Extrapolated full waveform inversion with convolutional neural networks

Hongyu Sun and Laurent Demanet, Massachusetts Institute of Technology*

SUMMARY

Computational low frequency extrapolation is in principle the most direct way to address the cycle skipping problem in full waveform inversion (FWI). We propose a method of extrapolated full waveform inversion (EFWI), where FWI is allowed to make use of data augmented by increasing its frequency band with a convolutional neural network (CNN). In extrapolated FWI with CNN (EFWI-CNN), the low-wavenumber components of the model are determined from the extrapolated low frequencies, before proceeding with a frequency sweep of the bandlimited data. The proposed deep-learning method of low-frequency extrapolation shows adequate generalizability for the initialization step of EFWI. Numerical examples show that the neural network trained on several submodels of the Marmousi model is able to predict the low frequencies for the BP 2004 benchmark model. Additionally, the neural network can robustly process seismic data with uncertainties due to the existence of noise, unknown source wavelet, and different finite-difference scheme in the forward modeling operator.

INTRODUCTION

It has been recognized that FWI can be a good method to invert for the subsurface structure with high resolution (Virieux and Operto, 2009). However, the missing low frequencies are probably the sole reason why the inverse problem is a highly nonconvex optimization problem (Li and Demanet, 2016, 2017). Our previous work has shown that casting bandwidth extension as a regression problem in supervised learning with a deep convolutional network can successfully extrapolate the low frequency data of band-limited recordings (Sun and Demanet, 2018).

In this note, we demonstrate the reliability of the extrapolated low frequencies to seed frequency-sweep FWI on the BP 2004 benchmark model (Billette and Brandsberg-Dahl, 2005). Two precautions are taken to ensure that trivial deconvolution of a noiseless record (by division by the high frequency (HF) wavelet in the frequency domain) is not an option: (1) add noise to the testing records, and (2) for testing, choose a hard bandpass HF wavelet taken to be zero in the low frequency (LF) band. In one numerical experiment involving bandlimited data above 0.6Hz from the BP 2004 model, the inversion results indicate that the predicted low frequencies are adequate to initialize conventional FWI from an uninformative initial model, so that it does not suffer from the otherwise-inherent cycle-skipping at 0.6Hz.

METHOD

CNNs are a method of supervised learning, i.e., inference of $\mathbf{y}_i$ from $\mathbf{x}_i$. We need to first train the CNN from a large number of samples $(\mathbf{x}_i, \mathbf{y}_i)$ to determine the coefficients of the network, and then use the network for testing on new $\mathbf{x}_i$. The object of this note is that the $\mathbf{x}_i$ can be taken to be seismograms bandlimited to the high frequencies, and $\mathbf{y}_i$ the same seismograms bandlimited to the low frequencies. Generating training samples means collecting, or synthesizing seismogram data from a variety of geophysical models, which enter as space-varying elastic coefficients in a wave equation.

The performance of the neural network is sensitive to the architecture and the hyperparameters, so we must design them carefully. Next, we illustrate the specific choice of architecture and hyperparameters for bandwidth extension, along with numerical examples involving synthetic data from community models.

In EFWI-CNN, the true unknown velocity model for FWI is referred to as the test model, since it is used to collect the test data set in deep learning. To collect the training data set, we create training models by randomly selecting nine parts of the full-size Marmousi2 P-wave velocity model (Figure 1) with different structure but the same number of grid points 166×451. We also downsample the original model to 166×451 pixels as the test model. We find that the randomized models produced in this manner are realistic and diversified enough to demonstrate the generalization of the neural network.

The following processing steps are used to collect each sample (i.e., shot record) in both the training and test data sets. The acquisition geometry of forward modeling on each model is the same. It consists of 30 sources and 451 receivers evenly spaced at the surface. We use a sixth order in space and second order in time finite-difference modeling method with PML to solve the 2D acoustic wave equation in the time domain, to generate the wavefield of both the training and test data sets. The sampling interval and the total recording time are 1.5ms and 5.25s, respectively. For each trace, we use a LF wavelet below 5Hz to synthesize the output, and a HF wavelet above 5Hz to synthesize the input of the neural networks. Both the low and high frequency wavelets are obtained by a sharp windowing of the same original Ricker wavelet with dominant frequency 15Hz.

The architecture of our neural network is a feed-forward stack of five sequential combinations of the convolution, batch normalization and PReLU layers, followed by one fully connected layer that outputs continuous-valued amplitude of the time-domain signal in the low frequency band. The numbers of filters in the five convolutional layers are 128, 64, 128, 64 and 1, respectively. The size of all the filters in our neural networks are 40×7. Other training paremetes are the same as our previous work (Sun and Demanet, 2018).

After training with 200 epochs, Figure 2 compares the shot gather between the bandlimited data ($5-30$ Hz), true and extrapolated low frequencies ($0.1-5$ Hz) where the source is located at the horizontal distance $x = 3.0km$ on the Marmousi2 model. The extrapolated results in Figure 2(c) show that the

proposed neural network can accurately predict the recordings in the low frequency band, which are totally missing before the test.

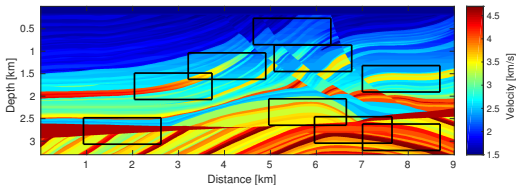

Figure 1: The nine training models randomly extracted from the Marmousi2 P-wave velocity model to collect the training data set. The test models are the Marmousi2 P-wave velocity model and the BP 2004 benchmark model.

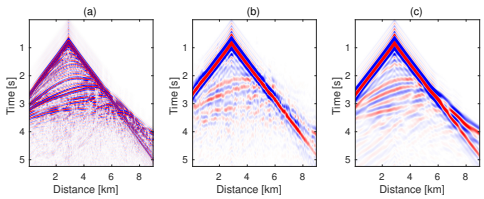

Figure 2: The extrapolation result on the Marmousi model: comparison among the (a) bandlimited recordings ($5-30$Hz), (b) predicted and (c) true low frequency recordings ($0.1-5$Hz).

UNCERTAINTY ANALYSIS

With a view toward dealing with complex field data, we investigate the stability of the neural network's predictive performance under three kinds of discrepancies, or uncertainties, between training and testing: additive noise; different finite difference operator in the forward modeling; and different source wavelet. In every case, we compare the extrapolation accuracy with the reference in Figure 2, where training and testing are set up the same way (noiseless bandlimited data, finite difference operator with second order in time and sixth order in space, and the high and low frequency wavelets which are bandpassed from the Ricker wavelet with 15Hz dominant frequency).

In the first case, the neural network is expected to extrapolate the low frequencies from the noisy bandlimited data. We add 25% Gaussian noise to the bandlimited data in both the training and test data sets. The low frequencies in the training set are noiseless as before. Even though noise will disturb the neural network to find the correct mapping between the bandlimited data with their low frequencies, Figure 3 shows that the proposed neural network can still successfully extrapolate the low frequencies of the main reflections. The neural network is able to perform extrapolation as well as denoising. Incidentally, we make the (unsurprising) observation that CNN has potential for the denoising of seismic data.

One main challenge of FWI is that the observed and calculated data come from different wave propagation schemes. For

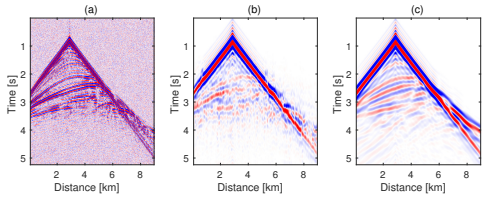

Figure 3: Noise robustness: comparison among the (a) bandlimited recordings ($5-30$Hz), (b) predicted and (c) true low frequency recordings ($0.1-5$Hz) on the Marmousi model.

example, under the control of different numerical dispersion curves, the phase velocity would have different behaviors if we used different finite difference (FD) operators to simulate the observed and calculated data. Therefore, it is necessary to study the influence of different discretization, or other details of the simulation, on the accuracy of low frequency extrapolation. In our case, the shot gathers in the test data set are simulated with a fourth order FD operator, but the neural network is trained on the samples simulated with a sixth order FD operator. With the lower order FD operator, the grid dispersion is clearly visible in Figure 4(a). However, the extrapolation result in Figure 4(b) shows that the neural network trained on the sixth order operator is able to extrapolate the low frequencies of the bandlimited data collected with the fourth order operator. The neural network appears to be stable with respect to mild modifications to the forward modeling operator, at least in the examples tried.

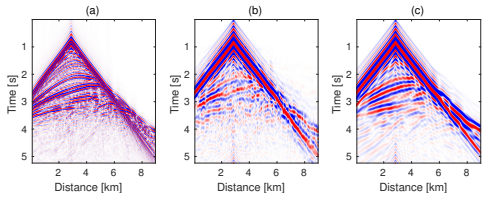

Figure 4: Forward modeling operator robustness: comparison among the (a) bandlimited recordings ($5-30$Hz), (b) predicted and (c) true low frequency recordings ($0.1-5$Hz) on the Marmousi model.

Another uncertainty is the unknown source wavelet. To check the extrapolation capability of the neural network in the context of data excited by an unknown source wavelet, we train the neural network with a 15Hz Ricker wavelet, but test it with a 10Hz Ricker wavelet. In this case, we bandpass the low and high wavelets from the Ricker wavelet with different dominant frequency when we synthesize the samples in the training and test data set. Figure 5 shows that the neural network trained on the data from the 15Hz source wavelet is able to extrapolate the data synthesized with the 10Hz source wavelet. The commonplace issue of the source wavelet being unknown or poorly known in FWI has seemingly little effect on the performance of the proposed neural network to extrapolate the low frequency data, at least in the examples tried.

Even though all of the uncertain factors hurt the accuracy of extrapolated low frequencies to some extent, the CNN's pre-

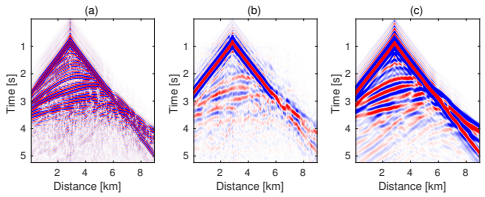

Figure 5: Unknown source wavelet robustness: comparison among the (a) bandlimited recordings ($5-30$Hz), (b) predicted and (c) true low frequency recordings ($0.1-5$Hz) on the Marmousi model.

diction has a degree of robustness that surprised us. All of the extrapolation results in the above numerical examples can be further improved by increasing the diversity of the training data set, because subjecting the network to a broader range of scenarios can fundamentally reduce the generalization error of the deep learning predictor.

EXTRAPOLATED FWI: BP MODEL

To check the validity of the method, we perform EFWI-CNN on the BP 2004 benchmark model (Figure 6) with the extrapolated low frequency data predicted by the neural network trained on the submodels of Marmousi2 P-wave velocity model (Figure 1). The objective function of the inversion is formulated as the least-squares misfit between the observed and calculated data in the time domain. We use the LBFGS (Nocedal and Wright, 2006) optimization method to update the model gradually from the lowest frequency that the data contained, to successively higher frequencies.

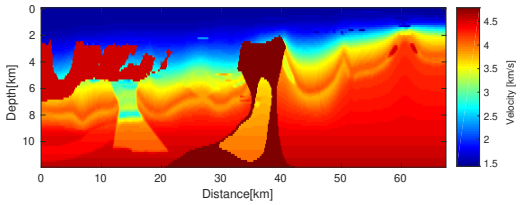

Figure 6: The 2004 BP benchmark velocity model used to collect the test data set for studying the generalizability of the proposed neural network.

To reduce the computation burden, we downsample the BP 2004 benchmark model to 80×450 grid points with a grid interval of 150m. Numerical examples show that, starting from the bad initial model (Figure 7(a)), the highest starting frequency to avoid cycle-skipping on this model is 0.3Hz. Therefore, we should extrapolate the bandlimited data to at least 0.3Hz to invert the BP model successfully.

In this example, we use a 7Hz Ricker wavelet as the source to simulate the fullband seismic records on both the training models (submodels of Marmousi2) and the test model (BP model). The sampling interval and the total recording time are 5ms and 10s, respectively. To collect the input of the CNN, a highpass filter where the low frequency band ($0.1-0.5$Hz) is exactly zero is applied to the fullband seismic data. The bandlimited data ($0.6-20$Hz) are fed into the proposed CNN model to extrapolate the low frequency data in $0.1-0.5$Hz trace by trace.

In addition to the training procedure described earlier, four modifications are used to improve the extrapolation accuracy on the BP model: (1)Increase the filter size to 200; (2)Increase the width of the last convolutional layer from 1 to 32; (3)Add one dropout layer (Srivastava et al., 2014) with a probability of 50% after the first convolution layer to further reduce the generalization error. (4)Rescale the input feature by dividing the bandlimited recording by its maximum amplitude and rescale the output of CNN by multiplying the amplitude of LF with a large constant.

The extrapolated low frequency data are used to invert the low-wavenumber velocity model with the conventional FWI method. We observe that the accuracy of extrapolated low frequency decreases as the offset increases, so we limit the maximum offset to 12km. Starting from the initial model (Figure 7(a)), Figure 7(b) and Figure 7(c) show the low-wavenumber inverted models using 0.3Hz extrapolated data and 0.3Hz true data, respectively. Compared to the initial model, the resulting model using the 0.3Hz extrapolated data reveals the positions of the high and low velocity anomalies, which is almost the same as that of true data. The low-wavenumber background velocity models can still initialize the frequency-sweep FWI in the right basin of attraction.

Figure 8 compares the inverted models from FWI using 0.6-2.4Hz bandlimited data, starting from the respective large-wavenumber models in the previous figure. In (a), the resulting model starts from the original initial model. In (b), the resulting model starts from the inverted low-wavenumber velocity model using 0.3Hz extrapolated data. In (c), the resulting model starts from the inverted low-wavenumber velocity model using 0.3Hz true data. With the low-wavenumber velocity model, FWI can find the accurate velocity boundaries by exploring higher frequency data. However, the inversion settles in a wrong basin with only the higher frequency components. The low frequencies extrapolated with deep learning are reliable enough to overcome the cycle-skipping problem on the BP model, even though the training data set is ignorant of the particular subsurface structure of BP – salt bodies. Therefore, the neural network approach has the potential to deal favorably with real field data.

So far, the experiments on BP 2004 have assumed that data are available in a band starting at 0.6Hz. We now study the performance of EFWI-CNN when this band starts at a frequency higher than 0.6Hz. We still start the frequency-sweep FWI with 0.3Hz extrapolated data, and the highest frequency is still fixed at 2.4Hz. With the increase of the lowest frequency of bandlimited data, Figure 9 compares the quality of the inverted models at each iteration for FWI using fullband data, EFWI-CNN, and FWI using only the bandlimited data. The norm of the relative model error is used to evaluate the model quality, as (Brossier et al., 2009)

$$mq = \frac{1}{N} \| \frac{\mathbf{m}_{inv} - \mathbf{m}_{true}}{\mathbf{m}_{true}} \|_2 \qquad (1)$$

where $\mathbf{m}_{inv}$ and $\mathbf{m}_{true}$ are the inverted model and the true model,

respectively. N denotes the number of grid point in the computational domain. The performance of EFWI-CNN of course decreases with the increase of the lowest frequency of the bandlimited data, because this leads to more extrapolated data involved in the frequency-sweep FWI. The more iterations of FWI with the extrapolated data, the more errors the inverted model will have before exploring the true bandlimited data. Overfitting of the unfavorable extrapolated data makes the inversion worse after several iterations with the extrapolated data. However, EFWI-CNN is still superior to using FWI with only bandlimited data. We observe that EFWI-CNN with the current architecture still helps to reduce the inverted model error on the BP model when the lowest available frequency is as high as 1.2Hz.

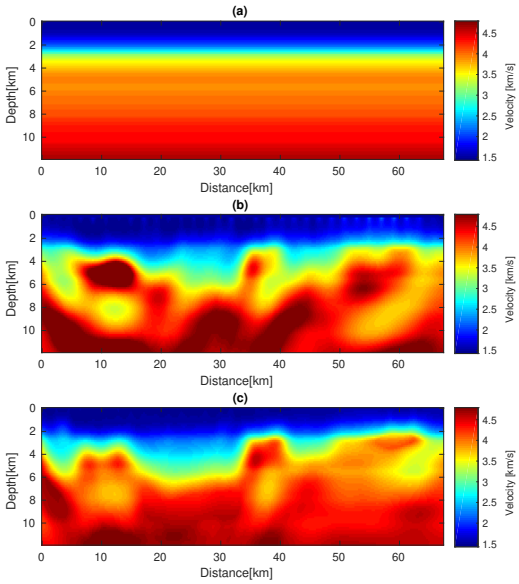

Figure 7: Comparison among (a) the initial model for FWI on the BP model, the inverted low-wavenumber velocity models using (b) 0.3Hz extrapolated data and (c) 0.3Hz true data. The inversion results in (b) and (c) are started from the initial model in (a).

Even though we are encouraged by the ability of a CNN to generate $[0.1, 0.5]$Hz data for the BP 2004 model, much work remains to be done to be able to find the right architecture that will generate data in larger frequency bands, for instance in the $[0.1, 1.4]$Hz band. Finding a suitable network architecture, hyperparameters, and training schedule for such cases remains an important open problem. Other community models, and more realistic physics such as elastic waves, are also left to be explored.

CONCLUSIONS

Extrapolated low frequency data from a CNN can be reliable enough to invert the low-wavenumber velocity model, which can then be used for initializing FWI on the bandlimited data without cycle-skipping. Training can be done on markedly different geophysical models than the scenario for which the network is then tested. Even though there is freedom in choosing the architectural parameters of the deep neural network, making the CNN have a large receptive field is necessary for low frequency extrapolation.

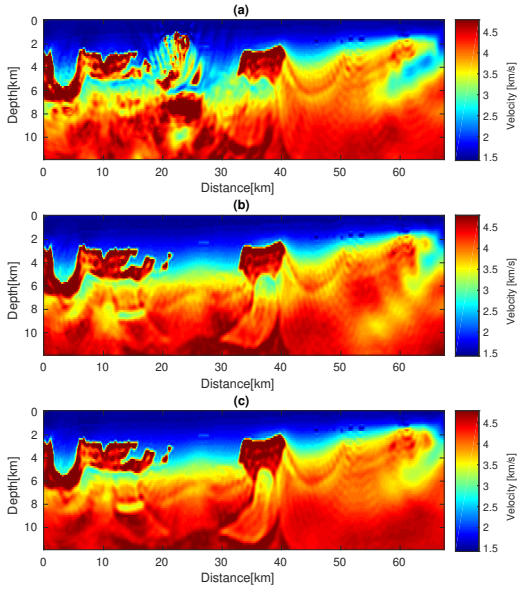

Figure 8: Comparison of the inverted models from FWI using 0.6-2.4Hz bandlimited data. In (a), resulting model starts from the original initial model. In (b), resulting model starts from the inverted low-wavenumber velocity model using 0.3Hz extrapolated data. In (c), resulting model starts from the inverted low-wavenumber velocity model using 0.3Hz true data.

ACKNOWLEDGMENTS

The authors thank Total SA for support. LD is also supported by AFOSR grant FA9550-17-1-0316. The Python Seismic Inversion Toolbox (PySIT) (Hewett et al., 2013) is used for FWI in this note.

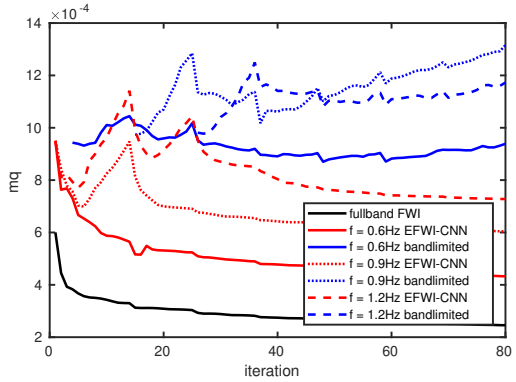

Figure 9: Comparison of the quality of the inverted models at each iteration for FWI using fullband data (black line), EFWI-CNN (red line) and FWI using only bandlimited data (blue line). f donates the lowest frequency of the bandlimited data.

REFERENCES

Billette, F., and S. Brandsberg-Dahl, 2005, The 2004 bp velocity benchmark: Presented at the 67th EAGE Conference & Exhibition.

Brossier, R., S. Operto, and J. Virieux, 2009, Seismic imaging of complex onshore structures by 2d elastic frequency-domain full-waveform inversion: Geophysics, **74**, WCC105–WCC118.

Hewett, R., L. Demanet, and the PySIT Team, 2013, PySIT: Python seismic imaging toolbox v0.5. (Release 0.5).

Li, Y., and L. Demanet, 2017, Extrapolated full-waveform inversion: An image-space approach: Society of Exploration Geophysicists.

Li, Y. E., and L. Demanet, 2016, Full-waveform inversion with extrapolated low-frequency data: Society of Exploration Geophysicists, **81**.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization 2nd.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, 2014, Dropout: a simple way to prevent neural networks from overfitting: The Journal of Machine Learning Research, **15**, 1929–1958.

Sun, H., and L. Demanet, 2018, Low frequency extrapolation with deep learning, *in* SEG Technical Program Expanded Abstracts 2018: Society of Exploration Geophysicists, 2011–2015.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.