Supervised and unsupervised learning for velocity model building

Stuart Farris and Mauricio Araya

ABSTRACT

We present a deep learning (DL) workflow that takes analog velocity models and realistic raw seismic waveforms as inputs and produces subsurface velocity models as output. When insufficient data is used for training, DL algorithms tend to either over-fit or fail completely. Gathering large amounts of labeled and standardized seismic data sets is not straight forward. We address this shortage of quality data by building a Generative Adversarial Network (GAN) (Goodfellow et al., 2014) to augment our original training data set, which then is used by the DL-driven seismic tomography as input. The DL tomographic operator predicts velocity models with high statistical and structural accuracy, after being trained with the GAN generated velocity models.

INTRODUCTION

There has been a recent increase in the number of upstream exploration problems that are being addressed with machine learning (ML) applications, including fault prediction (Araya-Polo et al., 2017), facies classification, (Roy et al., 2013), salt bodies detection (Waldeland and Solberg, 2017), digital rock (Waldeland and Solberg, 2017), lithology classification (Xie et al., 2018), and well log analysis (Bestagini et al., 2017). Unfortunately, the earth sciences as a whole suffer from a lack of labeled data, since we seldom know what truly exists beneath the surface. Therefore, supervised ML applications for earth sciences are not reaching their full potential. Also, concerns about generalization and statistical soundness arise when small heavily curated data are used to demonstrate or claim generic solutions. Industries experiencing the most drastic ML revolutions use labeled data sets with millions of samples (Lin et al., 2014). If exploration geophysics expects to replicate similar levels of success, the size and diversity of our labeled datasets must be comparable.

In this work, we leverage Generative Adversarial Networks (GANs) to learn a geologic representation from a finite number of model examples. We then sample from the learned distribution to obtain a large number of unique, geologically feasible models. In this way we mimic the velocity profiles of original models while simultaneously obtaining an abundance of models necessary to fulfill the statistical thoroughness needed for supervised learning problems. We introduce a workflow that uses earth models generated from a GAN to train a DL tomographic operator which reconstructs velocity models from raw seismic data. We aim to create earth models of high enough quality to be used as the precursor to structural imaging algorithms and for geological interpretation.

DEEP LEARNING TOMOGRAPHY

At the forefront of earth model building in exploration seismology is full waveform inversion (FWI). This technique iteratively simulates the seismic experiment and updates the earth model until the simulated seismic data matches the recorded seismic data in a least squares sense (Tarantola, 1984). By fully modeling how energy propagates through the subsurface, FWI is more likely than other methods to find accurate representations of the material properties. Despite this advantage, FWI suffers from many pitfalls. In order to remain stable, FWI must discretize wave propagation very finely in time and space which leads to high computational cost. Furthermore, the iterative computation of earth model updates from data residuals is highly nonlinear and frequently suffers from what is commonly referred to as cycle skipping when optimized with gradient-based inversion methods. FWI is notorious for failing when the initial inversion model is far from the truth. There is a need for an approach that exceeds the accuracy of FWI while simultaneously avoiding its computational complexity and inversion constraints.

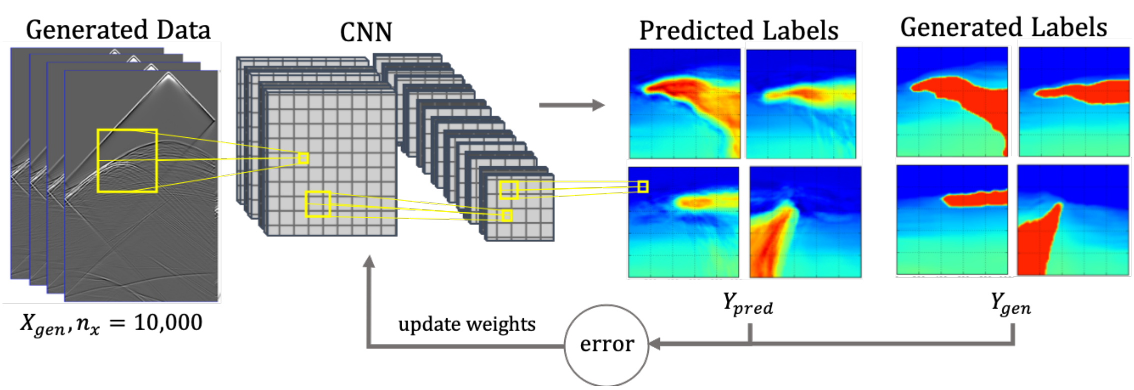

Figure 1: Illustration of the training process for the DL Tomography operator. [**NR**]

DL approaches are showing promising results in the ability to alleviate these restrictions. DL leverages all the signal content in the data for model predictions, frees the problem from incorrect assumptions about the physics of wave propagation, provides key computational advantages over FWI, and is nearly void of human bias. For example, Wang et al. (2018) and Araya-Polo et al. (2018) have each illustrated that convolutional neural networks are fully capable of accurately predicting velocity models with complex overburden profiles directly from seismic data at a fraction of the computational cost of FWI (Farris et al., 2018). While training DL models is notoriously expensive, this expense is incurred once. Then, assuming that the DL model is general enough, the cost of making subsequent model predictions is negligible. Alternatively, the full cost of FWI is incurred every time a new model is required. Here we expand on the CNN used in (Farris et al., 2018) used for mapping from seismic shots to velocity model. The training of our CNN is illustrated in Figure 1.

Despite the advantages over FWI, and subsequently other velocity model building techniques, DL-based applications have not replaced other earth model building techniques within the industry. Farris et al. (2018) notes a few, albeit difficult, remaining restrictions overshadow DL methods. Primarily, the results can still be outperformed by other earth model building approaches if one is willing to pay the computational price. This issue of performance highlights a key problem with supervised learning approaches such as DL, which is a lack of labeled data. The accuracy of DL is directly tied to the quality and quantity of data/label pairings used to train the weights within the network. One may, correctly, assume that the exploration industry has a plethora of seismic data and, therefore, training DL models to map from seismic data to earth models should be achievable. Unfortunately, though, the excess of seismic data that has been acquired by the industry over the last century is all unlabeled since we never actually know the true earth model that produced the data. Even when accurate interpretations are available as labels, they are not standardized between exploration regions, companies, or domain experts. This supervised learning problem is severely limited as we lack the volume of labeled data to build a network that can make accurate, general earth model predictions. Previous works have relied on pseudo-randomly generating on the order of thousands synthetic earth models, simulating seismic surveys in each model, and training DL models with this limited number of data-label pairings Araya-Polo et al. (2018); Farris et al. (2018). The problem remains that the ability to deterministically create a vast number of earth models that are diverse enough to represent a variety of geologic regimes is non-trivial. It would be advantageous if there was a way to create an unlimited number of earth models that are each unique, complex, and representative of realistic geology. If this set of earth models existed, it may be possible to train a DL algorithm to map from seismic data to earth models with a level of accuracy greater than all other physics-based approaches.

COMPLEX MODELS FROM GENERATIVE ADVERSARIAL NETWORKS

It has been shown empirically that GANs are the most successful class of generative models. They have been used by the ML community to generate images of human faces that are virtually indistinguishable from their real counterparts (Creswell et al., 2018), they are a promising approach to image super resolution (Ledig et al., 2016), and can produce high fidelity scientific data (Frid-Adar et al., 2018). Applications in the geophysics community are rapidly emerging. Siahkoohi et al. (2018) successfully used GANs for seismic data reconstruction and Richardson, (2018) suggests that when GANs are coupled with a FWI work flow more plausible results are obtained.

Generative models learn the intrinsic distribution function of the input data $p(x)$ (or $p(x, y)$ if there are multiple targets/classes in the dataset), allowing them to generate both synthetic inputs x and outputs/targets y, given a set of hidden parameters or latent variables z.

GANs are new class of generative models, implemented using two separate networks, that are trained simultaneously (Goodfellow et al., 2014). A Generator network aims at generating data that is indistinguishable from the original training set by learning its underlying probability distribution. It is analogous to a forger, that is continuously trying to convince the second network, the Discriminator, that the data it produces belongs to the original dataset. The Discriminator's job is to classify its input data as either 'real' or 'fake'. It is analogous to a human expert trained to detect fraud.

A neural network $G(z, \theta_1)$ is used to model the Generator. It maps a set of latent input variables z to the desired data space x (say velocity models). A second network $D(x, \theta_2)$ models the Discriminator, a binary classifier that maps the probability that its input came from the original data set. In both cases, we use θ_i to represent the weights or parameters that define each neural network. The Discriminator is trained to maximize $D(x)$, the probability that data x, sampled from the training set, is labeled as 'real'. At the same time, it must label data coming from the Generator as 'fake'. So it is trained by maximizing $D(G(z))$, or equivalently by minimizing $1 - D(G(z))$. We may formally write the generators loss function as (Goodfellow et al., 2014):

$$V(D, G) = log(D(x)) + log(1 - D(G(z))) \tag{1}$$

In practice, we use a log loss function because it heavily penalizes samples that we are confident are mislabeled (Salimans et al., 2016). The generator, on the other hand, is trained to continuously mislead the Discriminator, it is trained to maximize $D(G(z))$, so its lost function is simply $V = log(D(x))$.

Our GAN architecture resembles the one proposed by Radford et al. (2015). We implemented the generator as a series of 2D convolutional transpose operations on the input features, with a hyperbolic tangent function (tanh) activation function in its output layer **?**. The Discriminator uses a series of convolutional layers fully connected to dense output layer with a sigmoid activation function. We found that the most stable and consistent results are obtained when instead of using pooling layers we applied stridden convolutions. Also, it is not possible to obtain reliable results unless a batch normalization (Ioffe and Szegedy, 2015) scheme is applied to the output of the successive layers. Our training set was about 1000 2D slices from a Gulf of Mexico velocity cube that was the culmination of many other velocity estimation techniques. Once trained, our GAN was able to generate velocity models, examples seen in Figure

2, that, to an expert, were indistinguishable from model slices taken from the training set. This is sufficient to conclude our generator successfully learned the probability distribution of the training set and it is able to produce arbitrary velocity models using a small number of input variables.

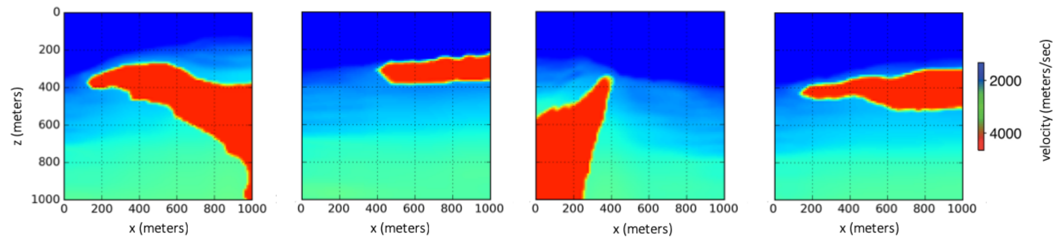

Figure 2: Four examples of unique, complex, and geologically realistic models produced by our GAN. [**NR**]

COMBINING SUPERVISED AND UNSUPERVISED DEEP LEARNING

The combination of the unsupervised learning for earth model generation and supervised learning for earth model prediction from seismic data, can be summarized in Figure 3. First, we train the GAN to generate realistic earth models containing complex salt overburden. This is illustrated in the top left of Figure 3. Once we are satisfied with the output of the GAN, we use it to generate approximately 10,000 earth models. We now move to the top right of Figure 3, in which we use wave propagation code to generate a seismic acquisition over each earth model. When this is complete we have a data/label pairing of sufficient size and geologic complexity to train an effective DL tomography operator. The training of the operator is shown in the bottom of Figure 3.

Once complete, a fully trained tomography operator is ready to be used for predictions of new velocity models when exposed to input data. To a certain extent, this workflow shares concepts with FWI. For instance wave equation modeling happens in both approaches, and both at their core solve an optimization problem. FWI community recognize the usefulness of some concepts developed in the DL community, for instance using similar loss functions, such as Wasserstein (Sun et al., 2019; Yang et al., 2016) which was first used in the DL community as a loss function in Frogner

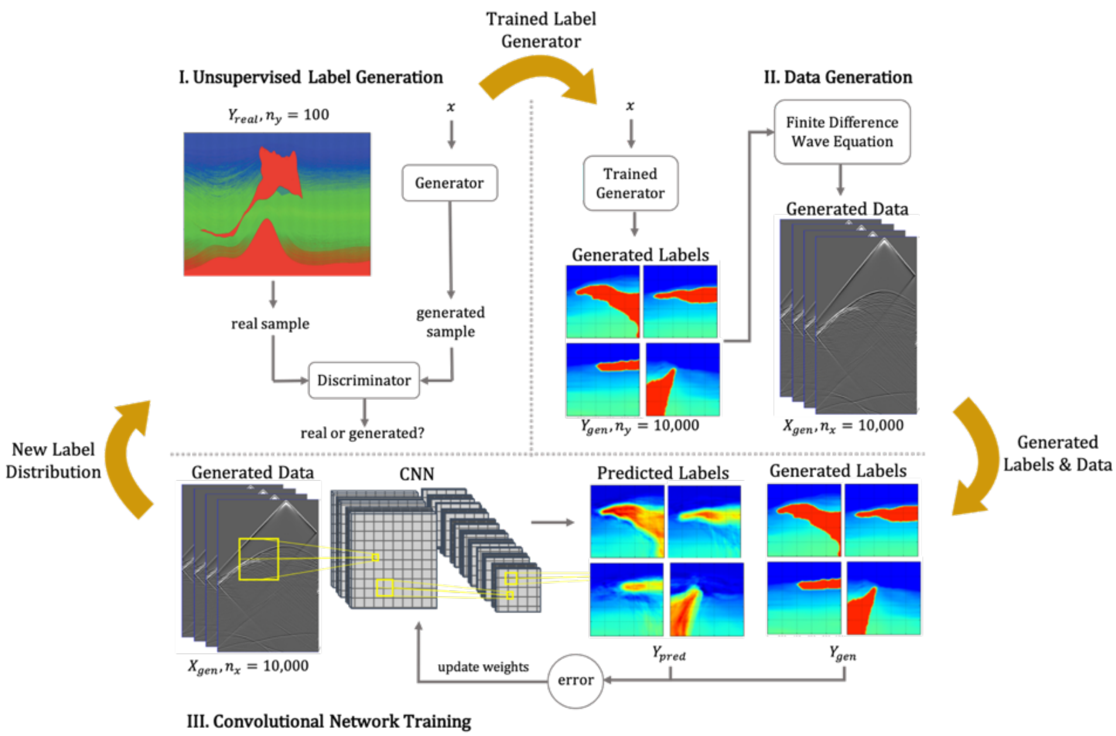

Figure 3: The combined workflow of unsupervised and supervised DL culminating in an accurate Tomography operator that maps from seismic shots to 2D velocity models. [**NR**]

et al. (2015) and then gained wide attention being integrated in GAN architectures (Gulrajani et al., 2017). The connection between these two techniques does not stop here, for instance, as demonstrated in Farris et al. (2018), models predicted with DL tomography can be used as advanced starting models for FWI, thus reducing optimization cost. We foresee more efforts from the community along the direction of combining these two methods, either DL being used as preconditioner or directly influencing FWI misfits.

CONCLUSION

By leveraging the power of GANs, we are able to fulfill our DL-based tomography workflow and reach impressive quantitative and qualitative results with trained tomography operator. This workflow gives us a strong foundation from which we can pursue more ambitious DL challenges, one of which is generalization to any possible geologic regime. To do this, we need GANs that can learn the underlaying statistical distribution of earth models with any possible structure, shape, and overburden trend. Furthermore, we need the ability to produce higher resolution models, with the possible extension into three dimensions. Overall, we believe GANs are a useful tool that helps to avoid issues with labeled data generation and quick turnaround of workflow based on Deep Learning in geosciences.

ACKNOWLEDGEMENT

This work was conducted at the Shell Technology Center in Houston, Texas. Many thanks to the individuals at Shell E&P who made this work possible and to the SEP sponsors for their continued funding.

REFERENCES

Araya-Polo, M., T. Dahlke, C. Frogner, C. Zhang, T. Poggio, and D. Hohl, 2017, Automated fault detection without seismic processing: The Leading Edge, **36**, 208–214.

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

Bestagini, P., V. Lipari, and S. Tubaro, 2017, A machine learning approach to facies classification using well logs: SEG Technical Program Expanded Abstracts 2017, Society of Exploration Geophysicists, 2137–2142.

Creswell, A., T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, 2018, Generative Adversarial Networks: An Overview: IEEE Signal Processing Magazine, **35**, 53–65.

Farris, S., M. Araya-Polo, J. Jennings, B. Clapp, and B. Biondi, 2018, Tomography: a deep learning vs full-waveform inversion comparison: EAGE Workshop on High Performance Computing.

Frid-Adar, M., E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, 2018, Synthetic Data Augmentation using GAN for Improved Liver Lesion Classification.

Frogner, C., C. Zhang, H. Mobahi, M. Araya, and T. A. Poggio, 2015, Learning with a Wasserstein Loss.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, 2014, Generative Adversarial Nets.

Gulrajani, I., F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, 2017, Improved Training of Wasserstein GANs.

Ioffe, S., and C. Szegedy, 2015, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.

Ledig, C., L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, 2016, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.

Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, 2014, Microsoft COCO: Common Objects in Context: Springer, Cham, 740–755.

Radford, A., L. Metz, and S. Chintala, 2015, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks.

Roy, A., V. Jayaram, and K. J. Marfurt, 2013, Active learning algorithms in seismic facies classification.

Salimans, T., I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, 2016, Improved Techniques for Training GANs.

Siahkoohi, A., R. Kumar, and F. Herrmann, 2018, Seismic Data Reconstruction with Generative Adversarial Networks: Presented at the .

Sun, B., , and Tariq, Alkhalifah1, 2019, A robust waveform inversion using a global comparison of modeled and observed data: The Leading Edge.

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

Waldeland, A., and A. Solberg, 2017, Salt Classification Using Deep Learning: Presented at the .

Wang, W., F. Yang, and J. Ma, 2018, Velocity model building with a modified fully convolutional network: SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 2086–2090.

Xie, Y., C. Zhu, W. Zhou, Z. Li, X. Liu, and M. Tu, 2018, Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances: Journal of Petroleum Science and Engineering, **160**, 182–193.

Yang, Y., B. Engquist, J. Sun, and B. D. Froese, 2016, Application of Optimal Transport and the Quadratic Wasserstein Metric to Full-Waveform Inversion.