Comparing Deep Neural Network and Full Waveform Inversion Velocity Estimation

Stuart Farris, Joseph Jennings, and Mauricio Araya Polo

ABSTRACT

We explore the feasibility of a deep learning approach for tomography by comparing it with the current velocity prediction techniques used in the industry. This is accomplished through quantitative and qualitative comparisons of velocity models predicted by a Machine Learning (ML) system and those of two variations of full-waveform inversion (FWI). Additionally, we compare computational aspects of the two approaches. The results show that the ML-based reconstructed models are competitive to the FWI-produced models in terms of selected metrics, and widely less expensive to compute.

INTRODUCTION

Velocity model building is a key step during seismic processing and interpretation, current industry tools of choice are tomography and full-waveform inversion (FWI). An alternative was presented in Araya-Polo et al. (2018), where the inverse problem is solved with a novel machine learning (ML) method. This approach has multiple advantages with respect to the classical methods. In principle, the main cost is the one-time upfront training of a neural network. Once the network is trained, velocity model prediction costs are negligible allowing cheap exploration of multiple scenarios. While the quality of model predictions from the ML scheme are highly dependent on the labeled data used to train the network, this reliance on labeled data also frees the trained model from human and methodological biases.

Until now, direct comparisons between this new ML prediction scheme and other velocity prediction methods have not been made. Our contribution here is to compare qualitatively and quantitatively the reconstructed model generated by the ML approach and by variations of the well-established FWI approach.

MACHINE LEARNING TOMOGRAPHY

In this section a brief summary of Araya-Polo et al. (2018) is introduced. The basic idea of using a machine learning approach for velocity estimation is to replace the following expression:

$$J(m) = ||d_m(m) - d_{obs}||_2^2, \tag{1}$$

where m is the optimal earth model that minimizes $J(m)$, d_m is a data vector that is modeled from a non-linear modeling operator $f(m)$, and d_{obs} is the recorded data vector, with a machine learning approach commonly expressed as:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L(V_i, T(X_i, \theta)), \tag{2}$$

where $T(X, \theta)$ is the tomography operator, parameterized by the coefficients vector θ, X is the input to the tomography operator, and its output is the reconstructed velocity model $\hat{V}$. In machine learning terminology, X is known as the input feature and V is known as the label. The loss function $L(V_i, \hat{V}_i)$ measures the difference between the ground truth velocity model and its reconstructed version. The loss function employed in this work is the squared error $L(V_i, \hat{V}_i) = (V_i - \hat{V}_i)^2$, which is frequently used in regression problems. Replacing the generic loss function with the squared-error loss, we can express the optimization problem in Equation 2 as

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} (V_i - T(X_i, \theta))^2. \tag{3}$$

This familiar regression problem can be solved with a gradient descent approach which iteratively updates the coefficients of θ. Both equations 2 and 3 can be seen as inverse problems, but the solution of the systems are reaching through very different approaches. The former is a deterministic optimization problem where the latter is a learning process in which a statistical mechanism helps minimize the loss function.

The novelty of the work presented in Araya-Polo et al. (2018) is that the tomographic operator $T(X, \theta)$ is implemented as a deep neural net (DNN) composed of layers of weighted nodes parameterized by θ. The input to the network is connected to the input layer which is followed by a varying number of hidden layers. The inputs of the hidden layer are activated by the outputs of the previous layer and eventually the output of the network is computed at the output layer. The output vector is a prediction of an earth model that would have modeled the input feature vector X. These networks are trained with examples per the statistical-learning approach in which the correct output (label) is known for a given input, and the weight parameters in the nodes of the network update due to the minimization of the error between the prediction and true value. Expressing the tomography operator as a DNN, it can be written as

$$T(X, \theta) = f_{out}(f_3(f_2(f_1(X, \theta_1), \theta_2), \theta_3)\theta_{out}), \tag{4}$$

where f_i is a layer within the DNN parameterized by the vector θ_i, respectively. Figure 1 illustrates this formulation and the workflow that moves from feature space to model prediction and Figure 2 shows the workflow used to train the weights of the DNN.

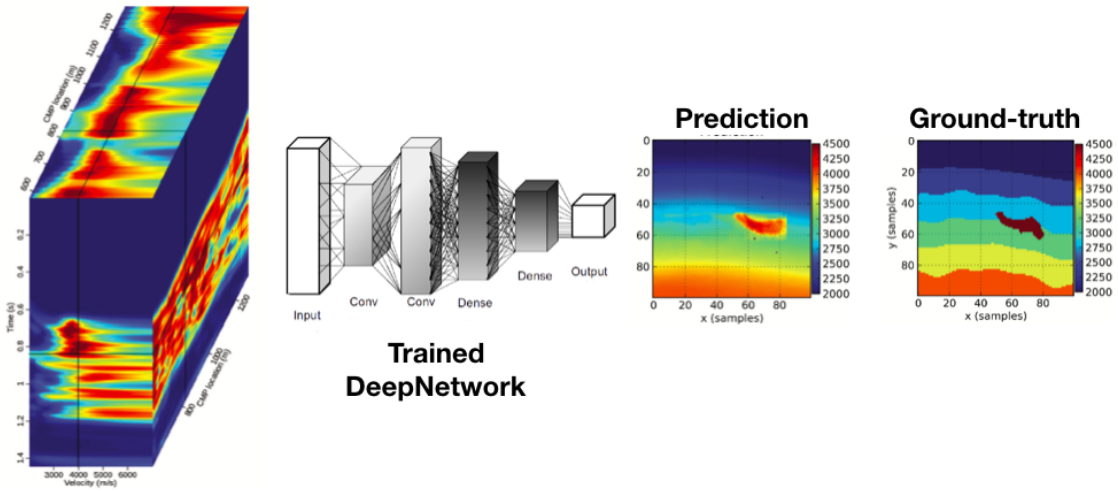


Figure 1: Prediction workflow. [**NR**]

Arguably the most crucial step in the DNN tomography formulation is the choice of input (X) used for the first layer of the DNN. It is necessary that we select a feature that reduces the input size but also amplifies the relevant changes in the data caused by the model parameters we wish to estimate. Velocity semblance (Taner and Koehler, 1969) was chosen by Araya-Polo et al. (2018) as the input feature to the DNN. This choice was made as velocity semblance gives a measure of apparent velocity with depth and is commonly the first step in estimating the velocity from reflection seismic data.

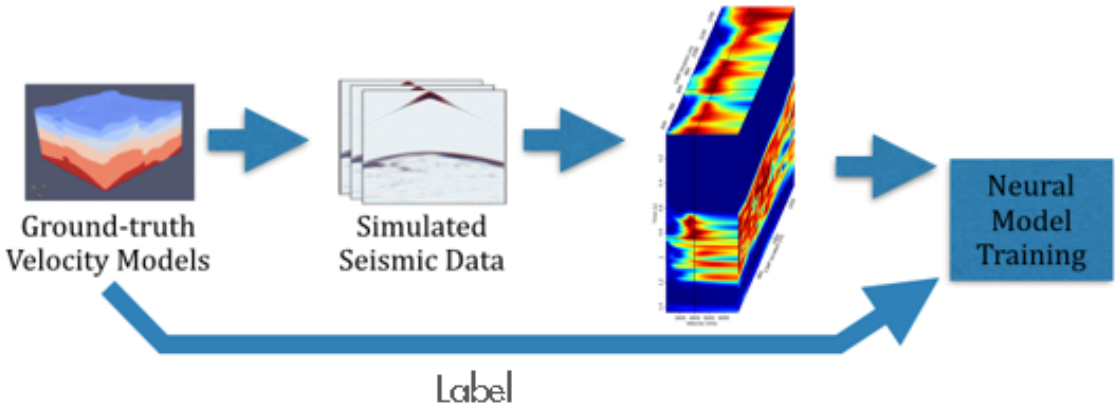


Figure 2: Training workflow. [**NR**]

BASELINE VELOCITY ESTIMATION TECHNIQUES

Full Waveform Inversion

In exploration geophysics, FWI is a topic of intense study and is at the forefront of earth model building from seismic data (Virieux and Operto, 2009). Despite numerous limitations (e.g., high computational cost, starting model sensitivity, unwanted convergence to local minimum) FWI is regarded as an area of development that may rectify the gap between low and high wavenumber earth model building and represent an all-inclusive solution to seismic exploration. For this reason we have chosen it as a baseline method to compare the predictions of the ML approach. Of course there are many other velocity prediction techniques used in the oil and gas industry that could be compared to the ML approach. In particular, inverting the Dix equation at CMP locations may be a more comparable method than FWI since it also brings the data into semblance space. But, we anticipate a 1D Dix inversion paired with lateral smoothing will be vastly outperformed by ML and FWI.

FWI is a nonlinear data fitting problem described in the iconic work of Tarantola (1984) which reformulates the exploding reflector concept of Claerbout (1971) as a local optimization problem. Minimizing the difference between observed and modeled seismic data with respect to some earth model can be written in the form of Equation 1 where $f(m)$ is the forward nonlinear wave equation modeling operator that maps the earth model space, m, into the data space, d_m, for a set of seismic experiments. This problem can be solved with a variety of gradient descent methods in which the model is updated iteratively from the gradient of the objective function J at the current model iteration m_j:

$$m_{j+1} = m_j + \alpha_j s_j. \tag{5}$$

The next model, m_{j+1}, is found by summing the current model, m_j, to the search direction, s_j, scaled by a step length, α_j. There are many ways to compute the search direction, s_j. Here we will implement the nonlinear conjugate gradient method in which:

$$s_j = s_{j-1} + \beta \nabla J(m_j), \tag{6}$$

where s_{j-1} is the previous search direction and β is the conjugate direction coefficient. Furthermore, $\nabla J(m_j)$ is the adjoint of the wave equation operator linearized around the current model iteration applied to the difference between the modeled and observed data.

Multi-scale Full Waveform Inversion

The high nonlinearity of FWI creates many local minimum of the objective function in Equation 1. These local minimum prevent the convergence of methods like conjugate gradient from reaching reasonable solutions unless beginning from models

fairly close to ground truth. Many methods exist and extensive research continues to find ways to avoid these convergence issues. A highly effective and widely accepted method is that of Bunks et al. (1995) which is referred to as Multiscale FWI. This technique decomposes the FWI problem by scale and performs conventional FWI with progressively higher bandpasses of the source wavelet and observed data.

COMPARISON SETUP

To compare the velocity model predictions of ML and FWI approaches, four synthetic seismic surveys are created and used as inputs for three velocity prediction methodologies: conventional FWI, Multiscale FWI, and a DNN. The intent is to keep the input data consistent in order to create a fair comparison between each method. Below describes the data generation, the parameters of the three experiments, and the quantitative methods used to compare model results.

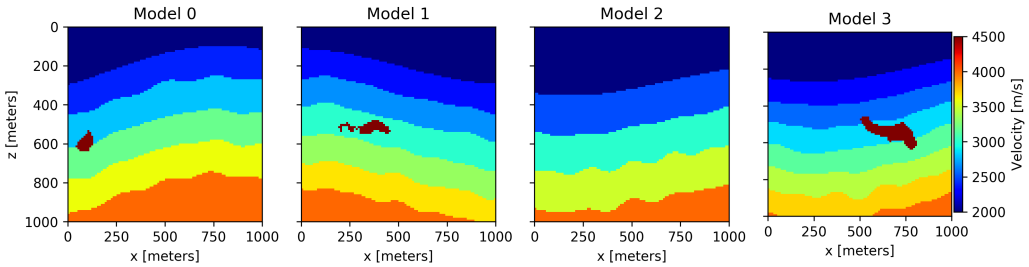


Figure 3: Ground-truth velocity models. The numbering is as they appear in Table 1. [**CR**]

Synthetic Seismic Data

The synthetic models used to generate the seismic data are 1.8 kilometers in the x direction and 1.4 kilometers in the z direction with grid cell discretization of 10 meters. The models parameter used is pressure wave velocity that increases with depth and contains salt bodies of varying shape and size. The velocities range from 2.0 km/s to 4.5 km/s.

The data itself is generated from 19 shots at the surface with 40 meter spacing in the x direction beginning at 520 meters. The shot wavelet is a 15 Hz peak Ricker. 144 receivers located at the surface record pressure data. They begin at 180 meters in x with 10 meter spacing. The wave propagation modeling assumes an acoustic, constant density earth and uses second order approximation in time and eighth order in space. Figure 3 illustrates the four models used to compare each method. Note, the data was generated on 1.8×1.4 km model but the velocity predictions were made on a 1.0×1.0 km subset of the original models.

Experiments

The first experiment is conventional FWI. 1000 iterations of nonlinear conjugate gradient are performed using all frequencies of all modeled shots. The starting model was a linear velocity gradient from 2.0 km/s to 4.5 km/s. A variation of this experiment is also conducted in which 200 conjugate gradient iterations are performed using the predicted model from the DNN as the starting model for FWI. The second experiment is Multiscale FWI which performed 150 conjugate gradient inversions over 5 bandpasses of the all modeled shots. The first 4 bandpasses of the data were smoothly tapered at 4Hz, 8Hz, 16Hz, and 32Hz. The fifth inversion used all frequencies. The starting model for the 4Hz inversion was a linear velocity gradient from 2.0 km/s to 4.5 km/s. Each progressively higher bandpass inversion uses the final model from the previous bandpass inversion. The third experiment results are obtained by exposing the trained neural network to unseen data, in our case, to unseen semblance cubes from velocity models created by our pseudo-random velocity model generator.

Comparison Metrics

In terms of quantitative metric for model quality comparison, we decided to recourse to the widely accepted standard metric in image dominated fields, the Structural Similitude Index Metric (SSIM) (see Wang et al. (2004)). SSIM differs from traditional objective metric since it is based on structural degradation rather than error or general distortion of the images. We also use Mean-Square-Error (MSE) and Signal-to-Noise ratio (SNR) as valid metrics since we know the ground-truth (GT) model used to generate the input data. We compute SNR as found in Johnson and Dudgeon (1992) using the correlation coefficient ρ as follows:

$$\rho = \frac{\text{cov}(I_{GT}, I_r)}{\text{stddev}(I_{GT}) * \text{stddev}(I_r)} \tag{7}$$

$$\text{SNR} = 10 \times \log_{10}(\frac{\rho^2}{(1-\rho^2)}) \tag{8}$$

RESULTS

We perform the comparative analysis on four seismic datasets generated from the velocity models in Figure 3. The comparison is limited to four datasets because of the high computational cost of FWI. In fact, retrieving one Multiscale FWI result takes more time than training the DNN used for the ML approach. After the upfront cost of creating the trained DNN, a single model prediction can be made almost instantaneously. This speaks to the computational cost of ML compared to FWI.

Three of the four compared models contain salt bodies and one model is a simple layer-cake model. Our selection process was not random, as interest on salt body detection is pervasive to the industry. Figure 4 shows the comparison between the

results of the DNN prediction, Multiscale FWI, and conventional FWI for model 0. When comparing the results of the three approaches, we observe that both the DNN prediction and Multiscale were able to recover the original velocity model with good accuracy while the conventional FWI approach cycle-skipped and was not able to recover a reasonable velocity solution. Additionally, the difference plots (third column) show that the multiscale approach performs best at resolving the interfaces between the layers. In fact, in general we find that the output of the DNN is smoother than the velocity estimated via Multiscale FWI. This likely due to the fact that when calculating the input semblance cubes, a smoothing occurs which limits the maximum frequency in the semblance cube. Multiscale FWI, however attempts to match modeled and predicted data that may have a broader range of frequencies.

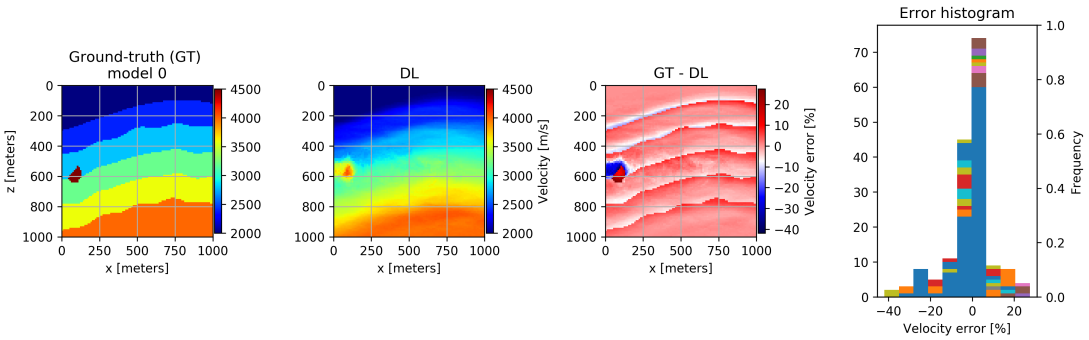
Possible our most interesting result can be drawn from model 0* which uses the DNN result as the starting model for conventional FWI. This example beats all other conventional FWI attempts and directly competes with the Multiscale FWI results with performing only 15% the number of the iterations.

Ground-Truth	DL Result			MS FWI Result			FWI Result		
	SSIM	MSE	SNR [dB]	SSIM	MSE	SNR [dB]	SSIM	MSE	SNR [dB]
model 0	0.66170	19070.4	13.8877	**0.84702**	**12149.2**	**16.2797**	0.49666	16510.5.0	7.2143
model 0*	-	-	-	-	-	-	0.82215	12761.2	15.6734
model 1	0.72079	14682.3	14.8517	**0.84896**	**13764.1**	**15.3158**	0.48856	275104.0	7.1058
model 2	0.78219	7250.57	18.0263	**0.87302**	**6094.7**	**19.1497**	0.32245	400733.0	6.0398
model 3	0.76253	**15402.4**	**14.6351**	**0.84241**	18620.7	13.9951	0.41005	452631.0	3.9613

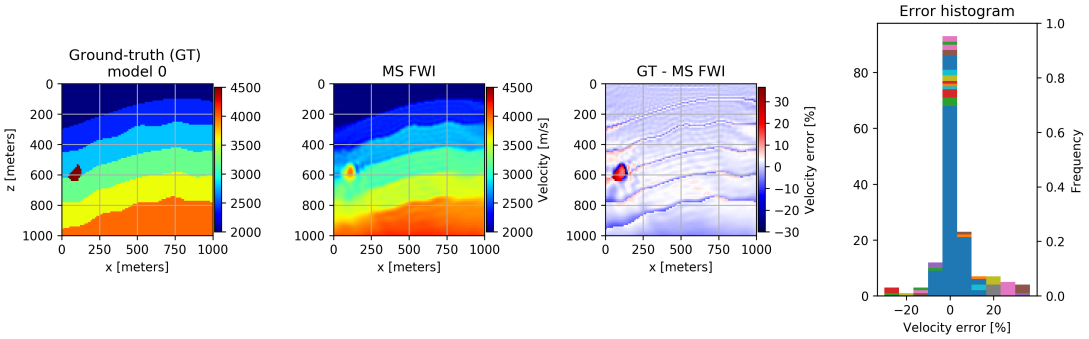
Table 1: Metrics summarizing results for deep learning prediction (DL), multi-scale FWI (MS FWI) and standard FWI for all models shown in Figure 3. In each metric category for each respective model, the best experiment result is bold. For example, for model 3 the best SSIM result was from MS FWI while the best MSE and SNR results came from DL. Model 0* uses the reconstructed model by DL as the initial velocity model for FWI.

CONCLUSION

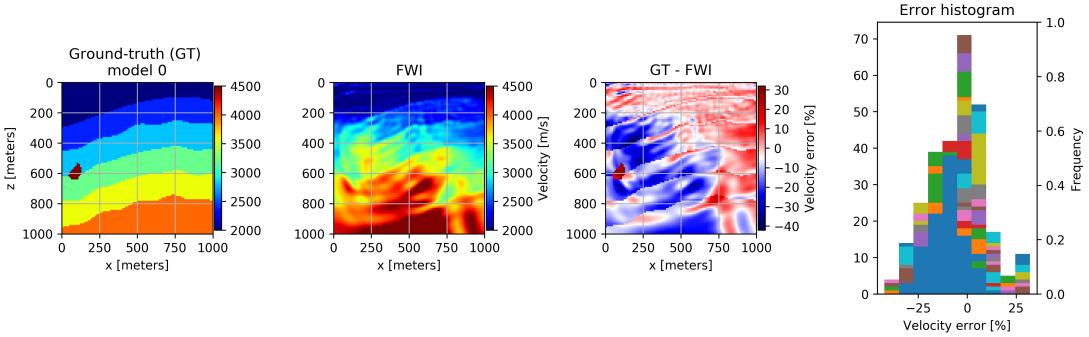
With respect to our three metrics, the DNN reconstructed models are competitive with the results Multiscale FWI. This demonstrates potential for use ML methods for velocity estimation applications in exploration seismology. Furthermore, the training of the DNN and the mode predicting take a fraction of FWI runtime, therefore opening real possibilities for multi-scenario analysis and effective uncertainty quantification efforts.



(a)



(b)



(c)

Figure 4: Comparison of tomography results from the DL and FWI for model 0. Leftmost column shows ground-truth (label), second from left shows the prediction from the DL (top), the Multiscale (MS) FWI result (middle) and the standard FWI result (bottom). Third column from left shows the difference between the ground truth and the prediction as a percentage of the velocity error. The last column shows the percentage of velocity errors for each sample binned and plotted in a histogram form. When comparing the prediction of the DNN to the MS FWI result, we observe that the DNN has difficulty in resolving sharp interfaces. Also note that a MS FWI approach was necessary to avoid cycle skipping that is apparent with the conventional FWI result. [**CR**]

ACKNOWLEDGMENTS

We wish to thank Shell International Exploration & Production Inc. for allowing the publication of this material.

REFERENCES

Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke, 2018, Deep-learning tomography: The Leading Edge, **37**, 58–66.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Claerbout, J. F., 1971, Toward a Unified Theory of Reflector Mapping: Geophysics, **36**, 467–481.

Johnson, D. H. and D. E. Dudgeon, 1992, Array signal processing: Concepts and techniques: Simon & Schuster.

Taner, M. T. and F. Koehler, 1969, Velocity spectradigital computer derivation applications of velocity functions: Geophysics, **34**, 859–881.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

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

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, 2004, Image quality assessment: from error visibility to structural similarity: IEEE Transactions on Image Processing, **13**, 600–612.