GEOPHYSICS

Toward improved urban earthquake monitoring through deep-learning-based noise suppression

Lei Yang[1,2], Xin Liu[1,3], Weiqiang Zhu[1], Liang Zhao[2]*, Gregory C. Beroza[1]*

Earthquake monitoring in urban settings is essential but challenging, due to the strong anthropogenic noise inherent to urban seismic recordings. Here, we develop a deep-learning-based denoising algorithm, UrbanDenoiser, to filter out urban seismological noise. UrbanDenoiser strongly suppresses noise relative to the signals, because it was trained using waveform datasets containing rich noise sources from the urban Long Beach dense array and high signal-to-noise ratio (SNR) earthquake signals from the rural San Jacinto dense array. Application to the dense array data and an earthquake sequence in an urban area shows that UrbanDenoiser can increase signal quality and recover signals at an SNR level down to ~0 dB. Earthquake location using our denoised Long Beach data does not support the presence of mantle seismicity beneath Los Angeles but suggests a fault model featuring shallow creep, intermediate locking, and localized stress concentration at the base of the seismogenic zone.

INTRODUCTION

Earthquake risk is highest in urban settings, owing to population density and to the presence of extensive and vulnerable infrastructure. Ideally, the intensive earthquake monitoring efforts in urban areas would be used to characterize the fault systems that pose the most immediate and direct threats to cities. However, the same factors—population and infrastructure—that cause risk exposure to be high also make earthquake monitoring difficult to carry out. This is due to the various types of seismic noise generated in cities and the logistical difficulties of instrument deployment.

The Los Angeles metropolitan area is located within an active plate boundary. The Newport-Inglewood Fault runs directly through Los Angeles (Fig. 1), as do other faults that either traverse or surround the Los Angeles Basin, including the Palos Verdes, Santa Monica–Hollywood, Sierra Madre, and Whittier Faults as well as some blind faults. Microseismic monitoring is important for this densely populated city because earthquake locations provide essential constraints on the location and geometry of active faults and the hazards they pose (1–3).

Traditional earthquake monitoring methods that use single-station measurements to detect wave arrivals for events above the noise floor on individual channels may fail to detect smaller events (4–6). Dense array data provide opportunities to detect and analyze these weak sources because adjacent stations have common signal attributes that can be exploited in detection algorithms. In 2011 and 2012, dense arrays with ~100-m spacing were deployed in two phases in Long Beach (Fig. 1A, blue and green polygons). Phase A (blue polygon) operated for the first 6 months of 2011, covering an area of 10 km by 7 km, and included approximately 5200 vertical 10-Hz geophones with a sampling frequency of 500 Hz. Phase B (green polygon) extended the original survey toward the east and covered a similar tectonic setting above a branch of the Newport-Inglewood

Fault Zone. It operated for the first 3 months of 2012, covered an area of 8.5 km by 4.5 km, and included approximately 2500 geophones.

Inbal et al. (7, 8), Li et al. (9), and Yang et al. (10) used the density of seismic wavefield data from these Long Beach arrays for microseismic monitoring. To suppress the strong anthropogenic noise from the Long Beach phase A data, Inbal et al. (7, 8) used downward continuation to back-propagate the wavefield recorded at the surface to a depth of 5 km and performed back-projection (BP) to locate/image seismic events below that. They detected and located widespread seismicity at depths greater than 20 km in the upper mantle, which is much deeper than the conventionally determined and widely accepted seismogenic depth limit for continental earthquakes in this region (11). Li et al. (9) used local waveform similarity to detect small events from the low signal-to-noise ratio (SNR) data. Their findings differ from those of Inbal et al. (7, 8) in that they only detected events with shallow origins. Yang et al. (10) applied a trace randomization procedure to assess the reliability of upper mantle earthquakes using back-projected Long Beach phase B data. By comparing the seismic localization results between the original and trace-randomized data, they inferred that the deep upper mantle events found by Inbal et al. (7, 8) may not be reliable event detections.

The discrepancies among these results are primarily due to the low SNR of the data. While methods such as downward continuation/ BP and local waveform similarity can decrease the detection threshold for small earthquakes, they are sensitive to both noise and uncertainties in the velocity structure. Seismic denoising has the potential to enhance detection sensitivity and the flexibility to improve results for a broad range of approaches to earthquake detection/location and for seismic structural imaging (12–14).

Traditional denoising methods based on simple spectral filtering fail when seismic signals and noise overlap within the same frequency bands. Time-frequency domain denoising can overcome this problem; however, the choice of a suitable thresholding function to map noisy data to optimally denoised signals is challenging. Machine learning techniques, especially deep learning (15), provide a powerful approach to learn complex functional relationships and to use them to extract useful patterns from very large datasets (16–21). This provides a promising approach for time-frequency denoising through

[1]Department of Geophysics, Stanford University, Stanford, CA, USA. [2]State Key Laboratory of Lithospheric Evolution, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China. [3]JAMSTEC/YES, 3173-25 Showa-machi, Kanazawa-ku, Yokohama 236-0001, Japan.
*Corresponding author. Email: zhaoliang@mail.iggcas.ac.cn (L.Z.); beroza@stanford.edu (G.C.B.)
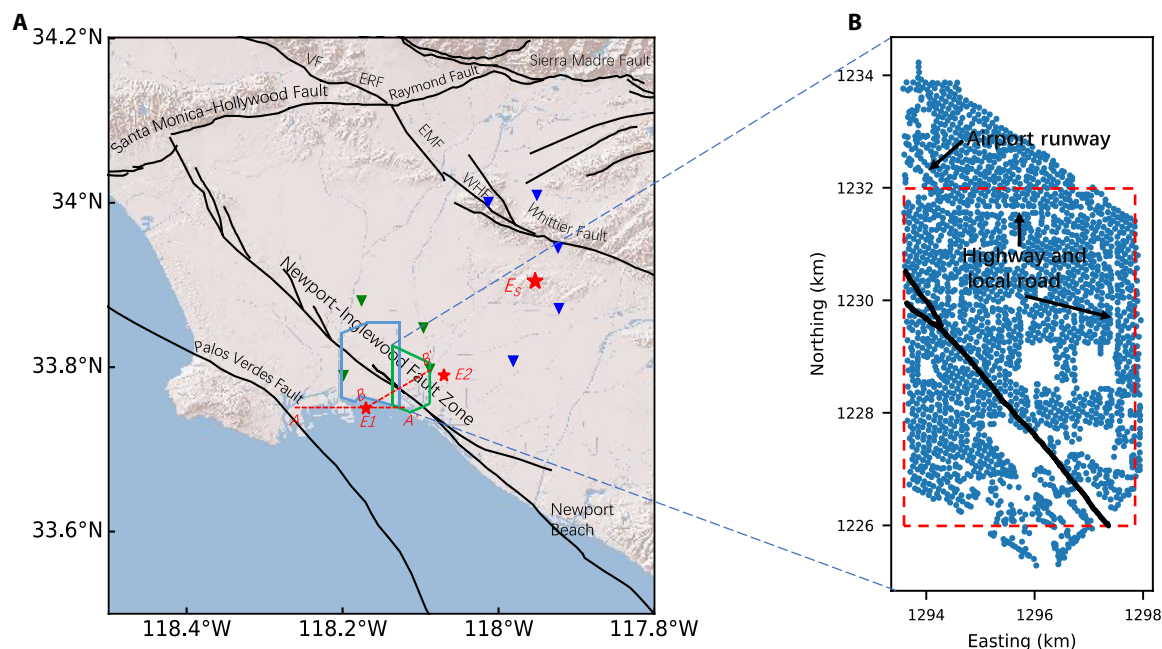
Fig. 1. Los Angeles Basin and Long Beach dense nodal deployment. (**A**) Map of the Los Angeles Basin showing the Newport-Inglewood Fault and other faults. VF, Verdugo Fault; ERF, Eagle Rock Fault; EMF, East Montebello Fault; WHF, Workman Hill Fault. Blue and green polygons outline the Long Beach phase A and B deployments, respectively. The red stars, *E1* and *E2*, show the epicenters of two earthquakes that occurred during the deployment of Long Beach phase B. *AA'* is a profile across the epicenter *E1*. *BB'* is a line across the epicenter *E1* and subvertical to the surficial trace of the Newport-Inglewood Fault Zone. The red star, E_S, shows the mainshock epicenter of the 2014 La Habra earthquake sequence. Green and blue inverted triangles represent the regional stations of the Southern California Seismic Network (SCSN) close to the Long Beach phase B deployment and the earthquake sequence, respectively. (**B**) Map of Long Beach phase B deployment. The strip of missing sensors in the top left is the Long Beach Airport runway. The narrow gap in the northern and eastern parts of the deployment tracks the highway and local roads. Black lines show the mapped surface trace of the Newport-Inglewood Fault. The red dashed rectangle is the superficial boundary of the three-dimensional (3D) imaging volume and horizontal slices we used in our analysis.

the sparse representation of data and improved signal versus noise. Zhu *et al.* (*22*) developed DeepDenoiser based on a deep neural network, which substantially improves the SNR with minimal changes in the waveform shape of interest. DeepDenoiser was originally trained on an extensive dataset from Northern California that was recorded on instruments deployed in unpopulated, low-noise settings. It effectively denoises independent seismic data recorded in that setting but does not generalize well to the Long Beach dataset, presumably because the noise sources differ from those of the Northern California seismic dataset on which the network was trained. Tibi *et al.* (*23*) modified the neural network architecture and developed a new model by training it with seismic data collected from the University of Utah Seismograph Stations network. The newly trained convolutional neural network model showed a success when it was applied to denoise regional distance seismic data in Utah.

The Long Beach dataset represents a rich data source for urban seismological noise (*24*). In this study, we retrained a machine-learning-based denoising algorithm by exploiting this rich noise resource, within the framework of DeepDenoiser, to filter out the strong noise levels present for seismic data recorded in urban settings. To explore the validity of the previously reported widespread seismicity down to the upper mantle beneath Long Beach, we specifically included high SNR seismic signals from the San Jacinto dataset (*5*) in the training dataset for the neural network to learn the seismic signature of real earthquakes recorded on the same instruments but in a low-noise environment. We then applied

UrbanDenoiser to the seismic data recorded by dense array and regional seismic networks and demonstrated that this deep-learning-based denoising tool can substantially improve the SNR for earthquakes across a range of magnitudes and enhance the detection rate by several folds. Thus, this method has the potential to improve the detection capacity of earthquake monitoring networks in urban settings.

RESULTS
Network training
We developed UrbanDenoiser by training a deep neural network using a waveform dataset that combines rich noise sources from the urban Long Beach dense array and high SNR earthquake signals from the rural San Jacinto dense array. The architecture of the neural network is based on that of the DeepDenoiser algorithm (Materials and Methods) (*22*). The dataset comprises 80,000 noise samples and 33,751 signal samples and were randomly split into training, validation, and test sets. We generated noisy waveforms at different SNR levels by repeatedly combining the signal training set with randomly selected noise samples from the noise training set and by randomly shifting the waveform in the window (*25*). The input for the neural network was the two-dimensional (2D) time-frequency representation of noisy waveforms, as determined by the short-time Fourier transform. Both the real and imaginary parts were input into the neural network so that it could learn from the time and phase information. The prediction targets were two masks for the recovered signal and noise. We generated seismic

waveforms for the validation set using the same procedure and applied them to fine-tune the hyperparameters of the network.

We extracted noise samples from the Long Beach data. These waveforms included various types of traffic sources (such as cars, airplanes, and helicopters), vibroseis events, wind interactions with surface obstacles, and other unknown activities (26, 27). We collected seismic recordings from all the receivers in the Long Beach phase B deployment on 27 and 48 Julian days 2012 and selected seismic noise samples from them, because there were fewer earthquakes during these 2 days in the Quake Template Matching (QTM) catalog (28). We segmented the data in 90-s-long time series and removed those containing earthquake signals from known seismic events in the QTM catalog or as determined by the PhaseNet algorithm (29).

The signal samples were extracted from the San Jacinto dataset, which was recorded by another dense array deployed on the active Clark branch of San Jacinto Fault from 7 May 2014 to 13 June 2014 (5). This deployment consisted of ~1108 geophones that collected high-quality seismic signals from small- to medium-magnitude local earthquakes. The two deployments used the same sensors with the same instrument responses. We selected the labeled signals under strict conditions. We ran PhaseNet on the continuous data, and the candidate earthquake signal waveforms were selected on the basis of their coherence across the seismic network. We selected only the signal windows with SNR > 10 (defined in Materials and Methods) as the labeled signals. We also included 30,000 seismic signal samples (SNR > 12) from the North California Seismic Network in the training dataset to increase the predictive power of deep neural network and reduce overfitting.

We tested the performance of the newly trained UrbanDenoiser on both the constructed waveforms from test set and real Long Beach waveform data that were not included in the training process and compared it with the original DeepDenoiser (Materials and Methods). Figures S1 to S3 show that UrbanDenoiser can achieve higher SNR and less waveform and amplitude distortion compared with DeepDenoiser on the constructed test dataset, while fig. S4 indicates that UrbanDenoiser also generalizes to the real Long Beach seismic data with improved earthquake SNRs. Figures S5 and S6 show concrete denoising results on Long Beach seismic recordings for different types of anthropogenic noise and earthquake waveforms. In these, we demonstrate that UrbanDenoiser performs better than the original DeepDenoiser for identifying noise and removing it from the seismic data, and in capturing seismic signal characteristics from the Long Beach data, to separate them from seismic noise.

Seismic location on the denoised Long Beach data with UrbanDenoiser

We applied UrbanDenoiser to 7 days of seismic data (1 to 7 March 2012) and performed BP on the denoised continuous data within a 4.4 km by 6.0 km by 25.0 km of imaging volume, the boundary of which is shown as a red dashed rectangle in Fig. 1B, to detect and locate the most likely seismic sources (Materials and Methods). Robust seismic denoising allows us to work on the entire day's data, not just during the night when anthropogenic noise is lower, as had been done previously (7, 10). This doubles the utility of existing data. Figure 2A shows a 1-day seismogram recorded by a randomly selected receiver that reveals a strong time-varying behavior in
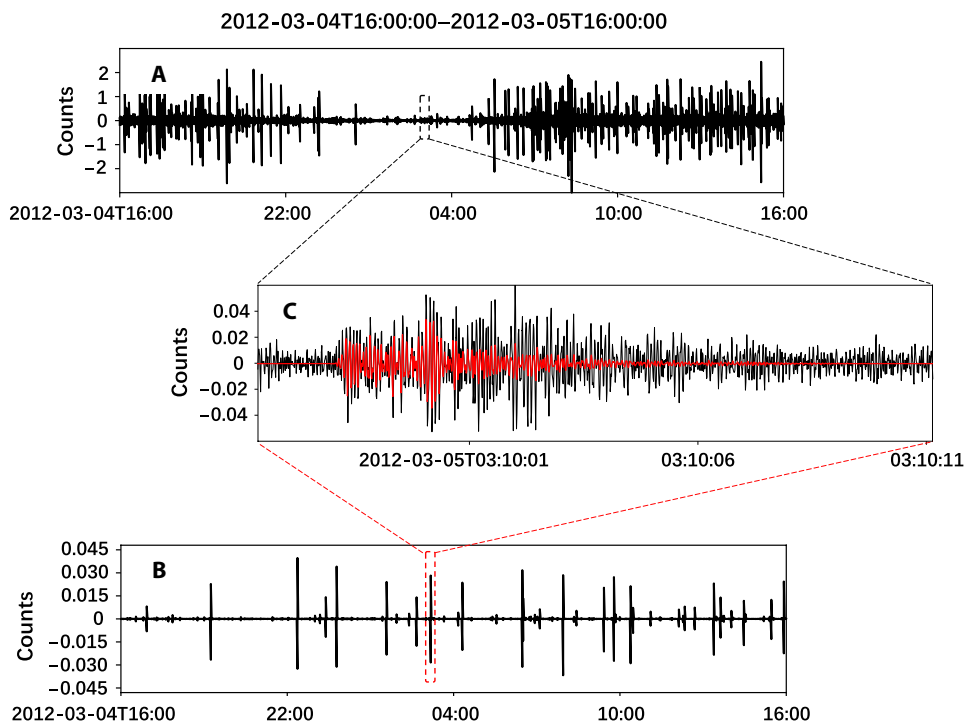
Fig. 2. Seismogram recorded by station R1134_5043 and the denoised version (local time). (**A**) One-day raw data. (**B**) Denoised version of (A). (**C**) Zoomed-in view of a microearthquake event in (A) (black) overlain by its denoised version from (B) (red). There is little phase distortion in the recovered signal, and the amplitude is well preserved in the earlier phases of body wave, while the later phases (e.g., surface wave or scattered phases) in coda may suffer from amplitude distortion. For the purpose of earthquake detection and localization, the later phases will not strongly influence the results. Some of the spikes in (B) are false positives, and we eliminated their influence on earthquake detection through waveform coherence across the dense array.

amplitude for local time 6:00 to 22:00 when the noise level is high versus 22:00 to 6:00[+1] when the noise level is low. Figure 2B shows the denoised results from Fig. 2A, which eliminates the daytime/nighttime variation.

Figure 3 shows the 7-day earthquake localization results from the original data (Fig. 3A) and the denoised data (Fig. 3B) in horizontal slices at different depth ranges. Each dot represents the detection and size scale with the back-projected amplitude. We removed the detections located at the boundary of the volume to avoid interference from regional events (among all the detections, the ratio between the detections inside and outside of the volume is about 1:0.71). The earthquake distribution pattern of the denoised data differs from that of the original data in the following aspects:

1) At 0 to 5 km depths, the detections of the original data were dispersed throughout the imaging area, while most of the detections by the denoised data were widely scattered around the fault trace;

2) At 15 to 20 km depths, the detections by the denoised data track the fault trace closely, while this trend in the original data is much weaker;

3) We found some detections deeper than 20 km in the original data, but very few of these were indicated in the denoised data.

We validated our detection/localization results in Fig. 3B by examining the seismic waveforms from the dense array dataset. We selected one detection at (1296.8, 1226.8, and 5 km) and plotted the seismic profiles spanning the duration of the earthquake. Figure 4A shows the seismic profiles of the raw Long Beach data, from which we can barely identify the seismic signals due to the strong background noise, relative to the weak earthquake energy; however, after denoising, we can see the seismic arrivals in Fig. 4B. Figure 4C shows a magnified view of the subset from Fig. 4B. The increased first arrival time on traces, sorted by the distances between each station and the determined epicentral location, supports the validity of the localization result. This detection was also validated after denoising the seismograms recorded by the isolated regional stations from the Southern California Seismic Network (SCSN) (fig. S7).

We performed BP to image a local magnitude (M) 2.1 earthquake on 27 March 2012 (red star $E1$ in Fig. 1A) with the original data and denoised data by DeepDenoiser and UrbanDenoiser (Materials and Methods) (see text S1). Although the localizations determined based on data processed by different procedures were similar, the resultant peak amplitudes varied with the UrbanDenoiser result showing the largest amplitude and the DeepDenoiser result showing the smallest amplitude (fig. S8, C to E). A seismic profile with denoised waveforms from 59 selected traces along a line intersecting the epicenter, perpendicular to the surface trace of the Newport-Inglewood Fault (red dashed line BB' in Fig. 1A), indicates that the P wave velocity to the west of the fault trace is faster than that to the east (fig. S8F) (30).
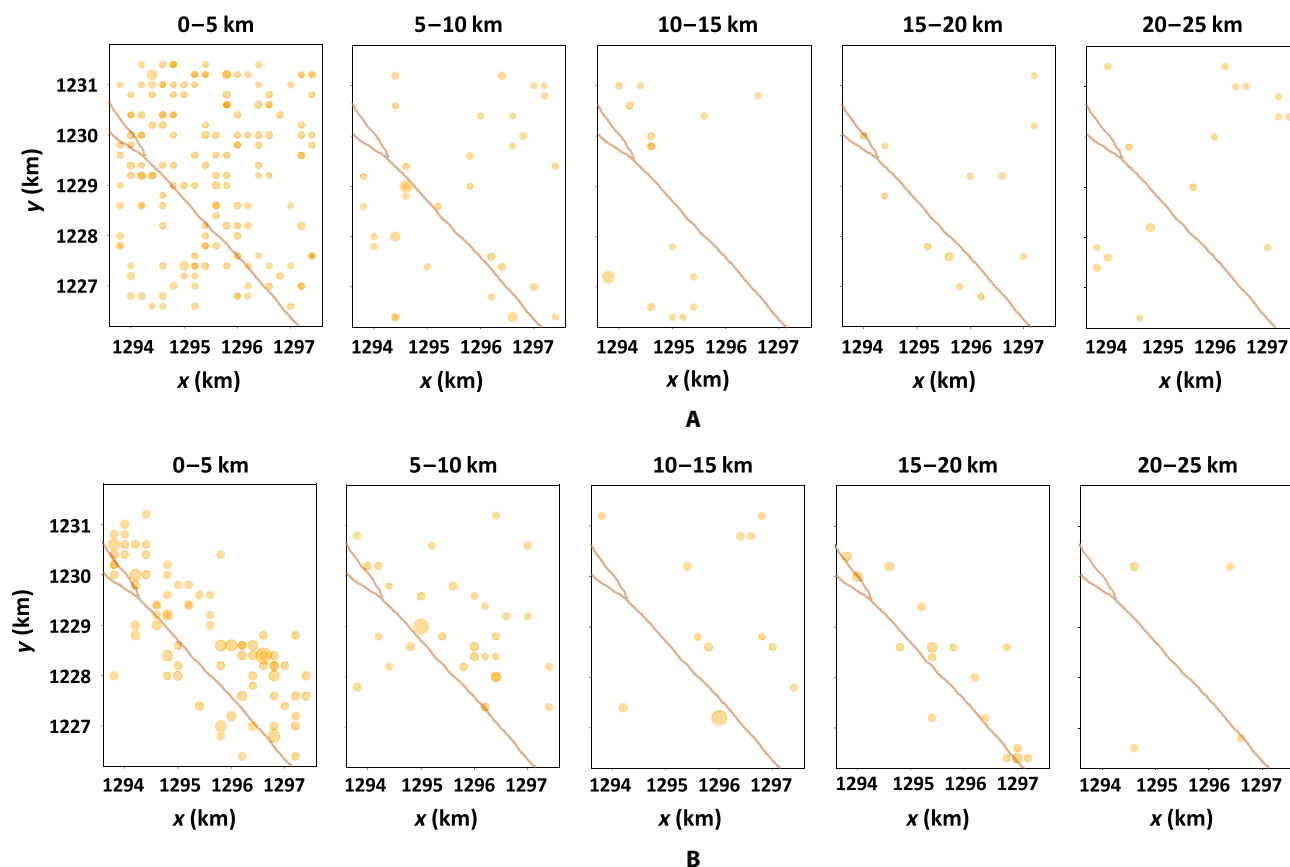
Fig. 3. Seven-day earthquake BP location results in horizontal slices at different depth ranges of 0 to 5 km, 5 to 10 km, 10 to 15 km, 15 to 20 km, and 20 to 25 km. (**A**) Original and (**B**) denoised data. The sizes of the dots scale with the number of median absolute deviation (MAD) by which the back-projected amplitudes exceed the detection threshold.
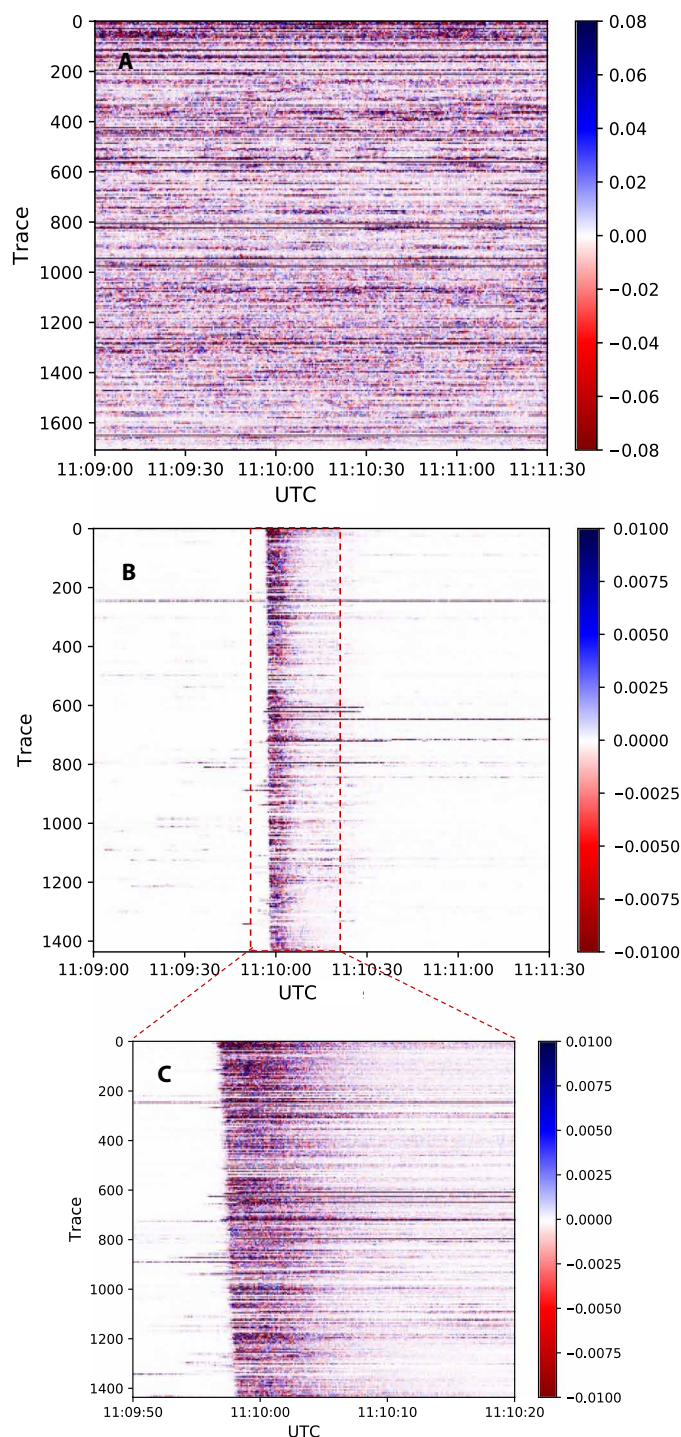
Fig. 4. Seismic profile containing information for a small earthquake. (**A**) Raw data. (**B**) Denoised data. (**C**) Zoomed view of the red rectangle in (B). The color bar indicates the amplitude of the seismic waveforms.

An earthquake was detected by the Long Beach nodal array on 7 March 2012. It occurred 2.5 km east of the imaging volume (Fig. 1A, red star $E2$); hence, the seismic energy (fig. S9B) was back-projected to a point on the east boundary of the imaging volume (fig. S9A), which is excluded from Fig. 3B. We visually inspected the

ground motion information recorded for this earthquake. Movie S1 shows the ground motion in the raw data (movie S1A) and in the denoised data (movie S1B). Comparing movie S1 (A and B), we found that, before the seismic wave arrived at the dense array, the background motion in the denoised data was clean, but that the raw data for the same contained substantial anthropogenic noise with the strongest noise appearing along the highway and local roads. The ground motion information in the denoised data has a much cleaner wavefront representing the passage of the seismic waves through the dense array deployment than that of the raw data.

Application of UrbanDenoiser to a regional seismic network for the La Habra earthquake sequence

An earthquake sequence struck urban La Habra with a mainshock magnitude of 5.1 (Fig. 1A, red star E_s) at 4:09:41 UTC on 29 March 2014. We chose the five stations from the SCSN closest to the sequence (Fig. 1A, blue inverted triangles) and applied UrbanDenoiser to the seismograms. Although the instrument response of the SCSN sensors differs from that of the nodal sensors, the denoised data revealed more detections for this earthquake sequence.

We confirmed an earthquake when the detected phases could be associated on two or more stations. By doing this, we found a total of 488 events in the 10 hours between 3:00 and 13:00. This amounts to more than 4.5 times the number (108) of detections in the SCSN catalog (*31*) and 10% greater than the number of detections listed in the QTM catalog (*28*), which is currently the most comprehensive catalog for southern California.

Figure 5 shows 40-min seismograms (03:20 to 04:00 UTC, 29 March 2014; vertical component only) from the five stations. This was a period between the *M* 3.57 foreshock and *M* 5.1 mainshock and was a relatively quieter window compared with the periods following the mainshock. There are no detections related to this sequence in the SCSN catalog during this time. The only detection listed in the QTM catalog was an *M* 0.67 earthquake at 3:40:59. On the basis of the denoised data, however, we found a total of nine events over this 40-min period. Figure 5 (A_{4-II} to D_{4-II}) shows the seismic signal waveforms detected both in the QTM catalog and on the denoised waveforms (marked by blue dashed rectangle), while Fig. 5 (A_{4-I} to E_{4-I} and B_{4-III} to D_{4-III}) shows another two examples detected on the denoised waveforms only (marked by red dashed rectangles). Comparing Fig. 5 (A_{4-I} to E_{4-I}, A_{4-II} to D_{4-II}, and B_{4-III} to D_{4-III}) with the raw data in Fig. 5 (A_{3-I} to E_{3-I}, A_{3-II} to D_{3-II}, and B_{3-III} to D_{3-III}), we found substantial enhancement of the SNR in the denoised version. This demonstrated that UrbanDenoiser can facilitate the detection of more small events in an urban setting.

Figure 6 shows a comparison between the SNR of the denoised signals versus nondenoised signals from Station CI.FUL for 102 events with $-0.16 < M < 5.1$. Although the SNR of both the non-denoised and denoised signals decreases with decreasing magnitude, the SNR of the denoised data is consistently higher, and the decreasing trend is slower. UrbanDenoiser is effective in denoising noisy signals with an SNR floor of approximately 0 dB (the minimum signal amplitude close to the noise level). On average, UrbanDenoiser enhanced the SNR by about 15 dB, with the most marked improvement of around *M* 1.5 to 3.8 (original SNR: 2.5 to 20 dB). This compares with a recently reported average increase of ~5 dB in SNR for a denoising applied to more typical seismological settings (*23*).
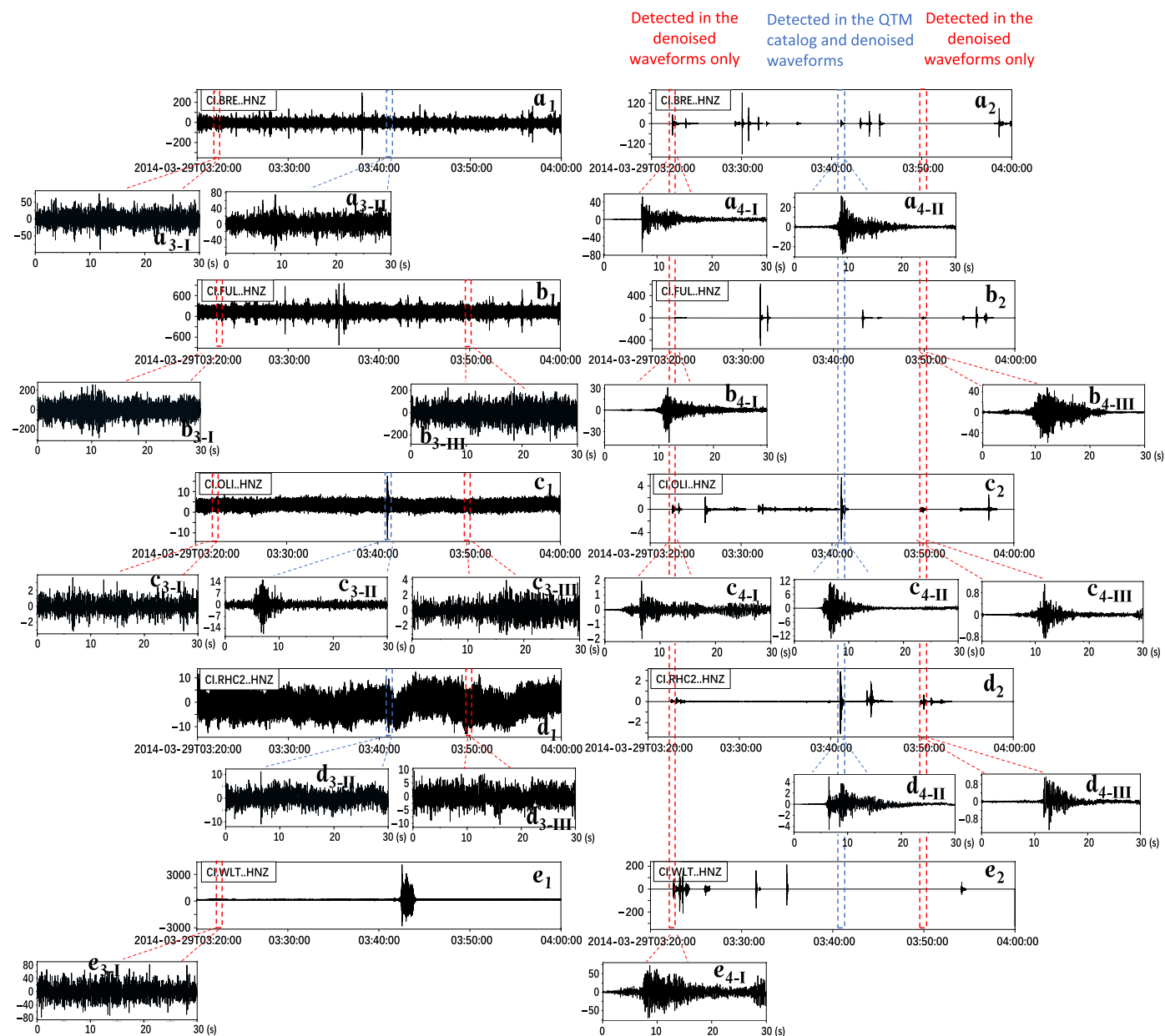
Fig. 5. Application of UrbanDenoiser to the 40-min seismograms (3:20 to 4:00 UTC, 29 March 2014; vertical component) from the five SCSN stations (stations CI.BRE, CI.FUL, CI.OLI, CI.RHC2, and CI.WLT). (**A1** to **E1**) Raw seismograms. (**A2** to **E2**) Denoised seismograms. (**A4** to **E4**) Zoomed-in view of the denoised potential earthquake waveforms compared with the raw waveforms (**A3** to **E3**). No detections are related to this sequence in the SCSN catalog during this time. The blue dashed rectangles mark the only detection in the QTM catalog, which was also detected in the denoised waveforms. The red dashed rectangles mark the seismic signals detected in the denoised waveforms only.

DISCUSSION

We applied UrbanDenoiser to both the dense array and regional seismic network. Application to the Long Beach dense array data allowed us to use a large portion of previously unusable seismic data for seismic analysis, such as daytime data dominated by anthropogenic noise. The earthquake localization results based on the denoised data show a different distribution pattern than the original data, which updates our knowledge on the seismogenic characteristics. The application of UrbanDenoiser to the data from regional seismic networks demonstrates that UrbanDenoiser can enhance the SNR

for earthquakes across different magnitudes and that it can recover seismic signals from noisy data, with an SNR floor close to 0 dB. For the La Habra earthquake sequence, the number of detections we observed in the denoised data was approximately four times more than the list of detections in the existing regional seismic catalog.

The detection/location results contained only earthquake events and excluded large-amplitude non-earthquake sources (e.g., Fig. 3B). Conventional detection methods detect pulses of energy with amplitudes that exceed the detection threshold and cannot differentiate earthquakes from other signals, such as waveforms generated by
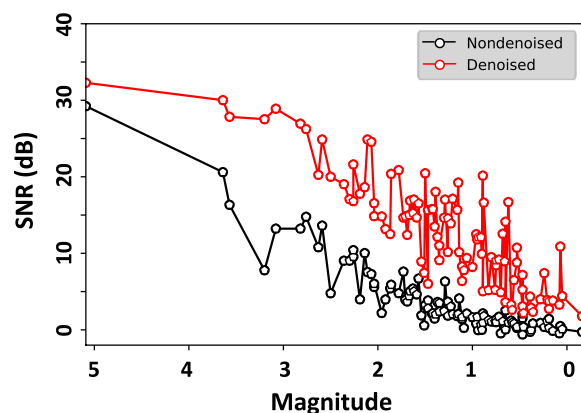
Fig. 6. SNR of the denoised signals versus nondenoised signals from Station CI.FUL for 102 events varying between *M* −0.16 and *M* 5.1.

vibroseis or traffic. Movie S2 (A and B) shows the ground motion for vibroseis events recorded by the Long Beach dense array on raw data and denoised data. In movie S2A, we observe the vibroseis operating at the northwest corner of the deployment. The waveforms generated by vibroseis and other noise sources were removed from the denoised data, so that the background was relatively cleaner in movie S2B. We performed BP on 8 hours of seismic data that included more than 5.5 hours of vibroseis operations. Figure S10 (A and B) shows the BP location results for the nondenoised and denoised data, respectively. From fig. S10A, we can see an anomalously large number of detection clusters in the northwest corner of the imaging volume from the vibroseis and many more detections throughout the imaging volume, especially at the bottom. In fig. S10B, however, the influence from the vibroseis is suppressed, and the earthquake location results are not strongly contaminated by anthropogenic activities.

UrbanDenoiser can effectively suppress the high noise levels, although false positives and false negatives in denoised data should still be expected to occur and need to be assessed. The influence of false positives in denoised data can be effectively filtered out using dense array data. False negatives occur when the seismic signal is too weak or when the target seismic phases and training signal samples are not similar enough to real earthquake waveforms. UrbanDenoiser is generally less capable in recovering coda than earlier phases. This will not affect the detection of *P* onset and other earlier phases, so it should have minimal impact on the earthquake detection and location results.

Although the aim of UrbanDenoiser is to separate earthquake signals from urban noise, it has the potential to be extended to denoising for vibroseis signals by training it using high SNR signal samples from vibroseis events. In the current model of UrbanDenoiser, the vibroseis waveforms are fed into the neural network as labeled noise samples, so that they are recovered in the class of noise (e.g., fig. S5, A3 and A4). For training "VibroseisDenoiser," the vibroseis waveforms were signals to our interest, while the earthquake waveforms would be grouped with the noise. The recovered cleaner vibroseis signals benefit the follow-on signal processing for seismic imaging.

For the most part, we do not have dense array deployments, such as Long Beach phases A and B, available to generate more complete earthquake catalogs. Instead, seismic monitoring implementations rely on isolated seismic instruments from regional seismic networks. The conventional short time average/long time average method can result in many false detections for phase identification (fig. S7, A2 to E2), which degrades the performance of phase association and event localization. UrbanDenoiser can remove most of the noise bursts from the raw data and markedly increase the SNR for seismic recordings in a single trace. This benefits subsequent earthquake detection processing and enhances the capacity of seismic monitoring in urban areas through regional seismic networks.

The earthquake detection results shown in Fig. 3B do not show evidence of widespread seismicity below 20 km in the upper mantle. We observed a weak tendency for events to follow the surface trace of the Newport-Inglewood Fault at 0 to 5 km. The seismicity at 5 to 15 km was more dispersed, which could be due to fault locking. The seismic location result is consistent with our previous study (*10*), which found that the seismicity concentration is greater at a 15- to 20-km depth range, the approximate root of the seismogenic zone, than at shallower depths of the probable locking part (5 to 15 km). This could be due to the stress concentration near the seismic-aseismic transition. On the basis of our analysis, we conclude that earthquake detection and location following preprocessing using deep learning to filter out noise should facilitate improved earthquake monitoring in urban environments.

MATERIALS AND METHODS
Network training and model evaluation
UrbanDenoiser is a transfer learning application using the same architecture and a pretrained model from DeepDenoiser (*22*). The network used an encoder and decoder architecture of 22 hidden layers. The input data were first encoded into dense features through a sequence of three by three convolutional layers with two by two strides for downsampling. The features were then decoded through a sequence of three by three convolutional layers, as well as three by three deconvolutional layers for upscaling. For each layer, a rectified linear unit activation function and batch normalization were applied. Skip connections were used to improve convergence. In the last layer, a softmax normalized exponential function was used to predict two masks for signal and noise, respectively.

The training data included 80,000 noise samples and 33,751 signal samples, which were randomly split into training, validation, and test sets (70%, 15%: 15%). The labeled waveform data were in 90-s-long time windows, which allowed the entire signal seismograms (<30 s) to be randomly shifted within the length of this window for data augmentation (*25*). The training parameters were the values for the three by three convolution or deconvolution filters in each of the layers. We used the Adam optimizer with a learning rate of 0.001, a batch size of 20, a maximum number of epochs of 80, and the cross-entropy loss function.

We assessed the performance of UrbanDenoiser on both the constructed waveforms from the test dataset and real seismic recordings from the Long Beach dataset. Three evaluation metrics including SNR, normalized correlation coefficient for measuring the similarity between the shapes of two waveforms, and signal-to-distortion ratio (SDR) for amplitude distortion were applied. The SNR and SDR are defined as follows (*23*)

$$ SNR = 10 \log_{10} \frac{A_S}{A_N} \qquad (1) $$

$$SDR = 10 \log_{10} \frac{\|W_{GT}\|}{\|W - W_{GT}\|} \qquad (2)$$

where A_S and A_N represent the seismic energy after and before the first arrival, respectively; W_{GT} is the amplitude array for the ground truth seismogram from the test set; and W is the amplitude array for the corresponding waveform recovered from the denoiser. Similar amplitudes between two waveforms result in high SDR values.

Figures S1 to S3 show the distribution of SNR, normalized correlation coefficient, and SDR for the recovered information (signals and noise) by UrbanDenoiser versus DeepDenoiser from the test dataset of constructed, noisy waveform. We observe that both the DeepDenoiser and UrbanDenoiser can improve these three metrics, while UrbanDenoiser outperforms DeepDenoiser in all these aspects.

We tested the performance of UrbanDenoiser on the real Long Beach dataset that were not included in the training process. Figure S4 shows the improvements in SNR of UrbanDenoiser for the noisy, urban seismic dataset. Figures S5 and S6 show examples in which UrbanDenoiser can better capture the seismic signal information and remove the noise waveforms from the Long Beach data compared with DeepDenoiser.

BP imaging
We performed BP to image the earthquake in two steps: (i) time shift of each seismogram and (ii) stacking. It can be expressed as (6)

$$\text{stack}_i(t) = \sum_{k=1}^{n} s_k(t - t_{ik}) \qquad (3)$$

where $s_k(t)$ is the seismogram recorded at the kth station, t_{ik} is the calculated travel time from the ith grid point to the kth station based on a known velocity model, n is the number of stations, and $\text{stack}_i(t)$ is the stacked seismogram for the ith grid-searching point.

We calculated the travel time between each grid point and each geophone at the surface based on the Southern California Earthquake Center Community Velocity Model (CVM-H 11.9.1) (32) and stored them in a travel time lookup table for computational convenience. We performed a grid search for each potential source location within the imaging region. For each grid point, all the seismograms were time-shifted on the basis of the corresponding travel time, and the aligned seismograms were stacked to a single representative time series. The largest amplitude value along the time series was set as the amplitude value at the imaging point.

Seismic data processing and BP localization with continuous data
We converted the original data from SEG-D format to NumPy format, decimated the time series from 500 to 100 Hz, and processed them with UrbanDenoiser. The denoised data were downsampled to 50 Hz and band-pass–filtered from 5 to 15 Hz. We normalized the data with their 1-hour maximum value to suppress the influence of any strong spatially dependent residual noise level and calculated the envelope by smoothing the data with a three-point median window on the squared waveforms to reduce the sensitivity to the inaccuracy of the velocity model.

We performed BP for a 4.4-km (X) by 6-km (Y) by 25-km (Z) 3D imaging volume with a grid spacing of 200 m in each dimension. The geographic boundary of the imaging volume is indicated by the red dashed rectangle in Fig. 1B. BP was performed as described above. We segmented the shifted-and-stacked time series for each grid point into 3-s time windows, and the maximum value within each time window was assigned as the BP value of this grid point. Thus, we obtained a 3D imaging volume for each 3-s time window. If the maximum BP value through the entire space within a time window exceeded the detection threshold, then we marked the corresponding grid point as a detection.

We set 10 times the median absolute deviation (MAD) as the detection threshold for BP earthquake detection using the denoised Long Beach dense array data. We fitted the peak amplitude values from all time windows at each imaging point with a generalized extreme value distribution. Figure S11 shows the peak amplitude value distribution for 28,800 time windows from 1 day of stacked seismograms for a randomly selected imaging point. The possibility of exceeding 10-MAD is 2.38×10^{-5}, which means that, under this detection threshold, we expect the number of false detections to be smaller than one per day. For the original data, we only used nighttime data and set eight times the MAD as the detection threshold.

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abl3564

REFERENCES AND NOTES

1. J.-L. Got, J. Fréchet, F. W. Klein, Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea. *J. Geophys. Res.* **99**, 15375–15386 (1994).
2. F. Waldhauser, A double-difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California. *Bulletin of the Seismological Society of America.* **90**, 1353–1368 (2000).
3. H. Kao, S.-J. Shan, The source-scanning algorithm: Mapping the distribution of seismic sources in time and space. *Geophys. J. Int.* **157**, 589–594 (2004).
4. K. K. Davenport, J. A. Hole, D. A. Quiros, L. D. Brown, M. C. Chapman, L. Han, W. D. Mooney, Aftershock imaging using a dense seismometer array (AIDA) after the 2011 Mineral, Virginia, earthquake, in *The 2011 Mineral, Virginia, Earthquake, and Its Significance for Seismic Hazards in Eastern North America* (Geological Society of America, 2015).
5. Y. Ben-Zion, F. L. Vernon, Y. Ozakin, D. Zigone, Z. E. Ross, H. Meng, M. White, J. Reyes, D. Hollis, M. Barklage, Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone. *Geophys. J. Int.* **202**, 370–380 (2015).
6. G. D. Beskardes, J. A. Hole, K. Wang, M. Michaelides, Q. Wu, M. C. Chapman, K. K. Davenport, L. D. Brown, D. A. Quiros, A comparison of earthquake backprojection imaging methods for dense local arrays. *Geophys. J. Int.* **212**, 1986–2002 (2018).
7. A. Inbal, R. W. Clayton, J.-P. Ampuero, Imaging widespread seismicity at midlower crustal depths beneath Long Beach, CA, with a dense seismic array: Evidence for a depth-dependent earthquake size distribution. *Geophys. Res. Lett.* **42**, 6314–6323 (2015).
8. A. Inbal, J. P. Ampuero, R. W. Clayton, Localized seismic deformation in the upper mantle revealed by dense seismic arrays. *Science* **354**, 88–92 (2016).
9. Z. Li, Z. Peng, D. Hollis, L. Zhu, J. McClellan, High-resolution seismic event detection using local similarity for Large-N arrays. *Sci. Rep.* **8**, 1646 (2018).
10. L. Yang, X. Liu, G. C. Beroza, Revisiting evidence for widespread seismicity in the upper mantle under Los Angeles. *Sci. Adv.* **7**, eabf2862 (2021).
11. J. J. Nazareth, The seismogenic thickness of the southern California crust. *Bull. Seismol. Soc. Am.* **94**, 940–960 (2004).
12. D. Bonar, M. Sacchi, Denoising seismic data using the nonlocal means algorithm. *Geophysics* **77**, A5–A8 (2012).
13. S. Gaci, The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces. *IEEE Trans. Geosci. Remote Sensing.* **52**, 4558–4563 (2014).
14. S. M. Mousavi, C. A. Langston, S. P. Horton, Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform. *Geophysics* **81**, V341–V355 (2016).
15. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
16. P. M. R. DeVries, F. Viégas, M. Wattenberg, B. J. Meade, Deep learning of aftershock patterns following large earthquakes. *Nature* **560**, 632–634 (2018).
17. T. Perol, M. Gharbi, M. Denolle, Convolutional neural network for earthquake detection and location. *Sci. Adv.* **4**, e1700578 (2018).
18. Z. E. Ross, Y. Yue, M. Meier, E. Hauksson, T. H. Heaton, PhaseLink: A deep learning approach to seismic phase association. *J. Geophys. Res. Solid Earth.* **124**, 856–869 (2019).

19. Q. Kong, D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, P. Gerstoft, Machine learning in seismology: Turning data into insights. *Seismol. Res. Lett.* **90**, 3–14 (2019).

20. M. Meier, Z. E. Ross, A. Ramachandran, A. Balakrishna, S. Nair, P. Kundzicz, Z. Li, J. Andrews, E. Hauksson, Y. Yue, Reliable real-time seismic signal/noise discrimination with machine learning. *J. Geophys. Res. Solid Earth.* **124**, 788–800 (2019).

21. S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, G. C. Beroza, Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat. Commun.* **11**, 3952 (2020).

22. W. Zhu, S. M. Mousavi, G. C. Beroza, Seismic signal denoising and decomposition using deep neural networks. *IEEE Trans. Geosci. Remote Sensing.* **57**, 9476–9488 (2019).

23. R. Tibi, P. Hammond, R. Brogan, C. J. Young, K. Koper, Deep learning denoising applied to regional distance seismic data in Utah. *Bull. Seismol. Soc. Am.* **111**, 775–790 (2021).

24. D. Snover, C. W. Johnson, M. J. Bianco, P. Gerstoft, Deep clustering to identify sources of urban seismic noise in long Beach, California. *Seismol. Res. Lett.* **92**, 1011–1022 (2021).

25. W. Zhu, S. M. Mousavi, G. C. Beroza, Seismic signal augmentation to improve generalization of deep neural networks. *Adv. Geophys.* **61**, 151–177 (2020).

26. H. Meng, Y. Ben-Zion, C. W. Johnson, Detection of random noise and anatomy of continuous seismic waveforms in dense array data near Anza California. *Geophys. J. Int.* **219**, 1463–1473 (2019).

27. C. W. Johnson, H. Meng, F. Vernon, Y. Ben-Zion, Characteristics of ground motion generated by wind interaction with trees, structures, and other surface obstacles. *J. Geophys. Res. Solid Earth.* **124**, 8519–8539 (2019).

28. Z. E. Ross, D. T. Trugman, E. Hauksson, P. M. Shearer, Searching for hidden earthquakes in Southern California. *Science* **364**, 767–771 (2019).

29. W. Zhu, G. C. Beroza, PhaseNet: A deep-neural-network-based seismic arrival time picking method. *Geophys. J. Int.* **216**, 261–273 (2018).

30. J. H. Shaw, A. Plesch, C. Tape, M. P. Suess, T. H. Jordan, G. Ely, E. Hauksson, J. Tromp, T. Tanimoto, R. Graves, K. Olsen, C. Nicholson, P. J. Maechling, C. Rivero, P. Lovely, C. M. Brankman, J. Munster, Unified structural representation of the southern California crust and upper mantle. *Earth Planet. Sci. Lett.* **415**, 1–15 (2015).

31. K. Hutton, J. Woessner, E. Hauksson, Earthquake monitoring in southern California for seventy-seven years (1932-2008). *Bull. Seismol. Soc. Am.* **100**, 423–446 (2010).

32. A. Plesch, C. Tape, R. Graves, J. H. Shaw, P. Small, G. Ely, Updates for the CVM-H including new representations of the offshore Santa Maria and San Bernardino basin and a new Moho surface, paper presented at Southern California Earthquake Center Annual Meeting, Palm Springs, CA, 11 to 14 September 2011.