

Seismic Intelligent Deblending via Plug and Play Method With Blended CSGs Trained Deep CNN Gaussian Denoiser

Weiwei Xu^{ID}, Yanhui Zhou, *Member, IEEE*, Dawei Liu^{ID}, Xiaokai Wang^{ID}, *Member, IEEE*, and Wenchao Chen^{ID}

Abstract—Deblending can extract good quality seismic data from blended seismic data. Generally, the deblending methods can be categorized as model- and data-driven methods. The model-driven deblending methods usually suffer from a massive computational burden, while the data-driven ones need to implement a forward blending process to construct training samples for network training. To solve these issues, we develop a plug and play (PnP) method that integrates a trained convolutional neural network (CNN) Gaussian denoiser as the prior for seismic intelligent deblending. Specifically, we propose to use acquired blended common shot gathers (CSGs) as training dataset for the CNN Gaussian denoiser training to avoid constructing or collecting any extra data. According to the theory and dedicated designed experiments, this training mode can greatly improve network performance. Then, the trained CNN Gaussian denoiser is plugged into the alternating direction method of multiplier (ADMM) algorithm to solve the deblending problem. Furthermore, based on the l_2 norm data fidelity term and the special structure of the network architecture, the PnP-ADMM method converges to a fixed point. Experiments on synthetic and field data demonstrate that the presented PnP method with blended CSGs trained deep CNN Gaussian denoiser has superior deblending performances over the dictionary learning and discriminative deblending CNN methods.

Index Terms—Alternating direction method of multipliers (ADMMs), blended common shot gathers (CSGs), deep learning, plug and play (PnP), seismic deblending, spatial reciprocity.

I. INTRODUCTION

SEISMIC blended acquisition is an increasingly developing technique which fires multiple shots at different spatial locations in a very short time and thus can radically shorten exploration duration, reduce economic costs, and even boost data quality [1], [2]. Nevertheless, it also leads to mutual interference between shot records, which bring a great challenge to later traditional processing and imaging workflows. To make the blended acquisition applied to practice, many attempts have been made to deal with the blended seismic data reasonably. A straightforward way is direct imaging and waveform inversion [3]. For example, Tang and Biondi [4] imaged

blended data directly using the least-squares migration (LSM) technique, which improved the final image resolution. Based on the LSM, Dai *et al.* [5] developed a multisource LSM algorithm to simultaneously increase the computational efficiency and improve the image quality. Chen *et al.* [6] proposed a least-square reverse time migration (LSRTM) method to process blended data by introducing local singular spectrum analysis (SSA) operator within iterations. Cao and Blacquiére [7] extended the joint migration inversion (JMI) method to directly obtain the reflectivity data and propagation velocities from the blended data. Although the above direct methods can eliminate the impact of crosstalk without any pre-separation operation of the blended data, sometimes we still need the information contained in the unblended single shot data. Furthermore, the existing conventional processing workflows have been carefully planned for unblended single shot data. Therefore, it is really necessary to develop deblending methods, which aim to separate the blended data into individual shots.

A lot of advanced algorithms have been successfully applied to deblending. During the last decade, model-driven methods consisting of the blending noise model and the prior knowledge of useful seismic signal have been developed rapidly [8]–[22]. Since Huo *et al.* [8] utilized the multidirectional vector-median filter to suppress blending noise in common midpoint domain, several median filtering-based methods emerged. For example, Chen [9] presented a space-varying median filter algorithm which detected blending noise by using signal reliability (SR) as a reference. Huang *et al.* [10] utilized structuring median filtering (SMF) to promote the iterative deblending. Rank reduction methods, which assume that the useful seismic signal should have low rank property after proper data rearrangement processes, have also been used for seismic deblending. Cheng *et al.* [11] used the projected gradient technique with the SSA filter as the projection operator to present an iterative deblending algorithm. Zhou *et al.* [12] developed an iterative deblending method combining the low rankness and threshold constraints. Sparsity promoting is also a practical scheme on the basis of the sparse representation of the useful seismic data in appropriate transform domains, such as Fourier transform [13], generalized windowed transforms [14], curvelet transform [15], and learned dictionary [16]–[18]. It is worth mentioning that taking advantage of low-rankness and sparsity priors, some robust methods have been developed for separating the blended data. Cheng *et al.* [19] proposed a

Manuscript received August 18, 2021; revised February 3, 2022; accepted April 5, 2022. Date of publication April 8, 2022; date of current version April 21, 2022. This work was supported by the National Natural Science Foundation of China under Grant 41974131 and Grant 41774135. The work of Weiwei Xu and Dawei Liu was supported by the China Scholarship Council. (Corresponding author: Wenchao Chen.)

The authors are with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: 544274644@qq.com; zhoyuh@xjtu.edu.cn; liudawei2015@stu.xjtu.edu.cn; xkwang@xjtu.edu.cn; wencchen@xjtu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2022.3166029

robust principal component analysis (RPCA) method adopting the nuclear norm regularization to yield the low-rank approximation of the useful signal and utilizing an l_1 norm constraint to reasonably model the blending noise. Ibrahim and Sacchi [20] demonstrated that deblending can be effectively accomplished by regarding blending noise as outliers and modeling the data with a Radon transform. Zhao *et al.* [21] devised a normalized shaping regularization in Curvelet domain to make the inversion-based deblending method robust when erratic noise is present. Wang *et al.* [22] introduced a robust sparse Radon transform (RSRT)-based algorithm to reduce erratic noise influence. Although the model-driven deblending methods have shown their outstanding performance, the effectiveness of these techniques relies on the handcrafted prior assumptions imposed on the useful signal to be recovered. Therefore, they could not constantly adapt to different seismic data with highly complex structures. Furthermore, these methods often suffer from a huge calculation burden, particularly when dealing with enormous amounts of seismic data [25].

In recent years, the data-driven methods based on deep learning techniques have also been introduced to the area of deblending [23]–[27]. In particular, Sun *et al.* [23] collected more than 20 000 field training examples to train a convolutional neural network (CNN) for seismic deblending. Zu *et al.* [24] embedded the CNN trained on synthetic and field datasets into an iterative strategy to further increase the deblending effect. Wang *et al.* [25] proposed a U-net-based deblending algorithm that used parts of labeled field data to further fine-tune the U-net trained by labeled synthetic data and incorporated an iterative framework with the fine-tuned U-net. Nakayama and Blacqui re [26] used a supervised deep CNN to suppress blending noise, reconstruct missing traces, and extrapolate low frequencies simultaneously. Wang *et al.* [27] presented a physics-augmented deep learning approach which utilized the trained CNN to predict the deblending results on the field data and then used them to update the velocity model, which would be used to generate synthetic training datasets for updating the CNN in turn. It can be seen that the data-driven methods have evident efficiencies as a result of their strong ability to learn from massive seismic data. However, most of them, on the one side, need to collect training datasets from other work areas or construct appropriate synthetic training datasets; on the other side, most of them include a discriminative deblending CNN trained on the pseudo-deblended data and conventional unblended data, necessitating the forward blending process.

The plug and play (PnP) methods [28]–[34] separate the data fidelity and regularization terms into different simple subproblems via various variable splitting techniques, such as projection onto convex sets (POCSs) [34], alternating direction method of multipliers (ADMMs) [35], and half quadratic splitting (HQS) [36]. In addition, the regularization term relates to a Gaussian denoising subproblem, which can be solved via trained deep CNN Gaussian denoisers. As a result, PnP methods [28]–[34] can combine the advantages of both model- and data-driven approaches. In this article, we present a

PnP method that incorporates a blended common shot gathers (CSGs) trained deep CNN Gaussian denoiser as the prior for seismic intelligent deblending. Moreover, based on the l_2 norm data fidelity term and the special structure of the network architecture, the PnP-ADMM method converges to a fixed point. Specifically, our contributions are summarized as follows.

- 1) To the best of our knowledge, we are the first to use PnP method with a trained deep CNN Gaussian denoiser for seismic deblending. Specifically, by considering the seismic deblending as an inversion problem, in this method, instead of a discriminative deblending CNN which is trained on pseudo-deblended data and conventional unblended data, a trained deep CNN Gaussian denoiser prior is integrated into the ADMM algorithm.
- 2) We propose to use acquired blended CSGs as training dataset for CNN Gaussian denoiser training, in order to avoid constructing or collecting any extra data. And we analyze the rationality and superiority of this training mode theoretically and experimentally.
- 3) Experiments on synthetic and field datasets show that our proposed method outperforms dictionary learning and discriminative deblending CNN methods.

The remainder of this article is organized as follows. In Section II, we introduce the presented method in great detail. In Section III, the experiments on synthetic and field seismic datasets are demonstrated. Finally, conclusion and future prospects are drawn in Section IV.

II. METHOD

A. Problem Formulation

In the concept of seismic blended acquisition, multiple shots are fired in a certain order with the given random dithering code, such that all the individual source data are blended and further received by the same receiver arrays [17]. Although interference noise is unavoidably introduced during the process of blended acquisition, it can be effectively suppressed in common shot domain before deblending [21]. Therefore, ignoring the interference noise effects, for a fixed receiver, the blended seismic data $\mathbf{b}$ can be modeled as

$$\mathbf{b} = \Gamma \mathbf{d} \quad (1)$$

where $\mathbf{d} \in \mathbb{R}^{(m \times n) \times 1}$ is the vector form of the conventional unblended common receiver gather (CRG) data with n shots and m time samples. Because of the blending effect, the measurement $\mathbf{b}$ is a long vector of length f , which can be approximated by

$$f \approx \frac{m \times n}{\text{STR}} \quad (2)$$

where STR is survey time ratio, which can be used to estimate the time efficiency of used blended acquisition [15]

$$\text{STR} = \frac{\text{time of conventional shooting of } n \text{ shots}}{\text{time of blended shooting of } n \text{ shots}} \quad (3)$$

where $1 \leq \text{STR} \leq n$. Specially, $\text{STR} = 1$ and $\text{STR} = n$ represent the conventional acquisition and simultaneous

source acquisition, respectively. $\mathbf{\Gamma} \in \mathbb{R}^{f \times (m \times n)}$ represents the blending matrix, which is set as a block matrix

$$\mathbf{\Gamma} = [\mathbf{H}_1, \dots, \mathbf{H}_n] \quad (4)$$

where $\{\mathbf{H}_i\}_{i=1}^n \in \mathbb{R}^{f \times m}$ can also be expressed as a block matrix and constructed as follows:

$$\mathbf{H}_i = \begin{bmatrix} \mathbf{0}_{i1} \\ \mathbf{I}_m \\ \mathbf{0}_{i2} \end{bmatrix} \quad (5)$$

where $\mathbf{I}_m \in \mathbb{R}^{m \times m}$ is an identity matrix, and $\mathbf{0}_{i1}$ and $\mathbf{0}_{i2}$ denote the zero matrices of size $c \times m$ and $(f - m - c) \times m$, respectively. The integer c with $0 \leq c \leq (f - m)$ varies with the firing order of shots at different positions and the given random dithering code. Correspondingly, we have

$$\mathbf{H}_i^T \mathbf{H}_i = \mathbf{0}_{i1}^T \mathbf{0}_{i1} + \mathbf{I}_m^2 + \mathbf{0}_{i2}^T \mathbf{0}_{i2} = \mathbf{I}_m. \quad (6)$$

Furthermore, we can get

$$\begin{aligned} \mathbf{\Gamma}^T \mathbf{\Gamma} &= \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_n \end{bmatrix} [\mathbf{H}_1 \quad \dots \quad \mathbf{H}_n] \\ &= \begin{bmatrix} \mathbf{I}_m & \mathbf{H}_1^T \mathbf{H}_2 & \dots & \mathbf{H}_1^T \mathbf{H}_n \\ \mathbf{H}_1^T \mathbf{H}_2 & \mathbf{I}_m & \dots & \mathbf{H}_2^T \mathbf{H}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_1^T \mathbf{H}_n & \mathbf{H}_2^T \mathbf{H}_n & \dots & \mathbf{I}_m \end{bmatrix} \\ \mathbf{\Gamma}^T \mathbf{b} &= \mathbf{\Gamma}^T \mathbf{\Gamma} \mathbf{d} = \mathbf{d} + \begin{bmatrix} \mathbf{0}_m & \mathbf{H}_1^T \mathbf{H}_2 & \dots & \mathbf{H}_1^T \mathbf{H}_n \\ \mathbf{H}_1^T \mathbf{H}_2 & \mathbf{0}_m & \dots & \mathbf{H}_2^T \mathbf{H}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_1^T \mathbf{H}_n & \mathbf{H}_2^T \mathbf{H}_n & \dots & \mathbf{0}_m \end{bmatrix} \mathbf{d} \end{aligned} \quad (7)$$

where $\mathbf{\Gamma}^T \mathbf{b}$ is pseudo-deblended data, and the second term on the right in (8) denotes the blending noise, which is laterally incoherent in the non-shot domain based on the given random dithering code [1]. Therefore, it is reasonable to get the recovery $\mathbf{d}$ from pseudo-deblended data $\mathbf{\Gamma}^T \mathbf{b}$ via different filtering-based methods [8], [20], [23]–[25], [27].

Besides filtering-based methods, a more stable way is to construct the deblending process as an inversion problem. Generally, this problem is ill-posed. Therefore, a proper prior, also known as regularization, has to be added. From the Bayesian perspective, we can formulate this problem as a maximum *a posteriori* (MAP) estimation, which can be written as

$$\hat{\mathbf{d}} = \arg \max_{\mathbf{d}} \log p(\mathbf{b}|\mathbf{d}) + \log p(\mathbf{d}) \quad (9)$$

where $\log p(\mathbf{b}|\mathbf{d})$ denotes the log-likelihood of measurement $\mathbf{b}$, and $\log p(\mathbf{d})$ defines the probability distribution of the unblended data $\mathbf{d}$. Typically, solving the MAP problem in (9) corresponds to solve the below optimization problem

$$\hat{\mathbf{d}} = \arg \min_{\mathbf{d}} \frac{1}{2} \|\mathbf{b} - \mathbf{\Gamma} \mathbf{d}\|_2^2 + \lambda \mathfrak{R}(\mathbf{d}) \quad (10)$$

where $(1/2)\|\mathbf{b} - \mathbf{\Gamma} \mathbf{d}\|_2^2$ represents the data fidelity term, $\mathfrak{R}(\mathbf{d})$ denotes the regularization term, and λ is the nonnegative weighting parameter.

Regularization term plays a vital role in solving the optimization problem in (10) which constricts the desired property of the solution. As mentioned earlier, although the hand-crafted priors, including low rankness [11], [12] and sparsity [13]–[18], can bring a great deblending effect, they also result in a huge computational burden, which causes great difficulty in practical application. Compared with the hand-crafted priors, the trained CNNs can comparably or even better capture the features of the useful seismic signals and have a much faster testing speed owing to end-to-end training [29]. Notably, they have been successfully applied to seismic noise suppression [37], [38]. Therefore, in order to exploit fully the advantages of model- and data-driven methods, we use PnP-ADMM framework [28] to impose the trained deep CNN Gaussian denoiser as the regularization into the optimization problem.

B. PnP-ADMM Framework

In ADMM algorithm [35], by introducing an auxiliary variable $\mathbf{x}$ which is constrained to be equal to $\mathbf{d}$, the unconstrained optimization in (10) can be equivalently transformed into a constrained problem

$$(\mathbf{d}, \mathbf{x}) = \arg \min_{\mathbf{d}, \mathbf{x}} \frac{1}{2} \|\mathbf{b} - \mathbf{\Gamma} \mathbf{d}\|_2^2 + \lambda \mathfrak{R}(\mathbf{x}) \quad \text{s.t. } \mathbf{d} = \mathbf{x}. \quad (11)$$

The augmented Lagrangian function of (11) can be written as

$$\mathcal{L}(\mathbf{d}, \mathbf{x}, \mathbf{u}) = \frac{1}{2} \|\mathbf{b} - \mathbf{\Gamma} \mathbf{d}\|_2^2 + \lambda \mathfrak{R}(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{d} + \mathbf{u}\|_2^2 \quad (12)$$

where ρ is a positive penalty parameter, and $\mathbf{u}$ denotes the dual variable. The solution of (11) is the saddle point of augmented Lagrangian function $\mathcal{L}$, which can be gotten by solving the following subproblems in an alternating iterative manner:

$$\mathbf{x}^{(k+1)} = \arg \min_{\mathbf{x}} \frac{\rho}{2} \|(\mathbf{d}^{(k)} - \mathbf{u}^{(k)}) - \mathbf{x}\|_2^2 + \lambda \mathfrak{R}(\mathbf{x}) \quad (13)$$

$$\mathbf{d}^{(k+1)} = \arg \min_{\mathbf{d}} \frac{1}{2} \|\mathbf{b} - \mathbf{\Gamma} \mathbf{d}\|_2^2 + \frac{\rho}{2} \|\mathbf{d} - (\mathbf{x}^{(k+1)} + \mathbf{u}^{(k)})\|_2^2 \quad (14)$$

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \mathbf{x}^{(k+1)} - \mathbf{d}^{(k+1)} \quad (15)$$

where k is the iteration number. Building upon the PnP-ADMM framework [28], (13) is the proximal operator of the regularization $\mathfrak{R}(\mathbf{x})$ with penalty ρ/λ . And from a Bayesian perspective, (13) corresponds to denoising the intermediate $(\mathbf{d}^{(k)} - \mathbf{u}^{(k)})$ via a Gaussian denoiser [48]

$$\mathbf{x}^{(k+1)} = \text{Denoiser}((\mathbf{d}^{(k)} - \mathbf{u}^{(k)}), \sqrt{\lambda/\rho}) \quad (16)$$

where $\sqrt{\lambda/\rho}$ is the estimated noise standard deviation. Thus, the trained deep CNN Gaussian denoiser can be introduced because of its fast speed and excellent performance. As for (14), it is a concise quadratic optimization problem with a closed-form solution as

$$\mathbf{d}^{(k+1)} = (\mathbf{\Gamma}^T \mathbf{\Gamma} + \rho \mathbf{I})^{-1} (\mathbf{\Gamma}^T \mathbf{b} + \rho (\mathbf{x}^{(k+1)} + \mathbf{u}^{(k)})). \quad (17)$$

Here, in order to avoid matrix inversion, we use the highly efficient conjugate gradient (CG) method to solve (14) instead.

Some progresses have been made on the theoretical convergence of PnP-ADMM methods under explicit assumptions [31]–[33]. Especially, Sun *et al.* [33] proved that given

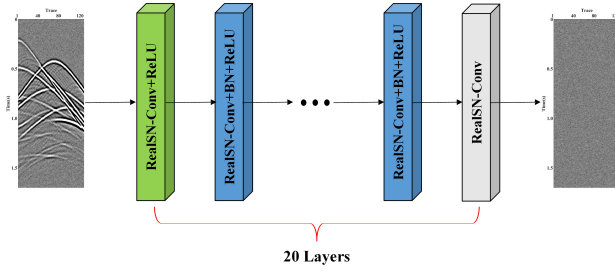


Fig. 1. RealSN-DnCNN network architecture.

the l_2 norm data fidelity term, the PnP-ADMM converges to a fixed point if the CNN denoiser is firmly nonexpansive, which is a guideline for the network architecture construction.

C. Learning CNN Denoiser Prior

1) *CNN Denoiser Architecture*: The CNN denoiser needs to have the ability to obtain good denoising effect in the case of keeping the nonexpansiveness [33]. Building a firmly nonexpansive CNN is equivalent to limiting the Lipschitz constant (LC) of CNN to be less than one [39]. Generally, if the LC of every module is upper-bounded, then the LC of the whole network is bounded [32]. Therefore, an intuitive but practical way to build a satisfying denoiser is to constrain the LC of modules in the existing effective and efficient CNN architectures, such as DnCNN [40]. For DnCNN architecture [40] with its batch normalization (BN) layers [41] removed, Ryu *et al.* [32] developed a technique called real spectral normalization (SN) to constrain the LC of the convolutional layers, thereby constraining the LC of the full network. Sun *et al.* [33] used the SN technique in [42] to control the global LC. It is worth noting that in [32], although not theoretically exploring the impact of BN layers [41] on global LC, the authors experimentally found that the full DnCNN regularized by real SN (RealSN-DnCNN) satisfies the global LC condition. In fact, Santurkar *et al.* [43] theoretically proved that inserting the BN layer [41] can reduce the LC significantly. Therefore, we use the RealSN-DnCNN with LC no more than one for effective denoiser prior modeling, which is illustrated in Fig. 1. It is the same as DnCNN-B [40] except that the convolutional layers are replaced by real SN-regularized convolutional layers.

2) *Training Set*: A good training set is key to train an effective CNN denoiser. For seismic deblending, the standard practice is to collect large amounts of appropriate field data [23] or to construct a synthetic dataset [24], [25], but neither is easy to implement. To address this problem, we explore the acquired blended CSGs and use them to construct the training samples. This means that no extra data are required. Considering our deblending process is implemented in common receiver domain, it is really necessary to analyze and probe into the rationality of using blended CSGs as training dataset from two aspects:

- 1) The relationship between CRG and CSG. According to spatial reciprocity of Green's function under certain initial and boundary conditions [44]–[46]

$$G_1(\mathbf{L}_r, t; \mathbf{L}_s) = G_2(\mathbf{L}_s, t; \mathbf{L}_r) \quad (18)$$

where $G_1(G_2)$ is Green's function which describes wave propagation from source location $\mathbf{L}_s$ (receiver location $\mathbf{L}_r$) to receiver location $\mathbf{L}_r$ (source location $\mathbf{L}_s$). t denotes the travel time. It can be seen from (18) that the CSGs have the same features the CRGs have under certain conditions. Even if the conditions are not met in field situations, they could have similar features [44]–[46];

- 2) The relationship between blended CSGs and unblended CSGs. In general, the underground structures in the same survey region are similar [38], which means that the features of different unblended CSGs are similar. Blended CSGs are the superposition of different unblended CSGs and present coherent status. Thus, the information of unblended CSGs is contained. Based on the above analysis, the unblended CRGs, which we aim to recover, have extremely similar features to the blended CSGs. Therefore, it is reasonable to use blended CSGs as the training dataset.

3) *Training Details*: The blended CSGs are used as the clean data $\mathbf{v}$. In order to simulate noisy data $\mathbf{y}$, we follow the common setting for Gaussian denoising and add additive white Gaussian noise (AWGN) $\mathbf{e}$ with noise standard deviation σ to clean data $\mathbf{v}$. From (16), we can see that σ needs to be equal to $\sqrt{\lambda/\rho}$, which is hard to determine in practice. A common practice is to set a range for σ [29], [30]. However, we empirically found that $\sigma = 5$ would get better performance. Instead of l_2 loss, the loss function is defined as

$$\text{loss} = \frac{1}{2N} \sum_{i=1}^N \|\mathcal{R}(\mathbf{y}_i; \Theta) - (\mathbf{y}_i - \mathbf{v}_i)\|_1 \quad (19)$$

where $\mathcal{R}(\mathbf{y}) \approx \mathbf{e}$ denotes the residual mapping and Θ is the trainable parameters in RealSN-DnCNN. $\{(\mathbf{y}_i, \mathbf{v}_i)\}_{i=1}^N$ is N noisy-clean training patch pairs and, here, we set the size as 50×50 . We train the network using the ADAM algorithm for 80 epochs, with a batch size of 128. The learning rate is 0.001 in the first 50 epochs, and then decreased to 0.0001.

D. Summary and Comparison

Based on the above analysis, the presented PnP-ADMM method with blended CSGs trained deep CNN Gaussian denoiser for seismic deblending is summarized in Algorithm 1. It is worth mentioning that the strategy of the existing deep learning-based iterative methods for seismic deblending is to train a discriminative deblending CNN which uses pseudo-deblended data and conventional unblended data as “noisy-clean” training pairs and embed it into an iterative framework [24], [25]. However, our method aims to train an effective deep CNN Gaussian denoiser and plug it into the ADMM algorithm as the regularization. Besides, the goal of using blended CSGs as the training dataset is only to obtain a highly effective Gaussian denoiser. In fact, as demonstrated in experiment part, any effective Gaussian denoiser can be used as the prior to achieve relatively good deblending results.

III. EXPERIMENTS

In this section, we designed a series of experiments on synthetic and field datasets to explore: 1) the superiority of

Algorithm 1 PnP-ADMM With Blended CSGs Trained Deep CNN Gaussian Denoiser for Seismic Deblending

1. **Input:** Trained RealSN-DnCNN prior model, the observed blended data $\mathbf{b}$, the blending matrix $\mathbf{\Gamma}$, and the iteration number K .
2. **Initialization:** Initialize $\mathbf{d}$ as $\mathbf{d}^0 = \mathbf{\Gamma}^T \mathbf{b}$, $\mathbf{u}$ as $\mathbf{u}^0 = \mathbf{0}$
3. **For** $k = 1, 2, \dots, K$ **do**
4. Update $\mathbf{x}^k$ by (13) using RealSN-DnCNN prior model.
5. Update $\mathbf{d}^k$ by (14) using CG method.
6. Update $\mathbf{u}^k$ by (15).
7. **end**
8. **Output:** The deblending result $\mathbf{x}^K$.

TABLE I
USED METHODS FOR SEISMIC DEBLENDING

Synthetic and Field Cases		
Method	Network Architecture	Training set
PnP-ADMM with blended CSGs trained denoiser	RealSN-DnCNN	Blended CSGs
PnP-ADMM with open dataset trained denoiser		Open Dataset
Blended CSGs trained deblending CNN	DnCNN-B [41]	Blended CSGs
Open dataset trained deblending CNN		Open Dataset
Dictionary learning based deblending method [17]	—	—

using blended CSGs as the training dataset; 2) the mechanism of the trained deep CNN Gaussian denoiser in the PnP framework; and 3) the overall performance of the presented method.

A. Comparative Methods and Quantitative Evaluation

1) *Comparative Methods:* As comparative methods, as shown in Table I, both deep learning-based deblending methods and dictionary learning-based deblending method [17] are considered. For deep learning-based methods, as mentioned earlier, the performances of the deblending effect largely depend on the training dataset. Thus, to verify the superiority of using blended CSGs as the training dataset, we used an open dataset from Society of Exploration Geophysicists (SEG) as the competing dataset, as shown in Fig. 2. Consequently, besides our presented PnP-ADMM with blended CSGs trained denoiser, we also constructed the PnP-ADMM with open dataset trained denoiser, which used open dataset as the training dataset. Furthermore, following the same training strategy in [24] and [25] which utilizes the pseudo-deblended CRGs and conventional CRGs as “noisy-clean” training pairs, we also constructed two discriminative deblending CNNs. The network architecture is DnCNN-B [40], and the blended CSGs (blended CSGs trained deblending CNN) or open dataset (open dataset trained deblending CNN) was regarded as the conventional CRGs.

2) *Quantitative Evaluation:* Because the conventional unblended synthetic and field datasets are obtainable, we use the signal-to-noise ratio (SNR) to assess the recovery

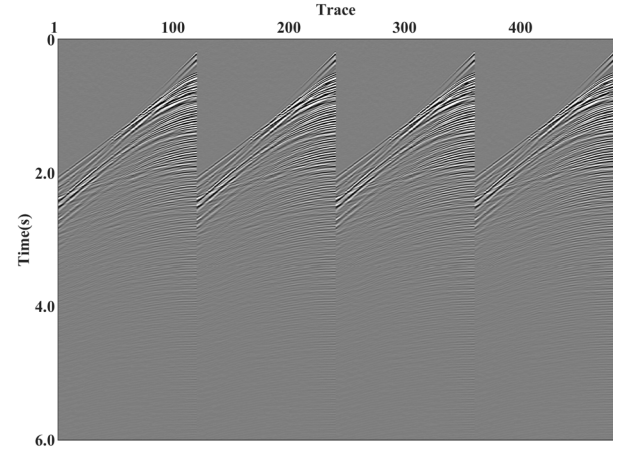


Fig. 2. Used open dataset from the Society of Exploration Geophysicists (SEG).

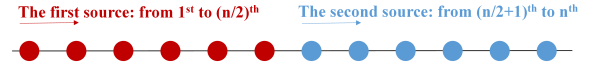


Fig. 3. Diagram of blending type for synthetic and the simple field dataset.

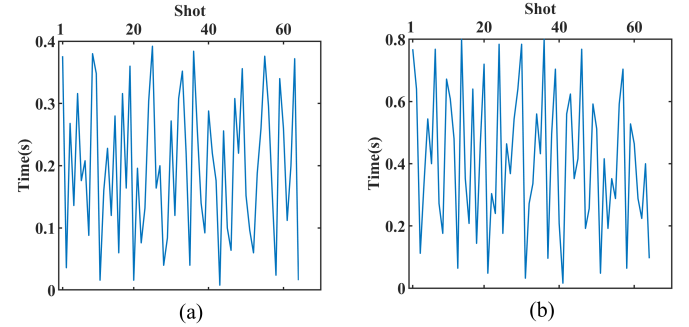


Fig. 4. Random dithering codes. (a) Random dithering code of synthetic dataset. (b) Random dithering code of the simple field dataset.

performance quantitatively

$$\text{SNR}(\mathbf{d}, \hat{\mathbf{d}}) = 10 \log_{10} \frac{\|\mathbf{d}\|_2^2}{\|\hat{\mathbf{d}} - \mathbf{d}\|_2^2} \quad (20)$$

where $\hat{\mathbf{d}}$ is the deblending result or Gaussian denoising result, and $\mathbf{d}$ is the conventional unblended CRG.

B. Synthetic and Simple Field Data Examples

1) *Numerical Simulation of Blended Data:* For both the synthetic and the simple field dataset, the STR is set as 2. The blending type is that two independent shooting sources move in the same direction and shoot successively with the given random dithering code, which is schematically shown in Fig. 3. In this case, the blended CSGs are the results of paired dislocation overlap of shots produced by different sources. Besides, it should be noted that the random dithering code is always set to be greater than zero to ensure that the first source always fires before the second.

The synthetic dataset includes 128 shots and 128 receivers. The spatial and temporal sampling intervals are 12 m and 2 ms,

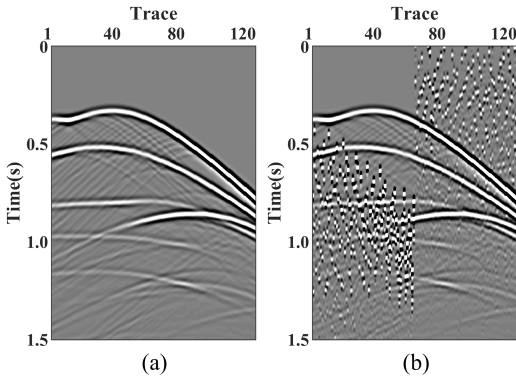


Fig. 5. Synthetic data. (a) 32nd unblended CRG. (b) Pseudo-deblended data of the 32nd unblended CRG.

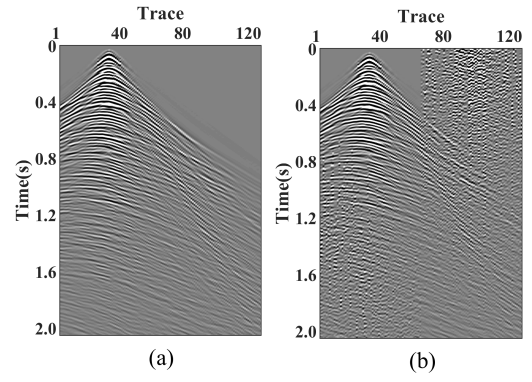


Fig. 7. Simple field data. (a) 32nd unblended CRG. (b) Pseudo-deblended data of the 32nd unblended CRG.

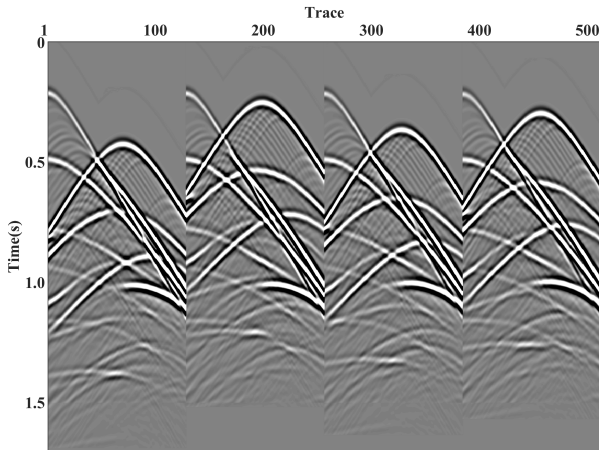


Fig. 6. Some of the blended CSGs in synthetic data example.

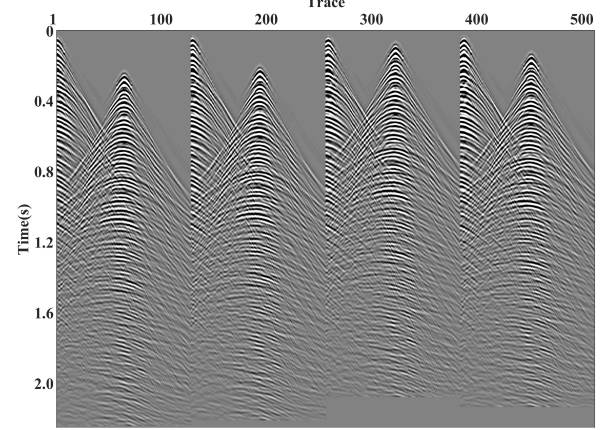


Fig. 8. Some of the blended CSGs in the simple field data example.

respectively. The number of time samples is 750. The random dithering code is predefined as $[0, 0.4 \text{ s}]$, which is shown in Fig. 4(a). The 32nd CRG prior to blending, and its pseudo-deblended data are shown in Fig. 5(a) and (b), respectively. Fig. 6 displays some of the blended CSGs, and it can be seen that they are coherent and have similar features to the conventional unblended CRG in Fig. 5(a).

The field dataset has 128 shots, 128 receivers, and 512 time samples. The spatial and temporal sampling intervals are 12.5 m and 4 ms, respectively. Similar to the synthetic data example, the random time delay lies in the time range $[0, 0.8 \text{ s}]$, as shown in Fig. 4(b). The 32nd CRG prior to blending and its pseudo-deblended data are shown in Fig. 7(a) and (b), respectively. Some of the blended CSGs are shown in Fig. 8.

2) *Experimental Results:* First, we tested the performances of different trained deep CNN Gaussian denoisers on Gaussian noise suppression. Table II summarizes the average SNR results of different denoisers on the synthetic and the simple field data for Gaussian noise levels 5. It can be seen that under the same conditions, using blended CSGs as training dataset achieves better average SNR results than using open dataset as training dataset, which illustrates that the blended CSGs have more similar features to the conventional unblended CRGs.

Then, Fig. 9(a) and (b) shows the SNR results of different methods in Table I on all the synthetic and the simple field CRGs, respectively. We can observe that whether in the

TABLE II

AVERAGE SNR (dB) RESULTS OF DIFFERENT DENOISERS FOR NOISE LEVEL 5 ON SYNTHETIC AND THE SIMPLE FIELD DATASET

Network	Noise Level	Synthetic Data		Simple Field Data	
		Blended CSGs	Open Datasets	Blended CSGs	Open Datasets
		SNR			
RealSN-DnCNN	$\sigma = 5$	27.88	20.96	18.18	16.65

synthetic data example or in the simple field data example, our method can achieve better SNR results across all the CRGs in contrast to other competing methods. For PnP-ADMM with open dataset trained denoiser, its performance is only marginally inferior to that of our method, demonstrating that, on the one hand, the PnP-ADMM framework is extremely effective due to the interaction between the data fidelity term and the trained deep CNN Gaussian denoiser prior. Using blended CSGs as the training dataset, on the other hand, can learn the features of conventional unblended CRGs to the greatest extent. For blended CSGs trained deblending CNN, it can get higher SNR results across almost all the CRGs compared with open dataset trained deblending CNN and dictionary learning-based deblending method, which also verifies the superiority of using blended CSGs as training dataset. Dictionary learning-based deblending method gets better performance than open dataset trained deblending CNN

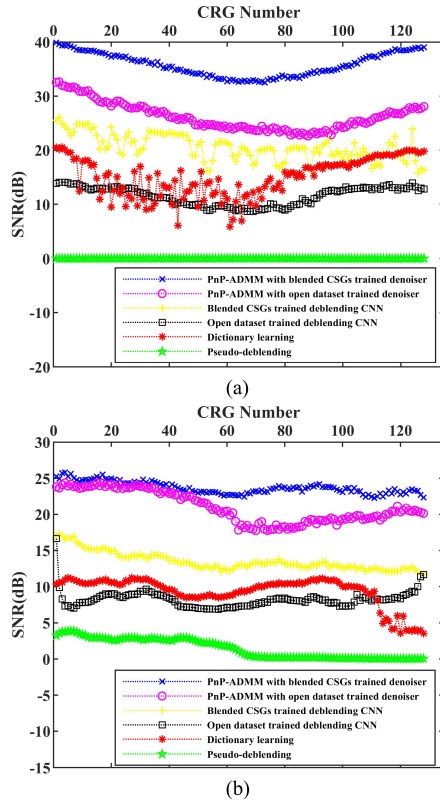


Fig. 9. SNR results of different methods in Table I on all the synthetic and the simple field CRGs. (a) Synthetic CRGs. (b) Simple field CRGs.

for most of CRGs in both synthetic and the simple field data examples. However, for some CRGs, its effect is poor.

Furthermore, to evaluate the deblending performance in more detail, the 32nd CRGs of the synthetic and the simple field dataset are used for quantitative and qualitative evaluation. Fig. 10(a)–(e) shows the deblending results of different methods on the 32nd synthetic CRG, and the corresponding errors are shown in Fig. 11(a)–(e), respectively. As shown in Fig. 10(a), our method effectively recovered the useful seismic data, and the corresponding error in Fig. 11(a) is small enough to be negligible. For PnP-ADMM with open dataset trained denoiser, although there is almost no damage to the useful signal in Fig. 10(b), we can still see the signal leakage in Fig. 11(b). From Fig. 10(c)–(e) to Fig. 11(c)–(e), it can be seen that discriminative deblending CNNs and dictionary learning-based deblending method cause different degrees of blending noise residual and signal leakage. The deblending results of different methods on the 32nd CRG of the simple field dataset are shown in Fig. 12(a)–(e), respectively. Fig. 13(a)–(e) shows the corresponding errors. As shown in Figs. 12(a) and 13(a), our method is capable of finely removing blending noise while also better preserving structure and details, resulting in better recovery results. Because the Gaussian noise suppression performances of blended CSGs trained denoiser and open dataset trained denoiser are similar, the final deblending effect of PnP-ADMM with open dataset trained denoiser is comparable with that of our method. From Fig. 12(b), we barely see the blending noise and there is almost no useful signal residue in Fig. 13(b). And from Fig. 12(c)–(e) to Fig. 13(c)–(e),

TABLE III
SNR (dB) RESULTS OF DIFFERENT METHODS ON THE 32ND CRGs OF SYNTHETIC AND THE SIMPLE FIELD DATASETS

Method	SNR (dB)	
	The 32 nd CRG of synthetic dataset	The 32 nd CRG of the simple field dataset
PnP-ADMM with blended CSGs trained denoiser	36.60	24.46
PnP-ADMM with open dataset trained denoiser	27.71	23.75
Blended CSGs trained deblending CNN	22.18	13.98
Open dataset trained deblending CNN	12.04	9.64
Dictionary learning based deblending method [17]	9.01	11.11
Pseudo-deblending	0.02	2.72

we can easily observe that discriminative deblending CNNs and dictionary learning-based deblending method not only failed to remove the blending noise but also caused to the useful signal degradation, especially the weak details in deep areas. Furthermore, the SNR results of different methods on the 32nd CRGs of synthetic and the simple field datasets are shown in Table III. We can see that the SNR results consistent with qualitative evaluation, and our method achieved much higher SNR than other comparative methods.

In summary, the utilization of blended CSGs as training dataset avoids inadequate capability of generalization and thus promotes the Gaussian and blending noise suppression ability of the CNN denoisers. And with the help of the interaction of data fidelity term and trained deep CNN Gaussian denoiser prior, the PnP-ADMM method draws on the advantages of model- and data-driven methods and surpasses other comparative methods.

3) *Analysis on the Mechanism of the Trained Deep CNN Gaussian Denoiser in the PnP Framework:* As mentioned above, from the Bayesian point of view, a Gaussian denoiser should be integrated into the PnP framework. However, considering that the input of the CNN Gaussian denoiser is initialized as pseudo-deblended data, and the existing blending noise is generally non-Gaussian distributed, one may raise doubts about the effectiveness of such a Gaussian denoiser in suppressing the non-Gaussian distributed noise. Actually, according to the existing literature, the difference between the input of the denoiser and the ideal useful signal is not required to strictly obey Gaussian distribution [29], [30], [34]. Furthermore, in order to deeply explore the mechanism of the trained CNN Gaussian denoiser in the PnP framework, we show the inputs ($\mathbf{d}^{(k)} - \mathbf{u}^{(k)}$), the outputs ($\mathbf{x}^{(k)}$) of the trained CNN Gaussian denoiser, and their differences ($\mathbf{d}^{(k)} - \mathbf{u}^{(k)} - \mathbf{x}^{(k)}$) at different iterations of the presented method on the 32nd CRGs of the synthetic and the simple field dataset in Figs. 14 and 15, respectively. It can be seen that in both cases, at the first few iterations, there is severe non-Gaussian distributed noise in the inputs of the denoiser due to the introduction of the blending noise at the initialization stage. At every iteration, the denoiser only suppresses a portion of the noise along with some useful

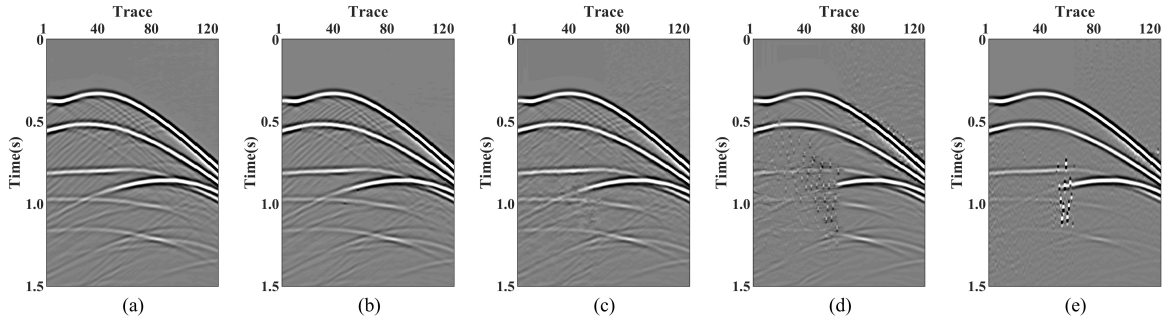


Fig. 10. Deblending results of different methods on the 32nd CRG of the synthetic dataset. (a) PnP-ADMM with blended CSGs trained denoiser. (b) PnP-ADMM with open dataset trained denoiser. (c) Blended CSGs trained deblending CNN. (d) Open dataset trained deblending CNN. (e) Dictionary learning-based deblending method [17].

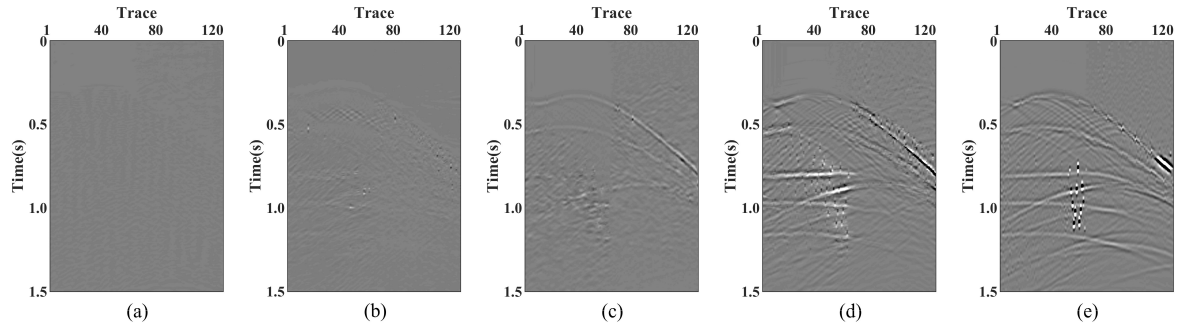


Fig. 11. Corresponding errors of different methods on the 32nd CRG of the synthetic dataset. (a) PnP-ADMM with blended CSGs trained denoiser. (b) PnP-ADMM with open dataset trained denoiser. (c) Blended CSGs trained deblending CNN. (d) Open dataset trained deblending CNN. (e) Dictionary learning-based deblending method [17].

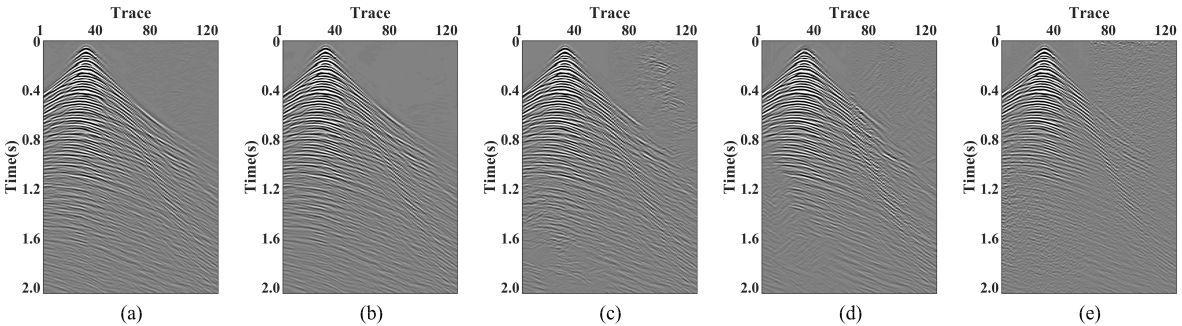


Fig. 12. Deblending results of different methods on the 32nd CRG of the simple field dataset. (a) PnP-ADMM with blended CSGs trained denoiser. (b) PnP-ADMM with open dataset trained denoiser. (c) Blended CSGs trained deblending CNN. (d) Open dataset trained deblending CNN. (e) Dictionary learning-based deblending method [17].

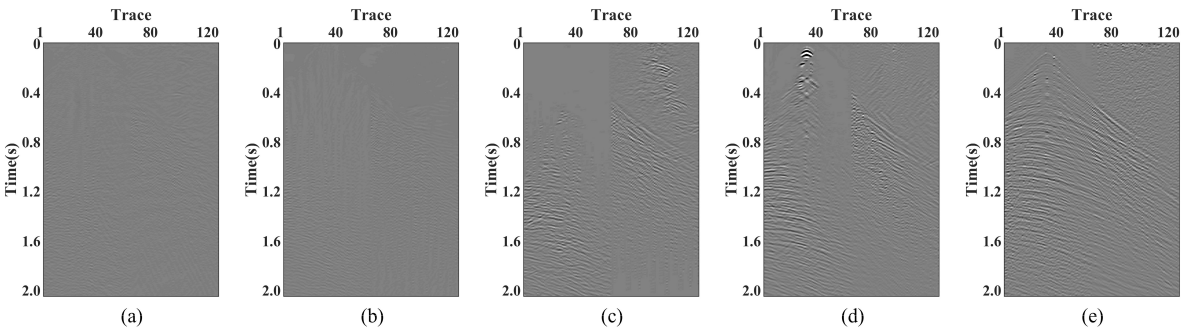


Fig. 13. Corresponding errors of different methods on the 32nd CRG of the simple field dataset. (a) PnP-ADMM with blended CSGs trained denoiser. (b) PnP-ADMM with open dataset trained denoiser. (c) Blended CSGs trained deblending CNN. (d) Open dataset trained deblending CNN. (e) Dictionary learning-based deblending method [17].

details, and the lost details along with some fake structures are introduced by subsequent subproblem. With the increasing of iterations, the noise in the inputs of the denoiser becomes rare.

Furthermore, we use the Kurtosis to measure if the distributions of the intermediate results are close enough to Gaussian distribution. The Kurtosis of Gaussian distribution is equal to

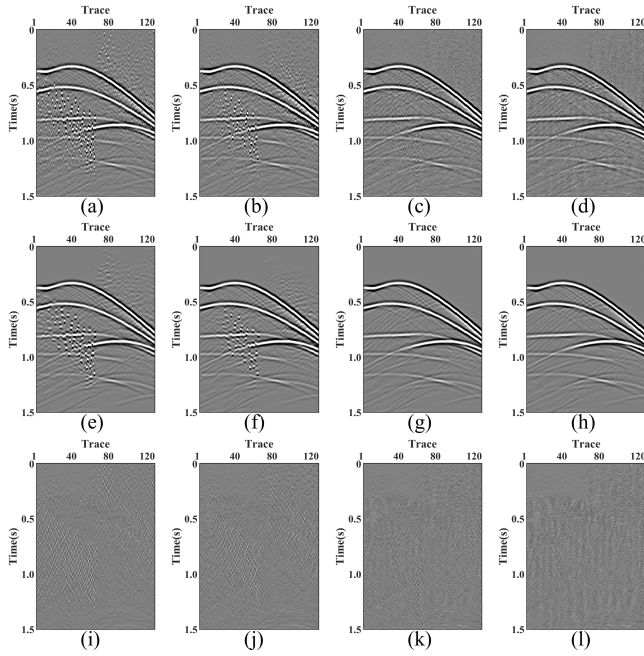


Fig. 14. Intermediate results of PnP-ADMM with blended CSGs trained denoiser at different iterations on the 32nd CRG of the synthetic dataset. (a)–(d) Inputs of the denoiser at the 4th, 8th, 18th, and 30th iterations. (e)–(h) Outputs of the denoiser at the 4th, 8th, 18th, and 30th iterations. (i)–(l) Differences between the input and the output at the 4th, 8th, 18th, and 30th iterations.

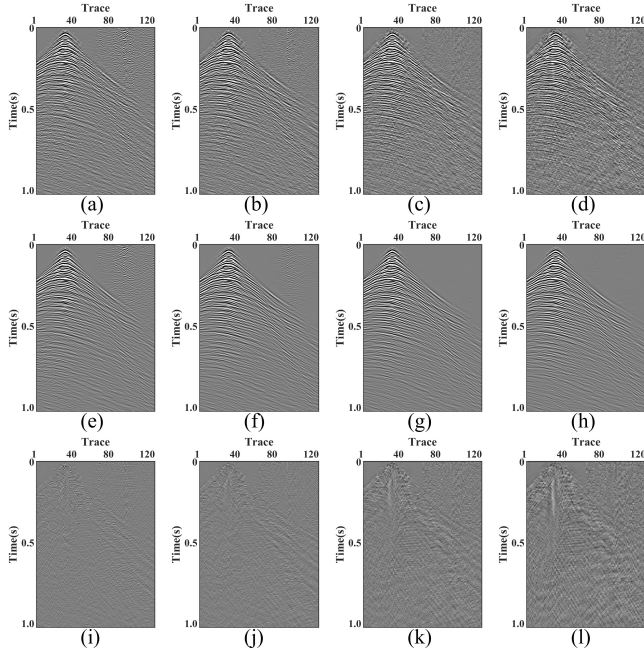


Fig. 15. Intermediate results of PnP-ADMM with blended CSGs trained denoiser at different iterations on the 32nd CRG of the simple field dataset. (a)–(d) Inputs of the denoiser at the 4th, 8th, 18th, and 30th iterations. (e)–(h) Outputs of the denoiser at the 4th, 8th, 18th, and 30th iterations. (i)–(l) Differences between the input and the output at the 4th, 8th, 18th, and 30th iterations.

zero. And the distribution with a higher kurtosis has a heavier tail. The Kurtosis values of: 1) the differences between the unblended conventional CRG and the inputs of the denoiser and 2) the differences between the inputs and the outputs of the denoiser, at different iterations on the 32nd CRGs

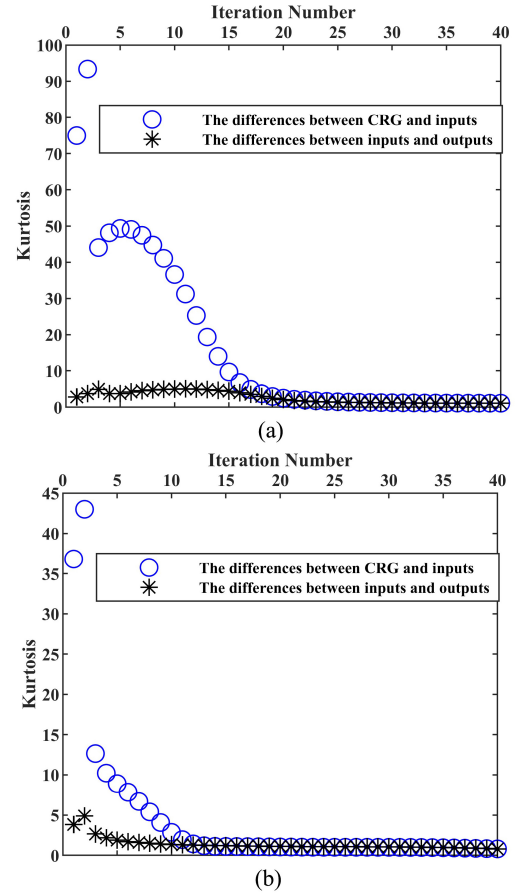


Fig. 16. Kurtosis values of: 1) the differences between the unblended conventional CRG and the inputs of the denoiser and 2) the differences between the unblended conventional CRG and the inputs of the denoiser, at different iterations. (a) 32nd CRG of synthetic dataset. (b) 32nd CRG of the simple field dataset.

of the synthetic and the simple field dataset, are shown in Fig. 16(a) and (b), respectively. We can observe that in these two examples, the Kurtosis values of the differences between the unblended conventional CRG and the inputs of the denoiser gradually tend to zero as the iteration process goes, which means that the differences are more and more Gaussian-like with the interaction of the subproblems in PnP-ADMM framework, and more suitable to be processed by CNN Gaussian denoiser. Besides, it can be seen that the Kurtosis values of the differences between the inputs and the outputs of the denoiser at different iterations are similar and all larger than zero, which implies that, on the one hand, the performance of the denoiser is stable. On the other hand, although the denoiser is trained for denoising Gaussian noise, the “noise” it removes is not exactly Gaussian distributed but Gaussian-like distributed.

C. More Challenging Field Data Example

In order to further test and verify the validity of the presented method, a more complicated blending scenario and a more complex field dataset are adopted. Specifically, the STR is increased from 2 to 3, and the blending type is that three independent shooting sources move in the same direction and shoot successively with the given random dithering codes in Fig. 17(a) and (b). The used field dataset contains 90 shots,

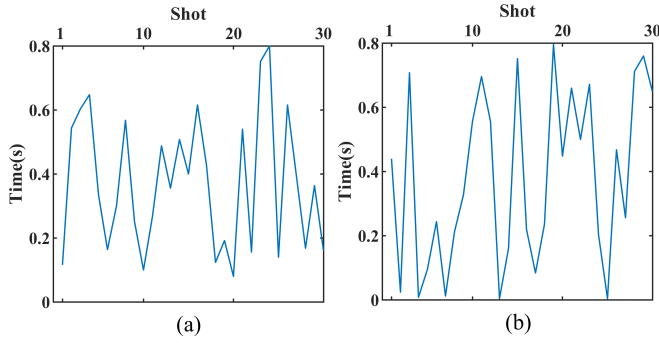


Fig. 17. Random dithering codes for more challenging field dataset. (a) Random dithering code of the second shooting source. (b) Random dithering code of the third shooting source.

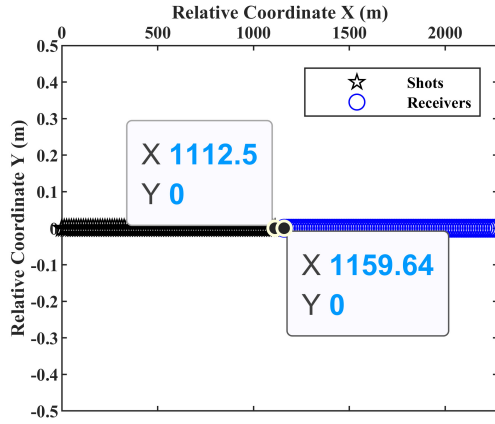


Fig. 18. Acquisition geometry of more challenging field dataset.

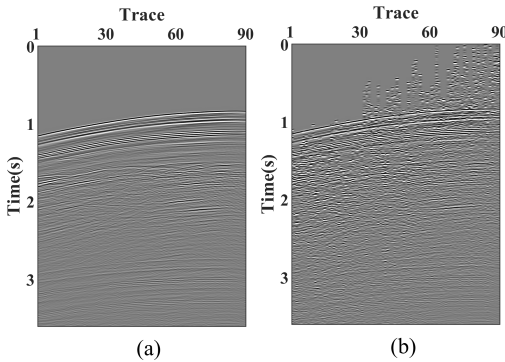


Fig. 19. More challenging field dataset. (a) 4th unblended CRG. (b) Pseudo-deblended data of the 4th unblended CRG.

90 receivers, and 900 time samples. The temporal sampling interval is 4 ms, and the spatial interval is 12.5 m. It should be noted that, as shown in Fig. 18, the acquisition geometry of this dataset is that the shots and the receivers are on separate sides, and the minimum offset is about 50 m. This kind of geometry cannot meet the conditions for (18); thus, the CSGs and the CRGs have similar rather than the same features.

The 4th CRG prior to blending is shown in Fig. 19(a). We can see that it has extremely complex subsurface structures with very weak events especially in deep area. Its corresponding pseudo-deblended data shown in Fig. 19(b) have

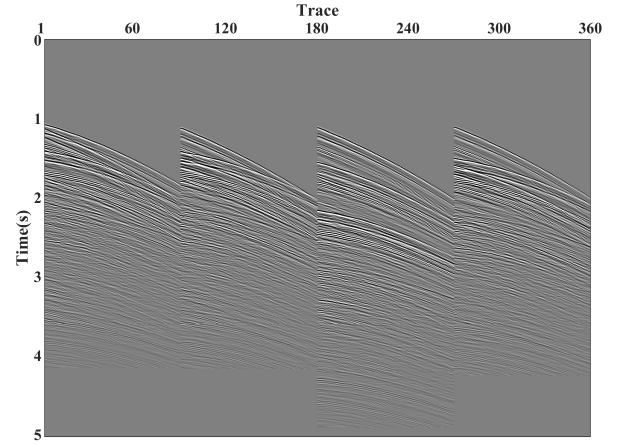


Fig. 20. Some blended CSGs in more challenging field data example.

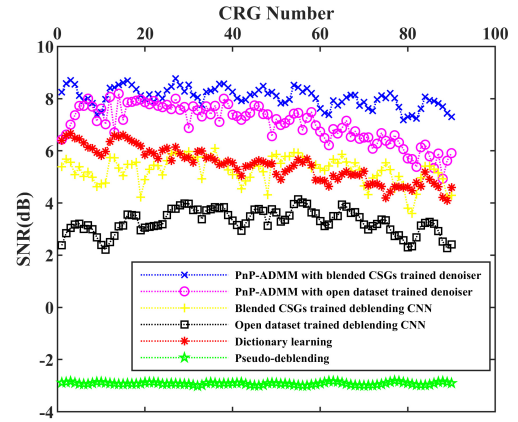


Fig. 21. SNR results of different methods in Table I on all the CRGs in more challenging field dataset.

the SNR value as -2.89 dB, and it can be seen that the blending noise severely masks the useful signals. In particular, some weak events are hardly visible, thereby greatly increasing the deblending difficulty. Fig. 20 shows some of the blended CSGs. The SNR results of different methods in Table I on all the CRGs in this field dataset are plotted in Fig. 21. Notice that the presented method can still achieve the best SNR results across all the CRGs compared with other competing methods. However, it did not achieve an overwhelming advantage as in the previous examples. There are three possible main reasons. First, the CRGs to be recovered have more complex events with diverse amplitudes. Second, the blending scenario is more complicated. Higher STR leads to fewer observations, which means that whether for the denoiser training or the final inverse problem optimization, far less information is available. Third, perhaps most importantly, the acquisition geometry dictates that the CSGs and CRGs are not that similar, thus using corresponding blended CSGs for the denoiser training may be unable to learn some features of the CRGs, resulting in final weak useful signals leakage. Considering the methods which use the open dataset as the training dataset, although the end-to-end deblending CNN acquired certain effect, it is the least effective among the five methods, which implies that

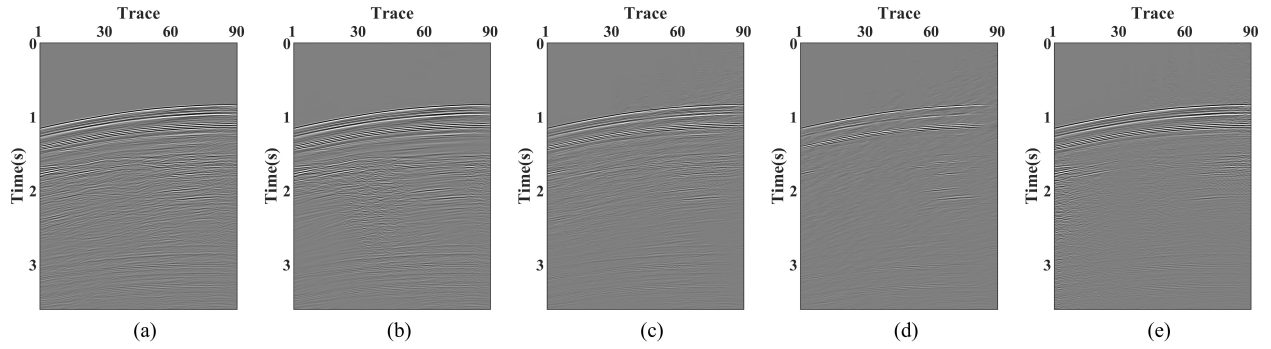


Fig. 22. Deblending results of different methods on the 4th CRG in more challenging field data example. (a) PnP-ADMM with blended CSGs trained denoiser. (b) PnP-ADMM with open dataset trained denoiser. (c) Blended CSGs trained deblending CNN. (d) Open dataset trained deblending CNN. (e) Dictionary learning based deblending method [17].

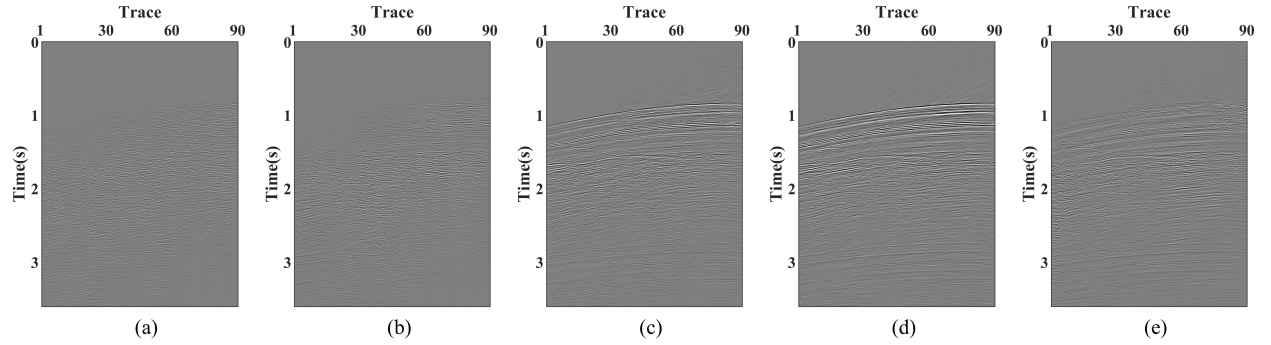


Fig. 23. Corresponding errors of different methods on the 4th CRG in more challenging field data example. (a) PnP-ADMM with blended CSGs trained denoiser. (b) PnP-ADMM with open dataset trained denoiser. (c) Blended CSGs trained deblending CNN. (d) Open dataset trained deblending CNN. (e) Dictionary learning-based deblending method [17].

the used open dataset only contains a small number of features of the CRGs to be recovered. Nevertheless, thanks to the interplay between the data fidelity term and the denoiser prior, the PnP-ADMM with open dataset trained denoiser is second only to the presented method in terms of the SNR results, which illustrates the excellence of the PnP-ADMM framework. The blended CSGs trained deblending CNN acquired better SNR results than the open dataset trained deblending CNN, which means that compared with the chosen open dataset, the blended CSGs are much more similar to the CRGs to be recovered. The dictionary learning-based deblending method [17] provides comparable performance to the blended CSGs trained deblending CNN.

In addition, Fig. 22(a)–(e) shows the deblending results of different methods on the 4th CRG in this field dataset, and the corresponding SNR values are 8.55, 7.37, 5.09, 3.18, and 6.49 dB, respectively. The corresponding errors are shown in Fig. 23(a)–(e). Our method has the highest SNR value. And consistent with SNR results, it can be seen that there is the least blending noise in Fig. 22(a) and the least signal leakage in Fig. 23(a), which means that our method outperforms other comparative methods.

IV. CONCLUSION AND FUTURE PROSPECTS

We developed a PnP method that combines the merits of model- and data-driven methods for seismic deblending.

Specifically, by treating the deblending process as an inversion problem, we utilized the ADMM algorithm for optimization, in which the proximal operator containing the regularization term can be replaced by a trained deep CNN Gaussian denoiser. In order to train an appropriate and effective CNN denoiser, we propose to use acquired blended CSGs as the training dataset to avoid constructing or collecting any extra data. In addition, we utilize a firmly nonexpansive CNN architecture to ensure theoretical convergence of the whole algorithm. Extensive experiments and analysis on synthetic and field data show the rationality of using blended CSGs as training dataset and the superior deblending performance of our presented method over the dictionary learning based deblending method and discriminative deblending CNNs.

It can be seen that the seismic acquisition geometry largely decides the similarity between CSGs and CRGs. And the firing type of shots greatly determines the difficulty of the subsequent deblending process. Therefore, considering the fact in field and designing an effective seismic survey for blended acquisition is the focus of future research, also a key to the effective application of the presented method.

It should also be mentioned that there are also other types of CNN architectures that have shown their great performances, such as the U-net [47]. Therefore, the future work may focus on designing a CNN Gaussian denoiser which can yield more excellent denoising performance and keep firmly nonexpansive

at the same time. In addition, besides the CNN denoiser prior, adding some traditional priors, such as sparsity, may achieve better performance.

ACKNOWLEDGMENT

The authors would like to thank Dr. Vincenzo Lipari, Prof. Paolo Bestagini, and Prof. Stefano Tubaro in the Image and Sound Processing Laboratory (ISPL), Politecnico di Milan, Milan, Italy, for generously providing the test data, and their constructive discussion and revision. They would also like to thank the editors and reviewers for their critical and constructive comments to significantly improve the quality of this article.

REFERENCES

- [1] A. J. Berkhou, "Changing the mindset in seismic data acquisition," *Lead. Edge*, vol. 27, pp. 924–938, Jul. 2008.
- [2] A. J. Berkhou, G. Blacquière, and E. Verschuur, "From simultaneous shooting to blended acquisition," in *Proc. SEG Tech. Program Expanded Abstr.*, Jan. 2008, pp. 2831–2838.
- [3] D. J. Verschuur and A. J. Berkhou, "Seismic migration of blended shot records with surface-related multiple scattering," *Geophysics*, vol. 76, no. 1, pp. A7–A13, Jan. 2011.
- [4] Y. Tang and B. Biondi, "Least-squares migration/inversion of blended data," in *Proc. SEG Tech. Program Expanded Abstr.*, Jan. 2009, pp. 2859–2863.
- [5] W. Dai, X. Wang, and G. T. Schuster, "Least-squares migration of multisource data with a deblurring filter," *Geophysics*, vol. 76, no. 5, pp. R135–R146, Sep. 2011.
- [6] Y. Chen, H. Chen, K. Xiang, and X. Chen, "Preserving the discontinuities in least-squares reverse time migration of simultaneous-source data," *Geophysics*, vol. 82, no. 3, pp. S185–S196, May 2017.
- [7] J. Cao and G. Blacquière, "Joint migration inversion for blended dispersed source array data," in *Proc. SEG Tech. Program Expanded Abstr.*, Sep. 2020, pp. 3088–3092.
- [8] S. Huo, Y. Luo, and P. G. Kelamis, "Simultaneous sources separation via multidirectional vector-median filtering," *Geophysics*, vol. 77, no. 4, pp. V123–V131, Jul. 2012.
- [9] Y. Chen, "Deblending using a space-varying median filter," *Explor. Geophys.*, vol. 46, no. 4, pp. 332–341, Dec. 2015.
- [10] W. Huang, R. Wang, X. Gong, and Y. Chen, "Iterative deblending of simultaneous-source seismic data with structuring median constraint," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 1, pp. 58–62, Jan. 2018.
- [11] J. Cheng and M. D. Sacchi, "Separation and reconstruction of simultaneous source data via iterative rank reduction," *Geophysics*, vol. 80, no. 4, pp. V57–V66, Jul. 2015.
- [12] H. Zhou *et al.*, "Deblending of simultaneous source with rank-reduction and thresholding constraints," in *Proc. 79th EAGE Conf. Exhib.*, Jun. 2017, pp. 1–5.
- [13] A. Mahdad, P. Doulgeris, and G. Blacquière, "Separation of blended data by iterative estimation and subtraction of blending interference noise," *Geophysics*, vol. 76, no. 3, pp. Q9–Q17, May 2011.
- [14] C. Li, C. C. Mosher, and Y. Ji, "An amplitude-preserving deblending approach for simultaneous sources," *Geophysics*, vol. 84, no. 3, pp. V185–V196, May 2019.
- [15] H. Mansour, H. Wason, T. Lin, and F. J. Herrmann, "Randomized marine acquisition with compressive sampling matrices," *Geophys. Prospecting*, vol. 60, no. 4, pp. 648–662, 2012.
- [16] Y. Zhou, W. Chen, and J. Gao, "Separation of seismic blended data by sparse inversion over dictionary learning," *J. Appl. Geophys.*, vol. 106, no. 7, pp. 146–153, 2014.
- [17] Y. Zhou, J. Gao, W. Chen, and P. Frossard, "Seismic simultaneous source separation via patchwise sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5271–5284, Sep. 2016.
- [18] S. Zu, H. Zhou, R. Wu, W. Mao, and Y. Chen, "Hybrid-sparsity constrained dictionary learning for iterative deblending of extremely noisy simultaneous-source data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2249–2262, Apr. 2019.
- [19] J. Cheng, K. Chen, and M. D. Sacchi, "Application of robust principal component analysis (RPCA) to suppress erratic noise in seismic records," in *Proc. SEG Tech. Program Expanded Abstr.*, Aug. 2015, pp. 4646–4651.
- [20] A. Ibrahim and M. D. Sacchi, "Simultaneous source separation using a robust radon transform," *Geophysics*, vol. 79, no. 1, pp. V1–V11, Jan. 2014.
- [21] Q. Zhao, Q. Du, X. Gong, X. Li, L. Fu, and Y. Chen, "Normalized shap-ing regularization for robust separation of blended data," *Geophysics*, vol. 84, no. 5, pp. V281–V293, Sep. 2019.
- [22] B. Wang, J. Wang, J. Li, and Z. Guo, "Iterative accurate seismic data deblending by ASB-based robust sparse radon transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–9, 2022, doi: [10.1109/TGRS.2021.3068795](https://doi.org/10.1109/TGRS.2021.3068795).
- [23] J. Sun, S. Slang, T. Elboth, T. L. Greiner, S. McDonald, and L.-J. Gelius, "A convolutional neural network approach to deblending seismic data," *Geophysics*, vol. 85, no. 4, pp. WA13–WA26, Jul. 2020.
- [24] S. Zu, J. Cao, S. Qu, and Y. Chen, "Iterative deblending for simultaneous source data using the deep neural network," *Geophysics*, vol. 85, no. 2, pp. V131–V141, Mar. 2020.
- [25] B. Wang, J. Li, J. Luo, Y. Wang, and J. Geng, "Intelligent deblending of seismic data based on U-Net and transfer learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8885–8894, Oct. 2021, doi: [10.1109/TGRS.2020.3048746](https://doi.org/10.1109/TGRS.2020.3048746).
- [26] S. Nakayama and G. Blacquière, "Machine-learning-based data recovery and its contribution to seismic acquisition: Simultaneous application of deblending, trace reconstruction, and low-frequency extrapolation," *Geophysics*, vol. 86, no. 2, pp. P13–P24, Mar. 2021.
- [27] S. Wang, W. Hu, Y. Hu, X. Wu, and J. Chen, "A physics-augmented deep learning method for seismic data deblending," in *Proc. SEG Tech. Program Expanded Abstr.*, Sep. 2020, pp. 3877–3881.
- [28] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 945–948.
- [29] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.
- [30] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jun. 14, 2021, doi: [10.1109/TPAMI.2021.3088914](https://doi.org/10.1109/TPAMI.2021.3088914).
- [31] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play admm for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 84–98, Jan. 2017.
- [32] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2019, pp. 5546–5557.
- [33] Y. Sun, Z. Wu, X. Xu, B. Wohlberg, and U. S. Kamilov, "Scalable plug-and-play ADMM with convergence guarantees," 2020, *arXiv:2006.03224*.
- [34] H. Zhang, X. Yang, and J. Ma, "Can learning from natural image denoising be used for seismic data interpolation?" *Geophysics*, vol. 85, no. 4, pp. WA115–WA136, Jul. 2020.
- [35] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jul. 2011.
- [36] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
- [37] S. Yu, J. Ma, and W. Wang, "Deep learning for denoising," *Geophysics*, vol. 84, no. 6, pp. V333–V350, Nov. 2019.
- [38] D. Liu, W. Wang, X. Wang, C. Wang, J. Pei, and W. Chen, "Poststack seismic data denoising based on 3-D convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1598–1629, Mar. 2020.
- [39] M. Terris, A. Repetti, J.-C. Pesquet, and Y. Wiaux, "Building firmly nonexpansive convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 8658–8662.
- [40] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [41] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.
- [42] H. Sedghi, V. Gupta, and P. M. Long, "The singular values of convolutional layers," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–13.

- [43] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 2483–2493.
- [44] K. Aki and P. G. Richards, *Quantitative Seismology*, 2nd ed. Sausalito, CA, USA: University Science Books, 2002.
- [45] J. Tromp, C. Tape, and Q. Liu, "Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels," *Geophys. J. Int.*, vol. 160, no. 1, pp. 195–216, Dec. 2004.
- [46] B. Wang, N. Zhang, W. Lu, J. Geng, and X. Huang, "Intelligent missing shots' reconstruction using the spatial reciprocity of Green's function based on deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1587–1597, Mar. 2020.
- [47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assisted Intervent.*, 2015, pp. 234–241.
- [48] W. Xu, Y. Zhou, X. Wang, and W. Chen, "MoG-based robust sparse representation for seismic erratic noise suppression," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.



Weiwei Xu received the bachelor's degree in communication engineering from Sichuan University, Chengdu, China, in 2017. He is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, China.

Since November, 2021, he has been a Visiting Scholar with the Image and Sound Processing Laboratory (ISPL), Politecnico di Milan, Milan, Italy. His research interests include seismic signal processing and machine learning.



Yanhui Zhou (Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2005 and 2011, respectively.

From September 2012 to December 2013, she was a Visiting Scholar with the Signal Processing Laboratory (LTS4), Lausanne, Switzerland, and the Swiss Federal Institute of Technology (EPFL), Lausanne. She is currently an Associate Professor with the School of Information and Communication Engineering, Xi'an Jiaotong University. Her research

interests include machine learning, seismic signal processing, and ground penetrating radar data processing.



deep learning, and time–frequency analysis and their applications in the seismic data processing.

Dawei Liu received the bachelor's degree in communication engineering from Chang'an University, Xi'an, China, in 2013, and the master's degree in electronic and communication engineering from Xi'an Jiaotong University (XJTU), Xi'an, in 2018, where he is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering.

He is also a Reviewer of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. His research interests include tensor decomposition,



Xiaokai Wang (Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in communication engineering from Xi'an Jiaotong University (XJTU), Xi'an, China, in 2006 and 2012, respectively.

He was a Post-Doctoral Fellow with the Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, from 2012 to 2014. He joined the Department of Computational Geophysics, XJTU, in 2015. He was a Visiting Scholar of the Bureau of Economic Geology, The University of

Texas at Austin, Austin, TX, USA, from 2016 to 2017. Since 2018, he has been with the School of Information and Communication Engineering, XJTU, where he is currently an Associate Professor. His research interests include seismic attributes extraction, time–frequency analysis, and high-speed rail seismology.



Wenchao Chen received the B.S. and M.S. degrees in seismic exploration and information technology from Chang'an University, Xi'an, China, in 1993 and 1996, respectively, and the Ph.D. degree in electromagnetic field and microwave technology from Xi'an Jiaotong University, Xi'an, in 2000.

From 2000 to 2002, he was a Post-Doctoral Fellow with the Department of Computation Science, Northwestern Polytechnical University, Xi'an. Since 2002, he has been with the School of Information and Communications Engineering, Xi'an Jiaotong University, where he is currently a Professor. He was also a referee for several journals, including the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Geophysics*, and *Interpretation*, and *Journal of Applied Geophysics*. His research interests include seismic and GPR signal processing, blind signal processing, and sparse representation.

Dr. Chen has served as a member for SEG and the Chinese Geophysical Society.