

Seismic Signal Denoising and Decomposition Using Deep Neural Networks

Weiqiang Zhu, S. Mostafa Mousavi and Gregory C. Beroza
 Department of Geophysics, Stanford University

Abstract—Denoising and filtering are widely used in routine seismic-data-processing 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 learn simultaneously 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. Our method properly handles 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 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 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—seismic denoising, decomposition, deep learning, convolutional neural network

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, etc. 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 data e.g. [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. Methods based on time-frequency denoising [20, 21] form a large class of seismic denoising techniques. In this approach, noisy time series are first transformed into the time-frequency domain using a time-frequency transform, such as a wavelet transform [22, 23, 24, 25, 26, 27], Short Time Fourier Transform (STFT) [28], S-transform [29], curvelet transform [30, 31, 32], dreamlet transform [33], contourlet

transform [34], shearlet transform [35], empirical mode decomposition [36, 37, 38, 37, 39, 40], etc. 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, 42, 43, 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, 48, 49, 50, 51, 52, 53, 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 an input data, DeepDenoiser produces two individual masks, one for seismic signal and the other for 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 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), $\hat{S}(t, f)$, from its noise contaminated version that minimizes the expected error between the true and estimated signal:

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

where $\hat{S}(t, f) = TFT^{-1}\{M(t, f)Y(t, f)\}$, 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:

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

$$M_N(t, f) = \left[\frac{\frac{|N(t, f)|}{|S(t, f)|}}{1 + \frac{|N(t, f)|}{|S(t, f)|}} \right] \quad (4)$$

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 2D convolutional layers with a ReLU (rectified linear unit) activation layer and batch normalization [56]. The convolution filter size is kept constant (3×3), but the feature space in the first half of the network is gradually shrunk using strides of 2×2 . These layers act as an effective feature extractor that can accelerate learning of a very sparse representation of input data at the bottleneck layer. In the second half of the network, deconvolution (transpose convolution layers) are 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 how to construct a sparse representation of data and 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 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 allows 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-second seismograms recorded by the high broadband channels (HN*) of the North California Seismic Network to train the neural network and test its performance. The dataset consists of 56,345 earthquake waveforms with very high signal-to-noise ratios (SNRs) as the signal samples and 179,233 seismograms associated with various types of non-earthquake waveforms as the noise samples. Both the signal and noise datasets are randomly split into training, validation and test sets. To form 'noisy' seismograms for training the neural network, we iterate through the signal training set repeatedly and in each iteration we add a randomly selected noise sample from the noise training set to the selected seismic signal to generate data of different SNR levels. The Short Time Fourier Transform (STFT) is applied to produce the time-frequency representation of the noisy waveforms. The 2D time-frequency matrices of the noisy waveforms are the input to our deep neural network. 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 each for signal and noise, composed with equation (3) and (4). The same procedure is used to generate validation and test sets. The validation set is only used for fine-tuning the hyper-parameters of the network. This helps to identify and prevent over-fitting. The test set is used for the final results in this paper.

IV. RESULTS

A. Test Set

We use the test set to analyze the final performance and visualize the denoising results. Fig. 3 demonstrates that the

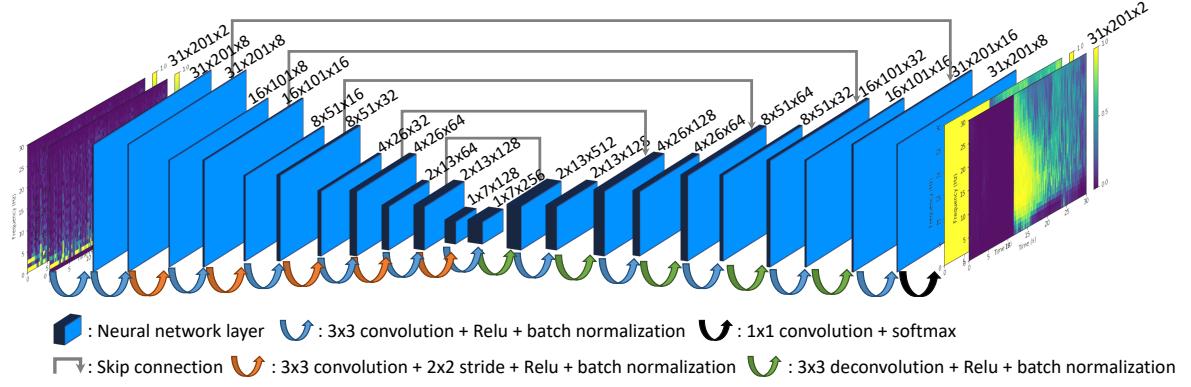


Fig. 1: Neural network architecture. Inputs are the real and imaginary parts of the time-frequency representation of noisy data. Outputs are two masks for signal and noise extraction. Blue rectangles represent layers inside the neural network. The dimension of each layer is presented above it, and contains "frequency bins \times time points \times channels". Arrows represent different operations applied to layers. The input data first go through 3×3 convolution layers with 2×2 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.

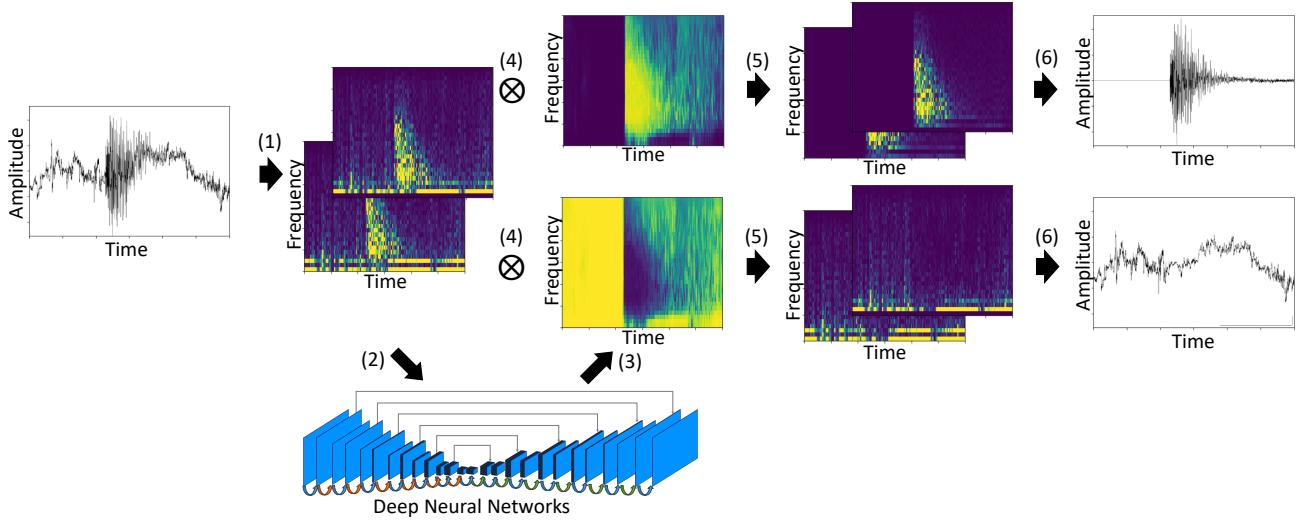


Fig. 2: Data flow diagram of our denoising method. (1) The noisy data is transformed into the time-frequency domain using Short Time Fourier Transform (STFT). (2) The real and imaginary parts of time-frequency coefficients are fed into our deep neural network. (3) The neural network produces two masks for signal and noise based on input data. (4, 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 time domain are obtained using inverse STFT.

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, 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, 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. The wide variety of noise sources along with their time-varying signatures makes it very difficult for hand-engineered features to well represent each type of noise. However, DeepDenoiser shows potential ability to learn a sparse feature representation for different types of noise. Some of them are recognized in our test dataset: The first type of noise is band-limited (Fig. 4(a, b)), with relatively strong values within narrow frequency bands; The second type of noise is low-frequency noise (Fig. 4(c, d))), which presents strong background fluctuations 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 band-pass or low-pass filtering given an accurate estimate of the noise frequency bands. The third type of noise is challenging for traditional denoising methods as the noise changes with time and its frequency band overlaps with the frequency band of the target signal. For all of 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 contents over time. The denoised signal and the separated noise (Fig. 4(b, d)(iv, v), Fig. 5(b, d)(iv, v)) support 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 renders DeepDenoiser with 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 heavily affect the cross-correlation results [58] and the current way of dealing with earthquake contamination is to simply discard the associated windows [59, 60].

B. Generalization

The neural network is trained on seismic data that is synthesized by superposing real noise signals onto the real high-SNR seismic signals. To test the network's generalizability, we applied it on the 91,000 samples of real 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 σ_{signal} and σ_{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 we can directly apply the deep neural network trained in this study 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 seismic waveform in "Trigger/Picker Tutorial"¹ of obspy [61] as a benchmark. The waveform is from the "EHZ" channel, which represents a short period high gain seismometer, recorded at station RTSH of the BayernNetz (BW) Network in Germany. We also cut a 30-second window prior to the event as the noise sample (Fig. 9). We scale the noise waveform to vary the noise level and stack it with the seismic waveform to generate a sequence of noisy signals with varying SNRs (Fig. 9(c)).

We used normal filtering and general cross-validation denoising [27] for the comparison. For normal filtering, we design the filter based on the smoothed frequency distribution of the clean signal (Fig. 12 in appendix), which is intended to emphasize the frequency band where the true signal resides. We measure the SNR improvements between the denoised and noisy signals, and 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. 10). The arrival time is picked using the same STA/LTA method in the section below.

Fig. 10 indicates 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. 10(a)). Fig. 10(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 in Fig. 9 when the SNR is 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. 10(c) further demonstrates the 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

¹https://docs.obspy.org/tutorial/code_snippets/trigger_tutorial.html

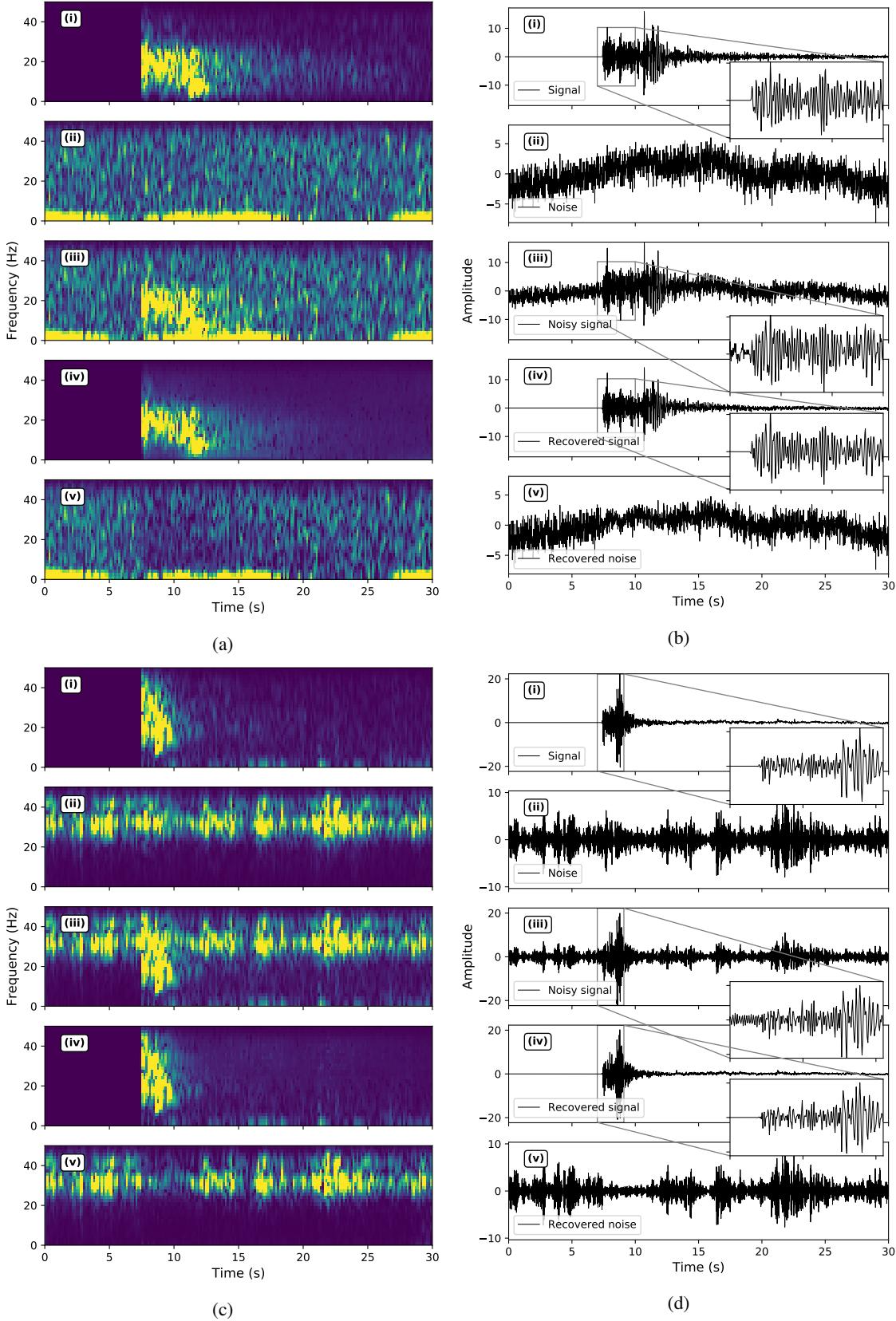


Fig. 3: Denoising examples: (a, c) time-frequency domain; (b, d) time domain. The time-frequency coefficients and waveforms of "clean" signal, real noise, and noisy signal are plotted in panels (i) (ii) (iii). Panels (a, c)(iv) show the denoised signal in the time-frequency domain, and panels (b, d)(iv) show its time domain waveform. The recovered noise is shown in panels (a, c)(v) and (b, d)(v).

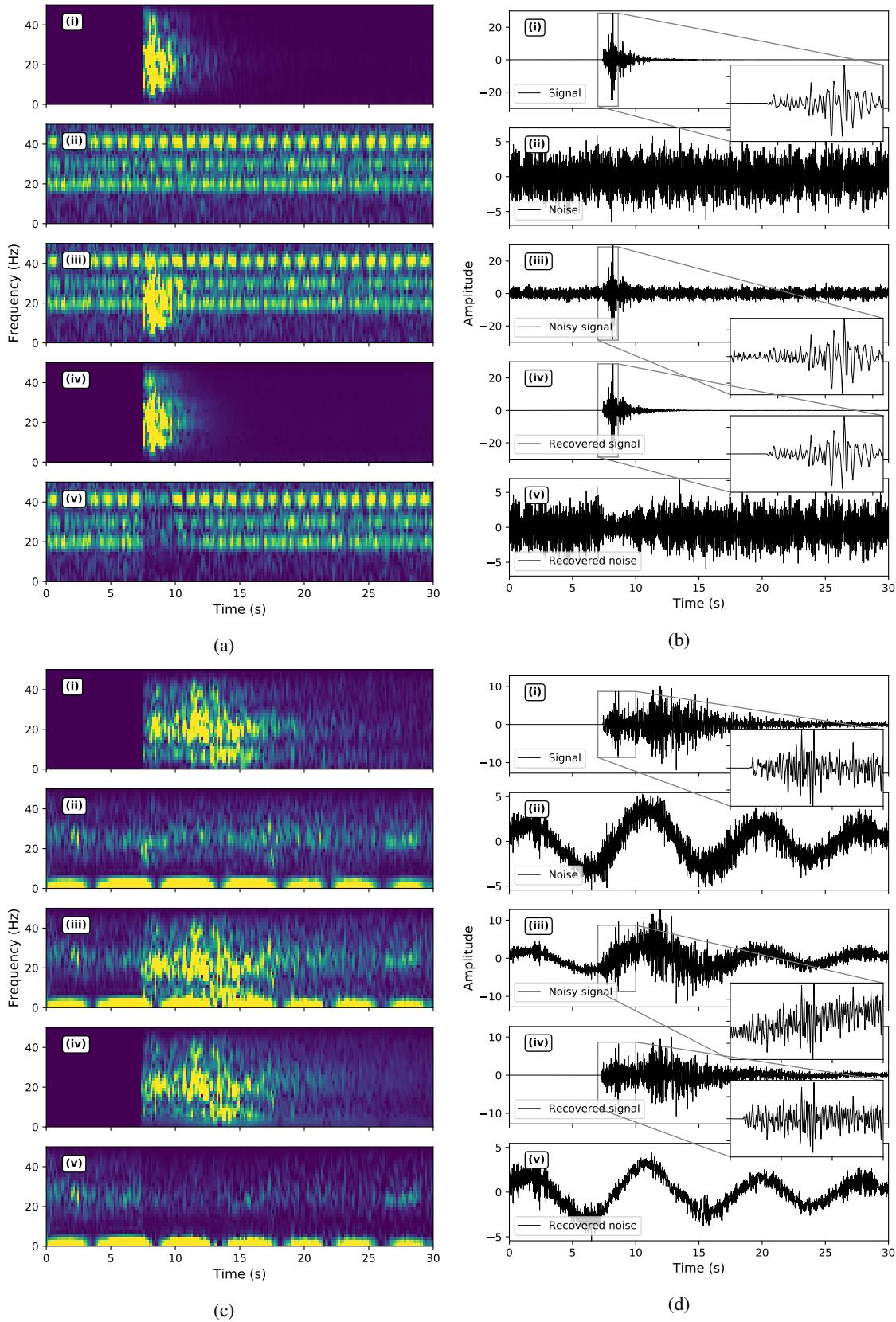


Fig. 4: Denoising performance in presence of strong colored noise. It is plotted in a same manner as Fig. 3.

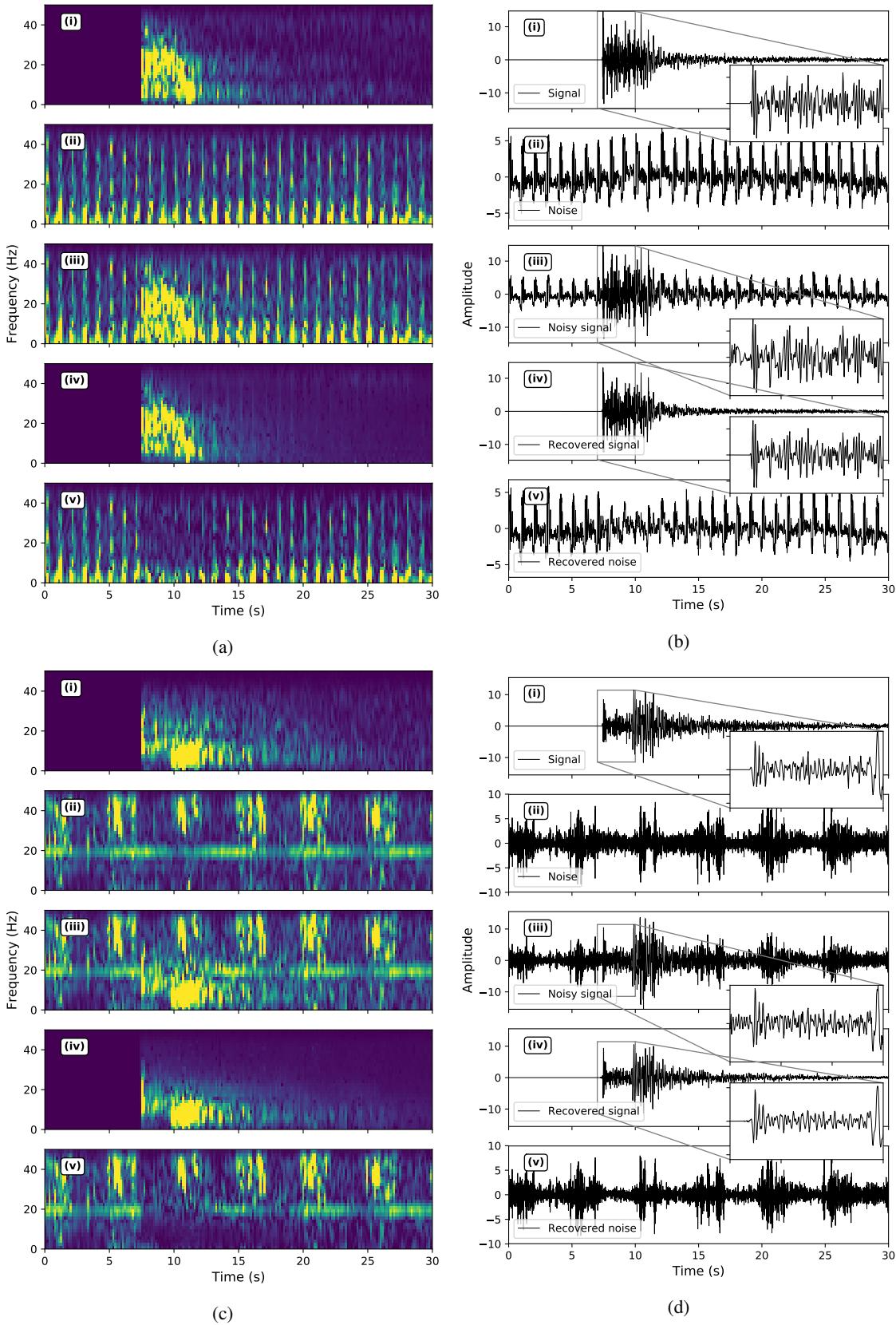


Fig. 5: Denoising performance in presence of cyclic non-seismic signals. It is plotted in a same manner as Fig. 3.

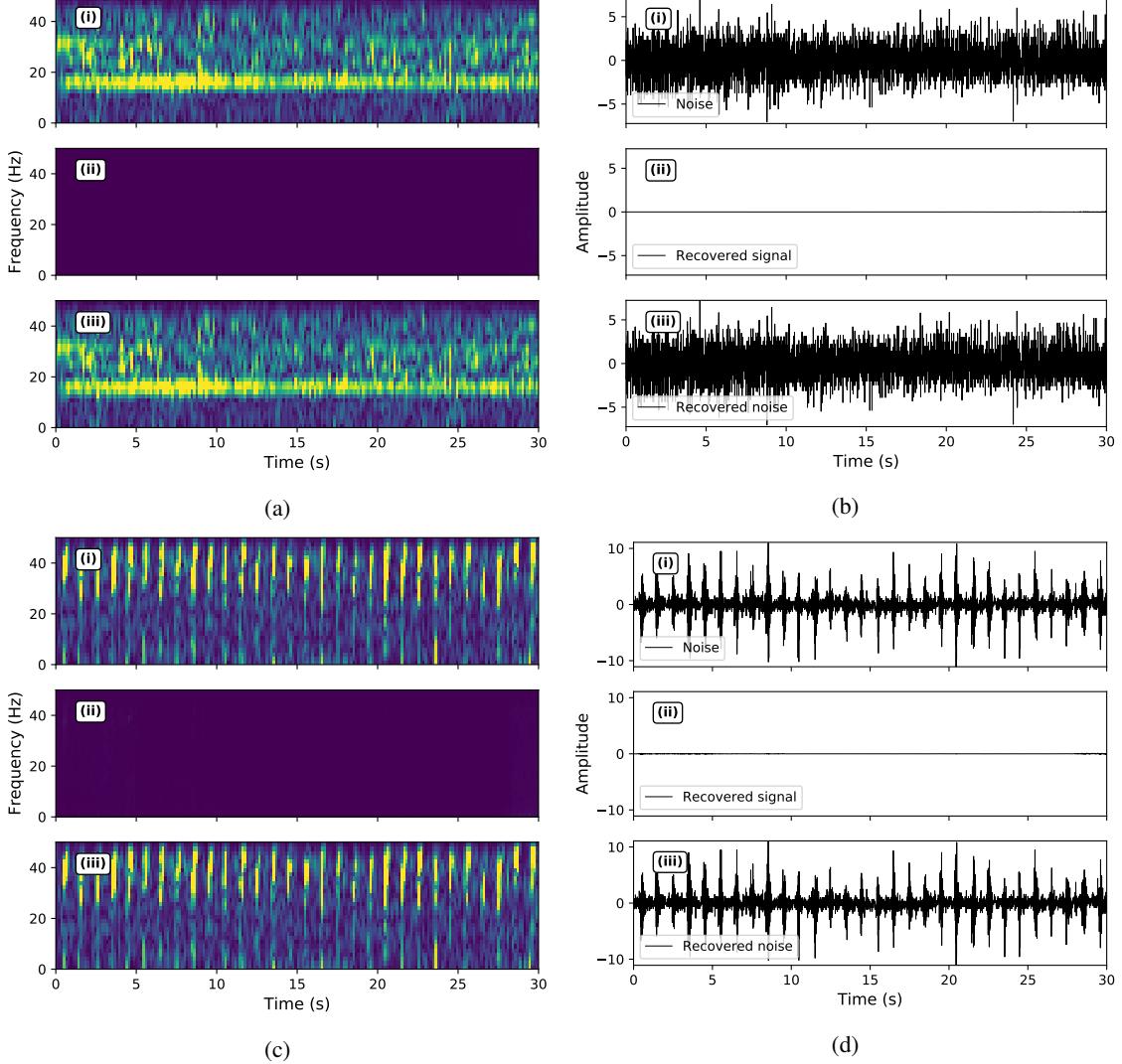


Fig. 6: Denoising examples of pure noise: (a, c) time-frequency domain; (b, d) time domain. The inputs are pure noise as shown in panels (i). The predicted signal masks in panels (a, c)(ii) and the recovered signal waveforms in panels (b, d)(ii) are mostly zeros. The predicted noise masks in panels (a, c)(iii) and recovered noise waveforms in panels (b, d)(iii) are same as the input noise in panels (i).

noisy signal. With a fixed activation threshold for STA/LTA method, the picked arrival times will have an systematically increasing delay and eventually not be recovered at increasing noise levels. Fig. 10(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 denoised signal by DeepDenoiser yields an accurate arrival time even at SNR of 2dB, which is also better than the other two denoising methods.

D. Application for Earthquake Detection

Background noise can significantly affect the performance of common detection algorithms such as STA/LTA (short-term averaging/long-term averaging) for detecting small and weak events [62]. Moreover, the presence of non-earthquake signals will increase the false positive rate, and as a result, degrade detection precision. In Fig. 11 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 seconds and 5 seconds. Fig. 11(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. 11(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 dataset of 10,800 samples before and after denoising. With a threshold of 5, we calculate the precision of earthquake detection to be 35.14%, 92.34% and the recall rates to be 17.69%, 93.76% for the noisy signal and denoised

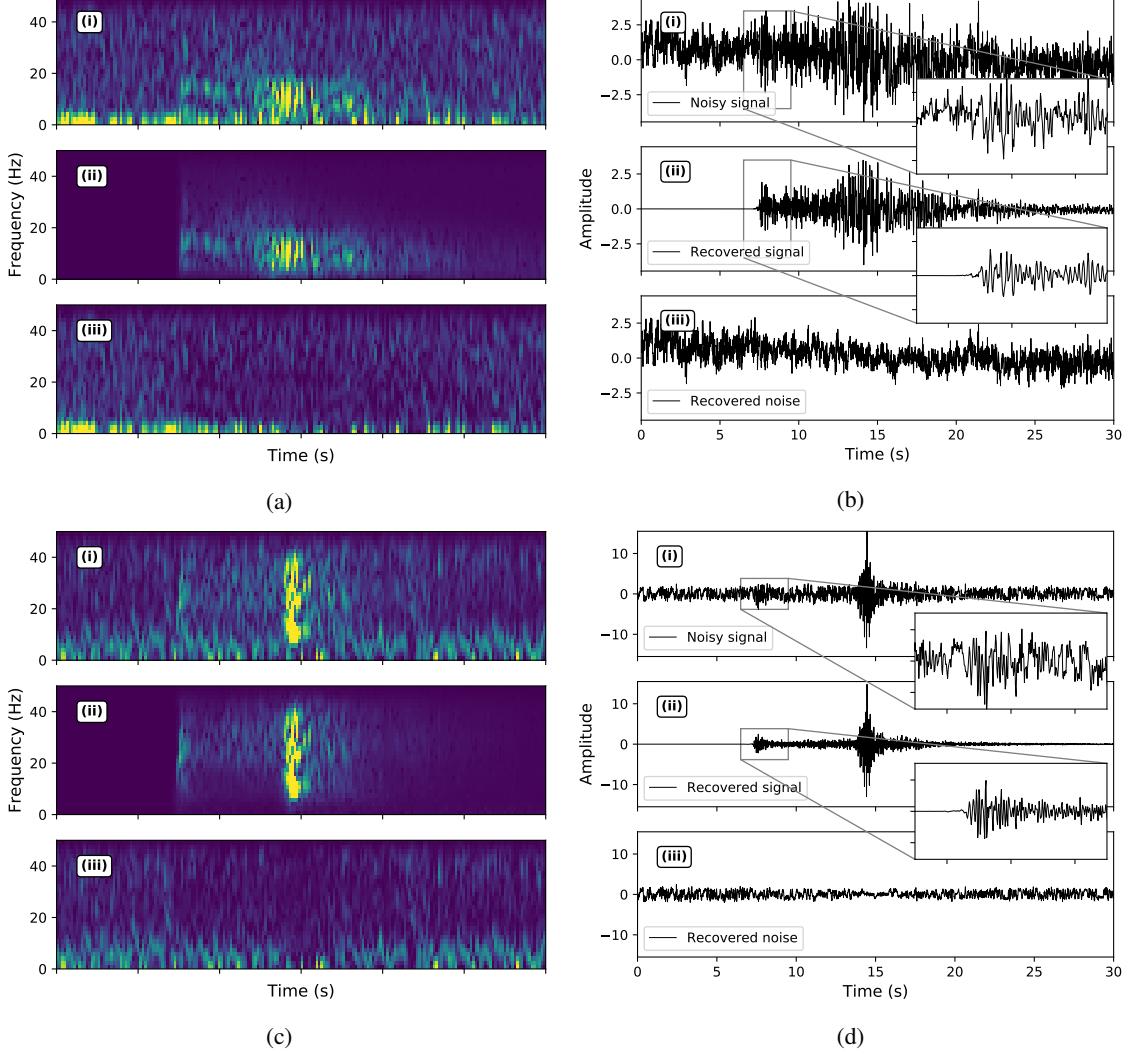


Fig. 7: Denoising performance on unseen seismograms: (a, c) time-frequency domain; (b, d) time domain. The time-frequency coefficients and waveforms of the unseen noisy signal are plotted in panels (i). Panels (a, c)(ii) show the denoised signal in the time-frequency domain, and panels (b, d)(ii) show its time domain waveform. The recovered noise is shown in panels (a, c)(iii) and (b, d)(iii).

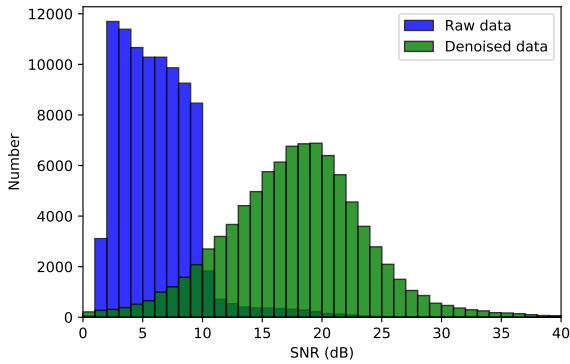


Fig. 8: Histogram showing SNR improvement of real noisy seismograms at Northern California.

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.

V. DISCUSSION AND CONCLUSION

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 reflect more accurately 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 present at each data point in the time-frequency space (mask). Learning a sparse representation of data and predicting two adaptive thresholding masks is performed simultaneously by optimizing

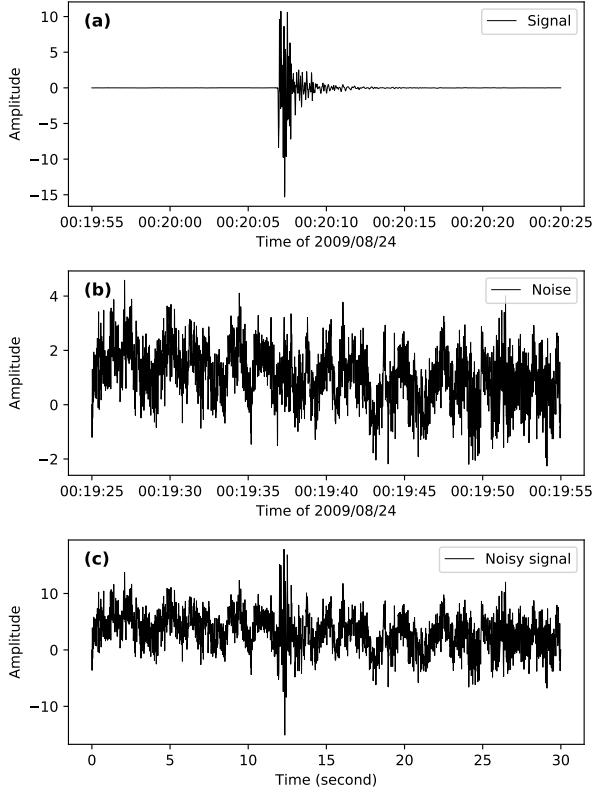


Fig. 9: The waveforms cut from "EHZ" channel of RTSH station, BayernNetz (BW) Network in Germany: (a) the "clean" signal; (b) the noise waveform obtained from the same station prior to the event origin time; (c) an example of generated noisy waveform with a stacking ratio of 3 for noise in (b) over signal in (a).

the loss function. The masks determined by the network are then used to effectively decompose the input data into a signal of interest and noise.

Our results show 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 can significantly improve the SNR with minimal change to the underlying signal of interest. DeepDenoiser preserves waveform shape more faithfully than other denoising methods, even in presence of high noise levels. The network appears to be able to generalize to datasets outside of 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.

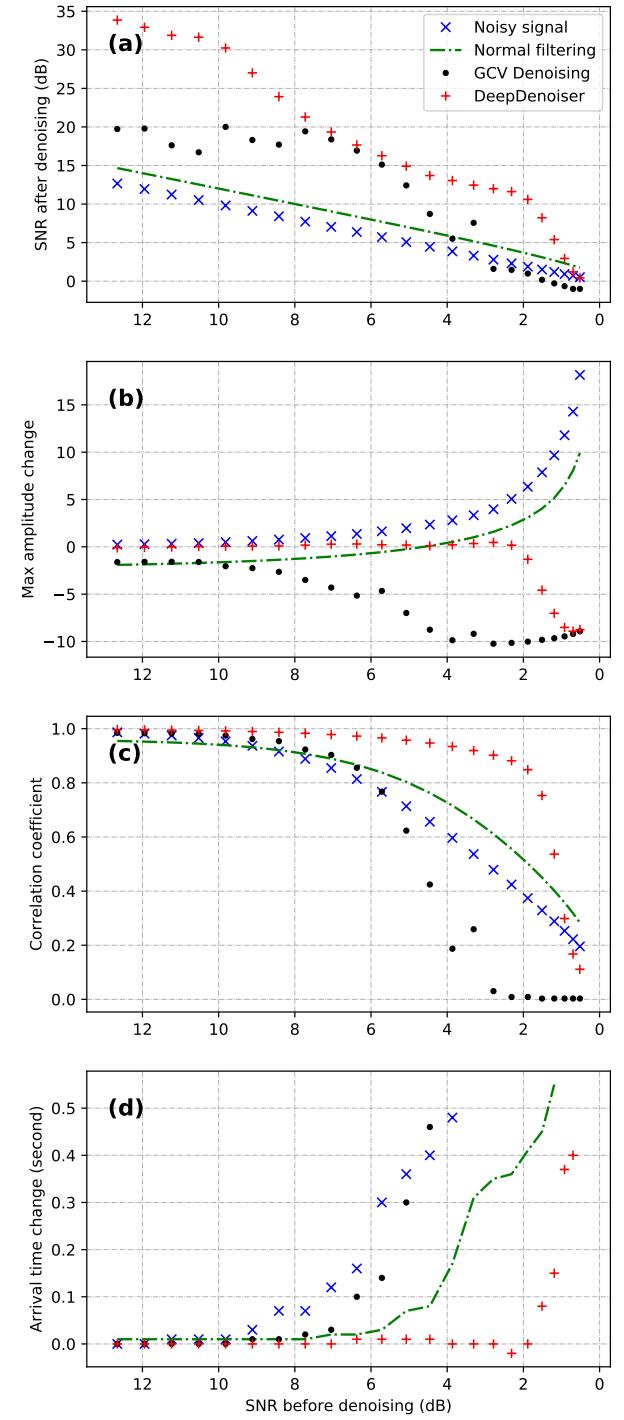


Fig. 10: Performance comparison between normal filtering, GCV denoising and DeepDenoiser: (a) improvement of SNR; (b) max amplitude changes; (c) correlation coefficient; (d) picked arrival time changes. Values in (b) and (d) are differences compared with the signal in Fig. 9(a). Values in (c) are calculated by zero-lag cross-correlation between the denoised waveforms and the signal in Fig. 9(a).

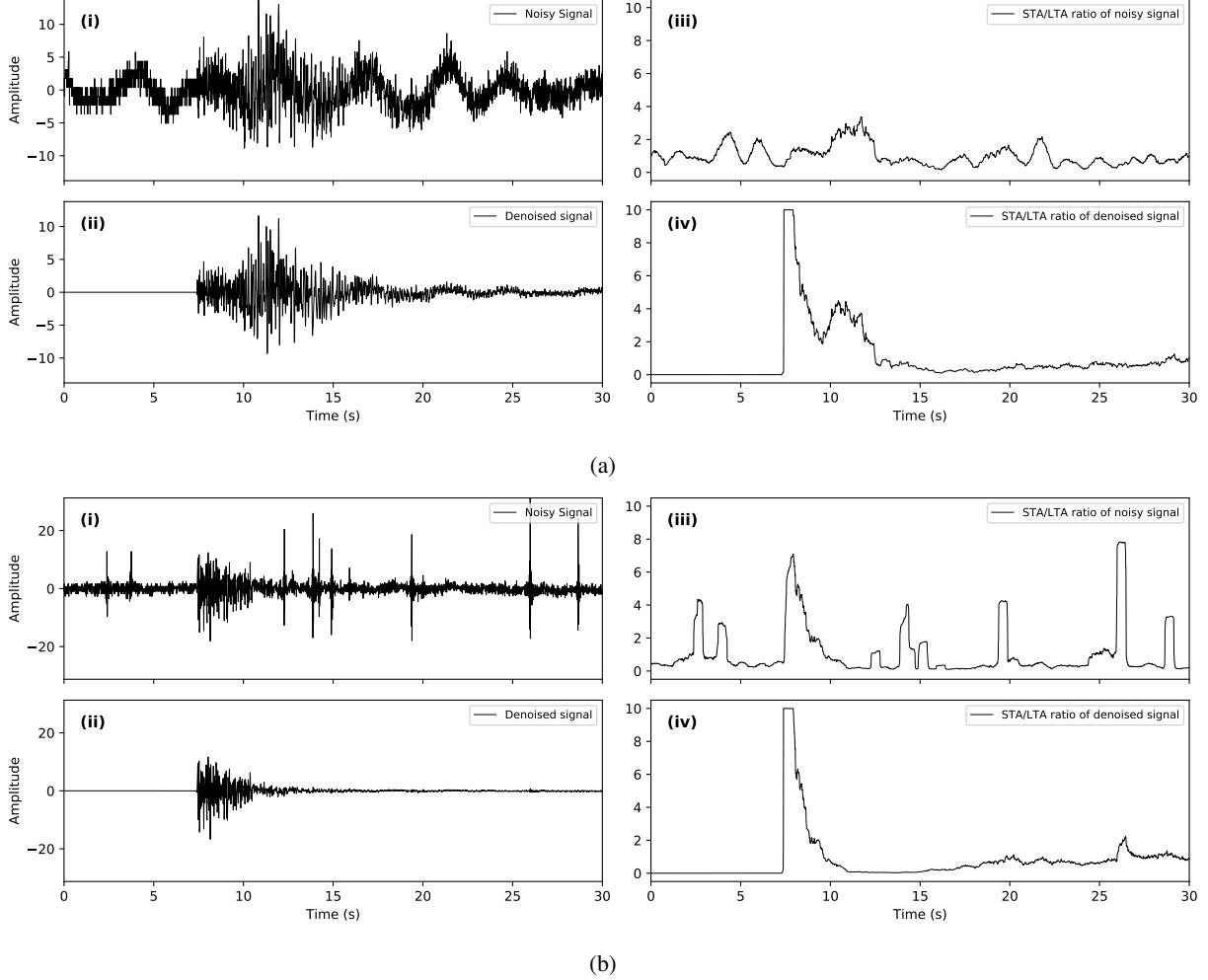


Fig. 11: Improvement of STA/LTA characteristic function after denoising: (i) noisy signals; (ii) denoised signals; (iii) and (iv) the corresponding STA/LTA characteristic functions.

Perfect separation of signal and noise in the time-frequency domain requires recovering two complex vectors for signal and noise from the complex vector of noisy signal, while DeepDenoiser uses the same mask for both the real and imaginary parts, and the mask, which has a value between [0, 1], can not recover signal values larger than the input noisy signal. Predicting the signal and noise values directly without the use of a mask may be a future direction for improvement.

ACKNOWLEDGMENTS

We thank Lind S. Gee and Stephane Zuzlewski for their help on downloading and processing the catalog and waveform data from NCEDEC. We thank William Ellsworth, Kaiwen Wang, Yixiao Sheng, Kai Sheng Tai and Kexin Rong for helpful discussions. Waveform data, metadata, or data products for this study were accessed through the Northern California Earthquake Data Center (NCEDEC). This research is supported by the National Science Foundation (NSF) grant number EAR-1818579.

APPENDIX A

Frequency distribution of the filter used for normal filtering in the comparison section are shown in Fig. 12. This is built based on the frequency distribution of clean signal in Fig. 9(a), so that it could keep the frequency band of the clean signal.

REFERENCES

- [1] R. Abma and J. Claerbout, "Lateral prediction for noise attenuation by tx and fx techniques," *Geophysics*, vol. 60, no. 6, pp. 1887–1896, 1995.
- [2] V. Orosez and M. Sacchi, "Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis," *Geophysics*, vol. 76, no. 3, pp. V25–V32, 2011. [Online]. Available: <http://library.seg.org/doi/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. [Online]. Available: <http://library.seg.org/doi/10.1190/geo2011-0235.1>
- [4] M. Naghizadeh, "Seismic data interpolation and denoising in the frequency-wavenumber domain," *Geophysics*,

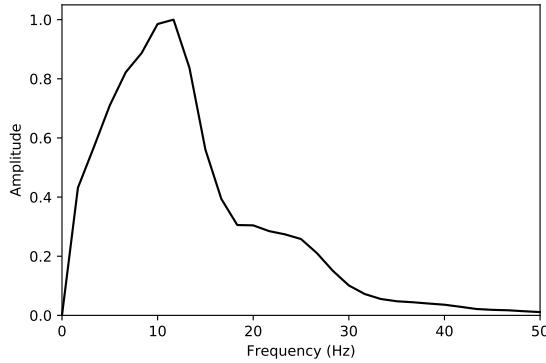


Fig. 12: Frequency distribution of the filter used for normal filtering in the comparison section

- vol. 77, no. 2, pp. V71–V80, 2012. [Online]. Available: <http://library.seg.org/doi/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. Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6449–6458, 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. WD1–WD9, 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, W. Huang, D. Zhang, and W. Chen, “An open-source matlab code package for improved rank-reduction 3d seismic data denoising and reconstruction,” *Computers & Geosciences*, vol. 95, pp. 59–66, 2016.
 - [11] Y. Chen, D. Zhang, Z. Jin, X. Chen, S. Zu, W. Huang, and S. Gan, “Simultaneous denoising and reconstruction of 5-d seismic data via damped rank-reduction method,” *Geophysical Journal International*, vol. 206, no. 3, pp. 1695–1717, 2016.
 - [12] W. Huang, R. Wang, Y. Chen, H. Li, and S. Gan, “Damped multichannel 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,” *Geophysical Journal International*, vol. 211, no. 3, pp. 1318–1340, 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 Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1633–1637, 2017.
 - [16] Y. Zhou, S. Li, D. Zhang, and Y. Chen, “Seismic noise attenuation using an online subspace tracking algorithm,” *Geophysical Journal International*, vol. 212, no. 2, pp. 1072–1097, 2017.
 - [17] Y. Chen, Y. Zhou, W. Chen, S. Zu, W. Huang, and D. Zhang, “Empirical low-rank approximation for seismic noise attenuation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4696–4711, 2017.
 - [18] W. Huang, R. Wang, and Y. Chen, “Regularized non-stationary morphological reconstruction algorithm for weak signal detection in microseismic monitoring: methodology,” *Geophysical Journal International*, vol. 213, no. 2, pp. 1189–1211, 2018.
 - [19] Y. Chen, “Non-stationary least-squares complex decomposition for microseismic noise attenuation,” *Geophysical Journal International*, 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,” *Journal of the american statistical association*, vol. 90, no. 432, pp. 1200–1224, 1995.
 - [22] S. Cao and X. Chen, “The second-generation wavelet transform and its application in denoising of seismic data,” *Applied geophysics*, vol. 2, no. 2, pp. 70–74, 2005.
 - [23] S. Gaci, “The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4558–4563, 2014.
 - [24] W. Liu, S. Cao, and Y. Chen, “Seismic time-frequency analysis via empirical wavelet transform.” *IEEE Geosci. Remote Sensing Lett.*, vol. 13, no. 1, pp. 28–32, 2016.
 - [25] S. M. Mousavi, C. A. Langston, S. Mostafa Mousavi, and C. A. Langston, “Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding,” *Bulletin of the Seismological Society of America*, 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. [Online]. Available: <http://library.seg.org/doi/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. [Online]. Available: <http://library.seg.org/doi/10.1190/geo2016-0433.1>

- [28] ——, “Adaptive noise estimation and suppression for improving microseismic event detection,” *Journal of Applied Geophysics*, vol. 132, pp. 116–124, 2016. [Online]. Available: <http://dx.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, otsus 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,” *Computing in Science & Engineering*, vol. 8, no. 3, pp. 16–25, 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,” *The Leading Edge*, vol. 27, no. 2, pp. 240–248, 2008.
- [32] G. Tang and J. Ma, “Application of total-variation-based curvelet shrinkage for three-dimensional seismic data denoising,” *IEEE geoscience and remote sensing letters*, vol. 8, no. 1, pp. 103–107, 2011.
- [33] B. Wang, R. Wu, X. Chen, and J. Li, “Simultaneous seismic data interpolation and denoising with a new adaptive method based on dreamlet transform,” *Geophysical Journal International*, vol. 201, no. 2, pp. 1180–1192, 2015.
- [34] H. Shan, J. Ma, and H. Yang, “Comparisons of wavelets, contourlets and curvelets in seismic denoising,” *Journal of Applied Geophysics*, vol. 69, no. 2, pp. 103–115, 2009. [Online]. Available: <http://dx.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 Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 896–900, 2014.
- [37] Y. Chen and J. Ma, “Random noise attenuation by fx 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 Sensing Lett.*, vol. 14, no. 1, pp. 18–22, 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. [Online]. Available: <http://library.seg.org/doi/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,” *Geophysical Journal International*, 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 Processing*, vol. 141, pp. 309–321, 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, p. 436, 2015.
- [46] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.
- [47] P. M. DeVries, F. Viégas, M. Wattenberg, and B. J. Meade, “Deep learning of aftershock patterns following large earthquakes,” *Nature*, vol. 560, no. 7720, p. 632, 2018.
- [48] T. Perol, M. Gharbi, and M. Denolle, “Convolutional neural network for earthquake detection and location,” *Science Advances*, vol. 4, no. 2, p. e1700578, 2018.
- [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,” *arXiv preprint arXiv:1810.01965*, 2018.
- [50] Z. E. Ross, M.-A. Meier, and E. Hauksson, “P-wave arrival picking and first-motion polarity determination with deep learning,” *Journal of Geophysical Research: Solid Earth*.
- [51] Z. E. Ross, M.-A. Meier, E. Hauksson, and T. H. Heaton, “Generalized seismic phase detection with deep learning,” *arXiv preprint arXiv:1805.01075*, 2018.
- [52] Z. E. Ross, Y. Yue, M.-A. Meier, E. Hauksson, and T. H. Heaton, “Phaselink: A deep learning approach to seismic phase association,” *arXiv preprint arXiv:1809.02880*, 2018.
- [53] J. Zheng, J. Lu, S. Peng, and T. Jiang, “An automatic microseismic or acoustic emission arrival identification scheme with deep recurrent neural networks,” *Geophysical Journal International*, vol. 212, no. 2, pp. 1389–1397, 2018. [Online]. Available: <http://dx.doi.org/10.1093/gji/ggx487>
- [54] W. Zhu and G. C. Beroza, “Phasenet: A deep-neural-network-based seismic arrival time picking method,” *arXiv preprint arXiv:1803.03211*, 2018.
- [55] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *Miccai*, pp. 234–241, 2015.
- [56] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” 2015. [Online]. Available: <http://arxiv.org/abs/1502.03167>
- [57] H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation,” *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015 Inter, pp. 1520–1528, 2015.
- [58] X. Liu, Y. Ben-Zion, and D. Zigone, “Frequency domain analysis of errors in cross-correlations of ambient seismic

- noise,” *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, vol. 207, no. 3, pp. 1630–1652, 2016.
- [59] G. Bensen, M. Ritzwoller, M. Barmin, A. Levshin, F. Lin, M. Moschetti, N. Shapiro, and Y. Yang, “Processing seismic ambient noise data to obtain reliable broadband surface wave dispersion measurements,” *Geophysical Journal International*, vol. 169, no. 3, pp. 1239–1260, 2007.
- [60] Y. Sheng, M. A. Denolle, and G. C. Beroza, “Multicomponent c3 greens functions for improved long-period ground-motion predictionmulticomponent c3 greens functions for improved long-period ground-motion prediction,” *Bulletin of the Seismological Society of America*, vol. 107, no. 6, pp. 2836–2845, 2017.
- [61] M. Beyreuther, R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann, “ObsPy: A Python toolbox for seismology,” *Seismological Research Letters*, vol. 81, no. 3, pp. 530–533, 2010.
- [62] M. Withers, R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo, “A comparison of select trigger algorithms for automated global seismic phase and event detection,” *Bulletin of the Seismological Society of America*, vol. 88, no. 1, pp. 95–106, 1998.