A progressive deep transfer learning approach to cycle-skipping mitigation in FWI

Wenyi Hu, Advanced Geophysical Technology, Yuchen Jin, Xuqing Wu, and Jiefu Chen, University of Houston*

Summary

To effectively mitigate the cycle-skipping phenomenon in full waveform inversion (FWI), we developed an inception-based deep learning neural network to reconstruct the absent low frequency data by exploiting the subsurface low wavenumber information buried in the acquired high frequency data. Two unique features of our deep learning network for the low frequency data reconstruction are 1) Dual Data Feed; 2) Progressive Transfer Learning. With the Dual Data Feed feature, not only the high frequency data, but also the corresponding Beat Tone data are fed into the neural network, significantly reducing the network complexity and the network training cost. The Progressive Transfer Learning technique enables us to initiate the deep learning network with only one training dataset. Unlike other deep learning approaches in this category, our training dataset is not fixed. Instead, the training dataset and the network accuracy are progressively enhanced in an iterative manner within the Progressive Transfer Learning workflow. The numerical experiments validated that the low frequency data reconstructed by this progressive transfer learning network are sufficiently accurate for a successful FWI inversion without suffering from the cycle-skipping without any *a priori* geological information.

Introduction

Full waveform inversion (FWI) is an advanced seismic processing technology for high resolution subsurface velocity models building through a data-fitting procedure based on nonlinear optimization algorithms. Because reliable low frequency components below 5 Hz do not practically exist in most acquired seismic datasets, FWI often suffers from the cycle-skipping phenomenon, inducing artifacts and producing wrong velocity models. In recent years, many research efforts have been devoted to overcome this challenge, i.e., suppressing the cycle-skipping phenomenon without acquiring low frequency data. Many of these research works can be classified into the following categories (Hu et al., 2018): 1) scattering angle based filtering methods (Alkhalifah, 2015). 2) FWI with extended velocity model space (Symes, 2008; Fu and Symes, 2015, 2017). 3) FWI with time shift minimization (Xu et al., 2012; Ma and Hale, 2013); 4) FWI with synthesized low frequency data (Shin and Cha 2008; Wu et al., 2014; Li and Demanet, 2016). 5) FWI resolving phase ambiguity (Hu, 2014; Choi and Alkhalifah, 2015). All of these research efforts helped mitigating the cycle-skipping issue to some extent. Recently, embracing the power of machine learning and data analytics, some researchers resort to the deep neural network approaches to predict the absent low frequency data by learning the underlying nonlinear relationship between the high frequency components and the low frequency components (Jin et al., 2018; Sun and Demanet, 2018; Ovcharenko et al., 2018). While the early stage testing results of these pure data driven methods are encouraging, some major issues need to be addressed before hitting a bottleneck in production. Two main concerns are accuracy and efficiency: 1) How to design the training datasets to yield unbiased training processes? 2) How to reduce the number of training datasets to make the algorithms computationally manageable?

In this work, we propose a novel Progressive Transfer Learning neural network to overcome these two main challenges. The Dual Data Feed structure featured in our deep learning network has two branches receiving the high frequency data and the corresponding Beat Tone data simultaneously, thus partially relieving the network from the heavy burden of feature extraction. Another unique feature of our method is the Progressive Transfer Learning technique, which integrates the machine learning module and the physics-based inversion module seamlessly to accelerate the training process and enhance the data reconstruction accuracy. With this novel technique, only one training dataset is required for the training process. Unlike other deep learning network training process where the training datasets are fixed, our method progressively updates the training velocity model iteratively to inject a continuously improved training dataset to the deep learning network. Consequently, the network training cost is substantially reduced and the data prediction accuracy is significantly improved, while the quality of the training process can be quantitatively monitored, showing great potential in large scale projects.

Feasibility Study of Low Frequency Reconstruction

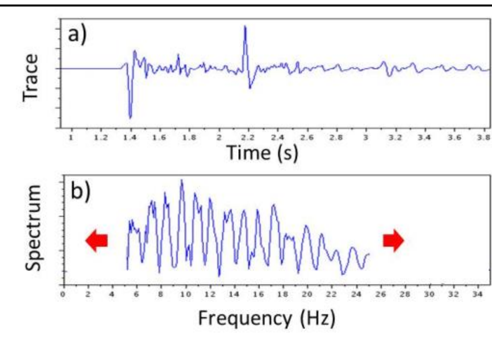

Figure 1: a) a seismic trace; b) spectrum of the trace with frequency components below 5 Hz and above 25 Hz abandoned.

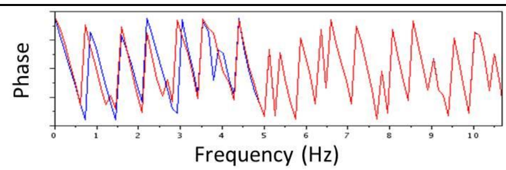

Figure 2: Comparison between phase spectra of original data (blue) and the reconstructed low frequency data (red).

There is an important question needs to be answered first: Is it possible to reconstruct the absent low frequency seismic data from the high frequency seismic data? For general random signals, reconstruction of the low frequency data from the high frequency data is impossible because there is no meaningful relationship between different frequency bands. However, an important fact needs to be pay attention to is that seismic data are not random signals. Seismic data are the earth response to a wideband excitation. Due to this fact, there is an implicit tunnel (nonlinear relationship) connecting the low frequency and the high frequency components and this relationship is dictated by the subsurface geophysical properties. Here, we aim to validate this statement empirically using a field data example. A seismic trace is plotted in Figure 1a and its truncated frequency spectrum (components below 5 Hz and above 25 Hz are removed) is shown in Figure 1b. We then employ a sparse inversion algorithm (with L_1-norm regularization) to reconstruct the data below 5 Hz and above 25 Hz. As shown in Figure 2, the strong agreement between the reconstructed low frequency data and the original low frequency data verifies the feasibility of the low frequency reconstruction provided that the sparsity assumption is justified.

Deep Learning Network with Dual Data Feed

A necessary condition for the success of the low frequency reconstruction in Figure 2 is the sparseness assumption of the subsurface reflection events, which is valid for this particular numerical example as observed in Figure 1a. In most FWI projects, the sparsity condition of the subsurface structure does not hold, and then the trace-by-trace based inversion method does not work because the inverse problem is highly underdetermined and often gives meaningless solutions. A valid strategy is to include the large quantity of traces in the surrounding areas to construct an extremely large sized nonlinear inverse problem that is often beyond our capability to solve. Moreover, *a priori* subsurface geological information needs to be incorporated to constrain the inversion process, which is usually not possible. In this context, we resort to the pure data-driven approach, the deep learning neural network.

In order to relieve the burden of feature extraction on the deep learning network, we further investigated the physical relationship between the low frequency data and the high frequency data to guide our network design. Figure 3 is a diagram qualitatively sketching this relationship, where we observe that the high frequency components are dominantly contributed from the subsurface high wavenumber structures while being vaguely connected to the low wavenumber structures via far offsets. To amplify and stabilize the connection between the low wavenumber and the high frequency, we introduce the Beat Tone data (Hu, 2014) into the network because Beat Tone data amplify the high wavenumber information buried in the high frequency data. Beat Tone data can be derived from high frequency straightforwardly, i.e., $\boldsymbol{\Phi_{BT}(S_2, S_1) = \Phi(S_2) - \Phi(S_1)}$, where $\boldsymbol{\Phi_{BT}}$ is the Beat Tone phase data and $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$ represent two single frequency domain datasets extracted at

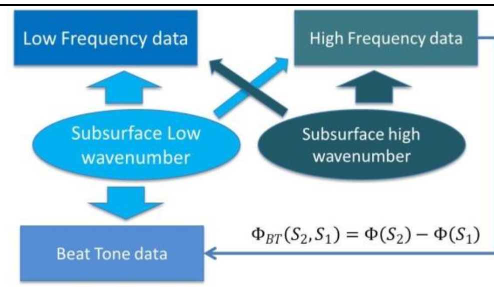

Figure 3: Diagram of relationship between low frequency,high frequency, and Beat Tone data through subsurface wavenumber.

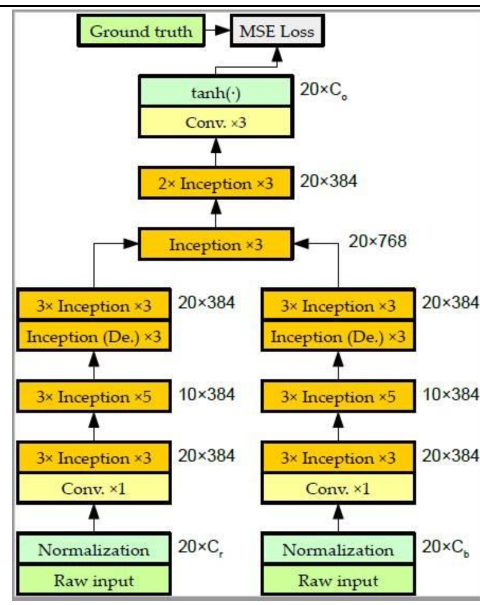

Figure 4: The deep learning network with Dual Data Feed for low frequency seismic data reconstruction

frequencies $\boldsymbol{f_1}$ an $\boldsymbol{f_2}$ with $\Delta \boldsymbol{f} = \boldsymbol{f_2} - \boldsymbol{f_1} \ll \boldsymbol{f_1}\ \boldsymbol{and}\ \boldsymbol{f_2}$. With the introduction of Beat Tone data, we establish a solid route connecting the low frequency components and the high frequency components through the subsurface low wavenumber structures, as sketched in Figure 3.

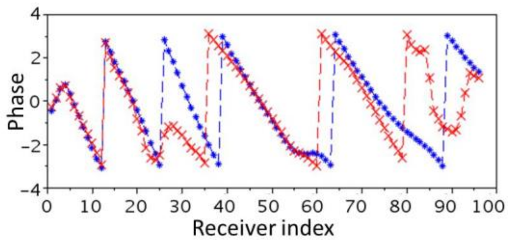

Figure 5: Comparison of true 3 Hz data (blue) and reconstructed 3 Hz data (red) from deep learning network. Receivers are deployed uniformly on the surface (receiver index 0 – 100 are corresponding to distance 0 – 4 km).

Based these observations, we designed an inception based deep learning network (Szegedy et al., 2017) featured a Dual Data Feed structure to receive both high frequency data and Beat Tone data simultaneously (Figure 4), to reduce the network structural complexity without sacrificing the prediction accuracy.

For proof of concept, we use the Marmousi model as an example to test the low frequency reconstruction performance of the network. First, high frequency data (≥10 Hz) are acquired for the Marmousi model as the production data. A training velocity model is selected arbitrarily, which is a truncated model of 1997 2.5D (The Carpathians thrusting over the North Sea) completely different from the Marmousi model. A set of training data is then generated by a forward modeling simulator, containing the high frequency components ranging from 10 Hz to 18 Hz with the interval of 0.5 Hz, and a low frequency (3 Hz) dataset, resulting in 18 frequencies in total. After that, we extract the Beat Tone data from the high frequency training dataset pairs (i.e., 10 Hz and 13 Hz, 10.5 Hz and 13.5 Hz, … etc.). Eventually, both the high frequency training data and the Beat Tone training data are input into the network for the supervised learning with the 3 Hz training data serving as the ground truth. The fully trained network is able to reconstruct the 3 Hz Marmousi data as plotted in Figure 5.

Progressive Transfer Learning

Although the testing result in Figure 5 generally validated that the fully trained network properly reconstructs the absent low frequency components, some remaining issues need to be addressed: 1) only one training model and the associated training dataset are used for the network training, which might be insufficient for accurate data reconstruction; 2) the training velocity model is selected arbitrarily, hence the training process might be biased, leading to the errors observed in Figure 5. One solution to the first problem is to train the network with multiple training models and training datasets, Unfortunately, to deal with general seismic data, the number of the training models needs to be very large (Ovcharenko et al., 2018), implying huge volumes of training data and extremely high network training cost. To address the second issue, we may incorporate the subsurface velocity *a priori* information in the training velocity model. However, we are facing a dilemma here: accurate subsurface velocity information is not available until the FWI is performed successfully.

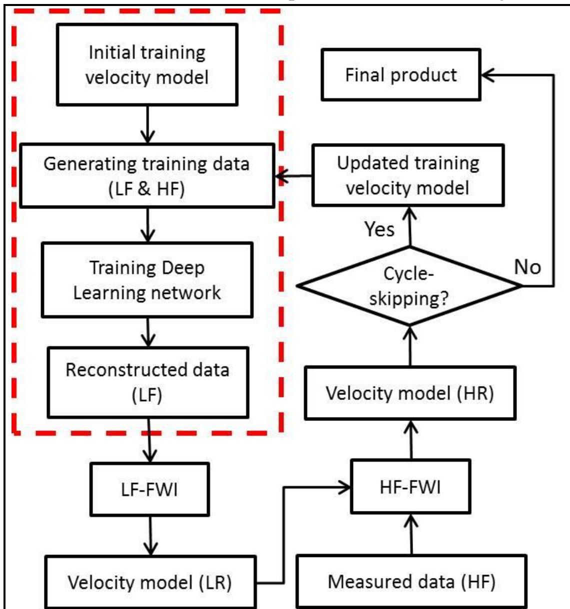

Figure 6: Workflow of Progressive Transfer Learning for low frequency data reconstruction. LF – low frequency; HF – high frequency; LR – low resolution; HR – high resolution.

To overcome these two difficulties, we propose a novel strategy called Progressive Transfer Learning. With this strategy, the deep learning network is always trained by a single training dataset but in an iterative manner. This training dataset evolves with the iteration proceeds. Each

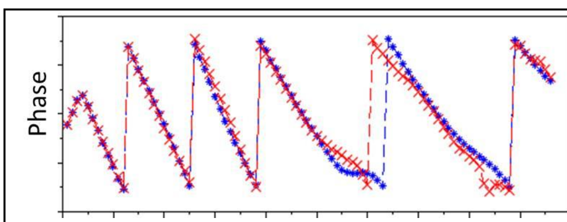

Figure 7: Comparison of true 3 Hz data (blue) and reconstructed 3 Hz data (red) by Progressive Transfer Learning Network.

Progressive Transfer Learning iteration consists of three modules: 1) network training module; 2) data reconstruction module; 3) training data generation module. The workflow of the Progressive Transfer Learning is shown in Figure 6.

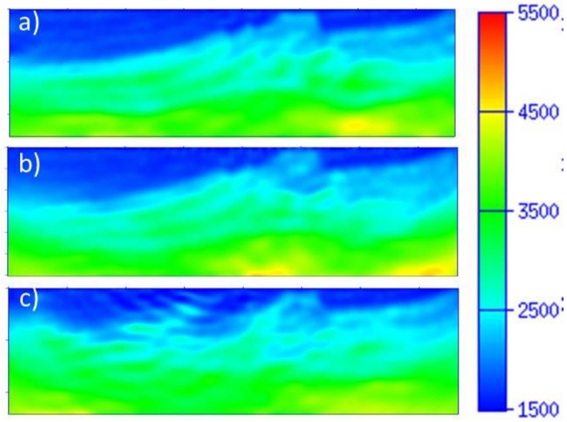

Figure 8: a) reference solution (FWI of 3 Hz, 6 Hz, and 10 Hz data); b) FWI result by inverting 3 Hz data reconstructed from high frequency data (≥10 Hz) by Progressive Transfer Learning network, followed by 10 Hz data FWI; c) FWI result of 10 Hz data.

The red block in Figure 6 represents the original deep learning network workflow discussed in the previous section, which is initiated by an arbitrarily selected training velocity model. The reconstructed low frequency data are then input into the FWI engine, providing a starting model for the subsequent high frequency FWI to complete the first iteration. After the first iteration of deep learning, the resulting high resolution FWI velocity model is still likely to be heavily contaminated by cycle-skipping-induced artifacts, partially due to the single arbitrary training velocity model. Next, instead of treating it as the final product, we send this velocity model back into the network training process, acting as the updated training velocity model to enter the second Progressive Transfer Learning iteration. In the every subsequent Progressive Transfer Learning iteration, the neural network injects a set of reconstructed low frequency data with increased prediction accuracy to the FWI engine. On the other hand, the FWI procedure continuously offers an improved training velocity model to the neural network. In other words, the deep learning network and the FWI engine complement each other alternatingly, progressively increasing the low frequency data prediction accuracy and eliminating the cycle-skipping phenomenon with the iteration proceeds.

The 3 Hz Marmousi data reconstructed by two iterations of the Progressive Transfer Learning plotted in Figure 7 are substantially more accurate than the original reconstruction result in Figure 5, indicating the power of the Progressive Transfer Learning method. The FWI velocity model obtained by inverting the reconstructed 3 Hz data followed by 10 Hz inversion (Figure 8b) is very close to the reference solution (the true 3 Hz FWI followed by 6 Hz and 10 Hz FWI in Figure 8a) and no cycle-skipping phenomenon is observed. On the other hand, the direct inversion of the 10 Hz data ends up with severe cycle-skipping artifacts (Figure 8c).

Conclusions and Discussion

A novel Progressive Transfer Learning method was developed to reconstruct the absent low frequency data from the acquired high frequency seismic data by learning the nonlinear relationship between different frequency bands. This pure data-driven approach does not require any *a priori* information of the subsurface geological structures. A training velocity model can be arbitrarily selected or constructed to initiate the network training process. After the initial network training and network prediction, each of the subsequent Progressive Transfer Learning iteration provides an improved training velocity model and training dataset to enhance the low frequency data reconstruction accuracy progressively until the training model converges to the true velocity model. The numerical example validated this approach and shown the robustness. The reconstructed low frequency data match the true data with high accuracy and the final FWI velocity model is immune to the cycle-skipping-induced artifacts.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Science, SC-1, under Award Number DE-SC0019665.

References

Alkhalifah, T., 2014. Scattering-angle based filtering of the waveform inversion gradients. Geophysical Journal International, 200(1), pp.363-373.

Choi, Y. and T. Alkhalifah, 2015. Unwrapped phase inversion with an exponential damping. Geophysics, 80(5), pp.R251-R264.

Fu L. and W. W. Symes, 2017, A discrepancy-based penalty method for extended waveform inversion, Geophysics, 82(5), R287-R298.

Fu L. and W. W. Symes, 2015, Reducing the cost of extended waveform inversion by multiscale adaptive methods, SEG Expanded Abstracts, 1127-1131.

Hu, W., 2014. FWI without low frequency data—Beat tone inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 1116–1120.

Hu, W., J. Chen, J. Liu, and A. Abubakar, 2018, Retrieving Low Wavenumber Information in FWI: An Overview of the Cycle-Skipping Phenomenon and Solutions, IEEE Signal Processing Magazine, 35(2), 132-141.

Jin, Y., W. Hu, X. Wu, and J. Chen, 2018. Learn low wavenumber information in FWI via deep inception based convolutional networks. In *SEG Technical Program Expanded Abstracts 2018* (pp. 2091-2095). Society of Exploration Geophysicists.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), p.436.

Li, Y. E. and L. Demanet, 2016, Full-waveform inversion with extrapolated low-frequency data. Geophysics, 81(6), pp.R339-R348.

Ma, Y. and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion, Geophysics, 78(6), pp.R223-R233.

Ovcharenko, O., Kazei, V., Peter, D., Zhang, X. and Alkhalifah, T., 2018, June. Low-Frequency Data Extrapolation Using a Feed-Forward ANN. In *80th EAGE Conference and Exhibition 2018*.

Shin, C. and Y. H. Cha, 2008, Waveform inversion in the Laplace domain. Geophysical Journal International, 173(3), pp.922-931.

Symes, W.W., 2008, Migration velocity analysis and waveform inversion. Geophysical prospecting, 56(6), pp.765-790.

Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A.A., 2017, February. Inception-v4, inception-resnet and the impact of residual connections on learning. In AAAI (Vol. 4, p. 12).

Sun, H. and Demanet, L., 2018. Low frequency extrapolation with deep learning. In *SEG Technical Program Expanded Abstracts 2018* (pp. 2011-2015). Society of Exploration Geophysicists.

Wu, R. S., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model. Geophysics, 79(3), pp.WA13-WA24.

Xu, S., F. Chen, G. Lambaré, Y. Zhang, and D. Wang, 2012, Inversion on reflected seismic wave, In 2012 SEG Annual Meeting. Society of Exploration Geophysicists.