A convolutional neural network approach for ghost removal

Khalid Almuteri and Paul Sava, Center for Wave Phenomena, Colorado School of Mines

SUMMARY

In towed-streamer acquisition, ghost reflections severely reduce the quality of seismic data. Ghost reflections limit the usable bandwidth, which hampers geophysical applications that require broadband data such as full waveform inversion. Different processing and acquisition solutions have been proposed to mitigate the ghost reflections problem. Satisfactory solutions often require complex acquisition configurations that come with complicated processing workflows. Receiver-side deghosting is commonly carried on common shot gathers, whereas source-side deghosting is carried on common receiver gathers. Since shot gathers are sparsely acquired in towed-streamer acquisitions, the source-side ghost remains an unsolved problem. We propose a solution capable of removing source-side and receiver-side ghosts in the shot domain for pressure data conventionally acquired at a single reference surface. Our solution uses a supervised learning approach, namely, convolutional neural networks, to iteratively learn to map a ghost-contaminated input to a ghost-free output. This method does not require any particular acquisition configuration and does not come with an involved processing workflow. However, it requires the acquisition geometry of field data and the bathymetry of the ocean floor to be known to create training data with the approximate configuration. We demonstrate with a numerical example the ability of our solution to remove ghost reflections and show that it is robust to random noise, even when noise-free data are used for training.

INTRODUCTION

Ghost reflections pose a challenge for geophysical applications such as full waveform inversion (FWI) and impedance inversion because they attenuate low frequencies, introduce frequency notches into the amplitude spectrum, and alter the source signature phase. Frequency notches limit the usable bandwidth between the first and second notch, lowering the seismic resolution. Since time delays of ghost reflections depend on the acquisition geometry, different traces in a shot gather have different frequency content (Aytun, 1999). Such discrepancies in the frequency content can complicate FWI and amplitude-versus-offset analysis.

Three types of ghost reflections co-exist in marine data: source-side ghosts, receiver-side ghosts, and source-receiver ghosts. Various solutions have been proposed to attenuate ghost reflections. In general, receiver-side deghosting is carried in the shot gather domain, whereas source-side deghosting is carried in the common receiver domain. Source-side deghosting is more challenging because shot gathers are sparsely acquired in towed-streamer acquisitions; source-side ghosts remain an open problem (Amundsen et al., 2017; Ikelle and Amundsen, 2018).

Receiver-side deghosting solutions can be classified by their implementation into processing-based and acquisition-based. Soubaras (2010) proposes a joint deconvolution approach to improve the data in the image domain. Amundsen and Zhou (2013) derive a low-frequency deghosting filter based on a truncated approximation of the ghost model. Beasley et al. (2013a,b) suggest an iterative wave equation solution employing the causality principle to remove ghost reflections. However, since the solution requires solving the wave equation iteratively, it is sensitive to noise in the data.

Acquisition solutions either require complementary measurements for upgoing/downgoing wavefield separation or improve data only poststack. Multiple measurement systems are favored over processing solutions because they are more robust and over poststack solutions because some applications such as full wavefield inversion or migration velocity analysis become feasible. Sønneland et al. (1986), Posthumus (1993), and Moldoveanu et al. (2007) propose an over/under acquisition configuration, i.e. acquiring pressure data at two different reference surfaces enables ghost removal through a wavefield separation process. Tenghamn et al. (2007) and Carlson et al. (2007) propose measuring pressure and particle motion data simultaneously to separate upgoing and downgoing wavefields. However, particle motion sensors are sensitive to mechanical noise below 20 Hz (Tenghamn et al., 2007; Carlson et al., 2007; Day et al., 2013), which interfere with the reconstructed data. Prestack acquisition solutions complicate field operations and have an involved processing workflow. Further, such solutions are not applicable to conventional pressure data measured at a single reference surface.

In this work, we present a seismic deghosting solution using convolutional neural networks (CNNs). Our solution does not require any particular acquisition configuration, but requires knowledge of the acquisition geometry and of the ocean floor bathymetry to create training data that mimic field data. Our solution can remove source-side and receiver-side ghosts in the shot domain for conventionally acquired data. Further, our method is robust to random noise in the data without the need to use noise-contaminated training data.

CNN-BASED DEGHOSTING

Theory

CNNs are powerful machine learning (ML) models inspired by the visual cortex in the brain (Géron, 2019). CNNs use learnable convolutional filters and non-linear activation functions to map inputs to outputs. A supervised learning approach is commonly used in the framework of artificial neural networks (LeCun et al., 2015), to find the optimal coefficients that minimize an objective function given a set of training examples. Such examples come in input-output pairs: $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, \ldots, N$, where the relationship between inputs $(\mathbf{x}_i)$ and outputs $(\mathbf{y}_i)$ is unknown and represented by a function f such

that $\mathbf{y}_i = f(\mathbf{x}_i)$. Training aims to find a function $h_{\mathbf{w}}$, represented by a CNN and parameterized by the coefficients of the convolutional filters, that approximates the function f. In our work, $h_{\mathbf{w}}$ represents the CNN-based deghosting operator, $\mathbf{x}_i$ is a ghost-contaminated input gather, and $\mathbf{y}_i$ is a true ghost-free output gather. We arrive at the deghosting operator ($h_{\mathbf{w}}$) by iteratively minimizing the objective function

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

with respect to the coefficients of the CNN ($\mathbf{w}$).

We use the Nesterov's Accelerated Gradient (NAG) method (Sutskever et al., 2013; Géron, 2019) to update the coefficients of the neural network according to

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mathbf{m}_{k+1}, \tag{2}$$

where

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

is the momentum ($\mathbf{m}_0 = 0$), $\beta \in [0,1]$ is the momentum weight, and α is the learning rate. To train our neural network we use $\beta = 0.9$ and $\alpha = 0.1$. The NAG method speeds up convergence, compared to gradient descent, by computing the gradient of the objective function at the current location of weights ($\mathbf{w}_k$) with a slight slide in the momentum direction ($\mathbf{m}_k$).

CNN architecture

Our choice of neural network is comprised of a three-times repeated sequence of three convolutional layers followed by a masking layer for a total of 12 layers (Figure 1). We use 16 filters in each convolutional layer. The convolutional filters in the first two layers in the sequence have a spatial width of three traces and a temporal length of 101 coefficients. The convolutional filters in the third layer have the same spatial width as the preceding two layers, but only 11 temporal coefficients. Throughout our CNN, we use the tanh as the activation function because it is bounded between -1 and 1 and centered around zero. The parameters that define the architecture of our CNN (e.g., number of coefficients), known as hyperparameters, are set experimentally. Since our neural network operates in the time-space domain, we use temporally long convolutional filters to reconstruct the low-frequency components of the data, and short filters to recover the high-frequency information. By using filters with different lengths, we give the CNN the freedom to deghost at different scales.

Deghosting can be carried using 3D filters, i.e. utilize information from adjacent traces from the inline and crossline directions, or using 1D filters that run on individual traces. Since ghost reflections have offset dependency, 3D filters outperform 1D filters. Such dependency can be observed using the 3D ghost model in the Fourier domain (Amundsen, 1993) given by

$$G(k_z, z) = 1 + re^{2ik_z z}, \tag{4}$$

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

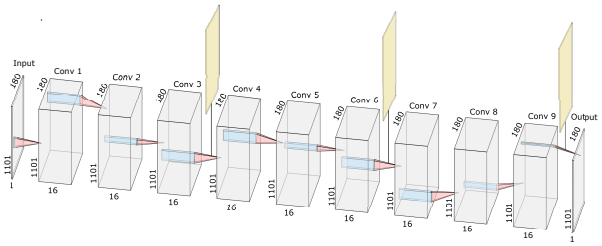

Figure 1: A schematic of the architecture of the network we use for seismic deghosting. Gray cuboid labeled with Conv represents a multi-channel feature map generated using N convolutional filters ($N = 16$). The height of each cuboid corresponds to the number of time samples ($n_t = 1101$) and the width corresponds to the number of traces in the gather ($n_r = 180$). Yellow cuboids represent temporal masking layers in the network.

The masking layers in the CNN serve two closely-related purposes. On one hand, they remove CNN-induced noise before any first-arrival. On the other hand, by removing CNN-induced noise, they help the CNN to learn deghosting; without masking layers, training fails because the neural network tries to fit the training examples by removing ghosting reflections and CNN-related noise simultaneously. A masking layer defines the regions where a seismic signal could exist in a shot gather. To create such a layer, we use the Modified Coppens's method (Sabbione and Velis, 2010) to pick along the time axis the first occurrence of any seismic event, i.e. a water bottom reflection or a refraction from a layer underneath. The masking layer mutes any signal not created by the source function above the first-arrival.

NUMERICAL EXAMPLE

CNN training

We generate 128 input-output pairs of ghost-free and ghost-contaminated shot gathers to train our CNN. We randomly extract 117 sub-models from the Marmousi model (Martin et al., 2002) and 11 sub-models from the Sigsbee model (Paffenholz et al., 2002). We add 0.35 km depth of water to each sub-model. We use sources at $z = 0.025$ km and $x = 0.303$ km, and 180 receivers at $z = 0.03$ km with a spacing of 0.005 km to mimic towed-streamer acquisition. To model ghost reflections we use a mirror source at $z = -0.025$ km and mirror receivers at $z = -0.03$ km. We model seismic data for 2.2 s with a sampling interval of 0.2 ms using source wavelet with a flat spectrum between $1 - 50$ Hz. As a preprocessing step, we normalize the data to the range of the activation function.

Seismic deghosting

We demonstrate the ability of our solution to remove ghost reflections in the shot domain on 2D data using a sub-model (Figure 2) we extract from the Amoco statics test model (O'Brien, 1994). We use the same acquisition parameters to generate

training and testing data. Our testing sub-model intrinsically differ from our training sub-models; the velocity values of the training models range between 1.03 km/s and 4.7 km/s.

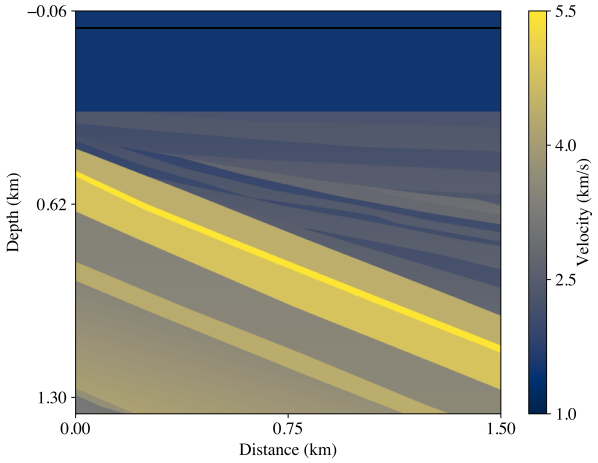

Figure 2: Sub-model from the Amoco statics test model we use to test our solution. We add a water layer with 0.35 km depth to the top. The black line corresponds to the air-water interface.

Figures 3a and 3b show the true ghost-free and ghost-contaminated data with added noise, respectively. Since noise is inevitably present in field data, we add Gaussian noise to the ghost-contaminated data, such that the signal-to-noise ratio (SNR) is 4. The waveform of the source signature in the ghost-contaminated gather has a reverberatory character that lowers seismic resolution, whereas the waveform of the ghost-free data is more compact. Ghost reflections attenuate frequencies at 0 Hz and 34 Hz, and amplify frequencies at 16 Hz, as indicated by the average amplitude spectra of two gathers (Figure 4). The attenuation at low frequencies is more significant than at high frequencies; the amount of energy loss at 34 Hz is 6 dB and 21 dB at 1 Hz.

Applying our CNN-based deghosting operator to the noisy ghost-contaminated gather gives the ghost-free prediction shown in Figure 3c. Even though our training data are noise-free, the CNN is able to predict ghost-free data accurately. The average amplitude spectra of the true and predicted ghost-free data match closely, demonstrating the robustness of our proposed solution to deghost noise-contaminated data even when training is done with noise-free data. More importantly, the CNN recovers the low frequencies needed for FWI and MVA with a great degree of accuracy.

Figure 5a shows individual traces from the gathers shown in Figure 3 for a receiver 0.1 km away from the source. The true ghost-free and ghost-contaminated traces have a low correlation coefficient of 0.39, whereas the true and predicted ghost-free traces correlate with a coefficient of 0.87. Since stronger amplitudes contribute the most to the cost function, they are better fit by the neural network during training. Inspecting the amplitude spectra (Figure 5b) and phase spectra (Figure 5c) indicate that the CNN reconstructs the phase of the true ghost-free data better than its amplitude. Still, the amplitude of the

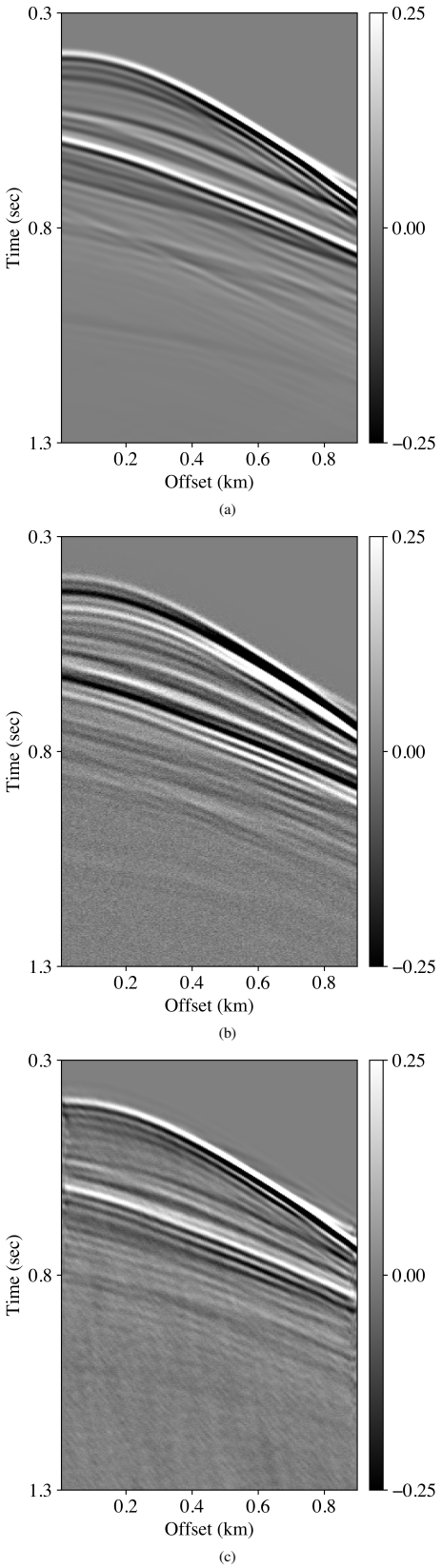

Figure 3: Comparison between (a) the true ghost-free data, (b) the ghost-contaminated data with random noise (SNR=4), and (c) the ghost-free prediction.

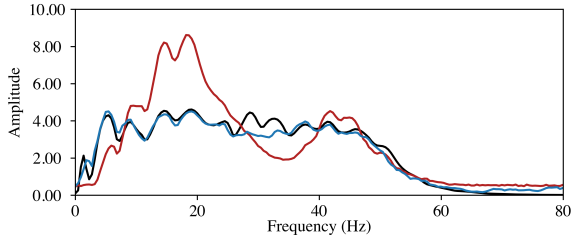

Figure 4: Comparison between the average amplitude spectra of the ghost-free data (black line), the ghost-contaminated data with random noise (red line), and the ghost-free prediction data (blue line).

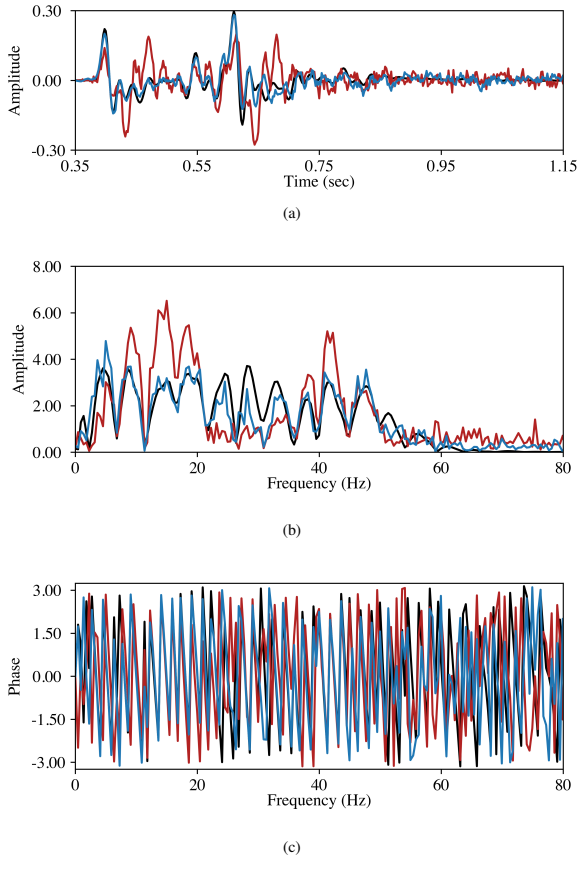

Figure 5: (a) Comparison between ghost-free (black line), ghost-contaminated with noise (red line), and ghost-free prediction (blue line) traces recorded at a receiver that is 0.1 km away from the source. (b) Comparison between the amplitude spectra of the shown traces. (c) Comparison between the phase spectra of the shown traces.

true and predicted ghost-free data exhibit the same characteristics.

DISCUSSION AND CONCLUSIONS

We demonstrate the ability to remove source-side and receiver-side ghosts for noise-contaminated pressure data acquired at a single reference surface in the shot domain. The solution does not require any specific acquisition configuration, but requires the acquisition geometry and the bathymetry of the ocean floor to be known to create training data that closely match field data.

We assume that data are acquired under a static and flat sea surface. Such an assumption is not always representative of the sea surface condition commonly associated with towed-streamer acquisitions. A dynamic sea surface alters the phase and amplitude of seismic data (Cecconello et al., 2018; Blacquière and Sertlek, 2019; Konuk and Shragge, 2020), making a time-varying sea surface important to consider in future work. We also assume the source function is known, i.e. we use the exact source function to generate training and testing data. Even though ghost reflections are independent of the source function, understanding the effects of wavelet uncertainties on deghosting is critical and remains open for future investigation. Such an understanding is important because our solution is based on data matching and because different source functions have different frequency content. We also leave investigating the consequences of different physics assumptions between training and testing data for future work. It is important to know whether training a CNN-based deghosting operator on acoustic data is adequate to accurately predict ghost-free viscoelastic data, given the Earth's intrinsic viscoelastic nature.

The numerical example demonstrates the ability of our solution to remove ghost reflections from legacy data acquired using conventional acquisition systems. Our solution reduces acquisition costs, simplifies the process of acquiring towed-streamer data, simplifies the seismic data processing workflow, and is robust to strong variations between training and testing velocity models. Exploiting such robustness reduces the computational complexity of generating training data. We show that our solution is able to remove ghost reflections from noise-contaminated data despite training the CNN with noise-free data. The presence of noise in the input data adds uncertainties to the accuracy of the ghost removal process, but our solution demonstrates the ability to remove ghost reflections and mitigate noise-related uncertainties.

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 Tugrul Konuk (CWP) for many useful discussions and for providing the forward-modeling code that made this research possible.

REFERENCES

Amundsen, L., 1993, Wavenumber-based filtering of marine point-source data: Geophysics, **58**, 1335–1348.

Amundsen, L., Ø. Pedersen, A. Osen, J. O. Robertsson, and M. Landrø, 2017, Broadband seismic over/under sources and their designature-deghosting: Geophysics, **82**, P61–P73.

Amundsen, L., and H. Zhou, 2013, Low-frequency seismic deghosting: Geophysics, **78**, WA15–WA20.

Aytun, K., 1999, The footsteps of the receiver ghost in the f-k domain: Geophysics, **64**, 1618–1626.

Beasley, C. J., R. Coates, Y. Ji, and J. Perdomo, 2013a, Wave equation receiver deghosting: A provocative example, *in* SEG Technical Program Expanded Abstracts 2013: Society of Exploration Geophysicists, 4226–4230.

Beasley, C. J., R. Coates, and C. Lapilli, 2013b, Wave equation receiver deghosting: 2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, 280–283.

Blacquière, G., and H. Ö. Sertlek, 2019, Modeling and assessing the effects of the sea surface, from being flat to being rough and dynamic: Geophysics, **84**, T13–T27.

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.

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**, T49–T68.

Day, A., T. Klüver, W. Söllner, H. Tabti, and D. Carlson, 2013, Wavefield-separation methods for dual-sensor towed-streamer data: Geophysics, **78**, WA55–WA70.

Géron, A., 2019, Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems: O'Reilly Media.

Ikelle, L. T., and L. Amundsen, 2018, Introduction to petroleum seismology: Society of Exploration Geophysicists.

Konuk, T., and J. Shragge, 2020, Modeling full-wavefield time-varying sea-surface effects on seismic data: A mimetic finite-difference approach: Geophysics, **85**, T45–T55.

LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436–444.

Martin, G. S., K. J. Marfurt, and S. Larsen, 2002, Marmousi-2: An updated model for the investigation of avo in structurally complex areas, *in* SEG Technical Program Expanded Abstracts 2002: Society of Exploration Geophysicists, 1979–1982.

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.

O'Brien, M., 1994, Amoco statics test: Amoco.

Paffenholz, J., J. Stefani, B. McLain, and K. Bishop, 2002, Sigsbee_2a synthetic subsalt dataset - image quality as function of migration algorithm and velocity model error: Presented at the 64th EAGE Conference & Exhibition, European Association of Geoscientists & Engineers.

Posthumus, B. J., 1993, Deghosting of twin streamer configuration: Geophysical Prospecting, **41**, 267–286.

Sabbione, J. I., and D. Velis, 2010, Automatic first-breaks picking: New strategies and algorithms: Geophysics, **75**, V67–V76.

Sønneland, L., L. E. Berg, P. Eidsvig, A. Haugen, B. Fotland, and J. Vestby, 1986, 2-d deghosting using vertical receiver arrays, *in* SEG Technical Program Expanded Abstracts 1986: Society of Exploration Geophysicists, 516–519.

Soubaras, R., 2010, Deghosting by joint deconvolution of a migration and a mirror migration, *in* SEG Technical Program Expanded Abstracts 2010: Society of Exploration Geophysicists, 3406–3410.

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, PMLR, 1139–1147.

Tenghamn, R., S. Vaage, and C. Borresen, 2007, A dual-sensor towed marine streamer: Its viable implementation and initial results, *in* SEG Technical Program Expanded Abstracts 2007: Society of Exploration Geophysicists, 989–993.