

# Robust seismic data denoising via self-supervised deep learning

Ji Li<sup>1</sup>, Daniel Trad<sup>1</sup>, and Dawei Liu<sup>2</sup>

## 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 data sets alongside corresponding clean ground truth, which is impractical for real-world seismic data sets. In contrast, unsupervised learning (UL) methods, which do not rely on ground truth during training, often fall short in performance when compared with their supervised or traditional denoising counterparts. Moreover, current unsupervised DL methods fail to address the specific challenges posed by erratic seismic noise adequately. 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

(N2N) and data augmentation principles, we develop a robust self-supervised denoising network called robust Noiser2Noiser. Our approach eliminates the need for paired noisy and clean data sets as required by supervised methods or paired noisy data sets as in N2N. Instead, our framework relies solely on the original noisy seismic data set. Our methodology generates two independent recorrupted data sets from the original noisy data set, using one as the input and the other as the training target. Subsequently, we use a DL-based denoiser, denoising convolutional neural network, for training purposes. To address various types of random and erratic noise, the original noisy data set is recorrupted with the same noise type. Detailed explanations for generating training input and target data for blended data are provided. We apply our network to synthetic and real marine data examples, demonstrating significantly improved noise attenuation performance compared with traditional denoising methods and state-of-the-art UL Codes are available on [https://github.com/Ji-seismic/N2N\\_deblending](https://github.com/Ji-seismic/N2N_deblending).

## INTRODUCTION

Seismic data denoising is important in seismic data processing. The seismic noise encompasses coherent and incoherent components, as elucidated by Abma and Claerbout (1995). Ground roll, as discussed by Deighan and Watts (1997), represents the prevalent form of coherent noise in land seismic data. In contrast, incoherent noise is divided into two distinct categories: broadband frequency, low-amplitude random noise and high-amplitude erratic noise. In previous decades, various methods 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 trans-

form (Hennenfent and Herrmann, 2008; Yu et al., 2015), wavelet transform (Cao and Chen, 2005; Mousavi and Langston, 2016), and curvelet transform (Hennenfent and Herrmann, 2006; Górszczyk et al., 2014). An alternative set of methods hinges on rank reduction, positing that seismic data exhibit low-rank characteristics in the frequency domain. These approaches, such as multichannel signal spectrum analysis (Oropeza and Sacchi, 2011; Siahzar et al., 2017) and singular value decomposition (Lari et al., 2019), seek to restore signals by reconstructing low-rank matrices.

However, erratic noise attenuation poses a more formidable challenge compared with random noise mitigation. Robust methodologies have sought to replace the  $\ell_2$ -norm cost function found in nonrobust methods with robust  $M$  estimators (Maronna, 1976). Notably, Guittot and Symes (2003) substitute the conventional  $\ell_2$  norm with the Huber norm, enhancing seismic data handling

Manuscript received by the Editor 18 December 2023; revised manuscript received 10 May 2024; published ahead of production 23 May 2024.

<sup>1</sup>University of Calgary, Department of Earth, Energy and Environment, Calgary, Alberta, Canada. E-mail: li.ji1@ucalgary.ca; daniel.trad@ucalgary.ca.

<sup>2</sup>University of Alberta, Department of Physics, Edmonton, Alberta, Canada. E-mail: 409791715@qq.com (corresponding author).

© 2024 Society of Exploration Geophysicists. All rights reserved.

in the presence of outliers. Similar efforts by Trickett et al. (2012) incorporate rank reduction filters to mitigate erratic noise. Li and Sacchi (2021) introduce 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 noise attenuation (Liu et al., 2018), deblending (Richardson and Feller, 2019; Wang and Hu, 2021), fault detection (Helbing and Ritter, 2018), and seismic interpolation (Oliveira et al., 2018; Wang et al., 2019; Kaur et al., 2021). DL techniques broadly fall into two categories: supervised learning (SL) and unsupervised learning (UL). SL requires vast quantities of paired clean and noisy data sets for training, a challenge given the scarcity of such paired data in the real seismic domain. Although methods exist that only require synthetic data for training (Othman et al., 2021), their performance across various types of noise in real data remains unclear. Traditional methods have been used 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) harness the generator convolutional neural network (CNN) to combat random noise in prestack seismic data and ground roll (Liu et al., 2023). Sun et al. (2022a) use self-supervised transfer learning to mitigate random noise in seismic data. Qian et al. (2022) introduce a deep convolutional autoencoder with the Welsch function to attenuate random and erratic noise. Saad et al. (2021) propose a UL approach based on deep image prior (DIP) (Ulyanov et al., 2018) to eliminate random noise in 3D seismic data. In contrast, Sun et al. (2022b) adopt DL methods to attenuate blended noise in the shot domain. Wang et al. (2023a) use unsupervised double-deep neural networks to iteratively remove blended noise. In addition, Wang et al. (2022) use a self-supervised blind-trace network to address the deblending problem. Self-SL, a vital branch of UL, generates labels directly from the training data set, with Noise2Noise (N2N) (Lehtinen et al., 2018) being a prominent strategy in this domain. Wang et al. (2023b) propose a self-supervised learning method based on the N2N strategy to reduce random noise. However, for self-supervised methods using the N2N principle, gathering pairs of training data sets with the same data but differing independent noise poses a challenge comparable with collecting pairs of clean and noisy data. As a result, most UL 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 with their supervised counterparts.

In this paper, we introduce a pioneering self-supervised framework designed to address random and erratic noise, with a specific focus on mitigating blended noise in simultaneous source acquisition data. Drawing inspiration from N2N (Lehtinen et al., 2018) and data augmentation methods, we present a robust variant called robust Noiser2Noiser, which is capable of attenuating random and erratic noise. Our framework operates in a zero-shot self-supervised manner, eliminating the need for data other than the original noisy data set. This approach involves independently recorrupting the original noisy data to generate two independent recorrupted data sets, using one as the training input and the other as the training

label. Subsequently, we use a commonly used denoising CNN (DnCNN) (Zhang et al., 2017) and train it with the input and label pairs. The specific recorruption method depends on the type of noise being addressed; for random noise, we recorrupt the data with additional random noise to create input-label pairs, whereas for erratic noise, we apply the same type of erratic noise and use a robust  $\ell_1$ -norm loss function instead of the conventional mean-square error (MSE) loss function. To further improve the useful signal extraction ability from similar pairs, particularly in data plagued by erratic noise, we use a symmetric loss function (Chen and He, 2021), integrating it with the residual learning strategy proposed by He et al. (2016).

Our method is applied to synthetic and real data examples featuring random noise and erratic noise. Notably, we use the proposed technique to remove blended 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 data sets containing noisy and clean data for SL. However, a significant breakthrough in this paradigm emerged with the introduction of N2N by Lehtinen et al. (2018). This pioneering work demonstrated that denoising networks trained on noisy/noisy image pairs can generate results that are remarkably close to those trained on noisy/clean data pairs from the same data set.

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

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{x} + \mathbf{e}_1, \\ \mathbf{y}_2 &= \mathbf{x} + \mathbf{e}_2, \end{aligned} \quad (1)$$

where  $\mathbf{e}_1$  and  $\mathbf{e}_2$  represent independent noise sources. A network  $F_\phi$  is then trained to minimize the Noise2Clean MSE loss function:

$$\begin{aligned} &\min E\{\|F_\phi(\mathbf{y}_1) - \mathbf{x}\|_2^2\} \\ &= \min E\{\|F_\phi(\mathbf{y}_1)\|_2^2 - 2\mathbf{x}^T F_\phi(\mathbf{y}_1)\}, \end{aligned} \quad (2)$$

where  $E$  is an expectation operator used to calculate the average loss over all possible pairs of data. This function aims to reduce the difference between the network's output when applied to  $\mathbf{y}_1$  and the clean signal  $\mathbf{x}$ . In addition, N2N uses a similar approach but with a different loss function:

$$\begin{aligned} &\min E\{\|F_\phi(\mathbf{y}_1) - \mathbf{y}_2\|_2^2\} \\ &= \min E\{\|F_\phi(\mathbf{y}_1) - \mathbf{x} - \mathbf{e}_2\|_2^2\} \\ &= \min E\{\|F_\phi(\mathbf{y}_1)\|_2^2 - 2\mathbf{x}^T F_\phi(\mathbf{y}_1) - 2\mathbf{e}_2^T F_\phi(\mathbf{y}_1)\}. \end{aligned} \quad (3)$$

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 zero. Consequently, the MSE loss functions for Noise2Clean and N2N yield equivalent results. This observation suggests that training a denoising network on noisy data samples

with the same clean signals can yield outcomes similar to training on clean/noisy pairs.

Nevertheless, it is essential to acknowledge that acquiring noisy data set pairs that precisely align with the clean data is often as challenging as obtaining corresponding noisy/clean pairs. As a result, recent research has focused on training DL models using noisy data sets exclusively, thus obviating the need for pairwise correspondence between noisy and noisy data samples. This transition marks a significant step forward in developing unsupervised and self-SL methods for noise removal in various applications, such as seismic data processing.

For noise removal without the need for paired noisy and clean data, several innovative self-supervised techniques have emerged. These approaches redefine the conventional SL paradigm and have shown promising results in various applications, such as image denoising. Here, we discuss 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 use 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) Recorruption-based frameworks: Another class of self-supervised techniques involves recorrupting noisy images to create even noisier versions. These augmented data sets 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](#)), Noisy-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) Subsampling: Recent developments such as Neighbor2-Neighbor ([Huang et al., 2021](#)) showcase innovative ways to leverage a single noisy data set. By subsampling the data set, they generate a pair of noisy data sets for training, exploiting data augmentation to enhance denoising performance. In addition, [Lequyer et al. \(2022\)](#) and [Mansour and Heckel \(2023\)](#) harness similar ideas, optimizing the training process for greater efficiency.
- 5) DIP: DIP ([Ulyanov et al., 2018](#)) is another influential self-supervised technique that capitalizes on the ability of CNNs to fit natural images more rapidly compared with noise. By using 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 data sets and offering versatile solutions for various applications, such as 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 us consider  $\mathbf{x}$  as our representation of clean data. The noisy data  $\mathbf{y}$  can be expressed as

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad (4)$$

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

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

Methods such as Noiser2Noise ([Moran et al., 2020](#)) and Noisy-as-Clean ([Xu et al., 2020](#)) use a noisier image as input, where they synthesize noise  $\mathbf{z}$  and then train the denoising network on the data set pair  $(\mathbf{y} + \alpha\mathbf{z}, \mathbf{y})$ . Meanwhile, [Pang et al. \(2021\)](#) train 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 have tested both augmentation methods and found the results are very close. For our robust Noiser2Noiser seismic denoising network, we adopt the same method as described by [Pang et al. \(2021\)](#). In all our tests, we use  $\alpha = 0.5$ , which means the training pair we use is  $(\mathbf{y} + 0.5 \times \mathbf{z}, \mathbf{y} - 2 \times \mathbf{z})$ . In addition, we use a symmetric loss function ([Chen and He, 2021](#)) to train a Siamese network. The Siamese network is especially useful when labeled training data are scarce or expensive, as they can learn to compare inputs directly without relying on explicit class labels.

The loss function is then defined as

$$\min \frac{1}{2} E\{\|F_\phi(\mathbf{y}_1) - \mathbf{y}_2\|_2^2\} + \frac{1}{2} E\{\|F_\phi(\mathbf{y}_2) - \mathbf{y}_1\|_2^2\}. \quad (5)$$

Our experiments have demonstrated that using 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

$$\min \frac{1}{2} E\{\|\mathbf{y}_1 - F_\phi(\mathbf{y}_1) - \mathbf{y}_2\|_2^2\} + \frac{1}{2} E\{\|\mathbf{y}_2 - F_\phi(\mathbf{y}_2) - \mathbf{y}_1\|_2^2\}. \quad (6)$$

In the case of a data set with erratic noise, we replace the  $\ell_2$  norm with the  $\ell_1$  norm to make the loss function more robust:

$$\min \frac{1}{2} E\{\|\mathbf{y}_1 - F_\phi(\mathbf{y}_1) - \mathbf{y}_2\|_1\} + \frac{1}{2} E\{\|\mathbf{y}_2 - F_\phi(\mathbf{y}_2) - \mathbf{y}_1\|_1\}. \quad (7)$$

Now that we have our training input, target pairs, and corresponding loss functions, the next step is to select the network architecture. The proposed robust self-supervised denoising framework is compatible with various network architectures, such as ResNet ([He et al., 2016](#)), DnCNN ([Zhang et al., 2017](#)), and U-net ([Ronneberger et al., 2015](#)). We opt for DnCNN in this study due

to its straightforward architecture and effective residual learning strategy. It has successfully addressed numerous seismic denoising challenges in recent years, such as 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 shown in Figure 1. We generate the training input and target pairs from the original

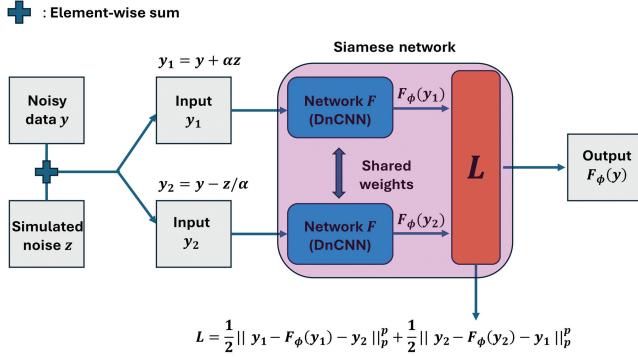


Figure 1. The architecture of the proposed robust self-supervised learning network. The training pair  $y_1 = \mathbf{y} + \alpha \mathbf{z}$  and  $y_2 = \mathbf{y} - \mathbf{z}/\alpha$  is used to train a Siamese network. After training, the original noisy data  $\mathbf{y}$  are used to estimate the clean data  $F_\phi(\mathbf{y})$ .

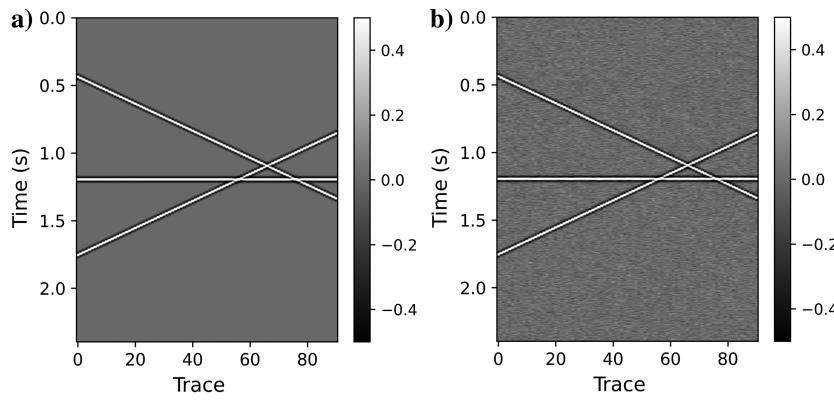


Figure 2. The 2D synthetic example with random noise: (a) clean data and (b) noisy data with random noise.

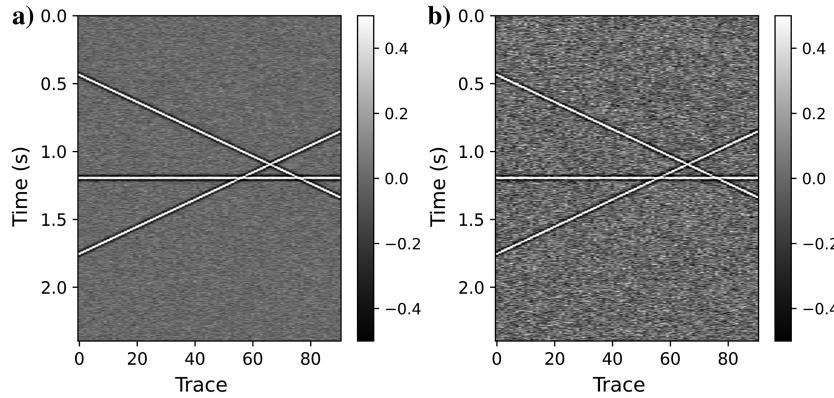


Figure 3. Recorrupted data pair with more independent random noise: (a) input and (b) label.

noisy data. Each training pair consists of  $\mathbf{y}_1 = \mathbf{y} + \alpha \mathbf{z}$  and  $\mathbf{y}_2 = \mathbf{y} - \mathbf{z}/\alpha$ . Here,  $\mathbf{z}$  represents the same type of noise as in the data set but is independent of the noise present in the original data. Our approach uses a Siamese network to train paired recorrupted noisy data. This network architecture is comprised of two identical DnCNN networks with shared parameters and weights. The outputs from these subnetworks are then leveraged to compute the final loss function. We use the  $\ell_2$  norm to attenuate random noise and the  $\ell_1$  norm for erratic noise. The training input and target should be normalized to ensure that the training process is more stable.

## Noise simulation

When dealing with random noise, generating an independent set of synthetic random noise is a straightforward process. We use additive white Gaussian noise, denoted as  $\mathbf{n}$ , with a distribution  $\mathbf{n} \sim N(0, \sigma^2 \mathbf{I})$ . 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 us 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; Berkhouit, 2008). Blended acquisition can be viewed as a time-shifting operation applied to data from individual sources, mathematically represented as

$$\mathbf{b} = \mathbf{B}\mathbf{D}, \quad (8)$$

where  $\mathbf{B}$  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 pseudodeblending operation:

$$\tilde{\mathbf{D}} = \mathbf{B}^* \mathbf{b}. \quad (9)$$

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

Suppose we have a pseudodeblended data set  $\tilde{\mathbf{D}}_1$ . 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 pseudodeblended data set  $\tilde{\mathbf{D}}_2$  that 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 synthetic and real marine data

examples. The network implementation is performed using PyTorch (Paszke et al., 2019), and all tests are conducted on a consumer-grade graphics processing unit with 4352 CUDA cores and 8 GB RAM.

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) Blended data created using the finite-difference method.
- 2) Real-world blended marine data collected through the simultaneous source acquisition method.
- 3) Real marine data that we numerically blended with much shorter time intervals and more blended noise.

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

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

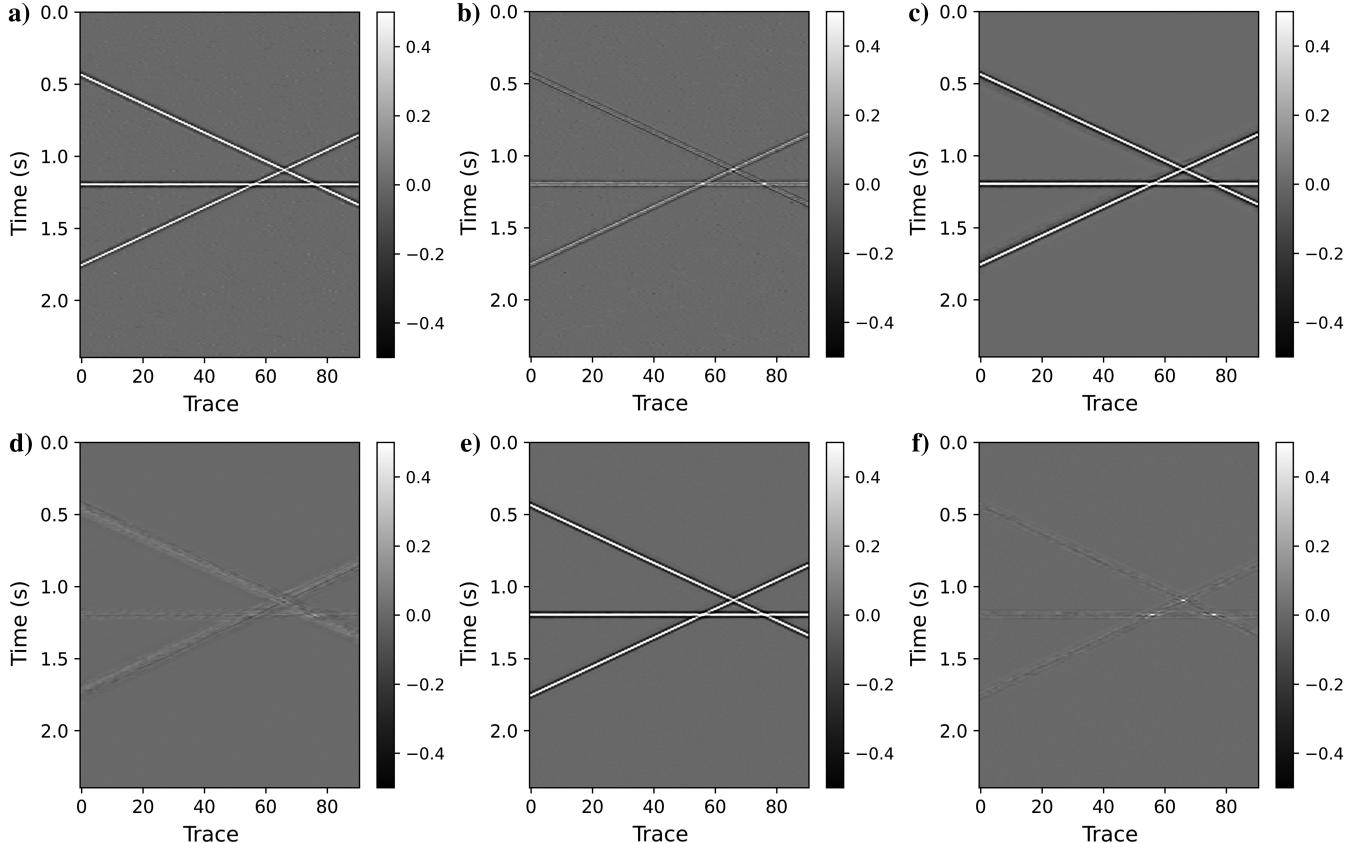


Figure 4. Denoising result of synthetic data with random noise. Noise2Void and robust Noiser2Noiser work for random noise. (a) Denoised via Neighbor2Neighbor ( $S/N = 10.2$  dB), (b) errors between (a) and clean data, (c) denoised via Noise2Void ( $S/N = 18.3$  dB), (d) errors between (c) and clean data, (e) denoised via robust Noiser2Noiser network ( $S/N = 18.1$  dB), and (f) errors between (e) and clean data.

$$S/N = 10 \log \frac{\|\mathbf{d}_c\|_2^2}{\|\mathbf{d}_c - \mathbf{d}_r\|_2^2}, \quad (10)$$

where  $\mathbf{d}_c$  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 S/N computation, we visualize the errors between the original data and the deblended result to illustrate any signal leakage.

## Synthetic examples denoising

### Simple 2D example

We explore the proposed framework by conducting tests on simple 2D synthetic data. These tests encompass scenarios involving 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 Neighbor2Neighbor (Huang et al., 2021).

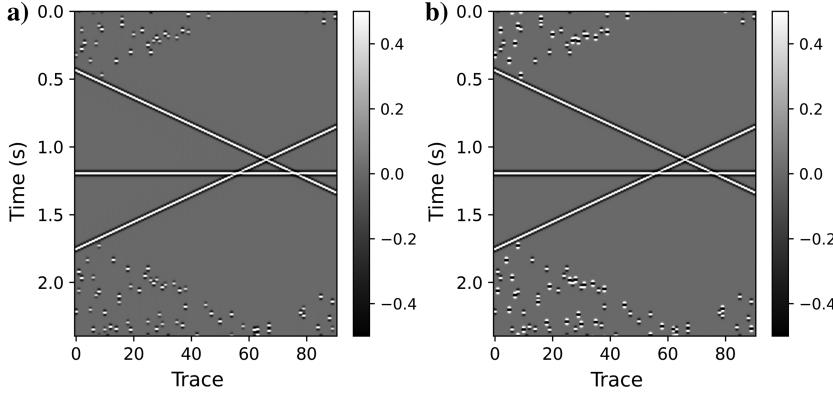


Figure 6. Recorrupted data pair with more independent blended noise: (a) input and (b) label.

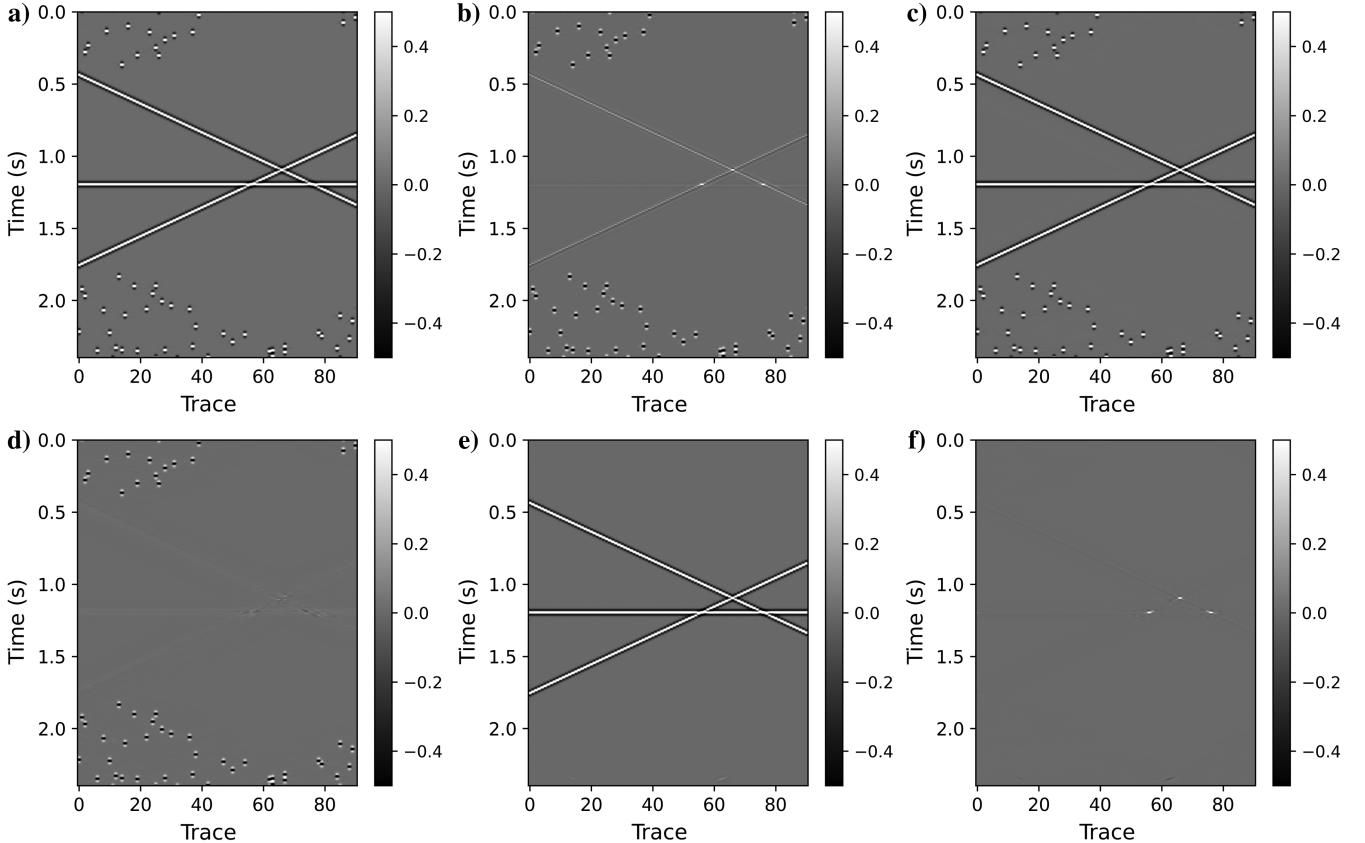


Figure 7. Denoising result of synthetic data with blended noise. Only the proposed robust Noiser2Noiser works for blended noise. (a) Denoised via Neighbor2Neighbor ( $S/N = 7.1$  dB), (b) errors between (a) and clean data, (c) denoised via Noise2Void ( $S/N = 7.4$  dB), (d) errors between (c) and clean data, (e) denoised via robust Noiser2Noiser ( $S/N = 22.1$  dB), and (f) errors between (e) and clean data.

Our initial test focuses on random noise, for which we use the  $\ell_2$  norm during training. Figure 2a and 2b shows the clean and noisy data, respectively. This 2D data section contains three linear events and operates at a sampling rate of 4 ms. To mitigate the impact of random noise, we manually recorrupt the noisy data set by introducing additional random noise, thereby generating training input and labels. Figure 3a and 3b shows the recorrupted training input and labels, respectively. The denoising results of these three methods are shown in Figure 4. In the case of random noise, Neighbor2Neighbor 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, blended noise. Figure 5b shows the same 2D synthetic data example but with blended noise manually introduced. To effectively address blended noise, we recorrupt the noisy data set, this time incorporating blended noise. The blending is performed with varying random time shifts. Figure 6 shows the generated training input and labels. To account for blended noise, we opt for the  $\ell_1$ -norm loss function, known for its robustness to erratic noise, to replace the  $\ell_2$  norm. This choice holds for all three methods. The deblending

outcomes from the three methods are visually shown in Figure 7, with the results revealing that only the proposed robust Noiser2Noiser denoising network effectively removes the blended noise.

### 2D finite-difference synthetic example

The previous example underscores the effectiveness of our proposed robust Noiser2Noiser framework in handling random and erratic noise scenarios. Consequently, in the subsequent tests, we focus solely on evaluating the deblending performance of our method across different data sets. 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 2D synthetic seismic shot generated using the finite-difference method. This 2D synthetic shot gather exhibits greater complexity compared with the previous example. As shown in Figure 8a and 8b, we present the clean data and the shot gather with blended noise, respectively. The S/N 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 rebledened training input and labels shown in Figure 8c and 8d.

Subsequently, Figure 8e and 8f shows the deblending results alongside error comparisons with the original clean shot gather. Notably, our approach leads to a remarkable improvement in the S/N, from  $-1.21$  dB to an impressive  $21.2$  dB.

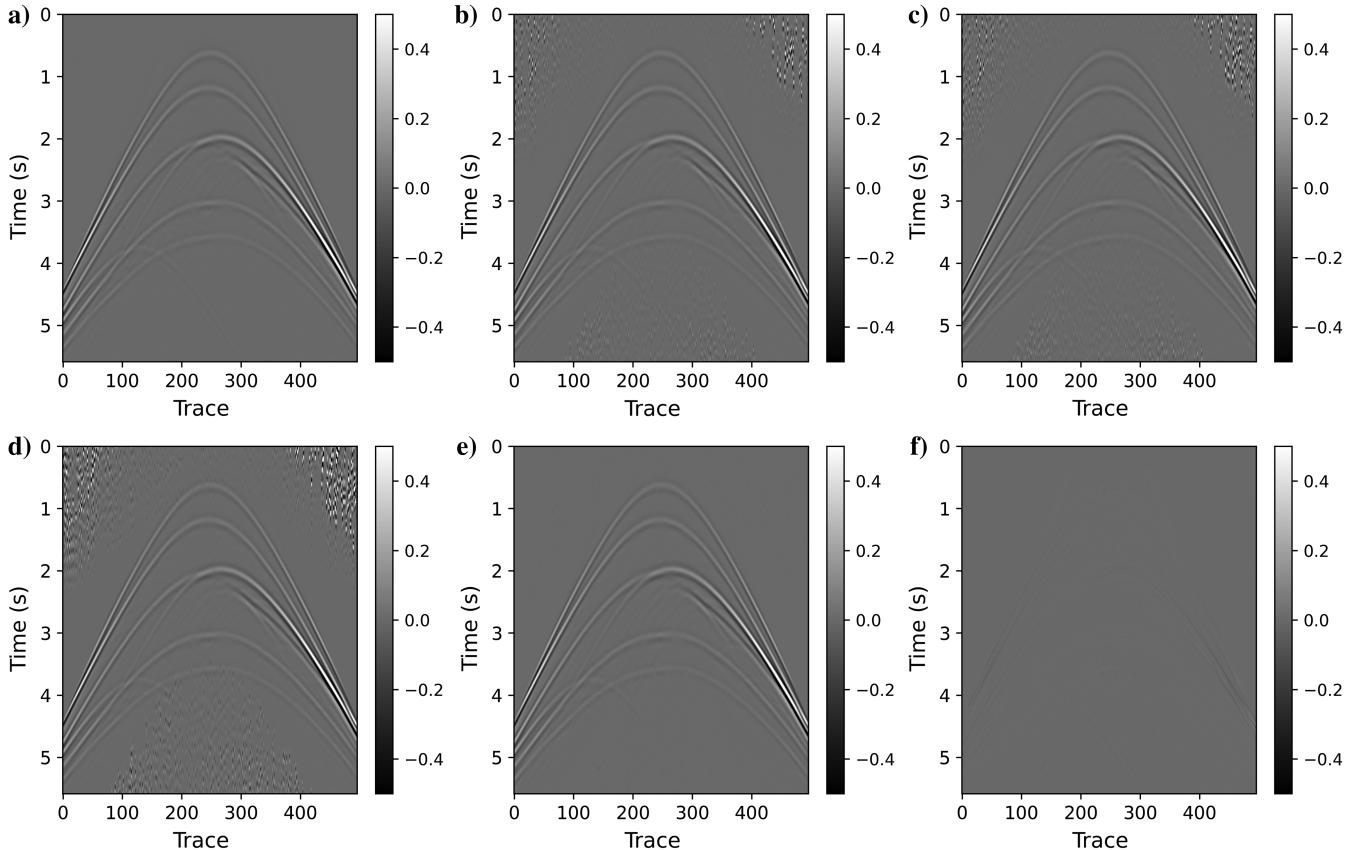


Figure 8. The 2D finite-difference synthetic example with blended noise. (a) Clean data, (b) noise data with blended noise ( $S/N = -1.21$  dB), (c) recorrupted training input, (d) recorrupted training label, (e) debledened result by the robust Noiser2Noiser network ( $S/N = 21.2$  dB), and (f) errors between debledened results and clean data.

Previously, we mentioned the effectiveness of using an  $\ell_1$ -norm loss function in conjunction with residual learning and a symmetrical term to enhance the denoising performance of our algorithm. In this subsection, we present the results of various tests conducted on this data set to assess the performance of our proposed method using different loss functions. Figure 9a shows the S/N values obtained with different data fitting methods. Notably, we observe that while the  $\ell_2$ -norm loss function is susceptible to erratic noise, the Huber norm exhibits higher resilience, albeit failing after extended training. Conversely, the  $\ell_1$  norm demonstrates significant robustness to erratic noise, achieving peak S/N and halting noise learning beyond a certain point. These trends are further evident in the loss curves shown in Figure 9b. The  $\ell_2$ -norm loss function continues to learn all data, including noise, until reaching zero loss, indicating overfitting. The Huber loss function's curve shows a slower decrease beyond a certain point, whereas the curve for the  $\ell_1$  norm stabilizes after learning all essential signals. Figure 9c shows the performance of our proposed method using the  $\ell_1$  norm with and without a symmetrical term as well as with and without residual learning. Notably, residual learning is crucial for handling blended noise, as evidenced by the failure of the network without it. In addition, incorporating a symmetrical term significantly enhances denoising performance.

### 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 data sets, which are typically 3D cubes containing time samples, receivers, and shots as dimensions, we observe that the blended noise in the common-shot domain is coherent. However, this translates into incoherent erratic noise in the common-receiver domain. Consequently, the deblending process can be conceptualized as removing erratic noise across all common-receiver gathers.

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

- 1) Training on whole data set (with patching): This method involves training on the entire data set, which comprises numerous shot gathers simultaneously. However, due to memory constraints, we must partition the data into smaller patches.

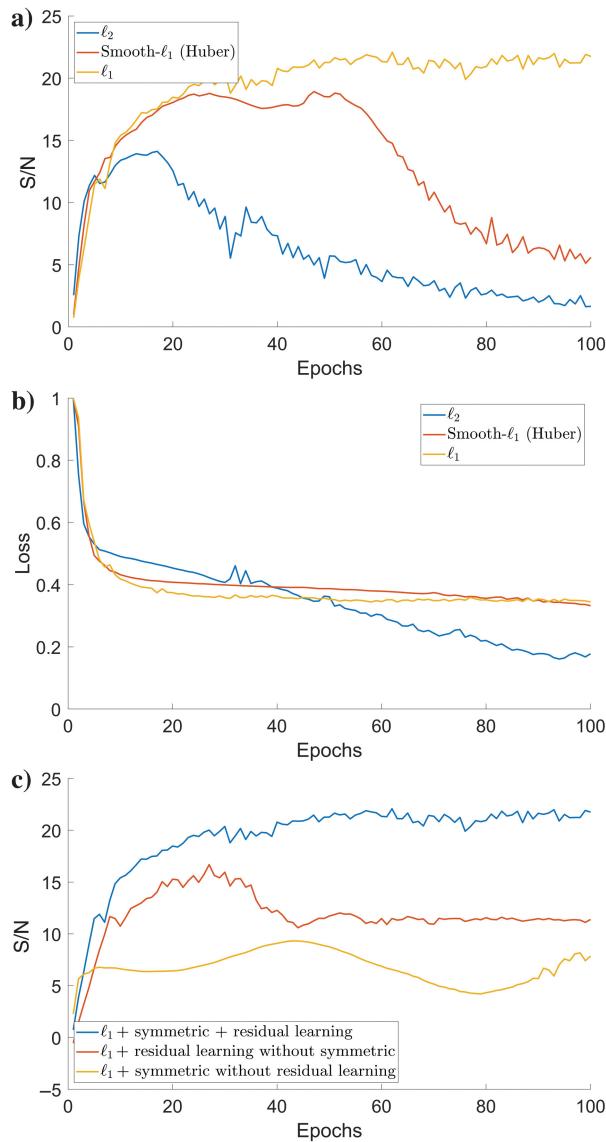


Figure 9. The proposed methods' behavior with different loss functions: (a) S/N curve for different data fit terms, (b) loss curve for different data fit terms, and (c) S/N curve for  $\ell_1$ -norm loss function with and without symmetric term.

- 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, we conducted experiments with a 3D cube of dimensions  $2000 \times 350 \times 350$ . When we divided the data set into  $32 \times 32$  patches with 50% overlap, training a single epoch on the entire data set took approximately 2 h. However, a significantly faster approach emerged when training on individual shot gathers, with each shot gather divided into  $32 \times 32$  patches. In this case, it required only approximately 1.2 s per epoch, totaling approximately 6 s for deblending one shot gather (equivalent to approximately 5 epochs).

When opting to train on the entire shot gather as a whole, without patching, it took approximately 0.4 s per epoch, with a total of 20 s (equivalent to 50 epochs) for deblending one shot gather. Although this method resulted in a longer training time compared with patch-based training, the improvement in deblending performance justified the investment. A total of 20 s 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 rectified linear unit activation functions. This initialization greatly accelerated the convergence of the training process. We used a learning rate of  $10^{-3}$  and used the Adam optimizer.

Regarding the DnCNN architecture, it is 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 first applying the Noiser2Noiser network to a synthetic blended data set meticulously crafted using the finite-difference method. This data set encompasses 350 shots and receivers sampled at a rate of 4 ms. We set the blending factor to three, signifying that each shot gather contains the source interference from two other shots.

Figure 10a and 10b shows the pristine and noisy common-receiver gathers. To train our network with utmost efficacy, we craft corresponding recorrupted common-receiver gathers, as shown in Figure 10c and 10d. Subsequently, Figure 10e and 10f unveils the deblended common-receiver gather in tandem with the error relative to its pristine counterpart.

We conducted statistical tests on the two noise sources to demonstrate that the newly generated blended noise is uncorrelated with the original noise. Figure 11a shows the blended noise extracted from the original gather (Figure 10b). In Figure 11b, we present the newly generated blended noise with different random time shifts, which was then used to create the training pairs, as shown in Figure 10c and 10d.

We calculated the correlation coefficient and covariance matrix between these two noise sources. The correlation coefficient is 0.0018 and both off-diagonal values of the covariance matrix are  $2.81 \times 10^{-5}$ . These values indicate that the newly generated blended

noise is statistically uncorrelated with the original noise, aligning well with the principles of the N2N theory.

We then iterate this denoising process across all common-receiver gathers and present the conclusive deblended outcomes for the common-shot gather in Figure 12.

#### Real simultaneous-source marine data set

To comprehensively evaluate the efficacy of our robust Noiser2Noiser network in tackling real-world challenges, we conducted further testing using real data. We test our proposed method rigorously using a real simultaneous source marine data set, generously provided by PGS and obtainable from the SEG website ([SEG Wiki, 2018](#)). This data set is a substantial  $2768 \times 256 \times 256$  matrix, encompassing temporal, receiver, and shot dimensions, sampled at a rate of 4 ms. Figure 13 shows one of the common-receiver gathers residing within this extensive data set. Following our established experimental protocol, we executed this data set's reblending and pseudodeblending procedures. The resulting training input and target pair corresponding to Figure 13 are shown in Figure 14a and 14b, respectively. Subsequently, Figure 14c and 14d visually rep-

resents the deblended outcome. It is important to note that, due to the absence of ground truth for this data set, we present the deblended result and the difference relative to the original data. This presentation helps demonstrate the effectiveness of noise attenuation and any potential signal leakage.

We systematically applied this process across the entire data set, diligently removing blended noise from all common-receiver

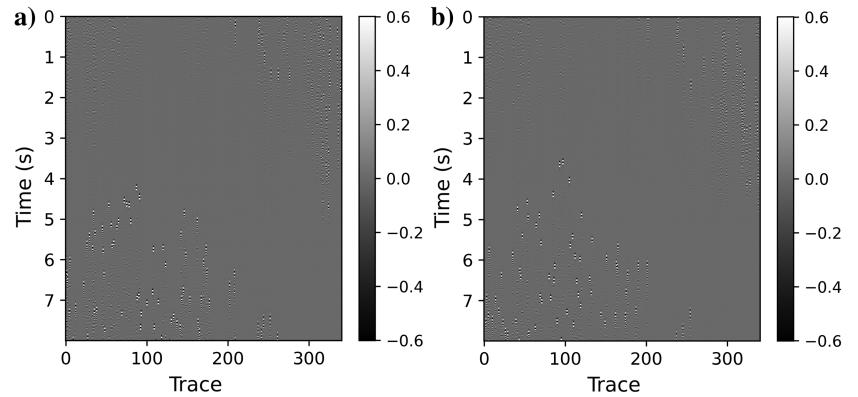


Figure 11. Blended noise in the common-shot gather. (a) Original blended noise in Figure 10b (the difference between Figure 10a and 10b) and (b) regenerated blended noise used to create Figure 10c and 10d.

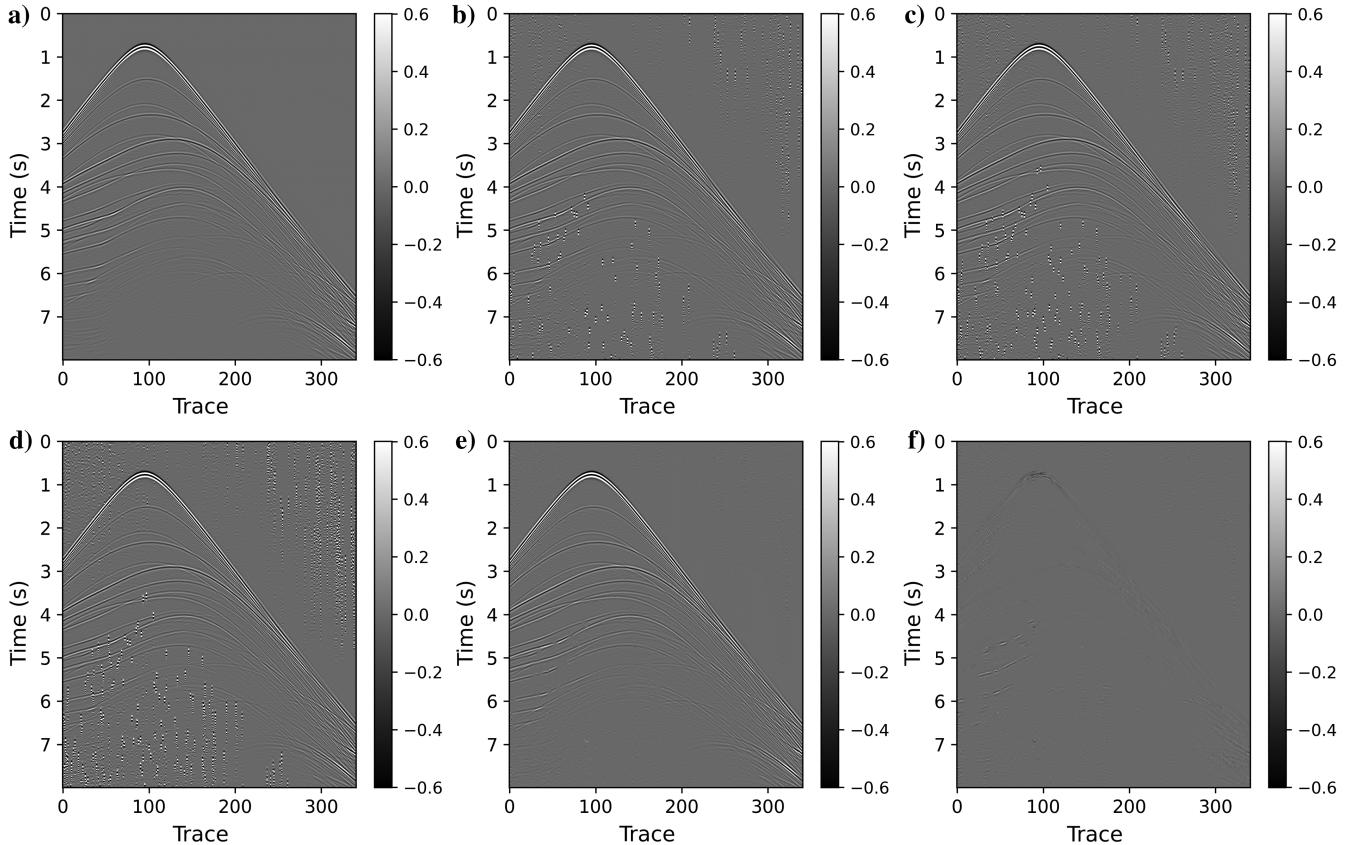


Figure 10. Common-receiver gather of finite-difference blended data. (a) Clean data, (b) noise data with blended noise ( $S/N = -0.84$  dB), (c) recorrupted training input, (d) recorrupted training label, (e) deblended result by the robust Noiser2Noiser network ( $S/N = 20.51$  dB), and (f) errors between deblended results and clean data.

gathers. The conclusive deblended result, achieved through our methodology, is shown in Figure 15.

#### Manually blended real marine data example

The blended noise in the previous real data example is far from the signal without much overlap between them. To further test the

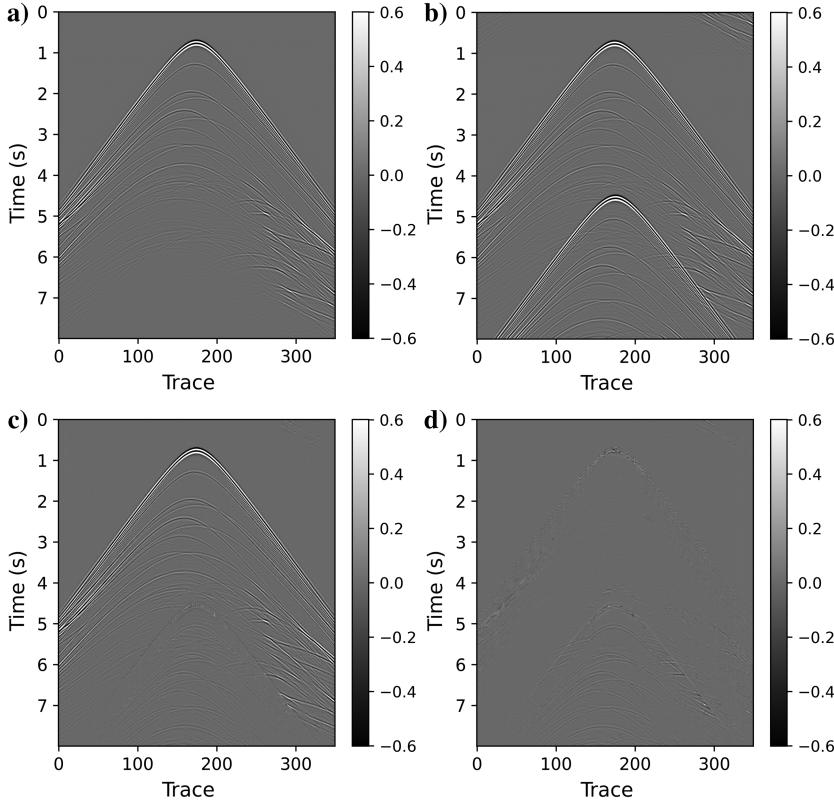


Figure 12. Common-shot gather of finite-difference blended data. All the source interference from the other shots is removed successfully. (a) Clean data, (b) pseudodeblended data ( $S/N = -0.20$  dB), (c) deblended result ( $S/N = 20.03$  dB), and (d) errors between deblended results and clean data.

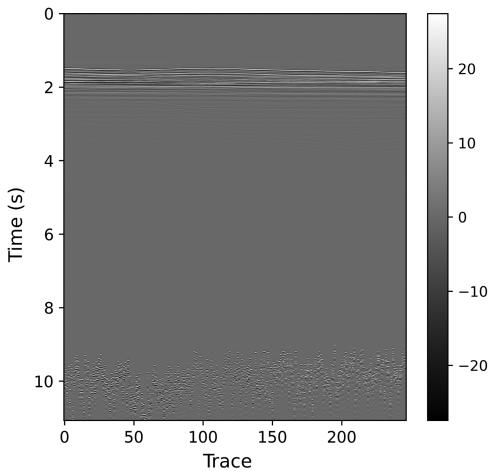


Figure 13. One common-receiver gather from the real simultaneous source acquisition data.

deblending performance of our algorithm, we run our algorithm on a manually blended real data example, which has much shorter time intervals and more overlap between the signal and blended noise. This specific data set, a 2D prestack marine example, is publicly accessible for download from the SEG website ([SEG Wiki, 2021](#)). In this data set, we manually fused four consecutive shots, giving rise to coherent source interference that appears in the common-shot gather.

In Figure 16a, we present an unaltered common-receiver gather for reference. In contrast, Figure 16b shows the same shot gather subjected to controlled blending, resulting in an  $S/N$  of approximately  $-1.23$  dB. In line with our previous experiments, we replicated the blending process using the pseudodeblended data set, introducing various time shifts. The resulting rebledened common-receiver gather, shown in Figure 16c and 16d, serves as the training input and the reference label for our model. The deblended outcome is shown in Figure 16e.

We adopted a similar strategy and trained the model on the entire data set to effectively mitigate all blended artifacts within the common-offset gather. Figure 17a and 17b provides a visual comparison between the original common-shot gather and the pseudodeblended counterpart. Conversely, Figure 17c and 17d emphasizes one of the final deblended common-shot gathers alongside the associated error analysis.

## DISCUSSION

### Why Noiser2Noiser works well on erratic noise

The proposed self-supervised learning network can be used to attenuate random and erratic noise. The first 2D synthetic example shows that the network exhibits better denoising performance with blended noise compared with random noise. Unlike traditional methods that often struggle to remove erratic noise (such as blended noise) due to its high amplitude, the network is more sensitive to the distribution of the noise rather than its amplitude alone. The network learns to distinguish between the signal and noise by understanding the differences between the training pairs. As demonstrated by these examples, although random noise has a smaller amplitude, it is pervasive. It corrupts all signals, making it noisier for the network than blended noise, which, in contrast, does not affect most parts of the signal. Therefore, it is easier to reconstruct signals corrupted by blended noise than those corrupted by random noise.

The original N2N paper ([Lehtinen et al., 2018](#)) demonstrated that the N2N method can effectively handle outliers when using the  $\ell_1$  loss function. This capability relies on having pairs of training data sets with the same signal but independently distributed noise. While adding more synthetic noise to the data is the most straightforward method to generate such pairs, it becomes challenging for real-world images where the noise distribution is often unknown. Alternative methods such as Noise2Void and Neighbor2Neighbor, which

do not require the generation of synthetic noise, have been proposed to address this challenge. These methods are based on the N2N theory but are not the same. For Neighbor2Neighbor, the ground truths of two downsampled noisy images are not the same. The Noise2Void uses the blind spot method instead of generating a pair of data with the same type of noise. The efficacy of these methods with erratic noise is not guaranteed. Most of the literature on these methods focuses on Gaussian and Poisson noise, with limited exploration of erratic noise scenarios. In the original Noise2Void paper, it is indeed mentioned that the method encountered challenges with outliers, but the specific reasons for this were not extensively elaborated in the paper. Most other methods do not mention erratic noise at all. Some of these methods probably work with the erratic noise by modifying the algorithm. However, the objective of this paper is not to adapt existing methods. Because we can generate independent blended noise from the original data, we opt for the most straightforward and theoretically aligned approach to remove the noise in our specific scenario. This choice allows us to stay true to the original N2N theory while effectively addressing the challenges of blended noise in seismic data processing.

In all of our examples, we consistently use  $\alpha = 0.5$ , indicating that the training pairs that we use are of the form  $(\mathbf{y} + 0.5 \times \mathbf{z}, \mathbf{y} - 2 \times \mathbf{z})$ . Consequently, the training target is always noisier than the training input. However, using a noisier training target is not a strict requirement. Due to the symmetric nature of the loss function we use, swapping the training input and target would yield identical results. Moreover, our method has a high tolerance for selecting  $\alpha$ . Using  $\alpha$  within a large range does not significantly alter the results. As shown in Figure 18, which illustrates the reconstructed result for the data presented in Figure 8b with different values of  $\alpha$ , using  $\alpha$  values between  $10^{-1}$  and  $10^1$  demonstrates negligible variations in performance. Even a broader range, such as  $10^{-2}$  to  $10^2$ , still yields acceptable results. In summary, while we consistently use  $\alpha = 0.5$  in our examples, the method remains robust to variations in this parameter.

### Proposed method versus traditional methods

Compared with traditional seismic denoising methods, particularly robust denoising methods for erratic noise, the proposed method offers several advantages. First, our method is fast and easier to implement. Traditional methods often rely on linear assumptions, limiting their applicability to linear data. This necessitates dividing data into smaller windows, increasing denoising costs significantly. In contrast, our method can operate directly on the entire shot gather. Second, traditional robust denoising methods often sacrifice weak events (those with amplitudes weaker than

noise in the transform domain) while attenuating erratic noise. However, our method preserves these weak events unless directly overlapped by strong erratic noise. Moreover, traditional methods are highly sensitive to parameters such as the minimum rank value in the rank reduction-based method or trade-off parameter  $\lambda$  in the regularization term of loss functions for sparse transform methods. Slight parameter adjustments can lead to significant result variations, posing challenges when processing real data in small windows. In contrast, our method requires only one parameter  $\alpha$ , and we have demonstrated its robustness across a wide range of values, minimizing the need for parameter optimization.

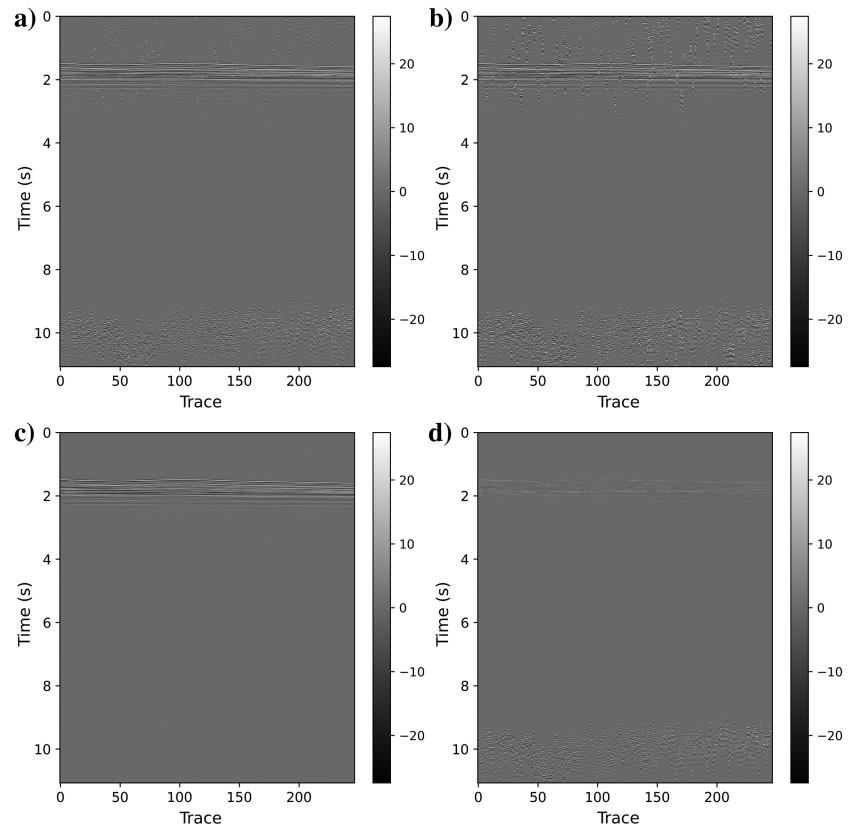


Figure 14. (a) Generated training input, (b) generated training input target, (c) deblended result, and (d) difference between (c) and Figure 13.

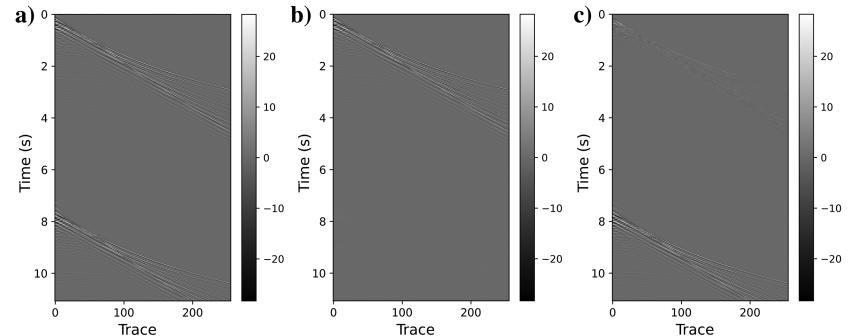


Figure 15. Common-shot gather of the final deblended result. The source interference from the other shot is separated successfully. (a) Original blended shot gather, (b) deblended shot gather, and (c) difference between (a and b).

## Limitations and future work

The finite-difference example shown in Figure 10 poses a significant challenge due to its curvature and the complete distortion of weak signals in the deep area. In this analysis, we compare our proposed method with two others. The first method is similar to ours but uses an  $\ell_2$  norm instead of the more robust  $\ell_1$  norm. Although the  $\ell_2$  norm is less robust against erratic noise, it typically results in signal reconstruction with less leakage than the  $\ell_1$  norm. The second method is a traditional approach based on the robust Radon transform (Ibrahim and Sacchi, 2014). This method addresses the deblending problem by minimizing an  $\ell_1 - \ell_1$  loss function using a robust solver such as iteratively reweighted least-squares (Chartrand and Yin, 2008) and alternating direction method of multipliers (Wen et al., 2016). For this method, we partitioned the data into small overlapping windows to adhere to the linear assumption of useful signals. We also optimized the trade-off parameter lambda for each window to balance deblending and signal leakage. To evaluate the deblended result's leakage, we present figures of the local similarity (Chen and Fomel, 2015) in this context. The local similarity measures the resemblance between the denoised result and the residual between the noise and denoised data, indicating signal leakage and retrieving useful signals from the residuals.

The findings shown in Figure 19 reveal significant insights. Although the  $\ell_2$  norm struggles to

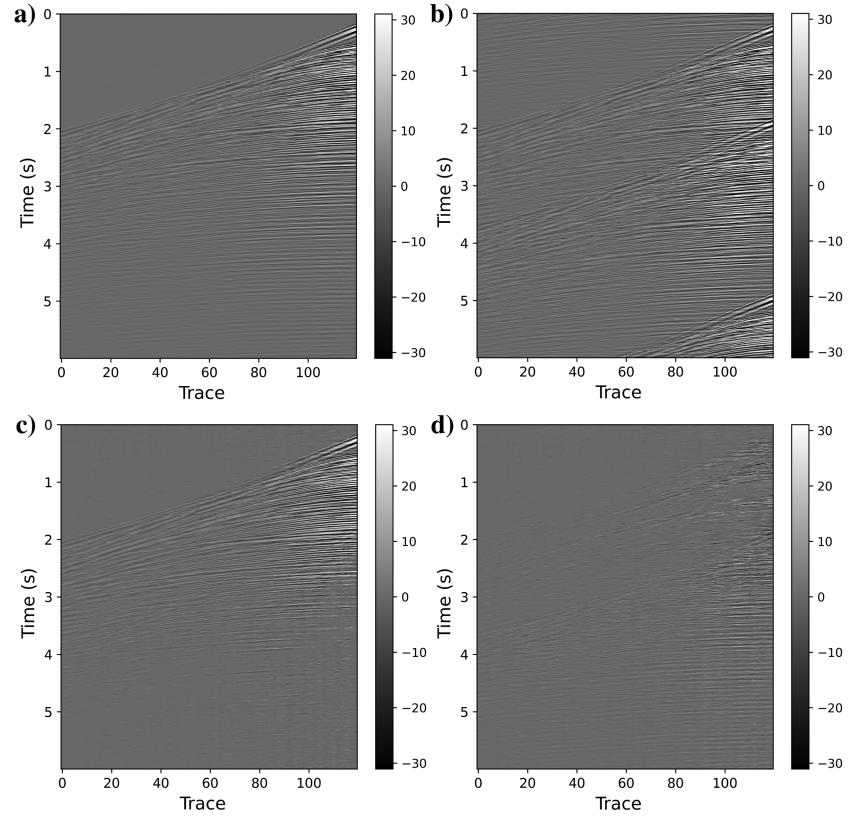


Figure 17. Common-shot gather of manually blended real marine data. Most blended noise is attenuated. (a) Clean data, (b) pseudodeblended data ( $S/N = -1.52$  dB), (c) deblended result ( $S/N = 14.3$  dB), and (d) errors between deblended results and clean data.

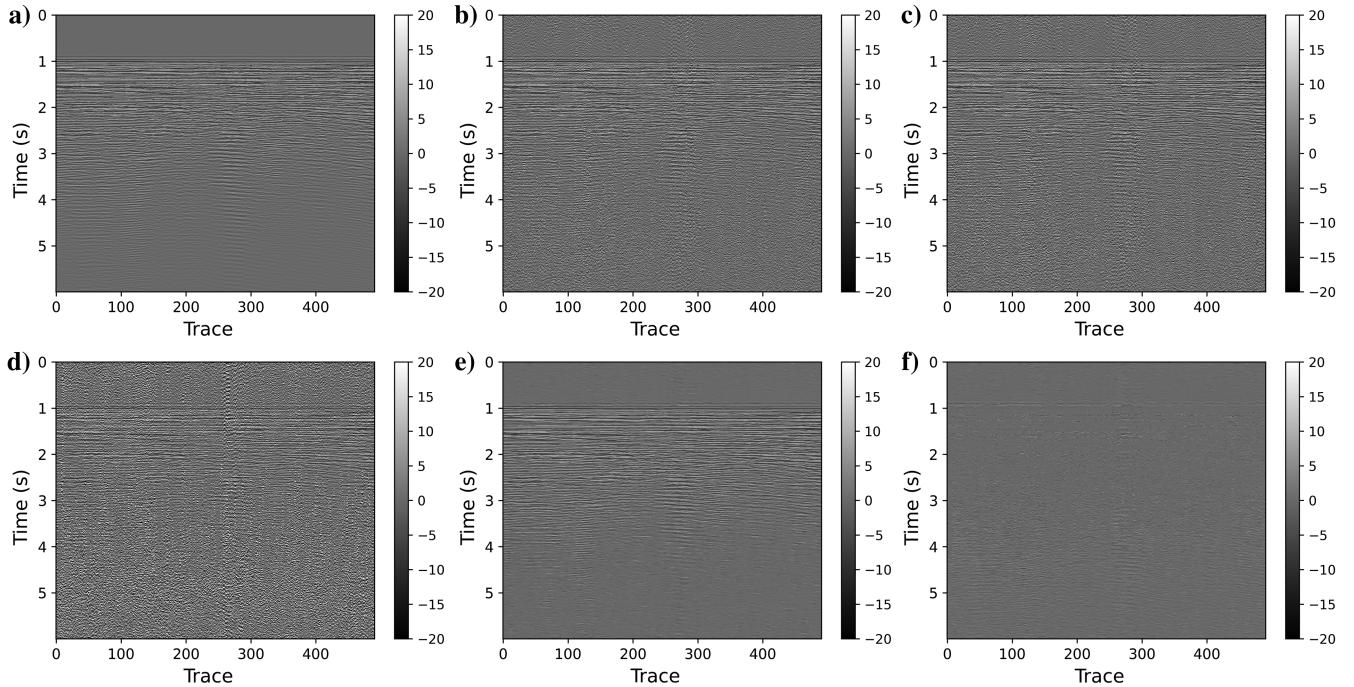


Figure 16. Common-receiver gather of manually blended real marine data. (a) Clean data, (b) noise data with blended noise ( $S/N = -1.23$  dB), (c) recorrupted training input, (d) recorrupted training label, (e) deblended result by the robust Noiser2Noiser ( $S/N = 15.4$  dB), and (f) errors between deblended results and clean data.

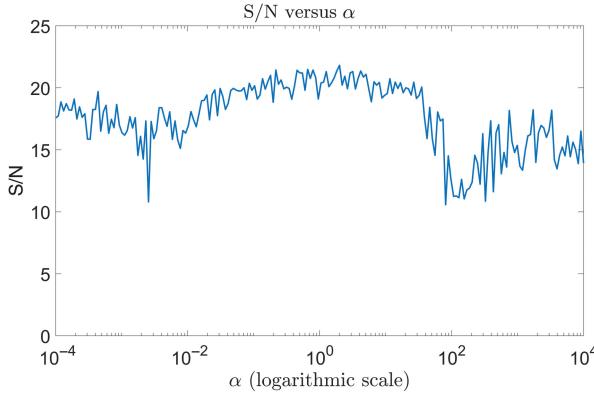


Figure 18. The S/N of reconstructed results for data in Figure 8 with different values of  $\alpha$ .

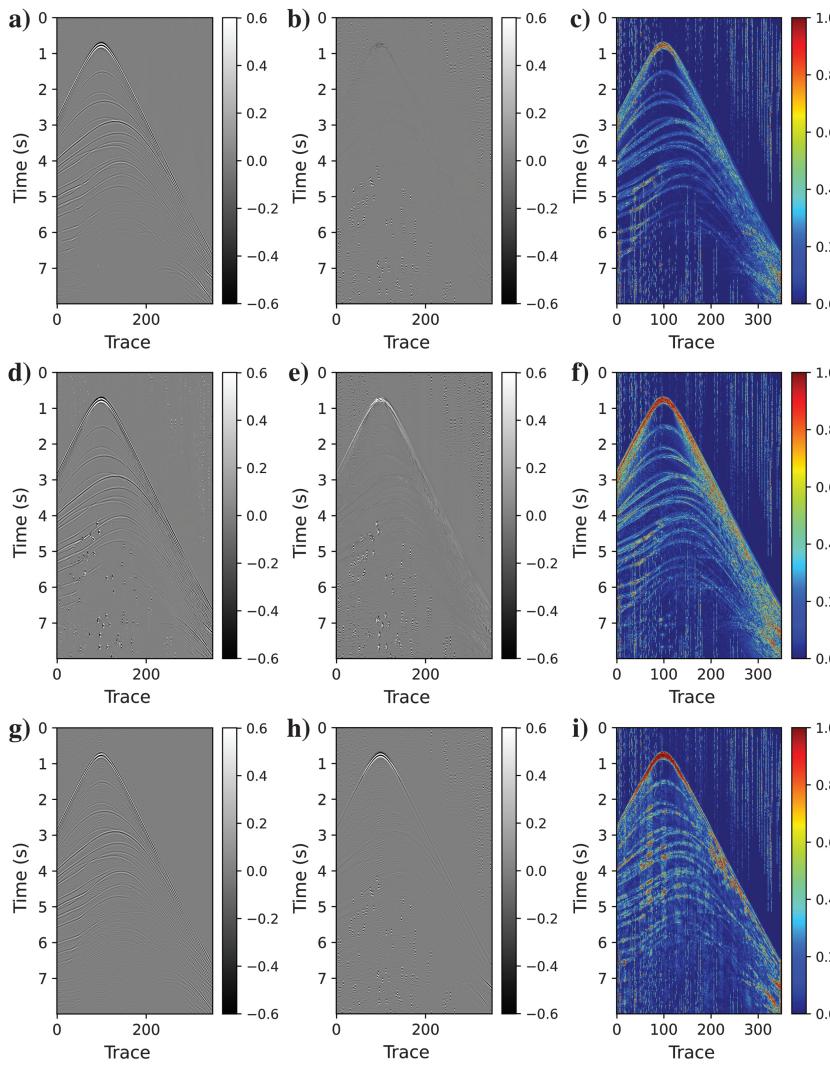


Figure 19. Comparison of the signal leakage for the example in Figure 10a with different methods. (a) Deblended result by the proposed method ( $S/N = 20.51$  dB), (b) residual between noise data in Figure 10 and denoised data (a), (c) local similarity between (a and b), (d) deblended result by the proposed method with  $\ell_2$  norm ( $S/N = 4.31$  dB), (e) residual between noise data in Figure 10 and denoised data (d), (f) local similarity between (d and e), (g) deblended result by the robust Radon method ( $S/N = 10.23$  dB), (h) residual between noise data in Figure 10 and denoised data, and (i) local similarity between (g and h).

robustly handle intense erratic noise, such as blended noise, evident in the deblended result and the local similarity graph, traditional methods such as the robust Radon transform are able to suppress most blended noise. However, they also may introduce distortions to the signal and generate undesired signals even in areas not directly affected by the noise. A notable advantage of the proposed method is its ability to reconstruct signals unaffected by noise fully. Nonetheless, in regions where strong blended noise obscures weak signals, the deblended result may exhibit discontinuities within the noisy areas.

Several potential solutions can be explored to address this limitation. One approach involves integrating the proposed method with traditional seismic data reconstruction techniques, leveraging the physical properties of seismic events to reconstruct data in these areas. Alternatively, introducing priors into the network could enhance the reconstruction results. These possibilities warrant further investigation.

Despite this limitation, our proposed method still stands out for its speed, simplicity, and robustness compared with traditional seismic denoising methods. Its effectiveness in handling erratic noise in seismic data makes it a promising approach for various applications.

## CONCLUSION

We introduce a novel zero-shot self-supervised framework designed to mitigate random and erratic noise effectively. In line with the foundational concept of N2N, we aim to reduce the dependency on clean training data. Our method uses a pair of recorrupted data sets during the training process, explaining how we synthesize random and blended noise (a specific subtype of erratic noise) for generating the training data.

To benchmark our approach, we compare it with two other N2N alternatives, conducting extensive evaluations on synthetic data. Subsequently, we apply our proposed framework to address the deblending challenge across synthetic and real marine data examples. To eliminate blended noise, we generate extra independent blended noise by applying the blending and pseudodeblending operators to the data set with different random time shifts. Then, the newly generated blended noise is used to construct a pair of training input and target. The trained model is finally deployed to detect and remove blended noise from the data set. These practical instances serve as compelling demonstrations of the framework's remarkable deblending performance, effectively eliminating blended noise without introducing significant signal leakage.

Notably, our self-supervised method achieves remarkable deblending results even without clean training data, comparable to, or surpassing, conventional algorithms and SL-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 is funded by CREWES industrial sponsors and the Natural Science and Engineering Research Council of Canada (NSERC) through the grant CRDPJ 543578-19. The contribution of the third author is supported by the industrial sponsors of the SAIG. We would like to acknowledge the use of ChatGPT, an artificial intelligence language model developed by OpenAI, for grammar checking and proofreading assistance during the preparation of this paper.

## DATA AND MATERIALS AVAILABILITY

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

## REFERENCES

- Abma, R., and J. Claerbout, 1995, Lateral prediction for noise attenuation by  $t$ - $x$  and  $f$ - $x$  techniques: *Geophysics*, **60**, 1887–1896, doi: [10.1190/1.1443920](https://doi.org/10.1190/1.1443920).
- Batson, J., and L. Royer, 2019, Noise2Self: Blind denoising by self-supervision: International Conference on Machine Learning, 524–533.
- Beasley, C. J., R. E. Chambers, and Z. Jiang, 1998, A new look at simultaneous sources: 68th Annual International Meeting, SEG, Expanded Abstracts, 133–135, doi: [10.1190/1.1820149](https://doi.org/10.1190/1.1820149).
- Berkhout, A. G., 2008, Changing the mindset in seismic data acquisition: *The Leading Edge*, **27**, 924–938, doi: [10.1190/1.2954035](https://doi.org/10.1190/1.2954035).
- Cao, S., and X. Chen, 2005, The second-generation wavelet transform and its application in denoising of seismic data: *Applied Geophysics*, **2**, 70–74, doi: [10.1007/s11770-005-0034-4](https://doi.org/10.1007/s11770-005-0034-4).
- Chartrand, R., and W. Yin, 2008, Iteratively reweighted algorithms for compressive sensing: IEEE International Conference on Acoustics, Speech and Signal Processing, 3869–3872.
- Chen, X., and K. He, 2021, Exploring simple siamese representation learning: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 15750–15758.
- Chen, Y., and S. Fomel, 2015, Random noise attenuation using local signal-and-noise orthogonalization: *Geophysics*, **80**, no. 6, WD1–WD9, doi: [10.1190/geo2014-0227.1](https://doi.org/10.1190/geo2014-0227.1).
- Deighan, A. J., and D. R. Watts, 1997, Ground-roll suppression using the wavelet transform: *Geophysics*, **62**, 1896–1903, doi: [10.1190/1.1444290](https://doi.org/10.1190/1.1444290).
- Górsczyk, A., A. Adamczyk, and M. Malinowski, 2014, Application of curvelet denoising to 2D and 3D seismic data — Practical considerations: *Journal of Applied Geophysics*, **105**, 78–94, doi: [10.1016/j.jappgeo.2014.03.009](https://doi.org/10.1016/j.jappgeo.2014.03.009).
- Guitton, A., and W. W. Symes, 2003, Robust inversion of seismic data using the Huber norm: *Geophysics*, **68**, 1310–1319, doi: [10.1190/1.1598124](https://doi.org/10.1190/1.1598124).
- He, K., X. Zhang, S. Ren, and J. Sun, 2015, Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification: Proceedings of the IEEE International Conference on Computer Vision, 1026–1034.
- He, K., X. Zhang, S. Ren, and J. Sun, 2016, Deep residual learning for image recognition: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.
- Helbing, G., and M. Ritter, 2018, Deep learning for fault detection in wind turbines: *Renewable and Sustainable Energy Reviews*, **98**, 189–198, doi: [10.1016/j.rser.2018.09.012](https://doi.org/10.1016/j.rser.2018.09.012).
- Hennenfent, G., and F. J. Herrmann, 2006, Seismic denoising with nonuniformly sampled curvelets: *Computing in Science & Engineering*, **8**, 16–25, doi: [10.1109/MCSE.2006.49](https://doi.org/10.1109/MCSE.2006.49).
- Hennenfent, G., and F. J. Herrmann, 2008, Simply denoise: Wavefield reconstruction via jittered undersampling: *Geophysics*, **73**, no. 3, V19–V28, doi: [10.1190/1.2841038](https://doi.org/10.1190/1.2841038).
- Huang, T., S. Li, X. Jia, H. Lu, and J. Liu, 2021, Neighbor2neighbor: Self-supervised denoising from single noisy images: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14781–14790.
- Ibrahim, A., and M. D. Sacchi, 2014, Simultaneous source separation using a robust Radon transform: *Geophysics*, **79**, no. 1, V1–V11, doi: [10.1190/geo2013-0168.1](https://doi.org/10.1190/geo2013-0168.1).
- Kaur, H., N. Pham, and S. Fomel, 2021, Seismic data interpolation using deep learning with generative adversarial networks: *Geophysical Prospecting*, **69**, 307–326, doi: [10.1111/1365-2478.13055](https://doi.org/10.1111/1365-2478.13055).
- Krull, A., T.-O. Buchholz, and F. Jug, 2019, Noise2void-learning denoising from single noisy images: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2129–2137.
- Lari, H. H., M. Naghizadeh, M. D. Sacchi, and A. Gholami, 2019, Adaptive singular spectrum analysis for seismic denoising and interpolation: *Geophysics*, **84**, no. 2, V133–V142, doi: [10.1190/geo2018-0350.1](https://doi.org/10.1190/geo2018-0350.1).
- Latif, A., and W. A. Mousa, 2016, An efficient undersampled high-resolution Radon transform for exploration seismic data processing: *IEEE Transactions on Geoscience and Remote Sensing*, **55**, 1010–1024, doi: [10.1109/TGRS.2016.2618848](https://doi.org/10.1109/TGRS.2016.2618848).
- Lehtinen, J., J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, 2018, Noise2noise: Learning image restoration without clean data: arXiv preprint, doi: [10.48550/arXiv.1803.04189](https://doi.org/10.48550/arXiv.1803.04189).
- Lequyer, J., R. Philip, A. Sharma, W.-H. Hsu, and L. Pelletier, 2022, A fast blind zero-shot denoiser: *Nature Machine Intelligence*, **4**, 953–963, doi: [10.1038/s42256-022-00547-8](https://doi.org/10.1038/s42256-022-00547-8).
- Li, H., W. Yang, and X. Yong, 2018, Deep learning for ground-roll noise attenuation: 88th Annual International Meeting, SEG, Expanded Abstracts, 1981–1985, doi: [10.1190/segam2018-2981295.1](https://doi.org/10.1190/segam2018-2981295.1).
- Li, J., and M. D. Sacchi, 2021, An  $l_p$ -space matching pursuit algorithm and its application to robust seismic data denoising via time-domain Radon transforms: *Geophysics*, **86**, no. 3, V171–V183, doi: [10.1190/geo2020-0136.1](https://doi.org/10.1190/geo2020-0136.1).
- Li, S., B. Liu, Y. Ren, Y. Chen, S. Yang, Y. Wang, and P. Jiang, 2019, Deep-learning inversion of seismic data: arXiv preprint, doi: [10.48550/arXiv.1901.07733](https://doi.org/10.48550/arXiv.1901.07733).
- Liu, D., Z. Deng, C. Wang, X. Wang, and W. Chen, 2020, An unsupervised deep learning method for denoising prestack random noise: *IEEE Geoscience and Remote Sensing Letters*, **19**, 1–5, doi: [10.1109/LGRS.2020.3019400](https://doi.org/10.1109/LGRS.2020.3019400).
- Liu, D., M. D. Sacchi, X. Wang, and W. Chen, 2023, Unsupervised deep learning for ground roll and scattered noise attenuation: *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 5920613, doi: [10.1109/TGRS.2023.3325324](https://doi.org/10.1109/TGRS.2023.3325324).
- Liu, D., W. Wang, W. Chen, X. Wang, Y. Zhou, and Z. Shi, 2018, Random noise suppression in seismic data: What can deep learning do? 88th Annual International Meeting, SEG, Expanded Abstracts, 2016–2020, doi: [10.1190/segam2018-2998114.1](https://doi.org/10.1190/segam2018-2998114.1).
- Mansouri, Y., and R. Heckel, 2023, Zero-shot noise2noise: Efficient image denoising without any data: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14018–14027.
- Maronna, R. A., 1976, Robust  $m$ -estimators of multivariate location and scatter: *The Annals of Statistics*, **4**, 51–67, doi: [10.1214/aos/1176343347](https://doi.org/10.1214/aos/1176343347).
- Matharu, G., W. Gao, R. Lin, Y. Guo, M. Park, and M. D. Sacchi, 2020, Simultaneous source deblending using a deep residual network: SEG 2019 Workshop: Mathematical Geophysics: Traditional vs Learning, 13–16, doi: [10.1190/iwmg2019\\_04.1](https://doi.org/10.1190/iwmg2019_04.1).
- Moran, N., D. Schmidt, Y. Zhong, and P. Coady, 2020, Noisier2noise: Learning to denoise from unpaired noisy data: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12064–12072.
- Mousavi, S. M., and C. A. Langston, 2016, Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding: *Bulletin of the Seismological Society of America*, **106**, 1380–1393, doi: [10.1785/0210150345](https://doi.org/10.1785/0210150345).
- Oliveira, D. A., R. S. Ferreira, R. Silva, and E. V. Brazil, 2018, Interpolating seismic data with conditional generative adversarial networks: *IEEE Geoscience and Remote Sensing Letters*, **15**, 1952–1956, doi: [10.1109/LGRS.2018.2866199](https://doi.org/10.1109/LGRS.2018.2866199).
- Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: *Geophysics*, **76**, no. 3, V25–V32, doi: [10.1190/1.3552706](https://doi.org/10.1190/1.3552706).
- Othman, A., N. Iqbal, S. M. Hanafy, and U. B. Waheed, 2021, Automated event detection and denoising method for passive seismic data using residual deep convolutional neural networks: *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1–11, doi: [10.1109/TGRS.2021.3054071](https://doi.org/10.1109/TGRS.2021.3054071).
- Pang, T., H. Zheng, Y. Quan, and H. Ji, 2021, Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2043–2052.
- Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, and A. Desmaison, 2019, PyTorch: An imperative style, high-performance deep learning library: *Advances in Neural Information Processing Systems*.
- Qian, F., W. Guo, Z. Liu, H. Yu, G. Zhang, and G. Hu, 2022, Unsupervised erratic seismic noise attenuation with robust deep convolutional autoencoders: *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1–16, doi: [10.1109/TGRS.2022.3158389](https://doi.org/10.1109/TGRS.2022.3158389).
- Quan, Y., M. Chen, T. Pang, and H. Ji, 2020, Self2self with dropout: Learning selfsupervised denoising from single image: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1890–1898.

- Richardson, A., and C. Feller, 2019, Seismic data denoising and deblending using deep learning: arXiv preprint, doi: [10.48550/arXiv.1907.01497](https://doi.org/10.48550/arXiv.1907.01497).
- Ronneberger, O., P. Fischer, and T. Brox, 2015, U-Net: Convolutional networks for biomedical image segmentation: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, 234–241.
- Saad, O. M., Y. A. S. I. Oboue, M. Bai, L. Samy, L. Yang, and Y. Chen, 2021, Self-attention deep image prior network for unsupervised 3-D seismic data enhancement: IEEE Transactions on Geoscience and Remote Sensing, **60**, 1–14, doi: [10.1109/TGRS.2021.3108515](https://doi.org/10.1109/TGRS.2021.3108515).
- SEG Wiki, 2018, PGS simultaneous source marine line, [https://wiki.seg.org/wiki/PGS\\_Simultaneous\\_Source\\_Marine\\_Line](https://wiki.seg.org/wiki/PGS_Simultaneous_Source_Marine_Line), accessed 17 May 2018.
- SEG Wiki, 2021, Mobil AVO viking graben line 12, [10.48550/arXiv.1907.01497](https://doi.org/10.48550/arXiv.1907.01497), accessed 9 September 2021.
- Siahzar, M. A. N., S. Gholashiz, E. O. Torshizi, W. Chen, and Y. Chen, 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**, 1086–1090, doi: [10.1109/LGRS.2017.2697942](https://doi.org/10.1109/LGRS.2017.2697942).
- Sun, H., F. Yang, and J. Ma, 2022a, Seismic random noise attenuation via self-supervised transfer learning: IEEE Geoscience and Remote Sensing Letters, **19**, 1–5, doi: [10.1109/LGRS.2022.3146173](https://doi.org/10.1109/LGRS.2022.3146173).
- Sun, J., S. Hou, V. Vinje, G. Poole, and L.-J. Gelius, 2022b, Deep learning-based shot-domain seismic deblending: Geophysics, **87**, no. 3, V215–V226, doi: [10.1190/geo2020-0865.1](https://doi.org/10.1190/geo2020-0865.1).
- Trickett, S., L. Burroughs, and A. Milton, 2012, Robust rank-reduction filtering for erratic noise: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: [10.1190/segam2012-0129.1](https://doi.org/10.1190/segam2012-0129.1).
- Ulyanov, D., A. Vedaldi, and V. Lempitsky, 2018, Deep image prior: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 9446–9454.
- Wang, B., N. Zhang, W. Lu, and J. Wang, 2019, Deep-learning-based seismic data interpolation: A preliminary result: Geophysics, **84**, no. 1, V11–V20, doi: [10.1190/geo2017-0495.1](https://doi.org/10.1190/geo2017-0495.1).
- Wang, K., and T. Hu, 2021, Deblending of seismic data based on neural network trained in the CSG: IEEE Transactions on Geoscience and Remote Sensing, **60**, 1–12, doi: [10.1109/TGRS.2021.3121806](https://doi.org/10.1109/TGRS.2021.3121806).
- Wang, K., T. Hu, and S. Wang, 2023a, Iterative deblending using unsupervised learning with double-deep neural networks: Geophysics, **88**, no. 3, V187–V205, doi: [10.1190/geo2022-0299.1](https://doi.org/10.1190/geo2022-0299.1).
- Wang, S., W. Hu, P. Yuan, X. Wu, Q. Zhang, P. Nadukandi, G. O. Botero, and J. Chen, 2022, A self-supervised deep learning method for seismic data deblending using a blindtrace network: IEEE Transactions on Neural Networks and Learning Systems, **34**, 3405–3414, doi: [10.1109/TNNLS.2022.3188915](https://doi.org/10.1109/TNNLS.2022.3188915).
- Wang, X., S. Fan, C. Zhao, D. Liu, and W. Chen, 2023b, A self-supervised method using noise2noise strategy for denoising CRP gathers: IEEE Geoscience and Remote Sensing Letters, **20**, 7503505, doi: [10.1109/LGRS.2023.3285951](https://doi.org/10.1109/LGRS.2023.3285951).
- Wen, F., P. Liu, Y. Liu, R. C. Qiu, and W. Yu, 2016, Robust sparse recovery in impulsive noise via  $L_0$ - $L_1$  optimization: IEEE Transactions on Signal Processing, **65**, 105–118, doi: [10.1109/TSP.2016.2598316](https://doi.org/10.1109/TSP.2016.2598316).
- Xu, J., Y. Huang, M.-M. Cheng, L. Liu, F. Zhu, Z. Xu, and L. Shao, 2020, Noisy-as-clean: Learning self-supervised denoising from corrupted image: IEEE Transactions on Image Processing, **29**, 9316–9329, doi: [10.1109/TIP.2020.3026622](https://doi.org/10.1109/TIP.2020.3026622).
- Yu, S., J. Ma, X. Zhang, and M. D. Sacchi, 2015, Interpolation and denoising of high-dimensional seismic data by learning a tight frame: Geophysics, **80**, no. 5, V119–V132, doi: [10.1190/geo2014-0396.1](https://doi.org/10.1190/geo2014-0396.1).
- Zhang, F., D. Liu, X. Wang, W. Chen, and W. Wang, 2018, Random noise attenuation method for seismic data based on deep residual networks: International Geophysical Conference, 1774–1777.
- Zhang, K., W. Zuo, Y. Chen, D. Meng, and L. Zhang, 2017, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising: IEEE Transactions on Image Processing, **26**, 3142–3155, doi: [10.1109/TIP.2017.2662206](https://doi.org/10.1109/TIP.2017.2662206).
- Zheng, Y., Q. Zhang, A. Yusifov, and Y. Shi, 2019, Applications of supervised deep learning for seismic interpretation and inversion: The Leading Edge, **38**, 526–533, doi: [10.1190/tle38070526.1](https://doi.org/10.1190/tle38070526.1).

Biographies and photographs of the authors are not available.