IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 22, 2019, accepted July 29, 2019, date of publication August 1, 2019, date of current version August 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2932476

An Efficient Seismic Data Acquisition Based on Compressed Sensing Architecture With Generative Adversarial Networks

XIAOPU ZHANG[1], SHUAI ZHANG[1], JUN LIN[1,2], FENG SUN[1,2], XI ZHU[3], (Member, IEEE), YANG YANG[3], (Senior Member, IEEE), XUNQIAN TONG[1,2], AND HONGYUAN YANG[1,2]

[1]College of Instrumentation and Electrical Engineering, Jilin University, Changchun 130061, China
[2]Key Laboratory of Geophysical Exploration Equipment, Ministry of Education, Jilin University, Changchun 130061, China
[3]School of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia

Corresponding author: Xunqian Tong (txq@jlu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 41804167, in part by the National Key Research and Development Program of China under Grant 2018YFC0603204 and Grant 2018YFB1501803, and in part by the Ph.D. Interdisciplinary Research Funding Project of Jilin University under Grant 10183201836.

ABSTRACT Recently, large scale seismic data acquisition has been a critical method for scientific research and industrial production. However, due to the bottleneck on data transmission and the limitation of energy storage, it is hard to conduct large seismic data acquisition in a real-time way. So, in this paper, an efficient seismic data acquisition method, namely, compressed sensing architecture with generative adversarial networks (CSA-GAN), is proposed to tackle the two restrictions of collecting large scale seismic data. In the CSA-GAN, a data collection architecture based on compressed sensing theory is applied to reduce data traffic load of the whole system, as well as balance the data transmission. To make the compressed sensing procedure perform well in both data quality and compression ratio, a kind of generative adversarial networks is designed to learn the recovering map. According to our experiment results, a high data quality (about 30 dB) at the compression ratio of 16 is achieved by the proposed approach, which enables the system to afford 15 times more sensors and reduces the energy cost by means of data collection from $N(N+1)/2$ to $N^2/16$. These results show that the CSA-GAN can afford more sensors with the same bandwidth and consume less energy, via improving the efficiency seismic data acquisition.

INDEX TERMS Compressed sensing (CS), generative adversarial networks (GAN), dictionary learning, seismic data acquisition.

I. INTRODUCTION

Seismology has been a major geophysical tool to determine the geological structure of a given area for more than one and a half centuries. In general, seismology can be broadly classified into two categories: passive and active seismology. In passive seismology, the nature and location of an earthquake source is unknown, and the acquired data are usually used by scientists to monitor the variety of underground activities, such as some large-scale natural phenomena [1], [2], volcanic unrest [3], [4] and nuclear test [5]. In active seismology, artificial sources of seismic activity are used, and the acquired data have been thought as a solid proof to image the geological structure of a given area, which is of great interest for hydrocarbon resources exploration or civil construction [6]. As the society's continued demand for security, hydrocarbon and urban underground space, both passive and active seismology have an obviously growing trend in the density and scale of their sensor arrays to achieve a more accurate inference or higher resolution image. Unfortunately, with the increasing density and scale of seismic data acquisition, the bandwidth limitation in communication is going to become the roadblock, which prevents seismic monitoring or exploration system from transmitting data timely [7]. Moreover, deploying more sensors with traditional seismic acquisition system leads to exponentially increasing energy cost, consequently enlarging the battery capacity and equipment weight [8]. As a result of the limitation of

The associate editor coordinating the review of this manuscript and approving it for publication was Shuihua Wang.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

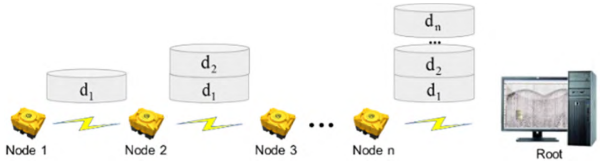


FIGURE 1. The basic logic of traditional data acquisition.

communication and energy, the traditional seismic acquisition method can hardly satisfy the requirement of acquiring data with an extremely large and dense sensor array.

During the procedure of seismic data acquisition, all sensors are placed on the surface to record vibrations produced by natural earthquake or artificial seismic sources. Each recorded signal from a sensor, termed as seismic trace, strands for the seismic wave through the sensor's placement. Because of the large scale of measured area, seismic data acquired by individual sensors are delivered to data center in a multi-hop form [9]. The basic logic of traditional data gathering method used in seismic data acquisition is shown in Fig. 1. Traditionally, sensors close to the data center need to afford more data to transmit, which results in more delaying time and consuming energy. Consequently, this unbalanced data gathering strategy make communicational and energy resources of the seismic data acquisition system consumed inefficiently.

Many attempts have been carried out to reduce the transmitting data in the system by implementing compression algorithms in individual sensors. These seismic data compression algorithms could be classified into two categories: dictionary-based and predictive methods [7]. The core idea of dictionary-based methods is transforming seismic data into another domain for compression, whereas predictive methods are aiming at using a predication algorithm to reduce data redundancy. However, both types of compression algorithms cost a large amount of computational resources for seismic sensors. In addition, these methods still cannot solve the inefficient problem caused by the unbalanced data collection architecture.

Compressed sensing (CS) theory is also leveraged to improve the efficiency of acquisition system [10]–[12]. As the spasity of seismic data is utilized, these CS-based methods are able to improve the efficiency of seismic data acquisition. However, most of them focused on the procedure of seismic signal recording, without regarding of the strategy of data collection. The remained unbalanced problem in the data retrieval has been a long-standing problem, which leads to inefficiency of data transmission and energy consumption.

Based on these observations, we propose a novel CS architecture in conjunction with generative adversarial networks (CSA-GAN), to improve the efficiency of seismic data acquisition system. This proposed method can not only solve the unbalanced data gathering problem, but also self-adaptively learn a sparse representation and reconstruction mapping for CS. Moreover, compared with state-of-the-art methods used in efficient seismic acquisition, there are two main advantages by applying CSA-GAN: first, only negligible computational cost is added to seismic sensors; second, it can be widely applied to different kinds of seismic acquisition. We also demonstrate that our method performs well on data recovery via performance assessment on both simulating data and real field data. To the best of our knowledge, it is the first work that utilizing CS theory combined with a deep learning algorithm for efficient seismic data acquisition.

The rest of this paper is organized as follows. In Section II, the recent works related to compressed and efficient seismic data acquisition are reviewed. Section III introduces the proposed method in detail. Then, the results, comparisons with other methods and related analyses are shown in Section IV. Finally, Section V concludes this paper.

II. RELATED WORK

In order to make the next generation of seismic data acquisition system have a great progress in density and scale, lots of researchers are mainly focusing on the two following approaches: compressing seismic data at individual sensors and improving acquisition efficiency by nonuniform sampling based on CS theory.

A. SEISMIC DATA COMPRESSION

To overcome the bandwidth limitation of communication, some in-field compression algorithms have been proposed. These algorithms reduce the amount of in-field data transfer by removing redundancy from the traces. To be specific, these seismic in-field compression algorithms can be generally summarized as two different kinds: dictionary-based and predictive methods.

Dictionary-based methods attempt to find a new domain where seismic signals have a compact and sparse representation. Early attempts for dictionary-based methods were mainly relying on fixed transforms, also termed as analytical predefined dictionaries, such as discrete cosine transform, wavelets, seislet [13]–[15]. Although analytical predefined dictionaries are applied to compress different types of seismic data, their recovery qualities have either a relative low signal-noise-ratio (SNR) or an unstable compression ratio. The main cause of this drawback is that analytical predefined dictionaries are only representative for the input data which is used to design the underlying mathematical models. Hence, it is impossible to select a fixed dictionary to perform stably and effectively on compressing the highly nonstationary seismic data. Latterly, some data-adaptive dictionaries are proposed to achieve more compact and sparse representation of the seismic signal, on the basis that seismic signals in one acquisition are almost following the same features and patterns. Liu *et al.* [16] proposed a scheme based on online dictionary learning, named sparse increment online dictionary learning (SIODL). To be specific, SIODL applies sparse constraints and a sliding window mechanism on the training process to speed up the convergence of online dictionary scheme. Compared with traditional online dictionary learning algorithm, this method consumes less computational power

and has a better compression result. In [17], a rate-optimized dictionary learning (Rate-Optimized DL) is proposed, which is aiming to minimize the overall bit rate in encoding the target domain coefficients directly instead of optimizing for sparsity of seismic signal representation. More specifically, Rate-Optimized DL attempts to represent seismic signals in a low entropy form, via considering the effect of quantization on the rate-distortion curve and the distributions of the coefficients in target domain. As encoding entropy is taken into account, Rate-Optimized DL has a significantly better performance than other dictionary-based methods used in seismic data compression. However, this consideration brings more computational complexity to Rate-Optimized DL as well, which makes it unavailable for seismic sensors.

Predictive methods are the most traditional compression algorithm used in time series data. These methods leverage the temporal or spatial correlation among consecutive samples to compress the residual signal from predictive coding. Like dictionary-based methods, the predictor in predictive methods are also have two types, known as fixed and adaptive predictors. Owing to the nonstationarity and diversity of seismic signals, an accurate prediction could be hardly achieved through a fixed number of predictor variables at fixed locations. Therefore, more recent researches related to predictive methods are focusing on utilizing the adaptive predictors. In [18], [19], a lightweight compression scheme for seismic sensor networks, namely adaptive linear filtering compression (ALFC), was proposed. This algorithm can learn a linear filter for predicting sample values followed by entropy coding of prediction residuals. Although ALFC is able to compress seismic data without loss, its compression ratio is only about 2. In [7], an oversampling recursive least square (OV-RLS) scheme was used to compress seismic data, whose aim is to alleviate the problems encountered in large alphabet signal compression and reduce the dynamic range of data. Particularly, seismic signals are oversampled, and a certain number of the samples are selected by a sequence of zero-crossings. Then, the selected samples are encoded into fewer bits by context tree weighting (CTW). At the receiver, the raw seismic can be recovered by a nonlinear method from the nonuniform samples. As shown in [7], OV-RLS and SIODL have a similar performance on seismic data compression.

However, all the state-of-the-art seismic data compressing methods require extra computational cost, compared with the original acquisition system. Even SIODL, ALFC and OV-RLS are relative lightweight among these seismic data compression algorithms, they still need to operate multiplications no less than hundreds of thousands of times per second at a sampling frequency of 500 Hz. For a seismic sensor, this computational cost is not negligible.

B. ACQUISITION EFFICIENCY IMPROVEMENT

To cover a larger acquisition area or achieve a higher resolution image with the same amount of equipment and cost, many researchers have studied seismic acquisition efficiency improvement, especially after appearance of the newly developed CS. Methodologies for the efficiency improvement are mainly designed through applying CS on two approaches: temporal and spatial domain.

The core idea of efficient acquisition methods used in temporal domain is that each seismic sensor records seismic signals in a random nonuniform and undersampled scheme by leveraging insights from CS. Following this approach, the data needed to be collected by the acquisition system can be reduced. This methodology was firstly proposed in [20], where seismic data were demonstrated to be successfully recovered from a well-designed nonuniform undersampling. In [21], a temporal domain CS-based acquisition system was implemented to monitor a bridge health for a long-term via recording its vibrations. More specifically, a pseudo-random generator is embedded into the central process unit (CPU) of seismic sensors, and the generator creates a subsampling time series. The analogue-to-digital (ADC) samples seismic data refer to the generated time series. So, the data recorded by seismic sensors and transmitted among the system is reduced. At the data center, the random subsampling timestamps and subsampled data are used for signal reconstruction via an iterative algorithm named CoSaMP [22]. As a result, this system can achieve the modal assurance criterion (MAC) values above 0.9 when only 20% of data volume is sampled and delivered. However, the method used in this system cannot be directly applied for other kinds of seismic acquisition, e. g., volcanic monitoring, seismic exploration, earthquake monitoring. The reason for this restriction is that the bridge vibrational signal is mainly focused on a very narrow frequency bandwidth (mentioned as 3-20 Hz), whereas the bandwidth of seismic signal varies significantly in other kind seismic acquisition. In detail, because of the specific structure of bridge and the common physical features of concrete, the signal of bridge vibration could be sparse enough in the transform domain, which is a very important factor that impacts the compression effect. In other kinds of seismic acquisition, geological structures, physical feature of geosphere and seismic sources are all aeolotropic and various, which results in the wide frequency band of seismic signals.

As another approach for improving the efficiency of seismic acquisition, the spatial domain CS-based methods have been proposed to optimize seismic sensors deployment. These methods aim to reduce acquisition costs, cover a larger area with fewer sensors, or increase the image resolution. One of the most representative researches focusing on such approach is presented in [23], which proposes a nonuniform sampling method for seismic survey, termed as Non-uniform Optimal Sampling (NUOS). This method combines CS, optimization algorithms and seismic wave theory together, to design an efficient sensor placement for a seismic survey. Specifically, a Monte-Carlo optimization scheme is used in NUOS to determine the location of each seismic sensor, relying on prior knowledge of the seismic survey. According to the result of field trials, a deployment based on NUOS can cover one time larger than a conventional survey with

the same image resolution in 2D seismic exploration. This result is equal to 50% data reduced, compared with traditional seismic acquisition method.

Although these summarized methods have improved the efficiency of seismic acquisition, there are still a few aspects remained to be regarded. First, as seismic data is correlative in both temporal and spatial domain, it could be more efficient once applying CS toward temporal-spatial domain. Second, data-driven methods combined with CS could be more effective on multiple seismic acquisition applications, due to the diversity of seismic data. At last, all these methods cannot solve the unbalanced data gathering problem.

III. THE PROPOSED METHOD

In this section, we first present our efficient acquisition architecture used in CSA-GAN. Second, the main idea and structure of the reconstruction algorithm are introduced in a high-level view. Finally, the two main parts of our reconstruction algorithm, namely the generator and discriminator, are described in detail, respectively.

A. EFFICIENT COMPRESSED SENSING ARCHITECTRUE IN CSA-GAN

According to the CS theory referred by [24], a vector $x \in \mathbb{R}^n$ can be conditionally recovered from an underdetermined linear measured vector $y \in \mathbb{R}^m$, described as following:

$$y = \boldsymbol{A}x + z \tag{1}$$

where $\boldsymbol{A}$ is a stochastic or deterministic measurement matrix with the size of $M \times N$ ($M < N$), and z is an unknown error term. There are two conditions for perfectly recovering x: first, x is a p-sparse vector and $p \leq M/2$; second, the matrix $\boldsymbol{A}$ satisfies the restricted isometry property (RIP) condition.

In most cases, seismic signal does not satisfy the sparsity required by CS theory directly. However, there is a sparse land for seismic data, referring to [25], [26]. So, the undersampling method stated in equation (1) can be applied for seismic data, after the seismic data is translated into its sparse domain. Then, equation (1) can be converted into the following form:

$$\boldsymbol{y} = \boldsymbol{s}\boldsymbol{\Psi}\boldsymbol{\Phi} + z \tag{2}$$

where $\boldsymbol{\Phi}$ is the new measurement matrix ($\boldsymbol{\Phi} \in \mathbb{R}^{N\times M}$), $\boldsymbol{\Psi}$ is the sparse dictionary of seismic data ($\boldsymbol{\Psi} \in \mathbb{R}^{D\times N}$), and N, M, D are the number of sensors, compressed signal dimension and sparse domain dimension, respectively. Compressed seismic data is represented as $\boldsymbol{y}$ ($\boldsymbol{y} \in \mathbb{R}^{T\times M}$). Original seismic data is termed as $\boldsymbol{x}$ ($\boldsymbol{x} \in \mathbb{R}^{T\times N}, \boldsymbol{x} = \boldsymbol{s}\boldsymbol{\Psi}$), where T means time length of seismic data and $\boldsymbol{s} \in \mathbb{R}^{T\times D}$ is the corresponding coefficients. The two recovery conditions are stated as follow: first, the sparsity of each row in the $\boldsymbol{s}$ is less than $M/2$; second, the matrix $\boldsymbol{\Psi}\boldsymbol{\Phi}$ satisfies the RIP condition.

In traditional seismic data acquisition, the sensor network is relying on a line-based multiple-hop structure. Considering the drawbacks of data collection in the line-based multiple hop network, equation (2) is applied to design an efficient acquisition architecture. To solve the bottleneck on data collection and allow each sensor to transmit data volume equally, the architecture make seismic acquisition system collect the compressed data $\boldsymbol{y}$ instead of the original data $\boldsymbol{x}$. The proposed acquisition architecture is shown in Fig. 2. Particularly, each sensor multiplies the recorded seismic data by its corresponding column vector in $\boldsymbol{\Phi}$, then adds this product with the vector received from its child node. At last, this seismic sensor sends the summed vector to its father node. For example, at time j, the seismic data x_{ji} is recorded in the sensor node i. Then, the measured vector for the node i can be computed via $x_{ji}\Phi_i$, where Φ_i denotes the i th column of $\boldsymbol{\Phi}$. After that, the sensor node i sums its own measured vector with the compressed vector of its child node, formulated as $\sum_{q=1}^{i} x_{jq}\Phi_q$. Finally, node i delivers its compressed vector to the sensor node $i+1$.

FIGURE 2. Compressed sensing architecture in CSA-GAN.

By utilizing this architecture, there are three improvements. First, the unbalanced problem in acquisition system is solved Communication bandwidth required is reduced from N to M and energy cost by data collection is changed from $N(N+1)/2$ to NM. Second, the compressed data could be reconstructed via time-frequency domain, since both temporal and spatial information are merged into the compressed data. Third, only negligible computational cost is introduced to the seismic sensors, as just M more multiplications are added to each sensor at one recording time. The extra computation is totally about thousands of multiplications per second at a sampling frequency of 500 Hz, which is much less than other seismic compression methods reviewed in Section II.

B. The MAIN IDEA OF RECONSTRUCTION ALGORITHM IN CSA-GAN

To recover the original signal $\boldsymbol{x}$ from equation (2), there are two parts must be determined: the measurement matrix $\boldsymbol{\Phi}$ and signal reconstruction mapping. For the measurement matrix, it is necessary to satisfy the RIP condition. As random Gaussian matrices has been proved to own an acceptable RIP and widely used in CS [27], a norm Gaussian matrix is used as the measurement matrix in our work. For the signal reconstruction mapping, many model-based methods, such as FISTA [28] and ADMM [29], have been proposed with theoretical guarantees. However, there are numerous uncertainties in these methods, such as transform domain, sparse regularizer in the transform domain, regularization parameters. These uncertainties are challenging to be determined properly, because various features of the input data must be considered. Inspired by the work introducing neural networks to image restoration and CS imaging tasks [30]–[32], we take a deep neural network to fit the reconstruction mapping with

the consideration of regularizations in CS, such as sparse transform domain and regularization. In addition, the deep neural network can handle varied features of seismic data in a data-driven way, which makes CSA-GAN be suitable for different kinds of seismic acquisition.

However, there are three non-trivial matters for a traditional neural network to be applied for CS image recovery. First, there is a prevailing problem caused by overfitting in a traditional neural network and other general machine learning algorithms. This problem restricts the reconstruction method based on traditional neural networks to a particular seismic acquisition. Second, traditional neural networks are not appropriate for seeking out the optimal solution to CS reconstruction problem, which consists of a certain amount of distributed subproblems. Moreover, these subproblems are considerably high-dimensional (such as representation learning and compressed data reconstruction), which make it difficult for traditional neural networks to directly reverse the compressed procedure and recover the original data in the high dimensional solution space. Third, the success of CS is relying on sparse representation, whereas the traditional conventional neural network is not designed for learning a reversible and sparse transform in an automatic form.

To deal with these three non-trivial issues simultaneously, a novel GAN-based CS reconstruction algorithm is proposed in this paper. First, an adversarial process is introduced to the reconstruction algorithm to provide the model an ability for generalization. As proven in [33], the adversarial framework is able to learn explicit and tractable features effectively, because the optimization is used directly on the likelihood the training data. Therefore, in the proposed reconstruction algorithm, an adversarial process is used to extract the common features of seismic data so as to avoid overfitting for a special case. Second, due to converging efficiency of the alternating direction method of multipliers (ADMM), insights from ADMM are taken into our reconstruction algorithm to tackle all the subproblems in CS reconstruction via a decomposition-coordination procedure. To be specific, the reconstruction is spilt into three jobs: transform domain learning, signal recovery learning and representation recovery learning. In the training process, the solutions to the three jobs are coordinated to find the final solution in an alternating direction mechanism. Third, we design a convolution neural network (CNN) module in the proposed model to explore the sparse domain of seismic signal via utilizing the connections between sparse modeling and deep learning architectures. The related foundations for the connections are presented in [34].

C. THE RECONSTRUCTION ALGORITHM STRUCTURE IN CSA-GAN

The proposed reconstruction algorithm can be formulated mathematically as a minimax game with two players. One is termed as a discriminator network D, whose aim is to learn a suitable transform according to CS theory; the other one is termed as a generator network G, which is kept finding a proper mapping from compressed data to original signal.

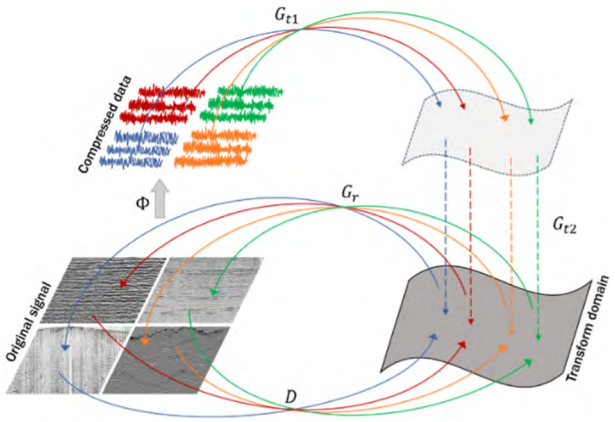


FIGURE 3. Reconstruction algorithm structure in CSA-GAN.

In order to match the procedure of CS imaging, a two-step generative learning scheme is designed. In the first step, a transforming module G_t is used to output the representation coefficients in transform domain learned by D, when the compressed data is given. Considering the different kinds of correlations between the compressed data and representation in the transform domain, there are two stages in G_t, namely G_{t1} and G_{t2}, which are designed to handle the global and local correlations. For the second step of the generator, a recovering module G_r is applied to recover the original signal from the transform domain.

The particular structure of our reconstruction algorithm structure is illustrated in Fig. 3 and detailed in the following two aspects: adversarial process and alternating direction mechanism.

1) ADVERSARIAL PROCESS

In the proposed reconstruction algorithm, the generator G and discriminator D play an adversarial game to search for the optimal solution to seismic signal reconstruction. More specifically, G wishes to learn a mapping with a less reconstruction error in the transform domain learned by D, while D attempts to find a transform domain where the reconstruction error could be measured as a larger value based on the mapping learned by G. To this end, G can generate a reconstruction mapping with less error, and D is able to underline the recovering error more clearly. Both G and D compute their costs according to the other one, and each of them can only control its own parameters to search for a better solution. Motivated by the improved stability of LSGAN [35], least square function is used to measure the distribution distances in our model. The specific adversarial process in our reconstruction algorithm can be formulated as following:

$$\min_{\theta_G}\max_{\theta_D}\left(\frac{\|D\left(\boldsymbol{x}-G\left(\boldsymbol{y};\theta_G\right);\theta_D\right)\|_2}{\|\boldsymbol{x}\|_2}\right)^2-\left(\frac{\|D(\boldsymbol{x})\|_2}{\|\boldsymbol{x}\|_2}-1\right)^2 \tag{3}$$

where θ_D and θ_G are the parameters of D and G to be learned, respectively.

Thus, D learns a new transform domain, which highlights the reconstruction error with respect to the mapping generated by G; G learns a new reconstruction mapping by minimized the reconstruction error in the transform domain updated by D. In other words, D keeps discovering the drawbacks of the solution proposed by G, and G is trying to update the new solution according to the suggestion given by D.

2) ALTERNATING DIRECTION MECHANISM

On the basis of CS theory, the reconstruction problem in this work can be spilt into three subproblems: transform domain learning, representation recovery learning and signal recovery learning. In the reconstruction algorithm, D is designed to learn the transform domain, while G_t and G_r aim to perform representation recovery learning and signal recovery learning, respectively. Besides, there two more requirements for the relationships among D, G_t and G_r. First, to ensure the transforming learned by D is reversible in a certain space and the signal recovering mapping learned by G_r can recover the original data from the transform domain, D and G_r are required to comprise an encoder-decoder framework. Second, to guide G_t to learn the correct representation coefficients with respect to corresponding G_r, the representation learned by D should act as the label to supervise G_t. Taking these aims and constraints into account, an alternating direction mechanism is introduced to the reconstruction algorithm to find the final optimal solution. In each iterative time, D, G_t and G_r are coordinate to updated for the solutions to their own subproblems with newest parameters of the other two. The specific procedure can be described as following:

$$\begin{cases} \theta_D^{(k+1)} = \min\limits_{\theta_D^{(k)}} \mathcal{L}_D\left(\boldsymbol{x}, \boldsymbol{y}; \theta_D^{(k)}, \theta_{Gr}^{(k)}, \theta_{Gt}^{(k)}\right) \\ \theta_{Gr}^{(k+1)} = \min\limits_{\theta_{Gr}^{(k)}} \mathcal{L}_{Gt}\left(\boldsymbol{x}, \boldsymbol{y}; \theta_D^{(k+1)}, \theta_{Gr}^{(k)}, \theta_{Gt}^{(k)}\right) \\ \theta_{Gt}^{(k+1)} = \min\limits_{\theta_{Gt}^{(k)}} \mathcal{L}_{Gr}\left(\boldsymbol{x}, \boldsymbol{y}; \theta_D^{(k+1)}, \theta_{Gr}^{(k)}, \theta_{Gt}^{(k+1)}\right) \end{cases} \tag{4}$$

in which k denotes k-th iteration. θ_{Gr} and θ_{Gt} are the parameters of G_r and G_t to be learned, respectively. $\mathcal{L}_D$, $\mathcal{L}_{Gt}$ and $\mathcal{L}_{Gr}$ are the cost functions of D, G_t and G_r, respectively.

D. THE DISCRIMINATOR IN CSA-GAN

The discriminator D aims at learning a transform domain with the consideration of reconstruction error and sparsity, by given the original seismic data. So, D is designed to implement an invertible and sparse projection by dealing with the correlations contained in a 2-dimensional matrix. The details in D are presented in the two aspects: the network structure and cost function.

1) THE NETWORK STRUCTURE OF DISCRIMINATOR

As the seismic profiles contain temporal-spatial correlations, the residual information remained in the reconstruction error also own a certain correlation neglected by the recovering mapping. Motived by leveraging the neglected correlation for more effective representation, D applies convolution operations to extract features of the residual information for the transform domain learning. To guarantee sparsity of the learned transform domain, a rectified linear unit (ReLU) is used as the sparsifying operation after every convolution operation. To be specific, D applies 3 convolution layers with ReLU to extract features, also termed as atoms in the dictionary learning area. There are 32 filters in the first layer, whose aim is to capture the local features. So, the size of these filter is 7×7, which is larger the minimum size of seismic data characters in temporal, spatial and frequency domain. The stride of first layer is 2. There are 128 and 512 filters in the second and third layer, respectively. They are aiming at global feature extraction, on account of their larger sensing field. The size of filters in the second layer is 5×5, and the size of filters in the third layer is 3×3. Their stride is 1. The neural network structure of D is illustrated in Fig. 4. Specifically, in Fig 4, t is the time length of seismic profiles and n represents the number of sensors in the seismic profile. Since values in the edges of seismic profile are around 0, the padding value used in the convolutional operations is 0.

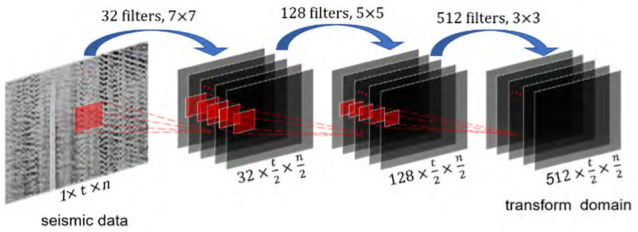


FIGURE 4. The neural network structure of discriminator.

2) THE COST FUNCTION OF DISCRIMINATOR

To guide D to learn a suitable transform domain for CS, the cost function must have some items regarding reconstruction error, sparsity and RIP, which also guarantee the consistency and stability of GAN. Taking the recovering precision and generalization into account, there are two key items needed to be included by the cost function. The first item is the reconstruction error, which must be minimized to achieve the requirement of recovering precision. The second one is the adversarial loss for D, which values reconstruction error in the learned transform domain. This adversarial loss leads D to capture the features of reconstruction error and provide an optimized direction for recovering mapping. Considering the sparse representation and RIP condition in CS, the cost function should also include regularization items related to the sparsity of representation and the incoherence between measurement matrix $\boldsymbol{\Phi}$ and the transformation. For the sparse regularization, ℓ_1-norm is used, since it is more tractable than ℓ_0-norm and lots of literatures has proven the CS problem can be solved under ℓ_1-norm regularization. Although random Gaussian matrices have been shown to satisfy an acceptable RIP, it is still necessary for high-quality reconstruction to

ensure the restricted isometric constant remains a relative low value. To guarantee a better RIP, the learned transformation needs to have a lower incoherence with $\mathbf{\Phi}$. So, ℓ_2-norm of the projection of $\mathbf{\Phi}$ in the learned transform domain is chosen as the RIP regularization. The cost function of D, denoted as $\mathcal{L}_D$, is represented as:

$$\mathcal{L}_D = \min_{\theta_D, \lambda_{1\sim4}} \lambda_1 \mathcal{E}_E + \lambda_2 \mathcal{L}_{GANd} + \lambda_3 \mathcal{R}_{RIP} + \lambda_4 \mathcal{R}_{sparse} \tag{5}$$

$$\mathcal{E}_E = \|X - G_r\left(D(\boldsymbol{x}; \theta_D); \theta_{Gr}\right)\|_2 \tag{6}$$

$$\mathcal{L}_{GANd} = \left(\frac{\|D(\boldsymbol{x})\|_2}{\|\boldsymbol{x}\|_2} - 1\right)^2 - \left(\frac{\|D\left(\boldsymbol{x} - G\left(\boldsymbol{y}; \theta_G\right); \theta_D\right)\|_2}{\|\boldsymbol{x}\|_2}\right)^2 \tag{7}$$

$$\mathcal{R}_{RIP} = \|D(\mathbf{\Phi}; \theta_D)\|_2 \tag{8}$$

$$\mathcal{R}_{sparse} = \|D\left(\boldsymbol{x}; \theta_D\right)\|_1 \tag{9}$$

where $\mathcal{E}_E$ represents the recovering error of the encoder-decoder procedure composed of D and G_r. $\mathcal{L}_{GANd}$ is the loss of adversarial process for D. $\mathcal{R}_{RIP}$ and $\mathcal{R}_{sparse}$ are the regularization of RIP and sparsity, respectively. λ_1, λ_2, λ_3 and λ_4 are trade-off parameters and can be calculated via the following equations:

$$e_1 = tanh(\theta_{\lambda_1} \mathcal{E}_E + b_{\lambda_1}) \tag{10}$$

$$e_2 = tanh(\theta_{\lambda_2} \mathcal{L}_{GANd} + b_{\lambda_2}) \tag{11}$$

$$e_3 = tanh(\theta_{\lambda_3} \mathcal{R}_{RIP} + b_{\lambda_3}) \tag{12}$$

$$e_4 = tanh(\theta_{\lambda_4} \mathcal{R}_{sparse} + b_{\lambda_4}) \tag{13}$$

$$\lambda_{t_1} = \frac{exp(e_{t_1})}{\sum_{n=1}^{4} exp(e_n)} \tag{14}$$

where $t_1 = 1, 2, 3, 4$; $\theta_{\lambda_1}, \theta_{\lambda_2}, \theta_{\lambda_3}, \theta_{\lambda_4}$ and $b_{\lambda_1}, b_{\lambda_2}, b_{\lambda_3}, b_{\lambda_4}$ are weight and bias terms to be learned. All these parameters are trained during the backpropagation process.

E. THE GENERATOR IN CSA-GAN

There are two modules in the generator G, namely transforming module G_t and recovering module G_r. G_t aims to learn the transform from compressed data $\boldsymbol{y}$ to the representation $\boldsymbol{s}$ projected by D, by considering the connections between $\boldsymbol{y}$ and $\boldsymbol{s}$. G_r focuses on learning the mapping from the transform domain $\boldsymbol{s}$ to original seismic data $\boldsymbol{x}$, which also be thought as an inverse procedure of D.

1) THE NETWORK STRUCTURE OF GENERATOR

For the transforming module G_t, to allow $\boldsymbol{y}$ be projected as the same representation as its corresponding $\boldsymbol{x}$, a data-driven supervised representation learning framework is implemented. The sparse representation $\boldsymbol{s}$ in the transform domain is the label of its corresponding compressed data $\boldsymbol{y}$. To fully mine the information contained in $\boldsymbol{y}$, the features contained by $\boldsymbol{x}$ and the connections between $\boldsymbol{y}$ and $\boldsymbol{x}$ must be considered in the structure design of G_t. More specifically, since each element of $\boldsymbol{y}$ is calculated using all the elements of $\boldsymbol{x}$ and $\mathbf{\Phi}$, $\boldsymbol{y}$ should be processed in a global pattern to capture the original features remained in $\boldsymbol{x}$. Therefore, the first stage of the transforming module G_{t1} is designed to globally approximate the between-manifold projection from the compressed data domain to the sparse representation domain, by considering the latent global correlations between $\boldsymbol{y}$ and $\boldsymbol{s}$. There is one fully-connected layer with t × n neurons. After the mapping of G_{t1}, there are some local correlations left, which are the small-scale relationships between $\boldsymbol{y}$ and $\boldsymbol{s}$ at a similar position. So, in the second stage of transforming module G_{t2}, local correlations between $\boldsymbol{y}$ and $\boldsymbol{s}$ become the key to lead G_{t2} generate a representation similar to $\boldsymbol{s}$ output by D. G_{t2} uses a CNN structure that is the same with D to learn the mapping for the sparse representation based on the features extracted by G_{t1}. The specific structure of G_t is shown in Fig. 5 (a).

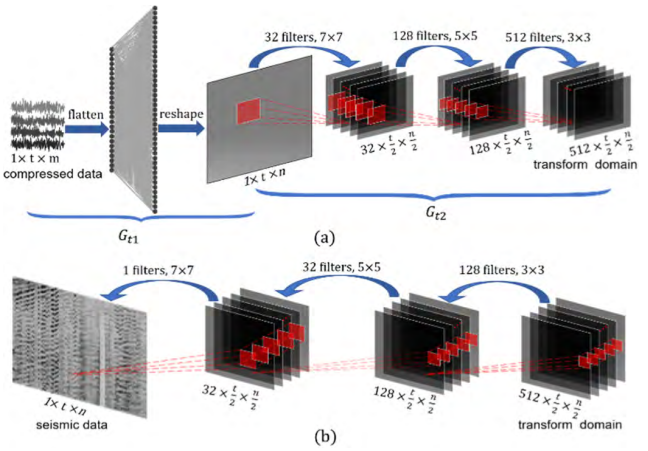


FIGURE 5. The neural network structure of generator. (a) The network structure of G_t. (b) The network structure of G_r.

To ensure the original data can be recovered from the representation $\boldsymbol{s}$ in the sparse transform domain, the recovering module G_r is designed to be a decoder for D. Since the decoder and encoder should appear in pair, deconvolutional operations with ReLU are utilized in G_r, and the structures of G_r and D are symmetric. In detail, the first deconvolutional layer in G_r corresponds to the last convolutional layer in D, the last deconvolutional layer in G_r corresponds to the first convolutional layer in D, and so on. The structure of G_r is shown in Fig. 5 (b).

2) THE COST FUNCTION OF GENERATOR

For implementing of the projection from compressed data $\mathbf{y}$ to its corresponding representation $\boldsymbol{s}$ in the learned transform domain, as well as taking the feedback from the discriminator D into consideration, G_t should be guided by both its corresponding label on representation and the adversarial loss. So, the cost function of G_t, denoted as $\mathcal{L}_{Gt}$, is represented as:

$$\mathcal{L}_{Gt} = \min_{\theta_{Gt}, \lambda_{5\sim6}} \lambda_5 \mathcal{E}_R + \lambda_6 \mathcal{L}_{GANg} \tag{15}$$

$$\mathcal{E}_R = \|G_t\left(\boldsymbol{y}; \theta_{Gt}\right) - D\left(\boldsymbol{x}; \theta_D\right)\|_2 \tag{16}$$

$$\mathcal{L}_{GANg} = \left(\frac{\|D\left(\boldsymbol{x} - G_r\left(G_t\left(\boldsymbol{y}; \theta_{Gt}\right); \theta_{Gr}\right); \theta_D\right)\|_2}{\|\boldsymbol{x}\|_2}\right)^2 \tag{17}$$

where $\mathcal{E}_R$ stands for the error of representation in the transform domain and $\mathcal{L}_{GANg}$ is the loss provided by adversarial process for G. λ_5 and λ_6 are trade-off parameters and calculated via the following equations:

$$e_5 = tanh(\theta_{\lambda_5}\mathcal{E}_R + b_{\lambda_5}) \tag{18}$$

$$e_6 = tanh(\theta_{\lambda_6}\mathcal{L}_{GANg} + b_{\lambda_6}) \tag{19}$$

$$\lambda_{t_2} = \frac{exp(e_{t_2})}{\sum_{n=5}^{6} exp(e_n)} \tag{20}$$

where λ_5 and λ_6 are trade-off parameters. $t_2 = 5, 6$. Moreover, θ_{Gt}, θ_{λ_5}, θ_{λ_6}, b_{λ_5} and b_{λ_6} are the learnable parameters and updated via the backpropagation.

To guide G_r to learn the mapping from transform domain to original seismic data as well as be a close part of the GAN, not only the recovering error caused by the encoder-decoder procedure (D and G_r), but also the adversarial loss from adversarial process should be used to update the learnable parameters. Thus, the cost function of G_r, denoted as $\mathcal{L}_{Gr}$, is represented as:

$$\mathcal{L}_{Gr} = \min_{\theta_{Gr}, \lambda_{7\sim 8}} \lambda_7\mathcal{E}_E + \lambda_8\mathcal{L}_{GANg} \tag{21}$$

$$e_7 = tanh(\theta_{\lambda_7}\mathcal{E}_E + b_{\lambda_7}) \tag{22}$$

$$e_8 = tanh(\theta_{\lambda_8}\mathcal{L}_{GANg} + b_{\lambda_8}) \tag{23}$$

$$\lambda_{t_3} = \frac{exp(e_{t_3})}{\sum_{n=7}^{8} exp(e_n)} \tag{24}$$

where λ_7 and λ_8 are trade-off parameters. $t_3 = 7, 8$. Besides, θ_{Gr}, θ_{λ_7}, θ_{λ_8}, b_{λ_7} and b_{λ_8} are the learnable parameters.

IV. EVALUATION

In this section, the performance of our efficient seismic data acquisition method is evaluated on both synthetic and real seismic data. First, the description of datasets and experimental setting used in our experiments are detailed. Next, synthetic seismic data is used to study the impacts of key hyperparameters used in the model, such as compression ratio and training epochs. The varied characters of synthetic seismic data also play a role to certify that our method has an ability to handle seismic data with different features. Finally, our method is tested on real seismic data and compared with several popular methods.

A. DATASETS AND EXPERIMENTAL SETTING

1) SYNTHETIC SEISMIC DATASET

Since there are many kinds of seismic data acquisition aiming at different applications, the characters of recorded seismic data are varied a lot. So, we use Ricker wavelets with different central frequency to generate various seismic data. According to the main frequency band of seismic wave recorded in the commonest use cases, Ricker wavelets with central frequency from 5Hz to 500Hz are used to generate seismic profiles. Some of these seismic profiles comprise the training set, while the others comprise the testing set. Each seismic profile includes 64 traces averagely distributed on a 640-meter line. The length of sampling time is 1.28s, and the sampling frequency is 1000Hz. Therefore, the sizes of these seismic profiles are all 1280×64.

In the training set, there are totally 1024 samples, whose central frequency are distributed evenly across the frequency domain from 5Hz to 500Hz. Due to the frequency bandwidth of Ricker wavelets in the training set, every two Ricker wavelets with adjacent central frequency have a significant overlap between each other in the frequency domain. Thus, 1024 training samples in the training set are able to cover main information in the frequency domain from 5Hz to 500Hz. Besides, the amplitudes of these seismic profiles and geophysical properties of seismic wave are also generated in a uniform distribution.

In the testing set, there are 512 samples, whose central frequency are at the middle of the two adjacent central frequency Ricker wavelets in the training set. So, the testing seismic profiles have the largest difference with training seismic profiles in this simulation, which would make the testing set be the most appropriate samples to examine effectiveness of the trained model. In addition, the amplitudes of testing seismic profiles and geophysical properties of seismic wave also obey the same uniform distribution with the training set.

2) REAL SEISMIC DATA

To prove our method is effective on the real application, a set of field seismic data is needed to evaluate the performance. Since the proposed data collection framework is used in each acquisition terminal, there should not be any data preprocess in the testing data, e. g., correction and stack. This kind of raw seismic data is not common in online open seismic dataset. So, the real seismic data used in our experiment is a set of seismic exploration data collected by our laboratory. The seismic data acquisition system proposed in [36] was used in this seismic exploration. Specifically, this set of seismic data was acquired for exploring the underground structure in a hilly land, and the place of this exploration was conducted in the southwest of Liaoning Province, China. There were 2400 traces averagely distributed on four 8-kilometer-length lines in this exploration. The sampling time length was 8s, and the sampling frequency was 500Hz.

In the real seismic data experiment, the training set is comprised of 1024 generative seismic profiles yielded by a forward seismic modeling of the target area, which is set up according to some hilly-related geological prior knowledge, such as the scope of wave speed, the possible underground structures, properties of the used seismic source.

All data acquired by the seismic exploration is used to comprise the testing set in the real seismic data experiment. To be specific, there are two types of seismic data recorded by our seismic data acquisition system. One is the regular data, collected in the traditional way, which is used as the ground truth in our experiment; the other one is the compressed data, which is collected by the proposed compressed sensing architecture and used as the input of our reconstruction algorithm.

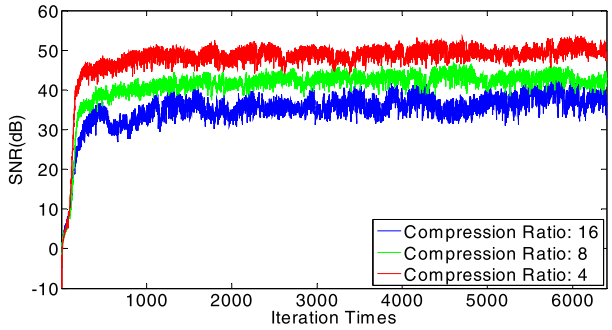


FIGURE 6. SNR curves with different compression ratios on the training set over 6400 iterations (200 epochs).

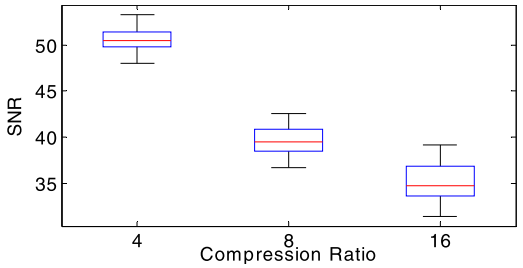


FIGURE 7. SNR with different compression ratios on the testing set.

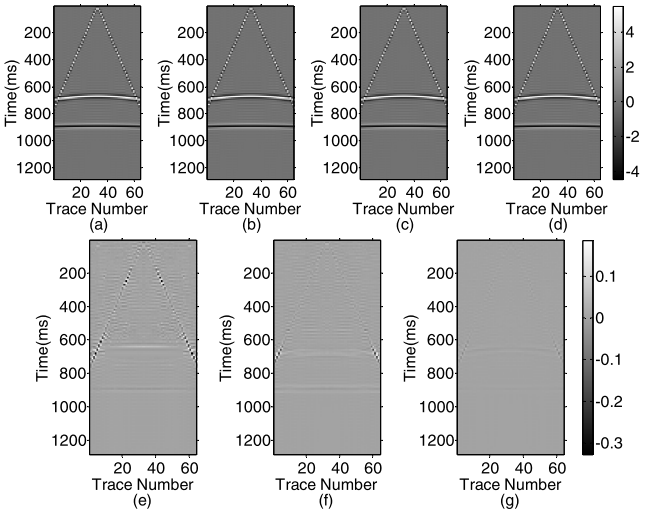


FIGURE 8. Simulating sample I and its recover results with different compression ratios. (a) Original seismic profile I. (b) The reconstruction image of 16-fold compression (SNR: 34.29dB). (c) The reconstruction image of 8-fold compression (38.42dB). (d) The reconstruction image of 4-fold compression (48.47dB). (e) The residual image of (b). (f) The residual image of (c). (g) The residual image of (d).

3) EXPERIMENTAL SETTING

Since most parameters of the proposed reconstruction algorithm can be learned by training, there are only two parameters need to be defined for the configuration of training process. They are batch size and learning rate. It is clear that the larger batch sizes need fewer epochs to get an acceptable result. However, there are only 32 samples in each batch for seismic training process, as the size of real seismic data is considerably large. Since a relatively small batch size is used here, an exponentially decaying learning rate start with a small value should be applied to minimize the noising scale brought by small batch size. Thus, the learning rate starts at 0.01 and decays every 100 steps with a base of 0.9. All cost functions in our method are optimized via the Adam algorithm. The proposed method and all the experiments in this work are implemented through TensorFlow and Python 3.

B. SYNTHETIC DATA RESULTS AND ANALYSIS

In the synthetic seismic data experiment, we aim to select a proper compression ratio for CSA-GAN, by taking both convergence speed and reconstruction performance into account. we first evaluated the convergence of our method with different compression ratios. The three different compression ratios are 4, 8 and 16. Since the training set has 1024 samples and the batch size is 32, there are 32 iterations in each epoch. The result of training procedure is shown in Fig. 6. Roughly speaking, the training process shows that our reconstruction method can reach a stable convergence stage within 2000 iterations (about 60 training epochs). In detail, when the compression ratio is low (e. g., 4), the reconstruction SNR can achieve a considerable value (e. g., 50 dB) within about 1000 iterations. When the compression ratio is relatively high (e. g., 16), the reconstruction SNR can still reach an acceptable value (e. g., 35 dB) within about 1000 iterations. So, Both the reconstruction SNR and rate of convergence have an acceptable decrease, until the compression ratio increases to 16.

According to Fig. 7, the testing result is consistent with the training result. Although the SNR tends to drop and diverge with the increasement of compression ratio, the reconstruction result of testing set is still acceptable. Specifically, with the compression ratios of 4, 8, and 16, the average values of SNR are 50.65dB 39.61dB and 35.15dB, respectively. When the compression ratio becomes 16, the range of SNR is between 31.43 to 39.17dB. According to the synthetic data results, our reconstruction method performs stably and effectively enough for seismic data acquisition when the compression ratio is 16.

For detailed analyses on the reconstruction effect, three different testing samples are shown in Figs 8, 9 and 10. To be specific, Fig. 8 (a), Fig. 9 (a) and Fig. 10 (a) are original profiles of these three testing samples. Figs 8 (b), (c) and (d) are reconstruction images of Fig. 8 (a), and their compression ratios are 4, 8, and 16, respectively. Figs 8 (e), (f) and (g) are residual images of Figs 8 (b), (c) and (d), respectively. Figs. 9 and 10 have the same format as Fig. 8. Since, the geophysical parameters (e. g., central frequency, seismic wave speed and underground structure) are different in these three original profiles, the reconstruction results prove that our method can recover the original seismic data under various geophysical circumstances. Based on the comparison of residual images with different compression ratios, it is easy to find two following phenomena: first, there is no obvious loss in the reconstruction results, when the compression ratio is less than 8; second,

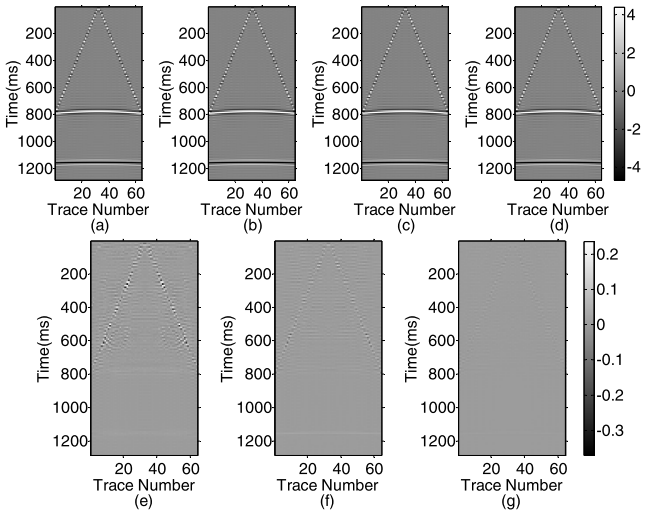


FIGURE 9. Simulating sample II and its recover results with different compression ratios. (a) Original seismic profile II. (b) The reconstruction image of 16-fold compression (34.61dB). (c) The reconstruction image of 8-fold compression (38.18dB). (d) The reconstruction image of 4-fold compression (49.06dB). (e) The residual image of (b). (f)The residual image of (c). (g) The residual image of (d).

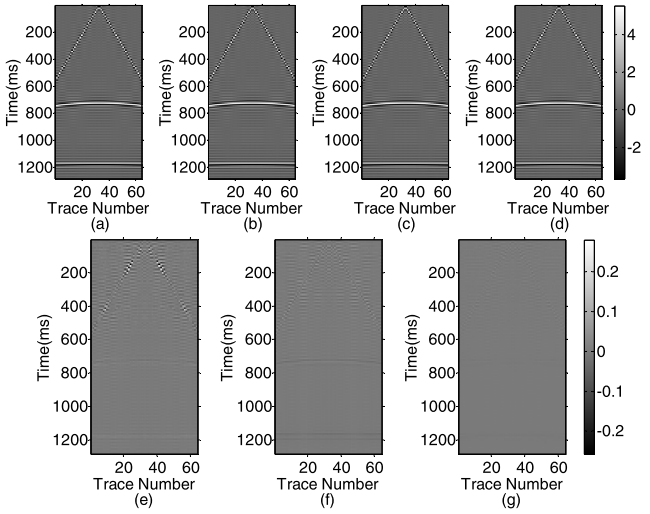


FIGURE 10. Simulating sample III and its recover results with different compression ratios. (a) Original seismic profile III. (b) The reconstruction image of 16-fold compression (33.98dB). (c) The reconstruction image of 8-fold compression (39.08dB). (d) The reconstruction image of 4-fold compression (51.15dB). (e) The residual image of (b). (f) The residual image of (c). (g) The residual image of (d).

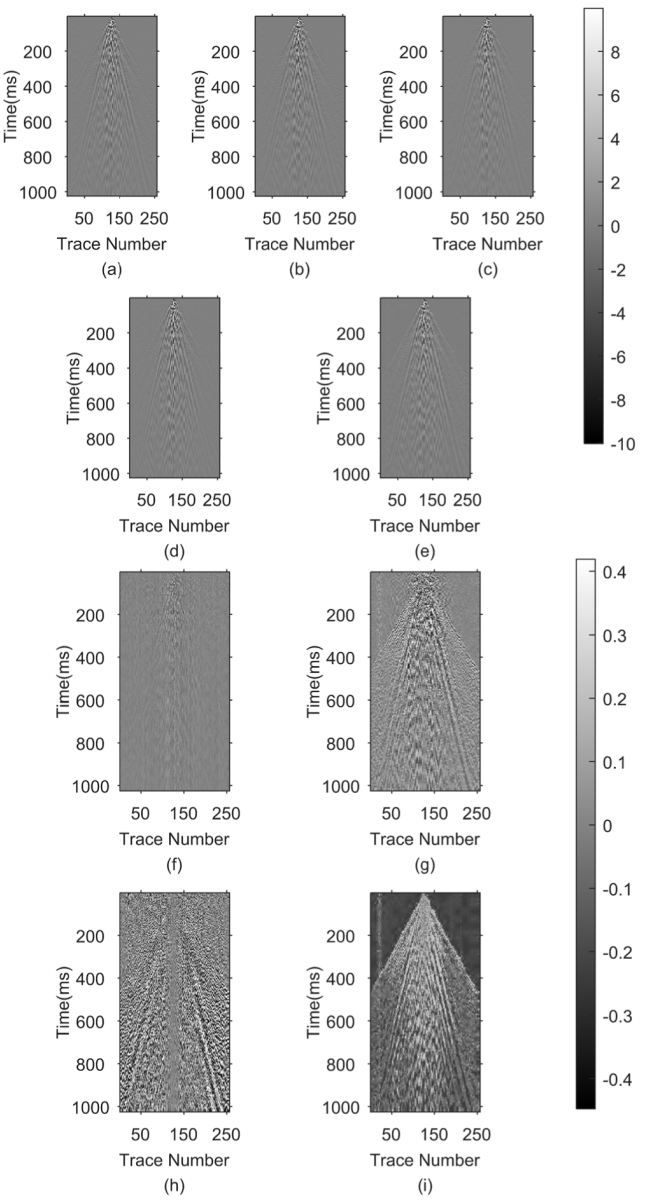


FIGURE 11. Measuring line I and its recover results with different methods. (a) Original seismic profile of measuring line I. (b) The reconstruction image (29.24dB) with CSA-GAN. (c) The reconstruction image (22.02dB) with SIODL. (d) The reconstruction image (18.73dB) with DPCA. (e) The reconstruction image (11.56dB) with JPEG2000. (f) The residual image of (b). (g) The residual image of (c). (h) The residual image of (d). (i) The residual image of (e).

when the compression ratio is large, such as 16, the main loss of recovery is concentrated in the direct wave part, which appears linear characters in the upper seismic profiles. The reason for the second phenomenon is that the direct waves has a great difference with the reflected and refracted waves, which appear quadratic characters in the bottom seismic profiles. The proposed model can easily extract the features of the reflected and refracted waves, whereas lose some features of the direct wave. However, it has little influence on seismic data acquisition, since the direct waves are not as import as the reflected and refracted waves in seismic data analysis.

C. REAL DATA RESULTS AND ANALYSIS

Considering the simulating results and resolution demands of seismic data acquisition, a compression ratio of 16 is chosen to implement the field seismic data experiment. To assess the data recovering performance of our reconstruction algorithm, three different seismic data compression methods are used as benchmarks. One of the benchmark methods is SIODL [16], which has been thought as a state-of-the-art dictionary learning algorithm used in seismic data compression [7]; another state-of-the-art seismic data compression method is distributed principal component

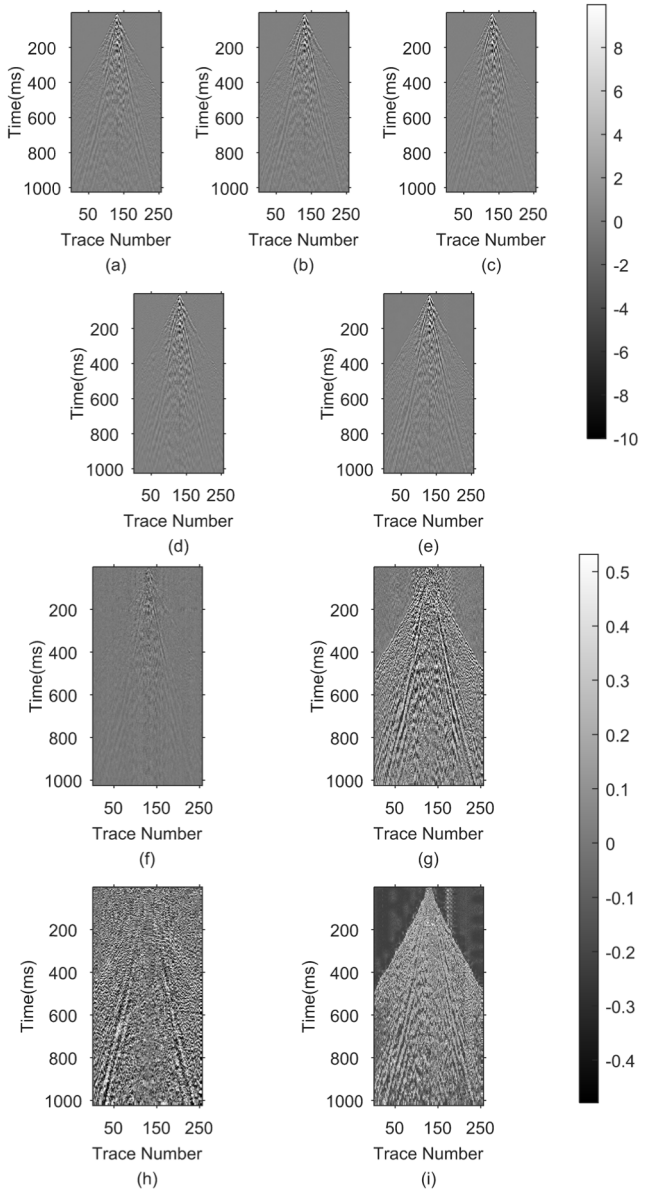


FIGURE 12. Measuring line II and its recover results with different methods. (a) Original seismic profile of measuring line II; (b) The reconstruction image (29.83dB) with CSA-GAN. (c) The reconstruction image (19.78dB) with SIODL. (d) The reconstruction image (16.27dB) with DPCA. (e) The reconstruction image (15.27dB) with JPEG2000. (f) The residual image of (b). (g) The residual image of (c). (h) The residual image of (d). (i) The residual image of (e).

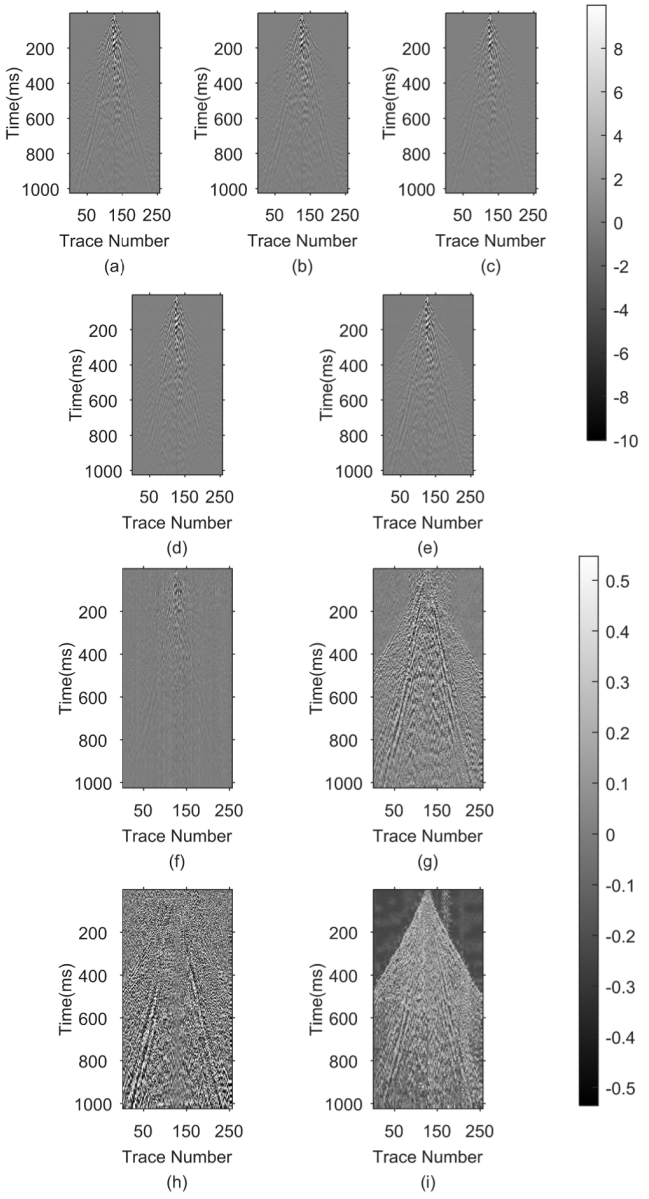


FIGURE 13. Measuring line III and its recover results with different methods. (a) Original seismic profile of measuring line III. (b) The reconstruction image (30.02dB) with CSA-GAN. (c) The reconstruction image (19.97dB) with SIODL. (d) The reconstruction image (16.04dB) with DPCA. (e) The reconstruction image (14.83dB) with JPEG2000. (f) The residual image of (b). (g) The residual image of (c). (h) The residual image of (d). (i) The residual image of (e).

analysis (DPCA), which is presented in [37]; the last baseline method is JPEG2000 Standard [38], which has been proved as an efficiency compression method for still image relying on the Cohen-Daubechies-Feauveau (CDF) 9/7 wavelet transform. In our experiment, all the three methods use the same compression ratio, which is 16.

To display recovery qualities of the four methods more clearly, only the most remarkable part of the data collected in the four measuring lines of this seismic exploration are shown in Figs11 to 14. Specifically, the data shown in the Fig. 11 (a), Fig. 12 (a), Fig. 13 (a) and Fig. 14 (a) are picked from the 256 sensors that are the nearest to the seismic source location in the four measuring lines. Besides, the sampling time selected to show is the first 1024ms after the seismic source exploded. This part of data has the most apparent features in the whole real seismic dataset of this exploration, since the energy of seismic source is decreasing with the increasing of spatial distance and temporal interval. Figs. 11 (b), (c), (d) and (e) are the reconstruction images of the profile Fig. 11 (a), which are recovered by different methods. Fig. 11 (b) is reconstructed by our method, while SIODL, DPCA and JPEG2000 are utilized in Figs. 11 (c), (d) and (e),

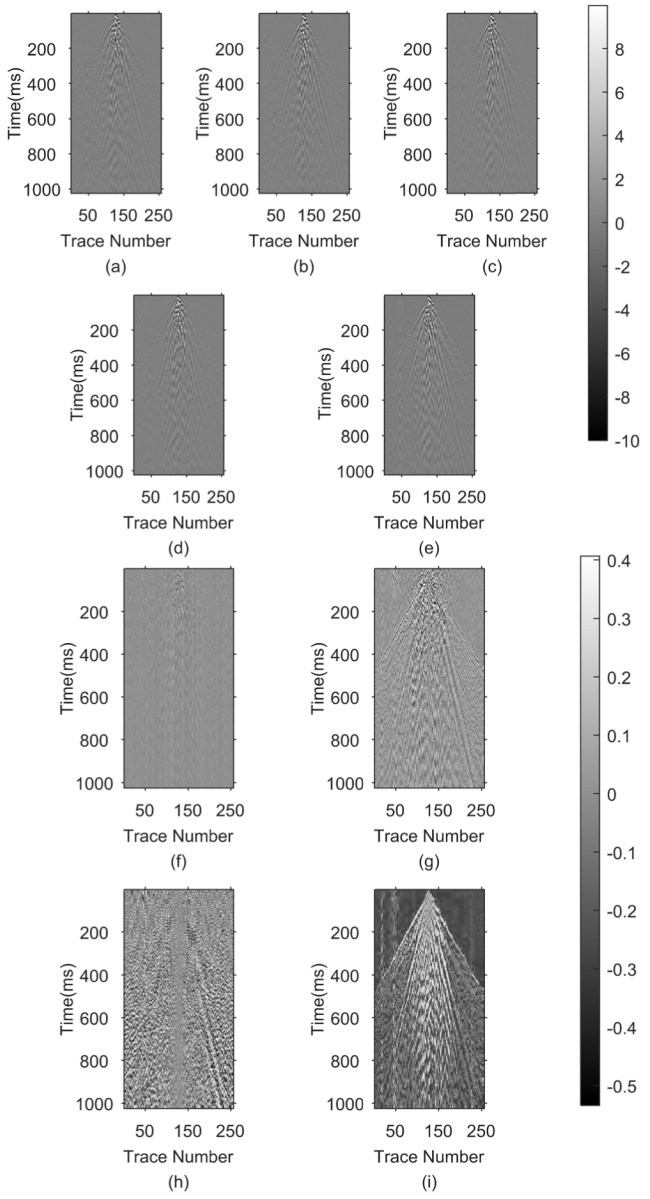


FIGURE 14. Measuring line IV and its recover result with different methods. (a) Original seismic profile of measuring line IV; (b) The reconstruction image (29.57dB) with CSA-GAN. (c) The reconstruction image (23.92dB) with SIODL. (d) The reconstruction image (21.02dB) with DPCA. (e) The reconstruction image (9.74dB) with JPEG2000. (f) The residual image of (b). (g) The residual image of (c). (h) The residual image of (d). (i) The residual image of (e).

respectively. Figs. 11 (f), (g), (h) and (i) are the residual images of Figs. 11 (b), (c), (d) and (e), respectively. Figs. 12 to 14 have the same format as Fig. 11. According to the four testing result figures, our reconstruction algorithm has a significantly better performance than the three benchmark methods. Broadly speaking, the reconstruction loss of our method is obviously less than that of the other three methods. Besides, as the residual images of our method illustrate, there is no structural information remained in the reconstruction error. In the SIODL's residual images, there is slight noise in the middle and bottom part of the seismic profile. In the residual images of DPCA, the noise is more obvious than that of SIODL, and only a narrow middle part is clean. JPEG2000's residual images have evident recovering errors distributed in the whole seismic profiles. In addition, the SNR values of our reconstruction results are higher and more stable than those of the three baseline results. The SNR values of our reconstruction results values are roughly about 30dB. By contrast, SIODL's SNR values range from 19dB to 23dB, which are the second-best result in the four methods. The SNR values of DPCA generally range from 16dB to 21dB, which are slightly worse than that of SIODL with respect to reconstruction quality and stability. JPEG2000's SNR values range from 9dB to 15dB, which are relatively low and fluctuant.

Based on the comparison of these four methods, the residual images show that our method owns not only a better recovering ability but also a more stable performance. Moreover, the residual images of JPEG2000 displays an obvious and global structure, which is similar to its original seismic profile. That means a lot of information has been lost, as the fixing dictionary used in JPEG2000 is not effective enough to represent seismic signal at that compression ratio. Since DPCA is a kind of data-driven method that can extract the most obvious components, there is nearly no noise in the middle part of seismic profile. However, lots of details in these seismic profiles have been lost, because this method only focuses on a few principal features. Although the dictionary of SIODL is self-adaptively learned from the input signal, there is still structural loss in several local regions of the residual images. In contrast to these three baseline methods, there is little structural noise in the whole residual images of our method, and this result certifies that generative adversarial network performs well in representing seismic data effectively and stably.

V. CONCLUSION

In this paper, we present a novel method, named CSA-GAN, for efficient seismic data acquisition by utilizing the sparsity and spatiotemporal features of seismic data. In this proposed method, compressed sensing theory and generative adversarial networks are combined to adaptively learn the effective representation of seismic data and compressively transmit seismic data with only negligible extra computation in the seismic terminals. Specifically, first, to tackle the data transmission bottleneck of traditional seismic acquisition network, a data collection architecture based on compressed sensing theory is designed and applied as the encoding procedure in CSA-GAN. Second, a GAN algorithm is used to learn the sparse transform domain and reconstruction mapping. Our approach is evaluated on both synthetic seismic data and real seismic data. The testing results of synthetic data show that our method has a stable training process and is able to handle seismic data with varied characters, when the compression ratio is 16. The experimental results of real seismic data show that CSA-GAN is able to stably get a high-quality seismic profile (about 30dB) a high compression ratio of 16,

compared with two state-of-the-art methods (SIODL and DPCA) and a standard compression method (JPEG2000). It means that the CSA-GAN can afford 15 times more sensors with the communication bandwidth and reduce the power cost by data collection from $N(N+1)/2$ to $N^2/16$, under the condition of only subtle noise introduced. Therefore, we could conclude that CSA-GAN is able to improve the efficiency of seismic data acquisition, which means it has great potential to be used in practice for large scale seismic data acquisition.

REFERENCES

[1] A. Mordret, T. D. Mikesell, C. Harig, B. P. Lipovsky, and G. A. Prieto, "Monitoring southwest Greenland's ice sheet melt with ambient seismic noise," *Sci. Adv.*, vol. 2, no. 5, May 2016, Art. no. e1501538.

[2] W. H. Bakun, B. Aagaard, B. Dost, W. L. Ellsworth, J. L. Hardebeck, R. A. Harris, C. Ji, M. J. S. Johnston, J. Langbein, J. J. Lienkaemper, A. J. Michael, J. R. Murray, R. M. Nadeau, P. A. Reasenberg, M. S. Reichle, E. A. Roeloffs, A. Shakal, R. W. Simpson, and F. Waldhauser, "Implications for prediction and hazard assessment from the 2004 parkfield earthquake," *Nature*, vol. 437, pp. 969–974, Oct. 2005.

[3] M. Malfante, M. D. Mura, J.-P. Metaxian, J. I. Mars, O. Macedo, and A. Inza, "Machine learning for volcano-seismic signals: Challenges and perspectives," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 20–30, Mar. 2018.

[4] F. Brenguier, N. M. Shapiro, M. Campillo, V. Ferrazzini, Z. Duputel, O. Coutant, and A. Nercessian, "Towards forecasting volcanic eruptions using seismic noise," *Nature Geosci.*, vol. 1, no. 2, pp. 126–130, Jan. 2008.

[5] B. Barker, M. Clark, P. Davis, M. Fisk, M. Hedlin, H. Israelsson, and R. North, "Monitoring nuclear tests," *Science*, vol. 281, pp. 1967–1968, Sep. 1998.

[6] A. Malehmir, R. Durrheim, G. Bellefleur, M. Urosevic, C. Juhlin, D. J. White, B. Milkereit, and G. Campbell, "Seismic methods in mineral exploration and mine planning: A general overview of past and present case histories and a look into the future," *Geophysics*, vol. 77, no. 5, pp. WC173–WC190, Sep. 2012.

[7] A. Payani, A. Abdi, X. Tian, F. Fekri, and M. Mohandes, "Advances in seismic data compression via learning from data: Compression for seismic data acquisition," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 51–61, Mar. 2018.

[8] F. J. Herrmann, M. P. Friedlander, and O. Yilmaz, "Fighting the curse of dimensionality: Compressive sensing in exploration seismology," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 88–100, May 2012.

[9] S. Savazzi, U. Spagnolini, L. Goratti, D. Molteni, M. Latva-aho, and M. Nicoli, "Ultra-wide band sensor networks in oil and gas explorations," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 150–160, Apr. 2013.

[10] R. G. Baraniuk and P. Steeghs, "Compressive sensing: A new approach to seismic data acquisition," *Lead. Edge*, vol. 36, no. 8, pp. 642–645, 2017.

[11] C. Mosher, C. Li, L. Morley, Y. Ji, F. Janiszewski, R. Olson, and J. Brewer, "Increasing the efficiency of seismic data acquisition via compressive sensing," *Lead. Edge*, vol. 33, no. 4, pp. 361–476, Apr. 2014.

[12] R. Zhang, Q. Hu, G. Wang, and B. Ye, "Distributed compressed sensing of microseismic signals through first break time extraction and signal alignment," *IEEE Access*, vol. 6, pp. 27408–27417, Apr. 2018.

[13] A. S. Spanias, S. B. Jonsson, and S. D. Stearns, "Transform methods for seismic data compression," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 3, pp. 407–416, May 1991.

[14] J. Ma, G. Plonka, and H. Chauris, "A new sparse representation of seismic data using adaptive easy-path wavelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 540–544, Jul. 2010.

[15] S. Fomel and Y. Liu, "Seislet transform and seislet frame," *Geophysics*, vol. 75, no. 3, pp. V25–V38, May 2010.

[16] E. Liu, A. Payani, and F. Fekri, "Seismic data compression using online double-sparse dictionary learning schemes," in *Proc. Data Compress. Conf. (DCC)*, Apr. 2017, p. 449.

[17] A. Abdi, A. Payani, and F. Fekri, "Learning dictionary for efficient signal compression," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 3689–3693.

[18] A. B. Kiely, M. Xu, W.-Z. Song, R. Huang, and B. Shirazi, "Adaptive linear filtering compression on realtime sensor networks," *Comput. J.*, vol. 53, no. 10, pp. 1606–1620, Dec. 2010.

[19] R. Huang, W. Z. Song, M. Xu, N. Peterson, B. Shirazi, and R. LaHusen, "Real-world sensor network for long-term volcano monitoring: Design and findings," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 2, pp. 321–329, Feb. 2012.

[20] F. J. Herrmann, "Randomized sampling and sparsity: Getting more information from fewer samples," *Geophysics*, vol. 75, no. 6, pp. WB173–WB187, Nov. 2010.

[21] S. M. O'Connor, J. P. Lynch, and A. C. Gilbert, "Compressed sensing embedded in an operational wireless sensor network to achieve energy efficiency in long-term monitoring applications," *Smart Mater. Struct.*, vol. 23, no. 8, Jul. 2014, Art. no. 085014.

[22] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[23] C. C. Mosher, S. T. Kaplan, and F. D. Janiszewski, "Non-uniform optimal sampling for seismic survey design," in *Proc. 74th EAGE Conf. Exhibit. Incorporating EUROPEC*, Jun. 2012, pp. 3953–3957.

[24] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[25] M. A. N. Siahsar, S. Gholtashi, A. R. Kahoo, W. Chen, and Y. Chen, "Data-driven multitask sparse dictionary learning for noise attenuation of 3D seismic data," *Geophysics*, vol. 82, no. 6, pp. V385–V396, Oct. 2017.

[26] B. She, Y. Wang, J. Liang, Z. Liu, C. Song, and G. Hu, "A data-driven amplitude variation with offset inversion method via learned dictionaries and sparse representation," *Geophysics*, vol. 83, no. 6, pp. R725–R748, Nov. 2018.

[27] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[28] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[29] 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, Jan. 2011.

[30] 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.

[31] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "ReconNet: Non-iterative reconstruction of images from compressively sensed measurements," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jan. 2016, pp. 449–458.

[32] V. Singhal, A. Majumdar, and R. K. Ward, "Semi-supervised deep blind compressed sensing for analysis and reconstruction of biomedical signals from compressive measurements," *IEEE Access*, vol. 6, pp. 545–553, Dec. 2017.

[33] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2014, pp. 2672–2680.

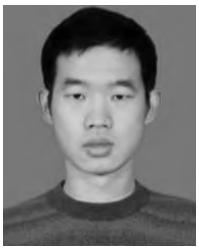
[34] V. Papyan, Y. Romano, J. Sulam, and M. Elad, "Theoretical foundations of deep learning via sparse representations: A multilayer sparse model and its connection to convolutional neural networks," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 72–89, Jul. 2018.

[35] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "On the effectiveness of least squares generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published. doi: 10.1109/TPAMI.2018.2872043.

[36] Y. Zhu, N. Li, F. Sun, J. Lin, and Z. Chen, "Design and application of a borehole–surface microseismic monitoring system," *Instrum. Sci. Technol.*, vol. 45, no. 3, pp. 233–247, Mar. 2017.

[37] B. Liu, M. Mohandes, H. Nuha, M. Deriche, and F. Fekri, "A distributed principal component analysis compression for smart seismic acquisition networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3020–3029, Jun. 2018.

[38] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 36–58, Sep. 2001.



XIAOPU ZHANG received the M.Eng. degree in instrumentation science and technology from Jilin University (JLU), China, in 2016, where he is currently pursuing the Ph.D. degree with the College of Instrumentation and Electrical Engineering.

His research interests include machine learning and signal processing. Specifically, he devotes himself to develop efficient algorithms for data compression and analysis in the edge devices.

SHUAI ZHANG received the B.S. degree from the College of Optoelectronic Engineering, Changchun University of Science and Technology, China, in 2017. He is currently pursuing the M.Eng. degree with the College of Instrumentation and Electrical Engineering, Jilin University.

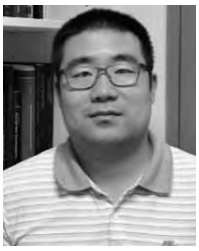
His research interests include seismic data transmission networks and signal processing.

JUN LIN received the B.S. and M.S. degrees in applied geophysics from the Changchun College of Geosciences, Changchun, China, in 1982 and 1987, respectively.

Since 2005, he has been the Chairman with the College of Instrumentation and Electrical Engineering, Jilin University, Changchun, where he is currently a Professor. He has authored ten books and more than 300 papers. His research interests include seismic instrument, nuclear magnetic resonance instrument, and electromagnetic instrument developments. He has received two key national invention awards as the Chief Director.

FENG SUN received the B.S. degree from the College of Electronic Science and Engineering, Jilin University, China, in 2003, and the Ph.D. degree from the College of Instrumentation and Electrical Engineering, Jilin University, in 2009.

He is currently an Associate Professor with the College of Instrumentation and Electrical Engineering, Jilin University. His research interests include development of mm-wave circuits, microseismic monitoring, microseismic location, and systems designs.



XI ZHU received the B.E. degree (Hons.) and the Ph.D. degree from the University of Hertfordshire (UH), Hertfordshire, U.K., in 2005 and 2008, respectively.

He is currently a Lecturer with the School of Computing and Communication, University of Technology Sydney, NSW, Australia. He has coauthored over 80 refereed publications in the abovementioned fields. His research interests include in the areas of analogue baseband, radio frequency (RF), and mm-wave circuit and system design.

YANG YANG (S'11–M'14–SM'17) received the Ph.D. degree from Monash University, Melbourne, Australia, in 2013.

From April to December 2016, he was a Research Fellow with the City University of Hong Kong. In December 2016, he joined the University of Technology Sydney, Australia. His research interests include RFIC, microwave and millimetre-wave circuits and systems, and millimetre-wave antennas. He is a winner of the CST University Publication Award 2018, by CST, Germany. He is currently an Associate Editor of IEEE Access.

XUNQIAN TONG received the Ph.D. degree from Jilin University, China, in 2016. He was a Cotutelle Student with Macquarie University, Australia, through the support from CSC, from 2014 to 2016.

He is currently a Lecturer with the College of Instrumentation and Electrical Engineering, Jilin University. His current research interests include routing protocol design and 5G spectrum analysis.

HONGYUAN YANG received the B.S. degree from the College of Electronic Science and Engineering, Jilin University, China, in 2003, and the Ph.D. degree from the College of Earth Exploration Science and Technology, Jilin University, in 2009.

He is currently an Associate Professor with the College of Instrumentation and Electrical Engineering, Jilin University. His research interests include development of seismic exploration instrument, wireless seismic data acquisition, and seismic data processing.