GEOPHYSICS, VOL. 88, NO. 3 (MAY-JUNE 2023); P. V113–V125, 15 FIGS.
10.1190/GEO2021-0521.1

Check for updates

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

Seismic deghosting using convolutional neural networks

Khalid Almuteri[1] and Paul Sava[1]

ABSTRACT

Ghost reflections deteriorate the quality of seismic data in towed-streamer acquisition, and various acquisition and processing solutions have been proposed to remove them from seismic data. A common issue with the proposed solutions is their limited ability to remove source-side ghosts because of the sparse source sampling. Satisfactory receiver-side deghosting solutions are facilitated by complementary measurements (e.g., particle motion data) for wavefield separation and also can be achieved using pressure data acquired at a single recording level only. We develop a solution based on convolutional neural networks (CNNs) to remove source- and receiver-side ghosts in the shot domain. The solution does not require complementary measurements, i.e., it can remove ghost reflections in conventional pressure data measured at a single recording level. Our method requires knowledge of the acquisition geometry to create training data that replicate the field acquisition geometry and require the ocean floor bathymetry to be known. A CNN learns to map ghost-contaminated gathers to corresponding ghost-free gathers through an iterative training process. We find that the CNN-based deghosting operator can remove ghost reflections from previously unseen data and demonstrate that the solution generalizes well when training is done on models unrelated to the actual field geology.

INTRODUCTION

Ghost reflections have been a long-standing issue in towed-streamer seismic acquisition. They introduce frequency notches into the amplitude spectrum, limiting the usable data bandwidth (Amundsen and Zhou, 2013; Dondurur, 2018), thus lowering the seismic resolution (Knapp, 1990). Furthermore, ghost reflections distort the source signature by altering its phase, leading to erroneous impedance inversion (Jovanovich et al., 1983). Ghost-free data are a prerequisite for many geophysical applications. Absent low frequencies make full-waveform inversion more susceptible to cycle skipping (Bunks et al., 1995; Virieux and Operto, 2009). Low frequencies help constrain and improve impedance inversion (Whitcombe and Hodgson, 2007; ten Kroode et al., 2013; Wang et al., 2017). Moreover, low frequencies are less prone to scattering and attenuation, making them ideal for subsalt imaging (Kapoor et al., 2005; ten Kroode et al., 2013).

Conventionally, streamers are towed at a shallow depth to broaden the bandwidth of the seismic signal (Wang et al., 2017). However, shallow tow attenuates low frequencies and yields data with a low signal-to-noise ratio (Kragh et al., 2010; He et al., 2013; Wang et al., 2017). A deep tow has reciprocal effects; it allows for the acquisition of low-frequency content and increases the signal-to-noise ratio, especially at the lower end of the spectrum, at the expense of a narrowband signal. There are two types of ghost reflections that coexist in marine data: source- and receiver-side ghosts. A source-side ghost starts its propagation from a source as an upgoing wave, whereas a receiver-side ghost ends its propagation at a receiver as a downgoing wave. Deghosting solutions fall into two main categories: receiver- and source-side solutions. The categories can be further classified into acquisition and processing based, as discussed next.

Acquisition deghosting

Acquisition solutions for receiver-side deghosting include slanted streamers (Bearnth and Moore, 1989; Soubaras and Dowle, 2010), over/under streamers (Sønneland et al., 1986; Posthumus, 1993; Moldoveanu et al., 2007), shallow-over/deep-under streamers (Kragh et al., 2010), and dual-sensor streamers (Carlson et al., 2007; Tenghamn et al., 2007). In a slanted-streamer acquisition, hydrophones vary in depth along the cable, which results in ghost notch diversity (Bearnth and Moore, 1989; Soubaras and Dowle, 2010). This diversity improves the bandwidth of poststack seismic data (Soubaras and Dowle, 2010; Provenzano et al., 2020), making the solution unsuitable for applications that require ghost-free

Manuscript received by the Editor 10 August 2021; revised manuscript received 12 December 2022; published ahead of production 10 January 2023; published online 29 March 2023.
[1]Colorado School of Mines, Center for Wave Phenomena, Department of Geophysics, Golden, Colorado, USA. E-mail: kalmuteri@mines.edu (corresponding author); psava@mines.edu.
© 2023 Society of Exploration Geophysicists. All rights reserved.

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

prestack data or applications that require receivers to have the same frequency content. An over/under streamer, at which pressure data are acquired at two different reference surfaces, enables removing receiver-side ghosts through an upgoing/downgoing wavefield separation process (Sønneland et al., 1986; Posthumus, 1993; Moldoveanu et al., 2007). However, it requires twice as many streamers with accurate positioning of vertically aligned cables (Reilly, 2016), which is difficult to implement in the field. Kragh et al. (2010) propose a shallow over with a sparse deep-under acquisition configuration. Shallow streamers contribute the mid- and high-frequency content to the seismic data, whereas deep streamers contribute to the low-frequency content. A shallow-over/deep-under acquisition requires fewer streamers than an over/under acquisition because deep streamers are sparse and contribute only low frequencies. However, a shallow-over/deep-under acquisition requires accurate positioning of the cables and a proper interpolation of the low-frequency data to compensate for the sparse acquisition.

Dual-sensor streamers (Carlson et al., 2007; Tenghamn et al., 2007) exploit the fact that upgoing/downgoing wavefield separation is possible if pressure and particle velocity data are measured simultaneously (Claerbout, 1976; Amundsen, 1993; Day et al., 2013; Ikelle and Amundsen, 2018). However, particle motion sensors are sensitive to mechanical noise below 20 Hz (Carlson et al., 2007; Tenghamn et al., 2007; Day et al., 2013). To overcome this issue, one needs to estimate particle motion data below 20 Hz from pressure data. When sources are below the streamer, estimating such data is done by applying a simple operator to pressure data (Amundsen, 1993; Amundsen et al., 1995). For sources shallower than the streamer, one needs to compute the direct arrival and its ghost to estimate the particle motion data. Computing the direct arrival and its ghost requires knowledge of the source function. Alternatively, Özdemir et al. (2012) and Rentsch et al. (2013) propose noise attenuation algorithms to enhance the low-frequency signal-to-noise ratio in the particle motion data. Özdemir et al. (2010), Vassallo et al. (2010), and Özbek et al. (2010) develop optimization-based algorithms to reconstruct and deghost the pressure wavefield using multicomponent marine streamer data. Still, data acquired with different sensor types require calibration with an unknown factor (Grobbe et al., 2016), which can be carried using a deterministic approach (Day et al., 2013) or a statistical approach (Cambois et al., 2009; Alexandrov et al., 2014).

One can extend the acquisition solutions for receiver-side deghosting to address source-side ghosts. Haavik and Landrø (2015) propose varying the source depth during acquisition to diversify frequency notches in the amplitude spectra of different shot gathers. This improves the bandwidth of the seismic signal poststack in a way similar to slanted-streamer acquisition. An over/under source configuration, in which data are acquired twice at the same inline location for sources at different depths (Moldoveanu, 2000), uses reciprocity to remove source-side ghosts in common-receiver gathers (CRGs) (Egan et al., 2007; Moldoveanu et al., 2007). However, source-side deghosting in the CRG domain requires dense shot sampling that towed-streamer acquisition generally lacks.

Processing deghosting

Soubaras (2010) proposes an imaging solution to deghost and improve seismic data poststack. The method works by jointly deconvolving two migration images: a conventional migration image and a mirror migration image. Amundsen and Zhou (2013) derive a low-frequency deghosting filter that can improve the data at least up to half the frequency of the second ghost notch. The solution can be applied in the time-space domain on a trace-by-trace basis (Amundsen and Zhou, 2013; Wang et al., 2017). However, it is only applicable to data acquired with horizontal streamers or streamers with mild variations in depth (Wang et al., 2017). Beasley et al. (2013a, 2013b) propose a physics-based solution to separate 1C measurements, e.g., pressure data acquired at a single reference surface, into upgoing and downgoing wavefields using the wave equation. The wavefield separation process is iterative and uses causality but is sensitive to noise. Zhang and Weglein (2005), Mayhan and Weglein (2013), and Weglein et al. (2013) propose wavefield separation methods based on Green's theorem that can accommodate streamers of arbitrary shapes (Amundsen et al., 2013).

King and Poole (2015) propose a processing workflow to remove receiver-side ghost reflections from pressure-only data that can handle a dynamic sea surface. The method requires an estimation of the sea-surface profile prior to deghosting. Furthermore, the deghosting method is implemented by solving a linear Radon system of equations. Grion et al. (2016) propose a deghosting method that works for rough sea surfaces and slanted streamers, given that the sea-surface profile is known. The method uses phase-shift redatuming operators to separate the up- and downgoing components of the wavefield. Processing solutions for receiver-side deghosting are applicable to source-side ghosts in the CRG domain. However, source-side deghosting is more challenging than receiver-side deghosting and remains mainly an unsolved problem (Amundsen et al., 2017; Ikelle and Amundsen, 2018) because shots are sparse. Vrolijk and Blacquière (2021) propose an approach that uses a convolutional neural network (CNN) to remove source-side ghost reflections in CRGs for coarsly acquired sources. Because the source deghosting process is carried in the common-receiver domain, this solution is applicable to horizontal and slated streamers alike.

In this paper, we propose seismic deghosting using CNNs, which is a special type of artificial neural networks (ANNs). Our solution does not require any particular acquisition configuration but requires that its geometry and the bathymetry of the ocean floor be known. It can remove source- and receiver-side ghosts simultaneously in the shot gather domain. This feature offers an opportunity to deghost conventional pressure data acquired at a single reference surface. We develop a CNN as a deghosting operator that inputs ghost-contaminated data and outputs a ghost-free prediction. In the following sections, we provide the theoretical foundation of CNNs, explain how they work, and discuss their benefits and pitfalls. We demonstrate how CNNs can remove ghost reflections using examples inspired by realistic acquisition and using field data.

CNN-BASED DEGHOSTING

Background

ANNs are machine learning (ML) models inspired by biological neurons found in the brain (Goodfellow et al., 2016; Russell and Norvig, 2016; Géron, 2019). Neural networks (NNs) are capable of tackling complex problems from various fields, such as natural language processing (Sutskever et al., 2014), object recognition (Krizhevsky et al., 2017), and speech recognition (Hinton et al., 2012) with a great degree of success (LeCun et al., 2015). The interest in the geophysics community is not new (van der Baan and

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

Jutten, 2000), but recent advances in ML have expanded their use in the field. Geophysical applications include velocity model building (Araya-Polo et al., 2018; Wu and McMechan, 2018; Yang and Ma, 2019), normal-moveout velocity picking (Biswas et al., 2018; Ma et al., 2018), petrophysical properties and impedance inversion (Biswas et al., 2019; Das et al., 2019; Das and Mukerji, 2020), and seismic deblending (Sun et al., 2020).

Various types of ANNs exist, the simplest of which is fully connected NNs, which consist of layers of neurons stacked together as shown in Figure 1a. The links that connect different neurons represent learnable parameters known as weights. One can find such weights through an iterative data-fitting procedure called training. The weights map inputs to outputs according to

$$a_j^l = g\left(\sum_{i=0}^{n} w_{ij}^l a_i^{l-1}\right), \tag{1}$$

where a_j^l is the value at the jth neuron at the lth layer, w_{ij}^l is the weight that connects the ith neuron at the $(l-1)$th layer with the jth neuron at the lth layer, n is the number of neurons in the $(l-1)$th layer, and g is a nonlinear activation function. Figure 1b shows a visual representation of equation 1. The activation function in the NN is critical; a linear activation function allows only linear mapping between inputs and outputs, whereas a nonlinear activation function enables the NN to learn nonlinear mappings.

A fully connected NN suffers from multiple issues that limit its usefulness in some applications. One such issue is fixed input and output size (LeCun et al., 1998, 2015; Goodfellow et al., 2016). Processing seismic data using NNs requires them to have a structure that can intrinsically handle data with variable recording lengths and variable number of traces. Also, the memory requirement of fully connected NNs with a large input size can be prohibitively expensive (LeCun et al., 1998; Goodfellow et al., 2016; Géron, 2019). In addition, as the input size to an NN increases, the number of learnable parameters increases as well. In such cases, one needs to use more training examples to avoid overfitting (LeCun et al., 1998, 2015; Géron, 2019). Fully connected NNs are not robust to any translation of features (e.g., faults in a stacked section or first break in a shot gather) in the inputs. This implies that if a fully connected NN is trained to detect faults in a stacked section, it will only detect faults in locations where they existed in training examples. LeCun et al. (1998) and Goodfellow et al. (2016) discuss in depth the limitations of fully connected layers and point out the advantages of using CNNs as an alternative. In CNNs, convolutional filters replace the weights characterizing a fully connected layer. CNNs generalize better than fully connected NNs, i.e., they outperform fully connected NNs on previously unseen data (LeCun et al., 2015). In addition, CNNs have fewer parameters than fully connected NNs, making them ideal for complex and large-scale problems.

In the framework of ANNs, a supervised learning approach is the most common form of learning (LeCun et al., 2015). The goal of supervised learning is to find the optimal weights that minimize an objective function given a set of training examples, in which the training examples are in input-output pairs: $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, \ldots, N$. The relationship between the input and output is given by an unknown function f such that $\mathbf{y}_i = f(\mathbf{x}_i)$. The goal of training is to find a function $h_{\mathbf{w}}$, which is represented by an ANN and parameterized by the learnable weights $\mathbf{w}$, which approximates the function f. In our work, the input $(\mathbf{x}_i)$ is a ghost-contaminated shot gather, the output $(\mathbf{y}_i)$ is a true ghost-free gather, and $h_{\mathbf{w}}$ is the deghosting operator. We arrive at function $h_{\mathbf{w}}$ by minimizing the objective function

$$E(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{y}_i - h_{\mathbf{w}}(\mathbf{x}_i)\|^2, \tag{2}$$

with respect to the weights $(\mathbf{w})$ of the NN. A common minimization method used is gradient descent, which iteratively updates the weights using the gradient of the objective function:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \alpha \frac{\partial E(\mathbf{w}_k)}{\partial \mathbf{w}_k}, \tag{3}$$

where α is the learning rate selected experimentally; a small learning rate results in slow convergence, whereas a large learning rate potentially results in divergence (Géron, 2019).

ANNs are useful and powerful ML models. According to the universal approximation theorem (Cybenko, 1989; Hornik et al., 1989; Hornik, 1991), ANNs are universal approximators, i.e., ANNs with a single hidden layer can approximate any continuous function up to an arbitrary order of accuracy. ANNs with two hidden layers, on the other hand, can approximate any function with an arbitrarily small error (Lapedes and Farber, 1987). An exact mapping of input-output pairs is achievable under some constraints (Huang and Huang, 1990; Sartori and Antsaklis, 1991; Tamura and Tateishi, 1997).

The ability to approximate any continuous function arbitrarily well is not unique to ANNs. One can approximate any continuous function up to an arbitrary order of accuracy with polynomials (including trigonometric polynomials) as well (Jeffreys and Jeffreys, 1950; Carothers, 1998; Stein and Shakarchi, 2011). However, the number of coefficients one needs to solve for in polynomial regression makes it prohibitive to approximate high-dimensional functions. The exact number of polynomial coefficients is given by

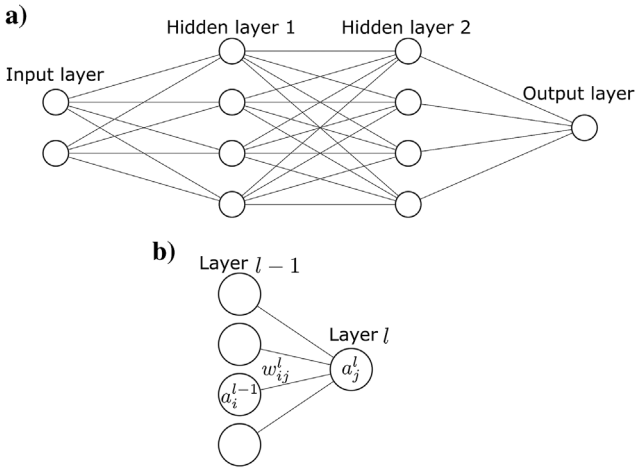

Figure 1. (a) An example of a fully connected NN with two input units, two hidden layers with four units in each layer, and one output unit. (b) A simple model of a neuron connected to four other neurons from a preceding layer. The neuron's output is $a_j^l = g\left(\sum_{i=0}^{n} w_{ij}^l a_i^{l-1}\right)$, where a_j^l is the value at the jth neuron at the lth layer, w_{ij}^l is the weight that connects the ith neuron at the $(l-1)$th layer with the jth neuron at the lth layer, n is the number of neurons in the $(l-1)$th layer, and g is a nonlinear activation function.

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

$$\frac{(n+d)!}{n!d!}, \tag{4}$$

where n is the number of dimensions and d is the degree of the polynomial (Géron, 2019). The prohibitive cost of approximating high-dimensional functions, which includes the number of training examples one needs to avoid overfitting, is commonly known as the "curse of dimensionality" (Hanka and Harte, 1997; Griebel, 2005; Géron, 2019; Grohs et al., 2019; E, 2020). The ability of ANNs to approximate any function, including high-dimensional nonlinear functions, makes them useful and gives rise to a wide range of applications that would not be possible otherwise.

The universal approximation theorem (Cybenko, 1989; Hornik et al., 1989; Hornik, 1991) only addresses the ability of ANNs to represent continuous functions and does not consider the learnability aspects of the problem. The existence of an ANN that can approximate an unknown function does not imply one can find it. Furthermore, one cannot characterize which functions can and cannot be approximated by a particular ANN structure (Russell and Norvig, 2016). An ANN architecture is given by a set of parameters (e.g., number of layers, learning rate, and the type of activation functions) commonly known as hyperparameters. One has to fine tune the hyperparameters, i.e., to experimentally search for the best set of parameters to solve a given problem with adequate accuracy (Goodfellow et al., 2016; Géron, 2019). In the context of seismic deghosting, we make use of the ability of ANNs to exploit any intrinsic relationship between inputs and outputs to construct a deghosting operator. Such an ANN-based operator takes time to train, but after training it can predict ghost-free data in a negligible amount of time. However, to construct the deghosting operator, we must first search for the best set of parameters. The hyperparameters that define the CNN used in our method are as follows: the number of layers, the number of filters, the size of each filter, the learning rate, and the activation function.

Neural network architecture

Our CNN consists of three convolutional layers followed by a masking layer, repeated three times for a total of 12 layers (Figure 2). Each convolutional layer has 16 filters. The first two convolutional layers in the sequence use filters with a temporal length of 101 coefficients and a spatial width of three traces in the inline and crossline directions. The third convolutional layer in the sequence has filters with 11 temporal coefficients and identical spatial width as the preceding two layers. We use tanh as an activation function throughout our NN because it is centered around zero and bounded between -1 and 1.

In CNNs, it is common to use filters with a small size (e.g., 3×3) because they usually outperform large filters and require fewer computations (Géron, 2019). Given that our objective is to remove ghost reflections in the time domain, we use temporally long filters of 101 samples to reconstruct the low-frequency information of the data. In our numerical example, a filter with 101 samples covers 12 times the dominant period of the data. Smaller filters can achieve the same goal if one uses many more layers. We use temporally short filters for every third convolutional layer of 11 samples to reconstruct the high-frequency information. By choosing temporally long and short filters, we give the CNN freedom to work at different scales. Our choices of filter sizes are experimental because a theoretically rigorous approach to select such a parameter remains an open problem in ML. As reported in the literature, optimizing such hyperparameters is an experimental process (Goodfellow et al., 2016; Géron, 2019).

Deghosting can be performed on individual traces using 1D filters or can use information from adjacent traces using 3D filters. The 3D filters should outperform 1D filters because ghost reflections have offset dependency, which can be observed using the 3D ghost function in the Fourier domain (Amundsen, 1993) given by

$$G(k_z, z) = 1 + re^{2ik_z z}, \tag{5}$$

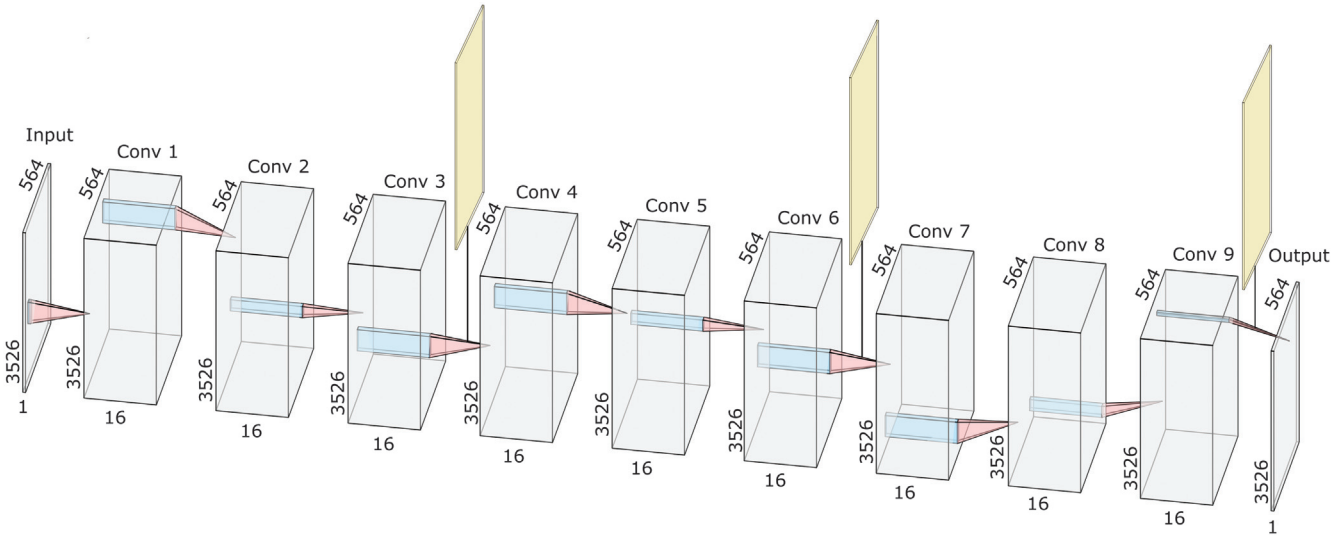

Figure 2. An illustration of the architecture of the network we use for seismic deghosting on 2D data. Each gray cuboid labeled with Conv represents a multichannel feature map generated using N convolutional filters ($N = 16$). The height of each cuboid corresponds to the number of time samples ($n_t = 3526$) and the width corresponds to the number of traces in the gather ($n_r = 564$). Each yellow cuboid represents a temporal masking layer in the network.

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

where $k_z = \sqrt{k^2 - k_x^2 - k_y^2}$ is the vertical wavenumber; $k = (\omega/c)$; $\omega = 2\pi f$ is the angular frequency; c is the water velocity; k_x and k_y are the horizontal wavenumbers in the inline and crossline directions, respectively; z is the source or receiver depth; and r is the reflection coefficient at the sea surface.

We update the weights $(\mathbf{w})$ using the Nesterov's accelerated gradient (NAG) method (Sutskever et al., 2013; Géron, 2019), which is a modified version of the gradient descent (equation 3),

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mathbf{m}_{k+1}, \tag{6}$$

where

$$\mathbf{m}_{k+1} = \beta \mathbf{m}_k - \alpha \frac{\partial E(\mathbf{w}_k + \beta \mathbf{m}_k)}{\partial \mathbf{w}_k} \tag{7}$$

is the momentum $(\mathbf{m}_0 = 0)$, $\beta \in [0,1]$ is the momentum coefficient, and α is the learning rate. The momentum coefficient (β) defines how much previous gradients contribute toward new updates, and if the momentum coefficient is zero, the method reduces to gradient descent. During the training, we use $\beta = 0.9$, i.e., we put much weight on previous directions to speed up the convergence rate. To further speed up convergence, the NAG method computes the gradient of the objective function not at the current location of weights $(\mathbf{w})$ but at the location of weights with a slide in the momentum direction $(\mathbf{w} + \beta\mathbf{m})$. We experimentally determine that the optimum learning rate to train the NN is $\alpha = 0.1$. We use a mini-batch of size 32 and run the training for 150 epochs.

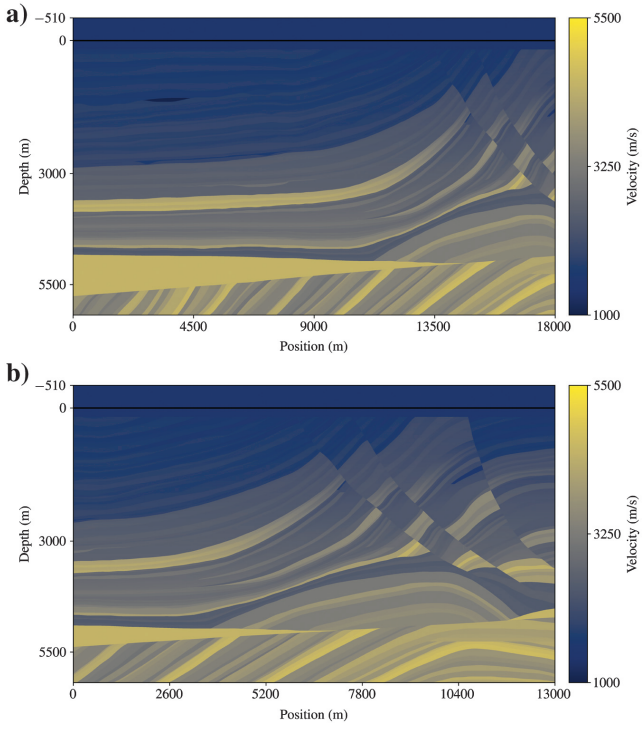

Figure 3. Submodels from the Marmousi model, with 710 m added water layer on top, used for training. The black line corresponds to the air-water interface.

The masking layers in our CNN are essential for two reasons: (1) to eliminate noise the NN produces before first arrivals and (2) to help the NN converge to a solution during training. Without masking layers, training fails because the NN tries to fit the data by removing noise the CNN generates and by removing ghost reflections. To create a masking layer, we use the modified Coppens's method (Sabbione and Velis, 2010) to pick the first seismic signal of any kind (whether it is a water-bottom reflection or refraction from a subsequent layer) along the time axis. The masking array defines the regions where a seismic signal could exist, and it accordingly mutes irrelevant signal, i.e., not generated by the source before the first arrival. Knowing the ocean floor bathymetry allows us to perform a quality control check of the automatic picks from the modified Coppens's method needed to build the masking layer.

NUMERICAL EXAMPLE

In this section, we demonstrate CNN ghost removal in the shot domain on 2D data. First, we describe the process of generating the training and testing data. Then, we illustrate the performance of CNN-based deghosting using a synthetic example and field data.

Neural network training data set

To train our CNN, we use 256 input-output pairs of ghost-free and ghost-contaminated shot gathers simulated using Marmousi submodels (Martin et al., 2002) (Figure 3a and 3b). We add a water layer with 710 m depth to the top of models. We simulate data using a primary source at $z = 5$ m and a mirror source for source-side ghost modeling at $z = -5$ m. In addition, we use a source at $z = -390$ m and a mirror source at $z = -400$ m to model the first-order surface-related multiples and their ghost reflections in data. Modeling first-order surface-related multiples and their ghost reflections increases the ghost-free and ghost-contaminated footprint in the training data, i.e., allows us to use fewer training pairs. We use 564 primary receivers at $z = 7$ m and 564 mirror receivers for receiver-side ghost modeling placed at $z = -7$ m, with a receiver spacing of 12.5 m. For all simulations, we use the same source wavelet with a flat spectrum between 2.5 Hz and 110 Hz. We model data using finite differences for 7 s with a sampling interval of 0.2 ms, later resampled to 2 ms. We model 260 shots from the first training model (Figure 3a) with a shot spacing of 12.5 m

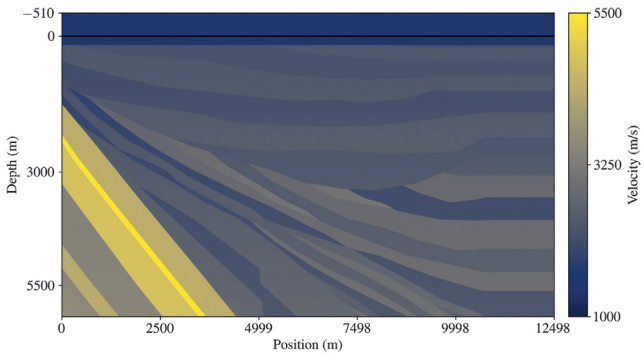

Figure 4. Submodel from the Amoco statics test model to test the CNN. The velocity values of this model are not within the range of the velocity values of the training models. We add a water layer with 710 m depth to the top. The black line corresponds to the air-water interface.

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

and 160 shots from the second training model (Figure 3b) with a shot spacing of 18.5 m. Then, we randomly extract 256 shots from the two modeled data sets for training. Before training the CNN, we normalize the data to the range of the chosen activation function. Our proposed CNN-based deghosting approach reconstructs the relative amplitude of the seismic data because our procedure normalizes the data (input and output) to the range of the chosen activation function.

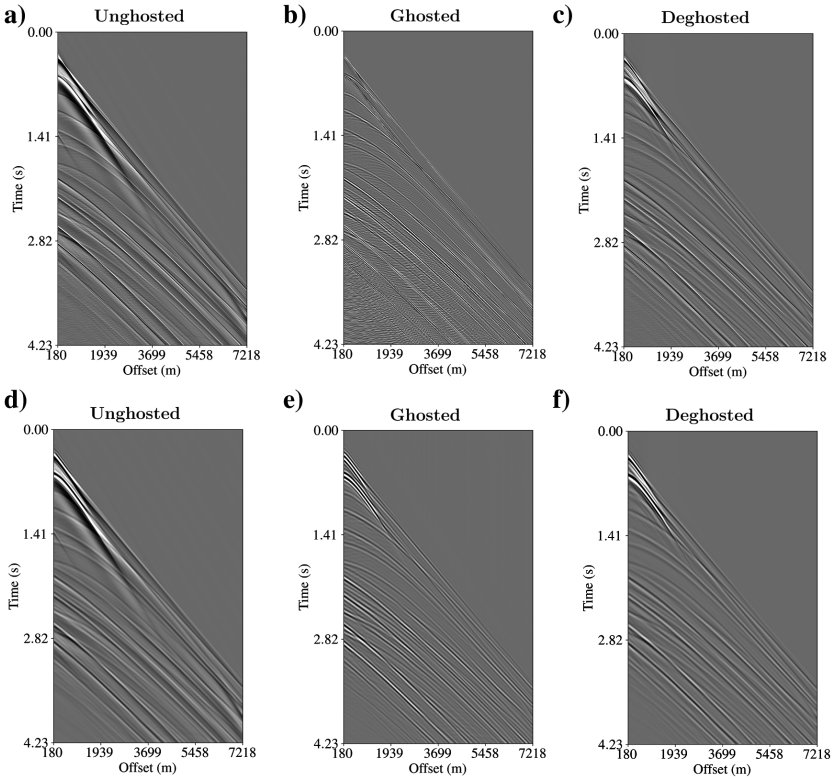

Figure 5. Comparison among (a) true ghost-free, (b) ghost-contaminated, and (c) ghost-free prediction shot gathers. (d–f) After applying 20 Hz high-cut filtering in (a–c), respectively.

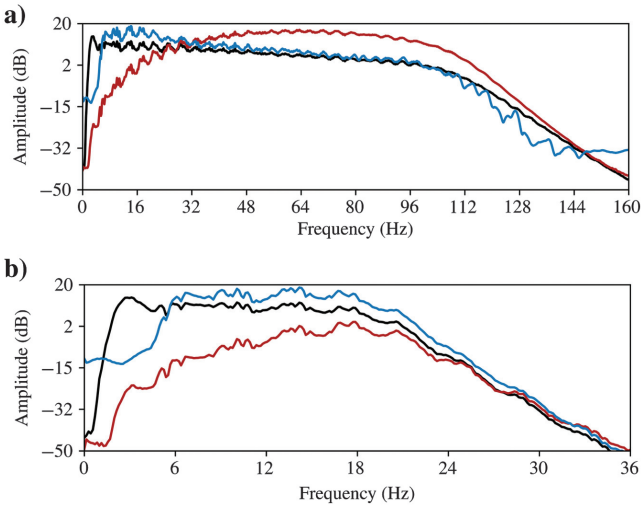

Figure 6. Comparison among the average amplitude spectra of the ghost-free data (black line), the ghost-contaminated data (red line), and the ghost-free prediction data (blue line). (a) Comparison between the full bandwidth data shown in Figure 5a–5c. (b) Comparison between the band-limited data shown in Figure 5d–5f.

Seismic deghosting: Synthetic data example

To test our method, we extract a submodel (Figure 4) from the Amoco statics test model (O'Brien, 1994). The velocity values of the testing submodel are not within the range of the training submodels. We choose this testing model to show that with simpler training models, one can remove ghost reflections from more complex models and lower the computational burden of generating training data simultaneously. By complex models, we mean models with stronger velocity contrasts and velocity values outside the range of that of the training models. We generate two gathers, with and without ghosts, for a source at $x = 2303$ m using the same source wavelet that we use to generate the training data. The source depth, receiver depth, receiver spacing, and modeling parameters of the testing data are similar to the training data.

Figure 5a and 5b shows the true ghost-free and ghost-contaminated data generated using the model shown in Figure 4, respectively. The ghost-free and ghost-contaminated gathers have different phases, and the low frequencies in the ghost-contaminated gather are attenuated by ghost reflections. The average amplitude spectra of the two gathers indicate that ghost reflections destructively interfere with the primary signal at 0 Hz and constructively interfere at 60 Hz (Figure 6a). The destructive interference is more significant at low frequencies than at high frequencies. The amount of energy loss at 5 Hz is 28 dB, whereas the amount of energy loss at 20 Hz is 5 dB.

Using the ghost-contaminated gather (Figure 5b) as an input to the CNN-based deghosting operator gives the ghost-free prediction shown in Figure 5c. The true and predicted ghost-free gathers closely match in character. The average amplitude spectra (Figure 6a) indicate that the CNN recovers the spectrum correctly within the constructive interference frequency bands but less accurately within the destructive interference bands. The relative ℓ^2 residual for the predicted ghost-free gather is 0.65, whereas it is 2.04 for the ghost-contaminated gather. Figure 7 shows the relative ℓ^2 residual for the predicted ghost-free data and ghost-contaminated data as a function of frequency band. The bandwidth of each band is 3.5 Hz, and the center of each band is changing by 3.5 Hz, starting from 4.5 Hz. Figure 7 shows the improvement across different frequency bands as a result of ghost removal. However, the frequency bands centered at approximately 4.5 Hz and 18.25 Hz show higher relative error compared with other bands. The relatively high ℓ^2 for the ghost-free prediction data is because the minimization problem is formulated to reconstruct the relative amplitude instead of the absolute amplitude.

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

Figure 5d–5f shows the true ghost-free, the ghost-contaminated, and the ghost-free prediction gathers after applying a 20 Hz low-pass filter to the gathers shown in Figure 5a–5c, respectively. The ghost-contaminated gather is characterized by weaker amplitudes, indicating that ghost reflections severely reduce the signal-to-noise ratio at low frequencies (Figure 6b). Our method can recover the low-frequency content in the data. Figure 8 shows the frequency-wavenumber spectra of the gathers shown in Figure 5. The low frequencies (below 30 Hz) are boosted, and the frequencies within the constructive interference band are reduced. However, frequencies below 5.5 Hz are not recovered well, and the frequencies between 5.5 and 30 are overcompensated.

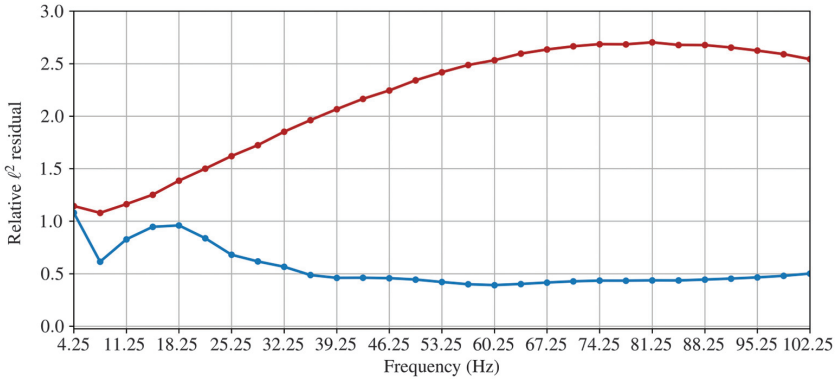

Figure 7. Relative ℓ^2 residual between the true ghost-free gather (Figure 5a) and the ghost-contaminated gather (Figure 5b) (red) over different frequency bands, with a bandwidth of 3.5 Hz for each band, and between the true ghost-free gather and the ghost-free prediction (Figure 5c) (blue).

Individual traces for a receiver 3930 m away from the source (Figure 9a and 9b) confirm the ability of the CNN to remove ghost reflections. The correlation coefficient between the true ghost-free and ghost-contaminated traces is 0.31, indicating that ghost reflections severely alter the waveform of the source signature, in terms of amplitude (Figure 9c) and phase. However, the correlation coefficient between the true and predicted traces is 0.91, indicating that the CNN-based deghosting operator can recover with a great degree of accuracy both the amplitude (Figure 9d) and phase of the true ghost-free data above 5.5 Hz. The ghost-free prediction improves when

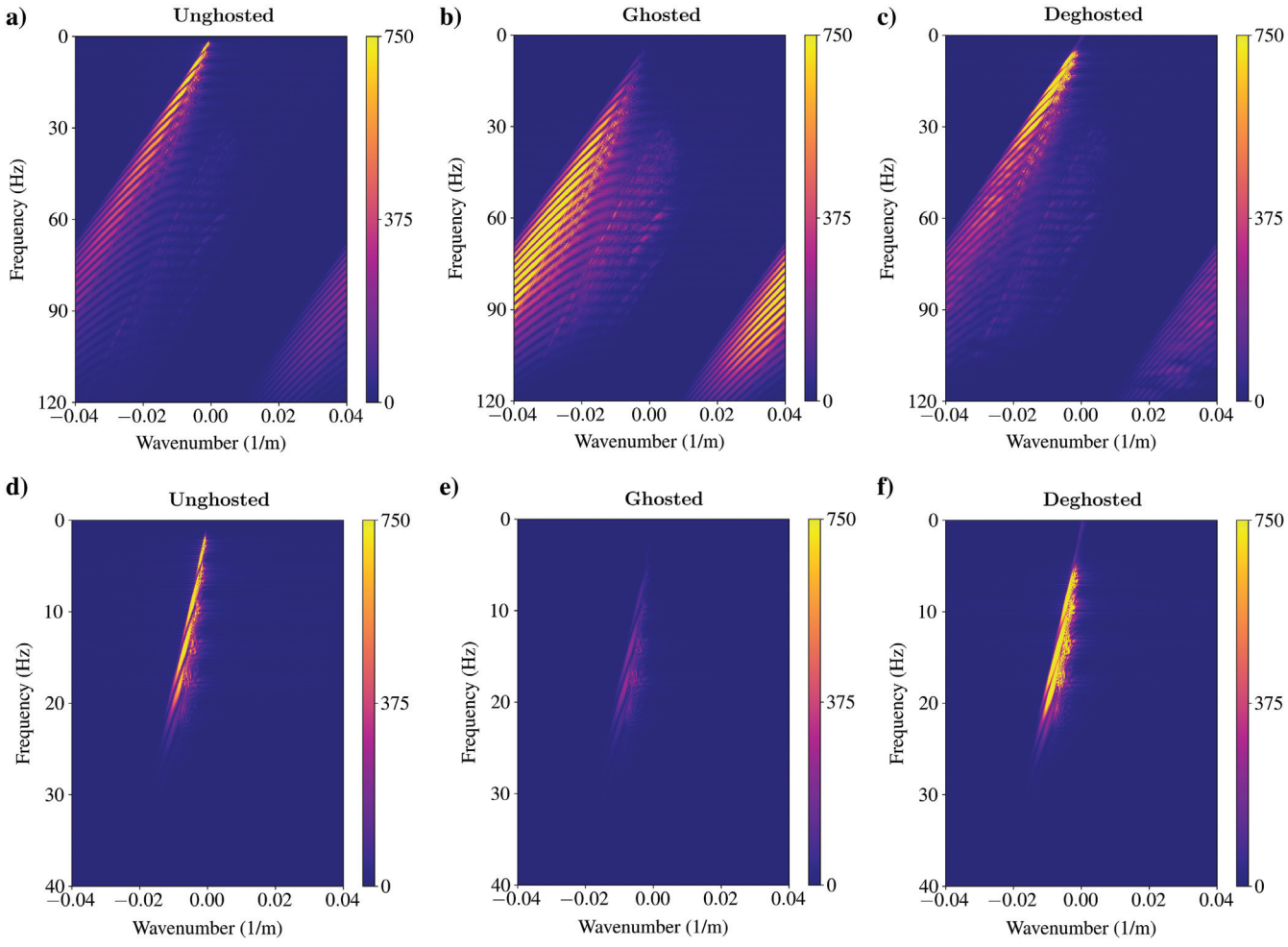

Figure 8. The frequency-wavenumber spectra of (a) Figure 5a, (b) Figure 5b, (c) Figure 5c, (d) Figure 5d, (e) Figure 5e, and (f) Figure 5f.

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

seismic events have comparable amplitude range. Therefore, removing direct arrival and ocean-bottom reflections improves the ghost-free prediction of reflections from subsequent layers and also improves the ghost-free prediction of refraction data.

Low-frequency traces from the true ghost-free, the ghost-contaminated, and the ghost-free prediction gathers for a receiver 3930 m away from the source are shown in Figure 10a and 10b. Figure 10c and 10d shows the corresponding amplitude spectra for the ghost-contaminated and ghost-free prediction traces, respectively. The reconstructed trace matches well the true ghost-free trace in amplitude and phase. The low-frequency ghost-free prediction is nearly noise free, indicating that noise present in the ghost-free prediction is mainly at high frequencies. However, such noise is not introduced by the CNN but appears to be high-frequency residuals from the ghost-contaminated input (Figure 9a and 9b).

The correlation coefficients between individual traces as a function of offset for the full bandwidth and band-limited data are shown in Figure 11a and 11b, respectively. The correlation coefficients indicate that ghost reflections introduce distortion that is a function of offset and frequency. The overall correlation coefficient for the full bandwidth traces between true and predicted ghost-free data is approximately 0.9. However, for traces recorded by receivers centered approximately 1939 m away from the source, there is a drop of approximately 20% in the correlation coefficients, which is likely caused by events crossing (i.e., events with different moveouts). The correlation coefficients between the band-limited traces indicate that low-frequency reconstruction is poor at offset shorter than 1939 m.

Seismic deghosting: Field data example

We test our CNN deghosting method on conventional streamer data acquired offshore of Australia. The data are acquired with a source spacing of 18.5 m and receiver spacing of 12.5 m. The source depth is 5 m, and the cable depth is 7 m. For this test, we use 608 shots to demonstrate the deghosting result pre- and poststack. We preprocess the field data by applying a 2.5–110 Hz band-pass filter.

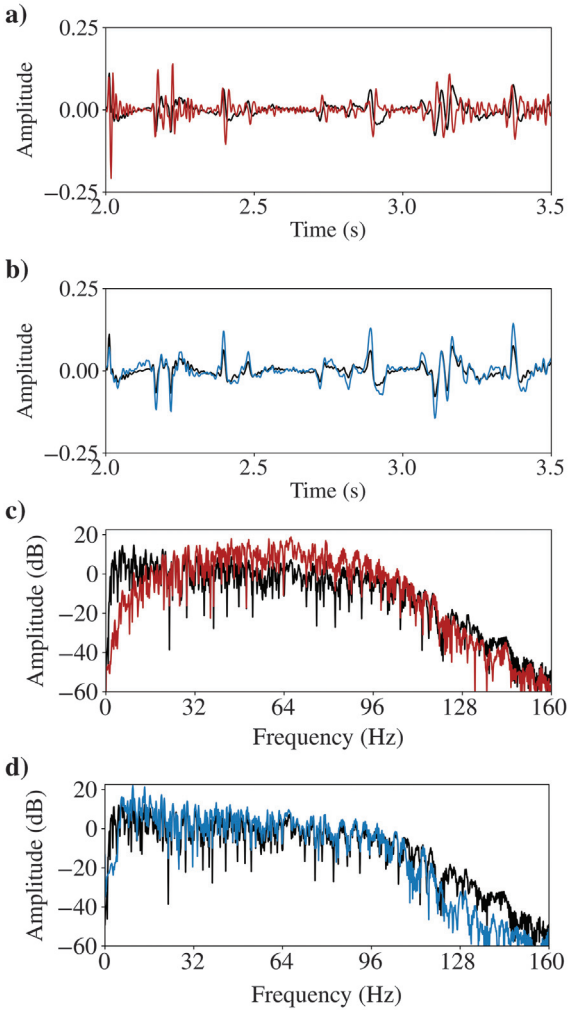

Figure 9. Comparison between traces recorded at a receiver that is 3930 m away from the source. (a) Comparison between ghost-free (black line) and ghost-contaminated (red line) traces. (b) Comparison between ghost-free and ghost-free prediction (blue line) traces. (c and d) Comparison between the amplitude spectra of the traces shown in (a and b), respectively.

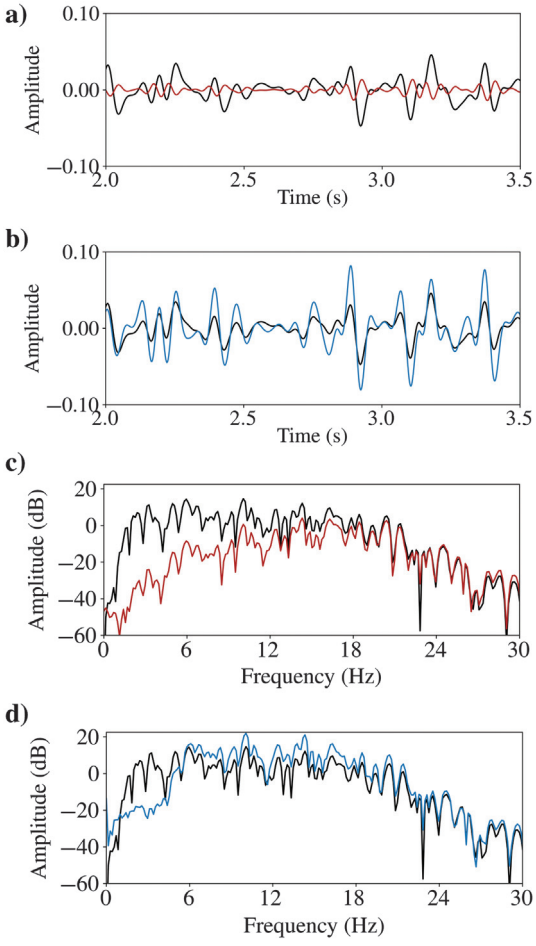

Figure 10. Comparison between traces recorded at a receiver that is 3930 m away from the source after applying a 20 Hz low-pass filter. (a) Comparison between ghost-free (black line) and ghost-contaminated (red line) traces. (b) Comparison between ghost-free and ghost-free prediction (blue line) traces. (c and d) Comparison between the amplitude spectra of the traces shown in (a and b), respectively.

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

Because the field data acquisition parameters match that of the training data, we use the same trained NN from the synthetic example to predict ghost-free data in this example.

Figure 12a and 12b shows a ghost-contaminated input gather to the CNN and the ghost-free prediction, respectively. The two gathers differ by the enhanced low-frequency content after the CNN deghosting (Figure 13a). The phase change in the data is not clearly visible in the full bandwidth gathers (Figure 12a and 12b), but after applying a 20 Hz low-pass filter (Figure 12c and 12d), the phase change becomes clear (e.g., at 1.4 s). The average amplitude spectra of the two band-limited gathers (Figure 13b) illustrate the change in frequency content for low frequencies. The amplitude spectra show that there is a 16 dB increase in amplitude in the range 2.5–20 Hz on average, compensating for the attenuated low frequencies due to ghost reflections.

Figure 14a and 14b shows the stacked data without and with CNN-based deghosting, respectively. The arrows point to improvements due to deghosting. Horizons are more coherent and appear continuous due to the ghost removal. The waveform in the ghost-contaminated section has a reverberatory character that lowers the resolution of the data, which is not observed in the ghost-free prediction section. Horizons in the deghosted sections are more compact and clearer, making the interpretation easier. Figure 15a shows the average amplitude spectra of the two stacked sections. Although the amplitude spectra are almost identical, the stacked sections are different in terms of horizons continuity and compactness. Part of the processing sequence that we use to generate the stacked sections is predictive deconvolution. As a consequence, the average amplitude spectra of the ghost-contaminated and ghost-free prediction are very similar. Nonetheless, the stacked sections are very different, with a clear improvement to the continuity of horizons after deghosting. Using the recorded source signature to deconvolve the deghosted data is not an option because the source signature has the source ghost. Using the recorded source signature to deconvolve the deghosted data requires that we train the CNN to remove receiver-side ghost reflections only.

Given the importance of low-frequency information in seismic data, we apply a 20 Hz low-pass filter to the stacked sections shown in Figure 14a and 14b (Figure 14c and 14d). For such low frequencies, the continuity and coherency of seismic events after CNN-based ghost removal become more evident. Figure 15b compares the average amplitude spectra of the stacked sections after a 20 Hz low-pass filter. The average amplitude spectra of the ghost-contaminated and ghost-free prediction are very similar, demonstrating the importance of phase correction in the ghost removal process. Furthermore, the band-limited sections show that ghost reflections alter the source signature more significantly at the lower end of the spectrum than at the higher end. This alteration is more significant at low frequencies because the phase of the source- and receiver-side ghosts coincides at 0 Hz. Because our training data are noise free, the field example shows that CNN deghosting is robust to random noise because the presence of such noise in field data is inevitable.

DISCUSSION

The numerical example demonstrates the ability of CNNs to remove source- and receiver-side ghosts from conventional pressure data acquired at a single reference surface in the shot gather domain. The advantage of using our approach is that source and receiver deghosting can be performed as a preprocessing step in the field by constructing the CNN-based deghosting operator prior to acquisition. This solution does not make assumptions about the acquisition geometry but requires it to be known to create training data that match the actual field acquisition geometry. We assume that data are acquired under a flat sea surface, which is not always a feature of marine acquisitions. Cecconello et al. (2018), Blacquière and Sertlek (2019), and Konuk and Shragge (2020) show that a dynamic sea surface affects the phase and amplitude of seismic data. However, the field example indicates that a flat sea surface assumption is still valid for ghost removal using a deep learning approach. More research is needed to assess the sea state conditions under which our proposed method is applicable. Our field example uses receivers at depth between 6.2 m and 7.7 m, with a nominal streamer depth of 7 m, showing that our

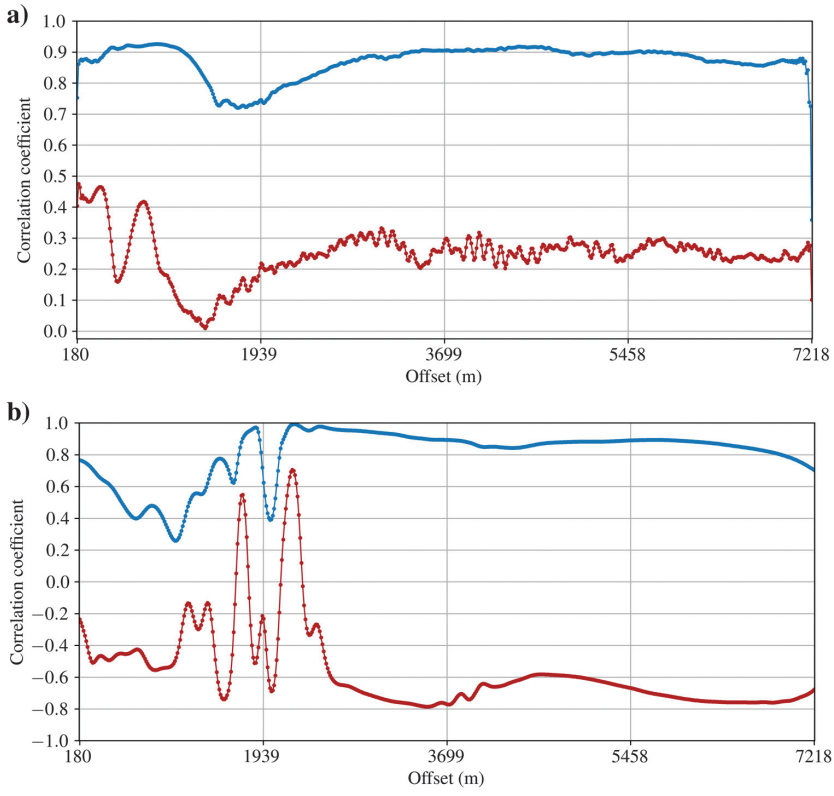

Figure 11. (a) The correlation coefficient between the true ghost-free data and ghost-contaminated data (red) and true ghost-free data and ghost-free prediction (blue) for the full bandwidth individual traces. (b) The correlation coefficient between the true-ghost free data and ghost-contaminated data (red) and true ghost-free data and ghost-free prediction (blue) for the band-limited individual traces.

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

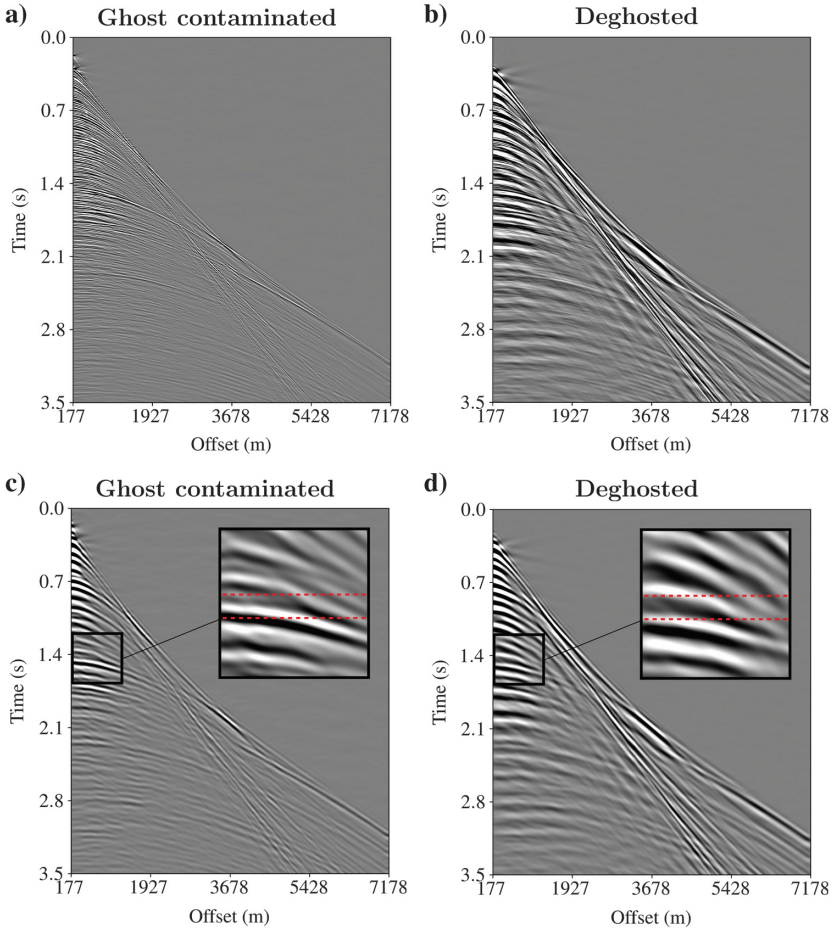

Figure 12. Comparison between (a) ghost-contaminated and (b) ghost-free prediction shot gathers. (c and d) After applying 20 Hz high-cut filtering in (a and b), respectively. The dashed red lines are for cross referencing.

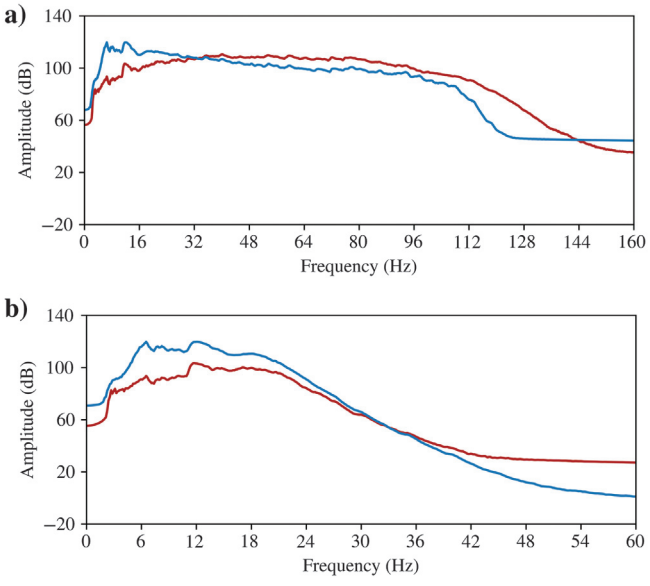

Figure 13. Comparison between the average amplitude spectra of the ghost-contaminated data (red line) and the ghost-free prediction data (blue line). (a) Comparison between the full bandwidth data shown in Figure 12a and 12b. (b) Comparison between the band-limited data shown in Figure 12c and 12d.

method can tolerate depth variations. Changing the acquisition geometry does not require rebuilding the CNN-based deghosting operator, and one can use transfer learning to retrain the NN for use in other field acquisition setups. In our work, we assume a horizontal streamer acquisition with no variation in geometry in the crossline direction. Thus, investigating 3D effects (e.g., feathering) under a 2D acquisition assumption would be needed and remains a subject for future work.

Although the ghost reflections are independent of the source function, they are dependent on the bandwidth of the seismic signal, as shown in equation 5. However, given that our solution is based on a data-fitting procedure, its applicability to field data depends on its ability to tolerate source wavelet uncertainties (i.e., depends on the ability of the CNN to learn to remove ghost reflections for arbitrary source functions). Our field data example shows that the proposed approach improves the data quality even though the source wavelet that we use to create the training data is different from the source wavelet of the field data. Moreover, our field data results use a purely statistical source signature deconvolution approach instead of a deterministic approach, thus demonstrating the improvement in the waveform character even when using a statistical deconvolution technique due to ghost removal.

The synthetic example shows that the NN is less capable of reconstructing frequencies below 5.5 Hz, indicating a need to increase the size of the convolutional filters, i.e., increase the number of trainable parameters in the NN. Prior publications by Sun and Demanet (2020), for low-frequency extrapolation, and by Almuteri and Sava (2021), for seismic deghosting, show that the ability to reconstruct data with a certain bandwidth is proportional to the capacity of the NN (i.e., the number of trainable parameters). Incorporating additional information (e.g., particle motion data) into our method could improve deghosting. Our formulation of the CNN reconstructs the relative amplitude of the seismic data rather than the absolute amplitude because of our choice of a bounded activation function, to whose range we normalize the input and output training data. We leave extending our approach to recover the absolute amplitude for future work. Using the method described in this paper, we can remove source-side ghost reflections in the shot domain without a need to transform the data to the common-receiver domain and without a need for shot interpolation because (1) the imprints of the source-side ghosts are present in shot gathers; otherwise, a mere domain transformation (to a CRG) would not unveil the source-side ghost, and (2) input-output pairs of ghost-contaminated and ghost-free data are the only way for the NN to learn through supervised learning how to map between the two, thus reducing the problem to amplitude and phase correction. Self-supervised learning is an alternative approach to supervised learning that removes the burden of generating input-output pairs of training data. It is applied in geophysics for noise

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

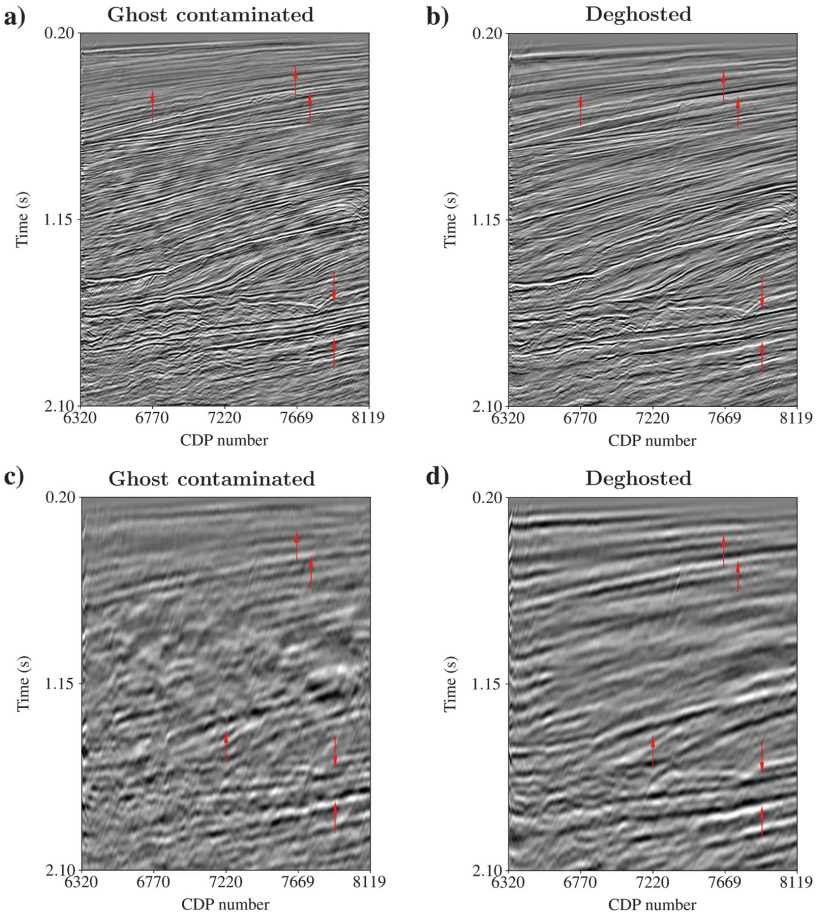

Figure 14. Comparison between (a) ghost-contaminated and (b) ghost-free prediction stacked sections. (c and d) Comparison between the sections shown in (a and b), respectively, after applying a 20 Hz low-pass filter.

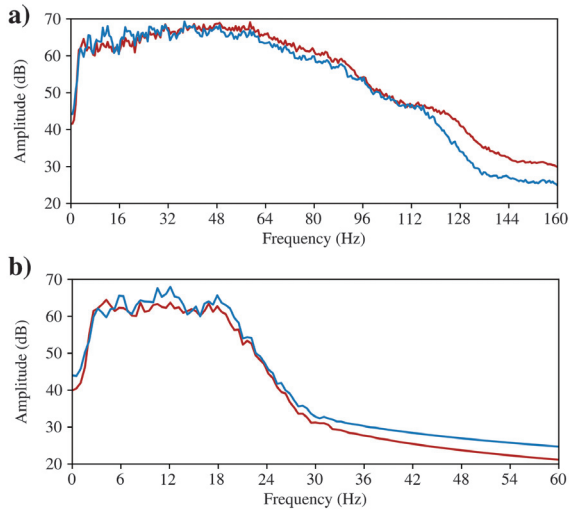

Figure 15. Comparison between the average amplitude spectra of the ghost-contaminated (red line) and ghost-free prediction stacked sections. (a) Comparison between the full bandwidth sections shown in Figure 14a and 14b. (b) Comparison between the band-limited sections shown in Figure 14c and 14d.

suppression (Birnie et al., 2021; Liu et al., 2022a, 2022b) and low-frequency extrapolation (Hu et al., 2020; Wang et al., 2020), and it is showing promising results. Self-supervised learning is applicable to seismic deghosting within a physics-guided NN framework.

CONCLUSION

We propose a method to remove source- and receiver-side ghosts in the shot domain. The solution works for pressure data acquired at a single recording level and uses a CNN with a supervised learning approach to map ghost-contaminated data to ghost-free data. Our solution alleviates acquisition and processing complexities associated with acquisition systems that use complementary measurements. Because our solution is based on learning through data fitting, the acquisition geometry and the bathymetry of the ocean floor must be known to create training data that closely mimic field data acquisition conditions. We demonstrate with a synthetic example and field data that CNNs can predict ghost-free data with a reasonable accuracy. Recovering the low frequencies remains a challenge because of the limited capacity of the CNN, i.e., its effectiveness increases proportionally with the number of trainable parameters. Still, we show that the solution is robust to data generated using a model with strong velocity contrasts and velocity values beyond that of the models used to generate the training data. We also show that the method is robust to inputs with random noise, i.e., field data, even though the training data are noise free. Finally, we show that a simple amplitude correction is not sufficient to recover the ghost-free data and highlight the importance of phase correction in the ghost removal process.

ACKNOWLEDGMENTS

The first author would like to thank Saudi Aramco for graduate study sponsorship. We thank the sponsors of the Center for Wave Phenomena at Colorado School of Mines for the support of this research. We also thank CGG for providing the field data and T. Konuk (CWP) for many useful discussions and for providing the forward-modeling code that made this research possible.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are confidential and cannot be released.

REFERENCES

Alexandrov, D., B. Kashtan, A. Bakulin, and P. Leger, 2014, Dual-sensor summation with buried land sensors: 84th Annual International Meeting, SEG, Expanded Abstracts, 1929–1933, doi: 10.1190/segam2014-1082.1.
Almuteri, K., and P. Sava, 2021, A convolutional neural network approach for ghost removal: First International Meeting for Applied Geoscience &

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

Energy, SEG, Expanded Abstracts, 2550–2554, doi: 10.1190/segam2021-3584059.1.
Amundsen, L., 1993, Wavenumber-based filtering of marine point-source data: Geophysics, **58**, 1335–1348, doi: 10.1190/1.1443516.
Amundsen, L., Ø. Pedersen, A. Osen, J. O. Robertsson, and M. Landrø, 2017, Broadband seismic over/under sources and their designature-deghosting: Geophysics, **82**, no. 5, P61–P73, doi: 10.1190/geo2016-0512.1.
Amundsen, L., B. G. Secrest, and B. Arntsen, 1995, Extraction of the normal component of the particle velocity from marine pressure data: Geophysics, **60**, 212–222, doi: 10.1190/1.1443749.
Amundsen, L., and H. Zhou, 2013, Low-frequency seismic deghosting: Geophysics, **78**, no. 2, WA15–WA20, doi: 10.1190/geo2012-0276.1.
Amundsen, L., H. Zhou, A. Reitan, and A. Weglein, 2013, On seismic deghosting by spatial deconvolution: Geophysics, **78**, no. 6, WA55–WA70, doi: 10.1190/geo2013-0198.1.
Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke, 2018, Deep-learning tomography: The Leading Edge, **37**, 58–66, doi: 10.1190/tle37010058.1.
Bearnth, R. E., and N. A. Moore, 1989, Air gun-slant cable seismic results in the Gulf of Mexico: 59th Annual International Meeting, SEG, Expanded Abstracts, 649–652, doi: 10.1190/1.1889556.
Beasley, C. J., R. Coates, Y. Ji, and J. Perdomo, 2013a, Wave equation receiver deghosting: A provocative example: 83rd Annual International Meeting, SEG, Expanded Abstracts, 4226–4230, doi: 10.1190/segam2013-0693.1.
Beasley, C. J., R. Coates, and C. Lapilli, 2013b, Wave equation receiver deghosting: 5th IEEE International Workshop on Computational Advances in Multi-sensor Adaptive Processing, 280–283, 10.1109/CAMSAP.2013.6714062.
Birnie, C., M. Ravasi, and T. Alkhalifah, 2021, Self-supervised learning for random noise suppression in seismic data: First International Meeting for Applied Geoscience & Energy, SEG, Expanded Abstracts, 2869–2873, doi: 10.1190/segam2021-3583248.1.
Biswas, R., M. K. Sen, V. Das, and T. Mukerji, 2019, Prestack and poststack inversion using a physics-guided convolutional neural network: Interpretation, **7**, no. 3, SE161–SE174, doi: 10.1190/INT-2018-0236.1.
Biswas, R., A. Vassiliou, R. Stromberg, and M. K. Sen, 2018, Stacking velocity estimation using recurrent neural network: 88th Annual International Meeting, SEG, Expanded Abstracts, 2241–2245, doi: 10.1190/segam2018-2997208.1.
Blacquière, G., and H. O. Sertlek, 2019, Modeling and assessing the effects of the sea surface, from being flat to being rough and dynamic: Geophysics, **84**, no. 2, T13–T27, doi: 10.1190/geo2018-0294.1.
Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473, doi: 10.1190/1.1443880.
Cambois, G., D. Carlson, C. Jones, M. Lesnes, W. Söllner, and H. Tabti, 2009, Dual-sensor streamer data: Calibration, acquisition QC and attenuation of seismic interferences and other noises: 79th Annual International Meeting, SEG, Expanded Abstracts, 142–146, doi: 10.1190/1.3255117.
Carlson, D. H., A. Long, W. Söllner, H. Tabti, R. Tenghamn, and N. Lunde, 2007, Increased resolution and penetration from a towed dual-sensor streamer: First Break, **25**, 71–77, doi: 10.3997/1365-2397.25.12.27722.
Carothers, N. L., 1998, A short course on approximation theory: Bowling Green State University.
Cecconello, E., E. G. Asgedom, O. C. Orji, M. W. Pedersen, and W. Söllner, 2018, Modeling scattering effects from time-varying sea surface based on acoustic reciprocity: Geophysics, **83**, no. 2, T49–T68, doi: 10.1190/geo2017-0410.1.
Claerbout, J. F., 1976, Fundamentals of geophysical data processing: McGraw-Hill.
Cybenko, G., 1989, Approximation by superpositions of a sigmoidal function: Mathematics of Control, Signals, and Systems, **2**, 303–314, doi: 10.1007/BF02551274.
Das, V., and T. Mukerji, 2020, Petrophysical properties prediction from pre-stack seismic data using convolutional neural networks: Geophysics, **85**, no. 5, N41–N55, doi: 10.1190/geo2019-0650.1.
Das, V., A. Pollack, U. Wollner, and T. Mukerji, 2019, Convolutional neural network for seismic impedance inversion: Geophysics, **84**, no. 6, R869–R880, doi: 10.1190/geo2018-0838.1.
Day, A., T. Klüver, W. Söllner, H. Tabti, and D. Carlson, 2013, Wavefield-separation methods for dual-sensor towed-streamer data: Geophysics, **78**, no. 2, WA55–WA70, doi: 10.1190/geo2012-0302.1.
Dondurur, D., 2018, Acquisition and processing of marine seismic data: Elsevier, doi: 10.1016/C2016-0-01591-7.
E, W., 2020, Machine learning and computational mathematics: Communications in Computational Physics, **28**, 1639–1670, doi: 10.4208/cicp.OA-2020-0185.
Egan, M., K. G. El-Kasseh, and N. Moldoveanu, 2007, Full deghosting of OBC data with over/under source acquisition: 77th Annual International Meeting, SEG, Expanded Abstracts, 31–35, doi: 10.1190/1.2792376.
Géron, A., 2019, Hands-on machine learning with scikit-learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems: O'Reilly Media.
Goodfellow, I., Y. Bengio, A. Courville, and Y. Bengio, 2016, Deep learning: MIT Press.
Griebel, M., 2005, Sparse grids and related approximation schemes for higher dimensional problems: SFB.
Grion, S., R. Telling, and S. Holland, 2016, Rough sea estimation for phase-shift deghosting: 86th Annual International Meeting, SEG, Expanded Abstracts, 5129–5133, doi: 10.1190/segam2016-13688475.1.
Grobbe, N., J. van der Neut, E. Slob, K. Wapenaar, C. A. Vidal, and G. Drijkoningen, 2016, Unified multi-depth-level field decomposition: Geophysical Prospecting, **64**, 361–391, doi: 10.1111/1365-2478.12290.
Grohs, P., A. Jentzen, and D. Salimova, 2019, Deep neural network approximations for Monte Carlo algorithms: arXiv preprint, doi: 10.48550/arXiv.1908.10828.
Haavik, K. E., and M. Landrø, 2015, Variable source depth acquisition for improved marine broadband seismic data: Geophysics, **80**, no. 3, A69–A73, doi: 10.1190/geo2014-0437.1.
Hanka, R., and T. P. Harte, 1997, Curse of dimensionality: Classifying large multi-dimensional images with neural networks, *in* M. Kárný and K. Warwick, eds., Computer intensive methods in control and signal processing: Springer, 249–260, doi: 10.1007/978-1-4612-1996-5_15.
He, J.-W., W.-K. Lu, and Z.-X. Li, 2013, An adaptive over/under data combination method: Applied Geophysics, **10**, 469–476, doi: 10.1007/s11770-013-0397-x.
Hinton, G., L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, 2012, Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups: IEEE Signal Processing Magazine, **29**, 82–97, doi: 10.1109/MSP.2012.2205597.
Hornik, K., 1991, Approximation capabilities of multilayer feedforward networks: Neural Networks, **4**, 251–257, doi: 10.1016/0893-6080(91)90009-T.
Hornik, K., M. Stinchcombe, and H. White, 1989, Multilayer feedforward networks are universal approximators: Neural Networks, **2**, 359–366, doi: 10.1016/0893-6080(89)90020-8.
Hu, W., Y. Jin, X. Wu, and J. Chen, 2020, Physics-guided self-supervised learning for low frequency data prediction in FWI: 90th Annual International Meeting, SEG, Expanded Abstracts, 875–879, doi: 10.1190/segam2020-3423396.1.
Huang, S., and Y. Huang, 1990, Bounds on number of hidden neurons of multilayer perceptrons in classification and recognition: IEEE International Symposium on Circuits and Systems, 2500–2503, doi: 10.1109/ISCAS.1990.112518.
Ikelle, L. T., and L. Amundsen, 2018, Introduction to petroleum seismology: SEG, doi: 10.1190/1.9781560803447.
Jeffreys, H., and B. Jeffreys, 1950, Methods of mathematical physics: Cambridge University Press.
Jovanovich, D. B., R. D. Sumner, and S. L. Akins-Easterlin, 1983, Ghosting and marine signature deconvolution: A prerequisite for detailed seismic interpretation: Geophysics, **48**, 1468–1485, doi: 10.1190/1.1441431.
Kapoor, J., C. Stork, and M. Egan, 2005, Benefits of low frequencies for subsalt imaging: 75th Annual International Meeting, SEG, Expanded Abstracts, 1993–1996, doi: 10.1190/1.2148099.
King, S., and G. Poole, 2015, Hydrophone-only receiver deghosting using a variable sea surface datum: 85th Annual International Meeting, SEG, Expanded Abstracts, 4610–4614, doi: 10.1190/segam2015-5891123.1.
Knapp, R. W., 1990, Vertical resolution of thick beds, thin beds, and thin-bed cyclothems: Geophysics, **55**, 1183–1190, doi: 10.1190/1.1442934.
Konuk, T., and J. Shragge, 2020, Modeling full-wavefield time-varying sea-surface effects on seismic data: A mimetic finite-difference approach: Geophysics, **85**, no. 2, T45–T55, doi: 10.1190/geo2019-0181.1.
Kragh, E., E. Muyzert, T. Curtis, M. Svendsen, and D. Kapadia, 2010, Efficient broadband marine acquisition and processing for improved resolution and deep imaging: The Leading Edge, **29**, 464–469, doi: 10.1190/1.3378309.
Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2017, ImageNet classification with deep convolutional neural networks: Communications of the ACM, **60**, 84–90, doi: 10.1145/3065386.
Lapedes, A., and R. Farber, 1987, How neural nets work: Conference on Neural Information Processing Systems, 442–456.
LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436–444, doi: 10.1038/nature14539.
LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998, Gradient-based learning applied to document recognition: Proceedings of the IEEE, **86**, 2278–2324, doi: 10.1109/5.726791.
Liu, S., C. Birnie, and T. Alkhalifah, 2022a, Coherent noise suppression via a self-supervised deep learning scheme: 84th Annual International Conference and Exhibition, EAGE, Extended Abstracts, 1–5, doi: 10.3997/2214-4609.202210382.
Liu, S., C. Birnie, T. Alkhalifah, and A. Bakulin, 2022b, Aiding self-supervised coherent noise suppression by the introduction of signal segmentation using blind-spot networks: Second International Meeting for Applied

Downloaded 01/12/24 to 138.67.12.184. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/page/policies/terms
DOI:10.1190/geo2021-0521.1

Geoscience & Energy, SEG, Expanded Abstracts, 2857–2861, doi: 10.1190/image2022-3742756.1.
Ma, Y., X. Ji, T. Fei, and Y. Luo, 2018, Automatic velocity picking with convolutional neural networks: 88th Annual International Meeting, SEG, Expanded Abstracts, 2066–2070, doi: 10.1190/segam2018-2987088.1.
Martin, G. S., K. J. Marfurt, and S. Larsen, 2002, Marmousi-2: An updated model for the investigation of AVO in structurally complex areas: 72nd Annual International Meeting, SEG, Expanded Abstracts, 1979–1982, doi: 10.1190/1.1817083.
Mayhan, J. D., and A. B. Weglein, 2013, First application of Green's theorem-derived source and receiver deghosting on deep-water Gulf of Mexico synthetic (SEAM) and field data: Geophysics, **78**, no. 2, WA77–WA89, doi: 10.1190/geo2012-0295.1.
Moldoveanu, N., 2000, Vertical source array in marine seismic exploration: 70th Annual International Meeting, SEG, Expanded Abstracts, 53–56, doi: 10.1190/1.1816117.
Moldoveanu, N., L. Combee, M. Egan, G. Hampson, L. Sydora, and W. Abriel, 2007, Over/under towed-streamer acquisition: A method to extend seismic bandwidth to both higher and lower frequencies: The Leading Edge, **26**, 41–58, doi: 10.1190/1.2431831.
O'Brien, M., 1994, 1994 Amoco statics test: SEG.
Özbek, A., M. Vassallo, K. Özdemir, D.-J. van Manen, and K. Eggenberger, 2010, Crossline wavefield reconstruction from multicomponent streamer data: Part 2 — Joint interpolation and 3D up/down separation by generalized matching pursuit: Geophysics, **75**, no. 6, WB69–WB85, doi: 10.1190/1.3497316.
Özdemir, A. K., B. A. Kjellesvig, A. Özbek, and J. E. Martin, 2012, Digital noise attenuation of particle motion data in a multicomponent 4C towed streamer: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, cp-293, doi:10.3997/2214-4609.20148301.
Özdemir, K., A. Özbek, D.-J. van Manen, and M. Vassallo, 2010, On data-independent multicomponent interpolators and the use of priors for optimal reconstruction and 3D up/down separation of pressure wavefields: Geophysics, **75**, no. 6, WB39–WB51, doi: 10.1190/1.3494621.
Posthumus, B. J., 1993, Deghosting of twin streamer configuration: Geophysical Prospecting, **41**, 267–286, doi: 10.1111/j.1365-2478.1993.tb00570.x.
Provenzano, G., T. J. Henstock, J. M. Bull, and G. Bayrakci, 2020, Attenuation of receiver ghosts in variable-depth streamer high-resolution seismic reflection data: Marine Geophysical Research, **41**, 1–15, doi: 10.1007/s11001-020-09407-9.
Reilly, J. M., 2016, Marine broadband technology: History and remaining challenges from an end-user perspective: The Leading Edge, **35**, 316–321, doi: 10.1190/tle35040316.1.
Rentsch, S., S. Bodyakin, L. P. Espezua, A. Özbek, W. G. Brouwer, and P. Watterson, 2013, Multidomain noise attenuation for multimeasurement towed streamer data: 75th Annual International Conference and Exhibition, EAGE, Extended Abstracts, cp-348, doi: 10.3997/2214-4609.20130887.
Russell, S., and P. Norvig, 2016, Artificial intelligence: A modern approach: Pearson Education Limited.
Sabbione, J. I., and D. Velis, 2010, Automatic first-breaks picking: New strategies and algorithms: Geophysics, **75**, no. 4, V67–V76, doi: 10.1190/1.3463703.
Sartori, M. A., and P. J. Antsaklis, 1991, A simple method to derive bounds on the size and to train multilayer neural networks: IEEE Transactions on Neural Networks, **2**, 467–471, doi: 10.1109/72.88168.
Sønneland, L., L. E. Berg, P. Eidsvig, A. Haugen, B. Fotland, and J. Vestby, 1986, 2-D deghosting using vertical receiver arrays: 56th Annual International Meeting, SEG, Expanded Abstracts, 516–519, doi: doi:10.1190/1.1893019.
Soubaras, R., 2010, Deghosting by joint deconvolution of a migration and a mirror migration: 80th Annual International Meeting, SEG, Expanded Abstracts, 3406–3410, doi: 10.1190/1.3513556.
Soubaras, R., and R. Dowle, 2010, Variable-depth streamer — A broadband marine solution: First Break, **28**, 89–96, doi: 10.3997/1365-2397.28.12.44692.
Stein, E. M., and R. Shakarchi, 2011, Fourier analysis: An introduction: Princeton University Press.
Sun, H., and L. Demanet, 2020, Extrapolated full-waveform inversion with deep learning: Geophysics, **85**, no. 3, R275–R288, doi: 10.1190/geo2019-0195.1.
Sun, J., S. Slang, T. Elboth, T. Larsen Greiner, S. McDonald, and L.-J. Gelius, 2020, A convolutional neural network approach to deblending seismic data: Geophysics, **85**, no. 4, WA13–WA26, doi: 10.1190/geo2019-0173.1.
Sutskever, I., J. Martens, G. Dahl, and G. Hinton, 2013, On the importance of initialization and momentum in deep learning: International Conference on Machine Learning, 1139–1147.
Sutskever, I., O. Vinyals, and Q. V. Le, 2014, Sequence to sequence learning with neural networks: Advances in Neural Information Processing Systems, 3104–3112.
Tamura, S., and M. Tateishi, 1997, Capabilities of a four-layered feedforward neural network: Four layers versus three: IEEE Transactions on Neural Networks, **8**, 251–255, doi: 10.1109/72.557662.
Tenghamn, R., S. Vaage, and C. Borresen, 2007, A dual-sensor towed marine streamer: Its viable implementation and initial results: 77th Annual International Meeting, SEG, Expanded Abstracts, 989–993, doi: doi:10.1190/1.2792571.
ten Kroode, F., S. Bergler, C. Corsten, J. W. de Maag, F. Strijbos, and H. Tijhof, 2013, Broadband seismic data — The importance of low frequencies: Geophysics, **78**, no. 2, WA3–WA14, doi: 10.1190/geo2012-0294.1.
van der Baan, M., and C. Jutten, 2000, Neural networks in geophysical applications: Geophysics, **65**, 1032–1047, doi: 10.1190/1.1444797.
Vassallo, M., A. Özbek, K. Özdemir, and K. Eggenberger, 2010, Crossline wavefield reconstruction from multicomponent streamer data: Part 1 — Multichannel interpolation by matching pursuit (MIMAP) using pressure and its crossline gradient: Geophysics, **75**, no. 6, WB53–WB67, doi: 10.1190/1.3496958.
Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC127–WCC152, doi: 10.1190/1.3238367.
Vrolijk, J.-W., and G. Blacquière, 2021, Source deghosting of coarsely sampled common-receiver data using a convolutional neural network: Geophysics, **86**, no. 3, V185–V196, doi: 10.1190/geo2020-0186.1.
Wang, M., S. Xu, and H. Zhou, 2020, Self-supervised learning for low frequency extension of seismic data: 90th Annual International Meeting, SEG, Expanded Abstracts, 1501–1505, doi: doi:10.1190/segam2020-3427086.1.
Wang, Y., A. C. Ramirez, and A. Osen, 2017, A low-frequency deghosting method: Analysis and numerical tests: Geophysics, **82**, no. 5, V285–V296, doi: 10.1190/geo2016-0116.1.
Weglein, A. B., H. Liang, J. Wu, J. D. Mayhan, L. Tang, and L. Amundsen, 2013, A new Green's theorem deghosting method that simultaneously: (1) avoids a finite difference approximation for the normal derivative of the pressure and, (2) avoids the need for replacing the normal derivative of pressure with the vertical component of particle velocity, thereby avoiding issues that can arise within each of those two assumptions/approaches: Theory and analytic and numeric examples: Journal of Seismic Exploration, **22**, 413–426.
Whitcombe, D., and L. Hodgson, 2007, Stabilizing the low frequencies: The Leading Edge, **26**, 66–72, doi: 10.1190/1.2431833.
Wu, Y., and G. A. McMechan, 2018, Feature-capturing full waveform inversion using a convolutional neural network: 88th Annual International Meeting, SEG, Expanded Abstracts, 2061–2065, doi: 10.1190/segam2018-2963265.1.
Yang, F., and J. Ma, 2019, Deep-learning inversion: A next-generation seismic velocity model building method: Geophysics, **84**, no. 4, R583–R599, doi: 10.1190/geo2018-0249.1.
Zhang, J., and A. B. Weglein, 2005, Extinction theorem deghosting method using towed streamer pressure data: Analysis of the receiver array effect on deghosting and subsequent free surface multiple removal: 75th Annual International Meeting, SEG, Expanded Abstracts, 2095–2098, doi: 10.1190/1.2148125.

Biographies and photographs of the authors are not available.