

# Seismic Signal Denoising and Decomposition Using Deep Neural Networks

Weiqiang Zhu<sup>ID</sup>, S. Mostafa Mousavi, and Gregory C. Beroza<sup>ID</sup>

**Abstract**—Frequency filtering is widely used in routine processing of seismic data to improve the signal-to-noise ratio (SNR) of recorded signals and by doing so to improve subsequent analyses. In this paper, we develop a new denoising/decomposition method, DeepDenoiser, based on a deep neural network. This network is able to simultaneously learn a sparse representation of data in the time–frequency domain and a non-linear function that maps this representation into masks that decompose input data into a signal of interest and noise (defined as any non-seismic signal). We show that DeepDenoiser achieves impressive denoising of seismic signals even when the signal and noise share a common frequency band. Because the noise statistics are automatically learned from data and require no assumptions, our method properly handles white noise, a variety of colored noise, and non-earthquake signals. DeepDenoiser can significantly improve the SNR with minimal changes in the waveform shape of interest, even in the presence of high noise levels. We demonstrate the effect of our method on improving earthquake detection. There are clear applications of DeepDenoiser to seismic imaging, micro-seismic monitoring, and preprocessing of ambient noise data. We also note that the potential applications of our approach are not limited to these applications or even to earthquake data and that our approach can be adapted to diverse signals and applications in other settings.

**Index Terms**—Convolutional neural networks, decomposition, deep learning, seismic denoising.

## I. INTRODUCTION

RECORDED seismic signals are inevitably contaminated by noise and non-seismic signals from various sources, including ocean waves, wind, traffic, instrumental noise, electrical noise, and so on. Spectral filtering (usually based on the Fourier transform) is frequently used to suppress noise in routine seismic data processing; however, this approach is not effective when noise and seismic signal occupy the same frequency range. Moreover, selecting optimal parameters for filtering is non-intuitive, typically varies with time, and may strongly alter the waveform shape, such that it degrades the analysis that follows.

Due to these limitations, numerous efforts have been made to develop more effective noise suppression in seismic

Manuscript received November 12, 2018; revised April 21, 2019; accepted June 26, 2019. Date of publication August 15, 2019; date of current version October 31, 2019. This work was supported by the National Science Foundation (NSF) under Grant EAR-1818579. (Corresponding author: Weiqiang Zhu.)

The authors are with the Department of Geophysics, Stanford University, Stanford, CA 94305 USA (e-mail: zhwq@stanford.edu).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2019.2926772

data [1]–[19]. Methods based on time–frequency denoising [20], [21] form a large class of seismic denoising techniques. In this approach, noisy time series is first transformed into the time–frequency domain using a time–frequency transform, such as the wavelet transform [22]–[27], the short-time Fourier transform (STFT) [28], the S-transform [29], the curvelet transform [30]–[32], the dreamlet transform [33], the contourlet transform [34], the shearlet transform [35], empirical mode decomposition [36], [37]–[40] and so on. The resulting time–frequency coefficients are modified (thresholded) to attenuate the coefficients associated with noise and to find an estimate of the signal coefficients. The modified coefficients are inverse transformed back into the time domain to reconstruct the denoised signal. The basic idea is to promote sparsity by transforming seismic data to other domains, where the signal can be represented by a sparse set of features so that signal and noise can be separated more easily.

These methods can suppress the noise even when it occupies the same frequency range as the signal; however, the choice of a suitable thresholding function to map the noisy data into optimally denoised signal can be challenging. Denoising performance of time–frequency methods can be improved in two ways: either by using a more effective sparse representation of the data or by using a more flexible and powerful mapping function. Machine learning techniques, such as dictionary learning, have been used to improve seismic denoising by learning better sparse representations [41]–[44]. The focus of this paper is on improving both sparsity and the mapping function using deep learning. Deep learning [45], [46] is a powerful machine learning technique that can learn extremely complex functions through neural networks. Deep learning has been shown to be a powerful tool for learning the characteristics of seismic data [47]–[54].

In this paper, we present DeepDenoiser, a novel time–frequency denoising method using deep neural networks. This network is able to simultaneously learn a sparse representation of the input data and a high-dimensional non-linear function that maps this representation into desired masks from the training data set. Given input data, DeepDenoiser produces two individual masks, one for the seismic signal and the other for the noise signal. The masks are further used to extract the corresponding waveforms from the input data. We use earthquake seismograms manipulated with various types of noise and non-earthquake signals recorded by seismic stations to train the network and demonstrate its performance. We apply the method to unseen noisy seismograms to illustrate

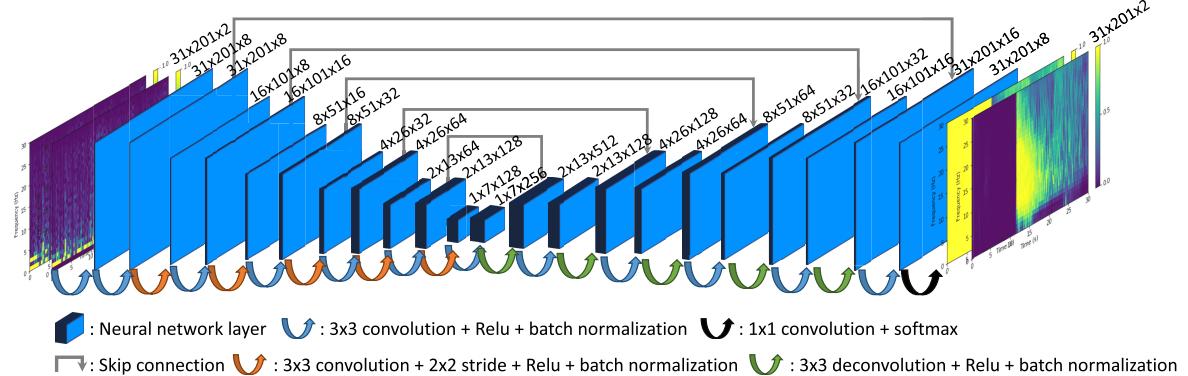


Fig. 1. Neural network architecture. Inputs are the real and imaginary parts of the time–frequency representation of noisy data. Outputs are two masks for signal and noise extraction. Blue rectangles represent layers inside the neural network. The dimension of each layer is presented above it and contains “frequency bins × time points × channels.” Arrows represent different operations applied to layers. The input data first go through  $3 \times 3$  convolution layers with  $2 \times 2$  strides for down-sampling and then go through deconvolution layers [57] for up-sampling. Batch normalization and skip connections are used to improve convergence during training. The softmax normalized exponential function is applied in the last layer to predict the masks for signal and noise.

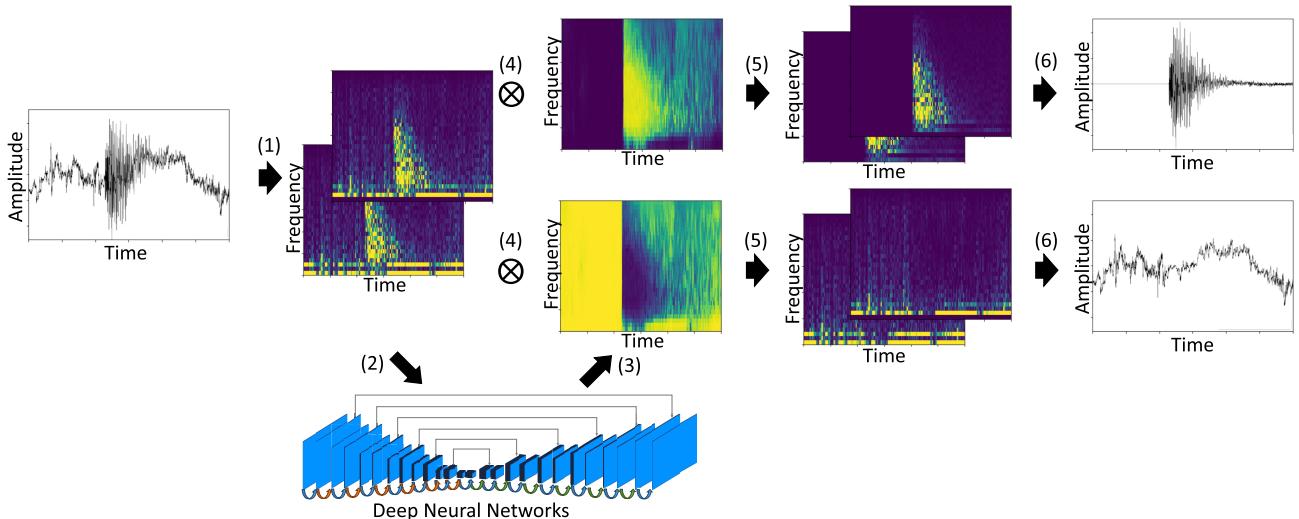


Fig. 2. Data flow diagram of our denoising method. (1) The noisy data are transformed into the time–frequency domain using the STFT. (2) The real and imaginary parts of time–frequency coefficients are fed into our deep neural network. (3) The neural network produces two masks for signal and noise based on input data. (4) and (5) The associated masks are applied to the noisy-signal coefficients to estimate the time–frequency coefficients of the seismic signal and noise. (6) The denoised signal and noise in the time domain are obtained using inverse STFT.

its generalizability, to compare its performance with other denoising methods, and to document its ability to improve earthquake detection results.

## II. METHOD

In the time–frequency domain, we represent recorded data,  $Y(t, f)$ , as the superposition of seismic signal,  $S(t, f)$ , and some additive natural/instrumental noise or non-seismic signals, collectively termed noise,  $N(t, f)$

$$Y(t, f) = S(t, f) + N(t, f). \quad (1)$$

The objective of denoising is to estimate the underlying seismic signal (i.e., the denoised signal),  $\hat{S}(t)$ , from its noise contaminated version that minimizes the expected error between the true and estimated signals

$$\text{error} = E\|\hat{S}(t) - S(t)\|_2^2 \quad (2)$$

where  $\hat{S}(t) = TFT^{-1}\{M(t, f)Y(t, f)\}$ ,  $TFT^{-1}$  is the inverse time–frequency transform,  $Y(t, f)$  is the

time–frequency representation of noisy data, and  $M(t, f)$  is a function that maps  $Y(t, f)$  to a time–frequency representation of the estimated signal. Donoho and Johnstone [20] showed that this mapping can be carried out through a simple thresholding in a sparse representation, where the thresholding value can be estimated from noise level assuming a Gaussian distribution.

Here, we cast the problem as a supervised learning problem, where a deep neural network will learn a sparse representation of the data to generate an optimal mapping function based on training samples of signal and noise data distribution. This is analogous to Wiener deconvolution as follows. We define our mapping functions as two individual masks,  $M_S(t, f)$  and  $M_N(t, f)$  for signal and noise, respectively

$$M_S(t, f) = \frac{1}{1 + \frac{|N(t, f)|}{|S(t, f)|}} \quad (3)$$

$$M_N(t, f) = \frac{\frac{|N(t, f)|}{|S(t, f)|}}{1 + \frac{|N(t, f)|}{|S(t, f)|}}. \quad (4)$$

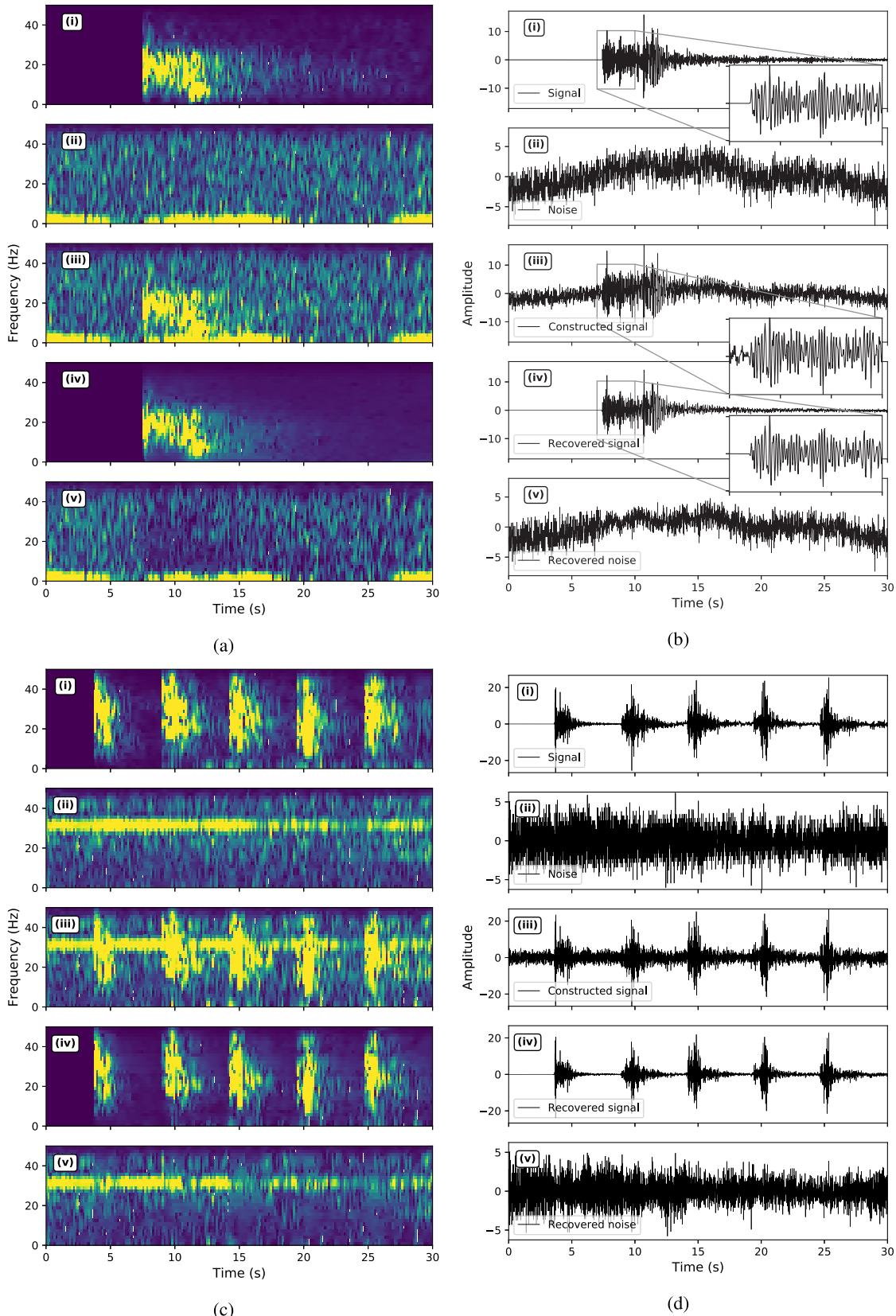


Fig. 3. Denoising examples. (a) and (c) Time–frequency domain. (b) and (d) Time domain. (i), (ii), and (v) ‘Clean’ signals, the real noise, and the constructed noisy signals, respectively. (iv) and (v) Recovered signals and noise by DeepDenoiser, respectively. The events in (a) and (b) are shifted to start from 7.5 s for plotting. Five events with increased shifting times are stacked together to construct the multi-event noisy samples in (c) and (d).

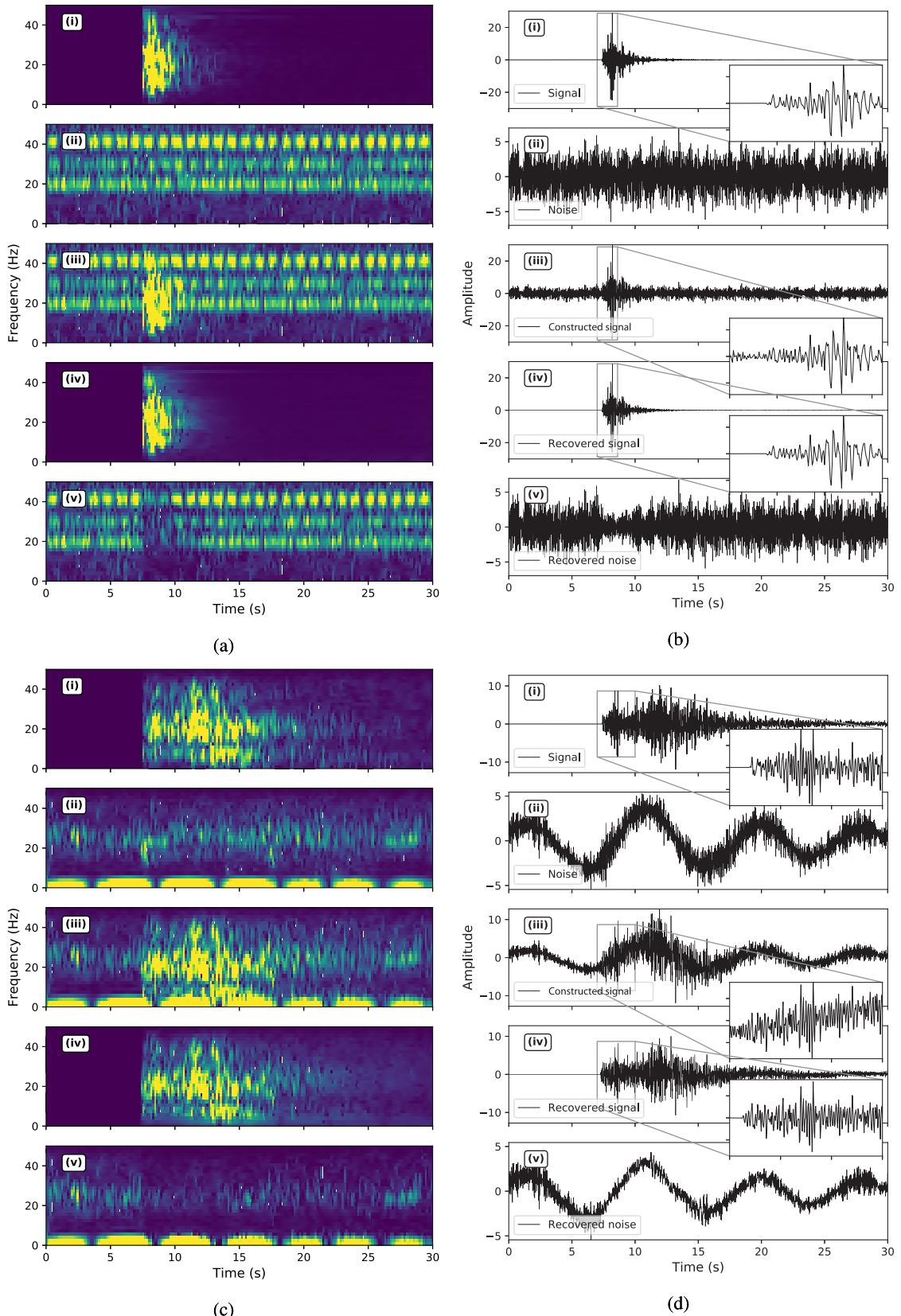


Fig. 4. Denoising performance in the presence of strong colored noise. (a) and (c) Time–frequency domain. (b) and (d) Time domain. (i), (ii), and (v) “Clean” signals, the real noise, and the constructed noisy signals, respectively. (iv) and (v) Recovered signals and noise by DeepDenoiser, respectively.

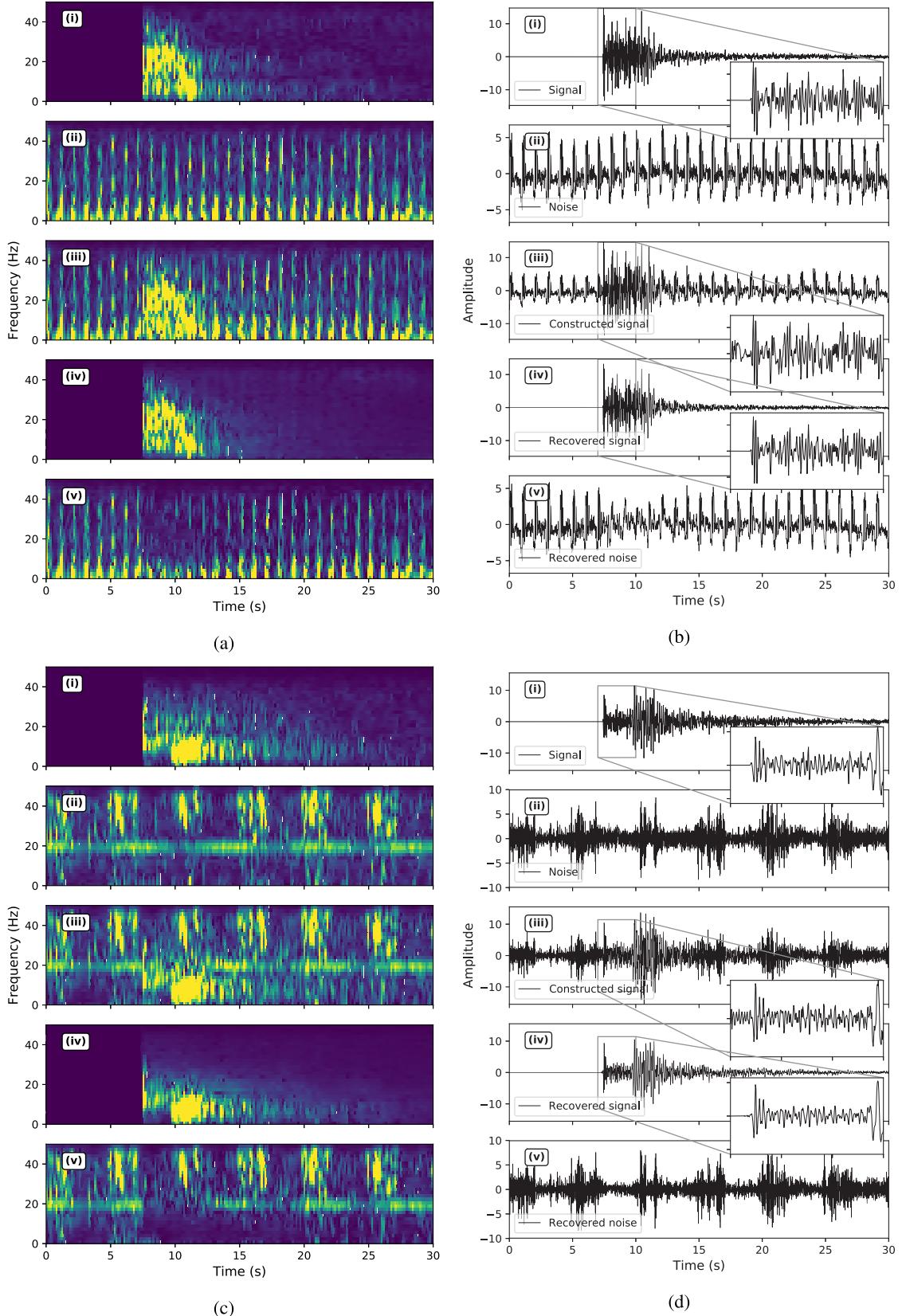


Fig. 5. Denoising performance in the presence of cyclic non-seismic signals. (a) and (c) Time–frequency domain. (b) and (d) Time domain. (i), (ii), and (v) “Clean” signals, the real noise, and the constructed noisy signals, respectively. (iv) and (v) Recovered signals and noise by DeepDenoiser, respectively.

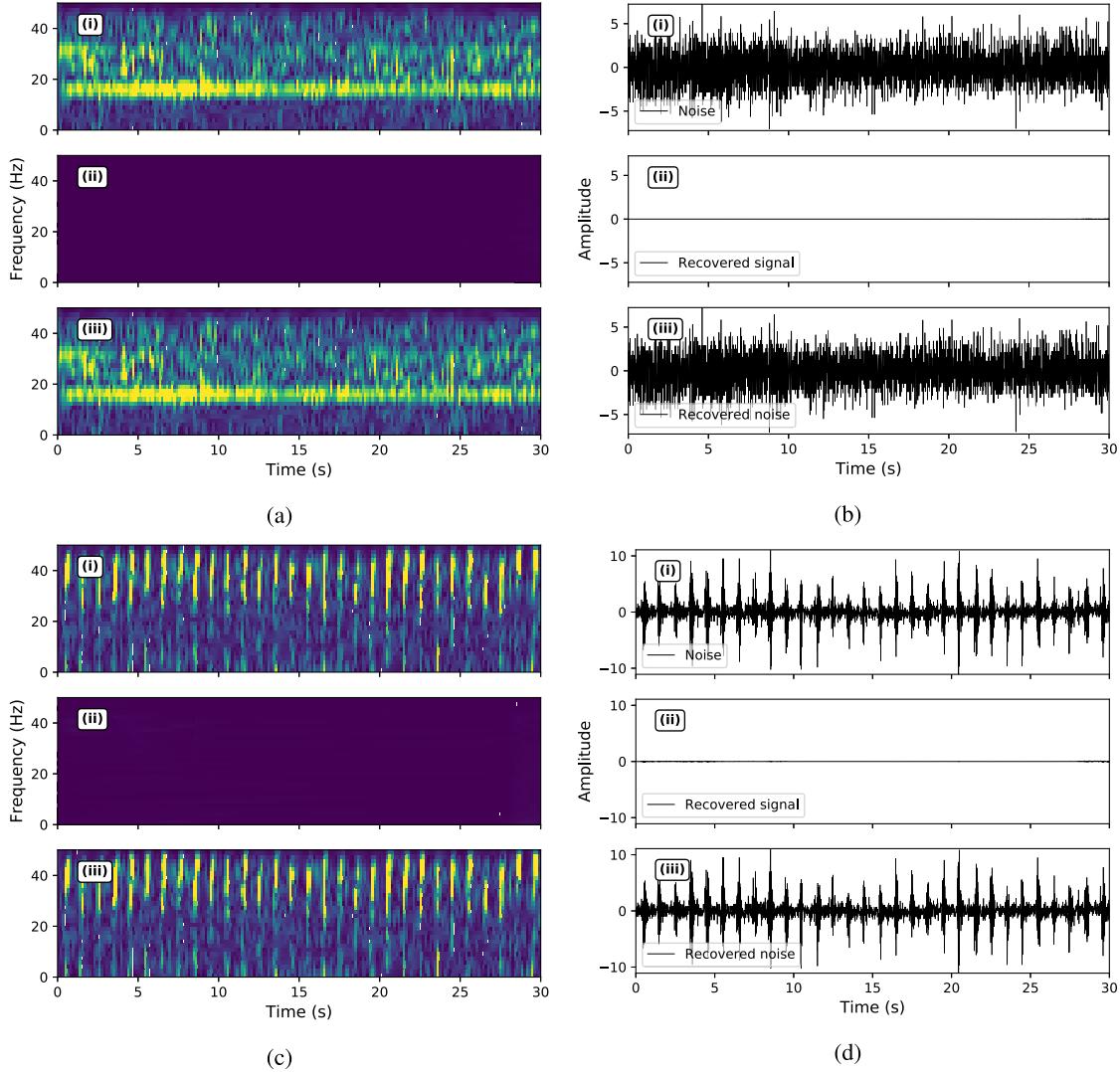


Fig. 6. Denoising examples of pure noise. (a) and (c) Time–frequency domain. (b) and (d) Time domain. (i) Input samples with pure noise. (ii) and (iii) Recovered signals and noise by DeepDenoiser, respectively. There is no artifact signals recovered by DeepDenoiser.

Each mask has the same size as the input time–frequency representation,  $Y(t, f)$ , and contains values between 0 and 1 that attenuate either noise or signal in time–frequency space.

Inspired by the capability of auto-encoders in learning a sparse representation of data with respect to an optimization objective, we designed our network in the form of a series of fully convolutional layers with descending and then ascending sizes (Fig. 1). Following Ronneberger *et al.* [55], we use skip connections to improve the convergence of training and prediction performance.

The inputs to the first layer are the imaginary and real parts of the time–frequency coefficients of the data,  $Y(t, f)$ . In the last layer, masks of signal and noise [ $M_S(t, f)$  and  $M_N(t, f)$ ] are provided as labels for training. The input time–frequency coefficients are processed and transformed through a series of 2-D convolutional layers with a rectified linear unit (ReLU) activation layer and batch normalization [56]. The convolution filter size is kept constant ( $3 \times 3$ ), but the feature space in the first half of the network is gradually reduced using strides of

$2 \times 2$ . These layers act as an effective feature extractor that accelerates learning of a very sparse representation of input data at the bottleneck layer. In the second half of the network, deconvolution (the transpose of the convolution layers) is used to generate a high-dimensional non-linear mapping of this sparse representation into output masks. In the last layer, a softmax normalized exponential function is used to produce masks. Through the training process, the network learns both the sparse representation of data and the optimal masks to separate signal from noise by optimizing a loss function (cross-entropy loss function).

Fig. 2 shows the data-flow diagram of our method. First, the seismic waveform is transferred into the time–frequency domain. The trained network takes the real and imaginary parts of time–frequency coefficients as the input and produces individual masks for both signal and noise as the outputs. The masks are the targets in optimizing the neural network during training. The estimated time–frequency coefficients of the seismic signal,  $\hat{S}(t, f)$ , and noise,  $\hat{N}(t, f)$ , are obtained

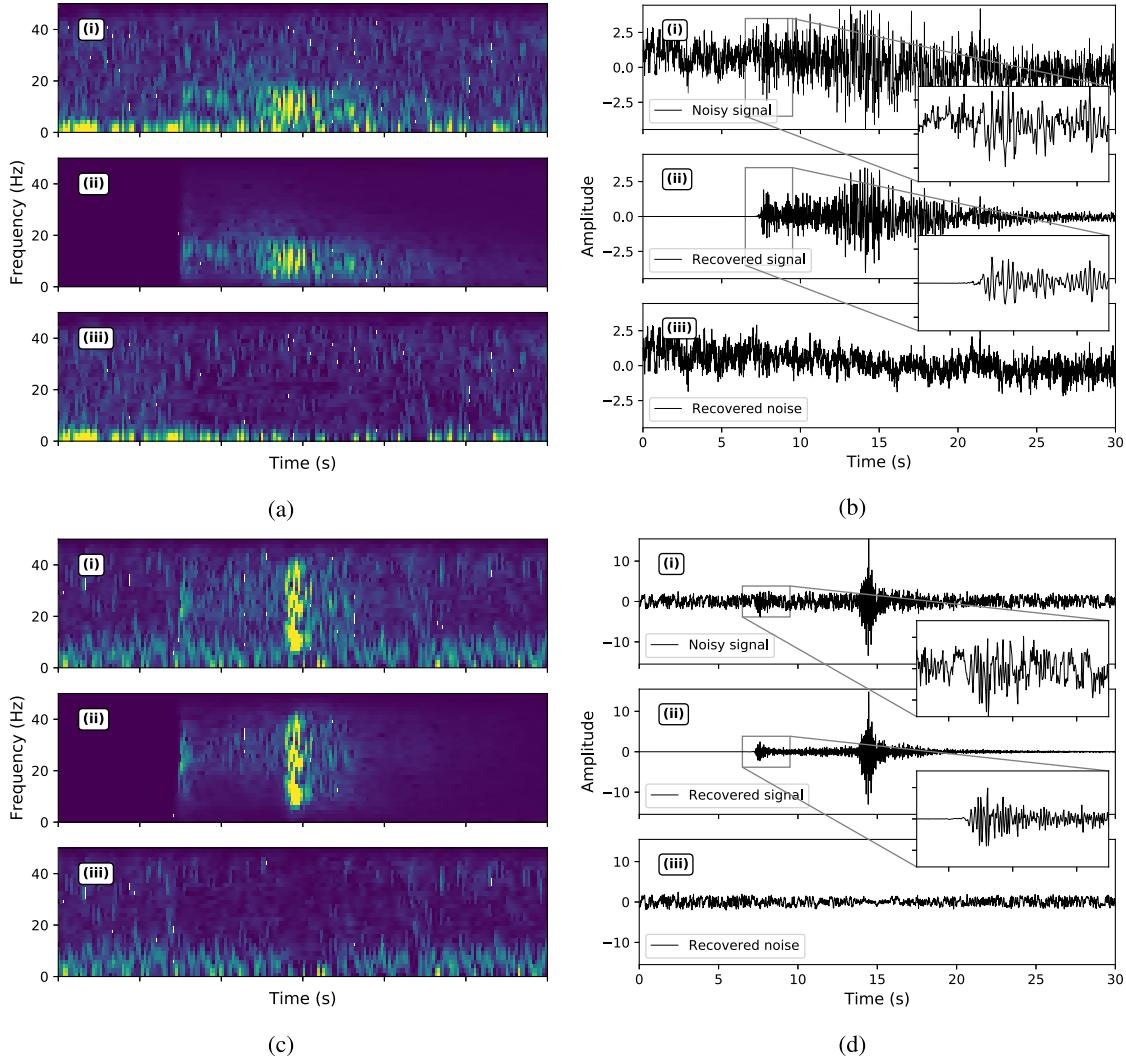


Fig. 7. Denoising performance on real noisy seismograms. (a) and (c) Time–frequency domain. (b) and (d) Time domain. (i) Real noisy seismograms. (ii) and (iii) Denoised signal and remaining noise using DeepDenoiser, respectively.

by applying the associated mask to the imaginary and real parts of the data coefficients,  $Y(t, f)$ . The final denoised signal and noise are obtained after inverse transforming  $\hat{S}(t, f)$  and  $\hat{N}(t, f)$  back into the time domain. Instead of defining different features and thresholds manually to enhance signal and attenuate noise, DeepDenoiser automatically learns richer features from semi-real seismic data that allow it to separate signal and noise in the time–frequency domain. Deep learning has the potential to provide a more effective and accurate automatic denoising tool for seismic data preprocessing, which can be applied to challenging tasks, such as micro-earthquake detection.

### III. NETWORK TRAINING

We use 30-s seismograms recorded by the high broadband channels (HN\*) of the North California Seismic Network to train the neural network and test its performance. The data set consists of 56345 earthquake waveforms with very high signal-to-noise ratios (SNRs) as the signal samples and 179233 seismograms associated with various types of

non-earthquake waveforms as the noise samples. Both the signal and noise data sets are randomly split into a signal training set with 45104 samples, a signal validation set with 5576 samples, a signal test sets with 5665 samples, a noise training set with 143403 samples, a noise validation set with 17890 samples, and a noise test set with 17940 samples. To construct “noisy” seismograms for training the neural network, we iterate through the signal training set repeatedly. In each iteration, we first randomly shift every signal sample and then randomly select one noise sample from the noise training set, randomly scale its amplitude, and stack it to the signal sample to generate training data of different SNR levels. This random combination between the signal and noise samples builds up a final training data set of over millions samples. The STFT is applied to produce the time–frequency representation of the noisy waveforms. The sampling frequency used in STFT is 100 Hz and the length of each segment is 0.3 s. We normalize the 2-D time–frequency matrices of the constructed noisy waveforms by removing the mean and dividing by the standard deviation. During prediction,

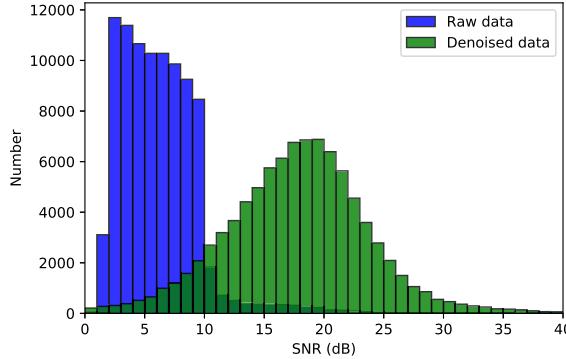


Fig. 8. Histogram showing SNR improvement of real noisy seismograms at Northern California.

the mean and standard deviation are recorded to recover the 2-D time–frequency matrices back to the original scale after denoising. The real and imaginary parts are fed to the neural network as two separate channels so that the network is able to learn from both the time and phase information. The prediction targets of the neural network are two masks: one for each signal and noise, constructed with (3) and (4). The same procedure is used to generate validation and test sets. The validation set is only used for tuning the hyper-parameters of the network, which helps to identify and prevent over-fitting. The test set is used to document performance. We train the deep neural network using the Adam optimizer [58] with a learning rate of 0.001 and a min-batch size of 200.

#### IV. RESULTS

##### A. Test Set

We use the test set to analyze the final performance and illustrate the denoising results. Fig. 3 shows that the network can successfully decompose the noisy inputs with different characteristics into denoised signal and noise. The algorithm can recover denoised signal with high accuracy [Fig. 3(b)(iv) and (d)(iv)]. As can be seen from the same example, signal leakage is minimal, and the waveform shape, frequency content, and amplitude characteristics are well preserved after denoising. These characteristics hold for the extracted noise as well [Fig. 3(b)(v) and (d)(v)].

One advantage of our learning-based denoising method is that the network not only learns the features of seismic signals but also the features of various noise sources. The wide variety of noise sources along with their time-varying signatures makes it very difficult for hand-engineered features to represent each type of noise effectively. However, DeepDenoiser shows the ability to learn a sparse feature representation for different types of noise. Some of them are recognized in our test data set. The first type of noise is band-limited [Fig. 4(a) and (b)], with relatively strong values within narrow frequency bands. The second type of noise is low-frequency noise [Fig. 4(c) and (d)], which exhibits strong background fluctuations that the signal rides on. The third type of noise is cyclic noise (Fig. 5), which is combined with different modes with the frequency band varying with time. The first two types of noise could be effectively attenuated by bandpass or low-pass filtering given an accurate estimate of the noise

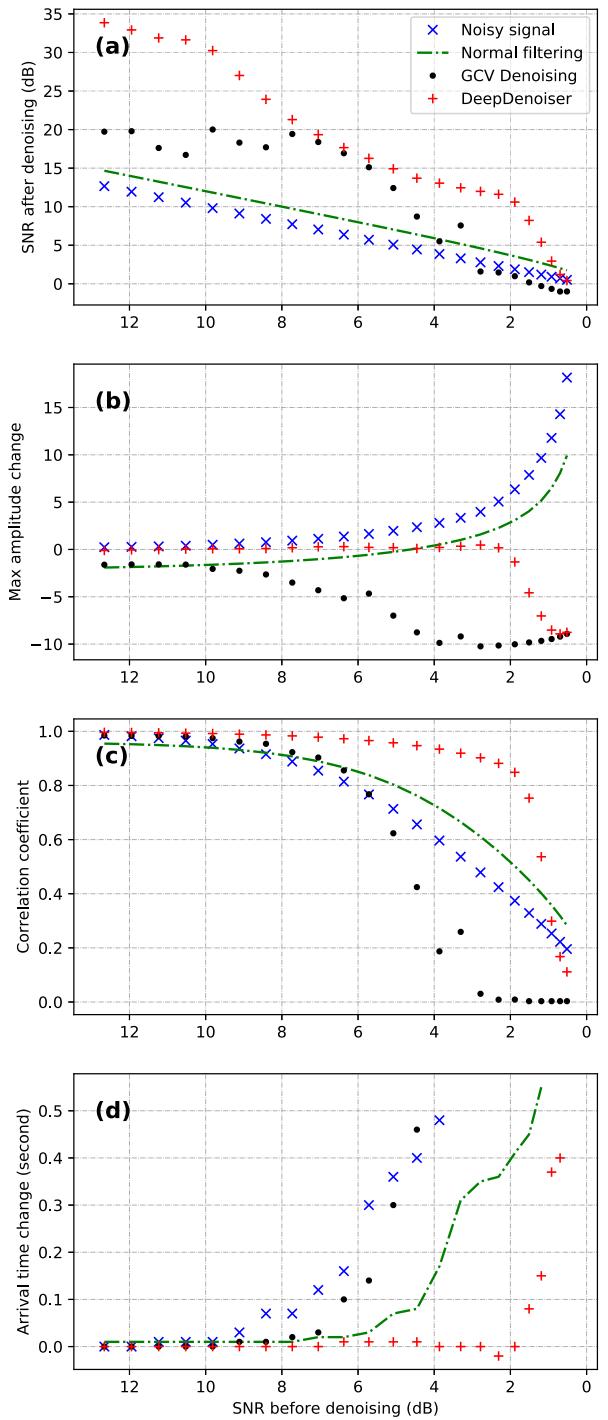


Fig. 9. Performance comparison between normal filtering, GCV denoising, and DeepDenoiser. (a) Improvement of SNR. (b) Max amplitude changes. (c) Correlation coefficient. (d) Picked arrival time changes. Values in (b) and (d) are differences compared with the signal in Fig. 12(a). Values in (c) are calculated by zero-lag cross correlation between the denoised waveforms and the signal in Fig. 12(a).

frequency bands. The third type of noise is challenging for traditional denoising methods because the noise changes with time and its frequency band overlaps with the frequency band of the target signal. For all these types of noise, DeepDenoiser automatically predicts a mask that adapts to the noise features. The mask not only estimates the noise frequency bands needed for denoising but also reflects the changes of frequency

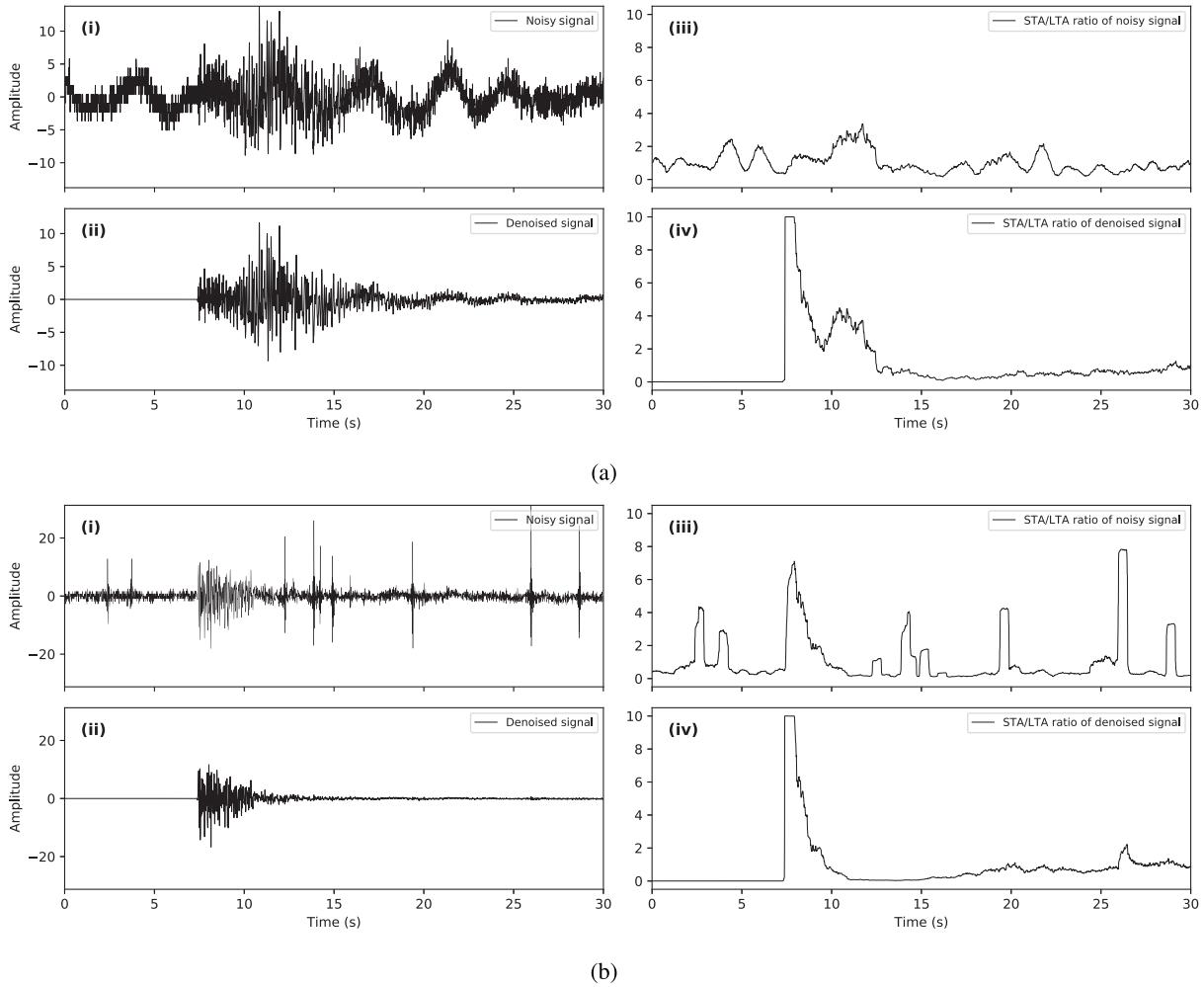


Fig. 10. Improvement of STA/LTA characteristic function after denoising. (i) Construed noisy signals. (ii) Denoised signals. (iii) and (iv) Corresponding STA/LTA characteristic functions. (a) and (b) two examples of earthquake detection.

content over time. The denoised signal and the separated noise [Figs. 4(b)(iv) and (v) and (d)(iv) and (v) and 5(b)(iv) and (v) and (d)(iv) and (v)] indicate that DeepDenoiser has robust denoising performance on various kinds of noise.

We also test DeepDenoiser on waveforms with pure noise. Fig. 6 shows that it accurately distinguishes these waveforms from those containing earthquake signals, i.e., no recovered signal is predicted and the recovered noise is equivalent to the input noise. With traditional denoising methods, the input noise waveform, on the other hand, could be contaminated. This important feature shows that DeepDenoiser has the ability to preserve noise signals with or without the presence of earthquakes at reasonable computational cost. It could have significant applications in data preprocessing for ambient seismic noise studies, where the contamination of noise data with earthquake signals can adversely affect the cross correlation results [59]. Current practice for dealing with earthquake contamination is to simply discard the associated windows [60], [61].

### B. Generalization

The neural network is trained on seismic data that are synthesized by superposing noise onto high-SNR seismic

signals. To test the network's generalizability, we applied it on the 91 000 samples of real noisy seismograms recorded in Northern California. These seismograms are from detected earthquakes in the Northern California Earthquake Catalog, but are contaminated heavily by noise [Fig. 7(i)] and, therefore, have low SNRs (Fig. 8).

The results suggest that DeepDenoiser successfully recovers clean seismic waveforms [Fig. 7(ii)], e.g., the first arrivals, and improves the SNR by around 15 dB (Fig. 8). The SNR is calculated as

$$\text{SNR} = 10 \log_{10} \left( \frac{\sigma_{\text{signal}}}{\sigma_{\text{noise}}} \right) \quad (5)$$

where  $\sigma_{\text{signal}}$  and  $\sigma_{\text{noise}}$  are the standard deviations of waveforms before and after the first arrival. Although DeepDenoiser is trained on synthetic data, it generalizes well to real seismograms. This suggests that we can directly apply the deep neural network trained in this paper to denoising tasks in real life whose performance relies on clean, undistorted seismic signals.

### C. Comparison With Other Methods

To compare the performance of our method with other denoising approaches, we select one independent seismic

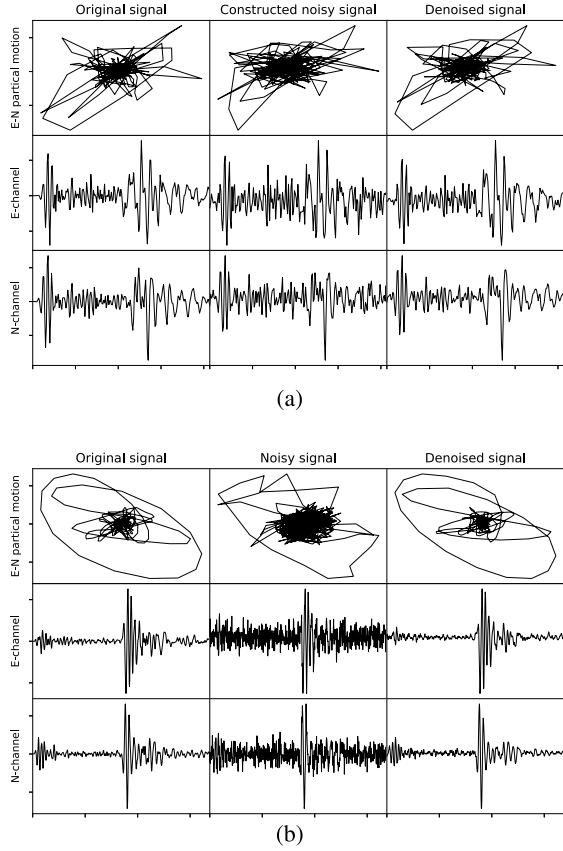


Fig. 11. Examples of particle motion between the east–west (E) and north–south (N) horizontal channels. (Left columns) Original high SNR signals. (Center columns) Constructed noisy signals. (Right column) Denoised signals by DeepDenoiser. First row: particle motions. Second and third rows: waveforms of E and N channels, respectively. (a) and (b) two examples of earthquake detection.

waveform of “EHZ” channel in obspy [62], which is recorded at station RTSH of the BayernNetz (BW) Network in Germany. We also cut a 30-s window prior to the event as the noise sample (Fig. 12). We scale the noise waveform to vary the noise level and stack it with the seismic waveform to generate a sequence of constructed noisy signals with decreased SNRs.

We used normal filtering and general cross-validation denoising (GCV) [27] for the comparison. For normal filtering, we design the filter based on the smoothed frequency distribution of the clean signal (Fig. 13), which is intended to emphasize the frequency band where the true signal resides. We measure the SNR improvements between the denoised and construed noisy signals, changes of the maximum amplitude, the correlation coefficient, and the picked arrival times as the bases of the comparison between the performance of these methods (Fig. 9). The arrival time is picked using the short-term averaging/long-term averaging (STA/LTA) method [63] the same as in Section IV-D for earthquake detection.

Fig. 9 shows that DeepDenoiser achieves a better denoising performance while introducing smaller distortion to the signal waveform. The SNR improvement of DeepDenoiser is more significant and more robust than the GCV method [Fig. 9(a)]. Fig. 9(b) shows that DeepDenoiser recovers the true amplitude of the signal more accurately. The max amplitude of the denoised signal is closer to the true signal when the SNR is

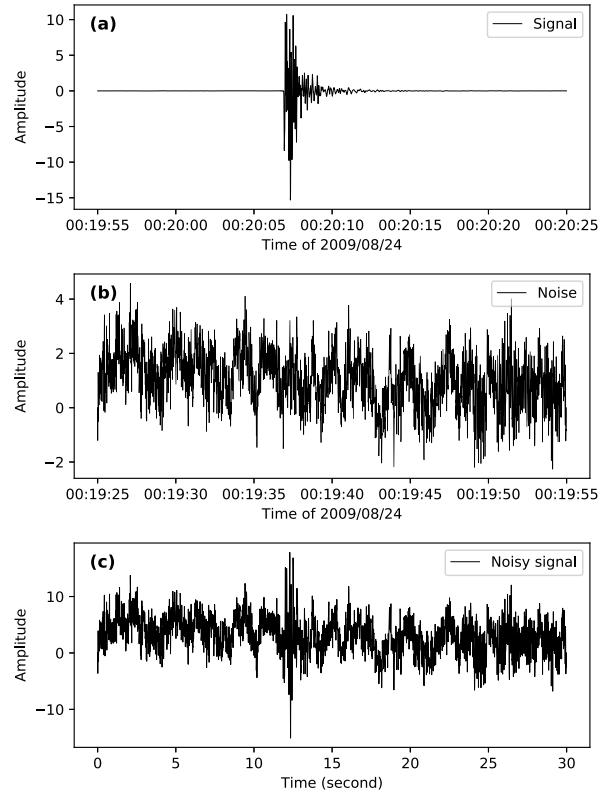


Fig. 12. Waveforms cut from “EHZ” channel of RTSH station, BayernNetz (BW) Network in Germany. (a) “Clean” signal. (b) Noise waveform obtained from the same station prior to the event origin time. (c) Example of generated noisy waveform with a stacking ratio of 3 for noise in (b) over signal in (a).

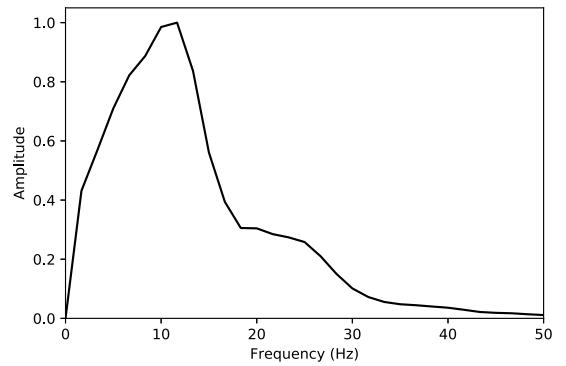


Fig. 13. Frequency distribution of the filter used for normal filtering in the comparison section.

larger than 2 dB, while the GCV method largely attenuates both noise and signal to achieve a good denoising performance. The higher correlation coefficients in Fig. 9(c) further demonstrate that DeepDenoiser introduces smaller waveform distortion during denoising. In contrast, the GCV method distorts the signal waveform significantly. DeepDenoiser is designed based on a way to separate signal and noise effectively, which enables it to improve SNR and preserve the signal waveform simultaneously. The denoised waveform also improves the recovery of arrival times from constructed noisy signal. With a fixed activation threshold for the STA/LTA method, the picked arrival times show a systematically increasing delay and eventual failure at increasing noise levels.

Fig. 9(d) shows that on the noisy signal, the arrival time has a delay of 0.5 s at SNR of 4 dB and fails to be recovered beyond this level (smaller dB). On the other hand, the signal after running DeepDenoiser yields an accurate arrival time even at the SNR of 2 dB, which is also superior to the other two denoising methods.

#### D. Application to Earthquake Detection

Background noise can significantly affect the performance of common detection algorithms such as STA/LTA (short-term average/long-term average) for detecting small and weak events. Moreover, the presence of non-earthquake signals will increase the false positive rate and, as a result, degrade detection precision. In Fig. 10, we present two examples of how DeepDenoiser can improve the STA/LTA characteristic function. The short and long time windows are set to be 0.5 and 5 s. Fig. 10(a) shows an example when the noise smooths out the sharp jump at the arrival of seismic waves and makes the earthquake undetectable. DeepDenoiser makes this earthquake easy to detect and increases the recall rate by making the STA/LTA characteristic function sharp after denoising. Fig. 10(b) shows another example when impulsive non-seismic signals bring sharp peaks to the STA/LTA characteristic function. These peaks will be falsely detected as earthquakes by the STA/LTA method. DeepDenoiser can effectively remove these non-seismic signals and increase the prediction precision. We compare the earthquake detection results on our test data set of 10 800 samples before and after denoising. With a threshold of 5, we calculate the precision of earthquake detection to be 35.14% and 92.34% and the recall rates to be 17.69% and 93.76% for the noisy signal and denoised signal, respectively. The accuracy is defined as the ratio of true positive detections over the total positive detections. The recall rate is defined as the ratio of true positives over the total true number of earthquakes. Both the accuracy and recall rate are improved significantly by DeepDenoiser.

We further investigate if DeepDenoiser would bias the timing information by measuring differential times via cross correlation, which determines the accuracy of earthquake locations [64]. We measure the cross correlation between the true signals and the denoised signals by DeepDenoiser. The cross correlation window is chosen to be 2.56 s centered on the P wave. We create three test data sets with low mean SNRs of 3.3, 1.4, and 0.5, where DeepDenoiser's performance starts to fail. The results show that 98.9%, 96.0%, and 85.8% of the samples have zero time-shifts. The standard deviations of time-shifts are 0.09, 0.14, and 0.29 s. Although expressing it using the standard deviation, the resulting distribution is not Gaussian and the values are strongly influenced by outliers with bad waveforms. This demonstrates that DeepDenoiser would not introduce systematic bias to the time information of seismic events.

#### V. CONCLUSION AND DISCUSSION

We have developed a novel deep-learning-based denoising algorithm for seismic data, DeepDenoiser. This learning-based

approach can learn a collection of sparse features with the aim of signal and noise separation from samples of data. These features more accurately reflect the statistical characteristics of a signal of interest and can be used for effective denoising or decomposition of input waveforms. The neural network automatically determines the percentage of signal (mask) present at each data point in the time–frequency space. DeepDenoiser learns the sparse representation of data to predict two adaptive thresholding masks simultaneously by optimizing the loss function. The masks determined by the network effectively decompose the input data into a signal of interest and noise.

Our results show that this algorithm can achieve robust and effective performance in denoising of data even when the signal and noise share a common frequency band. The denoising ability of our method is not limited to random white noise but performs well for a variety of colored noise and non-earthquake signals as well. Our tests indicate that DeepDenoiser significantly improves the SNR with minimal change to the underlying signal of interest. DeepDenoiser preserves waveform shape more faithfully than other denoising methods, even in the presence of high noise levels. Because we apply the same masks to both the real and imaginary parts of the input noisy signal [Fig. 2, step (5)], DeepDenoiser does not introduce a phase shift during denoising. Fig. 11 shows the particle motions between the two horizontal channels when we apply DeepDenoiser to the three channels of seismograms separately. The results further demonstrate that DeepDenoiser does not introduce large bias to the amplitude, time, and phase information after denoising. Although the network is trained using constructed noisy signals, it generalizes to real noisy data sets outside of the training set. We have only demonstrated the capability of our method in improving event detection; however, the potential applications of our approach are widespread. DeepDenoiser can be adapted for various types of seismic and non-seismic signals and different applications. Seismic imaging, micro-seismic monitoring, test-ban treaty monitoring, and preprocessing of ambient noise data are other potential applications of our method.

Although the performance of DeepDenoiser is impressive, it does not result in a perfect separation of signal and noise. Perfect separation of signal and noise in the time–frequency domain requires recovering two complex spectra for signal and noise from the complex spectra of the noisy signal. DeepDenoiser uses the same mask for both the real and imaginary parts of the spectrum, and the mask that has a value between [0, 1] cannot recover signal values larger than the input noisy signal. Predicting the signal and noise values directly without the use of a mask is a potential future direction for improvement. For three-component seismographs or seismic arrays, extending DeepDenoiser to consider all components simultaneously during denoising is another possibility.

#### APPENDIX

The waveform of the signal and noise used for section “comparison with other methods” is shown in Fig. 12. One example of the constructed noisy signal is shown in Fig. 12(c). Frequency distribution of the filter used for normal filtering

is shown in Fig. 13. This is built based on the frequency distribution of clean signal in Fig. 12(a), so that it retains the frequency bandwidth of the clean signal.

#### ACKNOWLEDGMENT

The authors would like to thank L. S. Gee and S. Zuzlewski for their help on downloading and processing the catalog and waveform data from the Northern California Earthquake Data Center (NCEDC). They would also like to thank W. Ellsworth, K. Wang, Y. Sheng, K. S. Tai, and K. Rong for helpful discussions. Waveform data, metadata, or data products for this paper were accessed through NCEDC.

#### REFERENCES

- [1] R. Abma and J. Claerbout, "Lateral prediction for noise attenuation by t-x and f-x techniques," *Geophysics*, vol. 60, no. 6, pp. 1887–1896, 1995.
- [2] V. Oropeza and M. Sacchi, "Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis," *Geophysics*, vol. 76, no. 3, pp. V25–V32, 2011. doi: [10.1190/1.3552706](https://doi.org/10.1190/1.3552706).
- [3] D. Bonar and M. Sacchi, "Denoising seismic data using the nonlocal means algorithm," *Geophysics*, vol. 77, no. 1, pp. A5–A8, 2012. doi: [10.1190/geo2011-0235.1](https://doi.org/10.1190/geo2011-0235.1).
- [4] M. Naghizadeh, "Seismic data interpolation and denoising in the frequency-wavenumber domain," *Geophysics*, vol. 77, no. 2, pp. V71–V80, Mar. 2012. doi: [10.1190/geo2011-0172.1](https://doi.org/10.1190/geo2011-0172.1).
- [5] G. Liu, X. Chen, J. Du, and K. Wu, "Random noise attenuation using f-x regularized nonstationary autoregression," *Geophysics*, vol. 77, no. 2, pp. V61–V69, 2012.
- [6] Y. Tian, Y. Li, and B. Yang, "Variable-eccentricity hyperbolic-trace TFPF for seismic random noise attenuation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6449–6458, Oct. 2014.
- [7] L. Zhu, E. Liu, and J. H. McClellan, "Seismic data denoising through multiscale and sparsity-promoting dictionary learning," *Geophysics*, vol. 80, no. 6, pp. WD45–WD57, 2015.
- [8] Y. Chen and S. Fomel, "Random noise attenuation using local signal-and-noise orthogonalization," *Geophysics*, vol. 80, no. 6, pp. WD45–WD57, 2015.
- [9] Y. Liu, S. Fomel, and C. Liu, "Signal and noise separation in prestack seismic data using velocity-dependent seislet transform," *Geophysics*, vol. 80, no. 6, pp. WD117–WD128, 2015.
- [10] Y. Chen, D. Zhang, W. Huang, and W. Chen, "An open-source MATLAB code package for improved rank-reduction 3D seismic data denoising and reconstruction," *Comput. Geosci.*, vol. 95, pp. 59–66, Oct. 2016.
- [11] Y. Chen *et al.*, "Simultaneous denoising and reconstruction of 5-D seismic data via damped rank-reduction method," *Geophys. J. Int.*, vol. 206, no. 3, pp. 1695–1717, May 2016.
- [12] W. Huang, R. Wang, Y. Chen, H. Li, and S. Gan, "Damped multi-channel singular spectrum analysis for 3d random noise attenuation," *Geophysics*, vol. 81, no. 4, pp. V261–V270, 2016.
- [13] W. Huang, R. Wang, Y. Yuan, S. Gan, and Y. Chen, "Signal extraction using randomized-order multichannel singular spectrum analysis," *Geophysics*, vol. 82, no. 2, pp. V69–V84, 2016.
- [14] W. Huang, R. Wang, S. Zu, and Y. Chen, "Low-frequency noise attenuation in seismic and microseismic data using mathematical morphological filtering," *Geophys. J. Int.*, vol. 211, no. 3, pp. 1296–1318, Dec. 2017.
- [15] Y. Zhou, C. Shi, H. Chen, J. Xie, G. Wu, and Y. Chen, "Spike-like blending noise attenuation using structural low-rank decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1633–1637, Sep. 2017.
- [16] Y. Zhou, S. Li, D. Zhang, and Y. Chen, "Seismic noise attenuation using an online subspace tracking algorithm," *Geophys. J. Int.*, vol. 212, no. 2, pp. 1072–1097, 2018.
- [17] Y. Chen, Y. Zhou, W. Chen, S. Zu, W. Huang, and D. Zhang, "Empirical low-rank approximation for seismic noise attenuation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4696–4711, Aug. 2017.
- [18] W. Huang, R. Wang, and Y. Chen, "Regularized non-stationary morphological reconstruction algorithm for weak signal detection in microseismic monitoring: Methodology," *Geophys. J. Int.*, vol. 213, no. 2, pp. 1189–1211, May 2018.
- [19] Y. Chen, "Non-stationary least-squares complex decomposition for microseismic noise attenuation," *Geophys. J. Int.*, vol. 213, no. 3, pp. 1572–1585, 2018.
- [20] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [21] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1995.
- [22] C. Siyuan and C. Xiangpeng, "The second-generation wavelet transform and its application in denoising of seismic data," *Appl. Geophys.*, vol. 2, no. 2, pp. 70–74, Jun. 2005.
- [23] S. Gaci, "The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4558–4563, Aug. 2014.
- [24] W. Liu, S. Cao, and Y. Chen, "Seismic time-frequency analysis via empirical wavelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 28–32, Jan. 2016.
- [25] S. M. Mousavi and C. A. Langston, "Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding," *Bull. Seismol. Soc. Amer.*, vol. 106, no. 4, pp. 1380–1393, 2016.
- [26] S. M. Mousavi, C. A. Langston, and S. P. Horton, "Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform," *Geophysics*, vol. 81, no. 4, pp. V341–V355, 2016. doi: [10.1190/geo2015-0598.1](https://doi.org/10.1190/geo2015-0598.1).
- [27] S. M. Mousavi and C. A. Langston, "Automatic noise-removal/signal-removal based on general cross-validation thresholding in synchrosqueezed domain and its application on earthquake data," *Geophysics*, vol. 82, no. 4, pp. V211–V227, 2017. doi: [10.1190/geo2016-0433.1](https://doi.org/10.1190/geo2016-0433.1).
- [28] S. M. Mousavi and C. A. Langston, "Adaptive noise estimation and suppression for improving microseismic event detection," *J. Appl. Geophys.*, vol. 132, pp. 116–124, Sep. 2016. doi: [10.1016/j.jappgeo.2016.06.008](https://doi.org/10.1016/j.jappgeo.2016.06.008).
- [29] G.-A. Tselentis, N. Martakis, P. Paraskevopoulos, A. Lois, and E. Sokos, "Strategy for automated analysis of passive microseismic data based on S-transform, Otsu's thresholding, and higher order statistics," *Geophysics*, vol. 77, no. 6, pp. KS43–KS54, 2012.
- [30] G. Hennenfent and F. J. Herrmann, "Seismic denoising with nonuniformly sampled curvelets," *Comput. Sci. Eng.*, vol. 8, no. 3, p. 16, May 2006.
- [31] R. Neelamani, A. I. Baumstein, D. G. Gillard, M. T. Hadidi, and W. L. Soroka, "Coherent and random noise attenuation using the curvelet transform," *Lead. Edge*, vol. 27, no. 2, pp. 240–248, Feb. 2008.
- [32] G. Tang and J. Ma, "Application of total-variation-based curvelet shrinkage for three-dimensional seismic data denoising," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 103–107, Jan. 2011.
- [33] B. Wang, R.-S. Wu, X. Chen, and J. Li, "Simultaneous seismic data interpolation and denoising with a new adaptive method based on dreamlet transform," *Geophys. J. Int.*, vol. 201, no. 2, pp. 1182–1194, May 2015.
- [34] H. Shan, J. Ma, and H. Yang, "Comparisons of wavelets, contourlets and curvelets in seismic denoising," *J. Appl. Geophys.*, vol. 69, no. 2, pp. 103–115, Oct. 2009. doi: [10.1016/j.jappgeo.2009.08.002](https://doi.org/10.1016/j.jappgeo.2009.08.002).
- [35] C. Zhang and M. van der Baan, "Multicomponent microseismic data denoising by 3D shearlet transform," *Geophysics*, vol. 83, no. 3, pp. A45–A51, 2018.
- [36] Y. Liu, Y. Li, H. Lin, and H. Ma, "An amplitude-preserved time-frequency peak filtering based on empirical mode decomposition for seismic random noise reduction," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 896–900, May 2014.
- [37] Y. Chen and J. Ma, "Random noise attenuation by f-x empirical-mode decomposition predictive filtering," *Geophysics*, vol. 79, no. 3, pp. V81–V91, 2014.
- [38] M. Bekara and M. Van der Baan, "Random and coherent noise attenuation by empirical mode decomposition," *Geophysics*, vol. 74, no. 5, pp. V89–V98, 2009.
- [39] W. Chen, J. Xie, S. Zu, S. Gan, and Y. Chen, "Multiple-reflection noise attenuation using adaptive randomized-order empirical mode decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 18–22, Jan. 2017.
- [40] J. Han and M. van der Baan, "Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding," *Geophysics*, vol. 80, no. 6, pp. KS69–KS80, 2015. doi: [10.1190/geo2014-0423.1](https://doi.org/10.1190/geo2014-0423.1).
- [41] Y. Chen, J. Ma, and S. Fomel, "Double-sparsity dictionary for seismic noise attenuation," *Geophysics*, vol. 81, no. 2, pp. V103–V116, 2016.

- [42] Y. Chen, "Fast dictionary learning for noise attenuation of multidimensional seismic data," *Geophys. J. Int.*, vol. 209, no. 1, pp. 21–31, 2017.
- [43] M. A. N. Siahzar, S. Gholtashi, V. Abolghasemi, and Y. Chen, "Simultaneous denoising and interpolation of 2D seismic data using data-driven non-negative dictionary learning," *Signal Process.*, vol. 141, pp. 309–321, Dec. 2017.
- [44] L. Liu, J. Ma, and G. Plonka, "Sparse graph-regularized dictionary learning for suppressing random seismic noise," *Geophysics*, vol. 83, no. 3, pp. V215–V231, 2018.
- [45] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [46] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.
- [47] P. M. R. DeVries, F. Viégas, M. Wattenberg, and B. J. Meade, "Deep learning of aftershock patterns following large earthquakes," *Nature*, vol. 560, no. 7720, pp. 632–634, Aug. 2018.
- [48] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Sci. Adv.*, vol. 4, no. 2, 2018, Art. no. e1700578.
- [49] S. M. Mousavi, W. Zhu, Y. Sheng, and G. C. Beroza, "Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection," Oct. 2018, *arXiv:1810.01965*. [Online]. Available: <https://arxiv.org/abs/1810.01965>
- [50] Z. E. Ross, M.-A. Meier, and E. Hauksson, "*P* wave arrival picking and first-motion polarity determination with deep learning," *J. Geophys. Res., Solid Earth*, vol. 123, no. 6, pp. 5120–5129, 2018.
- [51] Z. E. Ross, M.-A. Meier, E. Hauksson, and T. H. Heaton, "Generalized seismic phase detection with deep learning," May 2018, *arXiv:1805.01075*. [Online]. Available: <https://arxiv.org/abs/1805.01075>
- [52] Z. E. Ross, Y. Yue, M.-A. Meier, E. Hauksson, and T. H. Heaton, "Phaselink: A deep learning approach to seismic phase association," Sep. 2018, *arXiv:1809.02880*. [Online]. Available: <https://arxiv.org/abs/1809.02880>
- [53] J. Zheng, J. Lu, S. Peng, and T. Jiang, "An automatic microseismic or acoustic emission arrival identification scheme with deep recurrent neural networks," *Geophys. J. Int.*, vol. 212, pp. 1389–1397, Feb. 2018, doi: [10.1093/gji/ggx487](https://doi.org/10.1093/gji/ggx487).
- [54] W. Zhu and G. C. Beroza, "Phasenet: A deep-neural-network-based seismic arrival time picking method," Mar. 2018, *arXiv:1803.03211*. [Online]. Available: <https://arxiv.org/abs/1803.03211>
- [55] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
- [56] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015, *arXiv:1502.03167*. [Online]. Available: <https://arxiv.org/abs/1502.03167>
- [57] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1520–1528.
- [58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [59] X. Liu, Y. Ben-Zion, and D. Zigone, "Frequency domain analysis of errors in cross-correlations of ambient seismic noise," *Geophys. J. Int.*, vol. 207, no. 3, pp. 1630–1652, Dec. 2016.
- [60] G. Bensen *et al.*, "Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements," *Geophys. J. Int.*, vol. 169, no. 3, pp. 1239–1260, 2007.
- [61] Y. Sheng, M. A. Denolle, and G. C. Beroza, "Multicomponent C3 green's functions for improved long-period ground-motion prediction," *Bull. Seismol. Soc. Amer.*, vol. 107, no. 6, pp. 2836–2845, 2017.
- [62] M. Beyreuther, R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann, "ObsPy: A Python toolbox for seismology," *Seismol. Res. Lett.*, vol. 81, no. 3, pp. 530–533, 2010.
- [63] R. V. Allen, "Automatic earthquake recognition and timing from single traces," *Bull. Seismol. Soc. Amer.*, vol. 68, no. 5, pp. 1521–1532, 1978.
- [64] D. P. Schaff, G. H. R. Bokelmann, W. L. Ellsworth, E. Zanzerkia, F. Waldhauser, and G. C. Beroza, "Optimizing correlation techniques for improved earthquake location," *Bull. Seismol. Soc. Amer.*, vol. 94, no. 2, pp. 705–721, 2004.



**Weiqiang Zhu** received the B.S. and M.S. degrees in geophysics from Peking University, Beijing, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. in geophysics with Stanford University, Stanford, CA, USA.

Since 2016, he has been a Research Assistant with the Geophysics Department, Stanford University. His research interests include the application of deep learning in seismology and physics-based computational simulations of earthquakes.

Mr. Zhu is a member of the American Geophysical Union (AGU), the Seismological Society of America (SSA), and the Southern California Earthquake Center (SCEC).



**S. Mostafa Mousavi** received the M.S. degree in risk engineering from the University of Tehran, Tehran, Iran, in 2010, and the M.S. and Ph.D. degrees in geophysics from The University of Memphis, Memphis, TN, USA, in 2017.

He is currently a Post-Doctoral Research Fellow with Stanford University, Stanford, CA, USA. He has authored 20 journals and three conference papers. His research interests include machine learning, signal processing, statistics, and observational seismology.

Dr. Mousavi is a fellow of the National Elite Foundation of Iran. He received the Jeannine X. Kasper Award from the American Association of Geographers and the SEP/SEG Award from ExxonMobil in 2014.



**Gregory C. Beroza** received the B.S. degree in earth science from the University of California at Santa Cruz, Santa Cruz, CA, USA, in 1983, and the Ph.D. degree in geophysics from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1989.

After a year as a Post-Doctoral Researcher at MIT, he joined the Stanford Geophysics Department in 1990, where he now holds the Wayne Loel Professorship. He has authored over 170 publications. His research interests include analyzing seismograms to understand how earthquakes work and to quantify the hazards they pose. He is particularly interested in earthquake source processes for shallow earthquakes, intermediate-depth earthquakes, induced earthquakes, and slow earthquakes.

Dr. Beroza has been a fellow of the American Geophysical Union since 2008. He received the Presidential Young Investigator Award from NSF in 1991, has been named to the Brinson, RIT, IRIS/SSA, and Lawson Lectureships, and the Beno Gutenberg Medal from the European Geosciences Union in 2014 for outstanding contributions to seismology. Since 2007, he has been the Deputy Director and then the Co-Director of the Southern California Earthquake Center.