A Self-supervised Method Using Noise2Noise Strategy for Denoising CRP Gathers

Xiaokai Wang, *Member, IEEE,* Siyuan Fan, Chen Zhao, Dawei Liu and Wenchao Chen

***Abstract*—With the improvement of computing power and the rapid development of deep learning, deep-learning-based methods are widely used in the field of seismic data noise suppression. Supervised learning has proven to be effective but its performance largely relies on noise-free data labeling, which is often unavailable or an expensive process. Therefore, as a form of unsupervised learning, self-supervised learning emerged to overcome this difficulty, with its labels coming from the training dataset itself. In this letter, we propose a self-supervised learning method that requires only raw seismic data to train the model by using the Noise2Noise strategy, which takes advantage of the unpredictability of noises to regress from noisy data to clean data. Our method aims at improving the noise suppression effect for common-reflection-point (CRP) gathers. By comparing with conventional methods, both synthetic and field data show that the proposed framework is not only effective in suppressing random noise, but also remains effective for coherent noise.**

***Index Terms*—Convolutional neural network, self-supervised learning, common-reflection-point gathers, random noise suppression, coherent noise suppression**

I. Introduction

THE noise suppression in prestack seismic data is a classic problem and has always been a big challenge when processing seismic data in practice. In general, we can divide the seismic data noise suppression methods into two categories, conventional methods and deep-learning-based methods.

The conventional methods can be further divided into three main branches. The first branch is predictive filtering, such as $f-x$ [1] predictive filterings, which are based on the spatial predictability of the useful part of the signal in the frequency domain and the unpredictability of the noise. The second important branch is the noise suppression methods based on mathematical transforms. The transforms used in these methods are mainly wavelet transforms [2], ridgelet transforms [3], curvelet transforms [4] and other multi-scale transforms with local focusing properties. These methods usually use the mathematical transform to transform the data into the transform domain, then use the threshold function to suppress the transform coefficients, and finally obtain the noise-suppressed data with the corresponding inverse transform. The third branch is based on sparse representation of seismic data, which requires that seismic data can be represented as a linear combination of atoms from a dictionary [5] so that we can take advantage of the sparsity difference between the useful signal and noise to discriminate them. The conventional methods mentioned above have been proven to be effective in practice, but they share some common weaknesses, that is, their performance heavily relies on hand-crafted priors and appropriate parameter fine-tuning. In addition, the algorithms of these conventional methods are relatively time-consuming and considered inefficient when processing massive field data.

In recent years, the rapid development of deep learning, especially convolutional neural network, has given rise to many applications in seismic data noise suppression. As an efficient end-to-end method, convolutional neural network not only further improves the noise suppression effect, but also makes up for the weaknesses of conventional methods. The DnCNN [6] was initially used by Yu *et al.* [7] for noise suppression of seismic data, and the experiment proved that the DnCNN of 17 layers was sufficient for seismic data. Liu *et al.* [8] extended the DnCNN network to three dimensions (3D-DnCNN) to take advantage of the strong correlation of seismic data in three dimensions, and obtained a good effect of prestack and poststack data noise suppression. Li *et al.* [9] used the supervised residual learning CNN network to suppress the ground rolls. In 2022, Li *et al.* [10] used CNN to conduct super-resolution and noise suppression of seismic data. All the above belong to the research and application of supervised learning in seismic noise suppression. However, the main challenge with supervised learning is the requirement for noise-free data as labels. When dealing with field data, this is an expensive process or not even always available. Simultaneously, generalization performance is also a continuously discussed problem. Poor generalization performance of a trained model will make the improvement of computing efficiency in vain. In contrast, unsupervised learning eliminates the need for noise-free data labels while providing stronger generalization performance. In 2020, Liu *et al.* [11] applied the generator convolutional neural network(GCN) to prestack seismic data and obtained great results. As an important branch of unsupervised learning, self-supervised learning generates labels for training from inside the training dataset. Recently, the research of Sun *et al.* [12] used self-supervised learning method combined with transfer learning and well suppressed random noise in both prestack and poststack seismic data. Moreover, Noise2Noise [13] is a widely referenced strategy in self-supervised learning, which implies that, without overfitting, the network can merely recover useful signals when the useful signals of the input and labels are strongly coherent and the noises are independent. Based on the Noise2Noise strategy, Shao *et al.* [14] proposed a training data generation

This work was supported by the National Key RD Program of China (2021YFA0716903) and National Natural Science Foundation of China (41974131). (Corresponding author: Wenchao Chen.)

The authors are with the School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China (e-mail: xkwang@xjtu.edu.cn; fsy20000113@stu.xjtu.edu.cn; 978738268@qq.com; 409791715@qq.com; wencchen@xjtu.edu.cn).

This article has been accepted for publication in IEEE Geoscience and Remote Sensing Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LGRS.2023.3285951

strategy to suppress random noise.

In this letter, inspired by the Noise2Noise strategy and the characteristics of CRP gathers, we propose a promising paradigm, which is effective not only for random noise but also for coherent noise. To be more concrete, after processing by means of normal moveout (NMO) correction, the common reflection point gathers have a horizontal events in the space-time domain, because each point of the same time theoretically represents the reflection from the same point underground, which means that the signals in the horizontal direction of the time-space domain are highly self-similar. That characteristic of CRP gathers perfectly meets the requirements of the Noise2Noise strategy and allows us to accomplish our noise suppression task in an easier way, which means we don't need clean labels any more. On this basis, we propose an effective framework in terms of training dataset construction and model training. In section II, we briefly introduce the model formulation and the proposed framework. Section III elaborate on the implementation details and show the results of the proposed method on synthetic data and field data. Eventually, we conclude in section IV.

II. METHOD

A. Model Formulation

Seismic data can be modeled as a superposition of useful signals and noise.

$$\mathbf{x} = \mathbf{y} + \mathbf{n} \tag{1}$$

where $\mathbf{x}$ denotes the raw seismic data, $\mathbf{y}$ denotes the useful signal and $\mathbf{n}$ denotes the noises, including random noises and coherent noises. Supervised learning aims to find the best mapping between the raw data and the useful signal, which can be written as $\mathbf{y} \approx f_\theta(\mathbf{x})$, through a large amount of training with the model so that we can define the optimization objective as follows:

$$\arg\min_\theta \sum \{L(f_\theta(\mathbf{x}), \mathbf{y})\}. \tag{2}$$

Before further formulation, we briefly review the basic hypotheses of the Noise2Noise [13] strategy.

1) The input and label for training have strongly coherent useful signal.
2) The input and label for training have independent noise.

When the above hypothesis is satisfied, the optimization objective based on the Noise2Noise strategy can be rewritten as follows:

$$\arg\min_\theta \sum \{L(f_\theta(\mathbf{x}_1), \mathbf{x}_2)\}. \tag{3}$$

with $\mathbf{x}_1 = \mathbf{y} + \mathbf{n}_1$ and $\mathbf{x}_2 = \mathbf{y} + \mathbf{n}_2$.

Although the optimization objective is changed, our mapping objective remains the same, which is $\mathbf{y} = f_\theta(\mathbf{x})$, because of the hypothesis that the noise between the samples is independent of each other. Therefore, $f_\theta(\mathbf{x}_1)$ can be represented with $\hat{\mathbf{y}}$. To make clear the effect of noise in the labels on the optimization process, taking L2Loss as an example, we can formulate the loss function as

$$\begin{aligned}
L_2(\hat{\mathbf{y}}, \mathbf{y} + \mathbf{n}_2) &= \frac{1}{N}\sum_{n=1}^{N}[\mathbf{y}(n) + \mathbf{n}_2(n) - \hat{\mathbf{y}}(n)]^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\{[\mathbf{y}(n) - \hat{\mathbf{y}}(n)]^2 \\
&\quad + 2 \times \mathbf{n}_2(n) \times [\mathbf{y}(n) - \hat{\mathbf{y}}(n)] + \mathbf{n}_2(n)^2\} \\
&= L_2(\hat{\mathbf{y}}, \mathbf{y}) + \sigma(\mathbf{n}_2)^2 + \frac{2}{N}\sum_{n=1}^{N}\mathbf{n}_2(n)[\mathbf{y}(n) - \hat{\mathbf{y}}(n)]
\end{aligned} \tag{4}$$

The process of model training is the process of applying gradient descent method to the loss function in order to find the minimum. In general, we can divide noise into two categories, Gaussian random noise and other noise represented by coherent noise and non-zero random noise.

For Gaussian random noise, in the above equation, the random noise and useful signal in the third term are independent of each other with the mean value equal to 0. Therefore, when N is large enough, this term tends to be 0. The second term is the variance of noise, which is independent of the model parameter θ. The first remaining term is the loss function in the supervised learning case.

For other noises, coherent noise and non-zero random noise are not necessarily independent. In order to solve it, we return to the Noise2Noise denoising strategy, the core principle of which is to make use of the independence of Gaussian random noise in space-time domain so that its features can not be captured by CNN. When dealing with coherent noise and non-zero random noise, by randomly sliding sampling on the horizontal events as input or label, we can still make coherent noises or non-zero random noises irrelevant so that those features cannot be captured by CNN, which is an extension of the the Noise2Noise principle and the sliding sampling method will be detailed in the section II-B. Therefore, the noise will not affect the parameter optimization during the gradient descent, which means that the proposed strategy is theoretically feasible and reliable.

B. Proposed Framework

In the subsection II-A, we verify the feasibility of the Noise2Noise strategy through the formulation of the loss function and the optimization objective. However, in order to use Noise2Noise strategy, a framework that satisfies two hypotheses is indispensable, which is the key point of our method. On the one hand, the input and label should have strongly coherent useful signal, which is well satisfied in the case of the CRP gathers because of the horizontal events. On the other hand, the noise of input and label should be irrelevant, which is obvious for random noise, but for coherent noise, when we can eliminate the correlation of coherent noise between samples by random sampling. In the algorithm 1, we give the detailed process of the algorithm implementation.

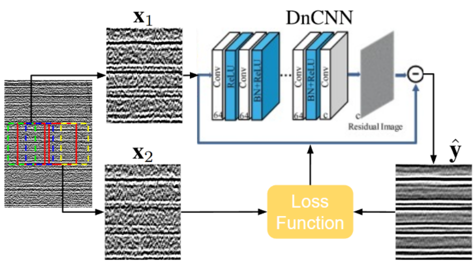

Fig. 1. Schematic of the proposed denoising framework

Algorithm 1 Construction of training data set by sliding sampling

Data: K CRP gathers with size [P, Q]; [L,W], size of samples; N, number of sample groups; M, Sliding repetition; S, Sliding step;
Result: N×(M+1) samples of length L and width W (N groups, each group containing M+1 samples)
Generate N coordinates (k, p, q) with Random function
for *the n-th coordinate (k, p, q)* **do**
  Generate a sample of size L×W centered on the current coordinate (k, p, q) in the k-th gather
  repeat
    Sample again after sliding the sampler along trace direction in step S (centered on the current coordinate (k, p, q+S×Times of repetition) and with size L×W)
  until *M sampling complete*;
  Pack M+1 samples into a group
end
Get N Groups of samples, in each of which there are M+1 samples with size L×W

When composing the training dataset, we use the random function again in order to randomly select two samples from each group of samples as input and label respectively, which ensures that the useful signals are highly similar while making the noise of input and label irrelevant in the vast majority of cases.

The network architecture we use is a 17-layer DnCNN network [6], which is a classical network proposed for image denoising without fully connected and pooling layers, and the padding operation makes the output after the convolutional operation consistent with the input dimension. All hidden layers use the ReLU activation function, and the hidden layers except the input layer are connected with a BN (Batch Normalization) layer after the activation function. The network uses a residual learning method, which means that the output of the output layer is the noise and the denoised data can be obtained by making the difference between the input and the residual. Previous studies have demonstrated that DnCNN is highly usable in seismic data noise suppression, so we selected this network to accomplish 2D seismic data noise suppression.

III. Experiments

In this section, we conducted experiments on synthetic data and field data. It should be noted that all experimental results are based on the same model trained with a synthetic data. First we specify the main parameters shared by both experiments. The block size of the sampler is 64×256, the number of sample groups in the training set is 10000, each DnCNN model is trained with 60 Epochs, with batch size equal to 24 and the "ADAM" optimizer, the initial learning rate is 0.003. Speaking of the comparison method, we chose f-x deconvolution, which is an efficient algorithm and it's widely used in industry as well.

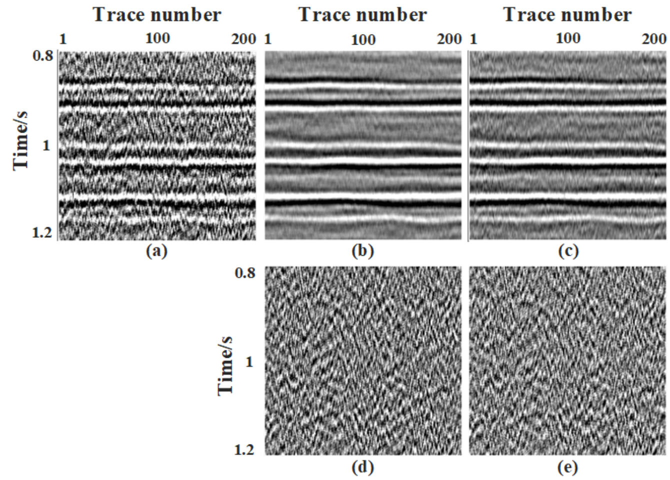

Fig. 2. Denoising comparisons of an enlarged area. Denoised results using (b) our method and (c) f-x deconvolution. Removed noise using (d) our method and (e) f-x deconvolution.

A. Synthetic Experiments

We first evaluate the denoising performance on synthetic data with 200 CRP gathers, each with a size of 200×3001. For synthetic data, we add two types of noise in general, random noise and coherent noise. Coherent noise added here is linear coherent noise. In order to prevent the details we want to keep from being suppressed in practical problems, we also add some fluctuations to clean synthetic data. Fig. 2 shows part of our synthetic data, from which we can clearly see the added noise and fluctuation.

In Fig. 2, the denoising results of our method and f-x deconvolution are displayed. To quantitatively evaluate the effectiveness of our noise suppression, we selected three metrics, peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [15] and denoising efficiency, which is given by timing the operation of the denoising program. The results corresponding to Fig. 2 are in the "Medium quality test sets" section of the table I. In terms of quantitative metrics, our method also has a clear advantage.

Table I shows the denoising results of the well trained model tested on test sets of different quality, including test sets of significantly higher quality than the training set and test sets of significantly lower quality than the training set in terms of PSNR and SSIM. Our method, on different quality test sets, performs better than f-x deconvolution, especially on the test sets with low signal-to-noise ratio to improve the denoising result the most, which indicates that our method has better generalization ability. Regarding the denoising ef-

TABLE I
DENOISING COMPARISON BETWEEN OUR METHOD AND F-X DECONVOLUTION ON NOISED SEISMIC DATA OF DIFFERENT QUALITY

Training sets: PSNR = 15.78, SSIM=0.1839			
High quality test sets: PSNR = 21.78, SSIM=0.4070			
Data type	PSNR(dB)	SSIM	Running time(s)
Our method	32.46	0.8188	99.77
Fx-deconvolution	31.83	0.8170	168.52
Medium quality test sets: PSNR = 14.95, SSIM=0.1953			
Data type	PSNR(dB)	SSIM	Running time(s)
Our method	26.59	0.5999	107.71
Fx-deconvolution	24.76	0.5243	168.80
Low quality test sets: PSNR = 10.32, SSIM=0.1386			
Data type	PSNR(dB)	SSIM	Running time(s)
Our method	22.19	0.5156	106.99
Fx-deconvolution	20.10	0.4246	169.85

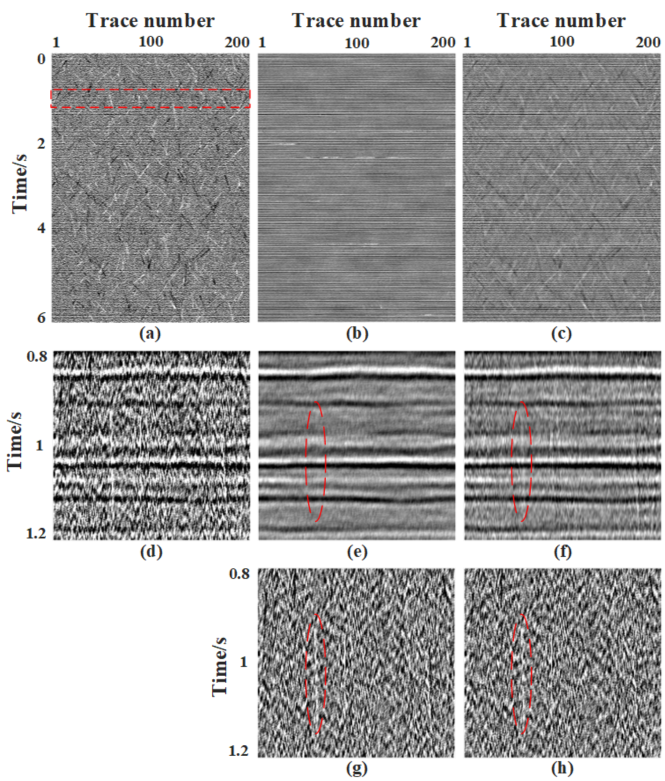

Fig. 3. Denoising comparisons of synthetic seismic data with dense coherent noise. (a) Raw data. Denoised results using (b) our method and (c) f-x deconvolution. (d) Enlarged raw data. Enlarged denoised results using (e) our method and (f) f-x deconvolution. Removed noise using (g) our method and (h) f-x deconvolution.

ficiency, our method reduces the running time by an average of 38% compared to f-x deconvolution thanks to our efficient model and the parallel computing power provided by CUDA. Considering that f-x deconvolution is an efficient algorithm, this is a considerable improvement.

In Fig. 3, to better verify the suppression of coherent noise by our method, we add dense linear coherent noise to the synthetic data for testing, even though this is not common in practice, but it should be valid as a verification of the method. In the enlarged areas in Fig. 3(e) and Fig. 3(f), we can clearly see that the f-x devonvolution results have a lot of linear coherence noise residuals, such as the circled area in Fig. 3(f), which is not the case with our method.

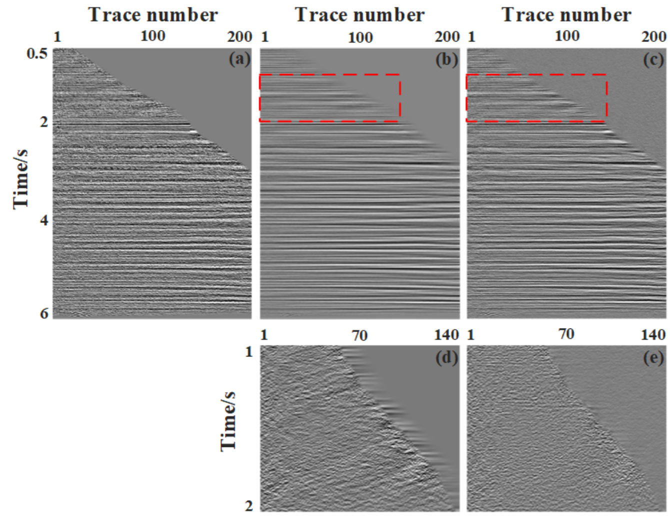

Fig. 4. Denoising comparisons of field seismic data. Denoised results using (b) our method and (c) f-x deconvolution. Removed noise using (d) our method and (e) f-x deconvolution of the enlarged area of the red boxes in (b) and (c).

B. Field Experiments

In this subsection, we conducted experiments on field data with 200 CRP gathers, each with a size of 200×3001. Fig. 4 shows the denoising results of our method and f-x deconvolution. On the one hand, from the noise removed in the enlarged areas in Fig. 4(d) and Fig. 4(e), the f-x deconvolution method shows some useful signal leakage, while our method leads to almost no useful signal leakage with random noise and especially coherent noise well suppressed. On the other hand, the energy of the removed noise of our method is significantly stronger. Simultaneously, in Fig. 4(b), the amplitude variations are well reserved and can not be found in removed noise, which is beneficial to the follow-up AVO analysis.

We further explore the denoising performance on the multichannel average amplitude spectrum. In general, noise in seismic data has mainly high-frequency energy. In Fig. 5, we plotted the multichannel average amplitude spectrum of the raw data and the denoised data. For example, between 50 Hz and 150 Hz, where most of the noise energy is concentrated, the spectrum of the three signals are significantly separated and the spectrum of our method lies below the spectrum of the f-x deconvolution method. And Fig. 6 shows the raw data and the denoised results filtered by 50 Hz high-pass filter. The signal-to-noise ratio of the raw data in Fig. 6(a) decreases significantly in the band superior to 50 Hz, indicating that the band is predominantly noisy, and in both Fig. 6(b) and Fig. 6(c), useful signals can be seen, but there is clearly more noise remaining in Fig. 6(c), which verifies that our method suppresses more of the noise energy. In terms of the removed

This article has been accepted for publication in IEEE Geoscience and Remote Sensing Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LGRS.2023.3285951

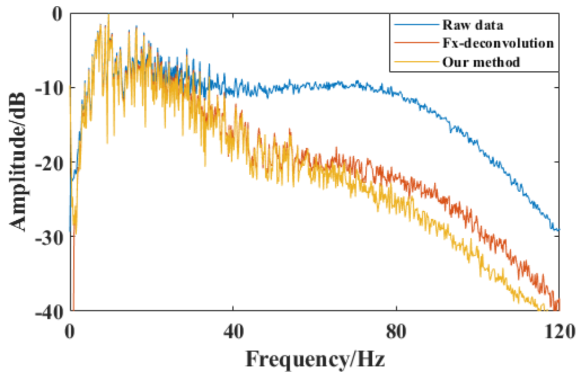

Fig. 5. Multichannel average amplitude spectrum of raw data and denoised data.

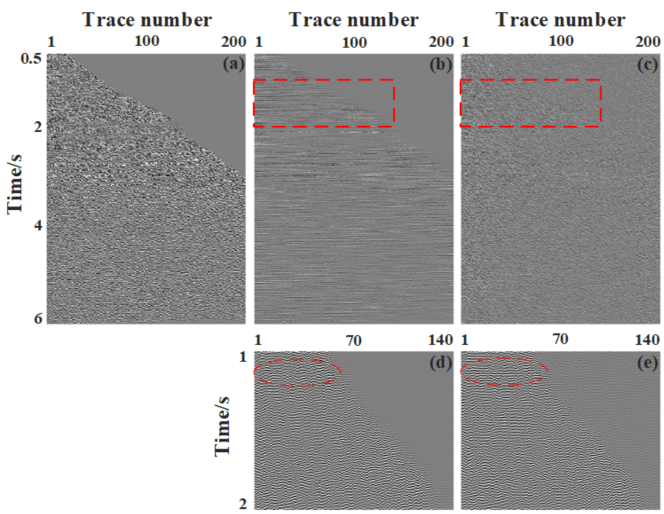

Fig. 6. The 50 Hz high-pass filter results of (a) raw data and denoised data using (b) our method and (c) f-x deconvolution. The 50 Hz high-pass filter results of removed noise using (d) our method and (e) f-x deconvolution of the enlarged area of the red boxes in (b) and (c).

noise higher than 50 Hz, in Fig. 6(d) and Fig. 6(e) we can see less useful signal leakage with our method. For example, in the area circled, less horizontal structures can be oberved in Fig. 6(d), which means less signal leakage.

As processing and interpretation in the pre-stack stage require high-quality CRP gathers, on which the denoising effect of proposed method is remarkable. In Fig. 7, the stack section of removed noise is displayed with the same value range. Our method still shows less useful signal leakage than f-x deconvolution, which is consistent with our previous analysis. The maximum value of the stack section of removed noise is 5.7% of the stack section of denoising result, which is much less than 11.5% of f-x deconvolution, which further illustrates the advantages of our method in terms of fidelity compared to f-x deconvolution.

IV. Conclusion

We proposed an effective self-supervised learning denoising method for both random noise and coherent noise through using and extending the Noise2Noise strategy. In making full use of the characteristics of common-reflection-point gathers,

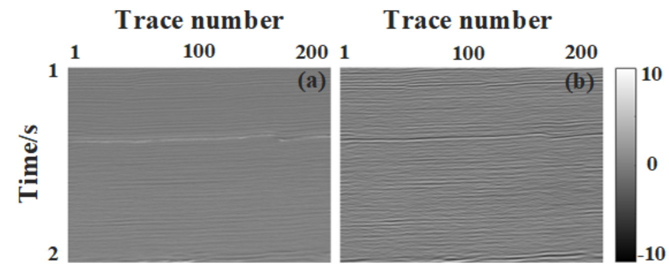

Fig. 7. The stack section of noise removed by (a) our method and (b) f-x deconvolution.

we managed to establish a simple but effective training set construction method, which only needs noisy seismic data. Compared with the methods commonly used in the industry, the proposed method performs better in terms of noise suppression on both synthetic and field data. Moreover, this result can be generalized to test data with significantly lower and significantly higher peak signal-to-noise ratios than the training data. Finally, leakage of useful signal is better limited by the proposed method as well.

References

[1] N. Gulunay, "Fxdecon and complex wiener prediction filter," in *Proc. SEG Tech. Program Expanded Abstr.*, 1986, pp. 279–281.

[2] R. Anvari, M. A. N. Siahsar, S. Gholtashi, A. R. Kahoo, and M. Mohammadi, "Seismic random noise attenuation using synchrosqueezed wavelet transform and low-rank signal matrix approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6574–6581, 2017.

[3] H. Zhang, T. Liu, and Y. Zhang, "Denoising of seismic data via multi-scale ridgelet transform," *Earthq. Sci.*, vol. 22, no. 5, pp. 493–498, 2009.

[4] R. Neelamani, A. I. Baumstein, D. G. Gillard, M. T. Hadidi, and W. L. Soroka, "Coherent and random noise attenuation using the curvelet transform," *Lead. Edge*, vol. 27, no. 2, pp. 240–248, 2008.

[5] 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, 2015.

[6] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017.

[7] S. Yu, J. Ma, and W. Wang, "Deep learning for denoising," *Geophysics*, vol. 84, no. 6, pp. V333–V350, 2019.

[8] D. Liu, W. Wang, W. Chen, X. Wang, Y. Zhou, and Z. Shi, "Random noise suppression in seismic data: What can deep learning do?" in *Proc. SEG Tech. Program Expanded Abstr.*, 2018, pp. 2016–2020.

[9] H. Li, W. Yang, and X. Yong, "Deep learning for ground-roll noise attenuation," in *Proc. SEG Tech. Program Expanded Abstr.*, 2018, pp. 1981–1985.

[10] J. Li, X. Wu, and Z. Hu, "Deep learning for simultaneous seismic image super-resolution and denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2021.

[11] D. Liu, Z. Deng, C. Wang, X. Wang, and W. Chen, "An unsupervised deep learning method for denoising prestack random noise," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2020.

[12] H. Sun, F. Yang, and J. Ma, "Seismic random noise attenuation via self-supervised transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[13] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, 2018, pp. 1–12.

[14] D. Shao, Y. Zhao, Y. Li, and T. Li, "Noisy2noisy: Denoise pre-stack seismic data without paired training data with labels," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

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