Journal of Applied Geophysics 175 (2020) 103979

Contents lists available at ScienceDirect

Journal of Applied Geophysics

journal homepage: www.elsevier.com/locate/jappgeo

Sparse seismic deconvolution via recurrent neural network

Deborah Pereg [a,*], Israel Cohen [a], Anthony A. Vassiliou [b]

[a] Technion – Israel Institute of Technology, Israel
[b] GeoEnergy, Inc, Houston, TX, USA

ARTICLE INFO

Article history:
Received 21 August 2019
Received in revised form 9 February 2020
Accepted 12 February 2020
Available online 17 February 2020

Keywords:
Seismic inversion
Seismic signal
Sparse reflectivity
Reflectivity estimation
Recurrent neural net

ABSTRACT

We propose a new efficient method to perform deconvolution of post-stacking post-migration seismic data. We employ recurrent neural networks (RNNs) to obtain super resolution reflectivity images. The network is designed and trained to take into account time and space relations. The robustness of the proposed method is experimentally validated for both synthetic data and real data with challenging structures and difficult signal-to-noise ratio (SNR) environment. We explore the system's behavior for different training and testing scenarios and discuss potential problems for future research. We show that training with synthetic data of simple structures solely can yield enhanced and detailed real data inversion results. The proposed method can be applied to large volumes of three-dimensional (3D) seismic data.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Acoustic waves imaging techniques are designed to give insight into the internal structure of a medium. Whether the media investigated are human organs or tissues, transition zones in the bottom of the ocean or the earth's surfaces, the objective is to be able to see, as clearly as possible, *into* a medium or an object, or, in other words, see into material. One of the challenges in echo-acoustic systems is to improve spatial resolution, both horizontally and vertically. Resolution limits are inherently determined by the wave length and fundamentally limited by diffraction. Since acoustic waves propagate in relatively low velocities, small wave lengths can be obtained at relatively low frequencies, exhibiting large penetration depth. Therefore, the optimal resolution of an echo-acoustic system is theoretically high. Yet, in practice, the actual resolution is far below optimal. Possible reasons could be acquisition conditions or limited data processing tools.

In different contexts the term super-resolution (Fernandez-Granda, 2013; Candès and Fernandez-Granda, 2013, Candès and Fernandez-Granda, 2014; Donoho, 1992) is used as an operation that enhances the resolution of a sensing system. Observing sub-wavelength features ("breaking diffraction limits") is a major challenge in imaging systems in astronomy, seismology, medical imaging, microscopy, spectroscopy, radar and more.

In reflection seismology our purpose is to reveal the structure and properties of the earth subsurface layers. By resolving the fine details of acquired seismic data we hope to be able to visually examine the internal structure of the subsurface, specifically geological structures such as layers, channels, traps and faults. Seismic images are of extreme importance as means for locating mineral deposits and energy sources (such as hydrocarbons, ores, water, geothermal reservoirs) (Sherif and Geldart, 1983), obtaining geological information for engineering, geothermal energy surveys, risk assessment of tsunamis, locating underground activity for security purposes such as tunnel detection (Gurbuz et al., 2006), and more.

When acquiring the data, a short acoustic pulse is transmitted into the ground. An array of geophones, placed on the ground, receives the acoustic pulses reflected back at layer boundaries. The data received is then further processed into two-dimensional (2D) seismic data (Sherif and Geldart, 1983). In a 2D section of seismic data, the horizontal direction is the spatial dimension referred to as in-line direction. The vertical dimension is the two-way travel-time corresponding to a reflector's depth in the ground.

Unfortunately, even after complex imaging procedures such as stacking and migration (Biondi, 2006), a 2D section of seismic data does not represent the actual 2D image of the subsurface. Each reflection is distorted during its propagation through the surface, in addition to coherent and random noise. Namely, everything that is not a seismic event that carries desirable information, including diffraction, multiples etc. is considered as noise. Inversion of the seismic data is an ill-posed problem primarily because the degrading pulse is band limited and non-stationary. Moreover, as mentioned, the data is often degraded by coherent and non-coherent interferences.

Many previous works try to simplify and solve the seismic inversion problem by breaking the data into independent vertical 1D deconvolution problems. An acoustic reflected pulse (the wavelet) is modeled as a 1D time-invariant kernel. Considering that reflections are generated at discontinuities in the medium impedance, the core

* Corresponding author.
E-mail address: deborahp@campus.technion.ac.il (D. Pereg).

https://doi.org/10.1016/j.jappgeo.2020.103979
0926-9851/© 2020 Elsevier B.V. All rights reserved.

assumption is that each recovered reflectivity channel is a sparse spike train (Wiggins, 1978; Taylor et al., 1979; Riel and Berkhout, 1985; Nguyen and Castagna, 2010; Zhang and Castagna, 2011; Gholami and Sacchi, 2012). In other words, each seismic trace (a column in the seismic 2D data) is modeled as a weighted superposition of one-dimensional (1D) pulses degraded by additive noise. In an attempt to estimate each reflectivity channel, some of these methods try to decompose the seismic data by applying an ℓ_1-norm minimization constraint to an optimization problem fitting the observed data to stable reflectivity solutions (Taylor et al., 1979; Riel and Berkhout, 1985; Nguyen and Castagna, 2010; Zhang and Castagna, 2011; Gholami and Sacchi, 2012; Pham et al., 2014; Repetti et al., 2015), assuming the smearing wavelet is either known or estimated in the process (blind deconvolution). Similarly to other atomic decomposition problems (Chen et al., 2001), investigated in other applications in signal processing and machine learning, such as ultasound (Bendory et al., 2016), image processing (Elad, 2010) and compressed sensing (Donoho, 2006).

Obviously, treating 2D data as a puzzle of 1D unrelated data pieces is inefficient. Multichannel seismic deconvolution methods (Idier and Goussard, 1993; Mendel et al., 1981; Kaaresen and Taxt, 1998; Heimer et al., 2007; Heimer and Cohen, 2009; 2008; Ram et al., 2010; Gholami and Sacchi, 2013) are tailored to take into account more than one trace in each channel estimation, thus using availabe spatial information to promote horizontal continuity of the estimated reflectivity image and to suppress noise. Heimer and Cohen, (2009) propose a variation of the Viterbi algorithm (Forney, 1973), based on modeling the reflectivity as a Markov-Bernoulli random-field (MBRF). The idea is to search for the most likely sequences of reflectors connected by defined legal transitions. Ram et al., (2010) propose an algorithm for multichannel blind deconvolution assuming a Markov-Bernoulli-Gaussian (MBG) reflectivity model, where each reflectivity channel is estimated from the corresponding observed seismic trace, and an estimate of adjacent reflectivity channels.

The above inversion methods assume a time-invariant convolution model. That is, the blurring wavelet is modeled as a 1D time-invariant signal that does not change neither in time nor in space (i.e., in both horizontal and vertical directions), thus ignoring time-depth variations in the waveform. However, in practice a typical seismic wavelet is time-variant. When seismic waves propagate through the earth subsurface their energy is absorbed by the media so that the wavelet's amplitude decreases with depth. In addition, due to velocity dispersion (i.e., different frequencies travel with different speeds) the wavelet's shape changes and broadens with increased traveltime. An attempt to make the seismic data stationary can be accomplished by amplitude balancing and spectral balancing. To address this issue seismic inverse Q-filtering (Kjartansson, 1979; Gelius, 1987; Hale, 1981; Wang, 2008) is applied in an attempt to invert phase and amplitude distortions of the propagating acoustic waves. Unfortunately, the process is of heavy computational burden and often impractical. Nonstationary deconvolution methods attempt to deconvolve the seismic data and also compensate for waves attenuation. For example, the Gabor deconvolution algorithm (Margrave et al., 2011), nonstationary sparse reflectivity inversion (NSRI) (Chai et al., 2014), and nonstationary deconvolution algorithm based on spectral modeling (Rosa and Ulyrch, 1991). Some of these methods are applied to pre-stacked data. In previous work (Pereg and Cohen, 2017) we show that the recovery of the seismic reflectivity can be efficiently done by solving a simple convex optimization problem based on earth Q attenuation model. We derive theoretical bounds on the recovery error and on the localization error, proving that estimated reflectors locations must be close to true reflectors locations.

Recently seismic data processing techniques concentrate more on 3D data designed to recover a 3D reflectivity function representing the earth's impulse response (Sherif and Geldart, 1983). In spite of degraded results, 1D methods are still often applied to 3D data due to practical convenience. In contrast, (Gholami and Sacchi, 2013) introduce a fast 3D blind seismic deconvolution algorithm. The algorithm iterates between two stages: 3D reflectivity estimation and source estimation. Also, in (Pereg et al., 2019) we propose a 3D time-variant deconvolution where we assess the relation between a point in the data and adjacent points, using discontinuity measures (Cohen and Coifman, 2002; Cohen et al., 2006). This approach enables the use of multiple traces in the seismic data for each channel estimation in noisy environments and highly attenuating media.

Unfortunately most inversion methods rely on optimization problem solving, which often requires accurate estimation of many parameters. In addition, when dealing with a large data set, or when real time processing is required, practical implementation becomes impossible. The computational complexity requires too much memory and running-time, and the solution becomes infeasible.

As described above, current methods are computationally heavy and sometimes complicated and impractical. Nowadays the extensive development of deep neural nets (DNNs) has had a profound impact on signal processing and image processing. In the field of exploration seismology attempts have been made to use feed forward neural network (FNN) to estimate normal moveout (NMO) velocity (CaldernMacas et al., 1998), to automate velocity picking (Fish and Kusuma, 2005) and first arrival picking (Murat and Rudman, 1992; McCormack and Rock, 1993). Moreover, recurrent neural nets (RNNs) were successfully used to estimate stacking velocity directly from seismic data for NMO correction (Biswas et al., 2018). Li et al., (2019) propose a method to bulid seismic velocity model from seismic data by DNNs. Various works explore the use of neural nets for seismic interpretation. To name a few, (Dorrington and Link, 2004) propose a genetic algorithm that uses neural network training to find optimal seismic attributes for well-log prediction; (Araya-Polo et al., 2017) and (Zhang et al., 2014) propose using convolutional neural nets (CNNs) for automatic fault detection from seismic data before migration; Yang et al., (2018) propose the use of CNN for detection of salt dome boundaries from pre-stack seismic data.

Generally speaking, DNN methods for image super-resolution usually employ some sort of architecture involving deep CNNs (e.g. (Nehme et al., 2018; Dong et al., 2016)), mainly because CNNs are considered to be able to extract features of the data assuming stationarity. Some of these methods employ a sparsity constraint. Inspired by the success of CNNs in image classification, learning methods are now also used to predict high resolution images from low resolution images.

In this paper, our main target is to reveal the ground truth reflectivity image. Our purpose is to develop an automatic, fast and efficient way to perform seismic deconvolution. We propose the use of RNN to deconvolve seismic data, in an attempt to address current issues in existing methods. We believe that RNN fits this task most because it incorporates the use of temporal information as well as spatial information. We suggest that each point in the estimated reflectivity can be inferred from a relatively small patch in the seismic data, which we will refer to as an analysis patch. For simplicity, we assume that the mapping from each patch to a reflectivity point is similar. That is, the data is stationary. As mentioned before, in practice this assumption does not always hold. Yet, assuming a time invariant wavelet is helpful for introducing the major concepts that affect the image quality, and is standard in the literature. As will be presented, this simplification does not necessarily lead to degraded results in comparison to recent time-variant deconvolution algorithms, as long as ground attenuation is not too aggressive.

The paper is organized as follows. In Section 2, we briefly review the signal model and problem formulation, and provide background for recurrent neural networks. In Section 3, we describe the proposed deconvolution method using RNN. Section 4 describes numerical experiments and real data results. Lastly, in Section 5, we conclude and discuss further study directions.

2. Problem formulation

2.1. Signal model

We consider an unknown 2D reflectivity signal $\mathbf{R} \in \mathbb{R}^{L_r \times J}$ representing an image section of the internal earth structure. Assuming a stratified structure, where reflections are generated at acoustic impedance boundaries, each 1D channel $r^{(l)}[k]$, that is, a column in the reflectivity, is formulated as a superposition of point sources. Namely, in the discrete setting, assuming a sampling rate F_s, and that the set of delays $T = \{t_m\}$ lies on a grid k/F_s, $k \in \mathbb{Z}$, i.e., $t_m = k_m/F_s$, a true reflectivity channel can be written as

$$r^{(l)}[k] = \sum_m c_m \delta[k - k_m], \qquad k \in \mathbb{Z}, \qquad c_m \in \mathbb{R} \tag{1}$$

for $l = 1, \ldots, J$, where $\delta[k]$ denotes the Kronecker delta function (see (Ricker, 1940)), and $\sum_m |c_m| < \infty$. $K = \{k_m\}$ is the set of discrete delays corresponding to the reflectors' locations.

Assuming a time-invariant model, an observed seismic discrete trace of channel l, in the observed seismic 2D data $\mathbf{S} \in \mathbb{R}^{L_s \times J}$, is of the form

$$s^{(l)}[k] = \sum_n r^{(l)}[n] g[k - n] + w^{(l)}[k], \qquad n \in \mathbb{Z} \tag{2}$$

where $g[k]$ is a known seismic wavelet of length L_g, and $w[k]$ is an additive white Gaussian noise. Clearly $L_s = L_r + L_g - 1$. The wavelet is assumed to be invariant in both time and space (i.e., both in horizontal and vertical directions). Accordingly, in this study, our goal is to find the true support $K = \{k_m\}$ and spikes' amplitudes $\{c_m\}$ given the observed seismic data. We also assume that the seismic signal is free of multiple reflections (multiples are events that have undergone more than one reflection (Sherif and Geldart, 1983)).

In matrix-vector form we can model the observed 2D seismic data image $\mathbf{S}$ of size $L_s \times J$ as

$$\mathbf{S} = \mathbf{G}\mathbf{R} + \mathbf{W} \tag{3}$$

where $\mathbf{G}$ is a convolution matrix of size $L_s \times L_r$ such that $G_{k,l} = g[k - l]$, and $\mathbf{W}$ is an additive i.i.d white Gaussian noise matrix independent of $\mathbf{R}$, with zero mean and variance σ_w^2. Note that we do not impose any prior knowledge of the structure or possible patterns in the reflectivity image.

2.2. Recurrent neural network

A recurrent neural network (RNN) is a type of neural network that employs feedback connections in addition to feedforward connections between nodes of the graph (Hopfield, 1982; Hochreiter and Schmidhuber, 1997). As known, in FNNs the signal travels in one direction - from input to output. In a recurrent network, the signal also travels backwards. Consequently, the output signal at a current state depends not only on current inputs, but also on outputs at previous states. In other words the network "remembers" its previous decisions.

In general, given an input sequence $\mathbf{x} = [\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{L_t - 1}]$, and a corresponding output sequence $\mathbf{y} = [\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{L_t - 1}]$, the RNN learns the mapping $f : \mathbf{x} \to \mathbf{y}$. At time step t, the output of the net can be formulated as

$$\mathbf{y}_t = \phi\left(\mathbf{W}_{xy}^T \mathbf{x}_t + \mathbf{W}_{yy}^T \mathbf{y}_{t-1} + \mathbf{b}\right) \tag{4}$$

where ϕ is an activation function, $\mathbf{W}_{xy}$ and $\mathbf{W}_{yy}$ are weight matrices and $\mathbf{b}$ is the bias vector. At time step $t = 0$ previous outputs are assumed to be zero. The function ϕ can be one of the known activation functions available such as sigmoid, rectified linear unit (ReLU) and hyperbolic tangent. In our experiments we use the ReLU activation function, ReLU $(z) = \max\{0, z\}$. A single recurrent neuron, or a layer of recurrent neurons, is considered as an RNN basic cell. In this work we used only one hidden layer of recurrent neurons, but it can be extended into multiple layers to form a deep neural network.

Now, suppose we have n_i inputs at time step t, i.e., $\mathbf{x}_t$ is of size $n_i \times 1$, and n_n neurons in an RNN cell. We know $\mathbf{W}_{xy}$ is of size $n_i \times n_n$, and $\mathbf{W}_{yy}$ is of size $n_n \times n_n$. During training of the network, in a single iteration we update the weights for multiple sequences of data known as mini-batch. The output of the whole mini-batch of m instances is

$$\mathbf{Y}_t = \phi\left(\mathbf{X}_t \mathbf{W}_{xy}^T + \mathbf{Y}_{t-1} \mathbf{W}_{yy}^T + \mathbf{b}\right) \tag{5}$$

where $\mathbf{Y}_t$ is of size $m \times n_n$, and $\mathbf{X}_t$ is of size $m \times n_i$.

As mentioned before, we say that the RNN has memory, because the output at time step t depends on outputs of previous time steps. Due to this property RNNs are often used to predict a future outcome. Since the RNN takes a time-series input and produces a time-series output, it is also applied to tasks where there is a need to apply temporal dynamic behavior, for example: handwriting recognition or speech recognition.

Now, the output $\mathbf{y}_t$ is of size $n_n \times 1$, and consists of non-negative values only, which typically does not fit the task at hand. In order to modify the size of the output to a desired size and allow negative values, we wrap the cell with a fully connected layer with the desired final output length M, which is written as

$$\mathbf{z}_t = \mathrm{FC}(\mathbf{y}_t) \tag{6}$$

such that $\mathbf{z}_t \in \mathbb{R}^{1 \times M}$.

We denote by $\mathbf{Z} \in \mathbb{R}^{L_t \times M}$ the matrix of predicted outputs, which is the concatenation of the vectors $\mathbf{z}_t$, $t = 0, 1, \ldots, L_t - 1$ as columns. During training, after each forward pass of a mini-batch, we update the weights using Adam-optimization. The gradient is calculated using back-propagation and the weights are updated with a defined learning rate value. We denote the weight matrices and the bias as $\theta = \{\mathbf{W}_{xy}, \mathbf{W}_{yy}, \mathbf{b}\}$. In our application, the loss function is the mean squared error of the predicted output and the expected output. Suppose $\mathbf{Z}_i$ and $\hat{\mathbf{Z}}_i$ are the expected output during training and current system's output, respectively, for input sequence $\mathbf{x}_i$. Denote the error matrix as $\mathbf{E}_i = \mathbf{Z}_i - \hat{\mathbf{Z}}_i$, then the loss term can be expresses as

$$J(\theta) = \frac{1}{m M L_t} \sum_{i=1}^{m} tr\left(\mathbf{E}_i^T \mathbf{E}_i\right) \tag{7}$$

where superscript ^{T} denotes the transpose of a vector or a matrix, and $tr(\cdot)$ denotes the trace of a square matrix. Other loss functions and regularization terms could be suggested. For example, it is possible to add a regularization term in order to enforce a small ℓ_1 norm of the output, because the solution is expected to be sparse. However, empirically, in our experience in this application, it does not seem to necessarily improve the results. Controlling the sparsity of the results in this manner in controversial, since it is not always clear if the level of sparsity versus the level of details of the solution is a result of true enhanced resolution or artifacts of an inaccurate sparsity weight constraint.

3. Seismic recovery using RNN

Let us denote $\tilde{\mathbf{S}} \in \mathbb{R}^{L_r \times J}$ as the cropped seismic image aligned with the reflectivity image $\mathbf{R}$, such that each reflector's location across the time-depth vertical axis corresponds to a the maximum value of a shifted wavelet $g[k - k_m]$ $\forall\, m$ in the seismic image $\tilde{\mathbf{S}}$. In other words, without loss of generality, assuming that $g[0] \geq g[k]$ $\forall\, k \in \mathbb{Z}$, that is, the pick of the wavelet is at its center, then

$$\tilde{\mathbf{S}}[k, n] = \mathbf{S}\left[k + \left\lfloor \frac{L_g}{2} \right\rfloor, n\right]$$

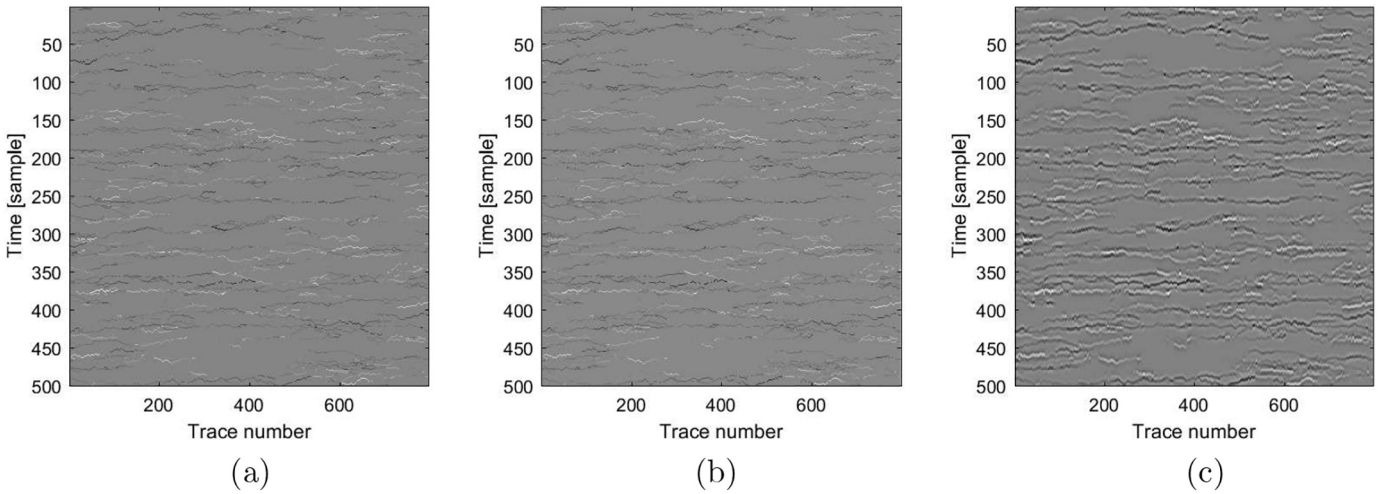

Fig. 1. Synthetic 2D recovery results: (a) Synthetic 2D reflectivity section; (b) Estimated 2D reflectivity, $\rho_{\mathbf{R},\hat{\mathbf{R}}} = 0.96$; (c) 2D noise free seismic data.

Definition 1 (Analysis Patch): We define an *analysis patch* as a 2D patch of size $L_t \times N$ enclosing L_t time (depth) samples of N consecutive traces of the observed seismic image $\mathbf{S}$. Assume $\{n_L, n_R \in \mathbb{N} : n_L + n_R = N - 1\}$. Then the analysis patch associated with a point at location (i,j) is

$$\mathbf{A}_{i,j} = \begin{pmatrix} \tilde{\mathbf{S}}_{i-L_t+1,j-n_L} & \cdots & \tilde{\mathbf{S}}_{i-L_t+1,j+n_R} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{S}}_{i,j-n_L} & \cdots & \tilde{\mathbf{S}}_{i,j+n_R} \end{pmatrix} \tag{8}$$

An analysis patch $\mathbf{A}_{i,j}$ is associated with a pixel $\mathbf{R}_{i,j}$ in the reflectivity image. To produce a point in the reflectivity $\mathbf{R}_{i,j}$, we set the input to the RNN as

$$\mathbf{x} = \mathbf{A}_{i,j} \tag{9}$$

Each time step input is a group of N neighboring pixels of the same corresponding time (depth). In other words, in our application $n_i = N$ and

$$\mathbf{x}_t = \left[\tilde{\mathbf{S}}_{t,j-n_L}, \ldots, \tilde{\mathbf{S}}_{t,j+n_R}\right]^T \tag{10}$$

We set the size of the output vector $\mathbf{z}_t$ to one expected pixel ($M = 1$), such that $\mathbf{Z}$ is expected to be the corresponding reflectivity segment,

$$\mathbf{Z} = \left[\mathbf{R}_{i-(L_t-1),j}, \ldots, \mathbf{R}_{i,j}\right]^T \tag{11}$$

Lastly, we ignore the first $L_t - 1$ values of the output $\mathbf{Z}$ and set the predicted reflectivity pixel $\hat{\mathbf{R}}_{i,j}$ as the last one, i.e., $\mathbf{z}_{Lt}$.

The analysis patch moves through the image and produces all expected reflectivity points in the same manner, which means each analysis patch and a corresponding reflectivity segment are an instance for the net. The size and shape of the analysis patch defines the geometrical distribution of traces and samples to be used for each point's computation. Typically we set L_t to be the approximated length of the wavelet $g[k]$. This way we ensure estimation of each point relies on suffice temporal information as well as spatial information. We believe RNN fits this task best because it is able to "remember" both in space and time dimensions.

The solution could be generalized to 3D images $\mathbf{R}_{3D} \in \mathbb{R}^{L_t \times J_x \times J_y}$ using a 3D analysis volume of size $L_t \times N_x \times N_y$. The analysis volume is then defined by N_x, N_y, the number of traces taken into account along the in-line and cross line axes, and L_t time-depth samples along the vertical axis. It can be defined to associate with a point in

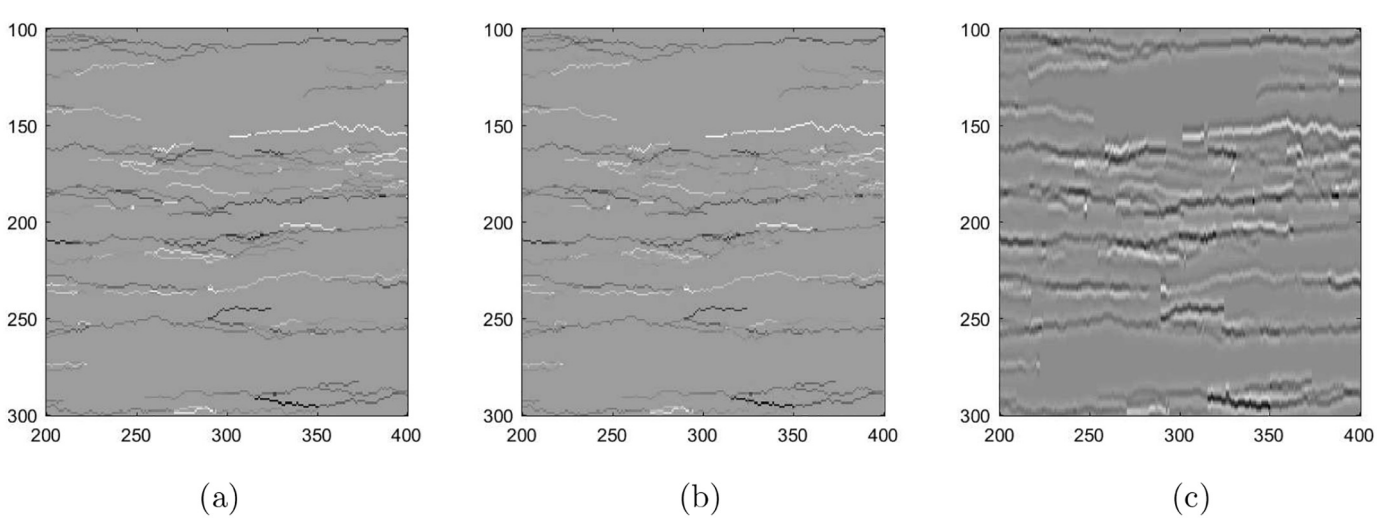

Fig. 2. A zoom into Fig. 2: (a) Synthetic 2D reflectivity section; (b) Estimated 2D reflectivity; (c) 2D noise free seismic data.

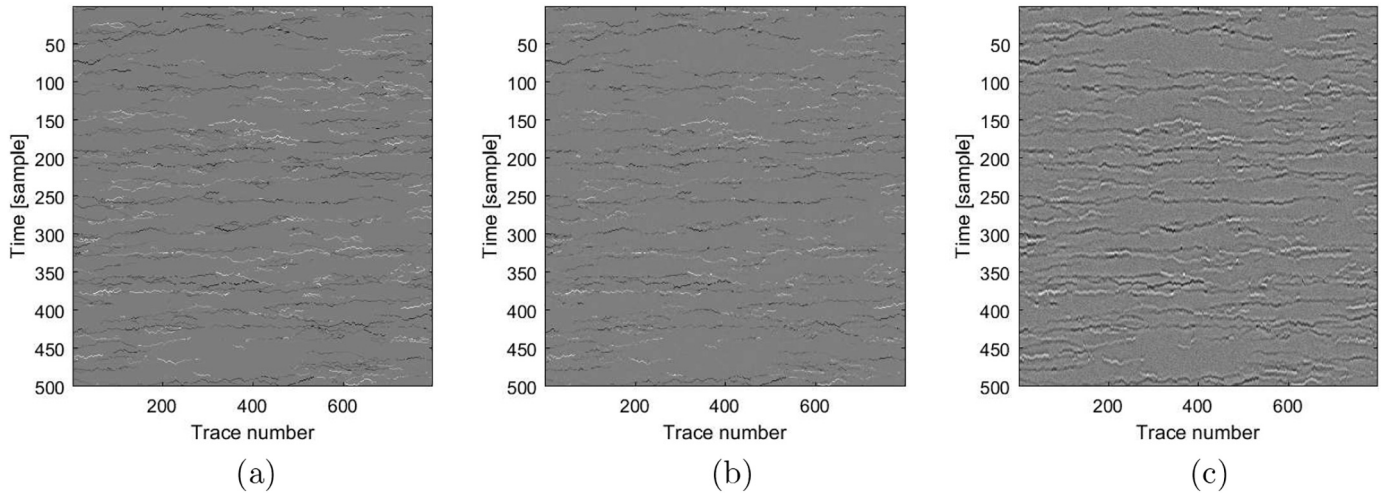

Fig. 3. Synthetic 2D recovery results: (a) Synthetic 2D reflectivity section; (b) Estimated 2D reflectivity, $\rho_{\mathbf{R}\hat{\mathbf{R}}} = 0.74$; (c) 2D seismic data (SNR = 5 dB).

its center, or in an asymmetrical manner. In a similar manner to the 2D configuration, for each reflectivity output voxel, the analysis volume would be an instance input to the RNN. Moving the analysis volume along the 3D observation image produces the entire 3D estimated reflectivity volume.

4. Experimental results

In this section, we provide synthetic and real data examples demonstrating the performance of the proposed technique. To implement the RNN we used TensorFlow (Abadi et al., 2015).

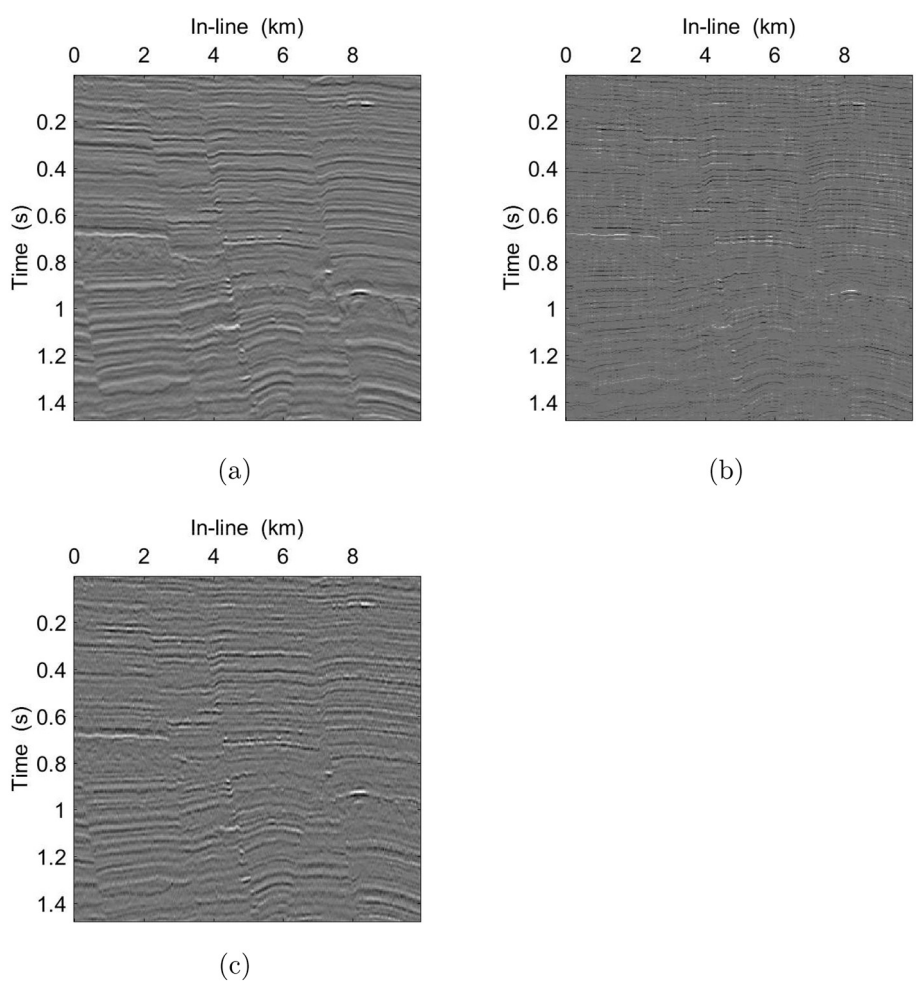

Fig. 4. Real data inversion results: (a) Seismic data; (b) Estimated reflectivity; (c) Reconstructed seismic data, $\rho_{\mathbf{S}\hat{\mathbf{S}}} = 0.87$.

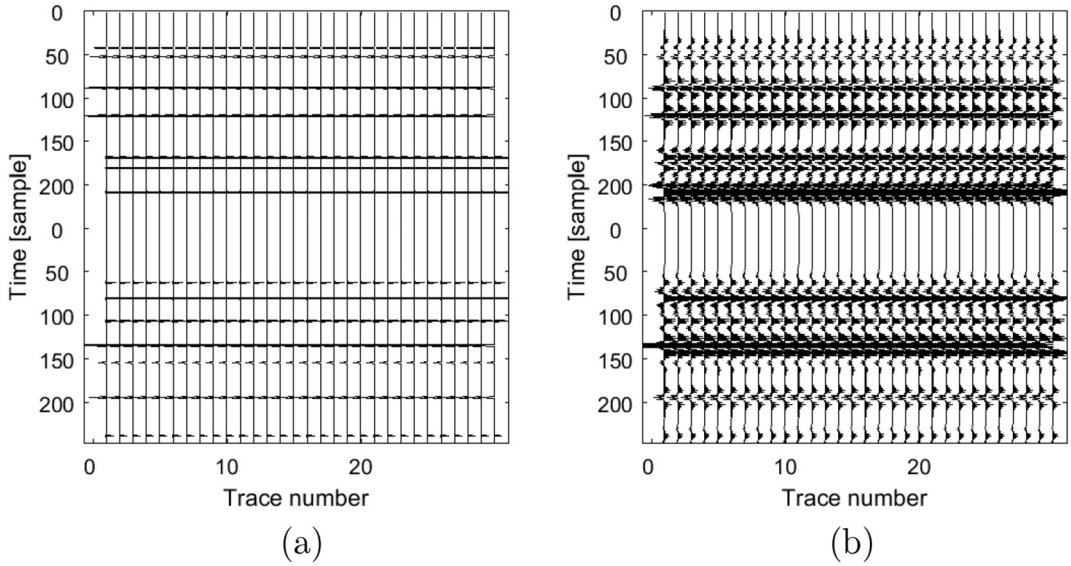

Fig. 5. Synthetic reflectivity and seismic data for RNN training: (a) Synthetic 2D reflectivity section of horizontal layers only; (b) 2D seismic data.

4.1. Synthetic data

First, we constructed a synthetic training and test data sets, simulating a 2D reflectivity of size $L_r \times J = 600 \times 800$. In other words, the reflectivity consists of 800 traces, each of 600 samples, with sampling interval of $T_s = 4$ms. We generate each 2D reflectivity $\mathbf{R}$ according to a Markov-Bernoulli random-field (MBRF) model, as described in (Ram et al., 2010). Namely, we model the reflectivity as a Bernoulli-Gaussian process. Assume $\mathbf{Q}$ is a matrix of location variables, indicating the position of layers boundaries. Accordingly, $q_{k,j} = 1$ if a reflector exists, and $q_{k,j} = 0$ otherwise. In addition the transition matrices $\mathbf{T}^{/}, \mathbf{T}^{-}, \mathbf{T}^{\backslash}$ indicate whether a reflector belongs to a layer boundary whose orientation is diagonally ascending, horizontal, or diagonally descending. We denote the transition variables as $t^{/}_{k,j}$, $t^{-}_{k,j}$, $t^{\backslash}_{k,j}$ of the (k,j) positions of $\mathbf{T}^{/}, \mathbf{T}^{-}, \mathbf{T}^{\backslash}$. Let $p(\cdot)$ denote a probability distribution function, we recall (Ram et al., 2010) the MBRF has the following properties:

1. Separability:

$$p\left(t^{/}_{k,j}, t^{-}_{k,j}, t^{\backslash}_{k,j}\right) = p\left(t^{/}_{k,j}\right) p\left(t^{-}_{k,j}\right) p\left(t^{\backslash}_{k,j}\right)$$

2. The jth columns of $\mathbf{Q}, \mathbf{T}^{/}, \mathbf{T}^{-}, \mathbf{T}^{\backslash}$, are white and Bernoulli distributed marginally from the rest of the field.
3. The Bernoulli distributions has the following parameters

$$\lambda = p\left(q_{k,j} = 1\right), \qquad \mu^{/} = p\left(t^{/}_{k,j} = 1\right)$$

$$\mu^{-} = p\left(t^{-}_{k,j} = 1\right), \qquad \mu^{\backslash} = p\left(t^{\backslash}_{k,j} = 1\right)$$

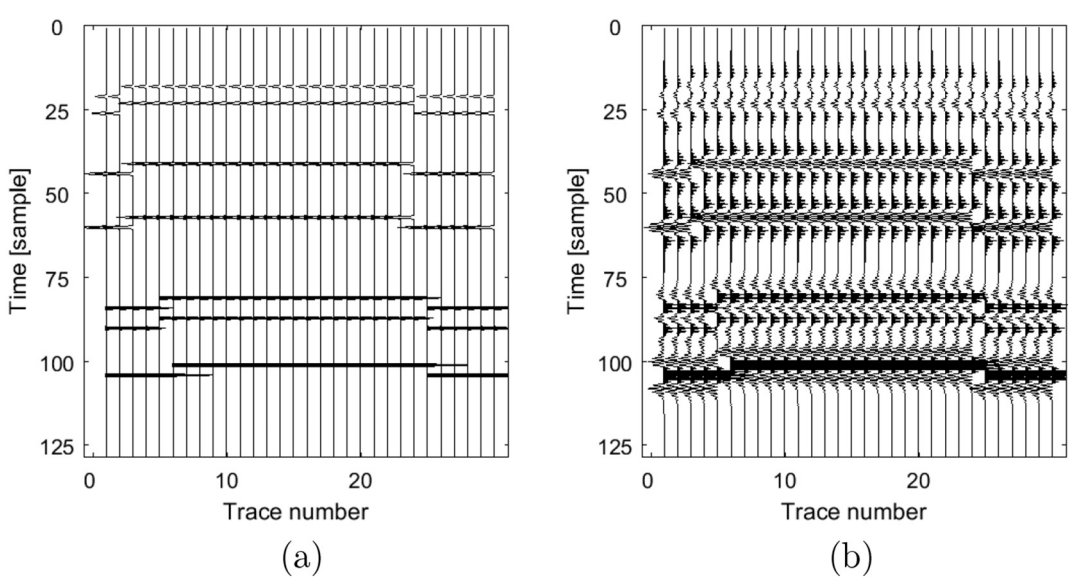

Fig. 6. Synthetic reflectivity and seismic data for RNN training: (a) Synthetic 2D reflectivity section of horizontal layers and two apparent faults; (b) 2D seismic data.

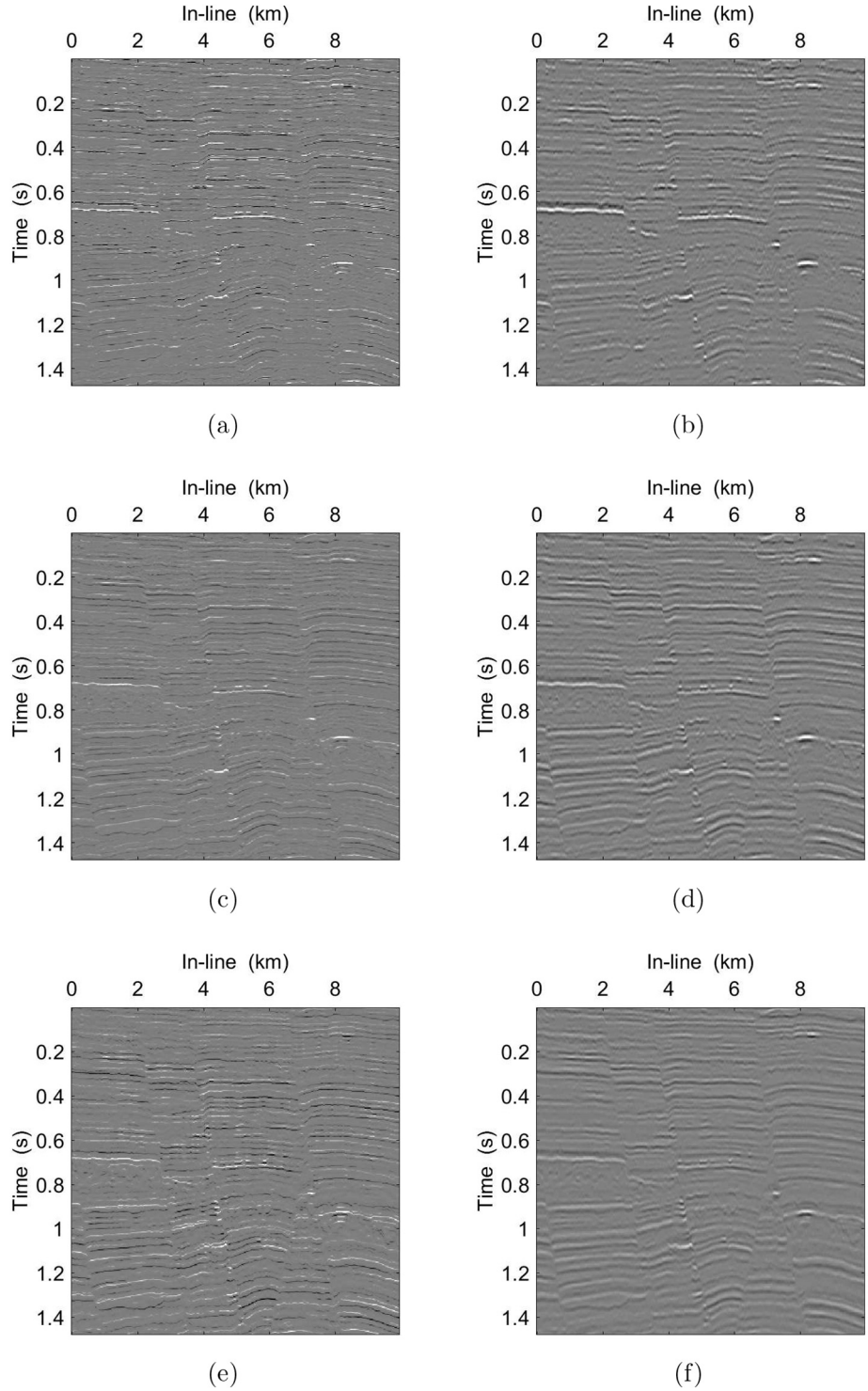

Fig. 7. Real data inversion results: (a)–(b) Estimated reflectivity when RNN is trained with the horizontal layers only, and corresponding reconstructed seismic data, $\rho_{\mathbf{s},\hat{\mathbf{s}}} = 0.77$; (c)–(d) Estimated reflectivity when RNN is trained with the horizontal layers and two faults, and corresponding reconstructed seismic data, $\rho_{\mathbf{s},\hat{\mathbf{s}}} = 0.88$; (e)–(f) Estimated reflectivity by 3D time-variant deconvolution (Pereg et al., 2019), and corresponding reconstructed seismic data, $\rho_{\mathbf{s},\hat{\mathbf{s}}} = 0.91$.

4. The probability for discontinuities along layers is given by

$$\varepsilon = p\left(q_{k,j} = 1 | t^{/}_{k,j} = 0, t^{-}_{k,j} = 0, t^{\backslash}_{k,j} = 0\right)$$

5. λ is related to $\{\mu^{/}, \mu^{-}, \mu^{\backslash}, \varepsilon\}$ by

$$\lambda = 1 - \left(1 - \mu^{/}\right)(1 - \mu^{-})\left(1 - \mu^{\backslash}\right)(1 - \varepsilon)$$

representing the *sparsity* of the expected reflectivity.

We generate reflectivity images with $\{\mu^{/} = \mu^{\backslash} = 0.016, \mu^{-} = 0.066, \varepsilon = 0.0005, \lambda = 0.0961\}$. A test reflectivity image is depicted in Fig. 1(a). Then, we build the seismic data as described in (2). The simulated seismic data is shown in Fig. 1(c). In our experiments we assume an initial source waveform $g(t)$ defined as the real-valued Ricker wavelet

$$g(t) = \left(1 - \frac{1}{2}\omega_0^2 t^2\right) \exp\left(-\frac{1}{4}\omega_0^2 t^2\right) \tag{12}$$

ω_0 is the most energetic (dominant) radial frequency (Wang, 2015), that plays a role of a scaling parameter, determining the width of the pulse. In this example the initial wavelet was a Ricker wavelet with $\omega_0 = 50\pi$, i.e., 25Hz, which is relatively a broad wavelet, so that the inversion process would be relatively more challenging. The number of traces taken into account in RNN's prediction for each pixel outcome is $N = 3$, and $L_t = 30$, in a symmetrical manner such that $n_L = n_R = 1$. In this case, training converges after 93,700 iterations with learning rate of 0.0001. Fig. 1(b) presents the estimated reflectivity image using the trained net. To asses the accuracy of the results we calculate the correlation coefficient between the ground truth known synthetic reflectivity image to the predicted reflectivity,

$$\rho_{\mathbf{R}\hat{\mathbf{R}}} = \frac{\hat{\mathbf{r}}_{cs}^T \mathbf{r}_{cs}}{\|\hat{\mathbf{r}}_{cs}\|_2 \|\mathbf{r}_{cs}\|_2}$$

where $\mathbf{r}_{cs}$ and $\hat{\mathbf{r}}_{cs}$ are column-stack vectors of the reflectivity images $\mathbf{R}$ and $\hat{\mathbf{R}}$ respectively, and $\|\cdot\|_2$ is the ℓ_2 norm. The correlation coefficient, between the original reflectivity and the estimated reflectivity, achieved by RNN deconvolution is $\rho_{\mathbf{R},\hat{\mathbf{R}}} = 0.96$, indicating promising expected performance of the RNN in this task. A zoom-into this example is depicted in Fig. 2.

To evaluate the recovery in noisy environment we added white Gaussian noise. Fig. 3(c) presents seismic data with SNR=5dB. The net was trained with noisy data with the same SNR level. The source reflectivity and the predicted reflectivity are shown in Fig. 3(a), (b) respectively. The correlation coefficient in this case is $\rho_{\mathbf{R},\hat{\mathbf{R}}} = 0.74$.

As can be seen in both examples, we used large images of considerable complex structure. Since the net is designed to process small analysis patches, it entails low complexity and is able to deconvolve the entire image in less than a minute. Generally speaking, the method is fast and efficient and therefore is suitable to large volumes of data. The proposed method could be easily extended to 3D seismic images, or other imaging data, such as: ultrasound imaging.

We conducted various experiments with the proposed system, with different parameters, which have led us to the following conclusions.

1. The accuracy of the prediction is *insensitive* to the exact *sparsity* of the training set, as long as the net has been trained with enough examples of different data patterns in an analysis patch, and that the data is sparse.
2. In contrast, the accuracy of the prediction is highly sensitive to correct choice of the wavelet's *scaling parameter* ω_0 for the training data. The wavelet's dominant frequency ω_0 can be estimated directly from the seismic data via spectral estimation (Zhang and Ulrych, 2002). Inaccurate choice of ω_0 results in degraded results.
3. Experimental results imply that the optimal choice for L_t is the approximated length of the corresponding wavelet.
4. Increasing the *sampling rate* F_s leads to degraded results as well, usually with many false spikes adjacent to each true spike location. Possible reasons can be due to higher computational complexity or "information overload" that tends to "confuse" the net. Essentially, any supervised learning method aims to capture the relevant information from each input variable and efficiently represent it in the output label. Information theoretic concept of data processing inequality (Cover and Thomas, 2006) reminds us that by processing the input, the network is only able to extract information. It is impossible for the net to create new information. Increasing the sampling rate, assuming that the network is able to recover higher frequencies that were mitigated during the waves propagation, data acquisition and imaging, is an unreasonable assumption, and therefore leads to degraded results.
5. Similarly, training with lower SNR levels or seismic data with attenuation according to Q earth model (see (Pereg and Cohen, 2017; Pereg et al., 2019)), does not lead to increased performance. The net is unable to learn the different patterns of a set of non-stationary wavelets corresponding to each reflector location (see (Pereg and Cohen, 2017)), as well as how to ignore noise interferences.
6. The number of traces along the spatial dimension N, in the analysis patch is user-dependent and can be determined according to the required resolution, the inline-spacing, and the type of geological features that is of interest to the interpreter. Increasing N, the number of neighboring traces taken into account, decreases the sensitivity to noise at the expense of decreasing the sensitivity to small-scale discontinuities.

4.2. Real data

We applied the proposed method, to real seismic data from a small land 3D survey (courtesy of GeoEnergy Inc., TX). The time interval is

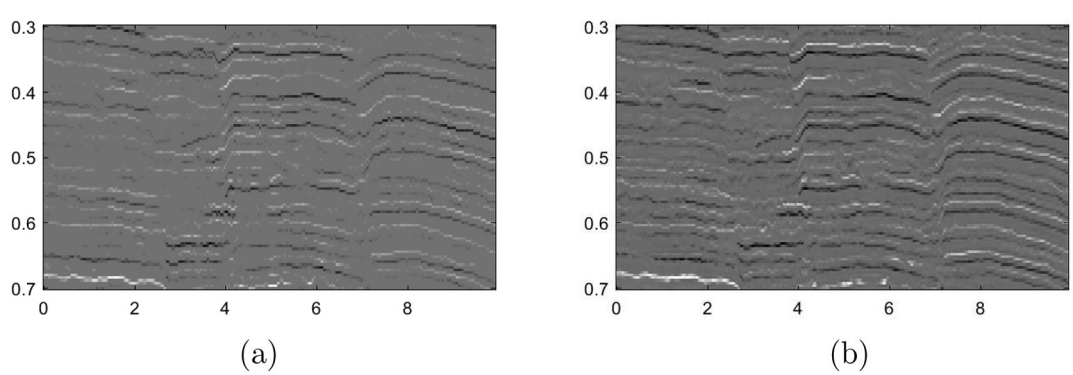

Fig. 8. A zoom into Fig. 8: (a) Estimated reflectivity by 3D time-variant deconvolution (Pereg et al., 2019); (b) Estimated reflectivity by RNN trained with horizontal layers and faults.

4 ms; inline and cross line trace spacing is 25 m. A small sub-image with inline distance of 10 km (401 traces) is used for demonstration. Each trace is 1.476 s in duration (369 samples). Fig. 4(a) shows a vertical cross section through the seismic data. As can be seen, the data chosen is non-stationary and of complex structure with multiple faults.

Assuming an initial Ricker wavelet with $\omega_0 = 50\pi$ (25Hz). The recovered reflectivity using an RNN of 1000 neurons, with an analysis patch of size 30×3 is shown in Fig. 4(b). Essentially, each reflectivity pixel is recovered by taking into consideration its relations to neighboring pixel: $L_t - 1$ preceding pixels of the same trace (spatial location) and $2L_t$ pixels from two neighboring traces - preceding and subsequent (in space).

Here, training data is *synthetic data* only. By conducting various experiments we have reached to the conclusion that training based on real data labeled by another algorithm, or combining both synthetic and real data leads to degraded results. In this example, we used for training simulated data with similar attributes as the data shown in Fig. 3. As mentioned before, estimating ω_0 accurately is of extreme importance. The seismic data reconstructed from the estimated reflectivity is shown in Fig. 4(c). We assume the given seismic data is noise-free. Since the ground truth is unknown, in order to evaluate the results in real data we build an estimated seismic data from the predicted reflectivity. Namely, we compute $\hat{\mathbf{S}} = \mathbf{G}\hat{\mathbf{R}}$, and then find the correlation coefficient between the reconstructed data to the given seismic data. In this example we have $\rho_{\mathbf{S},\hat{\mathbf{S}}} = 0.87$.

In order to analyze the net, and in hope of getting a better understanding of the net's learning process, we conducted the following experiment. We trained a net of 1000 neurons as described above, with a synthetic 2D reflectivity of size 248×30 of *horizontal layers* only, as depicted in Fig. 5. Then we tested the net with the same seismic data illustrated in Fig. 4(a). The recovered reflectivity and the reconstructed data are shown in Fig. 7(a) and (b) respectively. In this test the net achieved a score of $\rho_{S,\hat{S}} = 0.77$. Then, we trained the net with a synthetic data set, simulating 2D reflectivity with two apparent faults - one vertical fault, and one diagonal fault, as depicted in Fig. 6. The training reflectivity consists of 30 traces, each of 128 samples, with sampling interval of $T_s = 4$ms. We tested again for the real data. Results are shown in Fig. 7(c) and (d). In this test we got $\rho_{S,\hat{S}} = 0.88$. Fig. 7(e) and (f) present an estimated reflectivity of this seismic data section using recent 3D time-variant seismic deconvolution algorithm (Pereg et al., 2019), and the corresponding reconstructed data which achieved $\rho_{S,\hat{S}} = 0.91$. Fig. 8: shows a zoom into the reflectivity sections in time range 0.3–0.7 s.

Visually examining these reflectivity sections and taking into account the qualitative correlation score, it can be postulated that 77% of the information needed for the net to correctly estimate seismic reflections resides in its ability "to see" zero dip reflections (horizontal layers). Moreover, learning only two types more of possible structures - horizontal and diagonal faults, the estimated reflectivity exceeds the reflectivity achieved by training with complex structure, implying that applying more complex scenarios during training only "confuses" the net. Not only that, but when comparing to other novel time-invariant deconvolution algorithm, in spite of quantitative slightly lower score (correlation of 0.88 vs. 0.91), we can see that layer boundaries in the RNN's estimate are distinct and smooth. Both structural features (such as faults) and stratigraphic features (such as channels) can be observed in the images. Also, the reconstructed seismic data fits to the original given observation. These results are interesting because we would have expected that by training only with simple horizontal layers and simple faults, we would be underfitting the system. One would expect the model to be too simple for the net to be able to learn the underlying complex structures of the data. But comparing with training with complex structure, or in comparison to time-invariant deconvolution (Pereg et al., 2019), we observe more details in the RNN's estimate when trained with horizontal layers and faults. Yet, estimation does deteriorate in deeper layers due to attenuation and dispersion effects, which are not taken into account by the RNN.

Implementation of the above method entails low computational complexity. Training and evaluation were run on a standard workstation equipped with 32 GB of memory, an Intel(R) Core(TM) i7 - 6850 K CPU @ 3.60 GHz, and a NVidia GP102 GeForce GTX 1080 Ti GPU, with 12 GB of video memory. Network training for the last example converges after 17,000 iterations in 10 min only, and the net could be easily applicable to large volumes of data and can be adequate for real-time applications.

5. Conclusions

Recent advances in seismic acquisition has lead to a growing demand to process large amounts of 3D seismic data, in an attempt to provide high resolution subsurface images. Numerous algorithms were developed in an effort to address this problem, but further progress is still necessary. Current methods are often able to produce high-quality images usually depending on correct estimation of many parameters and in the expense of heavy computational burden.

In this work, we have presented a multichannel seismic deconvolution method based on the use of RNN. Namely, the RNN is designed to map a patch of seismic data into a point in the super-resolved reflectivity by learning spatial and temporal connections between close data samples. The performance of our method is demonstrated via experiments with both synthetic and real data. We also presented advantages and disadvantages comparing with recent 3D time-variant deconvolution method. Interestingly, experimental results imply that training using synthetic data of simple structure is superior to training using real data results of other algorithms, complex synthetic data, or a combination of both. Moreover, we do not assume any specific prior regarding the structure of the data. In addition, we have discussed the effect of several characteristics of the data, such as: sparsity, sampling rate and the wavelet's dominant frequency, on the predicted reflectivity. The proposed method is fast and efficient. Therefore, it can be applicable to 3D seismic data as well as other imaging data, such as medical ultrasound imaging.

Future research in exploration seismology can adapt the algorithm to handle non-constant Q layers model, and perhaps explore the use of discontinuity measures and different wavelets in the learning process. A blind or semi-blind deconvolution solution can also be investigated. In terms of the RNN's implementation considering skipping connections (Orhan and Pitkow, 2017), and using dropout (Srivastava et al., 2014) to avoid overfitting can also be studied.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Man, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Vigas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. http://download.tensorflow.org/paper/whitepaper2015.pdf.

Araya-Polo, M., Dahlke, T., Frogner, C., Zhang, C., Poggio, T., Hohl, D., 2017. Automated fault detection without seismic processing. Lead. Edge 36 (3), 208–214.

Bendory, T., Bar-Zion, S.D.A., Feuer, A., 2016. Stable support recovery of stream of pulses with application to ultrasound imaging. IEEE Trans. Signal Process. 64 (14), 3750–3759.

Biondi, B., 2006. 3D Seismic Imaging. Society of Exploration Geophysicists.

Biswas, R., Vassiliou, A., Stromberg, R., Sen, M.K., 2018. Stacking velocity estimation using recurrent neural network. SEG Technical Program Expanded Abstracts 2018, pp. 2241–2245.

CaldernMacas, C., Sen, M.K., Stoffa, P.L., 1998. Automatic NMO correction and velocity estimation by a feedforward neural network. Geophysics 63 (5), 1696–1707.

Candès, E.J., Fernandez-Granda, C., 2013. Super-resolution from noisy data. J. Fourier Anal. Appl. 19 (6), 1229–1254.

Candès, E.J., Fernandez-Granda, C., 2014. Towards a mathematical theory of superresolution. Commun. Pure Appl. Math. 19 (6), 1229–1254.

Chai, X., Wang, S., Yuan, S., Zhao, J., Sun, L., Wei, X., 2014. Sparse reflectivity inversion for nonstationary seismic data. Geophysics 79 (3), V93–V105.

Chen, S.S., Donoho, D.L., Saunders, M.A., 2001. Atomic decomposition by basis pursuit. SIAM Rev. 43 (1), 129–159.

Cohen, I., Coifman, R., 2002. Local discontinuity measures for 3 D seismic data. Geophysics 67 (6), 1933–1945.

Cohen, I., Coult, N., Vassiliou, A.A., 2006. Detection and extraction of fault surfaces in 3D seismic data. Geophysics 71 (4), P21–P27.

Cover, T.M., Thomas, J.A., 2006. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, New York, NY, USA.

Dong, C., Loy, C.C., He, K., Tang, X., 2016. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 38 (2), 295–307.

Donoho, D.L., 1992. Super-resolution via sparsity constraints. SIAM J. Math. Anal. 23 (5), 1309–1331.

Donoho, D., 2006. Compressed sensing. IEEE Trans. Inf. Theory 52 (4), 1289–1306.

Dorrington, K.P., Link, C.A., 2004. Geneticalgorithm/neuralnetwork approach to seismic attribute selection for welllog prediction. GEOPHYSICS 69 (1), 212–221.

Elad, M., 2010. Sparse and Redundant Representations. Springer.

Fernandez-Granda, C., 2013. Support detection in super-resolution. Proceedings of SampTA, pp. 145–148.

Fish, B.C., Kusuma, T., 2005. A neural network approach to automate velocity picking. SEG Technical Program Expanded Abstracts 1994, pp. 185–188.

Forney, D., 1973. The viterbi algorithm. Proc. IEEE, pp. 268–278.

Gelius, L., 1987. Inverse Q filtering. a spectral balancing technique. Geophys. Prospect. (35), 656–667.

Gholami, A., Sacchi, M.D., 2012. A fast and automatic sparse deconvolution in the presence of outliers. IEEE Trans. Geosci. Remote Sens. 50 (10), 4105–4116.

Gholami, A., Sacchi, M.D., 2013. Fast 3D blind seismic deconvolution via constrained total variation and GCV. SIAM J. Imag. Sci. 6 (4), 2350–2369.

Gurbuz, A.C., McClellan, J.H., Scott, W.R., Larson, G.D., 2006. Seismic tunnel imaging and detection. 2006 International Conference on Image Processing, pp. 3229–3232.

Hale, D., 1981. An Inverse Q Filter, Stanford Exploration Project Report. 26 pp. 231–243.

Heimer, A., Cohen, I., 2008. Multichannel blind seismic deconvolution using dynamic programmming. Signal Process. 88 (4), 1839–1851.

Heimer, A., Cohen, I., 2009. Multichannel seismic deconvolution using Markov-Bernoulli Random Field modeling. IEEE Trans. Geosci. Remote Sens. 47 (7), 2047–2058.

Heimer, A., Cohen, I., Vassiliou, A., 2007. Dynamic programming for multichannel blind seismic deconvolution. Proc. Society of Exploration Geophysicist International Conference, Exposition and 77th Annual Meeting, San Antonio, pp. 1845–1849.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory.

Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. 79 (8), 2554–2558.

Idier, J., Goussard, Y., 1993. Multichannel seismic deconvolution. IEEE Trans. Geosci. Remote Sens. 31 (5), 961–979.

Kaaresen, K., Taxt, T., 1998. Multichannel blind deconvolution of seismic signals. Geophysics 63 (6), 2093–2107.

Kjartansson, E., 1979. Constant Q-wave propagtion and attenuation. J. Geophys. Res. (84), 4737–4747.

Li, S., Liu, B., Ren, Y., Chen, Y., Yang, S., Wang, Y., Jiang, P., 2011. Deep Learning Inversion of Seismic Data, CoRR abs/1901.07733. arXiv:1901.07733. http://arxiv.org/abs/1901.07733.

Margrave, G., Lamoureux, M.P., Henleyl, D., 2011. Gabor deconvolution: estimating reflectivity by nonstationary deconvolution of seismic data. Geophysics 76 (3), W15–W30.

McCormack, M.D., Rock, A.D., 1993. Adaptive Network for Automated First Break Picking of Seismic Refraction Events and Method of Operating the Same.

Mendel, J., Kormylo, J., Aminzadeh, F., Lee, J., Habibi-Ashrafi, F., 1981. A novel approach to seismic signal processing and modeling. Geophysics 46, 1398–1414.

Murat, M.E., Rudman, A.J., 1992. Automated first arrival picking: a neural network approach. Geophys. Prospect. 40 (6), 587–604.

Nehme, E., Weiss, L.E., Michaeli, T., Shechtman, Y., 2018. Deep-storm: super-resolution single-molecule microscopy by deep learning. Optica 5 (4), 458–464.

Nguyen, T., Castagna, J., 2010. High resolution seismic reflectivity inversion. J. Seism. Explor. 19 (4), 303–320.

Orhan, A.E., Pitkow, X., 2017. Skip Connections Eliminate Singularities (arXiv: 1701.09175).

Pereg, D., Cohen, I., 2017. Seismic signal recovery based on earth Q model. Signal Process. 137, 373–386. https://doi.org/10.1016/j.sigpro.2017.02.016.

Pereg, D., Cohen, I., Vassiliou, A.A., 2019. Three-dimensional sparse seismic deconvolution based on earth Q model. Signal Process. 154, 97–107.

Pham, M.Q., Duval, L., Chaux, C., Pesquet, J.-C., 2014. A primal-dual proximal algorithm for sparse template-based adaptive filtering: Application to seismic multiple removal. IEEE Trans. Signal Process. 62 (16), 4256–4269.

Ram, I., Cohen, I., Raz, S., 2010. Multichannel deconvolution of seismic signals using statistical MCMC methods. IEEE Trans. Signal Process. 58 (5), 2757–2769.

Repetti, A., Pham, M.Q., Duva, L., Chouzenoux, E., Pesquet, J.-C., 2015. Euclid in taxicab: sparse blind deconvolution with smoothed ℓ_1/ℓ_2 regularization. IEEE Signal Process. Lett. 22 (5), 539–543.

Ricker, N., 1940. The form and nature of seismic waves and the structure of seismogram. Geophysics 5 (4), 348–366.

Riel, P.V., Berkhout, A.J., 1985. Resolution in seismic trace inversion by parameter estimation. Geophysics 50, 1440–1455.

Rosa, A., Ulyrch, T.J., 1991. Processing via spectral modeling. Geophysics 56 (8), 1244–1251.

Sherif, R., Geldart, L., 1983. Exploration Seismology. 2nd edition. Cambridge University Press, UK.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Taylor, H.L., Banks, S.C., McCoy, J.F., 1979. Deconvolution with the ℓ_1 norm. Geophysics 44, 39–52.

Wang, Y., 2008. Seismic Inverse Q-Filtering. Blackwell Pub.

Wang, Y., 2015. Frequencies of the Ricker wavelet. Geophysics 80 (2), A31–A37.

Wiggins, R., 1978. Minimum entropy deconvolution. Geoexploration 16, 21–35.

Yang, F., Wang, W., Ma, J., 2018. Automatic salt detection with machine learning. 80th European Association of Geoscientists and Engineers Conference and Exhibition Extended Abstracts, p. 912.

Zhang, R., Castagna, J., 2011. Seismic sparse-layer reflectivity inversion using basis pursuit decomposition. Geophysics 76 (6), 147–158.

Zhang, C., Ulrych, T., 2002. Estimation of quality factors from CMP record. Geophysics 67 (5), 1542–1547.

Zhang, C., Frogner, C., Araya-Polo, M., Hohl, D., 2014. Machine-learning based automated fault detection in seismic traces. Proceedings of 76th EAGE Conference and Exhibition.