Beyond Correlations: Deep Learning for Seismic Interferometry

Hongyu Sun, *Graduate Student Member, IEEE*, and Laurent Demanet

***Abstract*—Passive seismic interferometry (SI) is a vastly generalized blind deconvolution question, where different paths through the Earth correspond to different channels called Green's functions; the sources are completely incoherent and not shared by the channels, and the question is to estimate paths (channels) that are not present in the dataset. Seismic interferometry, turning noise to signal, has numerous applications, from monitoring industrial activities to crustal structure investigation. No standard method of signal processing will solve seismic interferometry. Instead, domain scientists resort to a simple cross-correlation operation, a.k.a. correlogram, which can retrieve the Green's function directly, but only under restrictive assumptions of ergodicity (energy equipartitioning) of the random process generating the seismic source. However, in practice, correlograms are not equal to the empirical Green's function, because these assumptions are generally far from being satisfied in realistic situations. In the framework of supervised learning, we propose to train deep neural networks to overcome two limitations of correlation-based SI: the temporal limitation of passive recordings, and the spatial limitation of the random source distribution. Deep neural networks are trained to implicitly find the relationship between the empirical Green's functions and the correlograms and then used to extract the correct Green's functions from ambient noise. The input of the network is correlograms (a virtual shot gather) and the desired output is the empirical Green's function (the active shot gather). The neural network can often retrieve Green's functions from five-minute passive recordings with acceptable accuracy in our synthetic example. Although an exact estimation of the source locations may not be necessary, a prior knowledge of the source directionality (through a preliminary beamforming step) is helpful when training the neural network to mitigate the challenges associated with inhomogeneous source distributions. In this work, all the numerical examples are based on the retrieval of P-wave reflections in the exploration scale, and are conducted on synthetic data. We use a modified ResNet in our numerical experiments.**

***Index Terms*—Ambient noise, deep learning, Green's function retrieval, seismic imaging, seismic interferometry (SI).**

I. Introduction

SEISMIC interferometry (SI) generally refers to the process of generating new seismic responses (Green's function retrieval) by correlating seismic noise recordings of different receivers [1], [2]. Since [3] proposed to use statistical properties of seismic noise to infer the Earth structure, SI applications have ranged from crust and upper-mantle structure investigation [4], to natural resources exploration, to urban monitoring.

Claerbout [5] pioneered the SI algorithm by autocorrelating a transmission seismogram to retrieve its reflection. Subsequently, the principle of correlation-based SI has been derived via normal-mode summation [6], time reversal [7], representation theorems [8] and stationary phase analysis [9] by assuming that the ambient wavefield is equipartitioned.

The challenges of correlation-based SI mainly concern extending the theory to account for real-world noise and media [10]–[12]. A correct retrieval of Green's functions relies on the prerequisite of uncorrelated and homogeneously distributed noise sources in media without intrinsic losses. A strong localized source outside the stationary zones can cause spurious arrivals resulting from imperfect cancellations of nonphysical amplitudes [13]. However, natural noise sources are always correlated and cannot illuminate the region of interest from all sides equally. Also, seismic waves in real-world media can suffer from geometrical spreading and attenuation (hence lowered amplitudes). In addition, there is the issue of statistical stability – large deviations in the case of small sample size (here, short recording window).

It is a major challenge to derive methods to obtain approximately correct Green's functions from the imperfect situations of SI. The statistical stability relative to the statistics of the noise sources can be controlled through the choice of a long enough recording time window or by stacking techniques, e.g., months or years in regional scale [4] and hours or days in exploration scale [14], [15]. However, the usage of long noise recording limits the application of SI for real-time monitoring of processes in the Earth's subsurface. In addition, to enhance signal-to-noise ratio (SNR), it is always necessary to perform time-domain weighting, frequency-domain weighting and spatial averaging. Even so, the results rely on subjective evaluations on the acceptance or rejection of a cross correlation. Thus, despite dedicated data processing, SI techniques still run into issues limiting their application for imaging or monitoring the subsurface.

In addition to cross-correlations, alternative methods for SI include: deconvolution [16], multi-dimension deconvolution [17], cross-coherence [18] and coda wave interferometry [19], [20]. Although interferometry by deconvolution removes the source signature by means of spectral division, it suffers from the same problems as correlation-based interferometry in terms

Manuscript received March 12, 2021; revised December 14, 2021 and April 1, 2022; accepted April 28, 2022. The authors thank TotalEnergies for support. Hongyu Sun acknowledges SEG scholarships and MIT MathWorks Science Fellowship for funding. *(Corresponding author: Hongyu Sun.)*

Hongyu Sun is with the Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (e-mail: hongyu-sun@outlook.com).

Laurent Demanet is with the Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, and also with the Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (e-mail: laurent@math.mit.edu).

of source distribution and recording length (see supplementary information I). More robust preprocessing and postprocessing methods are proposed to enhance the reliability of Green's function extraction from cross-correlation, e.g., directional balancing [21] and iterative denoising [22].

Deep learning [23] enables neural networks with multiple layers to discover intricate structure from large data sets and learn high-level representations for reasonable prediction. Over the past few years, it was used in geophysics for solving inverse problems [24], [25], low frequency extrapolation [26], [27], seismic data denoising [28]–[30] and velocity model building [31]–[33].

In passive data processing, researchers have leveraged machine learning in different ways. [34] use deep learning to measure the time lapse between two correlograms. [35] design an autoencoder for time-lapse monitoring with passive seismic data. [36] train an autoencoder to denoise noise correlation functions with a time resolution of 20 minutes. [37] extract dispersion curves from ambient noise correlations using deep learning. As for unsupervised learning, [38] use a clustering method to improve the quality of correlograms calulated via deconvolution. [39] use unsupervised deep learning to distinguish between earthquake signals and background noises. [40]–[42] apply machine learning for signal detection from ambient noise under a classification framework.

To handle the practical challenges in SI applications, we propose to use deep learning in a supervised manner to extract correct Green's functions from ambient noise wavefields with limited recording duration and localized source distribution. We focus on the Green's function retrieval in the exploration scale. In particular, we aim to retrieve reflected waves from ambient noise that agree with (here, simulated) active surveys. Direct mapping from noise to signal is optional but requires a careful design of the neural network. To relieve the difficulty of learning, we use raw correlograms as input to the network, and train the neural networks to output reflections, i.e., seismic events from an active source.

Although they were originally proposed in the computer vision community, many deep-learning architectures, such as ResNet [43], U-Net [44], DnCNN [45], VAE [46], DCGANs [47], Pix2Pix [48], CycleGAN [49], etc., have been successfully applied to solve geophysical problems. Here we consider the mapping from a virtual shot gather to an active shot gather as an image-to-image translation problem. Unlike our previous work [50], we choose to use convolutional neural networks with residual learning blocks (ResNet, [43]) as the tool for its better accuracy and efficiency (see supplementary information II). In addition, we make problem-specific modification to ResNet and use a wide kernel (11×11) on all convolutional layers to enhance its performance on seismic shot gathers. We train the neural network in a supervised setting with an L_1 loss function between the output and the label. Numerical examples have shown that deep learning with proper setting can overcome the temporal limitation of noise recording length and the spatial limitation of source distribution in correlation-based SI. We believe that seismic imaging with deep-learning-retrieved reflections from short noise recordings, and less-than-ideal source characteristics, offers an interesting new point of view for real-time monitoring.

The organization of this paper is as follows. We first present the basics of correlation-based SI and then propose to improve SI results using deep learning technologies. In the section of numerical results, we show the ability of deep neural networks to retrieve correct Green's functions from short noise recordings and directional wavefields (inhomogeneous source distribution). We further investigate the generalizability of the neural network on different geological structures. Finally, we discuss the challenges associated with field-data application, and the generalization of this method to diffusive media and regional scale. We also explain the limitations inherent to deep learning.

II. Data and Method

A. Review of correlation-based interferometry

Correlation-based SI retrieves Green's functions from N available passive measurements $u_{obs}(\mathbf{x}_A, t)$ and $u_{obs}(\mathbf{x}_B, t)$ using the relation [2], [8]:

$$\begin{aligned}&\{G(\mathbf{x}_A, \mathbf{x}_B, t) + G(\mathbf{x}_A, \mathbf{x}_B, -t)\} * S(t)\\ &\quad\approx \sum_i^N u^i_{obs}(\mathbf{x}_A, -t) * u^i_{obs}(\mathbf{x}_B, t).\end{aligned} \tag{1}$$

Here, $G(\mathbf{x}_A, \mathbf{x}_B, \pm t)$ is the Green's function between positions $\mathbf{x}_A$ and $\mathbf{x}_B$ in a homogeneous lossless acoustic medium. $\pm t$ denotes the causal and acausal parts of the Green's function. $S(t)$ stands for the autocorrelation of the noise sources, and the asterisk denotes convolution. Stacking the cross-correlation over N available measurements improves the statistical stability. This means that long recording time is necessary for a good retrieval of Green's functions.

The derivation of this relation relies on the assumption that ambient noise sources are uncorrelated and surround the area of interest from all directions. However, correlograms are not equal to Green's functions when the assumption is not fullfilled. In addition to an amplitude error, artifacts will be introduced in the reconstructed signal due to the fact that the source locations do not constitute a closed surface [9], [51], [52].

By correlating the passive recording of each receiver with that of the reference trace, a virtual common-shot gather can be obtained where the virtual source is located at the receiver position of the reference trace. To enhance the SNR, the causal and acausal parts are summed. The final result is the Green's function convolved with the autocorrelation of the noise sources $S(t)$. The imprint of source can be removed by deconvolution with a wavelet extracted from the autocorrelation trace using a narrow window around zero time [15].

B. Deep learning for seismic interferometry

We use deep learning to extract correct Green's functions from correlograms in a supervised manner to overcome the practical challenges of SI in imperfect situations: short recording time and inhomogeneous source distribution. The raw input $\mathbf{x}$ of the neural network is virtual shot gathers

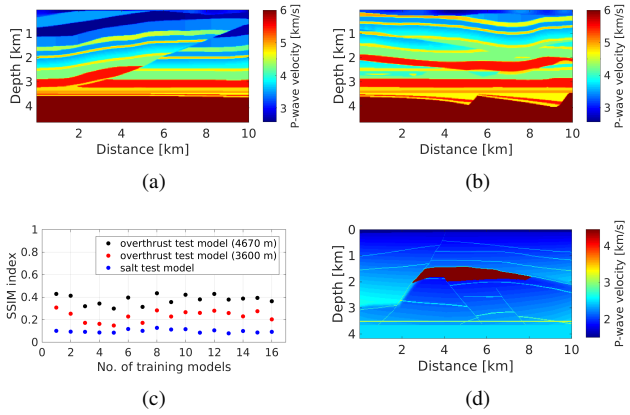


Fig. 1. Training and test models: (a) Overthrust test model. (b) One training model extacted from the 3D Overthrust model. In total, we have 16 training models. (c) Average SSIM index between the test models with all the training models. Black and red points show the comparison of the whole overthrust test model and the overthrust test model with a depth of less than 3.6 km, respectively. Blue points denote the comparsion of the salt test model. (d) Salt test model.

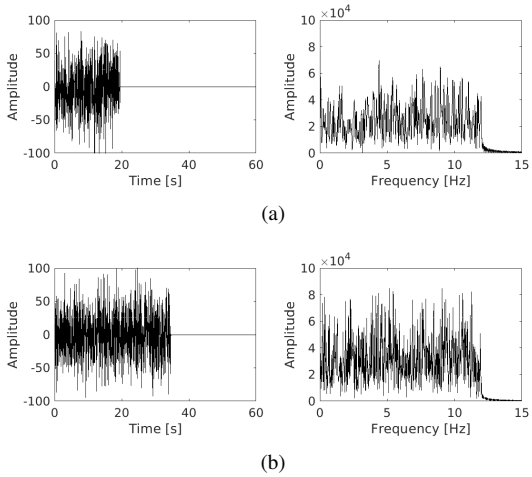


Fig. 2. Time series and its amplitude spectrum of two noise sources. The amplitude and duration of these noise sources are random, but the maximum frequency is 12 Hz. They are randomly triggered at random locations during the total modeling time.

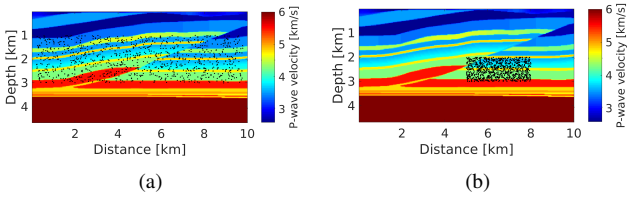


Fig. 3. Locations of the 1000 random noise sources in the rectangular region of (a) 200 m $\leq x \leq$ 9800 m and 1000 m $\leq z \leq$ 3000 m, and (b) 5000 m $\leq x \leq$ 8000 m and 2000 m $\leq z \leq$ 3000 m on the test model. In (a), the noise sources are assumed to be uniformly distributed in the subsurface. In (b), the noise sources are only localized in one part of the subsurface.

(correlograms) and the active shot gathers serve as the desired output (label $\mathbf{y}$).

1) Training and test data sets: In this work, we collect the training data set from a synthetic data set simulated on several 2D models extracted from the 3D SEG/EAGE Overthrust Model [53]. In total, we have 16 training models. Each model has a dimension of 10 km in the horizontal direction and 4.68 km in the vertical direction. To evaluate the performance of the neural network with a test data set, we extract one 2D model from the 3D SEG/EAGE Overthrust Model with structures different from the training models and use it as an overthrust test model (Fig. 1(a)) We use the structural similarity image metric (SSIM, [54]) to measure the similarity between a test model and a training model. Although it could be negative, the SSIM formula typically computes a continuous number between 0.0 and 1.0, where 1.0 corresponds to identical images and 0.0 corresponds to completely dissimilar images. Supplementary Fig. 1 shows the map of SSIM index between test models and the training model in Fig. 1(b). The average SSIM index between the overthrust test model and all the training models is 0.38 for the whole model and 0.23 for the model without the same high velocity layer below 3.6 km (Fig. 1(c)). The low similarity guarantees the suitability of the test dataset simulated on the overthrust test model for evaluating the trained neural networks. Furthermore, we test the neural network on a 2D salt test model (Fig. 1(d)) extracted from the 3D SEG/EAGE Salt Model to understand its generalizability over totally different geological structures. Another salt model (Supplementary Fig. 2) serves as the validation model.

We follow the method in [55] to simulate the ambient noise recordings. The wavefields are simulated by solving the 2D acoustic wave equation in the time domain using the finite difference method. The density of the training and test models are provided based on the P-wave velocity using the Gardner's relation [56]. A perfectly matched layer (PML) is applied to the bottom, right and left boundary of each model, while a free surface condition is applied to the top of the model. With a sampling rate of 8 ms, we record 300 s passive data on the test model and 300 s passive data on the training model. During the total recording time, 1000 noise sources are triggered with random starting time at random locations. Fig. 2 shows the amplitude spectra and the time series of two random noise sources. The amplitude and duration of these noise sources are random, but the maximum frequency of each noise source is 12 Hz. Here we choose to work on relatively low frequency band for two reasons: (1) Real ambient noise used for seismic interferometry generally shows low frequency characteristics [15], [18], [57]. (2) To efficiently synthesize training data sets, we use a relatively large grid spacing in forward modeling, so lower frequencies are preferred to avoid numerical dispersion.

We consider two source configurations to simulate the test datasets on the test model. In one case, the 1000 random noise sources are placed in the rectangular region of 200 m $\leq x \leq$ 9800 m and 1000 m $\leq z \leq$ 3000 m (Fig. 3(a)). The noise sources are assumed to be uniformly distributed in the subsurface. In the other case, the noise sources are only localized in one part of the subsurface. The 1000 random noise sources are placed in the rectangular region of 5000 m $\leq x \leq$ 8000 m and 2000 m $\leq z \leq$ 3000 m (Fig. 3(b)). To reduce the generalization gap, an ideal training data set should have the same source distribution as the test data set. However,

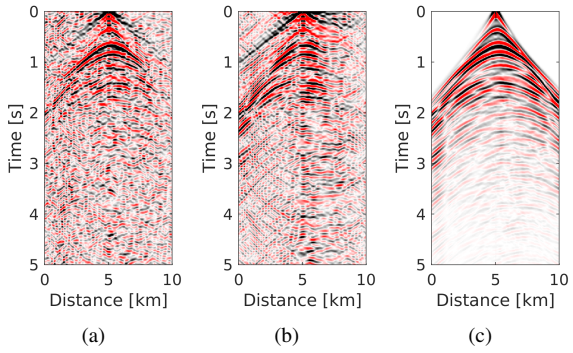


Fig. 4. Comparison of virtual shot gathers (passive correlograms) by correlating noise recordings from (a) homogeneously and (b) inhomogeneously distributed sources, and (c) the active shot gather. The reference trace of the virtual shot gathers is located at $x = 5000$ m, which is also the position of the active source. Due to the short duration of the ambient noise, the virtual shot gather only shows good agreement with the active one with two-way traveltime up to 2 s at the near offset. Traces at the far offset have extremely low SNR and are very difficult to distinguish between physical events and noise. In (b), due to the inhomogeneously distributed sources, the virtual shot gather is dominated by the direction of the energy flux. A limited number of physical events are visible on the virtual shot gather.

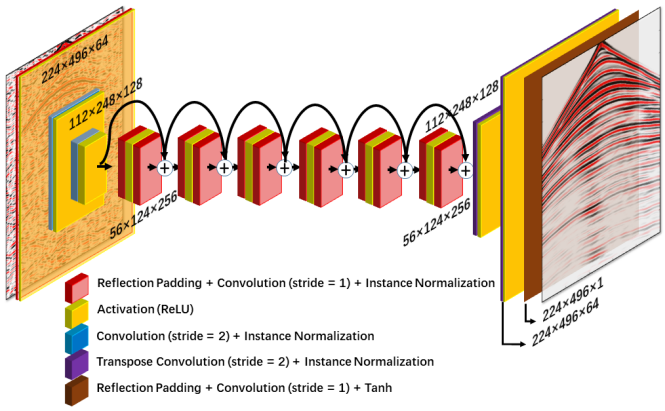


Fig. 5. Deep learning architecture (example for $nt = 496$ and $ntr = 224$). The network takes a virtual shot gather as input and transfers it to match the active shot gather with the same shot location. Each box between the input and output corresponds to a multichannel feature map. The operation on each layer is listed at the bottom of the architecture. The dimension of each feature map and the number of channels are provided around the box. For example, $224 \times 496 \times 64$ denotes that there are 64 feature maps with a size of 224×496. The arrows denote the flow of different operations.

the source configuration of the test dataset is usually unknown in practice. Therefore, we determine the source directionality of the passive recordings using a beamforming method, and use the estimated source locations to simulate the training datasets when the sources are localized in the subsurface.

Supplementary Fig. 3 shows 5 s passive recording measured by 251 receivers evenly placed on the top of the overthrust test model. The first receiver is located at 520 m and the interval between two receivers is 40 m. Fig. 4(a) and Fig. 4(b) show the virtual shot gather calculated by cross-correlating the noise recordings with the reference trace located at $x = 5000$ m. Here we cut the 5-minute noise recordings into five 1-minute time series without overlap (i.e., $N = 5$). Then the five measurements are stacked after cross-correlation for the virtual shot gathers. Fig. 4(c) shows the active shot gather where the active source is located at $x = 5000$ m on the test model. Throughout the paper, we use a Ricker wavelet with dominant frequency of 6 Hz to serve as the active source. The peak of the wavelet is placed at $t = 0$ s to make sure that the phases of the active shot gathers coincide with signals on the virtual shot gathers. In addition, both the active and virtual shot gathers have roughly the same maximum frequency.

2) Neural network architecture: We formulate reflection retrieval from ambient noise as an image-to-image translation problem using fully convolutional neural networks. In our application, the input domain contains 2D images of the virtual shot gather and the target domain is a collection of 2D images of the active shot gather. We train the neural network using a training set of aligned image pairs. Both input $\mathbf{x}$ and output $\mathbf{y}$ have equivalent dimensions of $nt \times ntr$ where nt and ntr are the numbers of recording points and receivers, respectively. The desired output is the active shot gather with the same shot location as the input $\mathbf{x}$.

Fig. 5 summarizes the neural network architecture (example for $nt = 496$ and $ntr = 224$). The deep learning model sequentially contains one convolutional block, two downsamping blocks, six residual blocks, two upsampling blocks and another convolutional block as the output layer. Instead of pooling layers, we use a convolution with a stride of two for downsamping. Likewise, the upsampling block employs a transposed convolutional layer with a stride of two. The kernel size is the same for all the convolutional layers and we have experimented with a kernel size of 3×3, 7×7 and 11×11. In addtion, we pad the input tensors by reflecting the input values across the border axis (Reflection Padding) before the convolution operation. We use instance normalization [58] following all the convolutional layers except the last one. We use a rectified linear unit (ReLU) for activation layers, except the last layer where a hyperbolic tangent function (Tanh) is used to output the retrieved shot gathers in $[-1, 1]$.

Building a deep learning model with the residual block [43] is proposed to ease the training of networks. Each residual block has two connections from its input, one going through a series of operations, such as convolutions, normalizations and activation functions, and the other simply performing identity mapping, and their outputs are added together as the input of the next residual block. The identity mapping between layers of the network allows large sections to be skipped if needed. Although the residual framework is expected to gain accuracy from considerably increased depth, we experimentally find that 9 residual blocks provide similar performance to 6 blocks, but 3 residual blocks are not sufficient.

3) Loss function for training: We train the neural network in a supervised fashion and let the output of the neural network be near the ground truth in an L_1 sense. Here L_1 is preferable as it encourages less blurring than L_2 for the image-to-image translation problem [48]. The supervised loss can be expressed as

$$\mathcal{L}_{L_1}(N_\theta) = \mathbb{E}_{\mathbf{x}\backsim p(\mathbf{x}),\mathbf{y}\backsim p(\mathbf{y})}[\|N_\theta(\mathbf{x}) - \mathbf{y}\|_1], \tag{2}$$

where N_θ denotes the neural network parameterized by θ. $\mathbf{x} \backsim$

$p(\mathbf{x})$ and $\mathbf{y} \backsim p(\mathbf{y})$ are the data distribution. Then, our final objective is

$$N_\theta^* = \arg\min_{N_\theta} \mathcal{L}_{L_1}(N_\theta). \tag{3}$$

During training, we use adaptive moment estimation (Adam, [59]) to update N_θ with a mini-batch of one. The learning rate is set as 2×10^{-5}.

III. Numerical experiments

In this section, we demonstrate the performance of the neural network for reflection retrieval in two imperfect situations of correlation-based SI: short recording length and inhomogeneous source distribution.

In the first case, we aim to retrieve reflections from short passive recordings for the application of SI for real-time monitoring of the subsurface. The total length is only 300 s for passive recordings in the training set and 300 s for the test set. In this case, the noise sources are assumed to be uniformly distributed in the subsurface (Fig. 3(a)). Fig. 4(a) and Fig. 4(c) compare the retrieved common-shot gather and the active-source common-shot gather where the virtual/active source is located at $x = 5000$ m. Due to the short recording time, the virtual shot gather only shows good agreement with the active one with two-way traveltime up to 2 s at the near offsets. Traces at far offsets have extremely low SNR so that physical events and noise are indistinguishable.

The second case aims to retrieve correct reflections under the imperfect condition of inhomogeneous source distribution. In addition to the same short recording length as the first case, the noise sources are localized in the small rectangular region of 5000 m $\leq x \leq$ 8000 m and 2000 m $\leq z \leq$ 3000 m (Fig. 3(b)) and, thus, not optimal for the correlation-based SI. Comparing with the active shot gather (Fig. 4(c)), the virtual shot gather (Fig. 4(b)) is dominated by the direction of the energy flux. A limited number of physical events are visible on the virtual shot gather retrieved by correlation, due to the directional wavefield.

The subsequent preprocessing steps are followed to generate each pair of images in the training and test data sets:

- The input $\mathbf{x}$ is a virtual short gather (correlograms);
- The desired output $\mathbf{y}$ (label) is the active shot gather generated by placing an active source at the position of the virtual source. Here we use a Ricker wavelet with 6 Hz dominate frequency and 12 Hz maximum frequency as the active source;
- Both the active and virtual shot gathers should be band-pass filtered to equalize the frequency components. Here both the active and virtual shot gathers have roughly the same frequency components so we omit this step;
- We taper the shot gathers with zero at the beginning of each time series using a time window of 0.16 s to remove the extreme large amplitude at the positions of the virtual/active sources;
- We normalize the input $\mathbf{x}$ and label $\mathbf{y}$ to $[-1, 1]$;

After preprocessing, we use each pair of the virtual and active shot gathers with the same shot location as the 2D images for the input and target of the neural network. Specifically, with the time dimension of $nt = 496$ and the distance dimension of $ntr = 224$, the shot gather to be processed is 3.96 s in time and $8920m$ in distance, respectively. As a result, for each case, we have a total of 3584 training image pairs (16 training models $\times$ 224 shots per model), 448 test image pairs (2 test models $\times$ 224 shots per model) and 224 validation image pairs (1 validation models $\times$ 224 shots per model). Furthermore, we retain the effect of the wavelet on the virtual and active shot gathers during training. Consequently, deconvolution of the source wavelet from the prediction of the neural networks is required to retrieve Green's functions from reflections. The source wavelet can be approximated by the Ricker wavelet used in the simulation of the active-source common shot gather.

A. Case 1: near real-time recording length

1) Results on the overthrust test model: We evaluate the performance of the neural network (NN) using the training and test data sets simulated under the case of near real-time recording length. In this case, we use the same source configurations (evenly distributed) to simulate the passive recordings on both training and test models. Supplementary Fig. 4 shows the learning curve of the neural network with a kernel size of 11×11 after training with five epochs. According to the performance on the validation data set, we choose to use the neural network trained after four epochs to predict the resulting shot gathers.

Fig. 6 compares the virtual shot gather, NN predicted and active shot gather where the virtual or active shot is located at $x = 6$ km on the test model. Overall, the predicted events show good coherence among different traces. Compared with the correlograms in the virtual shot gather, the neural network removes most of the spurious events and correlation noise. Although events at earlier arrival time are more accurate than those at the later arrival time, we notice that the NN prediction is quite comparable to the active shot gather. The evaluation metrics (SSIM and MAE) between the predicted and active shot gathers are improved compared with those between the virtual and active shot gathers (Table I).

A crucial hyperparameter that affects the performance is the kernel size of the convolutional operator. Here we compare the resulting shot gathers predicted by the neural network with a kernal size of 3×3 (Fig. 6(c)), 7×7 (Fig. 6(d)) and 11×11 (Fig. 6(e)). The comparision shows that the neural network with a larger kernel size provides better results, due to a larger receptive field [60]. However, a large kernel size also requires increasing memory and computing cost, so we end up choosing a kernel size of 11×11 in the following work.

Fig. 7 compares the traces extracted from the virtual shot gather, the resulting shot gather predicted by the neural network with a kernel size of 11×11, and the active shot gather in Fig. 6. Five traces located at 3, 4, 5, 6 and 7 km on the overthrust test model are plotted from each shot gather. Although minor event discrepancy exists between the predicted and the active shot gather, we find good agreement

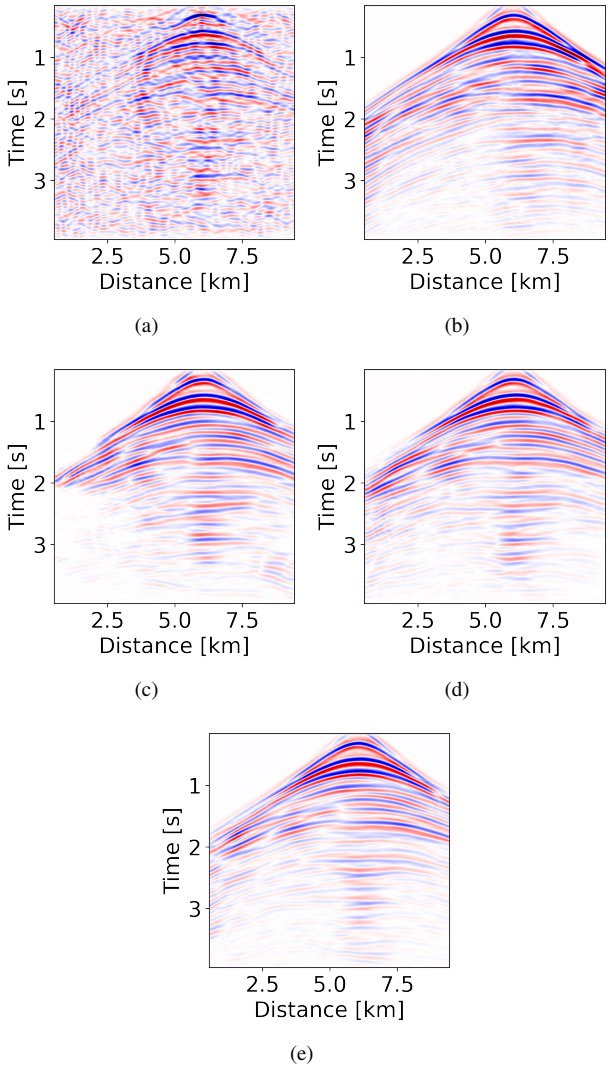


Fig. 6. Comparison of shot gathers under the case of *near real-time recording length*. (a) Virtual shot gather. (b) Active shot gather. The resulting shot gathers predicted by the neural network with a kernel size of (c) 3×3, (d) 7×7 and (e) 11×11. The virtual/active source is located at $x = 6$ km on the overthrust test model.

for most events. Compared with the traces on the passive correlograms, NN prediction has more visible physical events and is much cleaner. However, due to the larger geometric spreading of body waves, events at larger offet (distance between source and receiver) have lower SNR and show increasing prediction errors.

2) Seismic imaging: We migrate the virtual, predicted and active shot gathers to evaluate the performance of the passive recordings for seismic imaging. The normalized virtual and active shot gathers as well as the prediction without post-processing are directly used as the input for migration, because the overall amplitude of the shot gathers is not meaningful for seismic imaging purpose. 112 shots with an interval of 80 m are migrated using the phase shift plus interpolation (PSPI) method [61] shot-by-shot and then summed up for the final imaging results. The migration input model is a smoothing version of the overthrust test model (Supplementary Fig. 5 in the supplementary file). Any details below 1 km are removed

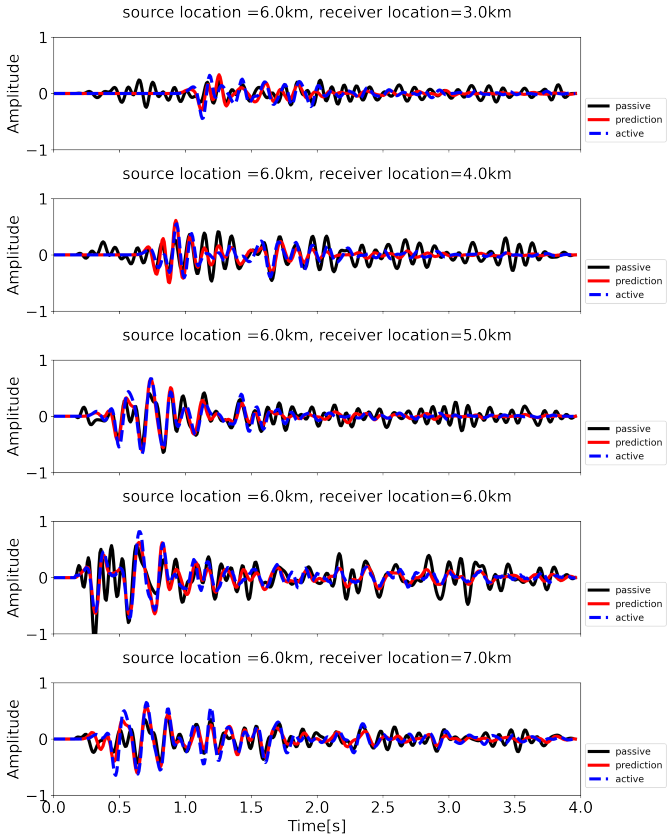


Fig. 7. Comparison of individual traces from the overthrust test model among the virtual (passive correlograms), active and resulting shot gathers predicted by the neural network with a kernel size of 11×11, under the case of *near real-time recording length*. The locations of the source and receiver are, respectively, labeled on the top of each panel. The accuracy decreases with increasing distance between source and receiver.

TABLE I
EVALUATION METRICS BETWEEN THE LISTED AND ACTIVE SHOT GATHERS ON THE *overthrust test model*[a]

shot gather	**SSIM**		**MAE**	
	single[b]	average[c]	single[b]	average[c]
passive (virtual)	0.192	0.159	0.089	0.103
predicted (case 1)	0.459	0.460	0.048	0.050
predicted (case 1, noisy label)	0.421	0.419	0.051	0.053
predicted (case 2, similar direction)	0.420	0.396	0.056	0.062
predicted (case 2, different direction)	0.395	0.374	0.060	0.064

[a] All the resulting shot gathers are predicted from the neural network with a kernel size of 11×11.
[b] This column shows the metric of the single shot located at $x = 6$ km.
[c] This column shows the average of all shot gathers.

from the migration input model.

Fig. 8 compares the prestack depth migration results using the virtual, predicted, and active shot gathers. The imaging result using the predicted shot gathers is comparable with that using the active shot gathers and is much cleaner compared to that using the raw virtual shot gathers. Note that reflectors below 3.5 km on the imaging results are the artifacts due to multiples since they are not removed from the shot gathers before migraion. However, the imaging result using predicted gathers shows good agreement with that using active shot gath-

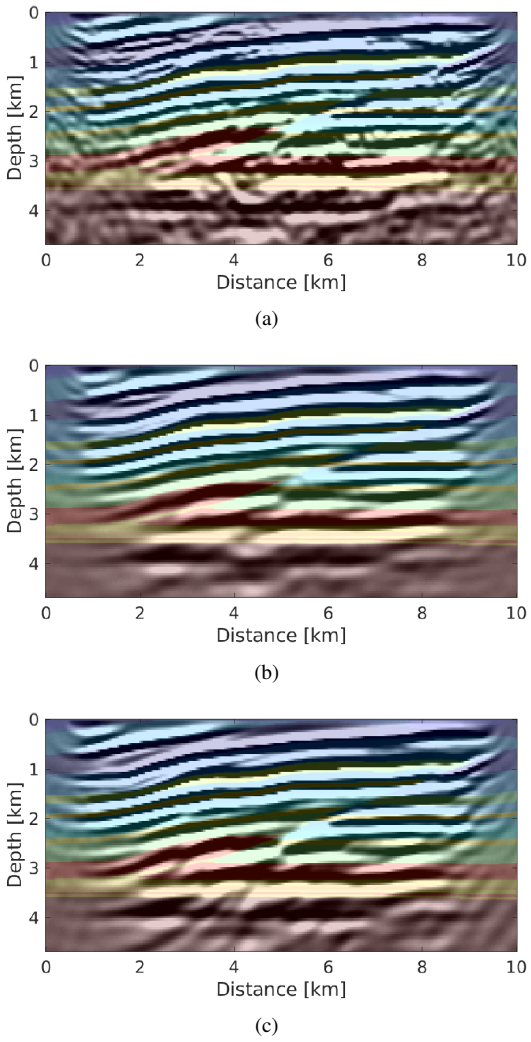


Fig. 8. Comparison of prestack depth migration results using (a) virtual, (b) predicted, and (c) active shot gathers. The passive recordings are simulated under the case of *near real-time recording length*. (The data comparison for one shot is plotted in Fig. 6.) The imaging result using the predicted shot gathers is comparable with that using the active shot gathers and is much less noisy than that using the raw virtual shot gathers.

ers. The similar pattern indicates that the neural network learns the mapping from virtual to active shot gathers successfully.

3) Robustness with random noise on labels: To evaluate the robustness of the method with random noise on labels, we add Gaussian white noise with mean $\mu = 0$ and standard deviation $\sigma = 0.5$ to the active shot gathers in the training data set. For comparison, we also add the same amount of random noise to the active shot gather in the test data set (Fig. 9(a)). Fig. 9(b) and Table I show the results predicted by the neural network with a kernel size of 11×11. Although discontinuity of events at far offset exists, the neural network is robust to random noise. However, the effect of other types of noise remains to be investigated [62], [63].

B. Case 2: inhomogeneous source distribution

We then evaluate the reliability of the neural networks to overcome the spatial limitation on noise source distribution by training the neural network with another training and test

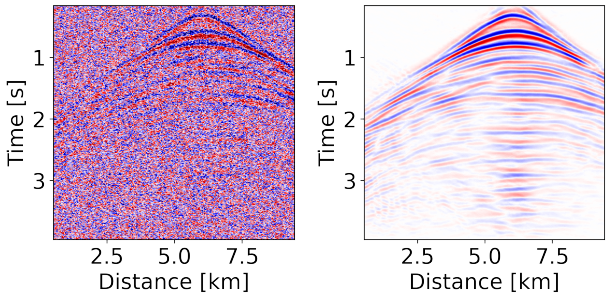


Fig. 9. Robustness when training with noisy labels. Comparison between (a) noisy active shot gather located at 6 km on the overthrust test model and (b) resulting shot gather predicted by the neural network with a kernel size of 11×11. The neural network is trained with active shot gathers contaminated with Gaussian white noise with mean $\mu = 0$ and standard deviation $\sigma = 0.5$.

data sets simulated under the case of inhomogeneous source distribution. In this case, the ambient noise recordings are dominated by the flux coming from the direction of noise sources. As seen in the comparison between Fig. 4(b) and Fig. 4(c), the virtual shot gather from cross correlation shows strong artifacts along with the direction of the energy flux. Although the special conditions of noise sources are necessary for correlation-based SI to be applicable, they are not necessary in order to have a relation between the Greens function and the cross correlation of ambient noise recordings [64]. The neural networks with proper setup can find this hidden relationship after being trained to extract Green's functions from correlograms based on events from an active survey.

1) Beamforming for source directionality: When the source distribution is inhomogeneous, the prior knowledge of the source location can be estimated from the passive recordings to collect a training data set with similar source direction as the test data set. Among the methods to estimate the angle of arrival, we choose MUltiple SIgnal Classification (MUSIC, [65]) for its higher resolution compared with classical beamforming approaches. This paper only considers the estimation of a single quantity, the angle of the leading direction of incidence of the noise, to then inform the training step of the network by providing noise distributions calibrated on this single parameter. We do not attempt to match the complete beam pattern, although this could be an interesting direction for future research.

On the 2D models with receivers placed on the surface, the azimuth of the sources relative to the center of the receiver array ranges from 0° to 180°. To satisfy the far-field approximation of the beamformer, we use recordings from only five receivers in the middle of the receiver array to calculate the pseudo-power of the recordings. Recordings from 16 s to 64 s are used to estimate the source directionality in the band of $0.1 - 12.0$ Hz.

Fig. 10 compares the source location on the test model, two designed source locations on each training model, and the corresponding beamforming results. On each model, the center of the maximum pseudo-power is roughly consistent with the true angle of the center of the source area indicated

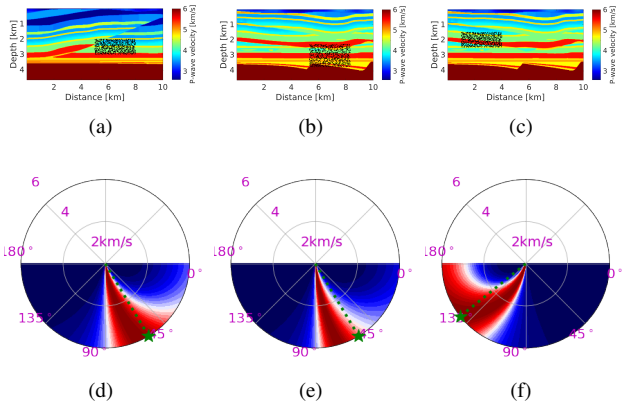


Fig. 10. Comparison of (a) source locations on the test model, (b) first choice and (c) second choice of source locations on the training model in Fig. 1(b) for simulating training datasets; (d)-(f) beamforming results of the noise recordings with source locations in (a)-(c), respectively. The green star indicates the true angle of the center of the source area in each panel.

by the green star. On the test model, the beamforming result (Fig. 10(d)) reveals that the source direction ranges from $45°$ to $90°$. The estimated source direction can be used to simulate a training dataset with a similar source direction as the test dataset. We randomly place 1100 sources in the rectangular region of $5330 \text{ m} \leq x \leq 8330 \text{ m}$ and $2300 \text{ m} \leq z \leq 3800$ m on each training model. The neural network is expected to achieve good performance when trained using the designed source configuration (Fig. 10(b)). For comparison, we put sources in another region of each training model (Fig. 10(c)) to investigate the importance of the prior knowledge of the source distribution. The beamforming result (Fig. 10(f)) also shows that the source direction is very different from that on the test model.

2) Results on the overthrust test model: Supplementary Fig. 6 shows the learning curves of the neural networks trained on the data set simulated using the estimated source direction (Fig. 10(b)) and a source direction different from the test data set (Fig. 10(c)). Fig. 11 compares the virtual, active, and resulting shot gathers predicted by the neural network trained with the two source directions in Fig. 10. Since the noise sources are localized in the subsurface and not suitable for correlation-based SI, only few physical events can be seen on the correlograms. However, the prediction generally shows good agreement with the active shot gather. As we would expect, the neural network trained under the situation of inhomogeneous source distribution predicts more errors than that trained with homogeneous sources. However, compared with the correlograms in the imperfect case of localized sources, the neural network still predicts reflections matching the active shot gather with enough degree of accuracy.

By comparison, Fig. 11(d) shows the results on the test model when the training data set is simulated without the prior knowledge of the source locations. Beamforming result shows that the angle of source direction ranges from $120°$ to $150°$ on the training models, but the angle ranges from $45°$ to $90°$ on the test models. In this case, the neural network generalizes worse and creates more artifacts on events at far

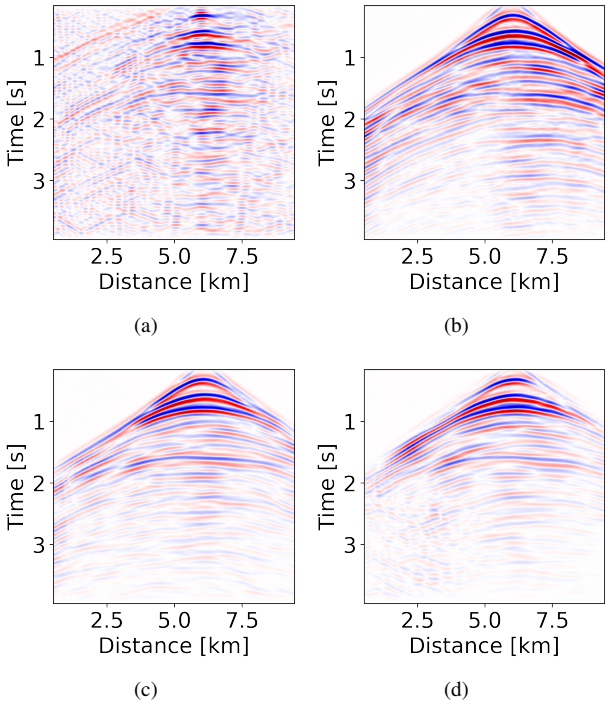
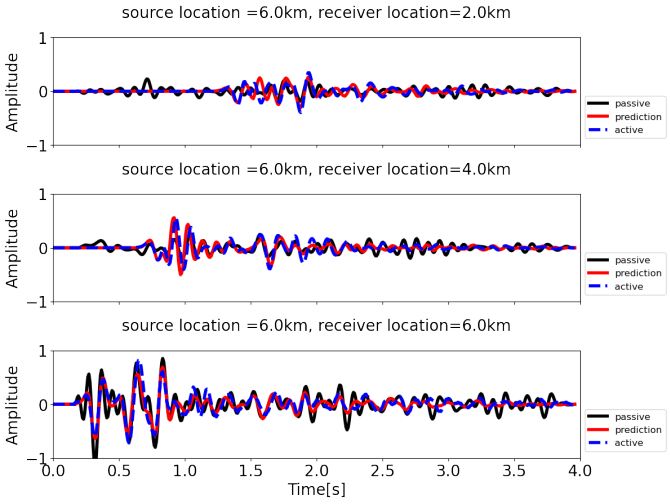


Fig. 11. Comparison of shot gathers under the case of *inhomogeneous source distribution*. (a) Virtual shot gather. (b) Active shot gather. The resulting shot gathers predicted by the neural network with a kernel size of 11×11 trained with (c) similar source direction to and (d) very different source direction from the test data set. The virtual or active source is located at $x = 6$ km on the overthrust test model.

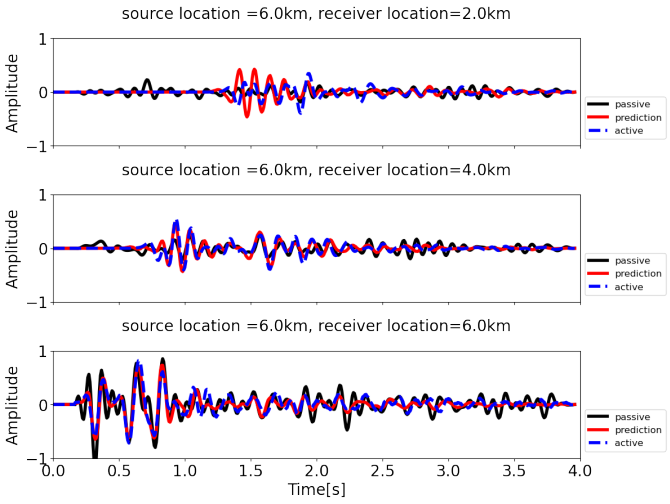
offset (Fig. 12). Although it is helpful to train the neural network with estimated source direction, an exact knowledge of the source positions may not be necessary regarding the acceptable degree of mismatch between prediction and its active reference (Table I).

3) Seismic imaging: Fig. 13 compares the prestack depth migration results using the passive recordings simulated under the case of inhomogeneous source distribution. The input data to migration in Fig. 13(a) are the virtual shot gathers. In contrast, the input data to migration in Fig. 13(b) and Fig. 13(c) are predicted by the neural network trained with the similar source direction (Fig. 10(b)) and the very different source direction (Fig. 10(c)), respectively. We see that, compared with imaging using the virtual shot gathers from a uniformly distributed sources, imaging with the localized source distribution is cleaner. However, the left area without sources is not illuminated and thus is not imaged by either type of input data. Imaging with the predicted shot gathers from the similar source direction removes the high frequency artifacts on the virtual shot gathers. Most reflectors on the image in Fig. 13(b) are comparable to the one using the active shot gathers (Fig. 8(c)). However, as we would expect, it is harder to train the neural network for finding the correct relationship between the virtual shot gathers and the active ones when the source is inhomogeneously distributed. As a result, the imaging quality of Fig. 13(b) is degraded compared with the imaging in Fig. 8(b).

Fig. 13(c) shows the imaging results using the shot gathers

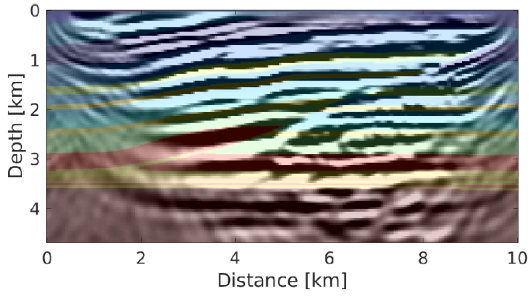


(a)

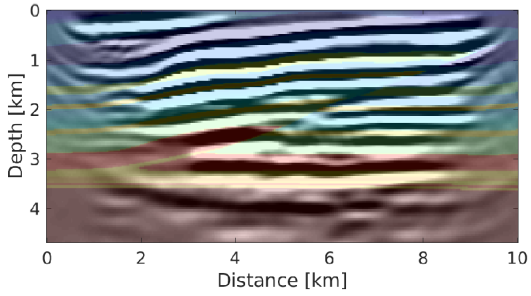


(b)

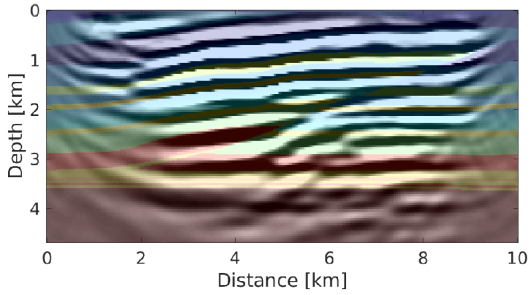
Fig. 12. Comparison of individual traces from the overthrust test model among the virtual (passive correlograms), active and resulting shot gathers predicted by the neural network with a kernel size of 11×11, under the case of *inhomogeneous source distribution*. (a) Training data set is simulated with a source direction similar to the estimated source direction from the test data set. (b) Source direction for the training data set is very different from the test dataset.



(a)



(b)



(c)

Fig. 13. Comparison of prestack depth migration results using passive recordings simulated under the case of *inhomogeneous source distribution*. (a) Migration input data are the virtual shot gathers. (b) Migration input data are predicted by the neural network trained with the similar source direction. (One shot of the prediction is plotted in Fig. 11(c).) (c) Migration input data are predicted by the neural network trained with the very different source direction. (One shot of the prediction is plotted in Fig. 11(d).)

predicted by the neural network trained with the different source locations (Fig. 10(c)). Although the overall imaging shows the layered structure of the subsurface, several major reflectors are curved up on the bottom-left part of the image, probably due to the imperfect prediction of events at far offset (Fig. 12(b)). The comparison of the imaging results in Fig. 13(b) and Fig. 13(c) demonstrates that the prior knowledge of the estimated source locations is helpful when dealing with passive recordings from a localized source distribution.

C. Generalization from overthrust to salt structures

Finally, we investigate the generalizability of the neural network from overthrust to salt structures. Fig. 14(a) and 14(b) show the virtual shot gathers on the salt test model calculated using ambient noise simulated under the two cases mentioned earlier. Fig. 14(c) shows the active shot gather as reference. Using the virtual shot gather in Fig. 14(a) as the input, Fig. 14(d) shows the result predicted by the neural network trained on the data set simulated on the overthrust training

TABLE II
EVALUATION METRICS BETWEEN THE LISTED AND ACTIVE SHOT GATHERS ON THE *salt test model*[a]

shot gather	**SSIM**		**MAE**	
	single[b]	average[c]	single[b]	average[c]
passive (virtual)	0.152	0.134	0.093	0.093
predicted (case 1)	0.454	0.468	0.045	0.045
predicted (case 1, noisy label)	0.419	0.391	0.048	0.049
predicted (case 2, similar direction)	0.411	0.403	0.058	0.056
predicted (case 2, different direction)	0.417	0.394	0.055	0.055

[a] All the resulting shot gathers are predicted from the neural network with a kernel size of 11×11.
[b] This column shows the metric of the single shot located at $x = 6$ km.
[c] This column shows the average of all shot gathers.

models under the case of *near real-time recording length*. Table II quantitatively measures the accuracy of the resulting shot gathers. Fig. 14(e) and 14(f) show the results predicted by the neural network trained under the case of *inhomogeneous source distribution*. In either case, although several main reflectors are well recovered, the direct arrival and events with small reflection coefficient are poorly retrieved. Furthermore, seismic imaging results (Supplementary Fig. 7 and 8) show that the generalization error is increased when training on the overthrust and testing on the salt model. However, the neural networks are still able to retrieve reflections with some degree of accuracy (Table II). To improve the accuracy, we may collect more training shot gathers simulated with some salt structures to improve the generalizability, or use transfer learning strategy.

DISCUSSION

The neural networks trained in the case of near real-time recording length can be seen as a denoising operator, which removes correlation noise and spurious events from correlograms. Hence, reflections with high SNR can be easily detected and recovered from correlograms by the neural network. In our numerical example, we use only 300 s passive noise recordings to generate the correlograms by correlation. The SNR of most physical events is extremely low. However, our result shows that the neural networks can reconstruct most of the physical events with satisfactory accuracy. To further improve the performance of the neural networks, we can use correlograms from noise measurements with longer recording time to enhance the SNR of physical events before feeding into the neural network.

As for the case of inhomogeneous source distribution, the neural network is not simply a denoising operator, because few physical events can be seen from the correlograms. The neural networks are trained to find a hidden relationship between the correlograms and Green's functions and use that relationship to reconstruct correct reflections. In this paper, we use MUSIC beamforming to estimate the source location in the subsurface. The estimated source directionality serves as the prior knowledge for simulating a training dataset with a similar source direction as the test dataset. The error in the estimation of the source directionality may bring its error in the predicted shot gathers if the neural network has difficulty generalizing among different source locations. However, by looking at the imaging result, we find that stacking of different shots may eliminate these errors and provide a roughly reasonable imaging result.

The theory of correlation-based SI is strictly valid for the media without attenuation. Although wavefields are assumed not to attenuate in most treatments of SI, spurious events will be visible on correlograms if the medium is dissipative. In this work, we do not deal with the imperfect situation of attenuation, but it would make sense to collect the training and test data sets using the viscoacoustic scheme of the wave equation and train the neural networks to remove the spurious events due to attenuation with a supervised learning framework.

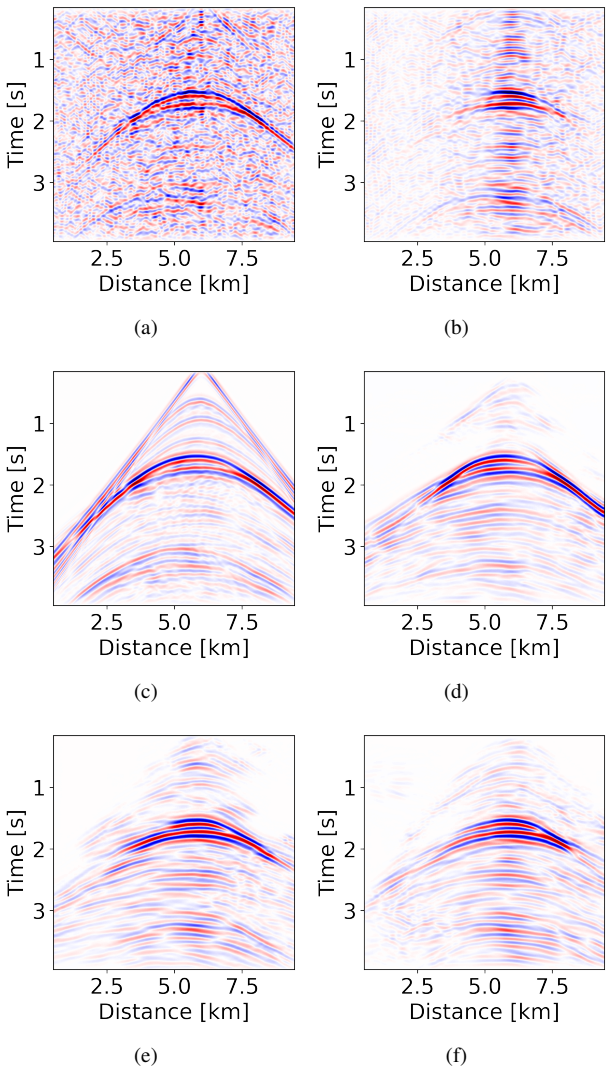


Fig. 14. Generalization from overthrust to salt structures. Virtual shot gather under the case of (a) *near real-time recording length* and (b) *inhomogeneous source distribution*. (c) Active shot gather. (d)-(f) The resulting shot gathers predicted by the neural network with a kernel size of 11×11. The virtual or active source is located at $x = 6$ km on the salt test model. In (d), the resulting shot gather is predicted using (a). In (e) and (f), the resulting shot gathers are predicted using (b) with the neural network trained with (e) similar source direction to and (f) very different source direction from the test data set.

The proposed method can be generalized to seismic interferometry on regional and global scales. However, the problems for regional or global scale are mainly focused on the difficulty of complete Green's function retrieval, since the surface wave part of the Green's function is much easier to reconstruct from ambient noise. Compared to the surface waves, the body waves have weaker energy, a more transparent (less scattered) property and faster attenuation. Nevertheless, this method may be used to relieve the unrealistic requirement of randomly-distributed sources for the application of SI in regional or global scale.

For the application to field data, we may rely on synthetic data for training and then consider to improve its generalizability from synthetic to field data. Alternatively, we may choose

to train the neural network with real passive and active shot pairs. How to collect a training data set containing both passive and active shot gathers at the same location may be a problem. For a purpose of monitoring, the passive data are recorded continuously with days to measure the seismic response of the subsurface. Then the active shots may be collected intermittently during monitoring. For example, if the ambient noise data are recorded by ocean bottom node, we may use the streamer data (may need a preprocessing of redatuming from ocean surface to floor) as the active shot reference. If the active shot gathers are unavailable, we may adopt architecture dealing with unlabeled real data, which is beyond the scope of this work. Moreover, instead of reflections, surface waves would dominate the Green's functions retrieved from real ambient noise data.

Finally, solving the problems of SI by deep learning will inevitably introduce issues of deep learning. Although we empirically find that the neural networks enjoy sufficient generalizability, the unreliability of the prediction will exist when the training data set is insufficient. Also, how the neural network learns and predicts is still difficult to unravel. Although the neural network trained in the case of short recordings may perform an operation of denoising, the hidden relationship found in the case of directional sources is still unexplained.

IV. Conclusion

We propose to use supervised learning to handle the practical challenges of seismic interferometry in realistic situations. By substituting ambient noise for an active source, we train deep neural networks to reconstruct the reflections from correlograms. To overcome the temporal limitation of the noise recording length, the neural networks are trained on correlograms using 300 s passive data to retrieve reflections from 300 s passive measurements. To overcome the spatial limitation of source distribution, the neural networks are trained on noise data generated from an inhomogeneous source distribution (directional wavefield). Numerical examples demonstrate that deep learning can help with seismic interferometry to extract reasonably accurate signals from ambient noise in realistic situations. Instead of simply estimating travel times or image reflectors from ambient noise, the retrieved reflections may be used in seismic imaging with the full Green's function. The new technology may help with near real-time monitoring of the Earth's dynamics in a wide ranges of areas, for example, urban environments.

Acknowledgment

The authors thank Zhilong Fang from MIT and Elita Li from Purdue University for helpful discussions. Tensorflow and Keras are used for deep learning.

References

[1] G. Schuster, J. Yu, J. Sheng, and J. Rickett, "Interferometric/daylight seismic imaging," *Geophysical Journal International*, vol. 157, no. 2, pp. 838–852, 2004.

[2] K. Wapenaar and J. Fokkema, "Green's function representations for seismic interferometry," *Geophysics*, vol. 71, no. 4, pp. SI33–SI46, 2006.

[3] K. Aki, "Space and time spectra of stationary stochastic waves, with special reference to microtremors," *Bulletin of the Earthquake Research Institute*, vol. 35, pp. 415–456, 1957.

[4] H. Yao, C. Beghein, and R. D. van der Hilst, "Surface wave array tomography in SE Tibet from ambient seismic noise and two-station analysis-II: Crustal and upper-mantle structure," *Geophysical Journal International*, vol. 173, no. 1, pp. 205–219, 2008.

[5] J. F. Claerbout, "Synthesis of a layered medium from its acoustic transmission response," *Geophysics*, vol. 33, no. 2, pp. 264–269, 1968.

[6] O. I. Lobkis and R. L. Weaver, "On the emergence of the Green's function in the correlations of a diffuse field," *The Journal of the Acoustical Society of America*, vol. 110, no. 6, pp. 3011–3017, 2001.

[7] P. Roux and M. Fink, "Green's function estimation using secondary sources in a shallow water environment," *The Journal of the Acoustical Society of America*, vol. 113, no. 3, pp. 1406–1416, 2003.

[8] K. Wapenaar, "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," *Physical Review Letters*, vol. 93, no. 25, p. 254301, 2004.

[9] R. Snieder, K. Wapenaar, and K. Larner, "Spurious multiples in seismic interferometry of primaries," *Geophysics*, vol. 71, no. 4, pp. SI111–SI124, 2006.

[10] A. Curtis, P. Gerstoft, H. Sato, R. Snieder, and K. Wapenaar, "Seismic interferometry - Turning noise into signal," *The Leading Edge*, vol. 25, no. 9, pp. 1082–1092, 2006.

[11] V. C. Tsai, "The relationship between noise correlation and the Green's function in the presence of degeneracy and the absence of equipartition," *Geophysical Journal International*, vol. 182, no. 3, pp. 1509–1514, 2010.

[12] A. Fichtner, "Source and processing effects on noise correlations," *Geophysical Journal International*, vol. 197, no. 3, pp. 1527–1531, 2014.

[13] R. Snieder, K. Van Wijk, M. Haney, and R. Calvert, "Cancellation of spurious arrivals in Green's function extraction and the generalized optical theorem," *Physical Review E*, vol. 78, no. 3, p. 036606, 2008.

[14] D. Draganov, K. Wapenaar, W. Mulder, J. Singer, and A. Verdel, "Retrieval of reflections from seismic background-noise measurements," *Geophysical Research Letters*, vol. 34, no. 4, 2007.

[15] D. Draganov, X. Campman, J. Thorbecke, A. Verdel, and K. Wapenaar, "Reflection images from ambient seismic noise," *Geophysics*, vol. 74, no. 5, pp. A63–A67, 2009.

[16] I. Vasconcelos and R. Snieder, "Interferometry by deconvolution: Part 1 - Theory for acoustic waves and numerical examples," *Geophysics*, vol. 73, no. 3, pp. S115–S128, 2008.

[17] K. Wapenaar, J. van der Neut, and E. Ruigrok, "Passive seismic interferometry by multidimensional deconvolution," *Geophysics*, vol. 73, no. 6, pp. A51–A56, 2008.

[18] N. Nakata, R. Snieder, T. Tsuji, K. Larner, and T. Matsuoka, "Shear wave imaging from traffic noise using seismic interferometry by cross-coherence," *Geophysics*, vol. 76, no. 6, pp. SA97–SA106, 2011.

[19] R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," *Physical Review E*, vol. 69, no. 4, p. 046610, 2004.

[20] N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface-wave tomography from ambient seismic noise," *Science*, vol. 307, no. 5715, pp. 1615–1618, 2005.

[21] A. Curtis and D. Halliday, "Directional balancing for seismic and general wavefield interferometry," *Geophysics*, vol. 75, no. 1, pp. SA1–SA14, 2010.

[22] P. Zhang, L.-G. Han, Q. Liu, Y.-H. Zhang, and X. Chen, "Interpolation of seismic data from active and passive sources and their joint migration imaging," *Chinese Journal of Geophysics - Chinese Edition*, vol. 58, no. 5, pp. 1754–1766, 2015.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[24] M. V. de Hoop, M. Lassas, and C. A. Wong, "Deep learning architectures for nonlinear operator functions and nonlinear inverse problems," *arXiv preprint arXiv:1912.11090*, 2019.

[25] Y. Khoo and L. Ying, "SwitchNet: a neural network model for forward and inverse scattering problems," *SIAM Journal on Scientific Computing*, vol. 41, no. 5, pp. A3182–A3201, 2019.

[26] H. Sun and L. Demanet, "Low frequency extrapolation with deep learning," in *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, 2018, pp. 2011–2015.

[27] ——, "Extrapolated full-waveform inversion with deep learning," *Geophysics*, vol. 85, no. 3, pp. R275–R288, 2020.

[28] W. Zhu, S. M. Mousavi, and G. C. Beroza, "Seismic signal denoising and decomposition using deep neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9476–9488, 2019.

[29] X. Dong, Y. Li, and B. Yang, "Desert low-frequency noise suppression by using adaptive dncnns based on the determination of high-order statistic," *Geophysical Journal International*, vol. 219, no. 2, pp. 1281–1299, 2019.

[30] X. Dong and Y. Li, "Denoising the optical fiber seismic data by using convolutional adversarial network based on loss balance," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 12, pp. 10 544–10 554, 2020.

[31] M. Araya-Polo, J. Jennings, A. Adler, and T. Dahlke, "Deep-learning tomography," *The Leading Edge*, vol. 37, no. 1, pp. 58–66, 2018.

[32] Y. Wu and Y. Lin, "InversionNet: An efficient and accurate data-driven full waveform inversion," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 419–433, 2019.

[33] Z. Zhang and Y. Lin, "Data-driven seismic waveform inversion: A study on the robustness and generalization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 6900–6913, 2020.

[34] J. Clancy, L. Demanet, J. Helland, and Z. Xu, "Deep learning for making sense of ambient seismic noise," in *AGU Fall Meeting Abstracts*. American Geophysical Union, 2018, pp. S13B–03.

[35] P. Bharadwaj, M. Li, and L. Demanet, "SymAE: An autoencoder with embedded physical symmetries for passive time-lapse monitoring," in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 1586–1590.

[36] L. Viens and C. Van Houtte, "Denoising ambient seismic field correlation functions with convolutional autoencoders," *Geophysical Journal International*, vol. 220, no. 3, pp. 1521–1535, 2020.

[37] X. Zhang, Z. Jia, Z. E. Ross, and R. W. Clayton, "Extracting dispersion curves from ambient noise correlations using deep learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8932–8939, 2020.

[38] L. Viens and T. Iwata, "Improving the retrieval of offshore-onshore correlation functions with machine learning," *Journal of Geophysical Research: Solid Earth*, vol. 125, no. 8, p. e2020JB019730, 2020.

[39] L. Seydoux, R. Balestriero, P. Poli, M. De Hoop, M. Campillo, and R. Baraniuk, "Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning," *Nature Communications*, vol. 11, no. 1, pp. 1–12, 2020.

[40] S. Jakkampudi, J. Shen, W. Li, A. Dev, T. Zhu, and E. R. Martin, "Footstep detection in urban seismic data with a convolutional neural network," *The Leading Edge*, vol. 39, no. 9, pp. 654–660, 2020.

[41] G. Binder and A. Tura, "Convolutional neural networks for automated microseismic detection in downhole distributed acoustic sensing data and comparison to a surface geophone array," *Geophysical Prospecting*, vol. 68, no. 9, pp. 2770–2782, 2020.

[42] A. L. Stork, A. F. Baird, S. A. Horne, G. Naldrett, S. Lapins, J.-M. Kendall, J. Wookey, J. P. Verdon, A. Clarke, and A. Williams, "Application of machine learning to microseismic event detection in distributed acoustic sensing data," *Geophysics*, vol. 85, no. 5, pp. KS149–KS160, 2020.

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[45] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[46] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[47] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[48] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[49] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[50] H. Sun and L. Demanet, "Seismic interferometry with neural networks," in *First International Meeting for Applied Geoscience & Energy*. Society of Exploration Geophysicists, 2021, pp. 1384–1389.

[51] K. Wapenaar, "Green's function retrieval by cross-correlation in case of one-sided illumination," *Geophysical Research Letters*, vol. 33, no. 19, 2006.

[52] J. Thorbecke and K. Wapenaar, "Analysis of spurious events in seismic interferometry," in *SEG Technical Program Expanded Abstracts 2008*. Society of Exploration Geophysicists, 2008, pp. 1415–1420.

[53] F. Aminzadeh, N. Burkhard, J. Long, T. Kunz, and P. Duclos, "Three dimensional SEG/EAEG models - An update," *The Leading Edge*, vol. 15, no. 2, pp. 131–134, 1996.

[54] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[55] J. W. Thorbecke and D. Draganov, "Finite-difference modeling experiments for seismic interferometry," *Geophysics*, vol. 76, no. 6, pp. H1–H18, 2011.

[56] G. H. F. Gardner, L. W. Gardner, and A. R. Gregory, "Formation velocity and density - The diagnostic basics for stratigraphic traps," *Geophysics*, vol. 39, no. 6, pp. 770–780, 1974.

[57] N. Nakata, J. P. Chang, J. F. Lawrence, and P. Boué, "Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry," *Journal of Geophysical Research: Solid Earth*, vol. 120, no. 2, pp. 1159–1173, 2015.

[58] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[60] H. Sun and L. Demanet, "Deep learning for low-frequency extrapolation of multicomponent data in elastic FWI," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, p. 5909511, 2022.

[61] J. Gazdag and P. Sguazzero, "Migration of seismic data by phase shift plus interpolation," *Geophysics*, vol. 49, no. 2, pp. 124–131, 1984.

[62] J. Frank, U. Rebbapragada, J. Bialas, T. Oommen, and T. C. Havens, "Effect of label noise on the machine-learned classification of earthquake damage," *Remote Sensing*, vol. 9, no. 8, p. 803, 2017.

[63] D. Karimi, H. Dou, S. K. Warfield, and A. Gholipour, "Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis," *Medical Image Analysis*, vol. 65, p. 101759, 2020.

[64] J. Garnier and G. Papanicolaou, *Passive imaging with ambient noise*. Cambridge University Press, 2016.

[65] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

Hongyu Sun (Graduate Student Member, IEEE) received the bachelor's and master's degrees in geophysics from Jilin University, Changchun, China, in 2014 and 2017, respectively. She is currently pursuing the Ph.D. degree in geophysics with the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA.

Her research interests range from computational to experimental studies of the Earth, including, but not limited to, machine learning, seismic imaging, seismic interferometry, inverse problems, signal processing, and geomechanics.

Laurent Demanet received the bachelor's degree in mathematical engineering and theoretical physics from the Université catholique de Louvain (UCLouvain), Ottignies-Louvain-la-Neuve, Belgium, and the Ph.D. degree in applied and computational mathematics, under Emmanuel Candès, from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 2006.

He is currently a Professor of applied mathematics with the Department of Mathematics, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. He is also the Director of the Earth Resources Laboratory, MIT. His research interests include signal processing, inverse problems, and wavepropagation.

Dr. Demanet was a recipient of the Sloan Research Fellowship, the CAREER Award from the National Science Foundation (NSF), and the Young Investigator Award from Air Force Office of Scientific Research (AFOSR).

Supplementary Information for "Beyond Correlations: Deep Learning for Seismic Interferometry"

Hongyu Sun, *Graduate Student Member, IEEE*, and Laurent Demanet

I. The difference between seismic interferometry (SI) and deconvolution

THE goal of this work is interferometry, which is substantially harder than deconvolution. The goal of blind deconvolution is to find the source wavelet $w(t)$ and the Green's function $G(x_r, x_s, t)$ between the source x_s and the receiver x_r given data $d_{r,s,t} = \int G(x_r, x_s, t-t')w(t')dt'$. In contrast, the goal of interferometry is to determine the Green's function $G(x_{r_A}, x_{r_B}, t)$ between two receivers x_{r_A} and x_{r_B} from ambient noise data $d_{r,s,t} = \int G(x_r, y, t-t')f(y,t')dt'dy$ where the noisy source signature $f(y,t') = \sum_\alpha \delta(y - x_\alpha)w_\alpha(t')$. With the seismic interferometry result $G(x_{r_A}, x_{r_B}, t)$, one of the receivers x_{r_A} or x_{r_B} is considered as the "virtual source" in seismic exploration.

Reasons that we are doing interferometry but not deconvolution include: (1) We have multiple sources (1000 sources in our numerical examples) at random locations in the subsurface whereas deconvolution is used to deal with recordings from one source at a *fixed location*. (2) The sources are triggered at random but nearby starting times, resulting in simultaneous wave propagation in the medium of interest. (3) There is no access to a theorem linking the expectation of the deconvolution in the general case to the Green's function $G(x_{r_A}, x_{r_B}, t)$ between the receivers x_{r_A} and x_{r_B}. In this regard, it is unclear that deconvolution is even better than cross-correlation.

As a result, it is only tangentially relevant to compare our method with any of the deconvolution approaches (such as Bharadwaj *et al.* [1]). In particular, Bharadwaj *et al.* [1] can be used for retrieval of Green's functions, but in a more contrived context: (1) Their paper performs a blind deconvolution in the presence of a *single* (point) noisy source. (2) Their paper is nowhere close to approaching the topic of virtual sources. By contrast, our work focuses on Green's function (of a virtual source) retrieval from ambient noise excited by multiple noisy sources at other locations.

There is one seismic interferometry approach performed by deconvolution in the frequency domain [2]:

$$D_{AB} = \frac{u(x_{r_A}, x_s)}{u(x_{r_B}, x_s)} = \frac{G(x_{r_A}, x_s)}{G(x_{r_B}, x_s)} = \frac{G(x_{r_A}, x_s)G^*(x_{r_B}, x_s)}{\mid G(x_{r_B}, x_s) \mid^2}. \tag{1}$$

Here the deconvolution result D_{AB} between two recordings $u(x_{r_A}, x_s)$ and $u(x_{r_B}, x_s)$ has the same phase information as the cross-correlation but the noisy source information is canceled out compared with cross-correlation. However, it suffers from the same problems as the correlation-based interferometry in terms of source distribution and recording length, as we discussed in the main text.

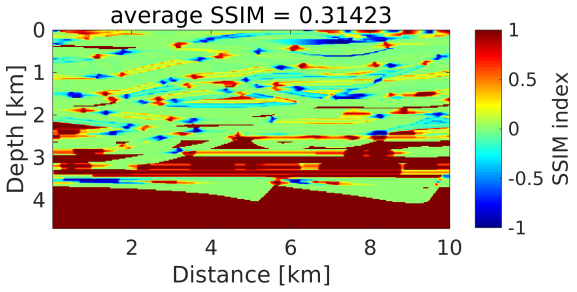
The authors are with the Massachusetts Institute of Technology, Cambridge, MA 02139, USA. (e-mail: hongyu-sun@outlook.com; laurent@math.mit.edu) *(Corresponding author: Hongyu Sun.)*

II. The choice of deep-learning model for SI

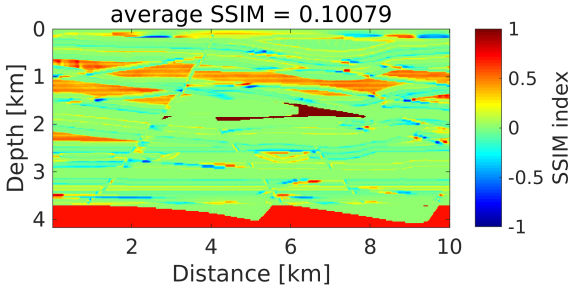
Green's function retrieval (of a virtual shot gather) in our experimental setup has been considered as an image-to-image translation problem. The input domain contains 2-D images of virtual shot gathers and the target domain is a collection of 2-D images of active shot gathers. The input and the output have equivalent dimensions: the numbers of recording points and receivers.

The paper is not testing the performance of various neural networks on the seismic interferometry problem but introducing deep learning and proposing a pipeline to show how to solve the problem using deep learning. The choice of the architecture is of course non-unique. A few classic neural networks can potentially be used to solve the image-to-image translation problem. People generally use convolutional neural networks (CNNs, such as ResNet [3]) and generative adversarial networks (GANs, such as Pix2Pix [4] and CycleGAN [5]) for this type of problem. (CycleGAN has been used for supervised training with an explicit supervised loss if the paired data are available. Then, the role of cycle-consistency loss can be considered as a double constraint for the supervised training.) However, VAE [6] or DCGANs [7] may not be good matches with our problem setup, because they are more suitable for image generation. Compared with image generation, the problem of image-to-image translation has no latent variables and is more constrained. Moreover, although the training of VAE is easier than that of GAN, VAE always generates poorer images than GAN, because the penalty of GAN is to generate realistic images [4]. Furthermore, there is no particular rationale for using residual neural network (RNN) in this context, because RNN is generally used for exhibiting temporal dynamic behavior.

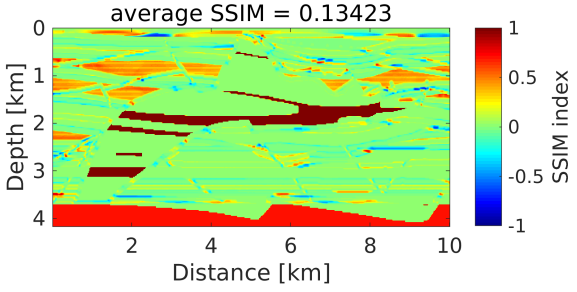
We compare the performance among a modified ResNet (Fig. 5), Pix2Pix and CycleGAN with and without an explicit supervised loss for paired images on this problem. For the GANs, the architecture of the generators is the same as that of the modified ResNet (Fig. 5). We employ the discriminator in Zhu *et al.* [5] as the discriminator architecture in our comparison. The output of the generator is a shot gather,



(a)



(b)

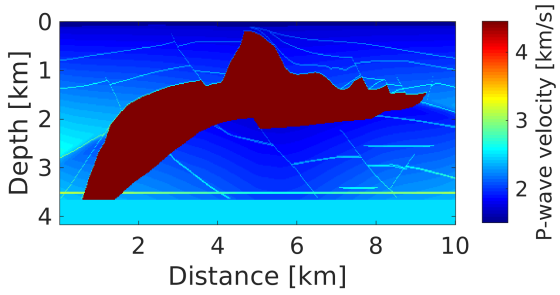


(c)

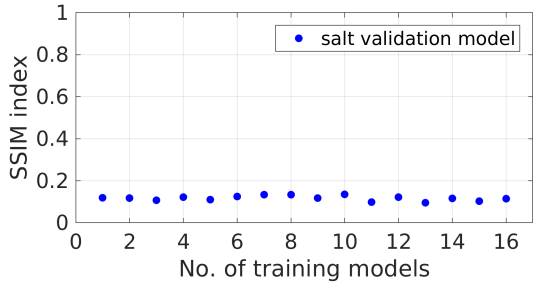
Supplementary Fig. 1. The map of SSIM index between the training model (Fig. 1(b)) and (a) the overthrust test model (Fig. 1(a)), (b) the salt test model (Fig. 1(d)), and (c) the validation model (Supplementary Fig. 2(a)).

whereas the output of the discriminator is considered as probabilities. For the loss calculated based on shot gathers, such as the cycle-consistency loss and the identity loss, we use L_1 norm, because it can prevent blur compared with L_2 norm. For the loss calculated based on probabilities, we use L_2 norm to train it under a framework of least-square GAN [8]. Although all the neural networks that we compared above seem to work for this problem, the modified ResNet is the best in terms of accuracy and efficiency among the deep-learning models that we investigated. Hence, we choose to use the modified ResNet.

In summary, ResNet is a good match with our problem for the following reasons: (1) We have paired data and would like to solve the problem using a supervised framework; (2) Compared with GANs, training a CNN, such as ResNet is easier and more efficient. Therefore, we chose to use a CNN with an L_1 loss (the modified ResNet) as the deep-learning model in the main text. It is much more efficient and can provide sufficient results for our problem. CycleGAN may be used in the future to deal with unlabeled real data on top of simulations.

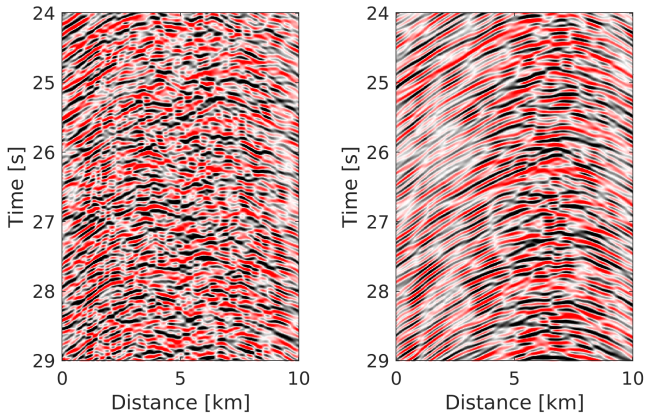


(a)



(b)

Supplementary Fig. 2. (a) The validation model. (b) The average SSIM index between the validation model and all the training models. In total, we have 16 training models extracted from the 3D SEG/EAGE Overthrust Model.



(a) (b)

Supplementary Fig. 3. The simulated noise recordings with 1000 random noise sources in the rectangular region of (a) 200 m $\leq x \leq$ 9800 m and 1000 m $\leq z \leq$ 3000 m, and (b) 5000 m $\leq x \leq$ 8000 m and 2000 m $\leq z \leq$ 3000 m on the overthrust test model. The total recording time is 300 s while only 5 s is shown here. In (a), the noise sources are uniformly distributed in the subsurface. In (b), the noise sources are localized in the subsurface (inhomogeneous source distribution).

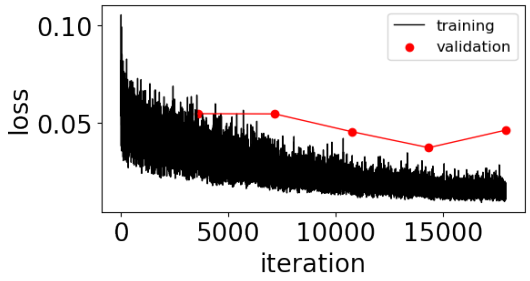
REFERENCES

[1] P. Bharadwaj, L. Demanet, and A. Fournier, "Focused blind deconvolution," *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3168–3180, 2019.

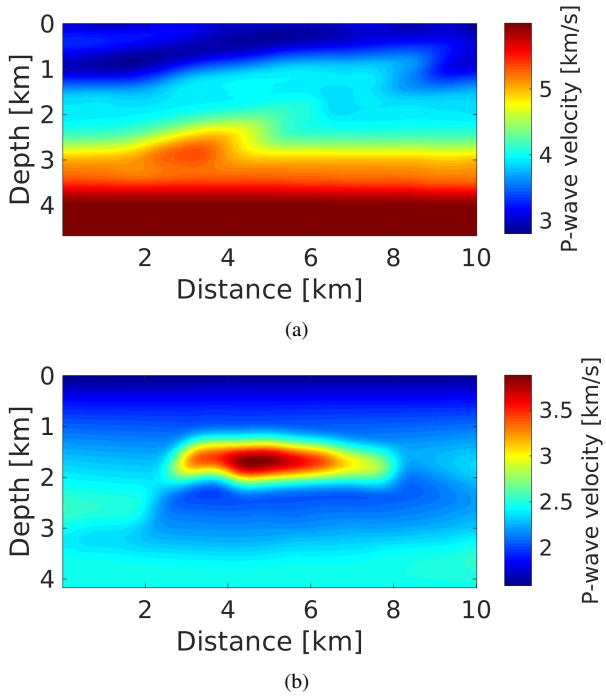
[2] I. Vasconcelos and R. Snieder, "Interferometry by deconvolution: Part 1-Theory for acoustic waves and numerical examples," *Geophysics*, vol. 73, no. 3, pp. S115–S128, 2008.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

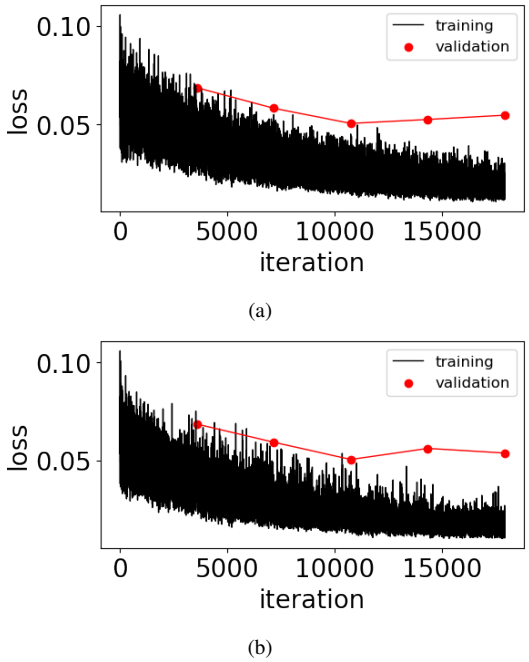
[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE*



Supplementary Fig. 4. The learning curves of the neural network with a kernel size of 11×11 trained under the case of *near real-time recording length*. According to the performance on the validation data set, we choose the neural network trained with four epochs to predict the resulting shot gathers.



Supplementary Fig. 5. Migration input model. The input velocity models for the seismic imaging results on (a) the overthrust test model and (b) the salt test model.



Supplementary Fig. 6. The learning curves of the neural network with a kernel size of 11×11 trained under the case of *inhomogeneous source distribution*. The training data sets are respectively simulated with (a) similar source direction to and (b) very different source direction from the test data set.

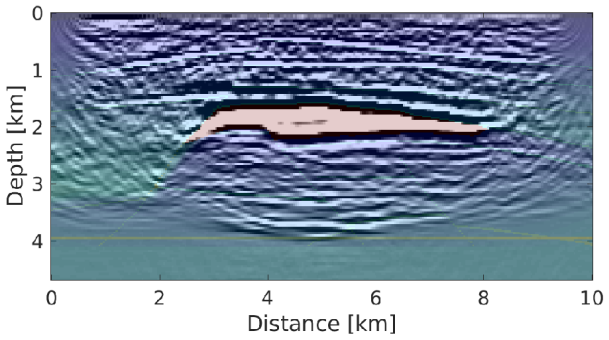
conference on computer vision and pattern recognition, 2017, pp. 1125–1134.

[5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

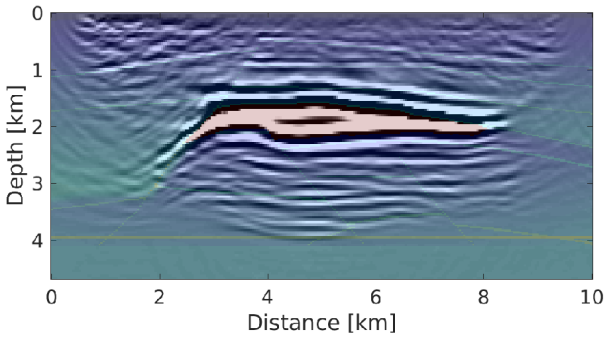
[6] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[7] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

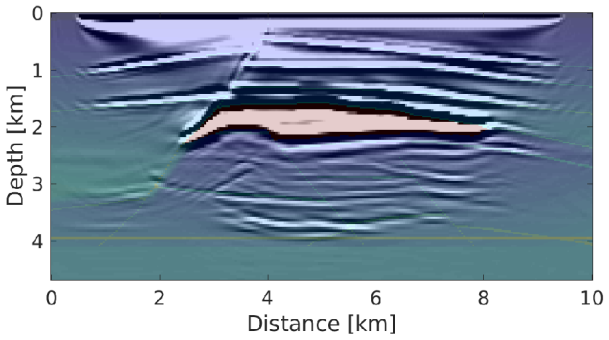
[8] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.



(a)

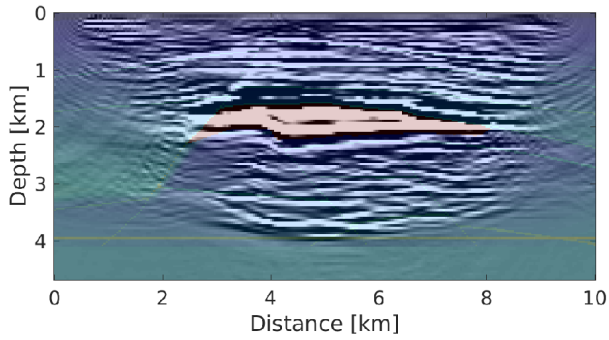


(b)

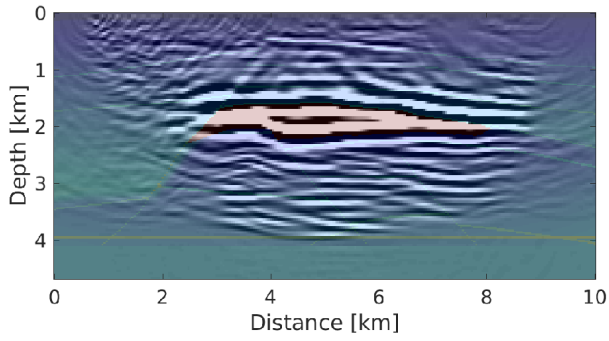


(c)

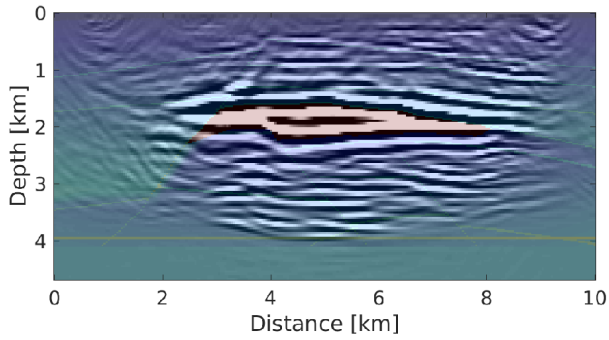
Supplementary Fig. 7. Seismic imaging result on the salt test model under the case of *near real-time recording length*. Comparison of prestack depth migration results using (a) virtual shot gathers (Fig. 14(a)), (b) predicted shot gathers (Fig. 14(d)), and (c) active shot gathers (Fig. 14(c)). The data comparison for one shot is plotted in Fig. 14. The imaging result using the predicted shot gathers is comparable with that using the active shot gathers and is much less noisy compared to that using the raw virtual shot gathers.



(a)



(b)



(c)

Supplementary Fig. 8. Seismic imaging result on the salt test model under the case of *inhomogeneous source distribution*. In (a), the migration input data are the virtual shot gathers (Fig. 14(b)). In (b), the migration input data are predicted by the neural network trained with the similar source direction (Fig. 14(e)). In (c), the migration input data are predicted by the neural network trained with the very different source direction (Fig. 14(f)). The data comparison for one shot is plotted in Fig. 14.