

An Unsupervised Deep Learning Method for Denoising Prestack Random Noise

Dawei Liu, Zheyuan Deng, Cheng Wang, Xiaokai Wang^{ID}, Member, IEEE, and Wenchao Chen^{ID}

Abstract—Deep-learning-based methods have been successfully applied to seismic data random noise attenuation. Among them, the supervised deep-learning-based methods dominate the unsupervised ones. The supervised methods need accurate noise-free data as training labels. However, the field seismic data cannot meet this requirement. To circumvent it, some researchers utilized realistic-looking synthetic data or denoised results via conventional methods as labels. The former ones encounter the problem of weak generalization ability because it requires the same distribution of test and training data. The latter ones encounter the issue of insufficient denoising ability because its denoising ability is difficult to significantly exceed the conventional methods which were used to generate labels. To avoid preparing noise-free labels, we propose a novel deep learning framework for attenuating random noise of prestack seismic data in an unsupervised manner. The prestack seismic data, such as common-reflection-point (CRP) gathers and common-midpoint (CMP) gathers after normal moveout (NMO) correction, have high self-similarity. It is because their events are coherent in the time-space domain and approximately horizontal from shallow to deep layers. The generator convolutional neural network (GCN) first learns self-similar features before any learning. The useful signals are more self-similar than random noise, which is incoherent and randomly distributed. Therefore, the GCN extracts features of useful signals before random noise. We select the specific training iteration and adopt the early stopping strategy to suppress random noise. Both synthetic and field prestack seismic data examples demonstrate the validity of our methods.

Index Terms—Attenuation, prestack seismic data, random noise, unsupervised.

I. INTRODUCTION

THE attenuation of prestack random noise in seismic processing is a classic but still not completely solved problem. Generally, the methods for attenuating random noise can be roughly divided into two categories, i.e., conventional model-based optimization methods and discriminative learning methods.

Manuscript received May 5, 2020; revised June 15, 2020 and July 26, 2020; accepted August 22, 2020. This work was supported in part by the National Science and Technology Major Project under Grant 2017ZX05049002-004 and in part by the National Natural Science Foundation of China under Grant 41774135 and Grant 41974131. The work of Dawei Liu was supported by the China Scholarship Council. (*Corresponding author: Wenchao Chen.*)

Dawei Liu, Zheyuan Deng, Xiaokai Wang, and Wenchao Chen are with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: liudawei2015@stu.xjtu.edu.cn; dzy1996@stu.xjtu.edu.cn; xkwang@xjtu.edu.cn; wencchen@xjtu.edu.cn).

Cheng Wang is with the Exploration and Development Research Institute, Daqing Oilfield Company Ltd., Daqing 163712, China (e-mail: wangcheng88@petrochina.com.cn).

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

Digital Object Identifier 10.1109/LGRS.2020.3019400

The model-based methods use explicit prior knowledge to formulate a model and solve the associated optimization problems to calculate the denoised seismic data. They have clear physical meaning and easy to generalize for handling different denoising problems. According to different prior knowledge, model-based methods can be roughly divided into three categories. The first type is based on prediction filtering [1], [2]. They utilize the prior assumption that seismic reflections have lateral continuity to distinguish reflections of interest from the background noise. The second is based on the low-rank prior, which comprises low-rank factorization [3]–[5], and nuclear norm minimization [6]. The third one is based on sparse representation, which merely requires that seismic data can be represented as a linear combination of several atoms from a dictionary [7]–[9]. The above methods have been successfully applied in practice. However, they have some common shortcomings. On the one hand, the hand-crafted priors need to be carefully designed, otherwise the priors may not be strong enough so that they cannot sufficiently distinguish some complicated structures of reflections from noise. On the other hand, most of them are time-consuming in the process of model optimization and require professional knowledge to realize parallel computing.

On the contrary, discriminant learning methods, such as deep learning, directly obtain prior knowledge from training data through end-to-end learning. The learned priors free us from the complex prior design. Deep learning is easy to carry out parallel computing with deep learning frameworks, which makes it widely used in the suppression of random noise. The common methods are based on 2-D network [10], and 3-D network [11], [12]. Although they achieve satisfying denoising results, there are several problems we are concerned about. The most crucial issue is that it requires a lot of noise-free data as labels, which are not always available for seismic field data. Besides, the generality of well-trained networks is limited, which means that its denoising ability may weaken when dealing with large-scale seismic data.

Unsupervised deep learning methods have both merits of model-based methods and discriminant learning methods. They are label-free and have a strong generalization ability with the high efficiency of parallel computing. Unsupervised deep learning method has been successfully applied to attenuate the noise of seismological data sets [13]. Zhang *et al.* [14] chose cross-entropy as the cost function and achieved unsupervised random noise suppression of seismic data. In this letter,

we propose a new unsupervised denoising method for prestack random noise. On the one hand, the events of common-reflection-point (CRP) gathers and common-midpoint (CMP) gathers after normal moveout (NMO) are coherent in the time–space domain and approximately horizontal from shallow to deep layers. Therefore, they have high self-similarity. On the other hand, the multiscale expression is beneficial to the processing of seismic data [15]. Inspired by Ulyanov *et al.* [16], We use a multiscale generator convolutional neural network (GCN) to extract multiscale self-similar features. The useful signals are more self-similar than random noise. Therefore, useful signals are extracted before random noise by the network. We select the network output of a specific iteration in the training process as denoised useful signals. In Section II, we introduce the model formulation, network architecture, and model training. In Section III, we use synthetic and field seismic data to prove the effectiveness of our method. Finally, we conclude in Section IV.

II. METHOD

A. Model Formulation

The seismic data, denoted by \mathbf{x}_0 , can be modeled as a superposition of useful signals and random noise

$$\mathbf{x}_0 = \mathbf{x} + \mathbf{n} \quad (1)$$

where \mathbf{x} denotes the useful signals and \mathbf{n} denotes random noise. We use a GCN to parameterize the useful signals as follows:

$$\mathbf{x} = f_\theta(\mathbf{z}) \quad (2)$$

where θ denotes the network parameters comprising the weights and bias, \mathbf{z} is a random vector, and f represents the nonlinear generator network which maps \mathbf{z} to \mathbf{x} . The parameterization can be seen as the reconstruction of \mathbf{x} . Inserting (2) to (1) leads to

$$\mathbf{x}_0 = f_\theta(\mathbf{z}) + \mathbf{n} \quad (3)$$

where θ and \mathbf{n} are the unknown parameters. Then, the task of random noise attenuation from the seismic data is equivalent to finding the optimal network parameters θ^* to minimize the energy function

$$\theta^* = \arg \min_{\theta} E(f_\theta(\mathbf{z}); \mathbf{x}_0). \quad (4)$$

The energy function $E(f_\theta(\mathbf{z}); \mathbf{x}_0)$ used in this letter is the following formulation:

$$E(f_\theta(\mathbf{z}); \mathbf{x}_0) = \|f_\theta(\mathbf{z}) - \mathbf{x}_0\|^2. \quad (5)$$

A unique advantage is that only the raw seismic data are employed, with no additional noise-free labels needed. Once θ^* is determined, we can obtain the recovered reflections quickly from the output of GCN $\mathbf{x}^* = f_{\theta^*}(\mathbf{z})$.

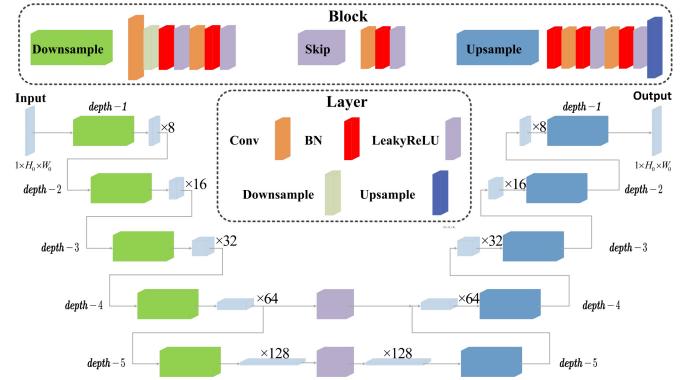


Fig. 1. Network architecture used in our method.

B. Network Architecture

The architecture of GCN is a U-Net type fully convolutional network comprising three parts: five downsample blocks, two skip blocks, and five upsample blocks. Each block consists of several basic units, including convolutional layers, batch normalization, downsampling layers, upsampling layers, and activation function layers, as shown in Fig. 1. To compromise calculation cost, the number of downsampling filters increases from 8 to 128 with the increase of depth i . The upsampling layer has a symmetrical output size. The downsampling blocks and upsampling blocks at different network depths reduce the dimensionality of the target seismic data by compressing common features and discarding useless information. This network architecture enables the GCN a multiscale self-similar feature extraction ability at multiple scales. The useful signals of seismic data have a multiscale feature in nature, and many seismic processing methods benefit from it [15], [17]. Therefore, we can apply the above ability to extract useful signals and denoise seismic data.

Also, we make several modifications based on the original U-Net. First, to reduce checkerboard artifacts caused by the upsample blocks, we substitute transposed convolution for bi-linear interpolation. Second, we adopt a skipping block strategy to avoid the gradient vanishing problem. Third, replacing rectified linear unit (ReLU) with leaky ReLU to prevent neuron annihilation.

C. Model Training

We parameterize the seismic data by the GCN in (2), which imposes a constrain that the reconstructed seismic data must be one possible solution described by the GCN. Equation (4) can be solved by an l_2 -norm optimization problem under this constraint. This problem is a nonconvex problem that the only unknown variable is θ . We can treat this problem as an unsupervised network training problem and apply the adaptive moment estimation (ADAM) optimizer to iteratively estimate a locally optimal θ^* . The network starts from a random initialization of θ_0 and an initialization of \mathbf{z} filled with uniform distribution. When training starts, the parameters θ_n of each iteration are mapped to a network output $\mathbf{x}'_n = f_{\theta_n}(\mathbf{z})$. As the training iteration increases, the energy function in (5) gradually converges. As stated in Section II-B,

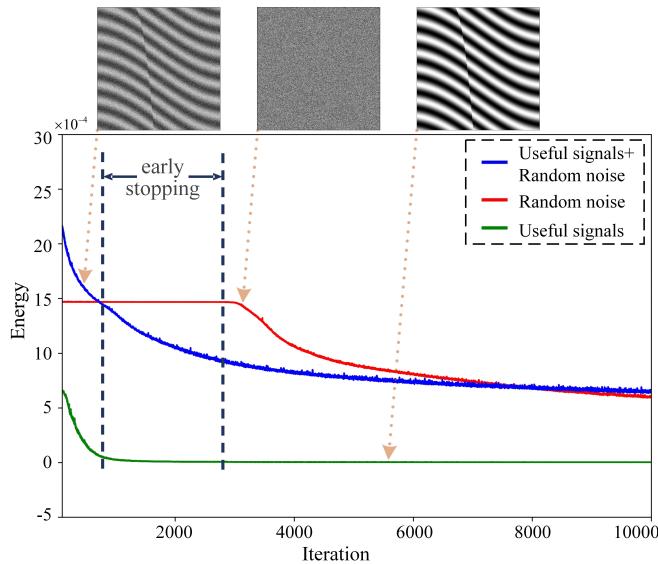


Fig. 2. Energy curves for the reconstruction task using: seismic useful signals, random noise, and noisy seismic data set.

the GCN prefers to extract self-similar features at multiple scales. The useful signals are coherent in space, while random noise is randomly distributed, which makes the useful signals more self-similar than random noise. Due to the constrain of GCN's solution space, the useful signals with self-similarities are more easily extracted by the network compared with random noise. To prove it, we use the GCN to reconstruct the noisy seismic data, random noise, and the useful signals, respectively. The noisy seismic data are obtained by adding the other two together. As shown in Fig. 2, if we want to reconstruct useful signals, 1500 iterations is enough because the GCN is easy to learn these self-similar features of useful signals. However, if we want to reconstruct random noise, more than 8000 iterations are needed because there are few features of self-similarity, so the network can only memorize it mechanically over and over again.

When we use the GCN to reconstruct raw seismic data, the output of the network first reconstructs useful signals and then starts to reconstruct random noise. In other words, there is a series of consecutive iterations in which the network has fit almost all the energy of useful signals, but only a few energy of random noise. By selecting a specific iteration and adopting an early stopping strategy, the network can filter out random noise from the useful signals. Although we could get the best results by a closer fine-tune of iterations, we found that a wide range of iterations gives us acceptable results. This robustness makes large-scale industrial applications accessible.

III. DATA EXAMPLES

A. Synthetic Data Example

We first evaluate the denoising performance on synthetic data. Fig. 3(a) shows the useful signals composed of three reflections modeled by hyperbolic events. The clean seismic data contain 9 gathers with 40 traces each gather and the spatial sampling interval is 40 m. Each trace has 500 time

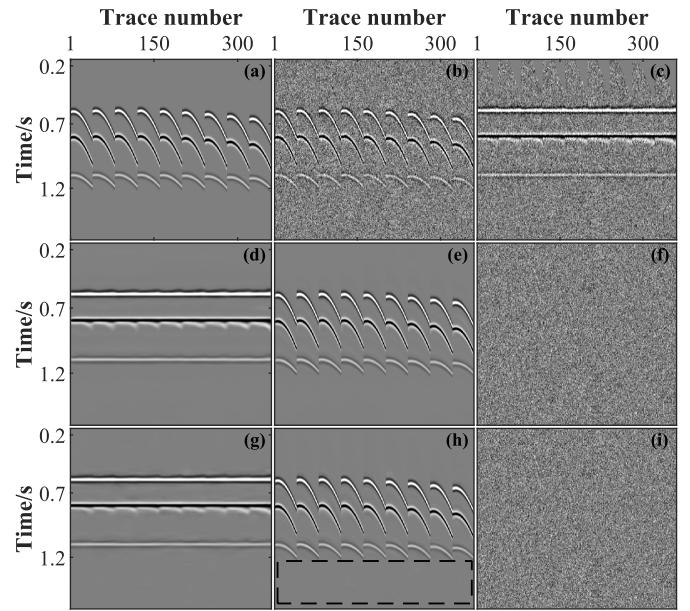


Fig. 3. Denoising comparisons of synthetic seismic data. (a) Clean data set. (b) Noisy data set ($\text{SNR} = -3.01 \text{ dB}$). (c) Noisy data set after NMO. Denoised results using (d) proposed method and (g) DDTF. Undoing NMO results of (e) proposed method and (h) DDTF. Removed noise using (f) proposed method and (i) DDTF.

TABLE I
SNR OF THE DENOISED RESULTS ON SYNTHETIC DATA

	DDTF	Our proposed method
Without NMO (SNR)	12.77 dB	11.68 dB
With NMO (SNR)	17.34 dB	18.47 dB

sampling points with 4 ms intervals. We add random noise to the clean data and obtain the noisy data shown in Fig. 3(b) with a signal-to-noise ratio (SNR) of -3.01 dB . We choose data-driven tight frame (DDTF) [7] with patch size 20 as the baseline method for comparison. The first row in Table I illustrates the denoised results without NMO. The SNR of our proposed method (11.68 dB) is lower than DDTF (12.77 dB). This is because these three events have three different velocities, which leads to different curvature of hyperbola and reduces the self-similarity. To further improve the denoising performance, we apply NMO to flatten useful signals. The flattened events, as shown in Fig. 3(c) has higher self-similarities. Fig. 3(d) and (g) display the corresponding denoising results of flattened data. After undoing NMO, the denoising results of both our methods (18.47 dB) in Fig. 3(e) and DDTF (16.95 dB) in Fig. 3(h) are improved compared with not using NMO. We see that our method can properly fit stretches generated by NMO and obtain the highest SNR. However, DDTF is hard to avoid producing weak artifacts after undoing NMO, as shown in the black box of Fig. 3(h), while our method can handle it well. Fig. 3(f) and (i) illustrate the corresponding noise removed by the two methods, where coherent useful signals can hardly be seen. The above results prove that the combination of NMO and data-driven denoising method we proposed is effective.

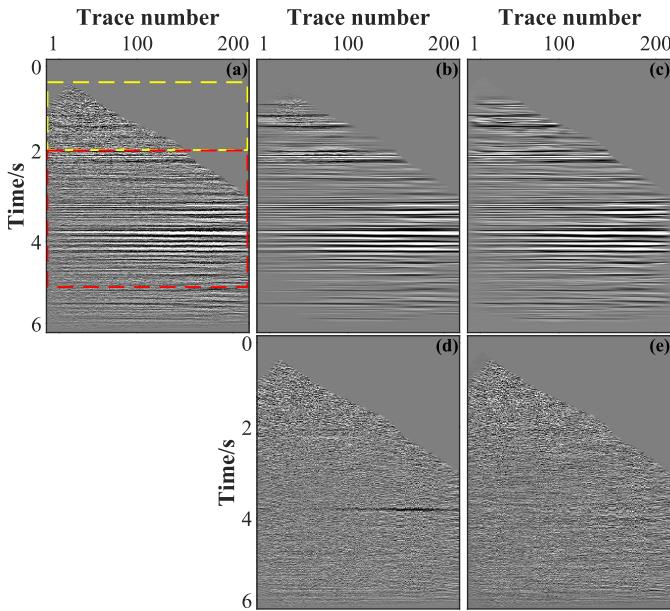


Fig. 4. Denoising comparisons of field seismic data. (a) Noisy data set. Denoised results using (b) DDTF and (c) proposed method. Removed noise using (d) DDTF and (e) proposed method.

B. Field Data Example

Then we apply our method to prestack field seismic data. The CRP gather shown in Fig. 4(a) consists of 200 traces with 2 ms time sampling interval. The useful signals are contaminated by a large amount of background noise, causing some weak signals to be blurred. Meanwhile, we note that the strongly reflected events are approximately horizontal, which both DDTF and our method are good at dealing with, as illustrated in the synthetic examples. Fig. 4(b) shows the denoised result of DDTF, and we can see that there is still some residual noise. Then we use the GCN to extract self-similar features from this raw seismic data. After 1500 iterations, we obtain the denoised result of the network, as shown in Fig. 4(c). It can be seen that there is no significant residual noise, and the events are clearer than the previous ones. Moreover, some coherent noise is also removed by our method without missing details of useful signals. In the corresponding removed noise sections of DDTF shown in Fig. 4(d), we find some leakage of useful signals. However, as shown in Fig. 4(e), we can hardly find obvious continuous events except for some unusual amplitude points, which indicate that our method does not seriously damage the useful signals during the denoising process. The computing time of our method (79.19 s) is longer than DDTF (51.30 s), but it is acceptable because our denoising results are better and parameters are easier to choose.

For a clearer comparison, we enlarge two areas indicated by the yellow box and red box in Fig. 4(a). The shallow-layer data of 0.5–2 s is seriously disturbed by random noise, which almost completely masks the useful signals, as shown in Fig. 5(a). The denoised results of DDTF in Fig. 5(b) still have noise residues. However, no significant random noise can be observed in the denoised results of our method shown in Fig. 5(c). It can also be seen from the noise section

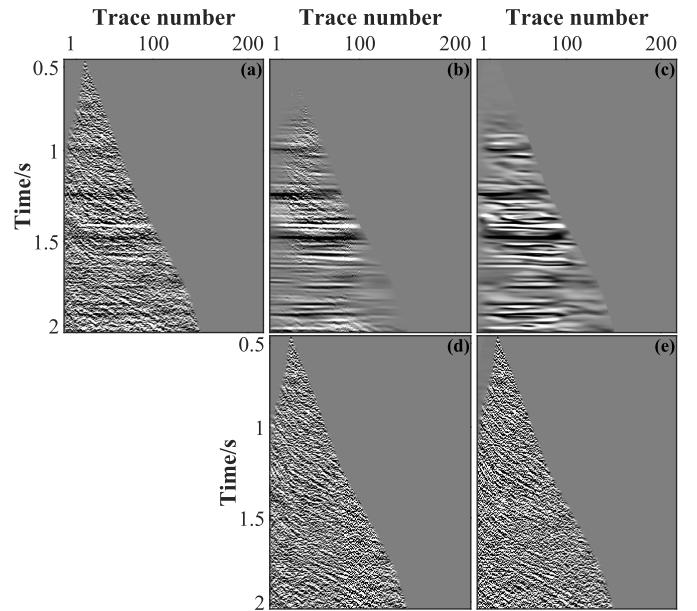


Fig. 5. Denoising comparisons of the enlarged area of the yellow box in Fig. 4. (a) Noisy data set. Denoised results using (b) DDTF and (c) proposed method. Removed noise using (d) DDTF and (e) proposed method.

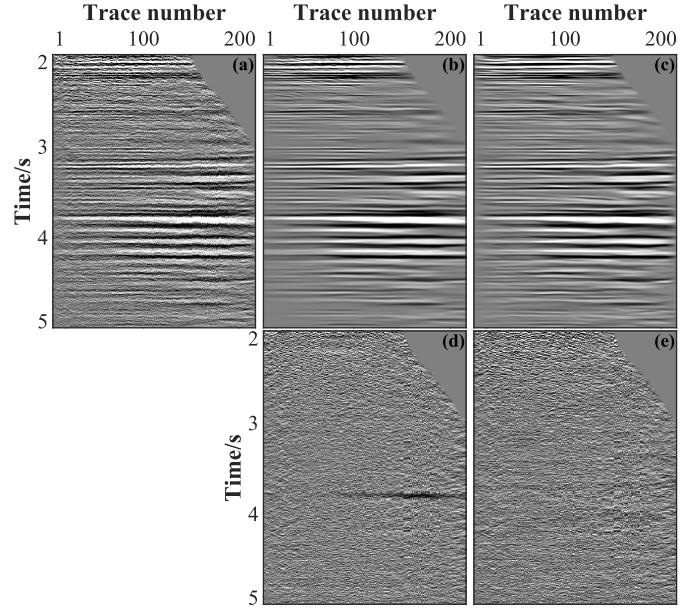


Fig. 6. Denoising comparisons of the enlarged area of the red box in Fig. 4. (a) Noisy data set. Denoised results using (b) DDTF and (c) proposed method. Removed noise using (d) DDTF and (e) proposed method.

in Fig. 5(e) that there is no obvious signal leakage. We can draw the same conclusion from the denoising results of 2–5 s in Fig. 6. In summary, our proposed method can effectively attenuate the noise under the premise of adequately preserving the signals.

We further explore the denoising performance on the multichannel normalized amplitude spectrum. As we all know, random noise has mainly high-frequency energy. We see that the high-frequency amplitude spectrum, such as higher than 60 Hz, denoised by the network is lower than that by DDTF,

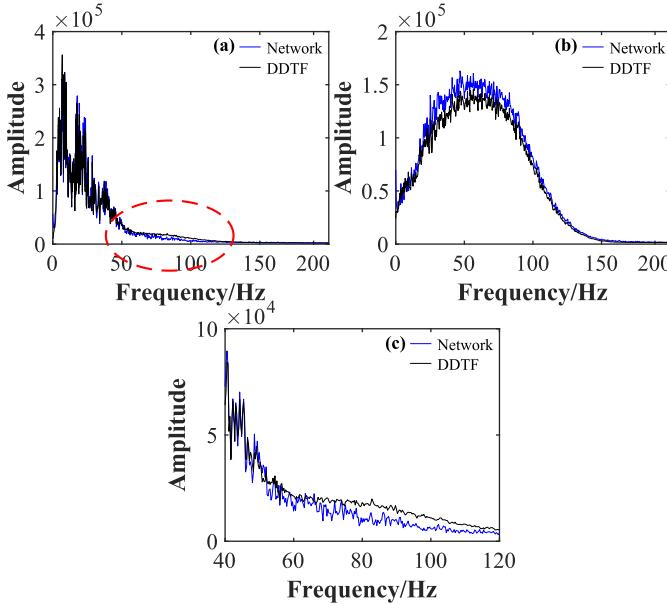


Fig. 7. Multichannel normalized amplitude spectrum of (a) denoised data, (b) removed noise, and (c) enlarged area of the red circle in (a).

as shown in Fig. 7(a). It means that the network suppresses more noise energy. We can also get the same conclusion from the amplitude spectrum of removed noise shown in Fig. 7(b). For a clearer comparison, Fig. 7(c) displays the enlarged area indicated by the red circle in Fig. 7(a). The denoising result of the network has lower energy at high frequencies. We can conclude that our method has higher fidelity and stronger noise suppression ability compared with DDTF.

IV. CONCLUSION

We propose an unsupervised method based on deep learning for random noise attenuation without requiring high-quality training labels. We use a GCN to reconstruct the raw seismic data. The network has a strong multiscale self-similarities feature extraction capability to recover the useful signals, but it is challenging to recover random noise. We can get the denoised useful signals by selecting a fixed iteration. Both synthetic and field data are utilized to demonstrate the effectiveness of the presented method. The proposed method

causes less damage to the useful signals while effectively suppressing random noise. The powerful self-similar features extraction ability gives us great hope in other prestack coherent noise suppression tasks for future research.

REFERENCES

- [1] N. Gulunay, "FXDECON and complex Wiener prediction filter," in *Proc. SEG Tech. Program Expanded Abstr.*, Jan. 1986, pp. 279–281.
- [2] R. Soubaras, "Signal-preserving random noise attenuation by the f-x projection," in *Proc. SEG Tech. Program Expanded Abstr.*, 1994, pp. 1576–1579.
- [3] S. R. Trickett, "F-xy eigenimage noise suppression," *Geophysics*, vol. 68, no. 2, pp. 751–759, 2003.
- [4] M. Bekara and M. Van der Baan, "Local singular value decomposition for signal enhancement of seismic data," *Geophysics*, vol. 72, no. 2, pp. V59–V65, Mar. 2007.
- [5] V. Orosez and M. Sacchi, "Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis," *Geophysics*, vol. 76, no. 3, pp. V25–V32, May 2011.
- [6] N. Kreimer and M. D. Sacchi, "A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation," *Geophysics*, vol. 77, no. 3, pp. V113–V122, May 2012.
- [7] J. Liang, J. Ma, and X. Zhang, "Seismic data restoration via data-driven tight frame," *Geophysics*, vol. 79, no. 3, pp. V65–V74, May 2014.
- [8] S. Yu, J. Ma, X. Zhang, and M. D. Sacchi, "Interpolation and denoising of high-dimensional seismic data by learning a tight frame," *Geophysics*, vol. 80, no. 5, pp. V119–V132, Sep. 2015.
- [9] M. Jian, G. Jinghuai, and C. Wenchoao, "On the denoising method of prestack seismic data in wavelet domain," in *Proc. SEG Tech. Program Expanded Abstr.*, Jan. 2006, pp. 2852–2856.
- [10] Y. Zhang, X. Tian, X. Deng, and Y. Cao, "Seismic denoising based on modified BP neural network," in *Proc. 6th Int. Conf. Natural Comput.*, Aug. 2010, pp. 1825–1829.
- [11] E. Wang and J. Nealon, "Applying machine learning to 3D seismic image denoising and enhancement," *Interpretation*, vol. 7, no. 3, pp. SE131–SE139, Aug. 2019.
- [12] D. Liu, W. Wang, X. Wang, C. Wang, J. Pei, and W. Chen, "Poststack seismic data denoising based on 3-D convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1598–1629, Mar. 2020.
- [13] Y. Chen, M. Zhang, M. Bai, and W. Chen, "Improving the signal-to-noise ratio of seismological datasets by unsupervised machine learning," *Seismol. Res. Lett.*, vol. 90, no. 4, pp. 1552–1564, 2019.
- [14] M. Zhang, Y. Liu, and Y. Chen, "Unsupervised seismic random noise attenuation based on deep convolutional neural network," *IEEE Access*, vol. 7, pp. 179810–179822, 2019.
- [15] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, "Multiscale seismic waveform inversion," *Geophysics*, vol. 60, no. 5, pp. 1457–1473, Sep. 1995.
- [16] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.
- [17] F. J. Herrmann, D. Wang, G. Hennenfent, and P. P. Moghaddam, "Curvelet-based seismic data processing: A multiscale and nonlinear approach," *Geophysics*, vol. 73, no. 1, pp. A1–A5, Jan. 2008.