Deep Seismic CS: A Deep Learning Assisted Compressive Sensing for Seismic Data

Naveed Iqbal, Mudassir Masood, *Member, IEEE*, Motaz Alfarraj, *Member, IEEE*, and Umair Bin Waheed

***Abstract*— For large-scale seismic exploration in areas that lack even basic infrastructure, wired geophones are impractical because of the huge effort involved and their high deployment and operating costs. A network of wireless geophones capable of recording and transmitting data could be an inexpensive solution. However, a typical seismic survey can generate hundreds of terabytes of raw seismic data per day. It takes a huge amount of energy to transmit this massive amount of data from geophones to the on-site data collection center, thus making the transformation from pre-wired to wireless geophones a significant challenge. To reduce data traffic to the data center without putting additional strain on the geophone, a standalone and lightweight compressive sensing (CS) method is proposed in this work. The method takes advantage of the inherent sparsity in the seismic data to enable the geophone to sense data in a compressed manner. This significantly reduces the amount of data that needs to be recorded/transmitted by the geophone, making it energy-efficient. However, instead of employing conventional optimization-based CS reconstruction methods, we propose an efficient implementation of a deep convolutional neural network (DCNN). This network processes the compressed data received at the collection center without any a priori assumptions about the underlying seismic signal statistics, making it appropriate for a wide range of seismic data. The use of CS for energy-efficient sensing and transmission combined with powerful DCNN for reconstruction yields a system that could achieve signal-to-noise ratio (SNR) of around 30 dB with a compression gain of 16 on a field dataset. Finally, when compared with existing methods, the proposed approach demonstrates significant superiority in maximizing compression gain and reconstruction quality for both synthetic and real-field datasets.**

***Index Terms*— Artificial intelligence (AI), compressive sensing (CS), deep neural networks, machine learning, seismic data.**

I. Introduction

LARGE arrays of sensors are used in oil and gas exploration [1]. These sensors record seismic signals produced by natural earthquakes (passive seismic) or artificial seismic sources (active seismic). These sensors are called geophones when used for inland exploration and hydrophones when deployed offshore. The signal recorded by a geophone/hydrophone is called a seismic trace. The raw seismic data acquired during exploration are transmitted to the in-field data center for further processing/storing.[1]

The Shannon–Nyquist sampling theorem is a fundamental theorem in the field of signal processing and states that a band-limited analog signal is fully determined by a set of samples taken at a rate that is more than twice the maximum frequency component in it. Violation of this principle may result in detrimental effects due to aliasing. Furthermore, driven by the desire to reconstruct high-resolution subsurface images, the seismic industry often oversamples the seismic traces. As a consequence, the number of sensors required and the amount of data to be collected per geophone increase significantly. This strains the efforts that need to be exerted for successful seismic acquisition and results in high costs in terms of time and money. Moreover, data storage becomes an issue and necessitates the use of compression. Data compression, in turn, opens the doors to several different artifacts in the data and might also result in information loss. In addition to that, seismic data recording/transmission consumes significant amounts of energy, which is of great concern in the case of battery-driven wireless geophones. Plagued with these shortcomings, the seismic acquisition process could be revolutionized by taking advantage of developments in the fields of compressive sensing (CS) and machine learning.

Furthermore, as the world continues to evolve toward wireless seismic data collection [2], [3], wireless channel bandwidth restrictions become a barrier to the density and scale of sensors that can be installed, hence lowering the Earth's subsurface sampling density. In response to these challenges, previous studies have espoused incorporating a form of artificial intelligence (AI), i.e., computing power, to compress data at field geophones. Specifically, it was proposed in [4], [5], [6], [7], and [8] to use high-performance components to enable on-geophone data compression. However, the extra computing capability introduces a burden on battery-driven wireless geophones and is not practical for most situations. It is important to note that these compression methods are different from acquiring compressed measurements as described by the CS theory. In this work, the focus is on the in-field

Manuscript received 20 March 2022; revised 6 October 2022 and 6 April 2023; accepted 16 May 2023. Date of publication 7 July 2023; date of current version 27 July 2023. This work was supported by the Deanship of Research Oversight and Coordination (DROC), King Fahd University of Petroleum and Minerals (KFUPM), under Project GTEC2013. *(Corresponding author: Naveed Iqbal.)*

Naveed Iqbal and Mudassir Masood are with the Department of Electrical Engineering and the Center for Communication Systems and Sensing, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia (e-mail: naveediqbal@kfupm.edu.sa; mudassir@kfupm.edu.sa).

Motaz Alfarraj is with the Department of Electrical Engineering and the SDAIA-KFUPM Joint Research Center for Artificial Intelligence, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia (e-mail: motaz@kfupm.edu.sa).

Umair Bin Waheed is with the Department of Geoscience, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia (e-mail: umair.waheed@kfupm.edu.sa).

Digital Object Identifier 10.1109/TGRS.2023.3289917

[1]Please note that while the term geophone is used most of the time, the approach also applies to hydrophones.

CS approach where we relax the processing power at geophones by eliminating complex compression algorithms and processes. The compression is achieved by acquiring compressed observations directly at a geophone and then utilizing a deep convolutional neural network (DCNN) at the data center to recover high-quality seismic signals. In this setup, the complex computations are offloaded to the data center, where computing power is not an issue. Furthermore, the transmission of compressed seismic data reduces the power requirements on geophones significantly and even allows for more efficient data transfer using a limited resource wireless channel. This is in contrast with the current seismic acquisition systems, which presume passive seismic sensors.

In this work, CS is used at the geophone to reduce the dimensions of the acquired data, while DCNN with 1-D convolutional (Conv) layers is utilized at the data center to recover seismic time series with a high signal-to-noise ratio (SNR). Hence, the proposed approach is termed deep seismic CS (DSCS). Unlike analytical CS recovery methods, for which the reconstruction problem is explicitly defined as the optimization process, machine learning techniques take advantage of large datasets in order to learn the key reconstruction parameters and priors. In particular, deep learning is used to develop a robust CS-inspired acquisition and reconstruction process from the existing datasets in advance. This allows us to benefit from the advantages provided by CS as well as the fast and efficient reconstruction features of deep learning techniques.

CS acquisition has been utilized in other fields, such as in medical for MRI images [9] and in radar for navigation [10]. In these fields, the images are reconstructed from the spatially sparse measurements. To the best of our knowledge, this is the first study to implement CS at the geophone level and integrate it with the 1-D DCNN for efficient seismic data acquisition. Our observations demonstrate that the proposed simplified approach produces a processing model that is very efficient and avoids computationally expensive estimation steps at the geophone. Moreover, the envisaged setup is generic, and it can be used for a variety of remote sensing applications [11], [12], not just seismic acquisition.

The main contributions of this work are highlighted below.

1) This study aims to use the relevant CS theory to highlight a practical acquisition strategy that is suitable for sampling seismic signals at a low rate. For this, various CS acquisition strategies are discussed and compared using synthetic and real seismic datasets.
2) Seismic signals are known to be sparse in the frequency domain. This property is utilized to design a DCNN-based reconstruction approach.
3) A detailed comparison of various CS reconstruction techniques in a seismic context as well as from a practical perspective is carried out.

II. CS Approaches and Limitations

CS-based acquisition strategies also pave the way for wireless geophones. It is well known that more than 50% of the seismic acquisition cost is related to the transport and maintenance of the heavy cables through which geophones are connected. While battery-powered wireless geophones are an answer to this problem, the prohibitive amount of energy required to operate high-precision analog-to-digital converters (ADCs), data processing, and transmission of data to central storage makes it impossible to materialize the idea. Although wireless geophone solutions exist [13], their high-energy requirement, small communication range, and low-data-rate transmission restrict their widespread adoption [14]. CS acquisition is not dependent on high-precision ADCs and results in data that are compressed inherently [15]. Therefore, the power required for sensing, acquisition, and transmission is reduced significantly. In addition to addressing the challenges highlighted above, the amount of data storage needed is also reduced. Thus, CS can be considered an enabler for wireless geophone technology.

In the ensuing, the CS acquisition techniques, sensing methods, and recovery procedures are briefly discussed.

A. CS Acquisition Strategies

CS replaces the requirements of high-resolution sampling and data compression by combining the two steps into a single low-resolution acquisition step. CS exploits the fact that a small set of nonadaptive linear measurements of a compressible signal carries enough information for reconstruction and processing [16], [17]. More specifically, CS acquisition involves randomized subsampling that turns the artifacts caused by coherent subsampling into easy-to-counter incoherent Gaussian noise. Seismic signals are compressible, and therefore, CS can be used to replace conventional sensing strategies [18]. To elaborate further, consider a real-valued discrete-time signal $\mathbf{x} \in \mathbb{R}^N$ representing a vectorized signal that has been recorded at a particular geophone. We may express the signal $\mathbf{x}$ as a linear combination of the basis vectors that form the columns of the $N \times N$ matrix $\mathbf{\Psi}$ as follows:

$$\mathbf{x} = \mathbf{\Psi}\mathbf{s}. \tag{1}$$

Note that $\mathbf{s}$ is another representation of the signal $\mathbf{x}$. In the context of CS, $\mathbf{\Psi}$ would be a sparsifying basis implying a sparse $\mathbf{s}$; i.e., $\mathbf{s}$ has only $K \ll N$ nonzero entries. In other words, $\mathbf{x}$ can be represented as a linear combination of just a few (K) columns of the $\mathbf{\Psi}$ matrix. Thus, we say that $\mathbf{x}$ has a sparse representation in the $\mathbf{\Psi}$ domain. The linear measurements $\mathbf{y} \in \mathbb{R}^M$ of $\mathbf{x}$ are given by

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{s} = \mathbf{\Theta}\mathbf{s} \tag{2}$$

where $\mathbf{\Phi}$ is an $M \times N$ measurement or sensing matrix that is independent of $\mathbf{x}$ to perform nonadaptive measurements. The advantage of CS is that a high-dimensional (N-dimensional) vector $\mathbf{x}$ can be recovered from a much smaller M-dimensional ($M < N$) observation vector $\mathbf{y}$. It was shown [16], [17] that if $\mathbf{\Phi}$ is made up of components drawn from a Gaussian density in an independent and identically distributed (i.i.d.) fashion, the measurements $\mathbf{y}$ would carry enough information to reconstruct $\mathbf{x}$.

A necessary and sufficient condition for reconstructing $\mathbf{x}$ is to satisfy

$$1 - \epsilon \leq \frac{||\mathbf{\Theta}\mathbf{s}||_2}{||s||_2} \leq 1 + \epsilon$$

for some $\epsilon > 0$. This condition is called the restricted isometry property (RIP). This condition ascertains that the matrix $\mathbf{\Theta}$ preserves the lengths of sparse vectors $\mathbf{s}$ when moving to the lower dimensional measurements domain. Satisfaction of this condition to a certain degree is desired for a successful reconstruction of $\mathbf{x}$ from $\mathbf{y}$. Fortunately, this property is satisfied by the random Gaussian matrix in addition to other structured and unstructured sensing matrices, as discussed in Section II-B.

B. Sensing/Measurement Matrix and Realization

The efficient design of sensing matrices is one of the most important aspects of the CS paradigm. Candes and Tao [19] and Lu et al. [20] proposed sensing matrices composed of i.i.d. Gaussian and Bernoulli random entries, respectively. However, reasonable use of these sensing matrices necessitates enormous computational expense and memory stockpiling requirements, making them unsuitable for large-scale applications. Lately, Gilbert and Indyk [21] constructed sparse matrices as an alternative to random dense matrices. These matrices, on the other hand, are unstructured, which makes hardware implementation difficult. The deterministic class of sensing matrices is created using partial Fourier matrices [22]. Regardless of the fact that deterministic sensing matrices are insufficient to meet the required RIP criteria, they are useful due to the sampler's deterministic concept and may achieve some compression ratio and less computational complexity. Another class is developed by Bajwa et al. [23] using Toeplitz and circulant-structured random sensing matrices. These matrices perform similar to i.i.d. CS matrices, such as Gaussian matrices, are very likely to satisfy the RIP.

Several practical implementations of CS mechanisms have emerged in the literature. The circuits designed for this purpose are based on the fundamental concept of aliasing the spectrum into baseband. Once the frequency content of signals is spread to low frequencies, the signals can be sampled using a lower rate ADC. One of the earliest such proposals is the random demodulator (Fig. 1) [24]. In this approach, the wideband input signal $x(t)$ is first multiplied with a pseudorandom sequence made up of $\pm 1'$ s. This is also known as chipping sequence $p_c(t)$. This operation is equivalent to convolution in the frequency domain and results in spreading the signal information all over the spectrum. After multiplication, the signal is passed through a low-pass filter and then sampled using a low-rate ADC. As the information was spread all over the spectrum, the resulting small number of samples $\mathbf{y}$ (termed measurements) contains all the useful information in the signal $\mathbf{x}$. Thus, $\mathbf{y}$ represents the compressed version of the high-dimensional seismic data. These samples can now be easily stored, transmitted, and processed. This opens up several possibilities for further enhancing the resolution of seismic data as well as shortening the processing and interpretation time of the data.

Completely random sensing matrices, such as those proposed by Candes and Tao [19] and Lu et al. [20], achieve optimal incoherence. However, Romberg [25], in his method of random convolution, proved that random circulant sensing matrices are equally effective as these matrices for signal recovery. The structured circulant random sensing matrix allows efficient hardware implementation [26] and results in faster and easier acquisition and decoding of signals. Several other circuits have been designed to acquire sparse signals in a compressed manner [27].

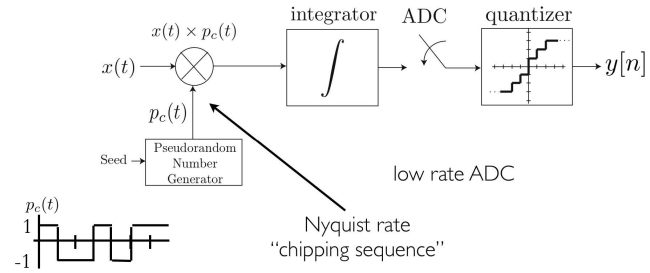


Fig. 1. Random demodulator [24] to be used in geophone.

Once geophones/hydrophones have acquired seismic traces in a compressed manner, the compressed signals need to be processed in order to restore the measurements.

C. CS Recovery Approaches

Once the measurement vector $\mathbf{y}$ has been acquired, the signal $\mathbf{x}$ or its sparse representation $\mathbf{s}$ needs to be reconstructed. A number of CS reconstruction approaches exist. Specifically, the following optimization problem can be solved:

$$\hat{\mathbf{s}} = \arg\min ||\mathbf{s}||_0 \quad \text{s.t. } \mathbf{y} = \mathbf{\Theta s}. \tag{3}$$

However, this is an NP-complete problem, and solving it is computationally prohibitive due to the ℓ_0 norm. It was shown that using ℓ_1 norm also provides exact recovery of the sparse vector $\mathbf{s}$

$$\hat{\mathbf{s}} = \arg\min ||\mathbf{s}||_1 \quad \text{s.t. } \mathbf{y} = \mathbf{\Theta s}. \tag{4}$$

This is a convex optimization problem known as basis pursuit. This type of convex optimization problem can be solved efficiently using low-complexity greedy search algorithms. These are iterative algorithms with reconstruction guarantees and are popular among the CS community.

While we propose a DCNN-based solution for the recovery of sparse representation of the seismic traces, we compare its performance with other existing approaches. Specifically, we have selected two greedy algorithms, namely, orthogonal matching pursuit (OMP) [28], [29] and support agnostic Bayesian matching pursuit (SABMP) [30], [31], for this purpose. These algorithms are briefly described in the following.

1) Orthogonal Matching Pursuit: OMP estimates the nonzero coefficients of the unknown sparse vector $\mathbf{s}$ one at a time. In every iteration, the support of $\mathbf{s}$ grows by one, as the algorithm selects one element (ith column) from $\mathbf{\Phi}$ that best correlates with the residual measurements vector $\mathbf{r}$. Note that the initial value of the residual measurement vector is simply the original measurement vector, i.e., $\mathbf{r} = \mathbf{y}$. In every step, all the previously selected elements of $\mathbf{\Phi}$ are used to solve a least squares (LS) problem to estimate a provisional solution of $\mathbf{s}$. The contribution of the approximated $\mathbf{s}$ is removed from the measurement vector before proceeding to the next

iteration. The process is repeated until the residual becomes small enough or a maximum number of iterations has been reached according to prespecified criteria.

2) Support Agnostic Bayesian Matching Pursuit: SABMP is used to perform Bayesian estimation of the unknown sparse signal. Bayesian approaches assume a prior distribution; however, given the complex and highly varying nature of the subsurface medium, it is usually impossible to characterize the distribution of the nonzero elements of $\mathbf{s}$. Furthermore, even if the distribution is known, it is very difficult to estimate the distribution parameters. SABMP provides the advantages of Bayesian estimation and yet is agnostic to the distribution of the unknown elements of $\mathbf{s}$. Therefore, the use of distribution agnostic Bayesian sparse signal recovery in SABMP is quite attractive and has shown robust performance in scenarios where the prior is non-Gaussian or unknown.

The algorithm computes the minimum mean square error (MMSE) estimate of $\mathbf{s}$ given the measurements $\mathbf{y}$ as shown below in an iterative manner

$$\hat{\mathbf{s}}_{\text{MMSE}} \triangleq \mathbb{E}\big[\mathbf{s} \mid \mathbf{y}\big] = \sum_{\mathcal{S}} p(\mathcal{S} \mid \mathbf{y}) \mathbb{E}\big[\mathbf{s} \mid \mathbf{y}, \mathcal{S}\big]. \tag{5}$$

Finding the MMSE estimate requires computing the expectation $\mathbb{E}[\mathbf{s} \mid \mathbf{y}, \mathcal{S}]$, where $\mathcal{S}$ is the set of support sets of $\mathbf{s}$ with significant posteriors. Since the distribution of $\mathbf{s}$ is unknown, the computation of $\mathbb{E}[\mathbf{s} \mid \mathbf{y}, \mathcal{S}]$ may be prohibitive, and it is replaced with the best linear unbiased estimate (BLUE). Furthermore, as the distribution of $\mathbf{s}$ is unknown, the computation of the posterior $p(\mathcal{S} \mid \mathbf{y})$ is done by annihilating the non-Gaussian component of $\mathbf{y}$ by projecting it onto the orthogonal complement space of $\mathbf{\Theta}$. The resultant greedy algorithm also builds the unknown sparse vector in an iterative fashion. However, in each step, the abovementioned strategy of MMSE estimation is used for estimating the nonzero elements of $\mathbf{s}$. In SABMP, the iterations stop when the change in the support size (number of nonzero elements in the sparse vector) in two subsequent iterations is not big enough (less than 2% [31]). Specifically, $(|R_i - R_{i-1}|/R_{i-1}) < 0.02$, where the sparsity rate is estimated in each iteration as $R_i = ||\mathbf{s}_i||_0/N$ and $\mathbf{s}_i$ is the estimate of the sparse vector in the ith iteration.

Multiple factors affect the choice of CS reconstruction algorithms. For example, seismic signals may not be perfectly sparse (i.e., compressible). Furthermore, even if they are sparse, their sparsity level cannot be determined. In addition, the noise levels and noise types (characteristics) may vary from one seismic acquisition setting to another. Most CS reconstruction algorithms make assumptions regarding one or more of these factors. Thus, a solution that is robust to these factors is needed for a practical CS seismic setup. Finally, the compression achieved by using the traditional CS recovery methods is not significant. All of these serve as motivation for the proposed DCNN-based recovery strategy given in this study.

III. Recovery Using DCNN

This section describes the 1-D DCNN layers and the training procedure for recovering compressed seismic measurements.

A. Architecture of DCNN

A 1-D DCNN is designed in this study as a cascade of Conv layers with descending (analysis path for encoder) and ascending (synthesis path for decoder) sizes. This is motivated by autoencoders, which have the ability to learn a sparse representation of the data in order to achieve an optimization goal. The number of layers in the encoder and decoder, the number of filters, and the dimensions of the filters are all customized for our application using sensitivity analysis. The DCNN used in this work is depicted in Fig. 2 and described in the sequel.

The network is trained to recover the uncompressed seismic trace from the compressed one. Hence, the network input is the compressed observation (e.g., length of 256 samples), and the output is an uncompressed seismic trace (e.g., the original length of 4096 samples). Furthermore, the output is estimated in the discrete cosine transform (DCT) domain. The reason for not recovering the time-domain signal directly using DCNN is twofold. First, seismic signals are known to be sparse in the frequency domain. Second, the time-domain traces need to be aligned [8] to maximize the performance, which is impractical. The network setup has seven Conv layers, three TConv layers, and a fully connected (FC) layer.

The Conv and TConv layers have the filter sizes of 3×1 each, and the number of filters ranges from 8 to 128, except for the last Conv layer, which has one filter of size 1×1. This Conv layer offers channelwise pooling, often known as feature map pooling or a projection layer. This simplistic approach is adopted for dimensionality reduction, i.e., decreasing the number of feature maps while retaining their salient features. The final layer of the network is an FC layer with 4096 neurons. Each one of the aforementioned layers is followed by the parametric leaky rectified linear unit (PLU) layer, except the last Conv and FC layers. The authors conclude from sensitivity analysis that the rectified linear unit (ReLU) layer does not perform well, because it has a zero slope for negative values. As a result, leaky ReLU can be used. However, in the case of a leaky ReLU, the parameter used for the negative slope must be manually adjusted. Hence, PLU is preferred to automatically deal with the negative slope. PLU is a type of leaky ReLU that, rather than having a predetermined slope such as 0.01, makes it a parameter for the neural network to figure out on its own. The activation function PLU is defined as follows:

$$f(z) = \begin{cases} \alpha z, & \text{for } z < 0 \\ z, & \text{for } z \geq 0 \end{cases} \tag{6}$$

where z represents the input to the activation layer.

The stride value of Conv layers is set to either 1 or 2 in the vertical direction, depending on the requirement for downsampling. Padding is added in order to have the same input and output size in the case of a stride equal to 1. The output is downsampled by an amount $\lceil (N_{\text{Size}}/N_{\text{St}}) \rceil$ when the stride is greater than 1, where N_{Size} is the input's size in the horizontal direction and N_{St} is the value of stride. If the padding has an even value, the same amount of padding is added to the top and bottom. Otherwise, it is added at the bottom only.

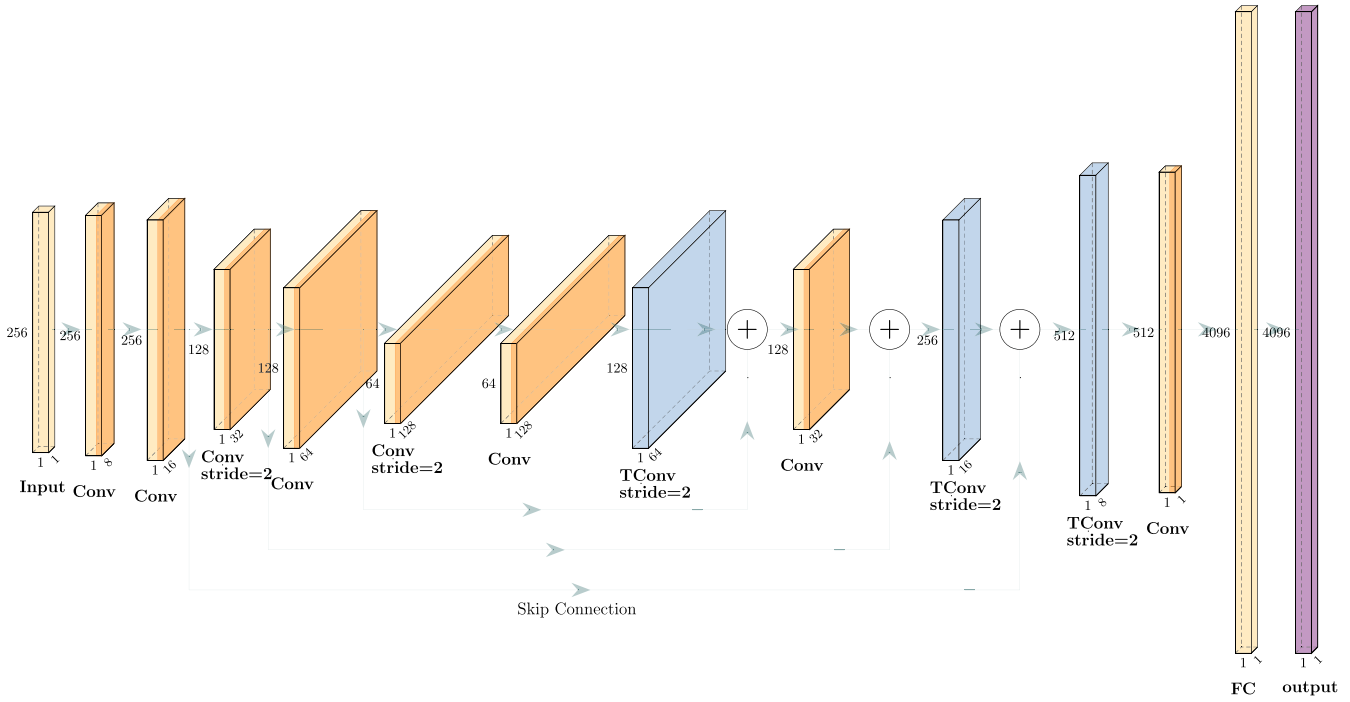


Fig. 2. Pictorial view of the 1-D DCNN. The dimension of each layer is indicated for a compression ratio of 16. The input is the compressed observation samples, whereas the output is the recovered trace in the DCT domain. The information is presented as "time samples × 1 × number of channels." For downsampling, the 1-D input data are first passed through 3 × 1 Conv layers with a stride of size 2, followed by upsampling through transpose convolutional (TConv) layers. The kernel size of the last Conv layer is 1 × 1, whereas the last layer of the network is an FC layer. The PLU is used after each layer except the last one. Skip connections are added to improve convergence during the training phase.

Similarly, in the case of TConv layers, depending on whether upsampling is required in the vertical direction, the stride is set to 1 or 2. The stride parameter determines how far the input is stretched in this case. A reduction in output size, i.e., cropping, is used to make the output size equal to $N_{\text{Size}} * N_{\text{St}}$.

The performance of deep neural nets decreases, as the architecture's depth increases. This is referred to as the degradation issue. The reason is that the gradient becomes vanishingly small, effectively preventing the weight from changing its value. As recommended by He et al. [32], skip connections are used in this study to improve predictive accuracy and learning convergence. While using skip connections, the signal from a specific layer is added to the output of a higher layer in the stack.

B. Parameters

For optimization, the following parameters are used: the weights of all the layers are initialized as stated in [33], whereas the biases are set to zero. The network is trained using the adaptive moment estimation optimizer [34], a 128 mini-batch size, and backpropagation with the gradient descent algorithm. The learning rate is set to $\text{lr} = 10^{-3}$ with $\epsilon = 10^{-8}$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$ [33]. The loss function is given as follows:

$$\mathbb{E}(\theta) = \sum_{i=1}^{4096} \frac{1}{2}\left[\mathcal{X}(i) - \hat{\mathcal{X}}(i)\right]^2 + \frac{\lambda}{2}\mathbf{w}\mathbf{w}^T \tag{7}$$

where $\hat{\mathcal{X}}(i)$ and $\mathcal{X}(i)$ denote the ith sample of the network's prediction and target output, respectively. In relation to the seismic data, $\mathcal{X}(i)$ $[\hat{\mathcal{X}}(i)]$ represents the ith DCT bin of ground-truth trace (estimated trace). A regularization term $\lambda = 0.00005$ is added to the loss function in order to reduce the overfitting problem [35].

IV. Training of DCNN

To train the DCNN, synthetic data are generated using the Marmousi model [36], which is commonly used as a standard case study in exploration seismology. The typical acquisition geometry used to generate the data consists of inline geophones with an interstation gap in the horizontal axis, and shots are generated sequentially at the same location as geophones. The seismic data are generated using the seismic source signature, i.e., Ormsby wavelet [37] with frequency $0 - 15 - 80 - 90$ Hz. The Ormsby wavelet has a trapezoidal frequency spectrum, which allows for more frequency-domain flexibility as compared with the Ricker wavelet. This suits well for both modern surveys and old land data because of the broader set of frequencies and more of a trapezoidal spectral shape. This aids in the reconstruction of seismograms comparable to real seismic data. The sampling frequency used in this work is 4 kHz. The MATLAB package [38] is used for the generation of synthetic data.

With the setup described above, approximately ten million traces are generated, with each trace containing 4096 samples. The survey consists of 10 000 in-line geophones, and the

distance between two successive geophones is 3 m. There are 1000 shots that are carried out at various random geophone locations. The target of the DCNN is the generated synthetic traces in the DCT domain, i.e., **s**, whereas the predictor is the observation **y** as in (2). The sparsifying basis Ψ is the Fourier matrix in the case of recovering using the CS recovery approaches. However, it is taken as DCT matrix for DCNN in this work, and the reason is to ease the DCNN by using only real numbers. This overall approach is termed DSCS, i.e., DSCS.

The training data are randomized before being fed to the DCNN and then shuffled after each epoch. The predictors **P** and targets **T** are z-normalized as follows:

$$\bar{\mathbf{P}} = \frac{\mathbf{P} - \mu_p}{\sigma_p}, \quad \bar{\mathbf{T}} = \frac{\mathbf{T} - \mu_t}{\sigma_t} \tag{8}$$

where μ and σ represent the mean and variance, respectively. The testing dataset for evaluation purposes is normalized by the mean and variance of the training dataset, so that the actual amplitude can be recovered. Moreover, 10% of the generated synthetic data is being used for validation. Training is considered complete when the validation loss exceeds or equals the previously lowest loss for five epochs.

V. Results and Discussion

The DCNN is trained on the synthetic data, which is generated using the Marmousi model, as previously explained. In order to assess the performance of the proposed method, the DCNN is tested on the synthetic data as well as field data from the Utah Tomography and Modeling/Mitigation Consortium (UTAM) [39]. Each trace in field data has 4096 samples with a sampling frequency of $f_s = 4$ kHz. To achieve high accuracy, the data to be recovered and the training dataset must have the same sampling frequency. During the preprocessing stage, the data are cleaned by removing bad traces caused by malfunctioning geophones. SNR is used as a criterion for performance evaluation and is defined as follows:

$$\text{SNR} = \frac{||\mathbf{s}||_2^2}{||\mathbf{s} - \hat{\mathbf{s}}||_2^2}. \tag{9}$$

The DSCS's performance is evaluated with the abovementioned datasets. The compression gains and SNR used in the experiments are calculated by taking the average of the respective values across all traces. The following formula is used to estimate the compression ratio (C):

$$C = \frac{S_t}{S_o} \tag{10}$$

where S_t and S_o are nonzero positive integers indicating the number of samples in a trace and compressed observation, respectively. Note here that, for a specific value of C, the DCNN is trained with the same value of C on the synthetic data to get the best results. Figs. 3 and 4 show the comparison of various sensing matrices (given in Table I) when the proposed DSCS is used. DSCS's performance degrades when the sensing matrix is sparse and nonrandom. This is due to the presence of zero elements in nearby locations in the matrix, making it unable to capture the details of the seismic signal.

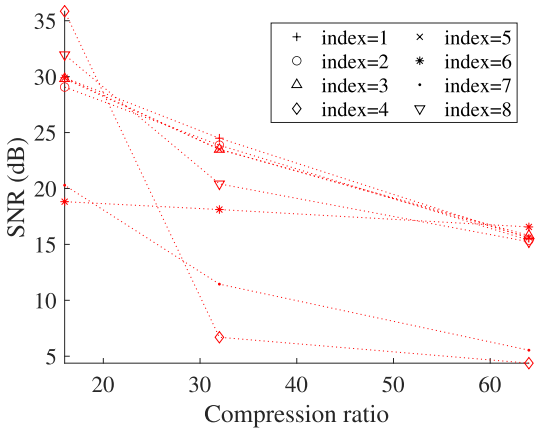


Fig. 3. SNR comparison of DSCS reconstructed synthetic traces for various sensor matrices and compression ratios.

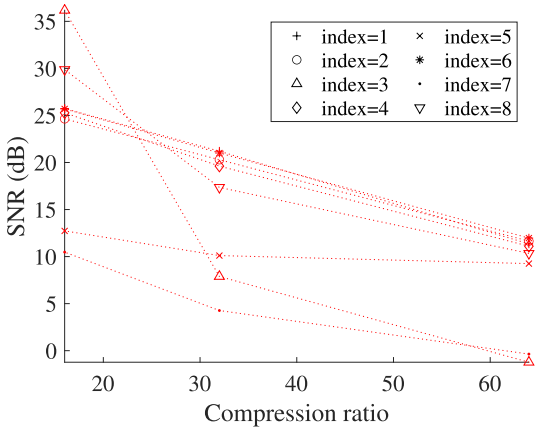


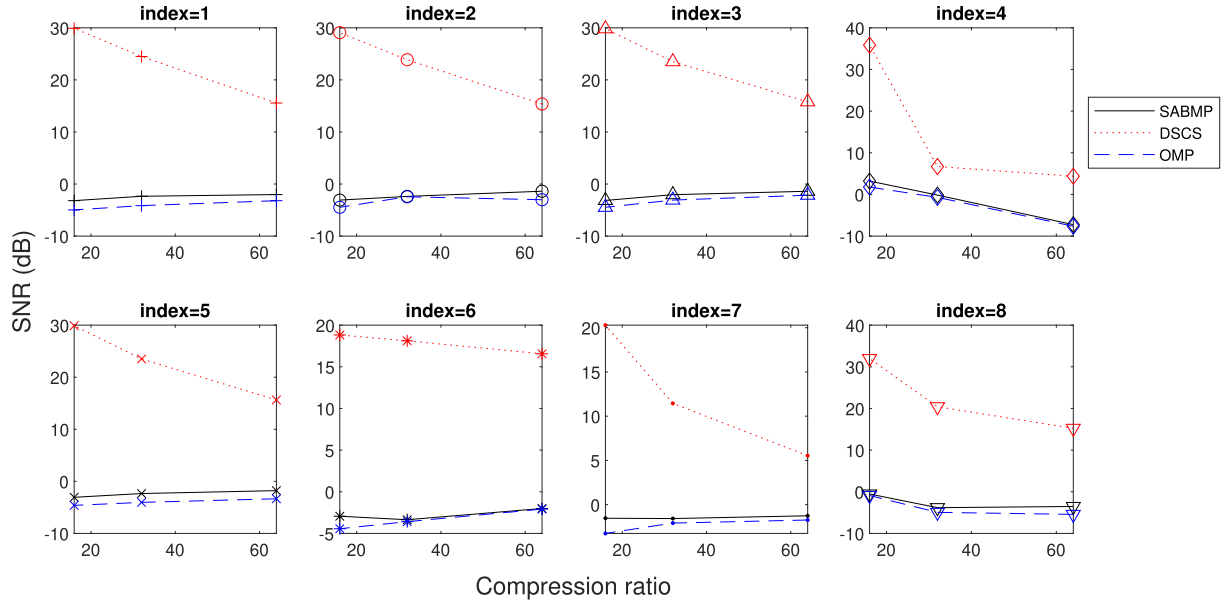
Fig. 4. SNR comparison of DSCS reconstructed field traces for various sensor matrices and compression ratios.

TABLE I
Various Forms of the Sensing Matrix Φ

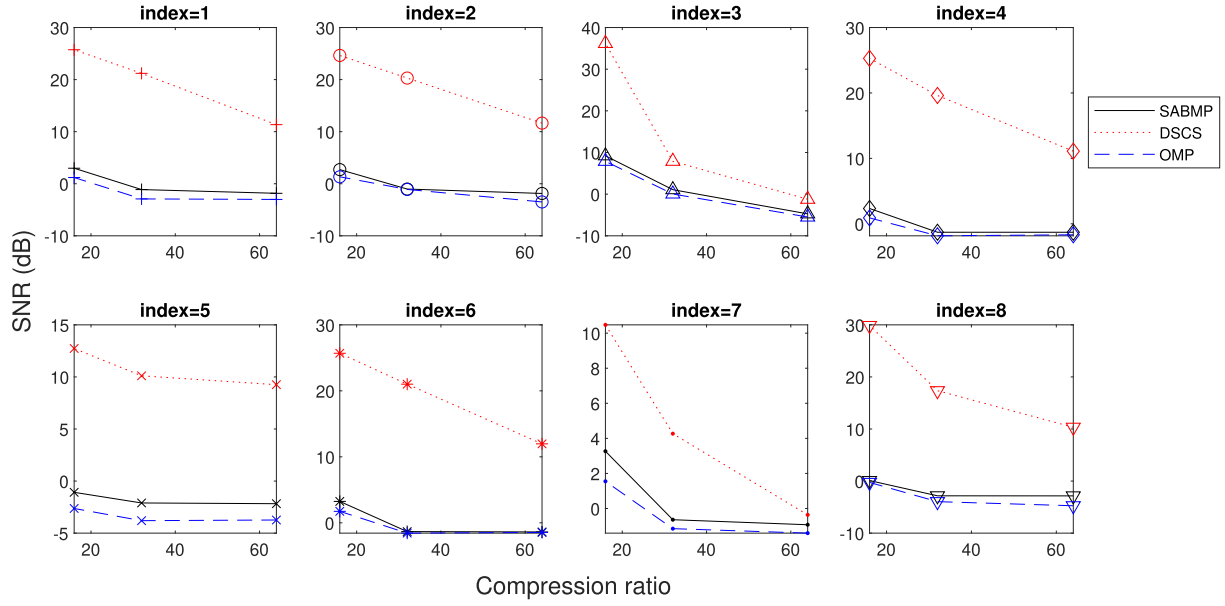
Index	Description
1	I.I.D. Gaussian random distribution [19]
2	I.I.D. Bernoulli random distribution [20]
3	Partial DCT (Similar to [22])
4	Partial DCT (randomized)
5	Sparse Dictionary using wavelets [40]
6	Sparse Dictionary using wavelets (columns randomized)
7	Random demodulator [24]
8	Random convolution [25]

Hence, in a practical situation, the random demodulator is not recommended due to the sensing matrix being block diagonal. In real-world scenarios, random convolution is advised in order to achieve the best possible results. Furthermore, partial DCT sensing matrix performance is better at low compression ratios. The reason is that the frequencies in the sensing matrix mimic the seismic signal's frequencies. However, at higher compression ratios, performance suffers greatly due to the loss of key frequencies in the partial DCT matrix. Figs. 3 and 4 also conclude that the randomness in the sensing matrices improves the performance.

According to Fig. 4, using the proposed DSCS method with already trained DCNN on synthetic data results in significant performance improvement, even with high compression gains.



(a)



(b)

Fig. 5. SNR comparison of reconstructed synthetic and field traces using DSCS, SABMP, and OMP methods for various values of compression ratio and sensing matrices. (a) Synthetic. (b) Field data.

For instance, a nearly 30-dB increase in SNR is attained at a compression gain of 16 using the random convolution-based sensing matrix. Fig. 5 shows a comparison among the DSCS and SABMP using various sensing matrices (given in Table I). It can be observed that the performance of the DSCS is the same and significantly superior to that of the SABMP with *random* sensing matrix of any distribution regardless of being sparse or nonsparse. Fig. 5 demonstrates that, while traditional CS *recovery* approaches fail in the presented case, this does not imply that the CS methods cannot be leveraged in seismic processing. The proposed deep learning-based recovery technique can still take advantage of CS theory for seismic recovery.

The performance of the DSCS using a random convolution-based sensing matrix is very close to that of the ideal (i.i.d.) random sensing matrix (even better at low compression ratio), as apparent from the previous figures. For the remaining figures, the random convolution-based sensing matrix is used due to its practical application and realization, as well as its better performance when compared with the random demodulator.

Though average SNR does not always provide a clear picture, a trace-by-trace comparison (Fig. 6) is demonstrated in order to justify DSCS's performance. As illustrated in the figure, the method's performance is commensurate, and SNR upticks can be seen across all the traces.

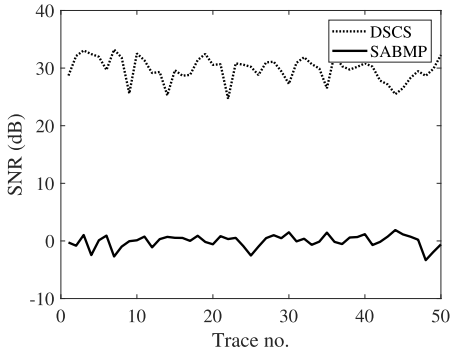


Fig. 6. Trace-by-trace SNR comparison of reconstructed field traces using DSCS and SABMP for a compression ratio of 16.

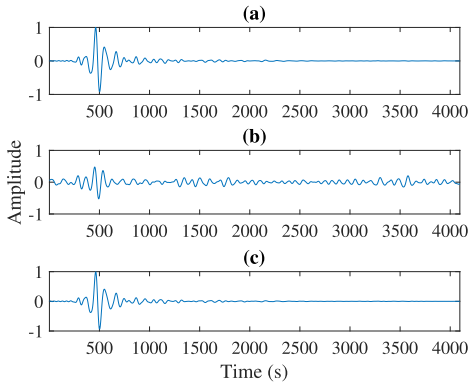


Fig. 7. Single trace comparison at a compression ratio of 16. (a) Original trace. (b) Reconstructed trace using SABMP. (c) Reconstructed trace using DSCS.

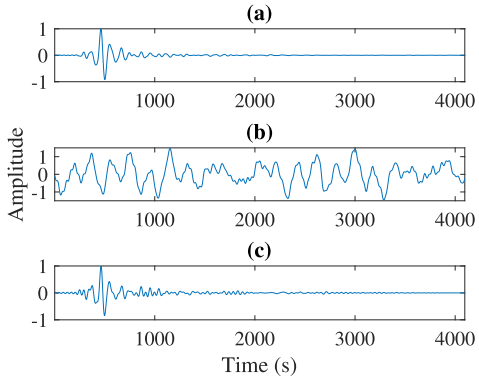


Fig. 8. Single trace comparison at a compression ratio of 64. (a) Original trace. (b) Reconstructed trace using SABMP. (c) Reconstructed trace using DSCS.

Next, a comparison of a reconstructed trace with the original trace at compression ratios of 16 and 64 is demonstrated in Figs. 7 and 8, respectively. The improvement is visible when DCNN is used for recovery.

It is also crucial to provide visualizations of seismic data in the frequency–space (f–x) domain, as displayed in Fig. 9. The noise attenuation is obvious from the figure. Furthermore, the method not only successfully removes high-frequency noise but also suppresses the low-frequency and in-band noise.

It can be observed from all the experimental results that the SABMP and OMP methods give the worst results using the realistic implementable random convolution matrix. This

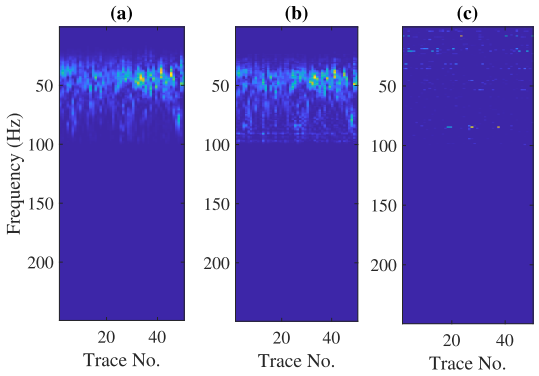


Fig. 9. Frequency spectrum at a compression ratio of 16. (a) f–x domain representation of the original traces. (b) f–x domain representation of the reconstructed signal using DSCS. (c) Difference of the original and recovered f–x domain representations.

also backs up the CS recovery methods' practicality concerns. Finally, it is important to argue that unlike other existing compression methods [4], [5], [6], [7], [8], [41], the proposed method does not rely on cooperation among the geophones to achieve the desired compression ratios. This avoids putting an extra processing burden on the battery-driven geophones. The proposed method works in a standalone manner, easing the processing load at a geophone.

VI. Conclusion

In this article, we have proposed an energy-efficient data acquisition and transmission method to sustain wireless geophone networks for large-scale oil and gas exploration. This is achieved by using CS at the geophone level. Numerous sensing mechanisms that could be implemented at the geophone have been investigated for their compression performance. Furthermore, a novel deep learning-based network for high-quality seismic signal reconstruction has been proposed that is able to recover signal from the compressed samples acquired by the geophones. For this purpose, an autoencoder architecture with skip connections is utilized to replace the conventional iterative CS reconstruction algorithms. The network is trained on carefully generated synthetic data so as to generalize it to a wide spectrum of real seismograms. The performance of the proposed method has been examined over several datasets that include both synthetic and real-field datasets. Various experimental results demonstrate that the proposed DSCS allows us to obtain significantly greater compression ratios coupled with high SNR performance when compared with the traditional iterative CS reconstruction algorithms. This implies that even if the traditional CS recovery approaches fail under certain conditions, CS methods cannot be leveraged. The visual analysis of the reconstructed signals also verifies the effectiveness of the proposed method. Specifically, DSCS's experimental results on the real-field seismic dataset concluded that it can produce a stable, high-quality seismic profile (SNR of around 30 dB) with a high compression gain of 16. However, the traditional approach fails to achieve this compression. A good subject for future investigation will be adapting the proposed method for the 3-D reconstruction of seismic volumes.

REFERENCES

[1] A. Othman, N. Iqbal, S. M. Hanafy, and U. B. Waheed, "Automated event detection and denoising method for passive seismic data using residual deep convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5900711.

[2] N. Iqbal, S. Al-Dharrab, A. Muqaibel, W. Mesbah, and G. Stüber, "Analysis of wireless seismic data acquisition networks using Markov chain models," in *Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2018, pp. 1–5.

[3] N. Iqbal, A. Zerguine, and S. Khan, "OFDMA-TDMA-based seismic data transmission over TV white space," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1720–1724, May 2021.

[4] A. Bilgin, M. W. Marcellin, and M. I. Altbach, "Compression of electrocardiogram signals using JPEG2000," *IEEE Trans. Consum. Electron.*, vol. 49, no. 4, pp. 833–840, Nov. 2003.

[5] 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.

[6] X. Zhang et al., "An efficient seismic data acquisition based on compressed sensing architecture with generative adversarial networks," *IEEE Access*, vol. 7, pp. 105948–105961, 2019.

[7] A. Payani, F. Fekri, G. Alregib, M. Mohandes, and M. Deriche, "Compression of seismic signals via recurrent neural networks: Lossy and lossless algorithms," in *SEG Technical Program Expanded Abstracts*. Houston, TX, USA: Society of Exploration Geophysicists, Aug. 2019, pp. 4082–4086.

[8] H. H. Nuha, A. Balghonaim, B. Liu, M. Mohandes, and F. Fekri, "Seismic data compression using deep neural network predictors," in *SEG Technical Program Expanded Abstracts*. Houston, TX, USA: Society of Exploration Geophysicists, Aug. 2019, pp. 258–262.

[9] J. C. Ye, "Compressed sensing MRI: A review from signal processing perspective," *BMC Biomed. Eng.*, vol. 1, no. 1, pp. 1–17, Dec. 2009.

[10] J. Yang, T. Jin, C. Xiao, and X. Huang, "Compressed sensing radar imaging: Fundamentals, challenges, and advances," *Sensors*, vol. 19, no. 14, p. 3100, Jul. 2019.

[11] O. Pradhan and A. J. Gasiewski, "Endfire synthetic aperture radar for a cryobot for exploration of icy moons and terrestrial glaciers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 1000114.

[12] S. Cui, A. Ma, L. Zhang, M. Xu, and Y. Zhong, "MAP-Net: SAR and optical image matching via image-based convolutional network with attention mechanism and spatial pyramid aggregated pooling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 1000513.

[13] *Wireless Seismic*. Accessed: Dec. 26, 2021. [Online]. Available: https://wirelessseismic.com/

[14] R. Kendall, *Cableless Seismic Acquisition*. Calgary, AB, Canada: Tesla Exploration, 2015.

[15] M. Mangia, F. Pareschi, V. Cambareri, R. Rovatti, and G. Setti, "Analog-to-information conversion," in *Proc. Adapt. Compress. Sens. Eff. Hardw. Implementations*. Cham, Switzerland: Springer, 2018, pp. 169–210.

[16] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Jan. 2006.

[17] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[18] D. J. Monk, *Survey Design and Seismic Acquisition for Land, Marine, and In-Between in Light of New Technology and Techniques*. Society of Exploration Geophysicists, Feb. 2020, doi: 10.1190/1.9781560803713.fm.

[19] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[20] W. Lu, W. Li, K. Kpalma, and J. Ronsin, "Compressed sensing performance of random Bernoulli matrices with high compression ratio," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1074–1078, Aug. 2015.

[21] A. Gilbert and P. Indyk, "Sparse recovery using sparse matrices," *Proc. IEEE*, vol. 98, no. 6, pp. 937–947, Jun. 2010.

[22] G. Xu and Z. Xu, "Compressed sensing matrices from Fourier matrices," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 469–478, Jan. 2015.

[23] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. IEEE/SP 14th Work. Stat. Signal Process.*, Aug. 2007, pp. 294–298.

[24] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2007, pp. 1959–1962.

[25] J. Romberg, "Compressive sensing by random convolution," *SIAM J. Imag. Sci.*, vol. 2, no. 4, pp. 1098–1128, Jan. 2009.

[26] W. Yin, S. Morgan, J. Yang, and Y. Zhang, "Practical compressive sensing with Toeplitz and circulant matrices," in *Proc. SPIE, Vis. Commun. Image Process.*, vol. 7744, Jul. 2010, Art. no. 77440K.

[27] M. Rani, S. B. Dhok, and R. B. Deshmukh, "A systematic review of compressive sensing: Concepts, implementations and applications," *IEEE Access*, vol. 6, pp. 4875–4894, 2018.

[28] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, 1993, pp. 40–44.

[29] S. K. Sahoo and A. Makur, "Signal recovery from random measurements via extended orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2572–2581, May 2015.

[30] T. Y. Al-Naffouri and M. Masood, "Distribution agnostic structured sparsity recovery algorithms," in *Proc. 8th Int. Workshop Syst., Signal Process. Appl. (WoSSPA)*, May 2013, pp. 283–290.

[31] M. Masood and T. Y. Al-Naffouri, "Sparse reconstruction using distribution agnostic Bayesian matching pursuit," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5298–5309, Nov. 2013.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, May 2016, pp. 770–778.

[33] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Stat. (PMLR)*, 2010, pp. 249–256.

[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[35] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[36] G. S. Martin, R. Wiley, and K. J. Marfurt, "Marmousi2: An elastic upgrade for marmousi," *Lead. Edge*, vol. 25, no. 2, pp. 156–166, Feb. 2006.

[37] *Ormsby Wavelet*. Accessed: Dec. 20, 2021. [Online]. Available: https://wiki.seg.org/wiki/Ormsby_wavelet

[38] G. F. Margrave and M. P. Lamoureux, *Numerical Methods of Exploration Seismology*. Cambridge, U.K.: Cambridge Univ. Press, Dec. 2018.

[39] *UTAM Seismic Data Library*. Accessed: Nov. 27, 2019. [Online]. Available: http://utam.gg.utah.edu/SeismicData/SeismicData.html

[40] M. Misiti, Y. Misiti, G. Oppenheim, and J. M. Poggi, "Wavelets toolbox™: User's guide: The Mathworks (R2021b)," Mathworks, Tech. Rep., 2021.

[41] 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.