Robust Seismic data denoising via zero-shot unsupervised deep learning

Ji Li, Daniel Trad

ABSTRACT

Seismic data denoising is a critical component of seismic data processing, yet effectively removing erratic noise, characterized by its non-Gaussian distribution and high amplitude, remains a substantial challenge for conventional methods and deep learning (DL) algorithms. Supervised learning frameworks typically outperform others, but they require pairs of noisy datasets alongside corresponding clean ground truth, which is impractical for real-world seismic datasets. On the other hand, unsupervised learning methods, which don't rely on ground truth during training, often fall short in performance when compared to their supervised or traditional denoising counterparts. Moreover, no existing unsupervised DL method adequately addresses the unique complexities of erratic seismic noise. This paper introduces a novel zero-shot unsupervised DL framework designed specifically to mitigate random and erratic noise, with a particular emphasis on blended noise. Drawing inspiration from Noise2Noise and data augmentation principles, we present a robust self-supervised denoising network named "Robust Noiser2Noiser." Our approach eliminates the need for paired noisy and clean datasets as required by supervised methods or paired noisy datasets as in Noise2Noise (N2N). Instead, our framework relies solely on the original noisy seismic dataset. Our methodology generates two independent re-corrupted datasets from the original noisy dataset, using one as the input and the other as the training target. Subsequently, we employ a deep-learning-based denoiser, DnCNN, for training purposes. To address various types of random and erratic noise, the original noisy dataset is re-corrupted with the same noise type. This paper is specifically focused on solving the problem of deblending in seismic data. Detailed explanations for generating training input and target data for blended data are provided in the paper. We apply our proposed network to both synthetic and real marine data examples, demonstrating significantly improved noise attenuation performance compared to traditional denoising methods and state-of-the-art unsupervised learning methods.

INTRODUCTION

Seismic data denoising stands as a fundamental pillar in the realm of seismic data processing. This noise encompasses coherent and incoherent components, as elucidated by Abma and Claerbout (1995). Ground roll, as elucidated by Deighan and Watts (1997), represents the prevalent form of coherent noise in land seismic data. On the other hand, incoherent noise is divided into two distinctive categories: high-frequency, low-amplitude random noise, and high-amplitude erratic noise. In previous decades, various methodologies have emerged to combat random noise, primarily relying on sparse transformation techniques. These methods transform seismic data into sparse domains, from which signals and noise are subsequently separated. Popular sparse transform domains in geophysics encompass the Radon transform (Ibrahim and Sacchi, 2014; Latif and Mousa, 2016), Fourier transform (Hennenfent and Herrmann, 2008; Yu et al., 2015), wavelet transform (Cao and Chen, 2005; Mousavi and Langston, 2016), and curvelet transform (Hennenfent and Her-

rmann, 2006; Górszczyk et al., 2014). An alternative set of methods hinges on rank reduction, positing that seismic data exhibits low-rank characteristics in the frequency domain. These approaches, such as multichannel signal spectrum analysis (MSSA) (Oropeza and Sacchi, 2011; Siahsar et al., 2017) and singular value decomposition (SVD) (Lari et al., 2019), seek to restore signals by reconstructing low-rank matrices.

However, the realm of erratic noise attenuation poses a more formidable challenge compared to random noise mitigation. Robust methodologies have sought to replace the ℓ_2 norm cost function found in non-robust methods with robust M-estimators (Maronna, 1976). Notably, Guitton and Symes (2003) substituted the conventional ℓ_2 norm with the Huber norm, enhancing seismic data handling in the presence of outliers. Similar efforts by (Trickett et al., 2012) incorporated rank reduction filters to mitigate erratic noise. Li and Sacchi (2021) introduced a sparse and robust Radon transform, estimated via Matching Pursuit, to address simultaneous source separation problems.

In recent years, deep learning (DL) techniques have gained substantial traction in geophysics, influencing seismic inversion (Li et al., 2019; Zheng et al., 2019), random and erratic seismic data denoising (Richardson and Feller, 2019; Liu et al., 2018; Wang and Hu, 2021), fault detection (Helbing and Ritter, 2018), and seismic interpolation (Wang et al., 2019; Kaur et al., 2021; Oliveira et al., 2018). DL techniques broadly fall into two categories: supervised learning (SL) and unsupervised learning (UL). SL mandates vast quantities of paired clean and noisy datasets for training, a formidable challenge given the scarcity of such paired data in the real seismic domain. Traditional methods have been employed to generate denoised data for training labels, but their imperfections and substantial computational demands are limiting factors.

To address these challenges, various unsupervised and self-supervised methods have emerged to tackle different forms of seismic noise. Liu et al. (2020) harnessed the generator convolutional neural network (GCN) to combat random noise in prestack seismic data. Sun et al. (2022) employed self-supervised transfer learning to mitigate random noise in seismic data. Qian et al. (2022) introduced a deep convolutional autoencoder with the Welsch function to attenuate random and erratic noise. Saad et al. (2021) proposed an unsupervised learning approach based on DIP (Deep Image Prior)(Ulyanov et al., 2018) to eliminate random noise in 3-D seismic data. Self-supervised learning, a vital branch of unsupervised learning, generates labels directly from the training dataset, with Noise2Noise (Lehtinen et al., 2018) being a prominent strategy in this domain. Wang et al. (2023) proposed a self-learning method based on the Noise2Noise strategy to reduce random noise. Nonetheless, most unsupervised learning methods are tailored exclusively for random noise, which does not pose a significant challenge in seismic data processing. Moreover, the performance of UL methods remains suboptimal compared to their supervised counterparts.

In this paper, we introduce a pioneering self-supervised framework designed to address both random and erratic noise, with a specific focus on mitigating blending noise in simultaneous source acquisition data. Drawing inspiration from Noise2Noise (Lehtinen et al., 2018) and data augmentation methods, we present a robust variant called robust Noiser2Noiser, which is capable of attenuating both random and erratic noise. Our framework operates in a zero-shot, self-supervised manner, eliminating the need for data other

than the original noisy dataset. This approach involves independently re-corrupting the original noisy data to generate two independent re-corrupted datasets, using one as the training input and the other as the training label. Subsequently, we employ a commonly used denoising convolutional neural network DnCNN (Zhang et al., 2017) to train the input and label pairs. The specific re-corruption method depends on the type of noise being addressed; for random noise, we re-corrupt the data with additional random noise to create input-label pairs, while for erratic noise, we apply the same type of erratic noise and utilize a robust ℓ_1 norm loss function instead of the conventional mean-square error (MSE) loss function.

Our method is applied to both synthetic and real data examples featuring random noise and erratic noise. Notably, we employ the proposed technique to remove blending noise stemming from simultaneous source acquisition, an erratic noise subtype.

THEORY

Unsupervised and Self-Supervised Denoising Networks

Traditionally, the training of denoising networks has heavily relied on accessing paired datasets containing noisy and clean data for supervised learning. However, a significant breakthrough in this paradigm emerged with the introduction of Noise2Noise by Lehtinen et al. (2018). This pioneering work demonstrated that denoising networks trained on noisy/noisy image pairs can perform remarkably close to those trained on noisy/clean data pairs from the same dataset.

To elaborate, consider a scenario where we have a pair of noisy data samples,

$$\mathbf{y}_1 = \mathbf{x} + \mathbf{e}_1,$$
$$\mathbf{y}_2 = \mathbf{x} + \mathbf{e}_2,$$

with $\mathbf{e}_1$,$\mathbf{e}_2$ representing independent noise sources. A network F_ϕ is then trained to minimize the Noise2Clean Mean Squared Error (MSE) loss function:

$$\begin{aligned}&\text{argmin}E\{\parallel F_\phi(\mathbf{y}_1) - \mathbf{x} \parallel_2^2\}\\ &= \text{argmin}E\{\parallel F_\phi(\mathbf{y}_1) \parallel_2^2 - 2\mathbf{x}^T F_\phi(\mathbf{y}_1)\}.\end{aligned}$$

This function aims to reduce the difference between the network's output when applied to $\mathbf{y}_1$ and the clean signal $\mathbf{x}$. Additionally, Noise2Noise employs a similar approach but with a different loss function:

$$\begin{aligned}&\text{argmin}E\{\parallel F_\phi(\mathbf{y}_1) - \mathbf{y}_2 \parallel_2^2\}\\ &= \text{argmin}E\{\parallel F_\phi(\mathbf{y}_1) - \mathbf{x} - \mathbf{e}_2 \parallel_2^2\}\\ &= \text{argmin}E\{\parallel F_\phi(\mathbf{y}_1) \parallel_2^2 - 2\mathbf{x}^T F_\phi(\mathbf{y}_1) - 2\mathbf{e}_2^T F_\phi(\mathbf{y}_1)\}.\end{aligned}$$

In this case, we aim to minimize the difference between the network's output for $\mathbf{y}_1$ and $\mathbf{y}_2$, or equivalently, the difference between the network's output for $\mathbf{y}_1$ and the clean signal $\mathbf{x}$ corrupted by $\mathbf{e}_2$.

Notably, when $\mathbf{e}_1$ and $\mathbf{e}_2$ are independent of each other, the term $2\mathbf{e}_2^T F_\phi(\mathbf{y}_1)$ simplifies to 0. Consequently, the MSE loss functions for Noise2Clean and Noise2Noise yield equivalent results. This observation suggests that training a denoising network on noisy data samples with the same clean signals can yield outcomes akin to training on clean/noisy pairs.

Nevertheless, it's essential to acknowledge that acquiring noisy dataset pairs that precisely align with the clean data is often as challenging as obtaining corresponding noisy/clean pairs. As a result, recent research endeavors have focused on training deep learning models using noisy datasets exclusively, thus obviating the need for pairwise correspondence between noisy and clean data samples. This transition marks a significant step forward in developing unsupervised and self-supervised learning methods for noise removal in various applications, including seismic data processing.

In the pursuit of noise removal without the need for paired noisy and clean data, several innovative self-supervised techniques have emerged. These approaches redefine the conventional supervised learning paradigm and have shown promising results in various applications, including image denoising. Here, we delve into some notable methodologies that have propelled the field forward.

1. **Blind Spot Prediction Techniques**: Noise2Void (Krull et al., 2019) and Noise2Self (Batson and Royer, 2019) are pioneering methods that employ blind spot prediction. They predict the value of a pixel by considering its surrounding context. These techniques operate under the assumption that the corruption is zero-mean and independent across pixels. The promising results from these methods have paved the way for more advanced self-supervised approaches.

2. **Self2Self for Single-Image Denoising**: Building upon the concept of blind spot prediction, Self2Self (Quan et al., 2020) has made substantial strides in achieving single-image denoising results that rival those of traditional fully trained methods. By leveraging the surrounding pixel information, Self2Self has demonstrated remarkable denoising capabilities without needing clean data pairs.

3. **Re-Corruption-Based Frameworks**: Another class of self-supervised techniques involves re-corrupting noisy images to create even noisier versions. These augmented datasets are then used to train networks to map the noisier images to the original noisy data. Noteworthy methods in this category include Noiser2Noise (Moran et al., 2020), Noise-as-clean (Xu et al., 2020), and Recorrupted-to-Recorrupted (Pang et al., 2021). These techniques show that increased noise levels in training data can improve the network's ability to denoise.

4. **Sub-Sampling**: Recent developments like Neighbour2Neighbour (Huang et al., 2021) showcase innovative ways to leverage a single noisy dataset. By sub-sampling the dataset, they generate a pair of noisy datasets for training, exploiting data augmentation to enhance denoising performance. Additionally, Lequyer et al. (2022) and Mansour and Heckel (2023) have harnessed similar ideas, optimizing the training process for greater efficiency.

5. **Deep Image Prior (DIP)**: DIP (Deep Image Prior)(Ulyanov et al., 2018) is another influential self-supervised technique that capitalizes on the ability of Convolutional Neural Networks (CNNs) to fit natural images rapidly compared to noise. By employing early stopping during training, DIP can effectively reconstruct a clean image before introducing noise.

These self-supervised methodologies collectively represent a paradigm shift in noise removal, eliminating the need for extensive paired datasets and offering versatile solutions for various applications, including seismic data processing and image denoising. As research in this domain continues to evolve, these techniques hold significant promise for addressing complex noise-related challenges.

Robust Noiser2Noiser

Let's consider $\mathbf{x}$ as our representation of clean data. The noisy data $\mathbf{y}$ can be expressed as

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

Here, $\mathbf{n}$ can be either random noise or erratic noise.

As previously discussed, denoising networks perform similarly when trained on noisy/noisy image pairs compared to noisy/clean data pairs from the same dataset. The question is how to construct a pair of noisy datasets $\mathbf{y}_1$ and $\mathbf{y}_2$ with independent noise from a single noisy dataset $\mathbf{y} = \mathbf{x} + \mathbf{n}$.

Methods such as 'Noiser2Noise' (Moran et al., 2020) and 'Noisy-as-Clean' (Xu et al., 2020) utilize a noisier image as input, where they synthesize noise $\mathbf{z}$, and then train the denoising network on the dataset pair $(\mathbf{y} + \alpha\mathbf{z}, \mathbf{y})$. Meanwhile, Pang et al. (2021) trained the denoising model on the pair $(\mathbf{y} + \alpha\mathbf{z}, \mathbf{y} - \mathbf{z}/\alpha)$, which results in a loss function more statistically connected to the supervised approach. We've tested both augmentation methods and found the results are very close. For our robust Noiser2Noiser Seismic Denoising Network, we adopt the same method as Pang et al. (2021). Additionally, we employ a symmetric loss function (Chen and He, 2021) to train a Siamese network. The loss function is then defined as:

$$\text{argmin}\frac{1}{2}E \parallel F_\phi(\mathbf{y}_1) - \mathbf{y}_2 \parallel_2^2 + \frac{1}{2}E \parallel F_\phi(\mathbf{y}_2) - \mathbf{y}_1 \parallel_2^2 . \tag{2}$$

Our experiments have demonstrated that employing the residual learning technique, as introduced by Zhang et al. (2017), leads to notable enhancements in denoising performance. With residual learning, the network is trained to optimize against the noise component rather than the raw image data. Consequently, this approach transforms the final loss function into:

$$\text{argmin}\frac{1}{2}E \parallel \mathbf{y}_1 - F_\phi(\mathbf{y}_1) - \mathbf{y}_2 \parallel_2^2 + \frac{1}{2}E \parallel \mathbf{y}_2 - F_\phi(\mathbf{y}_2) - \mathbf{y}_1 \parallel_2^2 . \tag{3}$$

In the case of a dataset with erratic noise, we replace the ℓ_2 norm with ℓ_1 to make the

loss function more robust:

$$\text{argmin}\frac{1}{2}E \parallel \mathbf{y}_1 - F_\phi(\mathbf{y}_1) - \mathbf{y}_2 \parallel_1^1 + \frac{1}{2}E \parallel \mathbf{y}_2 - F_\phi(\mathbf{y}_2) - \mathbf{y}_1 \parallel_1^1 . \tag{4}$$

Now that we have our training input and target pairs and the corresponding loss functions, the next crucial step is selecting the network architecture. The proposed Robust Self-Supervised denoising framework is compatible with various network architectures, such as ResNet (He et al., 2016), denoising convolutional neural network (DnCNN) (Zhang et al., 2017), and U-net (Ronneberger et al., 2015). In this study, we opt for DnCNN due to its straightforward architecture and effective residual learning strategy. It has successfully addressed numerous seismic denoising challenges in recent years, including random noise (Zhang et al., 2018), ground roll (Li et al., 2018), and blended noise (Matharu et al., 2020).

The architecture of our proposed network is depicted in Figure 1. We generate the training input and target pairs from the original noisy data. The training input is $\mathbf{y}_1 = \mathbf{y} + \alpha\mathbf{z}$, and the target is $\mathbf{y}_2 = \mathbf{y} - \mathbf{z}/\alpha$. Here, $\mathbf{z}$ represents the same type of noise as in the dataset but is independent of the noise present in the original data. These paired re-corrupted noisy data are trained using the DnCNN to minimize the specified loss function. We employ the ℓ_2 norm to attenuate random noise and the ℓ_1 norm for erratic noise. It's worth noting that the training input and target should be normalized to ensure that the training process is more stable.

Noise Simulatation

When dealing with random noise, generating an independent set of synthetic random noise is a straightforward process. We employ additive white Gaussian noise (AWGN), denoted as $\mathbf{n}$, with a distribution $\mathbf{n} \backsim \mathbf{N}(0, \sigma^2\mathbf{I})$. This noise is zero-mean and independent of the data $\mathbf{y}$. However, generating erratic noise is a more intricate task. To illustrate this, let's consider a specific example: blended noise. Blended acquisition, or simultaneous source acquisition, is an acquisition method used to reduce costs. It involves firing multiple seismic sources at short, random time intervals (Beasley et al., 1998; Berkhout, 2008). Blended acquisition can be viewed as a time-shifting operation applied to data from individual sources, mathematically represented as:

$$\mathbf{b} = \mathbf{B}\mathbf{D}, \tag{5}$$

Here, $\mathbf{B}$ represents the blending operator, $\mathbf{D}$ is the desired data cube obtained through a conventional seismic survey, and $\mathbf{b}$ represents the blended data. The adjoint operator, $\mathbf{B}^*$, corresponds to the pseudo-deblending operation:

$$\tilde{\mathbf{D}} = \mathbf{B}^*\mathbf{b}. \tag{6}$$

The resulting $\tilde{\mathbf{D}}$ represents pseudo-deblended data, which contains interference from other sources and is considered blended noise. This blended noise exhibits coherence in the common shot gather but appears as incoherent erratic noise in other types of gathers, such as common receiver gather, common offset gather, or common midpoint gather.

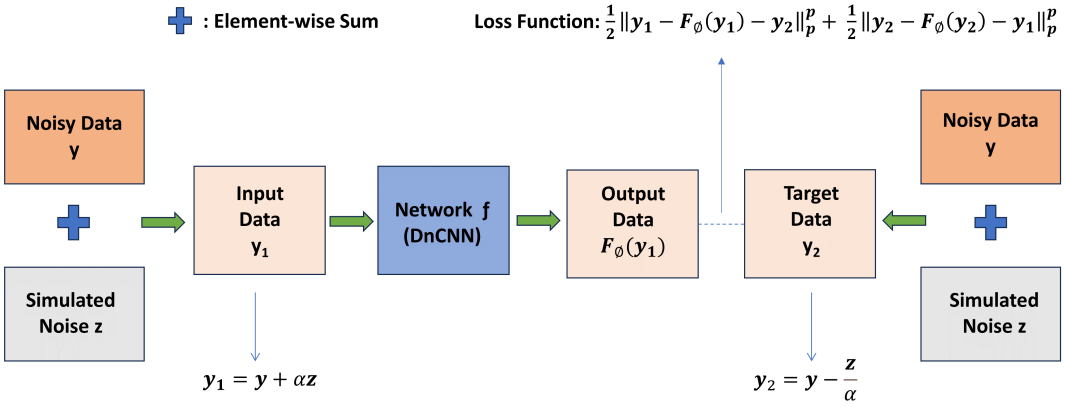

FIG. 1. The architecture of the proposed robust self-learning network. In our framework, we use the re-corrupted noisy data $\mathbf{y}_1 = \mathbf{y} + \alpha\mathbf{z}$ as the input, and re-corrupted noisy data $\mathbf{y}_2 = \mathbf{y} - \mathbf{z}/\alpha$ as target. After training, the original noisy data $\mathbf{y}$ is used to estimate the clean data $F_\phi(\mathbf{y})$.

Suppose we have a pseudo-deblended dataset $\tilde{\mathbf{D}}_1$. To generate synthetic blended noise for training purposes, we can apply the blending and deblending operators with random time shifts to $\tilde{\mathbf{D}}_1$, resulting in another pseudo-deblended dataset $\tilde{\mathbf{D}}_2$. $\tilde{\mathbf{D}}_2$ contains the same coherent seismic signal as $\tilde{\mathbf{D}}_1$ but with additional erratic blended noise. Subsequently, the synthetic blended noise is calculated as $\mathbf{BN} = \tilde{\mathbf{D}}_2 - \tilde{\mathbf{D}}_1$.

For training, we use $(\tilde{\mathbf{D}}_1 + \alpha\mathbf{BN}$ as the input and $\tilde{\mathbf{D}}_1 + \mathbf{BN}/\alpha)$ as the target."

EXAMPLES

In this section, we apply our proposed denoising framework to both synthetic and real marine data examples. The network implementation is done using PyTorch (Paszke et al., 2019), and all tests are conducted on a PC equipped with an NVIDIA GeForce GTX 4060 Ti GPU.

We begin by applying our method to two 2D synthetic examples: a simple one and a shot gather generated using the finite difference method. Subsequently, we tackle the challenging deblending problem using the Robust Noiser2Noiser framework. Our evaluation extends to three deblending scenarios:

1. The blended data created using the finite difference method.
2. Real marine data that we numerically blended
3. Real-world blended marine data collected through the simultaneous source acquisition method.

This comprehensive evaluation demonstrates the versatility and effectiveness of our approach across a range of datasets and scenarios.

To assess the performance of our denoising method on synthetic data examples and manually blended real data examples, we utilize the signal-to-noise ratio (SNR) as a metric for comparison. The SNR is defined as follows:

$$SNR_{out} = 10 \log \frac{\|\mathbf{d}_c\|_2^2}{\|\mathbf{d_c} - \mathbf{d_r}\|_2^2}, \tag{7}$$

Here, $\mathbf{d_c}$ represents the original clean data, and $\mathbf{d_r}$ is the denoised output. For the real simultaneous source seismic data, where we lack a ground truth for direct SNR computation, we visualize the errors between the original data and the deblended result to illustrate any signal leakage."

Synthetic examples denoising

Simple 2-D example

We explore the proposed framework by conducting tests on simple 2-D synthetic data. These tests encompass scenarios involving both random noise and erratic noise. To evaluate the denoising performance, we compare our approach with two other zero-shot unsupervised frameworks, namely Noise2Void (Krull et al., 2019) and Neighbour2Neighbour (Huang et al., 2021).

Our initial test focuses on random noise, for which we employ the ℓ_2 norm during training. Figure 2a and 2b respectively illustrate the clean and noisy data. This 2-D data section exhibits three linear events and operates at a sampling rate of 4 ms. To mitigate the impact of random noise, we manually re-corrupted the noisy dataset by introducing additional random noise, thereby generating training input and labels. Figure 3a and 3b present the re-corrupted training input and labels. The denoising results of these three methods are depicted in Figure 4. In the case of random noise, Neighbour2Neighbour exhibits the weakest denoising performance, whereas Noise2Void and the proposed robust denoising network deliver effective and comparable denoising results.

We then evaluate the robust Noiser2Noiser denoising algorithm on erratic noise, specifically, blending noise. Figure 5b showcases the same 2-D synthetic data example but with blending noise manually introduced by us. To effectively address blending noise, we re-corrupted the noisy dataset, this time incorporating blending noise. The blending is performed with varying random time shifts. Figure 6 displays the generated training input and labels. To account for blending noise, we opt for the ℓ_1 norm loss function, known for its robustness against erratic noise, to replace the ℓ_2 norm. This choice holds for all three methods. The deblending outcomes from the three methods are visually presented in Figure 7, with the results revealing that only the proposed robust Noiser2Noiser denoising network effectively removes the blending noise.

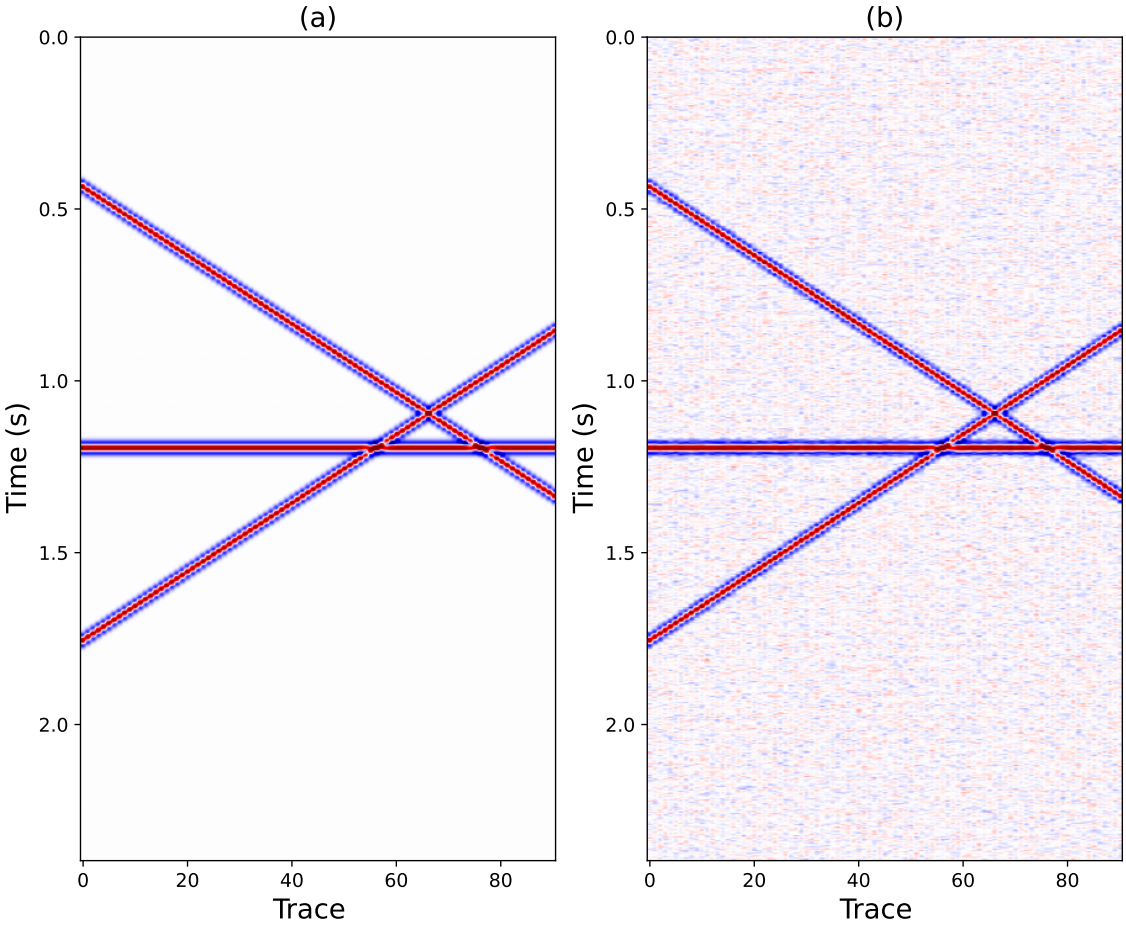

FIG. 2. 2-D synthetic example. (a) Clean data. (b) Noisy data with random noise.

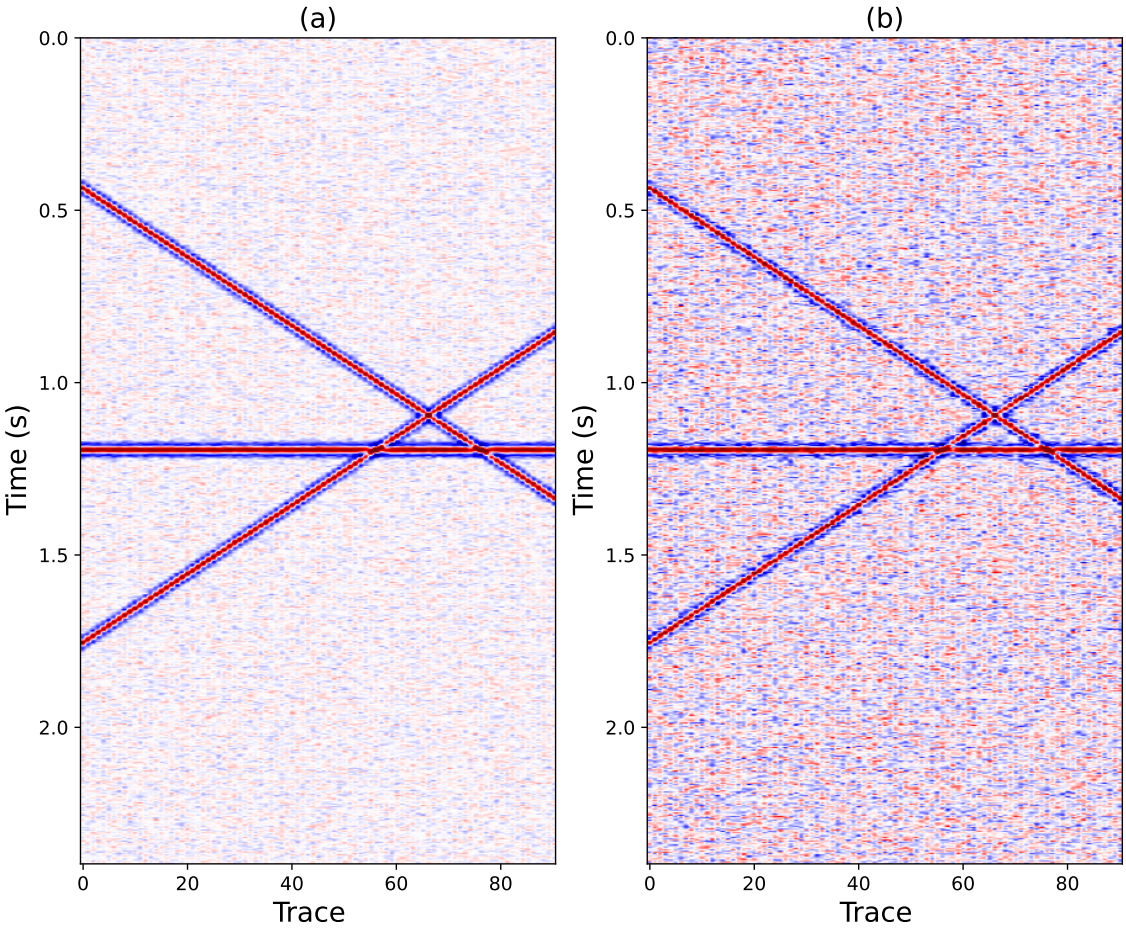

FIG. 3. Re-corrupted data pair. (a) Input. (b) Label.

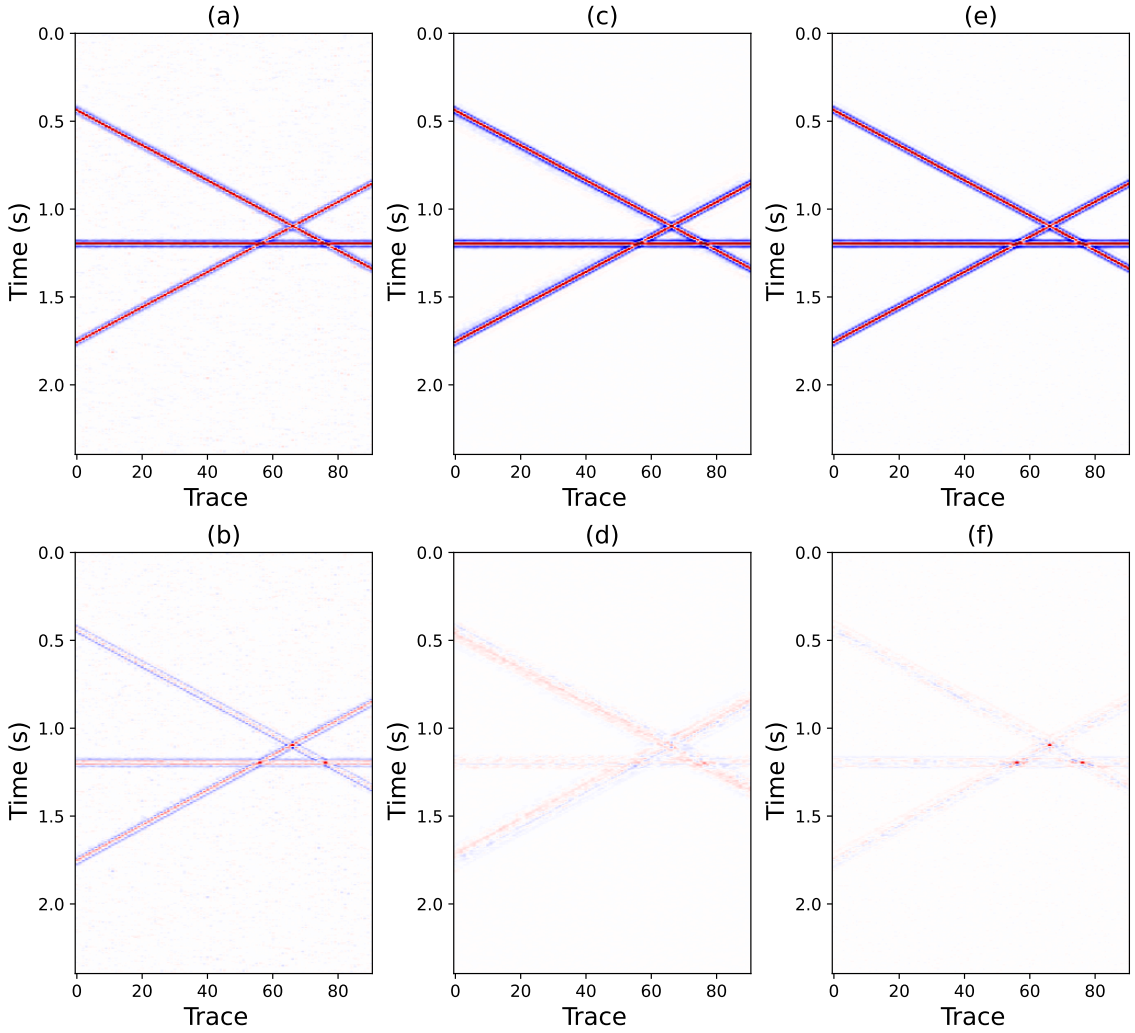

FIG. 4. Denosing result of synthetic data with random noise. (a) Neighbour2Neighbour (SNR=10.2 dB). (b) Errors between (a) and clean data. (c) Noise2Void (SNR=18.3 dB). (d) Errors between (c) and clean data. (e) Robust Noiser2Noiser network (SNR=18.1 dB). (f) Errors between (e) and clean data.

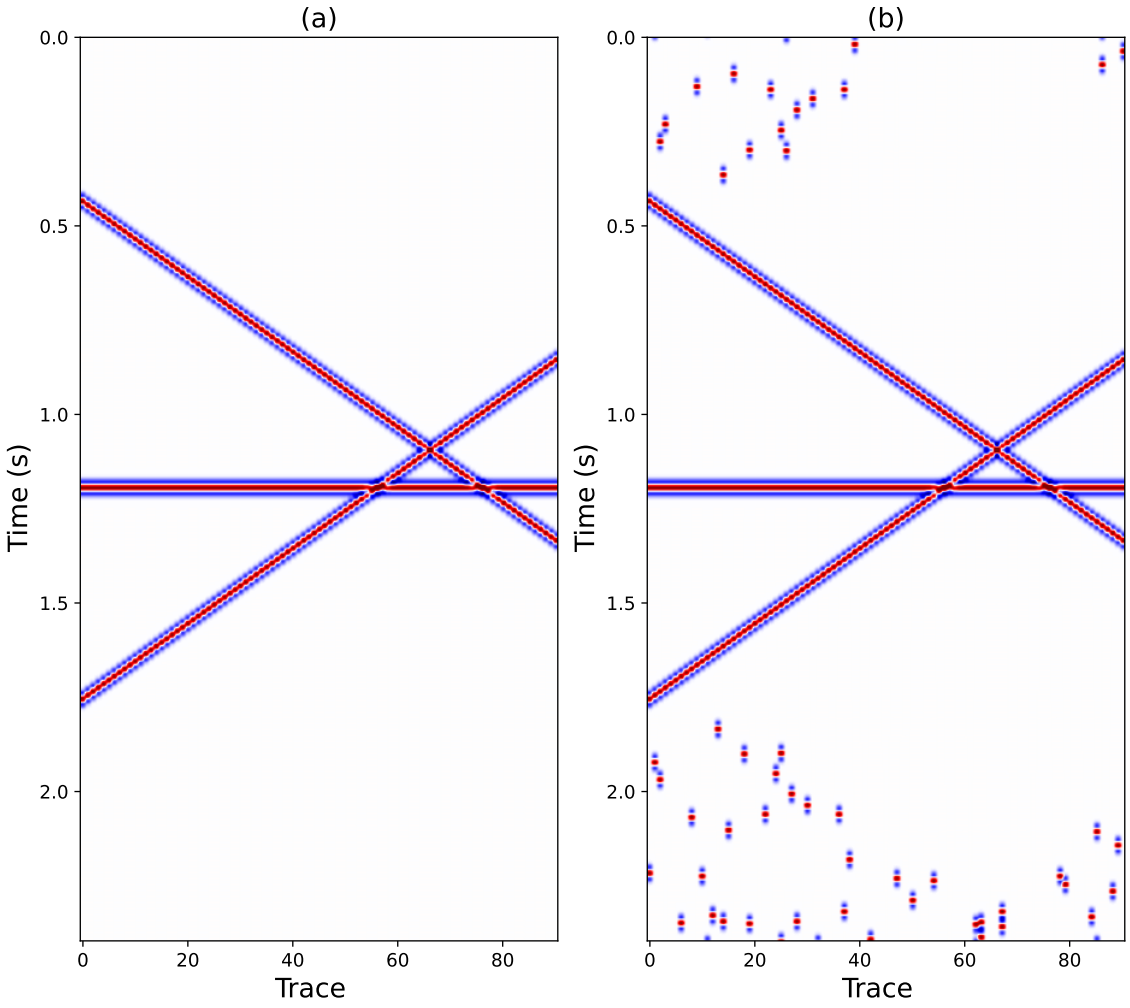

FIG. 5. 2-D synthetic example. (a) Clean data. (b) Noisy data with blending noise.

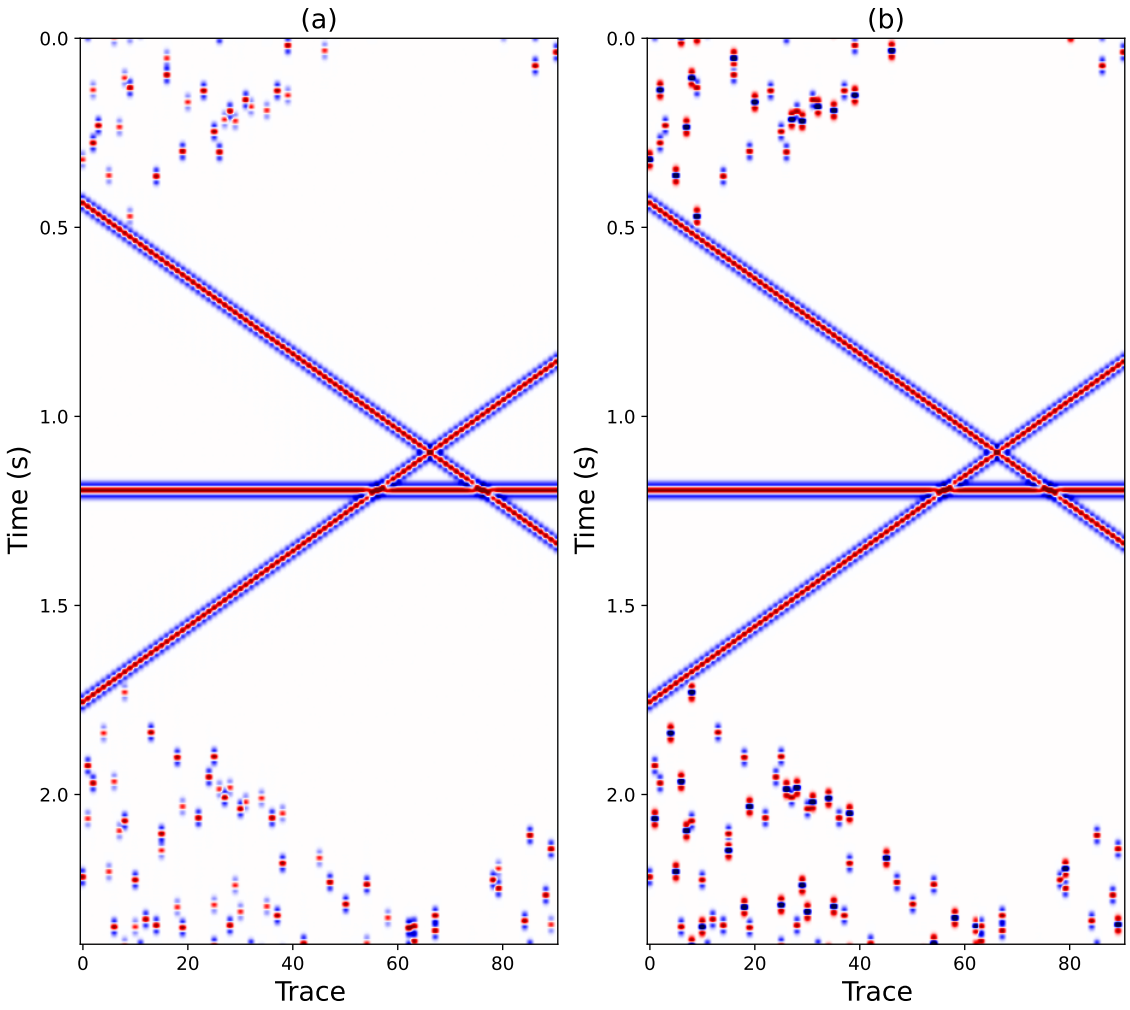

FIG. 6. Re-corrupted data pair. (a) Input. (b) Label.

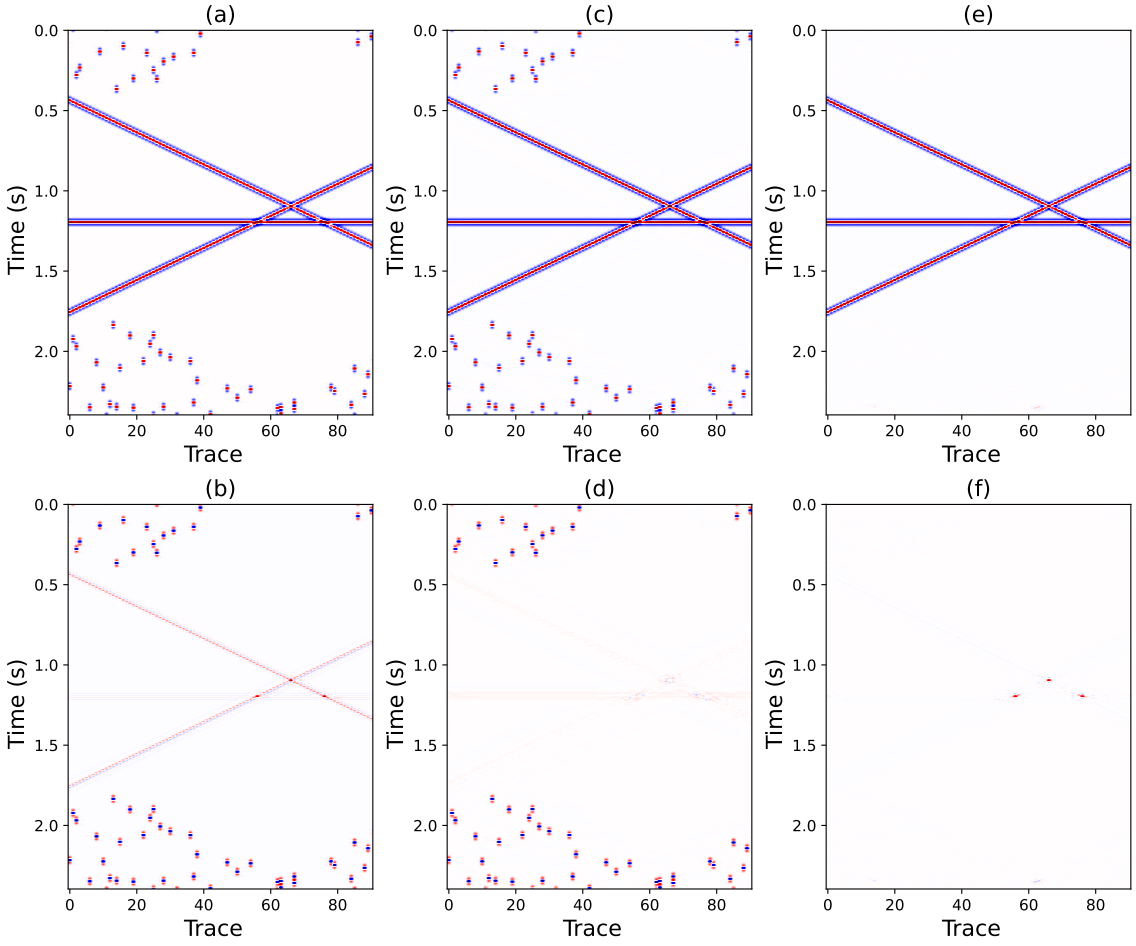

FIG. 7. Denosing result of synthetic data with blending noise. (a) Neighbour2Neighbour (SNR=7.1 dB). (b) Errors between (a) and clean data. (c) Noise2Void (SNR=7.4 dB). (d) Errors between (c) and clean data. (e) Robust Noiser2Noiser (SNR=22.1 dB). (f) Errors between (e) and clean data.

2-D finite difference synthetic example

The previous example underscores the effectiveness of our proposed robust Noiser2Noiser framework in handling both random and erratic noise scenarios. Consequently, in the subsequent tests, we focus solely on evaluating the deblending performance of our method across different datasets. Random noise attenuation, as demonstrated, poses no significant challenge in the context of seismic data denoising.

We now apply our robust Noiser2Noiser framework to a more complex 2-D synthetic seismic shot generated using the finite difference method. This 2-D synthetic shot gather exhibits greater complexity compared to the previous example. As depicted in Figure 8a and 8b, we present the clean data and the shot gather with blending noise. The signal-to-noise ratio (SNR) for the noisy shot gather is measured at -1.21 dB. To create the necessary training input and labels, we introduce random time shifts during the blending process, yielding the re-blended training input and labels showcased in Figure 8 c and d.

Subsequently, Figure 8 e and f showcases the deblending results alongside error comparisons with the original clean shot gather. Notably, our approach leads to a remarkable improvement in the signal-to-noise ratio, from -1.21 dB to an impressive 21.2 dB.

Deblending examples

The preceding two examples demonstrate the effectiveness of our robust Noiser2Noiser framework in successfully mitigating erratic noise. We now turn our attention to applying this methodology to address the deblending problem.

In the context of blended datasets, which are typically 3D cubes containing time samples, receivers, and shots as dimensions, we observe that the blended noise in the common shot domain exhibits coherence. However, this translates into incoherent erratic noise in the common receiver domain. Consequently, the deblending process can be conceptualized as the removal of erratic noise across all common receiver gathers.

To tackle the task of removing blending noise, we have three main approaches:

1. **Training on Whole Dataset (with Patching)**: This method involves training on the entire dataset, which comprises numerous shot gathers simultaneously. However, due to memory constraints, we must partition the data into smaller patches.
2. **Training on Shot Gathers Individually (with Patching)**: Here, we train on each shot gather separately, breaking each into smaller patches for training purposes.
3. **Training on Shot Gathers Individually (without Patching)**: In this approach, we train directly on each complete shot gather.

In our specific scenario, leveraging an RTX 4060 Ti GPU, we conducted experiments with a 3D cube of dimensions 2000 x 350 x 350. When we divided the dataset into 32 x 32 patches with a 50% overlap, training a single epoch on the entire dataset took approximately

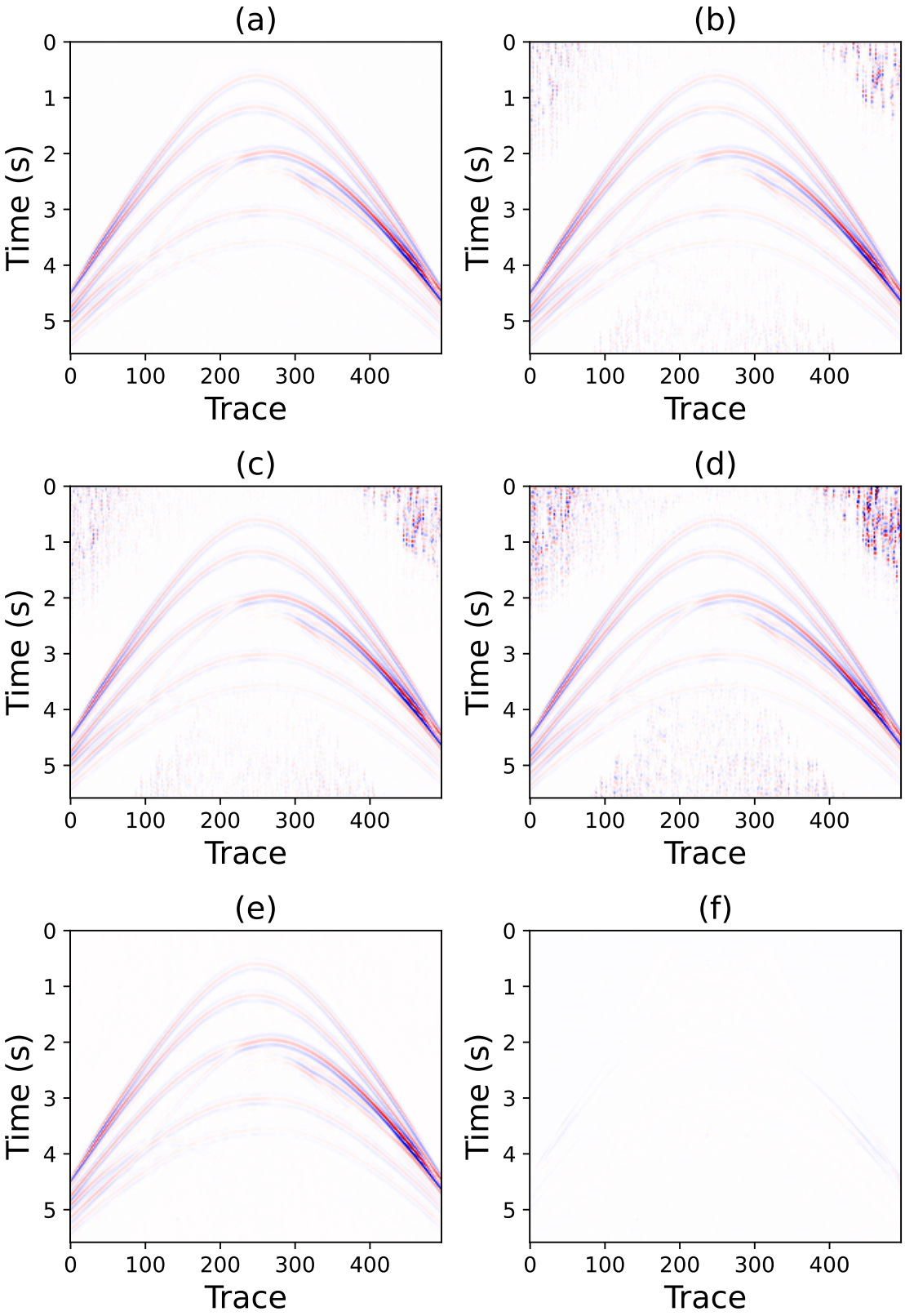

FIG. 8. 2-D finite difference synthetic example. (a) Clean data. (b) Noise data with blending noise (SNR=–1.21 dB). (c) Re-corrupted training input. (d) Re-corrupted training label. (e) Deblended result by the robust Noiser2Noiser Network (SNR=21.2 dB). (f) Errors between deblended result and clean data.

2 hours. However, a significantly faster approach emerged when training on individual shot gathers, with each shot gather divided into 32 x 32 patches. In this case, it required only about 1.2 seconds per epoch, totalling roughly 6 seconds for deblending one shot gather (equivalent to approximately 5 epochs).

When opting to train on the entire shot gather as a whole, without patching, the time dropped even further. Training took approximately 0.4 seconds per epoch, with a total of 20 seconds (equivalent to 50 epochs) for deblending one shot gather. While this method resulted in a longer training time compared to patch-based training, the improvement in deblending performance justified the investment. A total of 20 seconds per shot was deemed acceptable in exchange for superior results.

As a result, for all subsequent deblending experiments, we adopted the approach of applying our algorithm to shot gathers individually, without patching. We initialized the weights using the Kaiming initialization method (He et al., 2015), which is well-suited for ReLU activation functions. This initialization greatly accelerated the convergence of the training process. We employed a learning rate of $1e^{-3}$ and utilized the Adam optimizer.

Regarding the DnCNN architecture, it's common to use 17 layers. However, for our deblending problem, we found that reducing the number of layers to 10 still produced excellent deblending results while significantly reducing computational costs.

Finite difference deblending example

We initiate our exploration by applying the Noiser2Noiser network to a synthetic blended dataset, meticulously crafted using the finite difference method. This dataset encompasses 350 shots and receivers sampled at a rate of 4 ms. It's worth noting that we set the blending factor to three, signifying that each shot gather bears the influence of source interference from two other shots.

In Figure 9, panels (a) and (b) offer a visual glimpse into both the pristine and noisy common receiver gathers. To train our network with utmost efficacy, we craft corresponding re-corrupted common receiver gathers, as illustrated in panels (c) and (d). Subsequently, panels (e) and (f) unveil the deblended common receiver gather in tandem with the error relative to its pristine counterpart.

Upon iterating this denoising process across all common receiver gathers, we present the conclusive deblended outcomes for the common shot gather in Figure 10.

Mannully blended real marine data example

To comprehensively evaluate the efficacy of our robust Noiser2Noiser network in tackling real-world challenges, we conducted further testing using real data. Our initial experiment with real data involved a 2-D marine dataset in which we intentionally introduced manual blending to simulate complex acquisition scenarios. This specific dataset, a 2-D prestack marine example, is publicly accessible for download from the SEG website. In this dataset, we manually fused two consecutive shots, giving rise to coherent source inter-

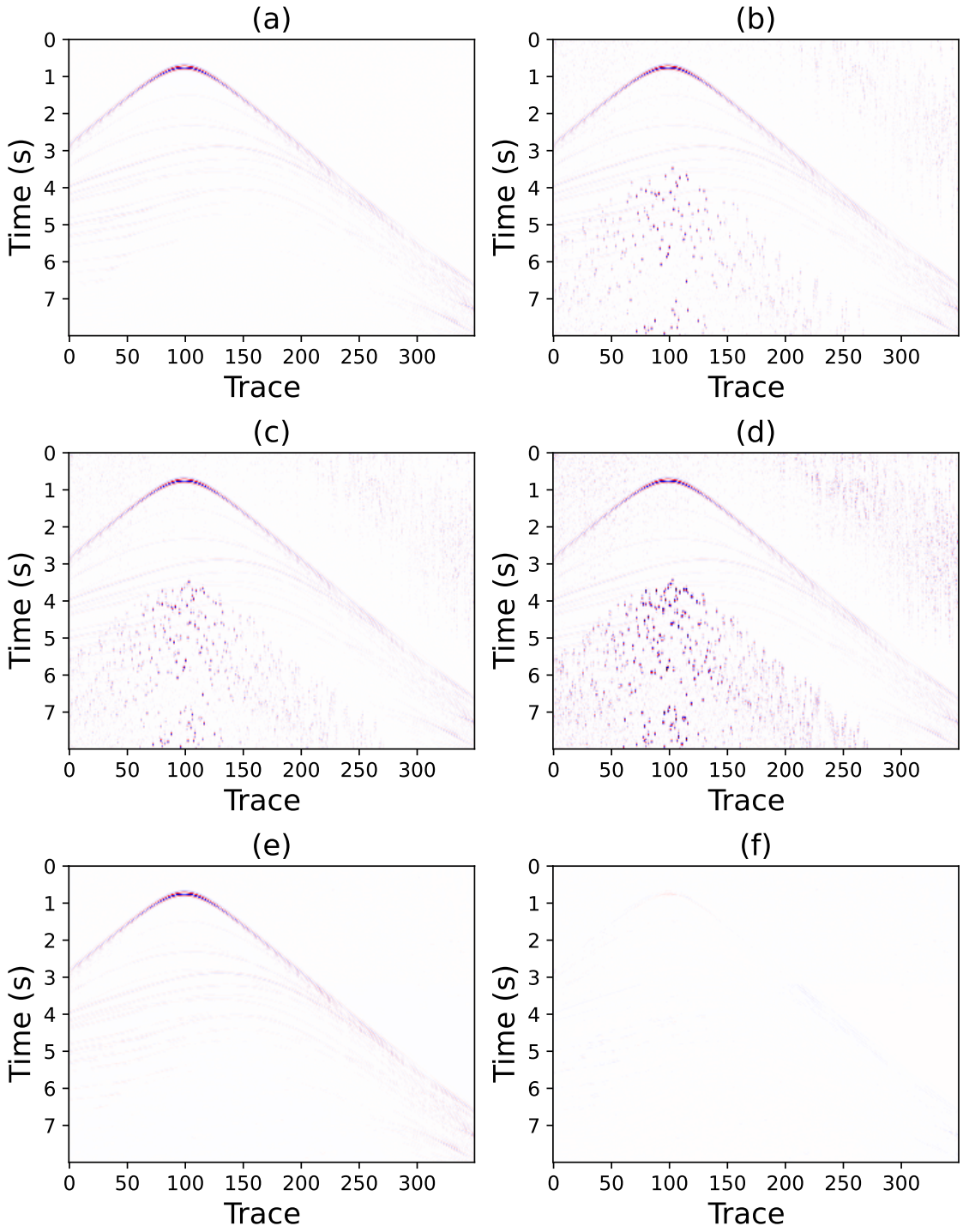

FIG. 9. Common receiver gather of finite difference blended data. (a) Clean data. (b) Noise data with blending noise (SNR=–0.84 dB). (c) Re-corrupted training input. (d) Re-corrupted training label. (e) Deblended result by the robust Noiser2Noiser Network (SNR=17.21 dB). (f) Errors between deblended result and clean data.

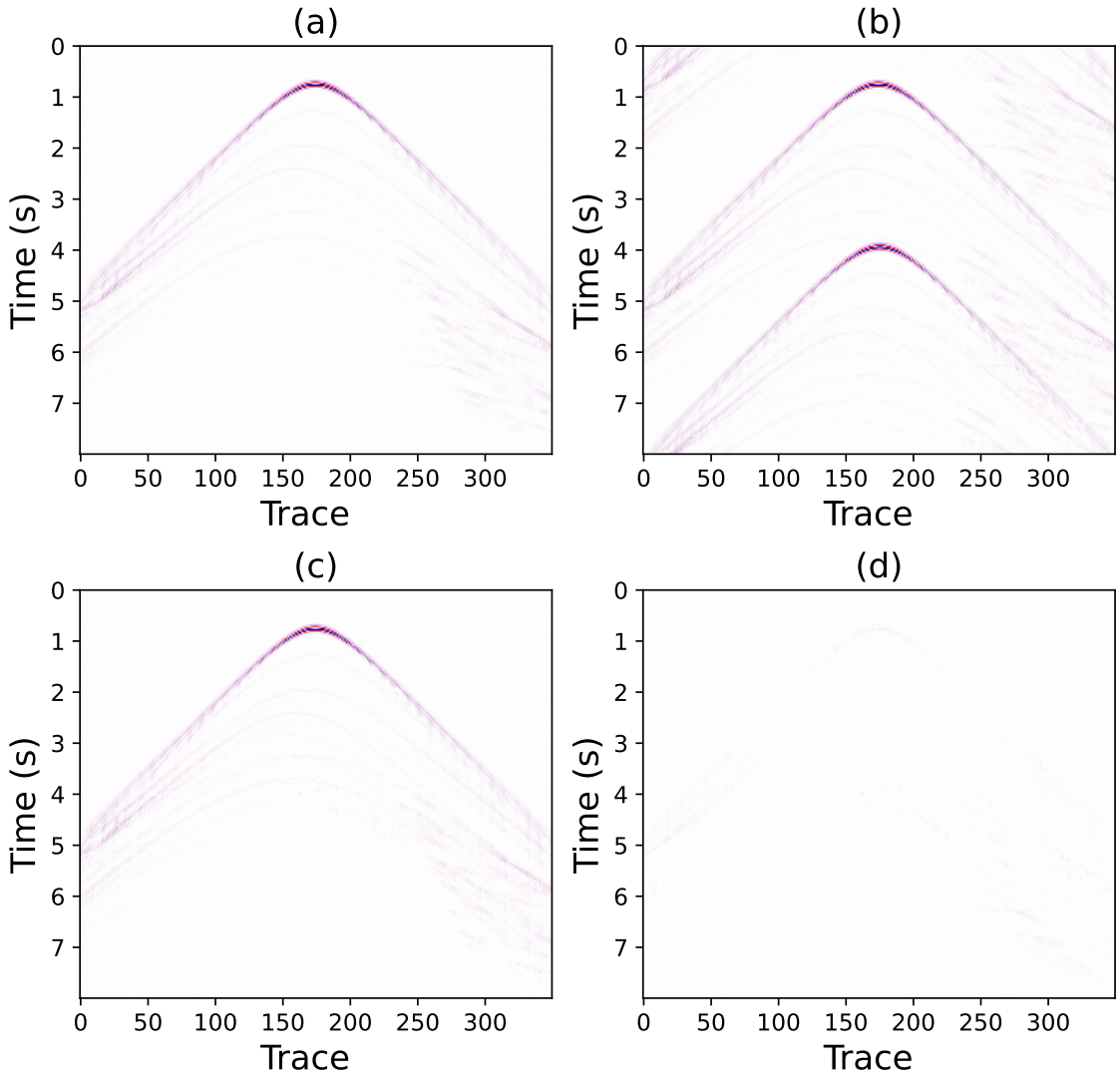

FIG. 10. Common shot gather of finite difference blended data. (a) Clean data. (b) Pseudo-deblended data (SNR=–0.20 dB). (c) Deblended result (SNR=17.65 dB). (d) Errors between deblended result and clean data.

ference that manifests in the common shot gather.

In Figure 11a, we present an unaltered common receiver gather for reference, while Figure 11b showcases the same shot gather subjected to controlled blending, resulting in a signal-to-noise ratio (SNR) of approximately -0.3 dB. In line with our previous experiments, we replicated the blending process using the pseudo-deblended dataset, introducing various time shifts. The resulting re-blended common receiver gather, shown in Figure 11c and 11d, serves as both the training input and the reference label for our model. The deblended outcome is visualized in Figure 11e.

We adopted a similar strategy and trained the model on the entire dataset to effectively mitigate all blending artifacts within the common offset gather. Figure 12a and 12b provide a visual comparison between the original common shot gather and the pseudo-deblended counterpart. Conversely, Figure 12c and 12d emphasize one of the final deblended common shot gathers, alongside the associated error analysis.

Real Simultaneous Source Marine dataset

In our final evaluation, we subjected our proposed method to rigorous testing using a real simultaneous source marine dataset, generously provided by PGS and obtainable from the SEG website. This dataset takes the form of a substantial 2768*256*256 matrix, encompassing temporal, receiver, and shot dimensions, sampled at a rate of 4 ms. Figure 13 provides a glimpse of one of the common receiver gathers residing within this extensive dataset. Following our established experimental protocol, we executed the re-blending and pseudo-deblending procedures on this dataset. The resulting training input and target pair corresponding to Figure 13 are depicted in Figure 14a and 14b, respectively. Subsequently, Figure 14c and 14d offer a visual representation of the deblended outcome. It's important to note that, due to the absence of ground truth for this dataset, we present both the deblended result and the difference relative to the original data. This presentation helps elucidate the effectiveness of noise attenuation and any potential signal leakage.

We systematically applied this process across the entire dataset, diligently removing blended noise from all common receiver gathers. The conclusive deblended result, achieved through our methodology, is showcased in Figure 15.

CONCLUSIONS

We introduce a novel zero-shot self-supervised framework designed to mitigate both random and erratic noise effectively. In line with the foundational concept of Noise2Noise, we aim to reduce the dependency on clean training data. Our method employs a pair of re-corrupted datasets during the training process, offering a detailed explanation of how we synthesize random and blended noise (a specific subtype of erratic noise) used for generating the training data.

To benchmark our approach, we compare it with two other Noise2Noise alternatives, conducting extensive evaluations on synthetic data. Subsequently, we apply our proposed framework to address the deblending challenge across synthetic and real marine data ex-

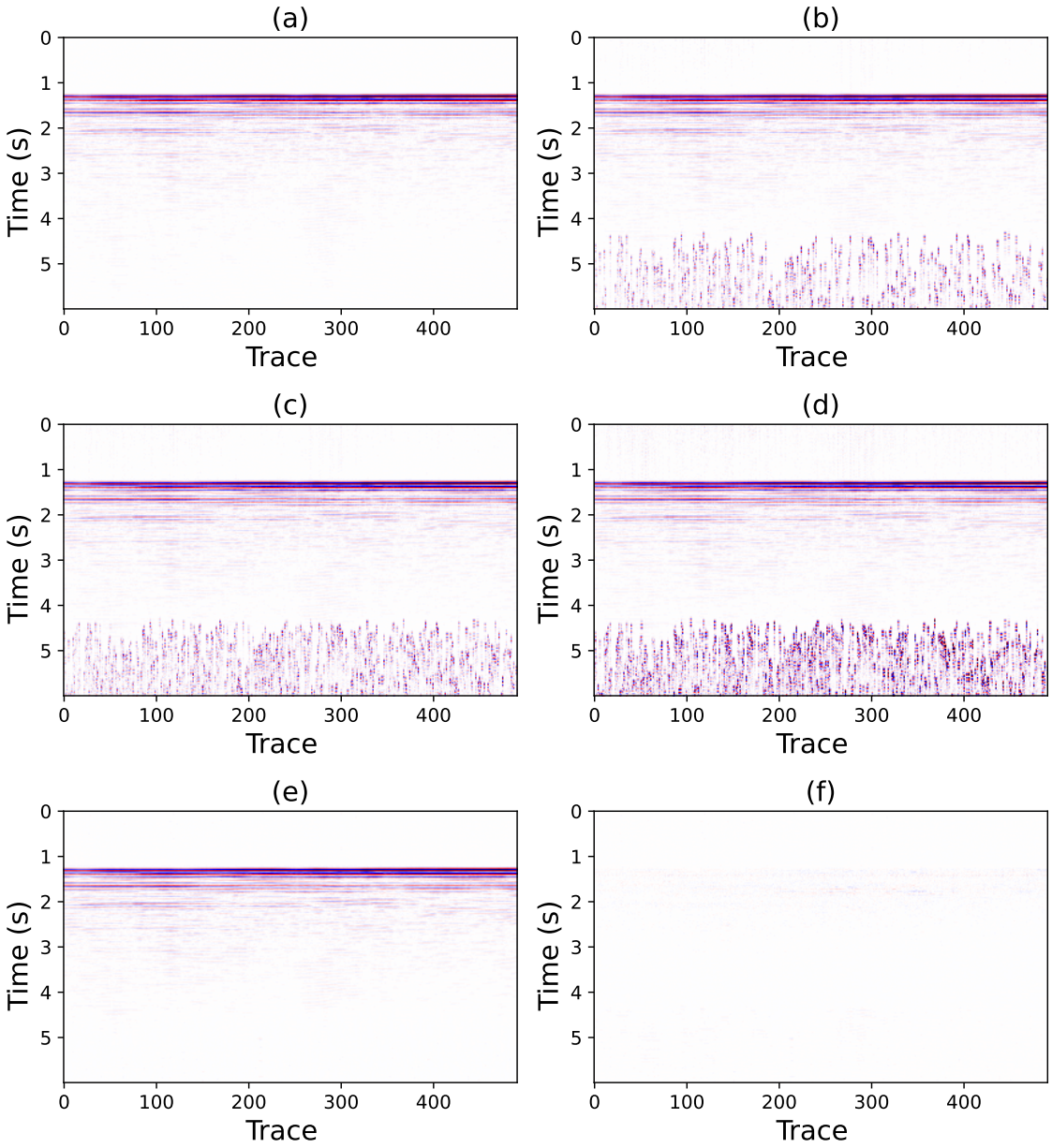

FIG. 11. Common receiver gather of manually blended real marine data. (a) Clean data. (b) Noise data with blending noise (SNR=–0.3 dB). (c) Re-corrupted training input. (d) Re-corrupted training label. (e) Deblended result by the robust Noiser2Noiser (SNR=17.3 dB). (f) Errors between deblended result and clean data.

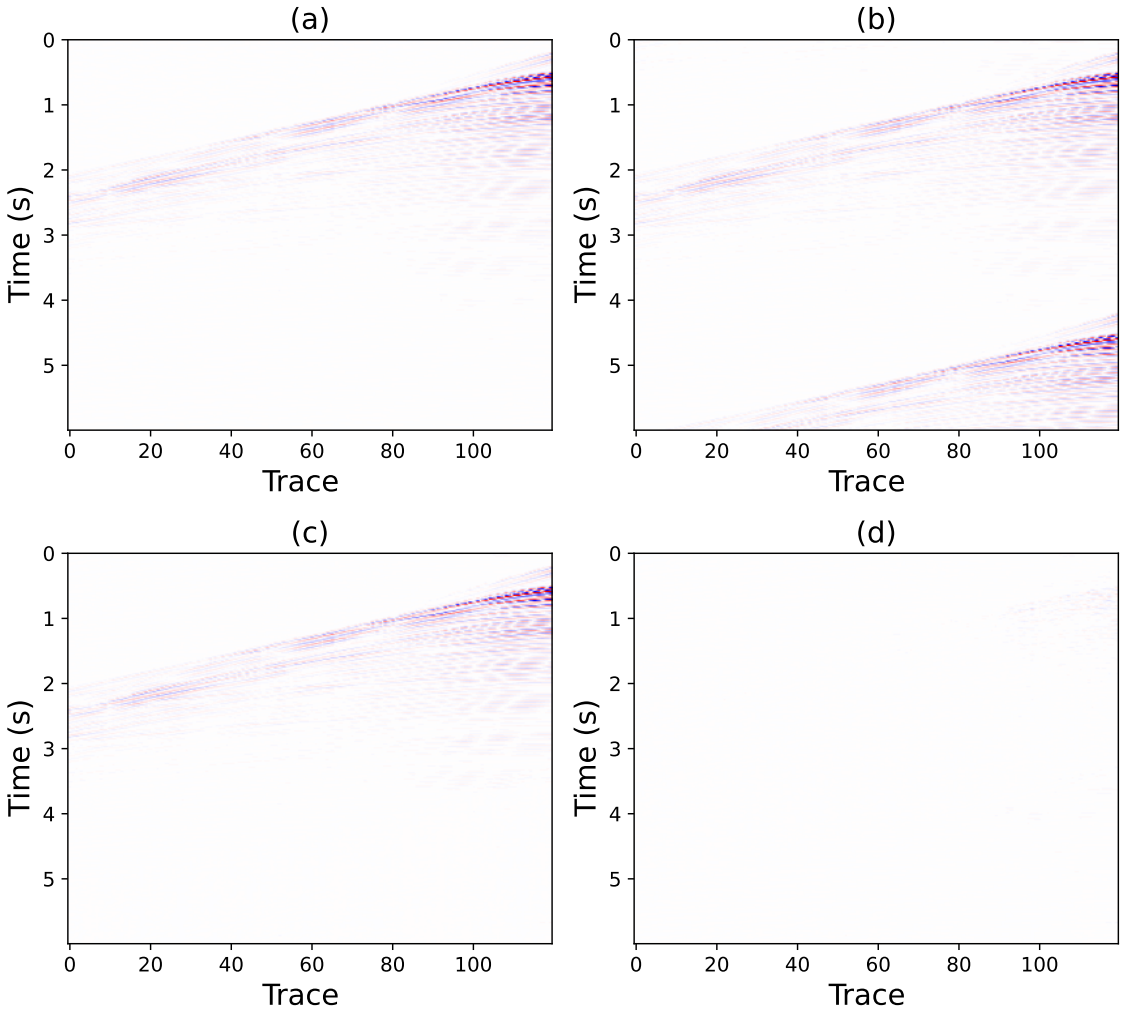

FIG. 12. Common shot gather of manually blended real marine data. (a) Clean data. (b) Pseudo-deblended data (SNR=–0.7 dB). (c) Deblended result (SNR=18.4 dB). (d) Errors between deblended result and clean data.

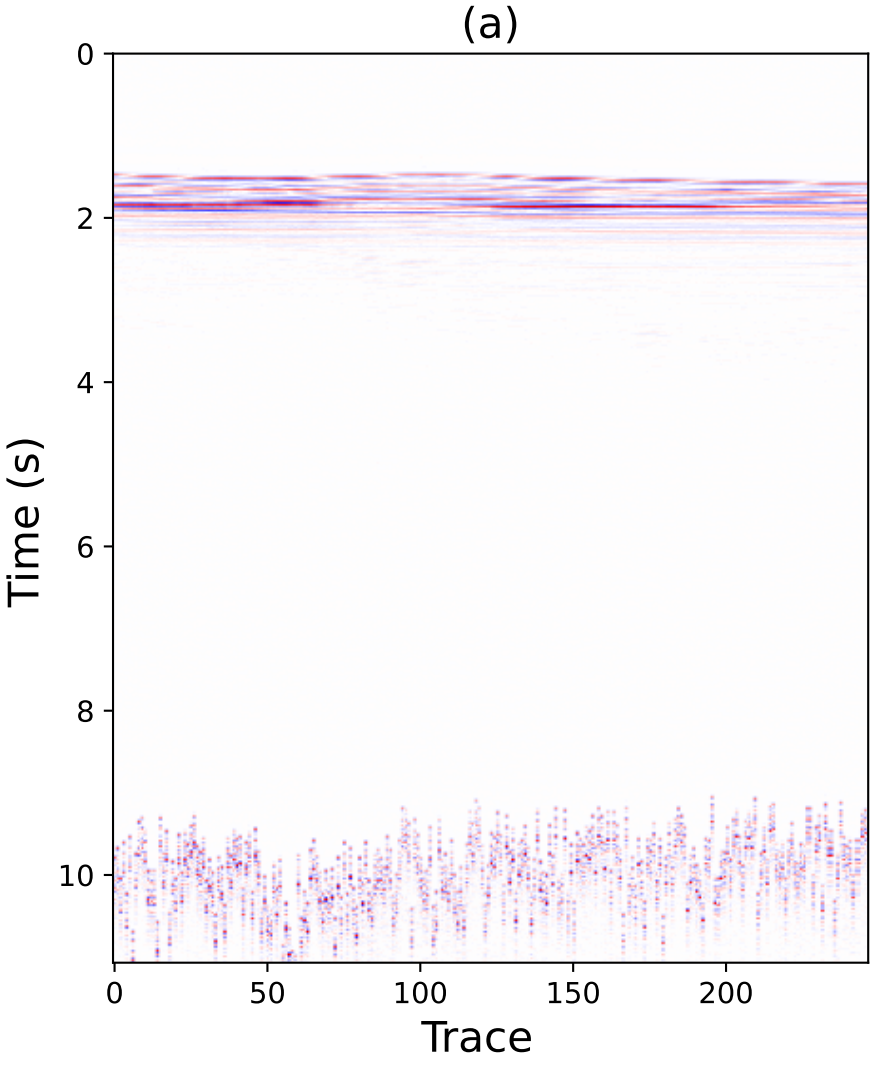

FIG. 13. One common receiver gather.

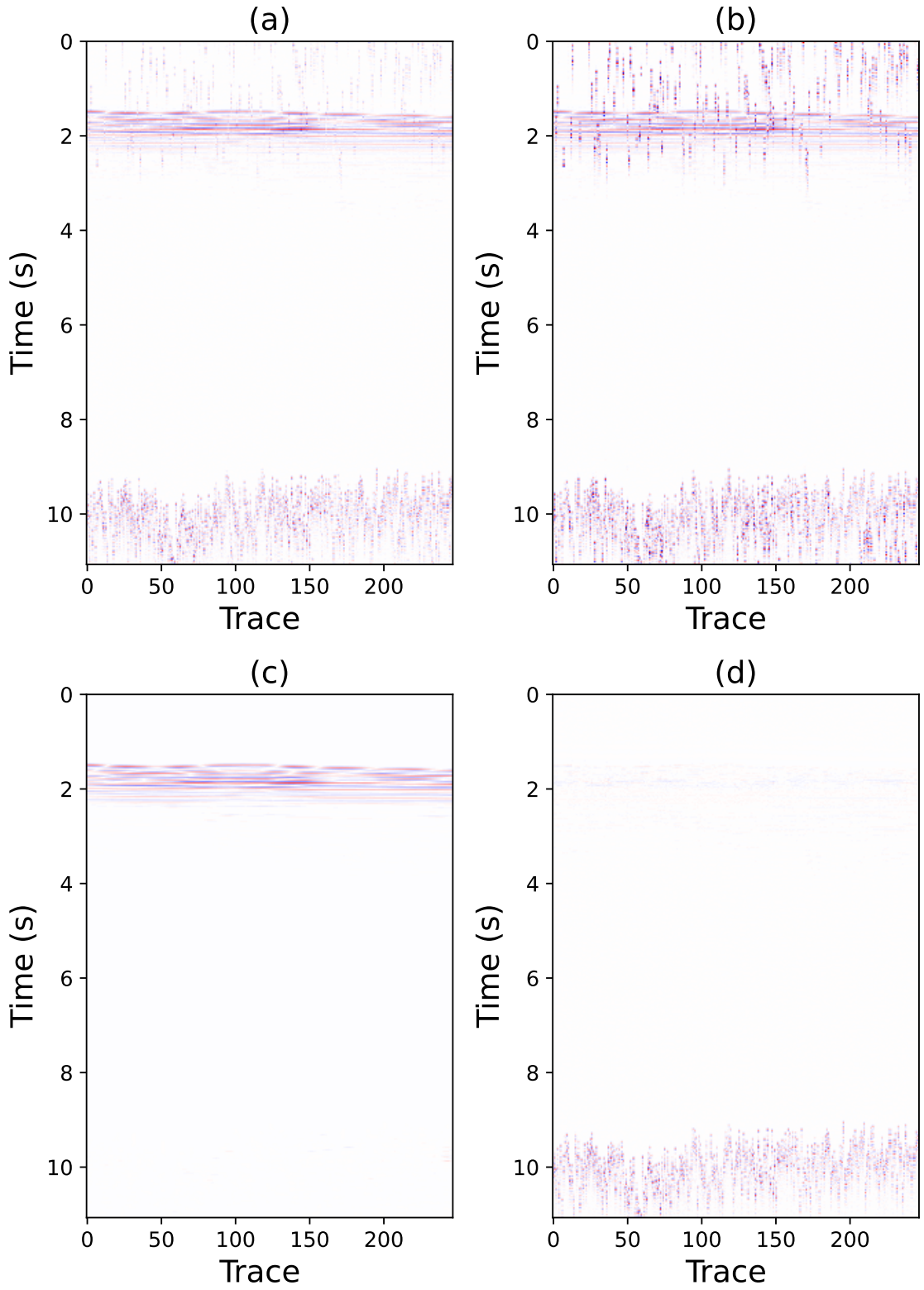

FIG. 14. (a) Generated training input. (b) Generated training input target. (c) Deblended result. (d) Difference between (c) and Figure 13.

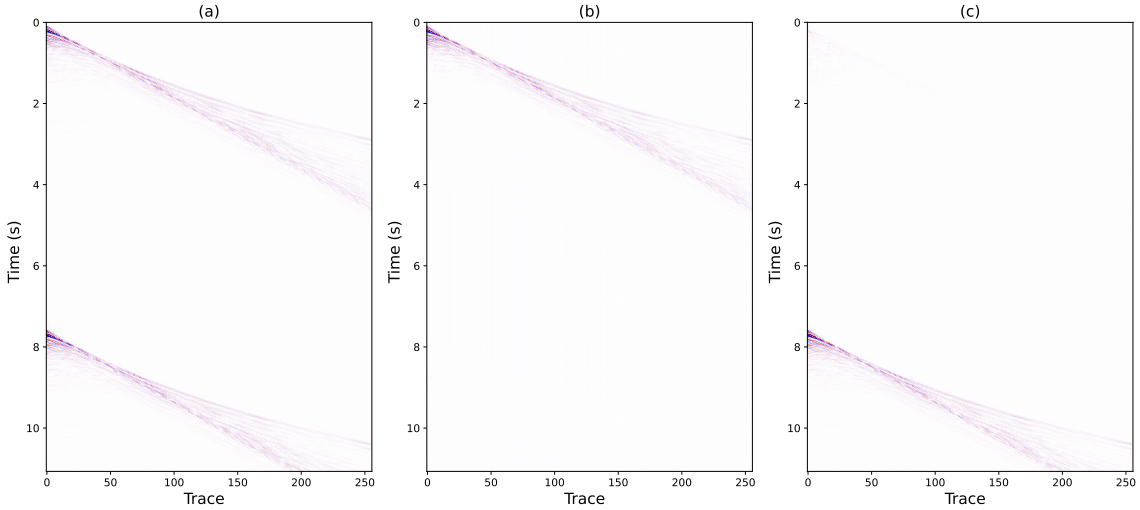

FIG. 15. Final deblended result of common shot gathers. (a) Original blended shot gather (b) Deblended shot gather (c) Difference between a and b.

amples. These practical instances serve as compelling demonstrations of the framework's remarkable deblending performance. It effectively eliminates blended noise without introducing significant signal leakage.

Notably, even without clean training data, our self-supervised method achieves noteworthy deblending results, surpassing conventional algorithms and supervised learning-based deblending methods. As we continue our research, we aim to extend the applicability of this framework to address various forms of erratic noise, presenting a promising avenue for future exploration.

ACKNOWLEDGMENTS

We thank the sponsors of CREWES for their continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19.

REFERENCES

Abma, R., and Claerbout, J., 1995, Lateral prediction for noise attenuation by tx and fx techniques: Geophysics, **60**, No. 6, 1887–1896.

Batson, J., and Royer, L., 2019, Noise2self: Blind denoising by self-supervision, *in* International Conference on Machine Learning, PMLR, 524–533.

Beasley, C. J., Chambers, R. E., and Jiang, Z., 1998, A new look at simultaneous sources, *in* Seg technical program expanded abstracts 1998, Society of Exploration Geophysicists, 133–135.

Berkhout, A. G., 2008, Changing the mindset in seismic data acquisition: The Leading Edge, **27**, No. 7, 924–938.

Cao, S., and Chen, X., 2005, The second-generation wavelet transform and its application in denoising of seismic data: Applied geophysics, **2**, 70–74.

Chen, X., and He, K., 2021, Exploring simple siamese representation learning, *in* Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 15,750–15,758.

Deighan, A. J., and Watts, D. R., 1997, Ground-roll suppression using the wavelet transform: Geophysics, **62**, No. 6, 1896–1903.

Górszczyk, A., Adamczyk, A., and Malinowski, M., 2014, Application of curvelet denoising to 2d and 3d seismic data—practical considerations: Journal of Applied Geophysics, **105**, 78–94.

Guitton, A., and Symes, W. W., 2003, Robust inversion of seismic data using the huber norm: Geophysics, **68**, No. 4, 1310–1319.

He, K., Zhang, X., Ren, S., and Sun, J., 2015, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, *in* Proceedings of the IEEE international conference on computer vision, 1026–1034.

Helbing, G., and Ritter, M., 2018, Deep learning for fault detection in wind turbines: Renewable and Sustainable Energy Reviews, **98**, 189–198.

Hennenfent, G., and Herrmann, F. J., 2006, Seismic denoising with nonuniformly sampled curvelets: Computing in Science & Engineering, **8**, No. 3, 16–25.

Hennenfent, G., and Herrmann, F. J., 2008, Simply denoise: Wavefield reconstruction via jittered undersampling: Geophysics, **73**, No. 3, V19–V28.

Huang, T., Li, S., Jia, X., Lu, H., and Liu, J., 2021, Neighbor2neighbor: Self-supervised denoising from single noisy images, *in* Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 14,781–14,790.

Ibrahim, A., and Sacchi, M. D., 2014, Simultaneous source separation using a robust radon transform: Geophysics, **79**, No. 1, V1–V11.

Kaur, H., Pham, N., and Fomel, S., 2021, Seismic data interpolation using deep learning with generative adversarial networks: Geophysical Prospecting, **69**, No. 2, 307–326.

Krull, A., Buchholz, T.-O., and Jug, F., 2019, Noise2void-learning denoising from single noisy images, *in* Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2129–2137.

Lari, H. H., Naghizadeh, M., Sacchi, M. D., and Gholami, A., 2019, Adaptive singular spectrum analysis for seismic denoising and interpolation: Geophysics, **84**, No. 2, V133–V142.

Latif, A., and Mousa, W. A., 2016, An efficient undersampled high-resolution radon transform for exploration seismic data processing: IEEE Transactions on Geoscience and Remote Sensing, **55**, No. 2, 1010–1024.

Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., and Aila, T., 2018, Noise2noise: Learning image restoration without clean data: arXiv preprint arXiv:1803.04189.

Lequyer, J., Philip, R., Sharma, A., Hsu, W.-H., and Pelletier, L., 2022, A fast blind zero-shot denoiser: Nature Machine Intelligence, **4**, No. 11, 953–963.

Li, H., Yang, W., and Yong, X., 2018, Deep learning for ground-roll noise attenuation, *in* SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 1981–1985.

Li, J., and Sacchi, M. D., 2021, An lp-space matching pursuit algorithm and its application to robust seismic data denoising via time-domain radon transforms: Geophysics, **86**, No. 3, V171–V183.

Li, S., Liu, B., Ren, Y., Chen, Y., Yang, S., Wang, Y., and Jiang, P., 2019, Deep-learning inversion of seismic data: arXiv preprint arXiv:1901.07733.

Liu, D., Deng, Z., Wang, C., Wang, X., and Chen, W., 2020, An unsupervised deep learning method for denoising prestack random noise: IEEE Geoscience and Remote Sensing Letters, **19**, 1–5.

Liu, D., Wang, W., Chen, W., Wang, X., Zhou, Y., and Shi, Z., 2018, Random noise suppression in seismic data: What can deep learning do?, *in* SEG International Exposition and Annual Meeting, SEG, SEG–2018.

Mansour, Y., and Heckel, R., 2023, Zero-shot noise2noise: Efficient image denoising without any data, *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14,018–14,027.

Maronna, R. A., 1976, Robust m-estimators of multivariate location and scatter: The annals of statistics, 51–67.

Matharu, G., Gao, W., Lin, R., Guo, Y., Park, M., and Sacchi, M. D., 2020, Simultaneous source deblending using a deep residual network, *in* SEG 2019 Workshop: Mathematical Geophysics: Traditional vs Learning, Beijing, China, 5-7 November 2019, Society of Exploration Geophysicists, 13–16.

Moran, N., Schmidt, D., Zhong, Y., and Coady, P., 2020, Noisier2noise: Learning to denoise from unpaired noisy data, *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12,064–12,072.

Mousavi, S. M., and Langston, C. A., 2016, Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding: Bulletin of the Seismological Society of America, **106**, No. 4, 1380–1393.

Oliveira, D. A., Ferreira, R. S., Silva, R., and Brazil, E. V., 2018, Interpolating seismic data with conditional generative adversarial networks: IEEE Geoscience and Remote Sensing Letters, **15**, No. 12, 1952–1956.

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

Pang, T., Zheng, H., Quan, Y., and Ji, H., 2021, Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising, *in* Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2043–2052.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. et al., 2019, Pytorch: An imperative style, high-performance deep learning library: Advances in neural information processing systems, **32**.

Qian, F., Guo, W., Liu, Z., Yu, H., Zhang, G., and Hu, G., 2022, Unsupervised erratic seismic noise attenuation with robust deep convolutional autoencoders: IEEE Transactions on Geoscience and Remote Sensing, **60**, 1–16.

Quan, Y., Chen, M., Pang, T., and Ji, H., 2020, Self2self with dropout: Learning self-supervised denoising from single image, *in* Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 1890–1898.

Richardson, A., and Feller, C., 2019, Seismic data denoising and deblending using deep learning: arXiv preprint arXiv:1907.01497.

Ronneberger, O., Fischer, P., and Brox, T., 2015, U-net: Convolutional networks for biomedical image segmentation, *in* Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 234–241.

Saad, O. M., Oboue, Y. A. S. I., Bai, M., Samy, L., Yang, L., and Chen, Y., 2021, Self-attention deep image prior network for unsupervised 3-d seismic data enhancement: IEEE Transactions on Geoscience and Remote Sensing, **60**, 1–14.

Siahsar, M. A. N., Gholtashi, S., Torshizi, E. O., Chen, W., and Chen, Y., 2017, Simultaneous denoising and interpolation of 3-d seismic data via damped data-driven optimal singular value shrinkage: IEEE Geoscience and Remote Sensing Letters, **14**, No. 7, 1086–1090.

Sun, H., Yang, F., and Ma, J., 2022, Seismic random noise attenuation via self-supervised transfer learning: IEEE geoscience and remote sensing letters, **19**, 1–5.

Trickett, S., Burroughs, L., and Milton, A., 2012, Robust rank-reduction filtering for erratic noise, *in* SEG International Exposition and Annual Meeting, SEG, SEG–2012.

Ulyanov, D., Vedaldi, A., and Lempitsky, V., 2018, Deep image prior, *in* Proceedings of the IEEE conference on computer vision and pattern recognition, 9446–9454.

Wang, B., Zhang, N., Lu, W., and Wang, J., 2019, Deep-learning-based seismic data interpolation: A preliminary result: Geophysics, **84**, No. 1, V11–V20.

Wang, K., and Hu, T., 2021, Deblending of seismic data based on neural network trained in the csg: IEEE Transactions on Geoscience and Remote Sensing, **60**, 1–12.

Wang, X., Fan, S., Zhao, C., Liu, D., and Chen, W., 2023, A self-supervised method using noise2noise strategy for denoising crp gathers: IEEE Geoscience and Remote Sensing Letters.

Xu, J., Huang, Y., Cheng, M.-M., Liu, L., Zhu, F., Xu, Z., and Shao, L., 2020, Noisy-as-clean: Learning self-supervised denoising from corrupted image: IEEE Transactions on Image Processing, **29**, 9316–9329.

Yu, S., Ma, J., Zhang, X., and Sacchi, M. D., 2015, Interpolation and denoising of high-dimensional seismic data by learning a tight frame: Geophysics, **80**, No. 5, V119–V132.

Zhang, F., Liu, D., Wang, X., Chen, W., and Wang, W., 2018, Random noise attenuation method for seismic data based on deep residual networks, *in* International Geophysical Conference, Beijing, China, 24-27 April 2018, Society of Exploration Geophysicists and Chinese Petroleum Society, 1774–1777.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L., 2017, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising: IEEE transactions on image processing, **26**, No. 7, 3142–3155.

Zheng, Y., Zhang, Q., Yusifov, A., and Shi, Y., 2019, Applications of supervised deep learning for seismic interpretation and inversion: The Leading Edge, **38**, No. 7, 526–533.