Deep Learning for Making Sense of Ambient Seismic Noise

Julien Clancy[1], Laurent Demanet[1], Jonathan Helland[2], Zongbo Xu[3]
MIT Department of Mathematics [1]
Boise State University Department of Geosciences[2]
Colorado School of Mines Department of Applied Mathematics[3]
julienc@mit.edu, laurent@math.mit.edu,
jhelland@mymail.mines.edu, zongboxu@u.boisestate.edu

SUMMARY

We apply recent advances in deep neural networks to three classes of geophysical problems stemming from ambient noise imaging: wavespeed inversion in homogeneous media in the presence of anisotropic sources, local wavespeed inversion in inhomogeneous media, and source directionality estimation in homogeneous media. Our networks are inspired by those commonly in the signal processing literature, such as convolutional networks and LSTMs, but use a training procedure that appears unique to physical problems: data is generated on the fly and only used to compute a single gradient, then discarded and never seen again. These techniques prove to be highly performant and quite flexible — they easily accomodate for data gathered from different sensor geometries, or for different priors in the data generating procedure. We also find preliminary evidence that, in simplified analogues of these problems, the nodes in deeper layers of our networks are computing physically meaningful quantities.

INTRODUCTION

Starting early in this decade, deep neural networks — most notably convolutional networks (ConvNets) (Lecun et al 1998), (Krizhevsky et al 2017), generative adversarial networks (GANs) (Goodfellow at al 2014), and long short-term memory networks (LSTMs) (Hochreiter and Schmidhuber 1997) — have taken over many areas of traditional signal processing, from computer vision to language translation, and have even trounced the best human players the board game Go (Silver et al 2016). The contribution of this paper is to bring these methods to bear on certain inversion problems in geophysics. While neural networks have been used in geophysics before, see e.g. (Asadi et al 2017), we believe we are the first to apply them to the inversion problems at hand. Further, our architectures leverage recent powerful innovations in the deep learning literature, and we see strong improvements in performance relative to "vanilla" artificial neural networks.

In this paper we study three problems:

- Wavespeed inversion in homogeneous media, when sources are anisotropic
- Wavespeed inversion in inhomogeneous media
- Recovery of the anisotropy of sources in homogeneous media, represented both by parameters in some predefined model and by pointwise intensity

In all three cases our neural networks achieve excellent performance, even in the presence of strong noise. We additionally study the toy problem of aligning two noisy signals, for which the cross-correlation is well-suited, and observe that neural networks trained to solve this problem appear to be computing something akin to the cross-correlation in their deep layers.

EXPERIMENTS, NETWORKS, AND TRAINING

All of our exprimentds were on synthetic data, in both noisy (with noise energy up to approximately up to 10% of the signal energy) and noiseless settings. We used three networks to achieve our results: a locally connected architecture, a so-called "relational network" (Santoro et al 2017), and convolutional LSTMs (Hochreiter and Schmidhuber 1997). These are shown below.

In the homogeneous case, we gathered data from two or three sensors located (randomly) near the origin, and generated Gaussian noise sources in the far-field (either uniformly or with some anisitropy in the intensity, depending on the experiment) and propagated them to the sensors with the Green's function. Combined with the simulated discrete sampling rate this only requires a simple FFT, so that we could generate new data on the fly for each gradient calculation. This means there is no distinction between training and testing data; the network only every saw any data point once. Since most theoretical guarantees for stochastic gradient descent only apply in this "true" setting, the success of networks in this setting indicate that they are doing something nontrivial (and are most definitely not memorizing data). The state of the art in a simpler setting — with isotropic noise — is known to be the cross-correlation. Specifically, the peak of the cross-correlation between two sensors gives the traveltime between them.

It is not practical to compute the inhomogeneous Green's function (and there is no guarantee that it could be quickly applied) so we used a direct solver instead. Our data generation process was otherwise similar, except that we sampled on a square grid of 2500 sensors to maximize the data gathered from each simulation. We then used the readings of four sensors in a cross shape, over a limited time window, as input to our networks. Out of twenty simulations, we usde fifteen for training and left five for testing, as well as the final 10% of each time trace at every sensor in every simulation. Both testing sets yielded similar results, indicating that the networks did not memorize data either from any simulation or from any particular sensor. Our experimental setups are shown below.

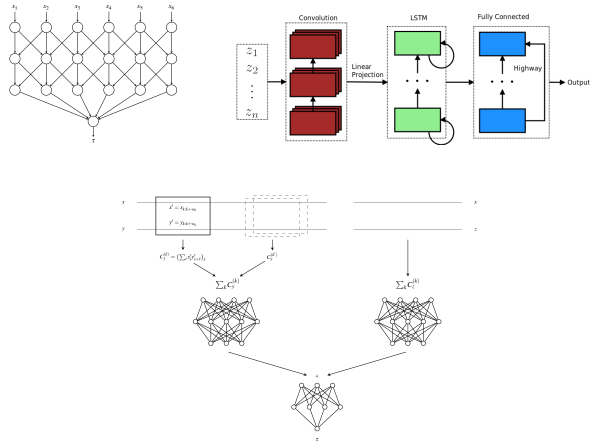

Figure 1: Examples of the networks we used. Cartoons of a locally connected architecture, a CLSTM, and a relational architecture with cross-correlations as its input.

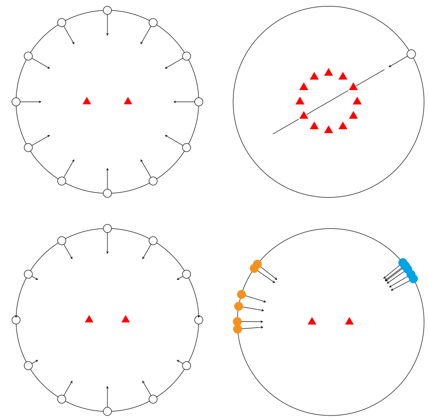

Figure 2: Examples of types of experiments where one would like to recover the underlying (homogeneous) wavespeed. They differ in the number of sensors and the assumption on the directionality of the noise. Top two: experiments amenable to the cross-correlation. Bottom two: experiments amenable to our methods.

RESULTS

Our results are best summarized in figures. The networks performed extremely well, notably in the case of two sensors. Considering the problem of inverting for wavespeed from directional sources, with three sensors, classical cross-correlations can achieve up to $1-\cos\pi/3 \approx 0.133$ relative error, while with just two they can achieve no approximation — for a wave coming in orthogonally to the line between the two sensors the peak of the cross-correlation will be at 0, and for small angles the situation is nearly as dire. In contrast, our neural networks were able to reliably invert for wavespeed *independently of the source distribution* using just two sensors.

We had similar success in recovering the source intensities themselves. Using two sensors we can never expect determine whether a wave came from angle θ or $-\theta$, but up to this sign ambiguity we were able to recover the source distribution. The use of a third sensor resolves this ambiguity in theory and in practice.

Our results in the inhomogeneous case were also quite good: we were able to reliably invert for wavespeed for velocity profiles that the network had never seen before. Notice that that the network only saw fifteen models and our model space is two-dimensional (with velocity of the form $ax+by+c$); this means that the network cannot be learning the model space, but must have identified something more basic about the relationship between the traces and the model.

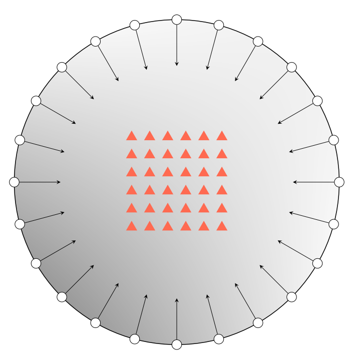

Figure 3: The experimental setup for inhomogeneous wavespeed recovery. The background shade represents the wave velocity. Our experiments used a dense grid of sensors for efficiency of generating a data set; we only fed the network four traces at a time, organized in a cross-shape.

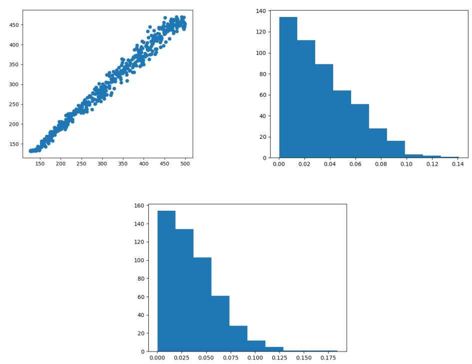

Figure 4: Top left: true (x-axis) vs. predicted (y-axis) wavespeed for the two-sensor case under the random wedges distribution. Top right: distribution of the relative absolute errors $|\hat{x} - x|/x$. Bottom: distribution of relative absolute error for a network trained on the random wedge model but tested on single sources located at angles $\pi/2$ or $3\pi/2$.

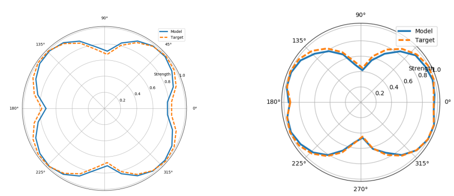

Figure 5: Examples from the test set, never seen by the network, of the source strength recovery problem. The dataset was composed of intensities taking the form $a \sin\theta + b \cos\theta$, where θ is the angle from which the wave originates (in the far field).

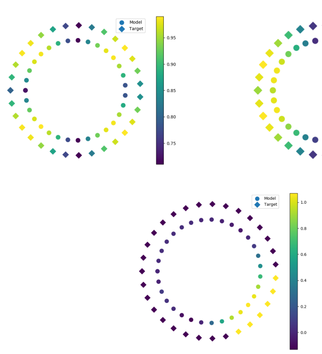

Figure 6: More strength recovery experiments, from a network trained on a mixture of the two models. Top: intensities of the form $a \sin\theta + b \cos\theta$. Bottom: a wedge-type intensity.

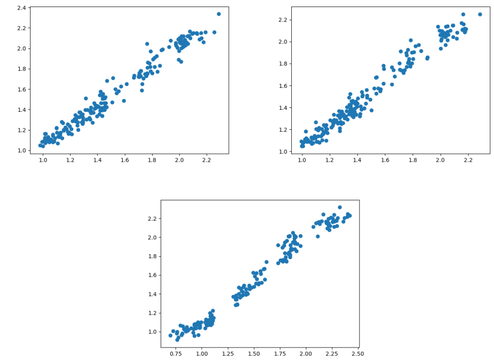

Figure 7: Results for the inhomogeneous case. Top left: predicted wavespeed against true wavespeed, on the training set. Top right: predicted wavspeed against true wavespeed, on the unseen 10000 time samples from the simulations that were used for training. Bottom: predicted wavespeed against true wavespeed on unseen simulations.

WHAT'S GOING ON?

The most important open question in the field of deep learning is arguably what networks are learning, and whether it is more meaningful than "finding a local minimum". The most satisfactory answers exist in the case of convolutional networks. On the one hand, community folklore holds that examining the filters in a convolutional network trained to classify images shows that the early layers learn to identify edges and curves, the next ones identify simple shapes like circles or rectangles, and so on; deeper layers compose the results of earlier ones these to recognize objects built out of simpler parts, in the way that faces or bicycles are composed of lines and circles. No such natural interpretability exists for other types of networks, as far as we know. [1]

Our attempt to answer this question for the networks we use is practical. The simplest version of the problems we address in this paper is to align two noisy signals: given $y = \tau_k(x) + \varepsilon$, where ε is noise and τ_k is the circular shift by k coordinates, how do we recover k? The cross-correlation peak is again the correct answer. We trained a very simple deep neural network to solve this inverse problem; a problem instance and a cartoon of the network we uesd are shown below. It achieved error rates competitive with cross-correlation, which is unremarkable, but when we regressed each of its inner nodes on the degree-2 monomial dictionary, we found that the network appeared to be computing some version of the coordinates of the cross-correlation. This remarkable result indicates that, in some sense, the simplest way to solve this problem is to compute the cross-correlation and then take the pointwise maximum. We found manifestations of this phenomenon in several completely different problems; to compute $x \cdot y$, for instance, our trained networks appeared to be computing $(x+y)^2$ and $(x-y)^2$ and subtracting them.

[1] It is worth noting that on the theoretical front, a recent line of work (Bruna and Mallat 2012, Mallat 2012) has shown that the general structure of iterating convolutions and nonlinearities automatically "linearizes" *all* diffeomorphisms, so that the resulting networks are naturally excellent image/sound classifiers.

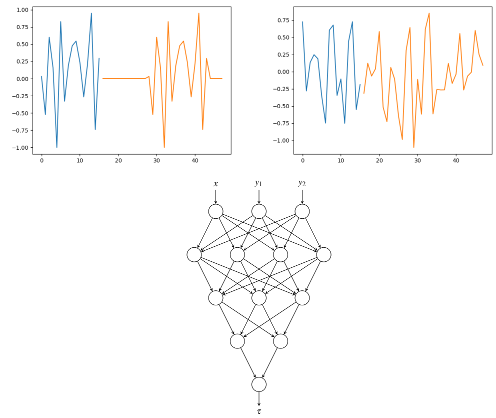

Figure 8: An instance of the problem of aligning noisy signals, and an example network that we used to solve it. The output of the network should be the amount that the signals are shifted by. Left: noiseless instance. Center: noisy instance; the reference signal is in blue, and the signal to be aligned is in orange. Right: the network we used.

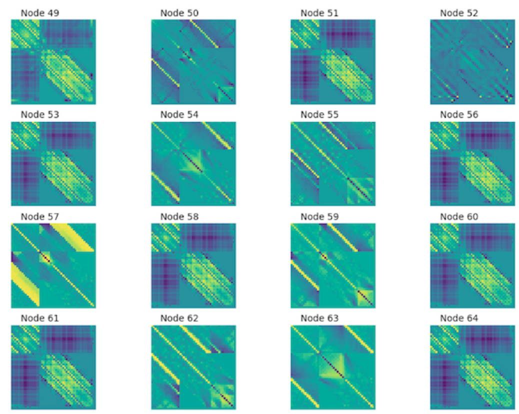

Figure 9: The regressed coefficients in the degree-2 dictionary for several nodes in the interior of the network. We show the coefficients of the monomials $x_i x_j$, where the two indices index the matrix of coefficients. The cross-correlation has entries $\sum x_i x_{i+\ell}$, which corresponds to an active offset diagonal in the matrix; note the similarity of the displayed coefficients to a shifted diagonal. In particular, nodes 50, 52, 54, 55, 59, 62, and 63 exhibit strong cross-correlation behavior. this is just one slice of all the nodes; at this particular level there are 128, and the others contain the other entries of the cross-correlation, are duplicates of the ones shown here, or are spurious. It is common for neural networks to duplicate or "deactivate" nodes.

FUTURE WORK

The first next step is to apply our techniques to real data; work on this is already underway. We would also like to improve our results for inhomogeneous wavespeed recovery, and have seen indications that this is well within reach. This points towards using neural networks for full-waveform inversion, though we have not yet touched the issue of detecting discontinuities in the wavespeed. However, neural networks have proven effective at sparse recovery, competitive with ℓ_1 minimization (Baraniuk and Mousavi, 2017), which supports the idea that they are capable of solving much more complex inverse problems.

An orthogonal direction of future progress would be to use our procedure of dictionary fitting to extract formulas or laws from trained networks. Since we treat the outputs of each node as functions, with no reference to their position in the network, this could provide an architecture-independent way to interpret what neural networks have learned (as opposed to the heavily architecture-dependent effect of convolutional networks building multiscale filterbanks for natural image processing).

ACKNOWLEDGEMENTS

This work was supported in part by AFOSR grant FA9550-17-1-0316, and NSF grant DMS-1255203.

REFERENCES

1. Y. Lecun, L. Bottou, Y Bengio, and P. Haffner. *Gradient-based learning applied to document recognition*. 1998.
2. A. Krizhevsky, I. Sutskever, and G. Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. 2017.
3. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, Sherjil, A. Courville, and Y. Bengio. *Generative Adversarial Nets*. 2014.
4. S. Hochreiter and J. Schmidhuber. *Long Short-Term Memory*. 1997.
5. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G v.d. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. *Mastering the game of Go with deep neural networks and tree search*. 2016.
6. A. Asadi, P. Dietrich, and H. Paasche. *Spatially continuous probabilistic prediction of sparsely measured ground properties constrained by ill-posed tomographic imaging considering data uncertainty and resolution*. 2017.
7. A. Santoro, D. Raposo, D. G. T. Barrett, M. Malinowski, R. Pascanu, P. W. Battaglia, and T. P. Lillicrap. *A simple neural network module for relational reasoning*. 2017.
8. R. Baraniuk and A. Mousavi. *Learning To Invert: Signal Recovery Via Deep Convolutional Networks*. 2017.
9. S. Mallat. *Group Invariant Scattering*. 2012.
10. J. Bruna and S. Mallat. *Invariant Scattering Convolutional Networks*. 2012.