GEOPHYSICS, VOL. 84, NO. 5 (SEPTEMBER-OCTOBER 2019); P. 1–11, 18 FIGS.
10.1190/GEO2018-0635.1

White noise attenuation of seismic trace by integrating variational mode decomposition with convolutional neural network

Hao Wu[1], Bo Zhang[1], Tengfei Lin[2], Fangyu Li[3], and Naihao Liu[4]

ABSTRACT

Seismic noise attenuation is an important step in seismic data processing. Most noise attenuation algorithms are based on the analysis of time-frequency characteristics of the seismic data and noise. We have aimed to attenuate white noise of seismic data using the convolutional neural network (CNN). Traditional CNN-based noise attenuation algorithms need prior information (the "clean" seismic data or the noise contained in the seismic) in the training process. However, it is difficult to obtain such prior information in practice. We assume that the white noise contained in the seismic data can be simulated by a sufficient number of user-generated white noise realizations. We then attenuate the seismic white noise using the modified denoising CNN (MDnCNN). The MDnCNN does not need prior clean seismic data nor pure noise in the training procedure. To accurately and efficiently learn the features of seismic data and band-limited noise at different frequency bandwidths, we first decomposed the seismic data into several intrinsic mode functions (IMFs) using variational mode decomposition and then apply our denoising process to the IMFs. We use synthetic and field data examples to illustrate the robustness and superiority of our method over the traditional methods. The experiments demonstrate that our method can not only attenuate most of the white noise but it also rejects the migration artifacts.

INTRODUCTION

Seismic noise attenuation is a key step to enhance the quality of seismic data. Seismic denoising not only lowers the effects of subjectivity in seismic interpretation but also improves the reliability of seismic inversion. In recent decades, numerous seismic denoising approaches have been developed and widely applied in practice. Methods for seismic denoising can be generally classified into four categories: The first category is based on building a prediction filter to remove the noise of seismic data. The commonly used algorithms within the first category include f-x predictive filtering (Canales, 1984), t-x predictive filtering (Abma and Claerbout, 1995), the forward-backward prediction approach (Wang, 1999), the polynomial fitting-based approach (Liu et al. 2011), and nonstationary predictive filtering (Liu et al., 2012). The second category projects the seismic data to a transformed domain and rejects the noise by applying a band-pass filter to the transformed data. The denoised seismic data are obtained by projecting the filtered data back to the time domain. The commonly used algorithms within the second category include the Fourier transform (Chen and Ma, 2014), curvelet transform (Herrmann et al., 2007), seislet transform (Fomel and Liu, 2010), shearlet transform (Kong and Peng, 2015), Radon transform (Trad et al., 2002; Xue et al., 2016), wavelet transform (Donoho and Johnstone, 1994), and dictionary learning-based sparse transform (Elad and Aharon, 2006). The third category decomposes the seismic traces into a set of components and then examines the time or frequency features of each decomposed component. Finally, we can obtain the clean seismic traces by rejecting the components that are regarded as "noise." The commonly used algorithms within the third category include empirical mode decomposition (Huang et al., 1998; Bekara and van der Baan, 2009),

Manuscript received by the Editor 30 August 2018; revised manuscript received 1 June 2019; published ahead of production 27 June 2019.
[1]University of Alabama, Department of Geological Science, Tuscaloosa, Alabama, USA. E-mail: hwu43@crimson.ua.edu; bzhang33@ua.edu (corresponding author).
[2]CNPC, Research Institute of Petroleum Exploration and Development, Department of Middle East E&P, Beijing, China. E-mail: lintf@petrochina.com.cn.
[3]University of Georgia, College of Engineering, Athens, Georgia, USA. E-mail: fangyu.li@uga.edu.
[4]Xi'an Jiaotong University, School of Electronic and Information Engineering, Xi'an, China. E-mail: lnhfly@163.com.

variational mode decomposition (VMD) (Dragomiretskiy and Zosso, 2014; Li et al., 2017), and singular-value decomposition-based approaches (Bekara and van der Baan, 2007). Yuan et al. (2018) propose a novel inversion-based denoising method. The method has the advantages of preserving 3D spatial edges and low-frequency signals. The fourth category is based on the rank-reduction reconstruction of seismic data. The commonly used algorithms within the fourth category include Cadzow filtering (Trickett, 2008) and singular spectrum analysis (Vautard et al., 1992; Oropeza and Sacchi, 2011).

Deep learning is a subset of machine learning that is based on learning data representation with multiple levels of abstraction (LeCun et al., 2015). The convolutional neural network (CNN) (LeCun et al., 1998) is one of the most popular and widely used deep-learning algorithms. CNN-based algorithms have already achieved great success in the field of computer vision. CNN is extremely efficient in learning the features of the images and labeling images. Numerous CNN-based algorithms have also been proposed to address the problem of image denoising. Jain and Seung (2009) successfully apply CNN to image denoising. Burger et al. (2012) denoise the images using multiple layer perceptron. Other popular CNN-based image denoising methods include the stacked sparse denoising autoencoder (Xie et al., 2012) and the trainable nonlinear reduction diffusion model (Chen et al., 2015). Zhang et al. (2017) propose denoising CNN (DnCNN) to learn the features of noise contained in the images. The main disadvantage of current CNN-based denoising methods is that these methods need clean data and the corresponding noisy data in the training process. Unfortunately, it is unfeasible to obtain clean seismic data for training in practice.

We propose the modified denoising CNN (MDnCNN), which uses user-generated white noise to simulate the white noise contained in the seismic data. Considering that the white noise contained in the seismic data is band-limited and the noise level is varying with the frequency bandwidth, we combine the VMD with the MDnCNN to accurately learn the feature of noise at different bandwidths. There are four main steps in the proposed workflow. The first step is to decompose the seismic volume into different intrinsic mode functions (IMFs) by using VMD. The second step is to add user-generated white noise to each decomposed component. The third step is to build the neural network hierarchy to learn the feature of additive white noise. The last step is to denoise seismic data by applying the well-trained network to the original seismic data. We use synthetic and field data to illustrate the robustness and superiority of our method over denoising methods such as f-x deconvolution and MDnCNN.

THEORY

There are many successful applications of image denoising by using CNN-based algorithms (Xie et al., 2012; Zhang et al., 2017). The main advantage of CNN-based denoising methods is that a CNN with multiple hidden layers (deep architecture) can recognize various features of the input data and classify them into corresponding categories.

Objective function

One 3D seismic volume can be reshaped into many 2D seismic sections in inline and crossline directions. Each seismic section can be treated as a 2D image. The image needing to be denoised can be defined as $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where $\mathbf{y}$ is the noisy image, $\mathbf{x}$ is the corresponding clean image, and $\mathbf{n}$ is the additive noise. The goal of image denoising is to build a model to recover the clean image $\mathbf{x}$ from the corresponding noisy image $\mathbf{y}$. According to the objective function, the image denoising methods using CNN can be classified into two categories.

The first category is modeling the clean image (Jain and Seung, 2009; Xie et al., 2012) by minimizing the following objective function $\mathbf{J}(\theta)\mathbf{J}(\theta)\mathbf{J}(\boldsymbol{\theta})$:

$$\mathbf{J}(\boldsymbol{\theta}) = \arg\min_{\theta} \frac{1}{M} \sum_{i=1}^{M} \|\mathbf{x}_i - R_\theta(\mathbf{y}_i)\|^2, \tag{1}$$

where $\{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^{M}$ represents the M noisy-clean image pairs, R_θ denotes the entire CNN with all trainable parameters (convolution filter and bias) $\boldsymbol{\theta}$, and $R_\theta(\mathbf{y}_i)$ is the predicted clean image by using the trained CNN R_θ.

The second category is modeling the noise by applying the residual learning formulation (Zhang et al., 2017):

$$\mathbf{J}(\boldsymbol{\theta}) = \arg\min_{\theta} \frac{1}{M} \sum_{i=1}^{M} \|R_\theta(\mathbf{y}_i) - (\mathbf{y}_i - \mathbf{x}_i)\|^2, \tag{2}$$

where $R_\theta(\mathbf{y}_i)$ is the predicted noise by using the trained CNN R_θ.

In seismic exploration, $\mathbf{x}$ and $\mathbf{y}$ can be regarded as the noise-free and noise-contaminated seismic images, respectively. Equations 1 and 2 require clean and corresponding noisy data in the training process. Unfortunately, it is unfeasible to obtain purely clean seismic data in practice. However, we assume that the white noise contained in the seismic data can be simulated by enough user-generated white noise realizations $\mathbf{n}'$ (Wu and Huang, 2009). The seismic image with additive white noise can be expressed as

$$\mathbf{y}' = \mathbf{x} + \mathbf{n} + \mathbf{n}'. \tag{3}$$

The new objective function of MDnCNN is given by

$$\begin{aligned} \mathbf{J}(\boldsymbol{\theta}) &= \arg\min_{\theta} \frac{1}{M} \sum_{i=1}^{M} \|R_\theta(\mathbf{y}_i) - (\mathbf{y}'_i - \mathbf{y}_i)\|^2 \\ \mathbf{J}(\boldsymbol{\theta}) &= \arg\min_{\theta} \frac{1}{M} \sum_{i=1}^{M} \|R_\theta(\mathbf{y}_i) - \mathbf{n}'_i\|^2. \end{aligned} \tag{4}$$

According to the statistical properties of white noise, the distribution of the original white noise $\mathbf{n}$ and the additive white noise $\mathbf{n}'$ are given by (Wu and Huang, 2004)

$$n_i \sim \eta(\mu_d, \sigma_d^2), \qquad n'_i \sim \eta(\mu_n, \sigma_n^2), \tag{5}$$

where η denotes the normal distribution, μ_d is the expectation, σ_d is the standard deviation of the noise contained in the seismic data, μ_n is the expected value, and σ_n is the standard deviation of the additive white noise. Equation 6 illustrates that we can simulate the noise contained in the seismic data if we have sufficient trials. To accurately simulate the noise contained in the seismic data, the noise level of additive white noise should be close to that of the original white noise. We use the peak signal-to-noise ratio (PS/N) method to compute the signal-to-noise ratio (S/N) for the input seismic data:

$$\mathrm{S/N} \approx \mathrm{PS/N} = \frac{\max(\mathbf{y}^2)}{\mathrm{MSE}}, \tag{6}$$

$$\mathrm{MSE} = \frac{1}{M}\sum_{i}^{M}(\mathbf{y}_i - \hat{\mathbf{y}})^2, \tag{7}$$

where $\hat{\mathbf{y}}$ denotes the mean value of the original seismic data and MSE is the mean-squared error (Li et al., 2017).

After adding sufficient additive white noise realizations, the expectation μ_n and standard deviation σ_n of certain realizations of simulated noise should approximately be equal to the expectation μ_d and standard deviation σ_d of noise contained in the seismic data (Wu and Huang, 2004):

$$\mu_n \approx \mu_d, \sigma_n \approx \sigma_d. \tag{8}$$

Then, we obtain

$$\begin{aligned}\mathbf{J}(\boldsymbol{\theta}) &= \arg\min_{\theta} \frac{1}{N}\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{N}\|R_\theta(\mathbf{y}_i) - \mathbf{n}'_{ik}\|^2 \\ &\approx \arg\min_{\theta} \frac{1}{M}\sum_{i=1}^{M}\|R_\theta(\mathbf{y}_i) - \mathbf{n}_i\|^2,\end{aligned} \tag{9}$$

where N is the number of additive white noise realizations and $\mathbf{n}'_{ik}$ represents the ith additive white noise image of the kth additive white noise realization. Equation 9 indicates that the proposed method does not require clean seismic data in the training process.

Architecture

The architecture of the proposed neural network is a sequence of nonlinear processing layers followed by a sigmoid classifier layer based on the architecture of MDnCNN (Figure 1). The input of the network is the original seismic data $\mathbf{y}$ and the seismic data with additive white noise $\mathbf{y}' = \mathbf{y} + \mathbf{n}'$. The network contains 17 layers in total. The first layer contains 64 convolution filters of size 3×3 and 64 rectified linear units (ReLUs) activation operators. The objective of the convolution filter is to generate feature maps of the input seismic data. The objective of ReLU is to activate the main

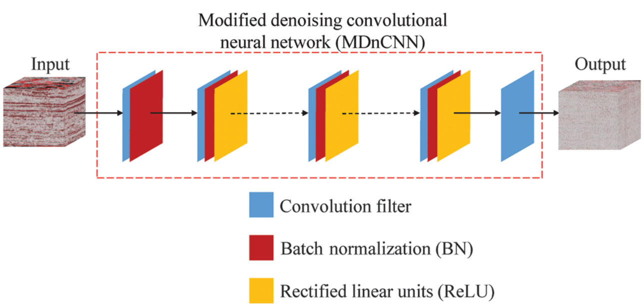


Figure 1. The overall architecture of MDnCNN.

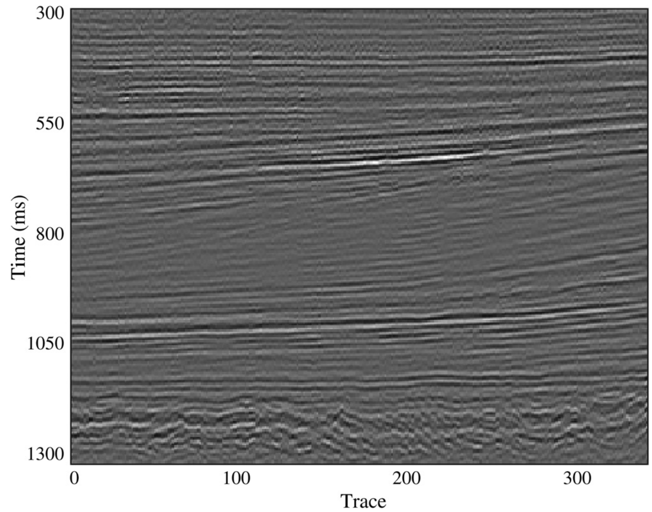


Figure 2. One section example of F3-block seismic survey.

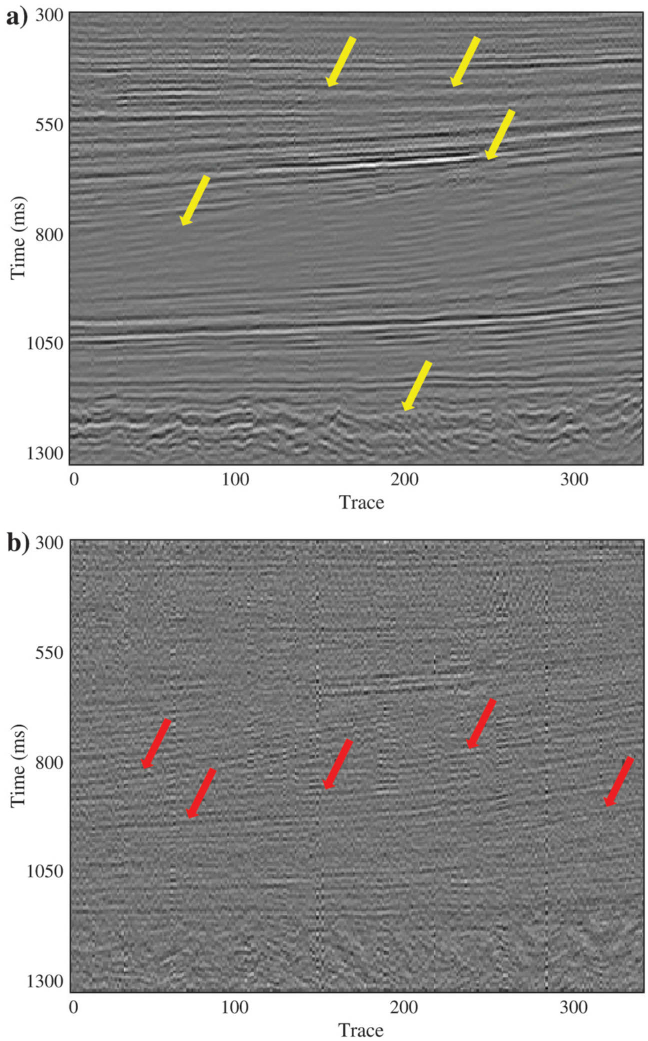


Figure 3. Illustration of the test on F3-block using MDnCNN. The yellow arrows indicate the remaining noise, and the red arrows indicate the rejected visible reflections. (a) The denoised result and (b) the rejected noise.

features contained in the feature map. Different from the first layer, a batch normalization (BN) (Ioffe and Szegedy, 2015) is added between the convolution filter and ReLU for the following 2–16 layers. BN is a reparameterization used to stabilize the updating of parameters and improve the learning process. The last layer only contains 64 convolution filters of size 3×3 to reconstruct the output. Then, the built neural network transforms the seismic noise attenuation procedure into an optimization problem by solving a sequence of nonlinear functions. A gradient-based optimization algorithm of adaptive moment estimation (Adam) (Kingma and Ba, 2015) is used to minimize the proposed objective function through iterative updating of the parameters of the network.

VMD-MDnCNN

Figure 2 shows a real seismic section of F3-block in the North Sea, Netherlands. Figure 3a and 3b shows the denoised result and the rejected noise by using the MDnCNN. The noise is indicated by the yellow arrows in Figure 3a, and rejected visible reflections are indicated by the red arrows in Figure 3b. The main reason for this phenomenon is that the seismic data are band limited and the S/N is varying with bandwidth. If we only use the MDnCNN to learn the feature of original white noise at full bandwidth, the MDnCNN cannot attenuate all the original white noise and will remove some visible reflections. To better learn the noise feature, we first decompose the seismic data into different components and then apply the MDnCNN to each decomposed component. In this work, we apply the VMD to decompose the seismic traces. We called this seismic noise attenuation procedure as VMD-MDnCNN.

VMD is an adaptive and nonrecursive signal decomposition method (Dragomiretskiy and Zosso, 2014). This method decomposes a signal into a series of IMFs. The frequency spectrum of each IMF is computed around the center frequency ω_i, and the sparsity of each IMF is constrained by its bandwidth in the frequency domain. In other words, VMD decomposes the signal into different IMFs and the frequency spectrum of each component is tuned around the center frequency ω_i. We obtain each IMF by recursively solving the following optimization problem:

$$\min_{\{u_i\},\{\omega_i\}} \left\{ \sum_i \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_i(t) \right] e^{-j\omega_i t} \right\|_2^2 \right\}, \qquad (10)$$
$$\sum_i^L u_k = s(t)$$

where u_i and ω_i are the modes and their center frequencies, respectively, $\delta(t)$ is a Dirac impulse, $s(t)$ is the signal to be decomposed, the constraint condition is that the summation over all modes should be the input signal, the term $(\delta(t) + (j/\pi t)) * u_i(t)$ is the Hilbert transform of u_i, and the parameter L is the user-defined decomposed number. The denoising objective function $\mathbf{J}(\boldsymbol{\theta})^{(j)}$ for the jth decomposed seismic component is given as

$$\mathbf{J}(\boldsymbol{\theta})^{(j)} = \arg\min_\theta \frac{1}{N}\frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{N} \| R_\theta^{(j)}(\mathbf{y}_i^{(j)}) - (\mathbf{y}_{ik}^{\prime(j)} - \mathbf{y}_i^{(j)}) \|^2, \qquad (11)$$

where $\mathbf{y}^{(j)}$ and $\mathbf{y}^{\prime(j)}$ represent the jth decomposed component and the components with additive white noise, respectively.

Figure 4 shows the proposed denoising workflow using VMD-MDnCNN. We first decompose the 3D seismic data (reshaped into 2D seismic sections in the inline and crossline directions) into several decomposed components and compute the S/N for each component. We produce a "noisier" IMF by adding additive white noise to each decomposed IMF and the residual component. The energy of the additive white noise is approximately equal to the energy of the white noise estimated within each decomposed component. We next learn the feature of white noise by minimizing the difference between additive noise and learned white noise from the noisier decomposed components (equation 11). We produce the denoised components by subtracting the learnt noises from the corresponding decomposed components. We finally obtain the denoised seismic data by integrating the ensemble of denoised components. Figure 5a and 5b shows the denoised results and rejected noise by using VMD-MDnCNN. It can be noted that our method successfully rejects the noise indicated by the yellow arrows in Figure 3a and preserves the seismic reflections indicated by the red arrows in Figure 3b.

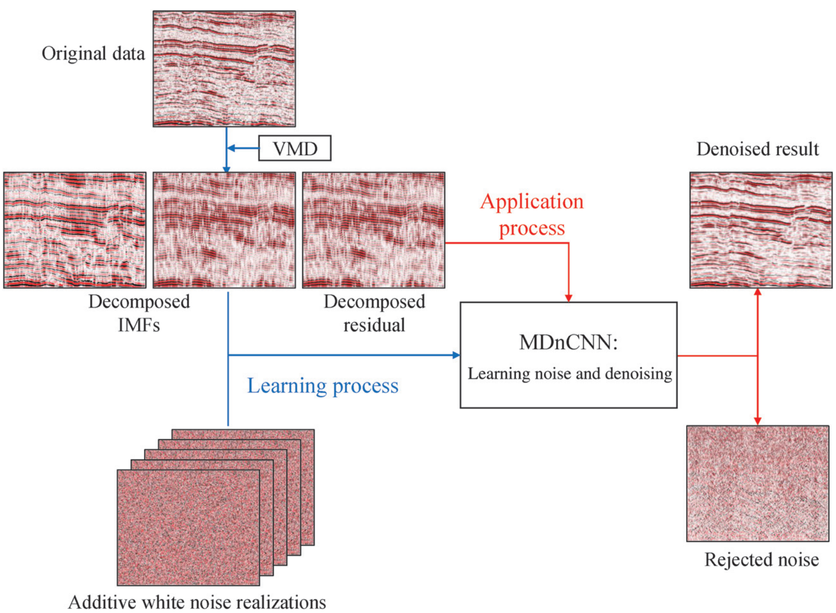


Figure 4. The workflow of VMD-MDnCNN for seismic noise attenuation.

SYNTHETIC EXAMPLE

To demonstrate the performance of VMD-MDnCNN, we first test our method using synthetic seismic data (Figure 6a) generated using the Marmousi model. We use a zero-phase Ricker wavelet to generate our synthetic seismic data. The dominant frequency of the Ricker wavelet is 30 Hz. The synthetic seismic data contain 128 traces, each with 128 time samples at

4 ms time interval. Figure 6b shows the noisy synthetic seismic data. The additive noise is Gaussian noise and the S/N is two. To ensure that the additive noise has the same frequency bandwidth with the seismic data, we applied a band-pass Butterworth filter (5–10–95–100 Hz) to the Gaussian noise before we add the noise to the noise-free synthetic seismic data. Figure 7a–7c shows the denoised seismic data using f-x deconvolution, MDnCNN, and VMD-MDnCNN, respectively. Figure 8a–8c shows the rejected noise using f-x deconvolution, MDnCNN, and VMD-MDnCNN, respectively. VMD-MDnCNN not only rejects the white noise (the yellow arrows in Figure 7a–7c), but it also preserves the visible reflections rejected by f-x deconvolution and MDnCNN (the red arrows in Figure 8a–8c).

The filter length is 12 sample points, and the cutoff frequency range is 5–100 Hz for the f-x deconvolution. The IMF number is determined by the performance of denoising and computation cost. In this work, we found that that there is no obvious difference between the denoised results for the synthetic and real seismic data if the IMF number is equal to or greater than two. The computation cost increases with the increasing of the IMF number. We set the IMF number as two in our synthetic testing. The moderate bandwidth constraint and the tolerance of the convergence criterion are 100 and 0.01, respectively, for VMD. The preset center frequencies for these two IMFs are 25 and 45 Hz, respectively. The input for our VMD-MDnCNN is the first decomposed IMF plus additive noise, the second decomposed IMF plus additive noise, and the residual component plus additive noise. Figure 9 shows the training and validation loss varying with optimization epochs. Considering that the data size used in deep learning is huge, we usually divide the learning data set into several small subsets (batch). The optimization procedure is implemented batch by batch, and one epoch means one optimization iteration of the full batches. We obtain the training and validation loss by applying the objective function shown in equation 12 to the training and validation data set, respectively. To overcome the overfitting problem in the training procedure, the training and validation seismic traces are randomly

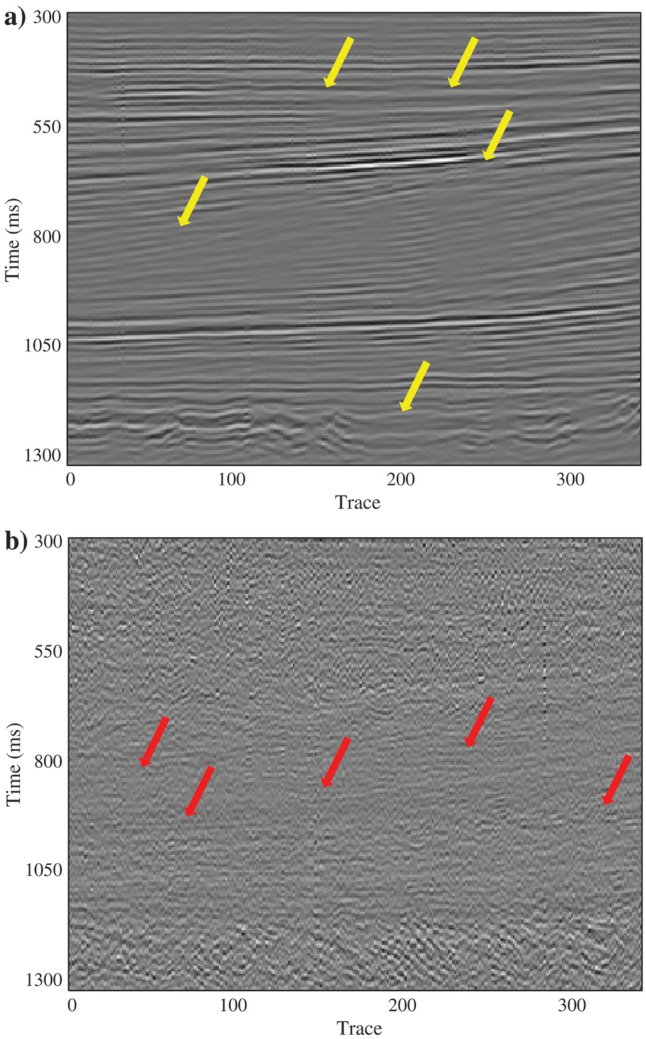


Figure 5. Illustration of the test on the F3-block using VMD-MDnCNN. The yellow arrows indicate the removed noise, and the red arrows indicate that there are no rejected visible reflections. (a) The denoised result and (b) the rejected noise.

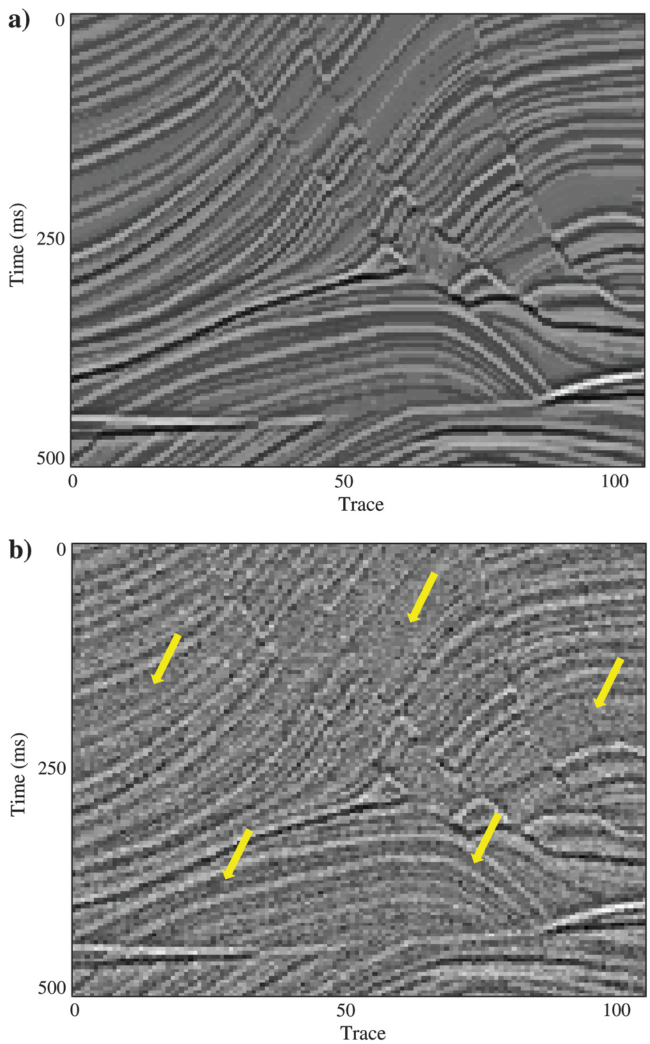


Figure 6. (a) The noise-free synthetic data of the Marmousi model. (b) The noisy synthetic data of the Marmousi model, where the yellow arrows point out several representative locations of white noise.

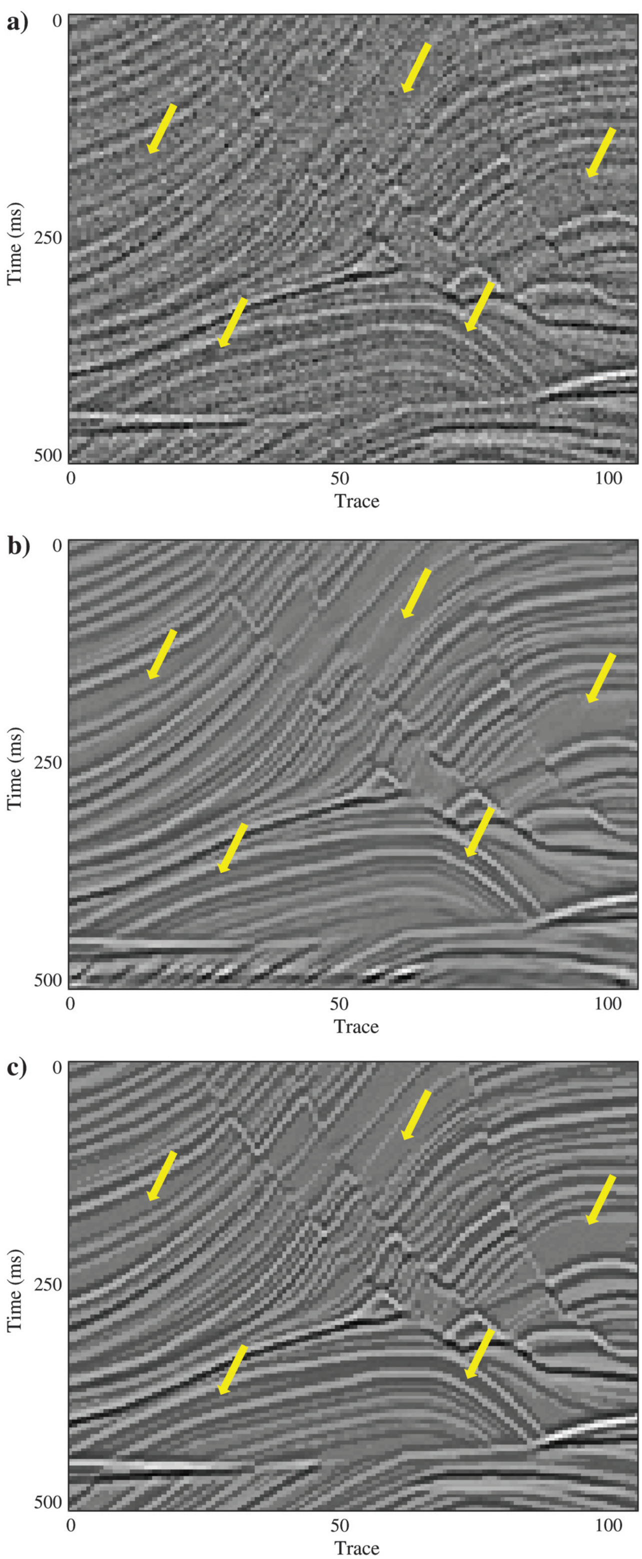


Figure 7. Illustration of the test on synthetic example. Note that the VMD-MDnCNN successfully rejects the white noise, whereas the f-x deconvolution and MDnCNN failed to reject the white noise (the yellow arrows in figures a-c). (a) The denoised result using f-x deconvolution. (b) The denoised result using MDnCNN. (c) The denoised result using VMD-MDnCNN.

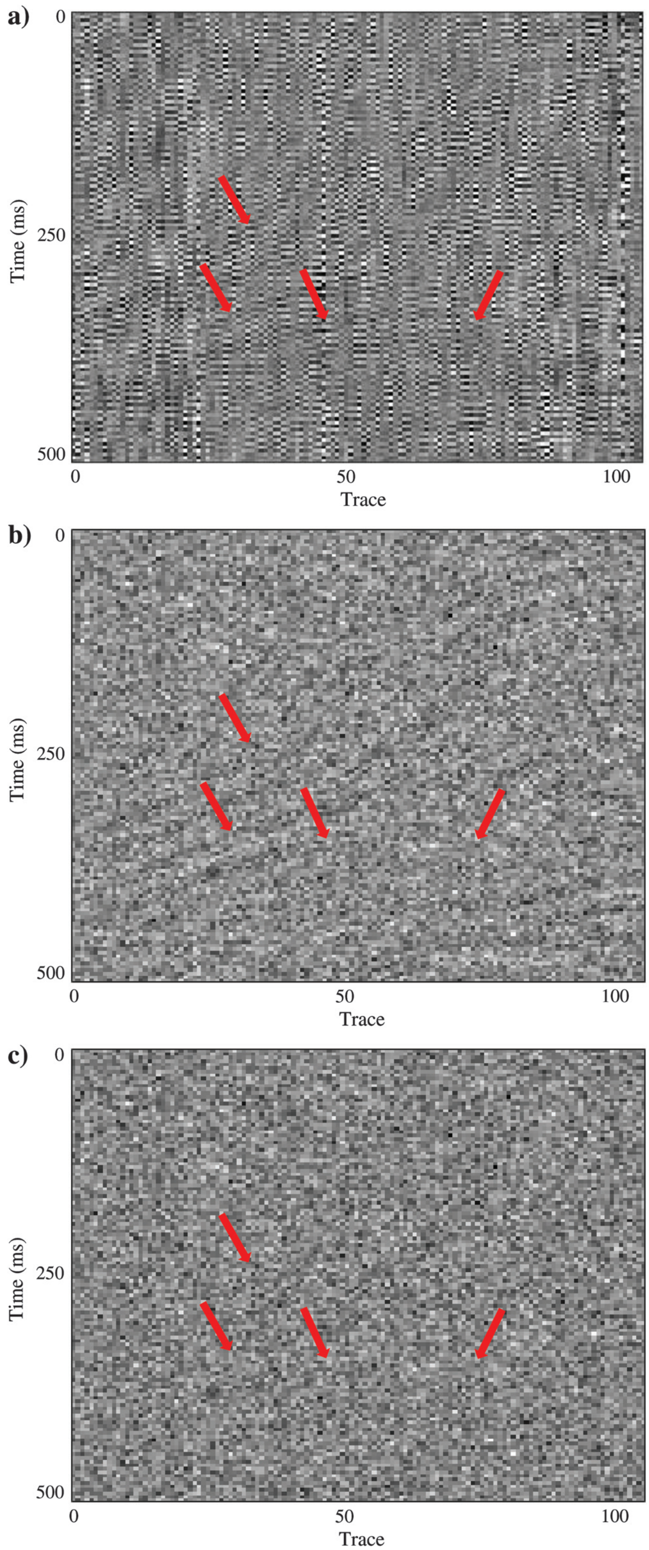


Figure 8. Rejected noise through different denoising methods on synthetic example. Note that the VMD-MDnCNN successfully preserves the visible reflections, whereas the f-x deconvolution and MDnCNN failed to preserve the visible reflections (the red arrows in figures a-c). (a) The rejected noise using f-x deconvolution. (b) The rejected noise using MDnCNN. (c) The rejected noise using VMD-MDnCNN.

selected during each optimization epoch. The percentage of the training and validation seismic traces in this paper is 70% and 30%, respectively, for synthetic and real seismic data. A specific seismic trace may belong to training seismic traces set in the current optimization epoch but may belong to validation seismic traces set in the next optimization epoch. Figure 9 illustrates that we obtain a stable neural network hierarchy after 50 epochs in the training procedure.

Figure 10 shows the average amplitude spectrum of the original seismic (black), the denoised result using f-x deconvolution (red), the denoised result using MDnCNN (blue), and the denoised result using VMD-MDnCNN (green). The amplitude spectrum of the denoised result using VMD-MDnCNN has a very good match with that of original seismic data. Unfortunately, the denoised result using f-x deconvolution and MDnCNN lost certain middle- and high-frequency content when compared to that of the original seismic data.

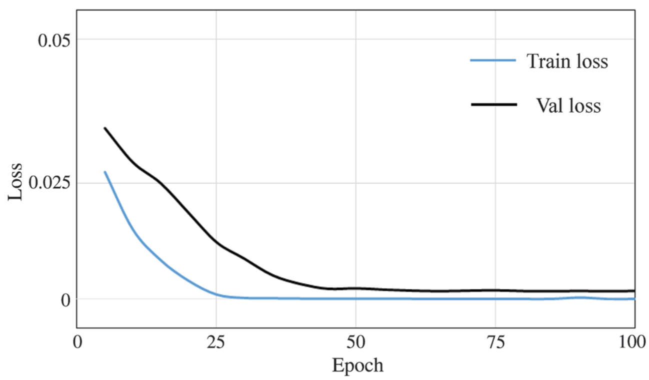


Figure 9. Train loss and validation loss varying with different epochs for synthetic example.

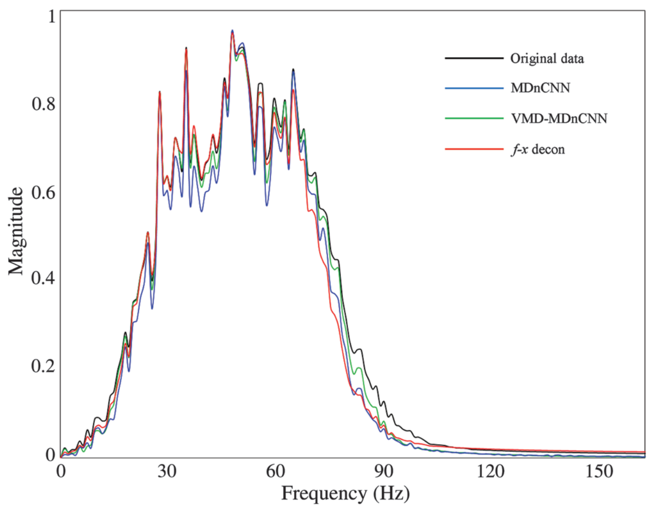


Figure 10. The frequency spectrum of the original data (the black curve), MDnCNN result (the blue curve), VMD-MDnCNN result (the green curve), and f-x deconvolution result (the red curve) for the synthetic test.

FIELD DATA EXAMPLE

We further apply VMD-MDnCNN to a public domain seismic survey (Penobscot) to illustrate the effectiveness of our proposed method. The Penobscot seismic survey was acquired over the Scotian shelf, offshore Canada. The seismic survey contains 601 inlines and 482 crosslines. The time increment of the seismic survey is 4 ms. We observe the residual noise and possible migration artifacts indicated by the yellow arrows in Figure 11.

We first generate a large number of band-limited additive white noise realizations. The bandwidth of the additive white noise is obtained according to the frequency spectrum of the seismic data. We then use VMD to decompose the original seismic data into two IMFs and a residual volume. Based on testing of the real data application, we found that we can successfully simulate the noise contained in the seismic data if the number of white noise realizations is greater than 2000. In this work, we choose 2000 as the number of white noise realizations. The simulated noise is then added to the

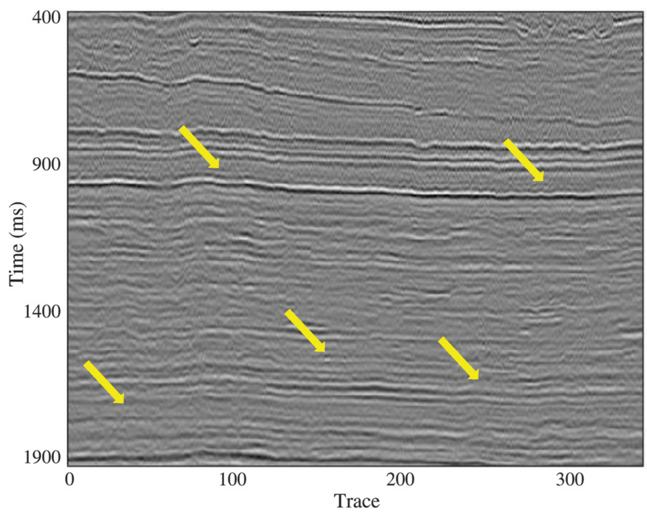


Figure 11. One representative inline section example of Penobscot. The yellow arrows point out several representative migration artifacts.

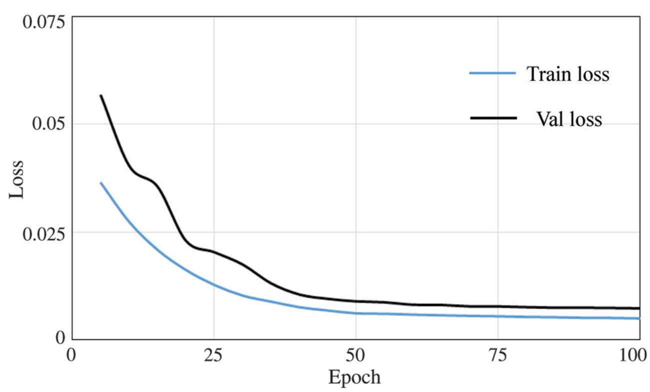
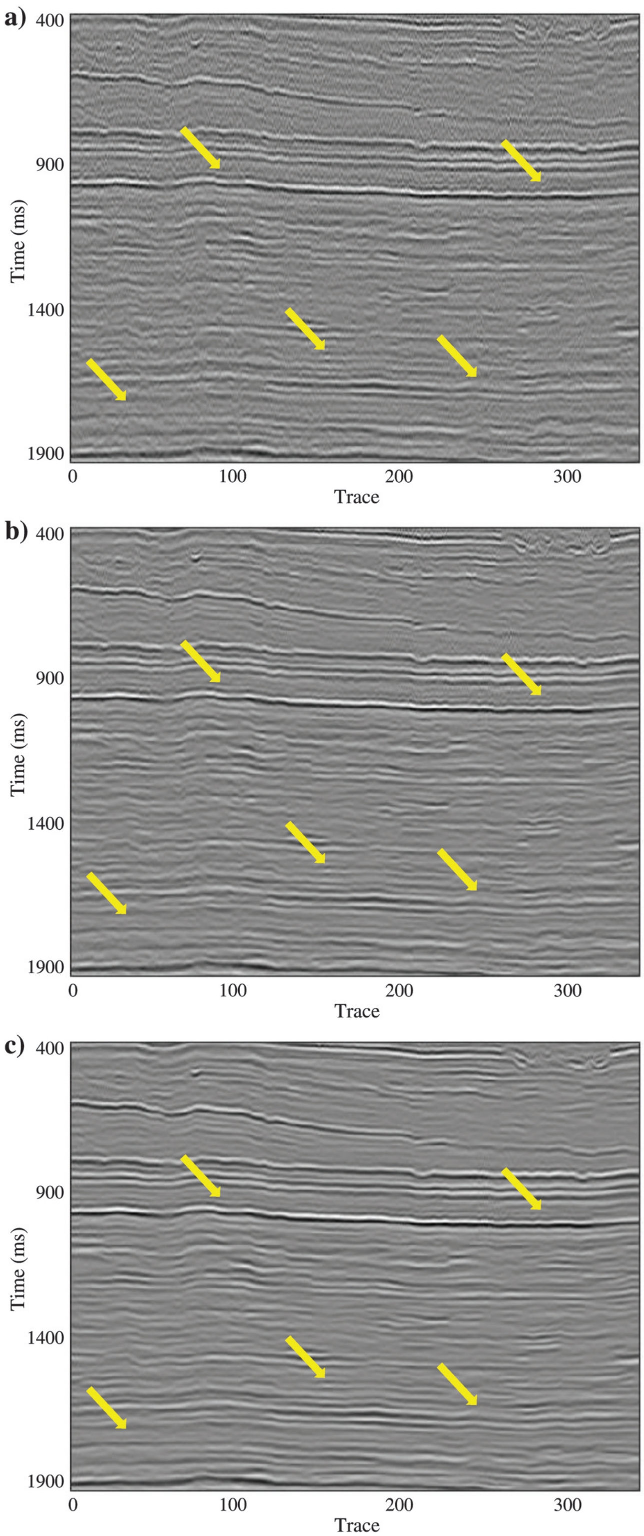


Figure 12. Train loss and validation loss varying with different epochs for the seismic survey of Penobscot.



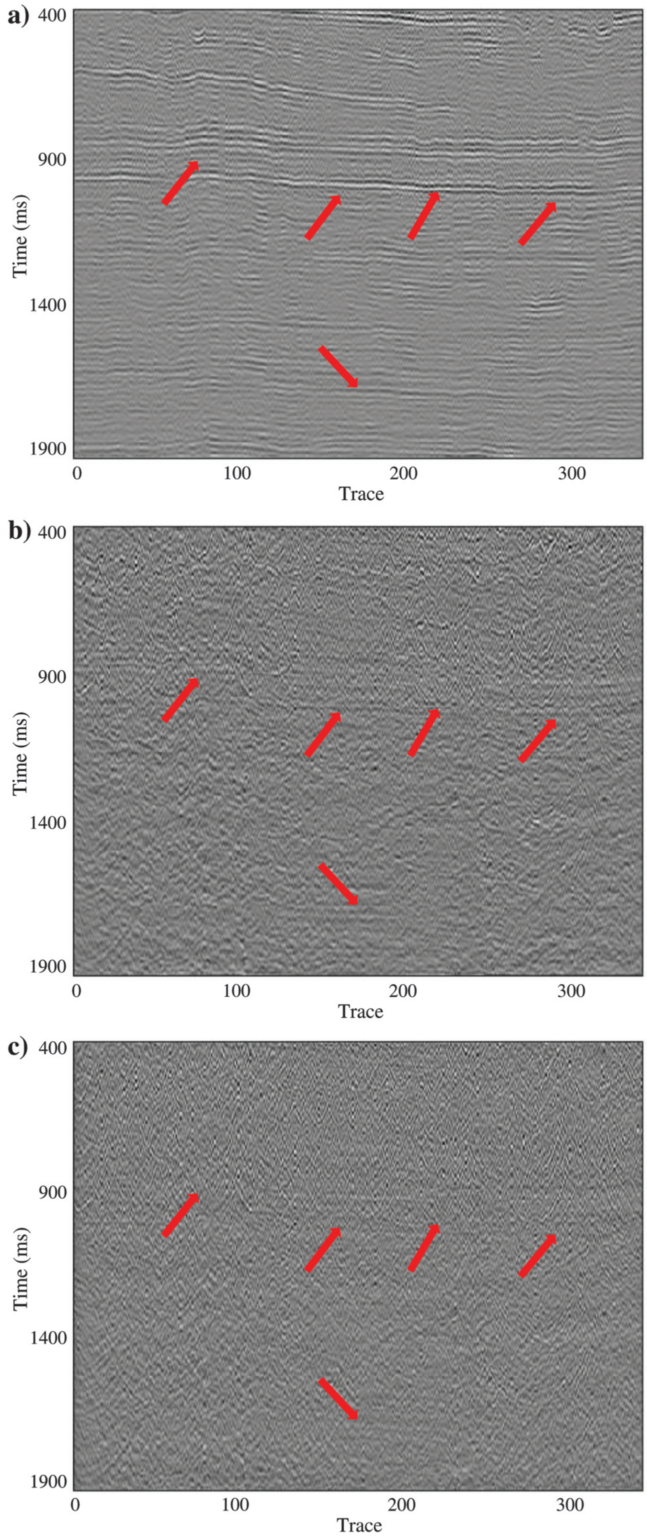


Figure 13. Illustration of the seismic noise attenuation test on the seismic survey of Penobscot. Note that VMD-MDnCNN successfully rejects the migration artifacts, whereas the f-x deconvolution and MDnCNN failed to reject the migration artifacts (the yellow arrows in figures a-c). (a) The denoised result using f-x deconvolution. (b) The denoised result using MDnCNN. (c) The denoised result using VMD-MDnCNN.

Figure 14. Rejected noise through different denoising methods on the seismic survey of Penobscot. Note that the VMD-MDnCNN successfully rejects the migration artifacts, whereas the f-x deconvolution and MDnCNN still preserve the migration artifacts (the red arrows in figures a-c). (a) The rejected noise using f-x deconvolution. (b) The rejected noise using MDnCNN. (c) The rejected noise using VMD-MDnCNN.

decomposed two IMFs and a residual component. Figure 12 illustrates that we obtain a stable neural network after 60 epochs in the training procedure.

Figure 13a–13c shows the denoised results using f-x deconvolution, MDnCNN, and VMD-MDnCNN, respectively. Figure 14a–14c shows the difference between the original seismic and denoised results using f-x deconvolution, MDnCNN, and VMD-MDnCNN, respectively. VMD-MDnCNN not only rejects the white noise and migration artifact indicated by the yellow arrows in Figure 13, but it also preserves the visible reflections rejected by the f-x deconvolution and MDnCNN indicated by the red arrows in Figure 14. Figures 15 and 16a–16c show the 3D cube of the original, denoised using f-x deconvolution, MDnCNN, and VMD-MDnCNN results, respectively. Figure 17a–17c shows the 3D rejected noise using f-x deconvolution, MDnCNN, and VMD-MDnCNN, respectively. VMD-MDnCNN has successfully rejected most of the white noise and migration artifacts indicated by the yellow arrows in Figure 15. However, the denoised results using f-x deconvolution and the MDnCNN still contain white noise and migration artifacts indicated by the yellow arrows in Figure 16. Figure 17a–17c illustrates that our proposed method not only attenuates white noise but it also preserves most of the useful seismic amplitude indicated by the red arrows. In addition, the visible reflections indicated by the red arrows in Figure 17b and 17c illustrate the superiority of VMD-MDnCNN over MDnCNN.

Again, we further compare the spectrum of the original and denoised seismic data to show the effectiveness of our method. Figure 18 shows the amplitude spectrum of the original seismic (black), the denoised result using f-x deconvolution (red), the denoised result using MDnCNN (blue), and the denoised result using VMD-MDnCNN (green). It should be noted that the average amplitude spectrum of denoised data using VMD-MDnCNN has a very good match with that of the original seismic data. Unfortunately, the denoised result using f-x deconvolution and MDnCNN lost certain middle- and high-frequency content when compared to that of the original seismic data.

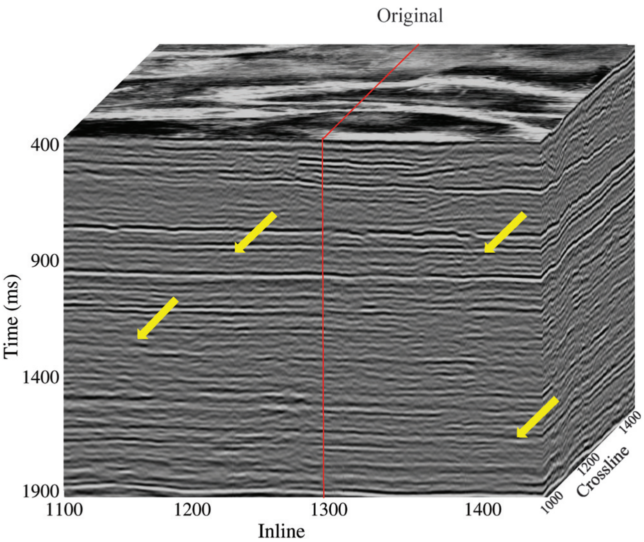


Figure 15. The original 3D seismic data of Penobscot. The yellow arrows point out several representative migration artifacts.

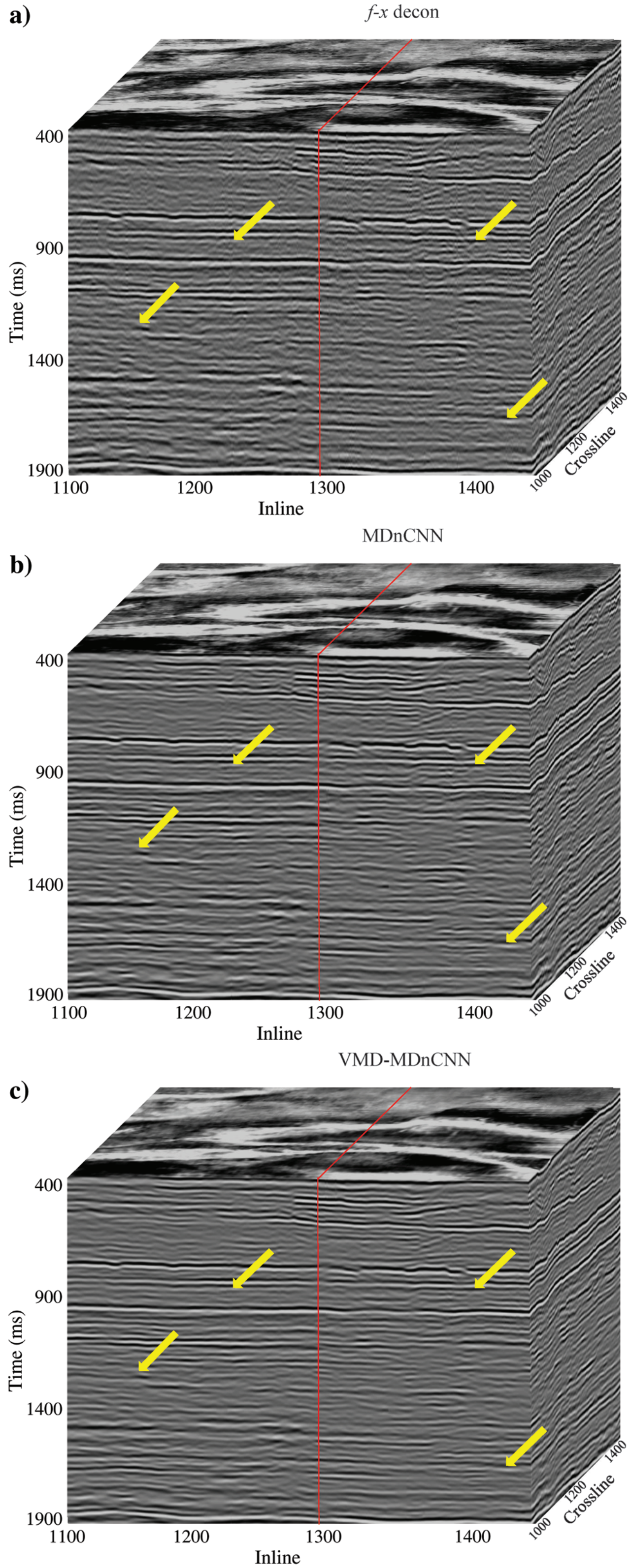


Figure 16. Illustration of the 3D volume denoised result on the seismic survey of Penobscot. Note that the VMD-MDnCNN successfully rejects the migration artifacts, whereas the f-x deconvolution and MDnCNN failed to reject the migration artifacts (the yellow arrows in figures a-c). (a) The denoised result using f-x deconvolution. (b) The denoised result using MDnCNN. (c) The denoised result using VMD-MDnCNN.

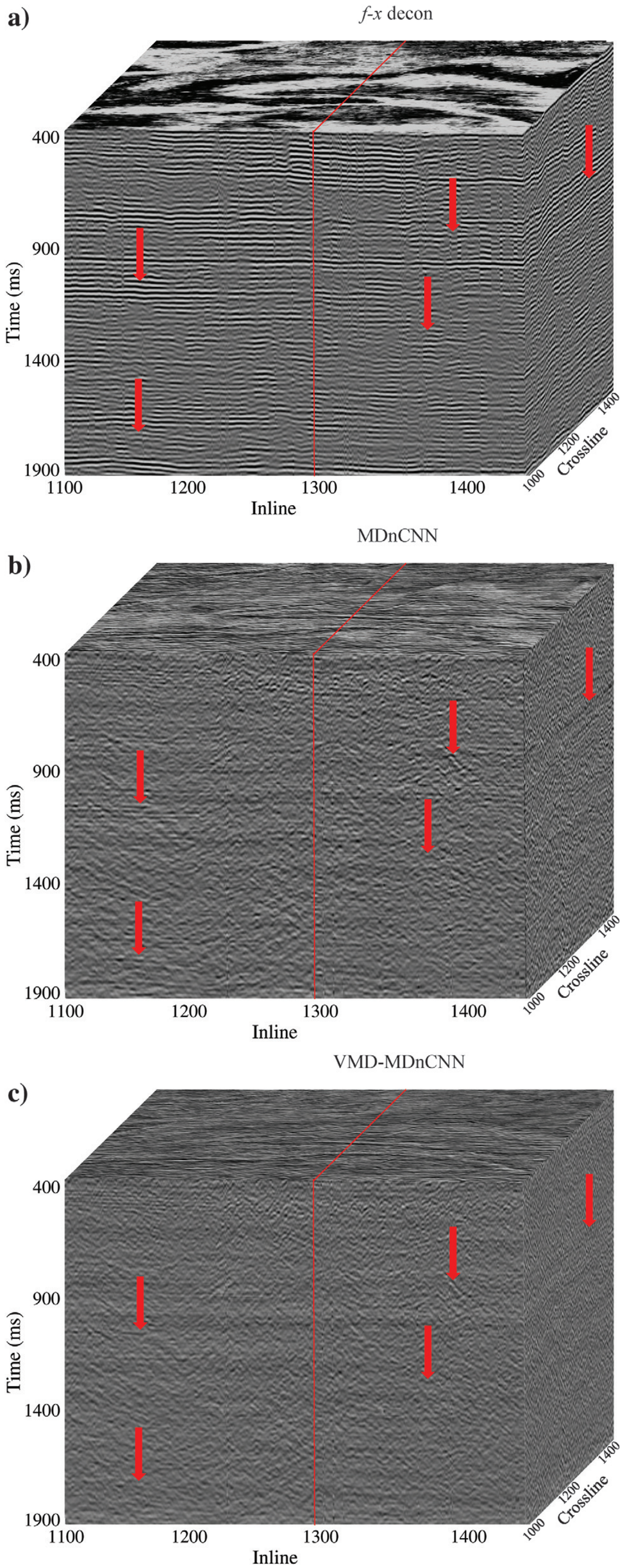


Figure 17. Rejected noise on a 3D volume through different denoising methods on the seismic survey of Penobscot. Note that the VMD-MDnCNN successfully rejects the migration artifacts, whereas the f-x deconvolution and MDnCNN still preserve the migration artifacts (the red arrows in figures a-c). (a) The rejected noise using f-x deconvolution. (b) The rejected noise using MDnCNN. (c) The rejected noise using VMD-MDnCNN.

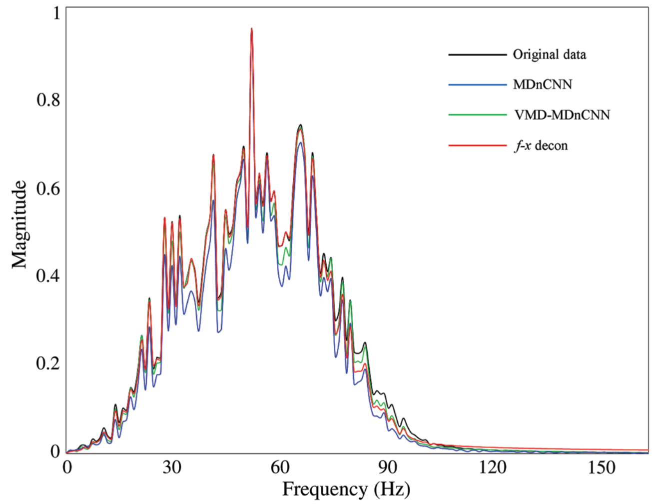


Figure 18. The frequency spectrum of the original data (the black curve), MDnCNN result (the blue curve), VMD-MDnCNN result (the green curve), and f-x deconvolution result (the red curve) for the seismic survey of Penobscot.

CONCLUSION

We propose a novel seismic noise attenuation method (VMD-MDnCNN) by integrating our MDnCNN with VMD. Current CNN-based denoising methods either require the label of clean seismic data or the label of noise contained in the seismic data, which is unfeasible in practice. The applications of our proposed method demonstrate that the white noise contained in the seismic can be simulated by a sufficient number of user-generated white noise realizations, which do not require clean seismic data. In addition, the applications demonstrate that the MDnCNN can obtain a more accurate estimation of the noise feature from the decomposed band-limited seismic data. The synthetic and real seismic data applications illustrate that our method is superior to the traditional denoising method of f-x deconvolution. The applications also demonstrate that our method effectively rejects not only the white noise but also the migration artifacts contained in the seismic data.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.

REFERENCES

Abma, R., and J. Claerbout, 1995, Lateral prediction for noise attenuation by t-x and f-x techniques: Geophysics, **60**, 1887–1896, doi: 10.1190/1.1443920.

Bekara, M., and M. van der Baan, 2007, Local singular value decomposition for signal enhancement of seismic data: Geophysics, **72**, no. 2, V59–V65, doi: 10.1190/1.2435967.

Bekara, M., and M. van der Baan, 2009, Random and coherent noise attenuation by empirical mode decomposition: Geophysics, **74**, no. 5, V89–V98, doi: 10.1190/1.3157244.

Burger, H., C. Schuler, and S. Harmeling, 2012, Image denoising with multi-layer perceptrons — Part 1: Comparison with existing algorithms and with bounds, arXiv: 1211.1544.

Canales, L., 1984, Random noise reduction: 54th Annual International Meeting, SEG, Expanded Abstracts, 525–527, doi: 10.1190/1.1894168.

Chen, Y., and J. Ma, 2014, Random noise attenuation by f-x empirical-mode decomposition predictive filtering: Geophysics, **79**, no. 3, V81–V91, doi: 10.1190/geo2013-0080.1.

Chen, Y., W. Yu, and T. Pock, 2015, On learning optimized reaction diffusion processes for effective image restoration: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 5261–5269.

Donoho, D. L., and I. M. Johnstone, 1994, Ideal spatial adaptation by wavelet shrinkage: Biometrika, **81**, 425–455, doi: 10.1093/biomet/81.3.425.

Dragomiretskiy, K., and D. Zosso, 2014, Variational mode decomposition: IEEE Transactions on Signal Processing, **62**, 531–544, doi: 10.1109/TSP.2013.2288675.

Elad, M., and M. Aharon, 2006, Image denoising via sparse and redundant representations over learned dictionaries: IEEE Transactions on Image Processing, **15**, 3736–3745, doi: 10.1109/TIP.2006.881969.

Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: Geophysics, **75**, no. 3, V25–V38, doi: 10.1190/1.3380591.

Herrmann, F. J., U. Boniger, and D. J. Verschuur, 2007, Non-linear primary multiple separation with directional curvelet frames: Geophysical Journal International, **170**, 781–799, doi: 10.1111/j.1365-246X.2007.03360.x.

Huang, N., Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, and H. Liu, 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, **454**, 903–995, doi: 10.1098/rspa.1998.0193.

Ioffe, S, and C. Szegedy, 2015, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv: 1502.03167.

Jain, V., and S. Seung, 2009, Natural image denoising with convolutional networks: Proceedings of the Advances in Neural Information Processing Systems, 769–776.

Kingma, D. P., and J. L. Ba, 2015, Adam: A method for stochastic optimization, arXiv: 1412.6980.

Kong, D., and Z. Peng, 2015, Seismic random noise attenuation using shearlet and total generalized variation: Journal of Geophysics and Engineering, **12**, 1024–1035, doi: 10.1088/1742-2132/12/6/1024.

LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436–444, doi: 10.1038/nature14539.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998, Gradient-based learning applied to document recognition: Proceedings of the IEEE, **86**, 2278–2324, doi: 10.1109/5.726791.

Li, F., B. Zhang, S. Verma, and K. Marfurt, 2017, Seismic signal denoising using thresholded variational mode decomposition: Exploration Geophysics, **49**, 450–461, doi: 10.1071/EG17004.

Liu, G., X. Chen, J. Du, and J. Song, 2011, Seismic noise attenuation using nonstationary polynomial fitting: Applied Geophysics, **8**, 18–26, doi: 10.1007/s11770-010-0244-2.

Liu, Y, N. Liu, and C. Liu, 2012, Adaptive prediction filtering in t-x-y domain for random noise attenuation using regularized nonstationary autoregression: Geophysics, **80**, no. 1, V13–V21, doi: 10.1190/geo2014-0011.1.

Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, no. 3, V25–V32, doi: 10.1190/1.3552706.

Trad, D., T. Ulrych, and M. Sacchi, 2002, Accurate interpolation with high-resolution time variant Radon transforms: Geophysics, **67**, 644–656, doi: 10.1190/1.1468626.

Trickett, S., 2008, F-xy Cadzow noise suppression: CSPG CSEG CWLS Convention, 303–306.

Vautard, R., P. Yiou, and M. Ghil, 1992, Singular-spectrum analysis: A toolkit for short, noisy chaotic signals: Physica D: Nonlinear Phenomena, **58**, 95–126, doi: 10.1016/0167-2789(92)90103-T.

Wang, Y., 1999, Random noise attenuation using forward-backward linear prediction: Journal of Seismic Exploration, **8**, 133–142.

Wu, Z., and N. Huang, 2004, A study of the characteristics of white noise using the empirical mode decomposition method: Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, **460**, 1597–1611, doi: 10.1098/rspa.2003.1221.

Wu, Z., and N. Huang, 2009, Ensemble empirical mode decomposition: A noise-assisted data analysis method: Advances in Adaptive Data Analysis, **1**, 1–41, doi: 10.1142/S1793536909000047.

Xie, J., L. Xu, and E. Chen, 2012, Image denoising and inpainting with deep neural networks: Proceedings of the Advances in Neural Information Processing Systems, 350–358.

Xue, Y., J. Yang, J. Ma, and Y. Chen, 2016, Amplitude-preserving nonlinear adaptive multiple attenuation using the high-order sparse radon transform: Journal Geophysics and Engineering, **13**, 207–219, doi: 10.1088/1742-2132/13/3/207.

Yuan, S., S. Wang, C. Luo, and T. Wang, 2018, Inversion-based 3-D seismic denoising for exploring spatial edges and spatio-temporal signal redundancy: IEEE Geoscience and Remote Sensing Letters, **15**, 1682–1686, doi: 10.1109/LGRS.2018.2854929.

Zhang, K., W. Zuo, Y. Chen, D. Meng, and L. Zhang, 2017, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising: IEEE Transactions on Image Processing, **26**, 3142–3155, doi: 10.1109/TIP.2017.2662206.