EAGE ANNUAL 2022
83RD CONFERENCE & EXHIBITION
MADRID I SPAIN

TOWARDS USING NEURAL NETWORKS TO COMPLEMENT CONVENTIONAL SEISMIC PROCESSING ALGORITHMS

T. Lesieur[1], G. Pignot[1], N. Al Kiyumi[2], J. Messud[1]

[1] CGG; [2] PDO

Summary

Convolutional-based neural network (CNN-based) architectures have shown promise in performing denoising tasks. However, it can be demonstrated that their predictions are of limited use for some tasks because they produce signal leakage. For these tasks, a possible improvement is to incorporate CNN-based architectures as one component of, rather than replacement for, the conventional denoising algorithms. In this paper, we formally define a class of denoising problems usually solved iteratively for which using CNN-based predictions as an initial solution can improve efficiency. We illustrate our points using a land data deblending example, for which the CNN-based prediction quality was higher than that of the conventional first iteration but lower than that of the final product. The CNN-complemented conventional deblending leads to satisfactory and efficient results.

Towards using neural networks to complement conventional seismic processing algorithms

Introduction

Deep Learning (DL) for seismic processing represents an active field of research. For many seismic processing applications, physics-based state-of-the-art algorithms are available, with rigorous criteria for success. Most of the current DL investigations consist of learning to mimic the results of these algorithms, i.e., to efficiently predict the processed outcomes from the data using neural networks (NN), especially convolutional-based (CNN-based) architectures (Richardson et al., 2019; Mandelli et al., 2019). These have the potential to add value, for instance, by learning the best of various existing workflows or decreasing turnaround time. However, using the neural network predictions in place of the conventionally processed data remains a challenge because, among other reasons, the predictions often produce signal leakage, a loss of signal amplitude, violating one of the process success criterion (Hou and Messud, 2021). In some cases, the leakage might be due to an imperfect tuning of DL parameters and thus eventually overcome. In other cases, however, the leakage could be due to a fundamental limitation of DL. Understanding the distinction between these cases is crucial in the quest for an efficient use of DL in seismic processing.

Peng et al. (2021) took a step in this direction. They propose a theoretical foundation for the use of CNN-based architectures with ReLU internal activations to predict the outcomes of a specific class of various seismic processing tasks. Their work can also be used to define classes of seismic processing tasks for which CNN-based architectures will encounter difficulties predicting outcomes rivaling those of conventional algorithms, i.e., they will always result in signal leakage. In these cases, a better possible approach is to use CNN-based architectures as a building block (or component) of the conventional algorithms, rather than trying to replace their outcomes. Here, we explore this approach, which has already been adopted in other application fields, e.g., Borgerding et al. (2017) and Pandit et al. (2020). We start by formally defining a general class of denoising problems (including deghosting and deblending among others), usually solved iteratively. Adapting the work of Peng et al. (2021), we then identify problems for which the use of CNN-based architectures (with ReLU internal activations) could be adapted to directly predict denoised outcomes. For some other problems, we propose to use the CNN-based predictions inside the denoising algorithm as an initial solution for the iterations. We illustrate our points on a land field data deblending example.

Denoising problems and CNN-based architectures

We consider processing algorithms that can be formulated in the following optimization problem:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{W}\,\boldsymbol{L}\,\boldsymbol{\alpha} - \boldsymbol{d}_{obs}\|_2^2 + \|\mathbf{b} \odot \boldsymbol{\alpha}\|_1 \,. \tag{1}$$

$\boldsymbol{d}_{obs}$ is a vector representing the observed (noisy) data and $\boldsymbol{W}$ is a matrix representing the noise modelling operator (e.g., blending); $\boldsymbol{W}$ is parameterized by prior physical information and, in many applications, leads to a strongly under-determined problem where constraints must be added. A sparsity constraint is often considered, acting on a vector $\boldsymbol{\alpha}$ which represents the denoised data vector $\boldsymbol{d}$ through

$$\boldsymbol{d} = \boldsymbol{L}\,\boldsymbol{\alpha}. \tag{2}$$

$\boldsymbol{L}$ is the matrix that represents the transformation from a certain domain, into which $\boldsymbol{\alpha}$ lives, to the data domain (e.g., $\boldsymbol{L}$ might represent the inverse of a curvelet transform). The l_1 sparsity constraint in the second term of eq. (1) enforces the sparsity of $\boldsymbol{\alpha}$, with $\mathbf{b}$ as the associated weight vector that allows to tune the sparsity level ($\odot$ denotes element-wise multiplication). The problem in eq. (1) is usually solved using an iterative gradient-descent-based method (as the l_1 constraint brings non-linearity). The iterations are often initialized with a noncommittal null vector $\boldsymbol{\alpha}^{(initial)} = 0$, in the absence of anything better. Once the final iteration vector $\boldsymbol{\alpha}^{(final)}$ is obtained, the denoised data outcome $\boldsymbol{d}$ is deduced by applying eq. (2).

We consider applications where $\boldsymbol{W}$ has a complex structure and is large. Then, many computationally costly iterations are usually required to converge. To mitigate this, some preconditioner matrix $\boldsymbol{P}$ can be added in the l_2-norm part of eq. (1), multiplying its argument on the left by $\boldsymbol{P}$. Detailing this procedure goes beyond the scope of this paper. Another possibility is to start iterating from a better $\boldsymbol{\alpha}^{(initial)}$, which interests us here. In particular, we would need an $\boldsymbol{\alpha}^{(initial)}$ of much better quality than

$(\boldsymbol{W}\,\boldsymbol{L})^{\boldsymbol{t}}\,\boldsymbol{d}_{obs}$ (which is close to the first gradient-descent iteration result starting from $\boldsymbol{\alpha}^{(initial)} = 0$; "$t$" denotes the transpose).

The use of CNN-based architectures to predict the outcomes of denoising problems of the form considered here has already been explored (Richardson et al., 2019; Mandelli et al., 2019). The goal is to train a CNN-based $\boldsymbol{F}_{\theta}$ (θ representing the convolutional kernels and bias parameters) with the hope that the corresponding predictions,

$$\boldsymbol{d}_{pr} = \boldsymbol{F}_{\theta}(\boldsymbol{d}_{obs}), \tag{3}$$

represent a good approximation of the conventional algorithm's outcomes $\boldsymbol{d}$, on a given denoising task. One prerequisite, to have a chance to succeed, is to be able to relate CNN-based architectures to a reparameterization of the problems in eqs. (1) and (2). Peng et al. (2021) describe situations where this can occur. To transfer their conclusions to the problem considered here, we use the variable change in eq. (2) to reformulate eq. (1) as (superscript "$-p$" denotes the pseudo-inverse)

$$\min_{\boldsymbol{d}} \|(\boldsymbol{W}\,\boldsymbol{L})\,\boldsymbol{L}^{-\boldsymbol{p}}\,\boldsymbol{d} - \boldsymbol{d}_{obs}\|_2^2 + \|\mathbf{b} \odot \boldsymbol{L}^{-\boldsymbol{p}}\,\boldsymbol{d}\|_1\,. \tag{4}$$

Applying the reasoning of Peng et al. (2021) to eq. (4), we deduce that CNN-based architectures with ReLU internal activations can represent a valid reparameterization of our problems if the noise modelling operator $\boldsymbol{W} \approx (\boldsymbol{W}\,\boldsymbol{L})\,\boldsymbol{L}^{-\boldsymbol{p}}$ is convolutional (i.e., has a Toeplitz matrix form) and can be split into a product of smaller convolution kernels related to sparse domains (see eq. (3) in Peng et al. (2021)). This hypothesis is often:

- True for denoising problems like deghosting (Peng et al., 2021). For these problems, it is pertinent to work on tuning the DL parameters to eliminate signal leakage and allow using $\boldsymbol{d}_{pr}$ directly as an approximation of $\boldsymbol{d}$. This could help reduce the cost of the conventional algorithm.
- False for problems like deblending, where $\boldsymbol{W}$ is far from having a Toeplitz matrix form (Guillouet et al., 2016). For these problems, CNN-based predictions $\boldsymbol{d}_{pr}$ fundamentally cannot provide an approximation of $\boldsymbol{d}$ sufficient for a direct use. In practice, even if $\boldsymbol{d}_{pr}$ contained many features of $\boldsymbol{d}$, it would tend to exhibit non-negligible signal leakage, whatever the DL parameter used. However, $\boldsymbol{d}_{pr}$ might lead to a better initializer of the conventional processing by taking

$$\boldsymbol{\alpha}^{(initial)} = \boldsymbol{L}^{-\boldsymbol{p}}\,\boldsymbol{d}_{pr} = \boldsymbol{L}^{-\boldsymbol{p}}\,\boldsymbol{F}_{\theta}(\boldsymbol{d}_{obs}). \tag{5}$$

This could help to significantly accelerate the convergence of the conventional algorithm; this observation represents the first main proposal of this article, with the second main point being the understanding of the class of denoising problems where this could apply. In such an application, DL complements the conventional algorithm, which can be considered as a post-processing of the DL prediction to eliminate the signal leakage. We next provide a field data illustration, in the context of a land acquisition dataset deblending problem.

Land deblending example

In the following land deblending application modelized using eq. (4), $\boldsymbol{L}$ represents the transform from the curvelet domain to the data domain and $\boldsymbol{W}$ represents a multi-convolutional matrix (that is Toeplitz only per blocks, each block containing the source signature associated with each shot) multiplied by a "restriction" matrix (Guillouet et al., 2016). Thus, $\boldsymbol{W}$ is far from having a Toeplitz matrix structure and has many more columns than rows.

Fig. 1a shows a blended data $\boldsymbol{d}_{obs}$ example from a 3D land, wide-azimuth, broadband acquisition in the north of the Sultanate of Oman (Al Kiyumi et al., 2021), obtained using simultaneous shooting (Shorter et al., 2017). The deblending is performed in the common receiver domain. The conventional algorithm requires many costly iterations to converge towards the deblended data $\boldsymbol{d}$ in Fig. 1b (top). As the deblending occurs early in the processing sequence, we must be very careful that the algorithm preserves all the signal. To ensure this, a blending noise model is inverted for (bottom of Fig. 1) and then subtracted from the blended data to get the deblended data (top of Figs. 1b, c, d, e); To avoid signal leakage, the reflectivity model used to derive the noise model needs to be QCed for residual blending noise.

We now consider using DL to learn to predict the deblending outcome. We use a CNN-based architecture with ReLU internal activations, adapted from Unet and representing a good compromise between quality and cost (Messud and Chambefort, 2020). The training data consists of a few blended data, $\boldsymbol{d}_{obs}$, and deblended data produced by the conventional deblending algorithm, $\boldsymbol{d}$. Less than 0.1% of the total data was selected for the training (unsupervised machine learning was used to classify the blended data into clusters, which were used to define a pertinent data subset). The effectiveness of using CNN-based predictions to deblend the rest of the data was then evaluated. Fig. 1c illustrates that it does not subtract enough blending noise (circled areas). Because of the complex structure of $\boldsymbol{W}$, the considerations above allow us to understand that CNN-based architectures may not be able to predict the deblending outcomes well enough for direct use. They also help us to infer that the use of other architectures like DUnet (Peng et al., 2021) should not dramatically improve the results.

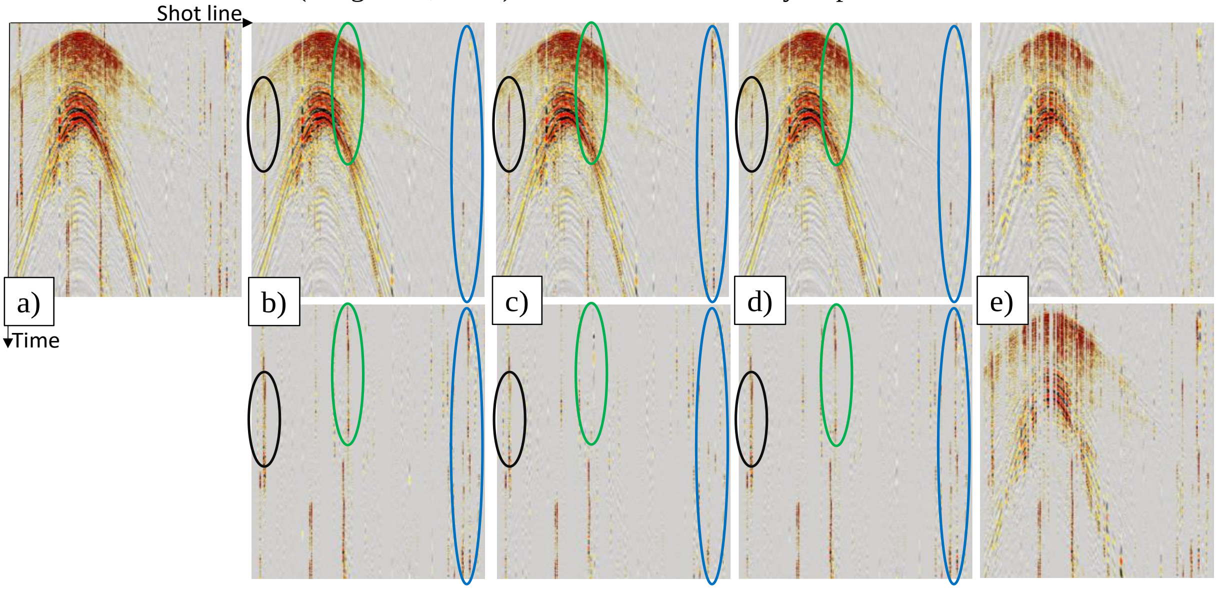

Figure 1: *Common receiver domain data. Top: a) Blended data $\mathbf{d_{obs}}$, b) conventionally deblended data $\boldsymbol{d}$, c) CNN-based prediction $\mathbf{d_{pr}}$, d) deblended data using (c) as an initializer of the conventional deblending (number of needed iterations reduced by 80% compared to (b)), e) Simply applying the transpose, i.e., $\boldsymbol{L}\,(\boldsymbol{W}\,\boldsymbol{L})^{\boldsymbol{t}}\,\boldsymbol{d_{obs}}$ (close to one conventional algorithm iteration). Bottom: Modelled blending noise subtracted from the data. The ovals highlight the amplitude leakage in the CNN-based blending noise predictions ((c), bottom).*

A better approach is to use CNN-based predictions within the conventional deblending algorithm, e.g., as an improved initial solution, following the proposal mentioned above. Fig. 1d shows that this produces a deblending result of similar quality to the conventional deblending, Fig. 1b, with a reduction of 80% in the number of iterations (Al Kiyumi et al., 2021). As each of these iterations are costly, the efficiency of the deblending is much increased (even accounting for DL training). Fig. 1e shows the result of applying $\boldsymbol{L}\,(\boldsymbol{W}\,\boldsymbol{L})^{\boldsymbol{t}}\,\boldsymbol{d_{obs}}$; we illustrate here that any remaining blending noise in the primary model will results in signal leakage once subtracted from the original data.

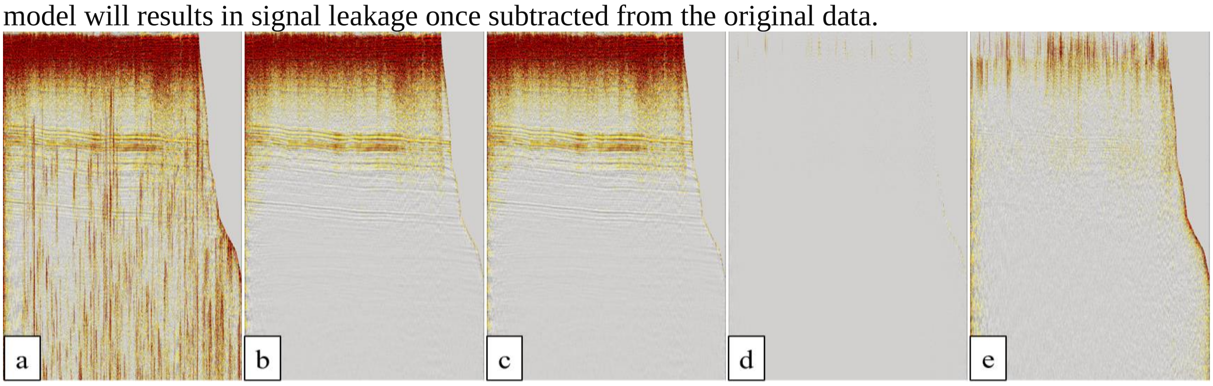

Figure 2: *NMO stack domain data. a) Blended data, b) conventionally deblended data, c) deblended data using the CNN-based prediction as an initializer of the conventional deblending, d) difference between (b) and (c), e) difference between (b) and (c) with a gain of 20dB.*

To further QC the results, Fig. 2 shows normal moveout (NMO) stack domain comparisons between the methods discussed here. This additional control confirms that the signal is as well preserved when using our proposed DL-complemented deblending as with the conventional method.

Conclusions and challenges

CNN-based architectures, with their versatility, appear to be a promising way to tackle denoising tasks. However, using their predictions directly may be insufficient for a class of various denoising tasks solved iteratively, including deblending, resulting in signal leakage regardless of the DL parameters. For these tasks, a possible improvement will be to use CNN-based architectures as a component of the conventional denoising algorithms instead of trying to replace their outcomes, for instance as an initial solution for the iterations, to bring better efficiency. We illustrated our points using a land data deblending example, on which an approach based solely on neural networks gives a result of insufficient quality compared to the conventional algorithm, while our DL-complemented approach provides satisfactory and efficient results.

We proposed one possible way to integrate CNN-based architectures into conventional denoising algorithms. Finding other effective, and possibly better, ways still represents an open question. For instance, we could consider using neural network predictions through an additional prior within the conventional algorithm, as already done in other fields (Borgerding et al., 2017). Indeed, the sparse inversion methods considered here can handle large amplitude variations in the data, which is crucial for deblending. However, it is not easy to reconstruct very low amplitude signal with conventional techniques (this requires a very fine tuning of the sparsity constraint, related to the **b** in eq. (1)), and modelling these low amplitudes is crucial to improve current deblending. Conversely, neural networks can succeed in preserving low amplitudes, but experience difficulties in handling large amplitude variations. A successful way to improve deblending could be to merge the benefits of both approaches, namely having a neural network prior powerful enough to preserve the low amplitudes, while tackling the problem of reconstructing events with widely different amplitudes using conventional techniques.

Acknowledgements

We are grateful to the Ministry of Energy and Minerals of the Sultanate of Oman, Petroleum Development Oman (PDO) and CGG for permission to present the data examples. We are also grateful to T. Ben Youssef, D. Carotti and G. Lambaré for their involvement.

References

Al Kiyumi, N., Pignot, G., Lesieur, T., Ben Youssef, T., Siyabi, O. and Al Harthi, A. [2021] Increased efficiency of deblending using deep learning: Case study from Sultanate of Oman. Presented at *SEG workshop: 2nd edition artificially intelligent earth exploration: Between expectations and the current situation.*

Borgerding, M., Schniter, P. and Rangan, S. [2017] AMP-Inspired Deep Networks for Sparse Linear Inverse Problems. *IEEE Transactions on Signal Processing*, **65**(16), 4293-4308.

Guillouet, M., Berthaud, A., Bianchi, T., Pignot, G., Mahrooqi, S. and Shorter, J. [2016] Recovery of Blended Data - A Sparse Coding Approach for Seismic Acquisition. *78th EAGE Conference & Exhibition.*

Hou, S. and Messud, J. [2021] Machine learning for seismic processing: The path to fulfilling promises. *1st International Meeting for Applied Geoscience & Energy*, Expanded Abstracts, 3204-3208.

Mandelli, S., Lipari, V., Bestagini, P. and Tubaro, S. [2019]. Interpolation and Denoising of Seismic Data using Convolutional Neural Networks. *arXiv: 1901.07927.*

Messud, J. and Chambefort, M. [2020] Understanding how a deep neural network architecture choice can be related to a seismic processing task. *1st EAGE Digitalization Conference*, Extended Abstracts, WeDIGI09.

Pandit, P., Sahraee-Ardakan, M., Rangan, S., Schniter, P. and Fletcher, A. K. [2020]. Inference with deep generative priors in high dimensions. *IEEE Journal on Selected Areas in Information Theory*, **1**(1) 336-347.

Peng, H., Messud, J., Salaun, N., Hammoud, I., Jeunesse, P., Lesieur, T. and Lacombe, C. [2021] Proposal of the DUnet neural network architecture: Deghosting example and theoretical analysis. *82nd EAGE Conference & Exhibition*, Extended Abstracts.

Richardson, A. and Feller, C. [2019]. Seismic data denoising and deblending using deep learning. *arXiv:1907.01497*.

Shorter, J., Mahrooqi, S., El-Taha, Y. C., and Clow, A. F. [2017] Ultra High Productivity: A minimally constrained blended acquisition methodology. Presented at *SEG workshop: Seismic with Simultaneous Sources: Where Does the Industry Stand?*