Convolutional Autoencoders: Application in denoising of geophysical data

Deepak K. Gupta

November 17, 2018

1 Domain Background

The domain of geophysics deals with data acquired through several physics-based experiments. Several teams of geophysicicsts work together on to clean and process the acquired data before it anything can be interpreted out of it. The traditional processing steps are extremely cumbersome, and among these, removal of noise is a very challenging task. The geological complexities inside the earth make the data so heterogeneous that it is extemely difficult to differentiate the noise from the signal component of the data. Moreover, denoising is a cumbersome task, and with traditional techniques, a significant amount of time needs to be spent on it.

The application of various machine learning techniques has been explored to automate the process of seismic data denoising. A direction of research along this line is to investigate the application of autoencoders on such problems. Autoencoders are a variant of the traditional machine learning methodology, where the input and output of the network are set to be the same [1]. During training, the network is thus forced to learn an internal representation of the input data in a different space. Recently, application of autoencoders has also been explored on geophysical data [6, 8].

In a recent work, the application of deep fully-connected autoencoders has been explored for the removal of noise and reconstruction of signal in geophysical data [2]. Remarkable reduction in noise could be achieved, which makes autoencoders a potential denoising tool. Although the reduction in noise is significant, there exist further possibilities of improvements. Moreover, due to the computational costs associated with the fully-connected layers, they are not so favorable for large datasets.

In the past two years, convolutional neural networks (CNN), a variant of the traditional neural networks, have proved to be enormously powerful tool for classification problems. During this course, the discussion on CNN has provoked my interest to explore its application for the purpose of denoising geopysical data. Some recent applications of CNN on geophysical problems are identification of geological features on seismic data [9], automated velocity model building [7], among others.

Convolutional autoencoders, a form of CNNs can be used or denoising. Convolutional autoencoders could possibly help to overcome the drawbacks of traditional autoencoders in the context of denoising. In particular, it is believed that the spatial pattern of the signal can be preserved, and noise could be removed at significantly lower computational costs. Thus, in this project, the application of convolutional autoencoders will be explored for removal of noise from synthetic and real geophysical images.

2 Problem Statement

The main objective of this project will be to explore the application of convolutional autoencoders for the reduction of noise in geophysical images. The focus of this project will be to explore the reduction of incoherent noise in particular. The test problems will include synthetic images containing random, normal and salt-pepper noise as well as real geophysical datasets with added noise.

3 Datasets and Inputs

For the problem stated above, the required input are synthetic and real geophysical images convoluted with noise as well as the noisefree versions. Some synthetic non-seismic test cases, including the published research paper and computer code can be accessed at [5, 4]. Further real datasets are openly available at [3], some of which together with additive noise will be used in this study. Smaller images will be cropped from the huge datasets available on these links for this study. Overall, most of the data will be synthesized or augmented based on existing sparse real data.

4 Solution Statement

As stated above, the problem that is considered in this project is that of denoising seismic data, and the proposed solution is 'Convolutional Autoencoders'. The two aspects to be considered are that denoising process needs to be automated and computational costs associated with the task should be reasonable. Moreover, the denoising process has to be scalable. All of these problems can hopefully be solved using convolutional autoencoders.

Convolutional autoencoder takes a noisy image as an input and can map it to a space of different dimensions through a series of convolution and pooling operations. Keeping the dimension of the projected space low, it is possible to preserve only a certain set of features of the input image and omit the rest. Assuming that the noise and signal components of an image contain different feature-components, the goal is to train the convolutional network to be able to preserve the signal part and forget the noise. Once the signal part is projected into the lower-dimensional space, the autoencoder learns to reconstruct the signal back from this dimension. This is how an autoencoder will be trained to forget noise.

5 Benchmark Model

For comparison purpose, the synthetic datasets (where the target is known) will also be analyzed using the traditional filtering techniques such as variational filtering, bilateral filtering and wavelet filtering [5]. These results will be the baseline model. Further, fully connected autoencoders of similar size will also be used for comparison purpose.

6 Evaluation Metrics

The problems to be studied here can be considered as regression problems in the sense that a good projection from noisy domain to noisefree domain needs to be learned. Thus, root mean evaluation metric would suffice. In particular L_1 and L_2 norms will be studied as well as the standard deviations will be analyzed.

7 Project Design

Although discussed in pieces in the above sections, a design methodology for the project is discussed here. The different steps to be taken in this project are:

1. Data preparation: This is an important and crucial step. First, several simple 2D synthetic images, making sense in the context of geophysics, will be generated. Different types of noise including random, normal and salt-pepper type will be added to them and noisy images will be generated through augmentation. Further, the real seismic dataset will be broken into smaller seismic images. From these, noisy seismic images will be generated and added to the datasets.

2. Data split: Data will be split for training. Unlike random split, here, a significant part of the input data used in training will be noisefree and only a certain fraction (less than 20%) will contain noise. The reason is that during training, the model needs to learn to identify patterns of signal and noisefree data is allowed to dominate the training set. On the contrary, most of the test examples will be noisy,

since our main goal is to remove noise from them. However, there will also be cases of false positives, where a noisefree signal becomes noisy. The model will be trained to avoid it.

3. Evaluation metric: Before beginning any experiments, the metric needs to be chosen. Details on this aspect have been provided above.

4. Baseline model: Next step is to validate the performance of the baseline model to establish the reference. Methods such as variational fitering, wavelet filtering and bilateral approach are employed and improvement is recorded. Further, shallow and deep fully-autoencoders are used and their performance as well is recorded. Together, these outputs are the reference and the trained convolutional autoencoder needs to perform better than this.

5. The convolutional autoenoder will be implemented using python, Tensorflow and keras and a deep variant will be tested against the baseline models. Hyperparameters will be tuned with the goal of outperforming the other methods. The results will be reported.

6. Results will be analyzed, drawbacks will be identified and final conclusions will close the project.

References

[1] Y. Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.

[2] D. Bhowmick, D. K. Gupta, S. Maiti, and U. Shankar. Stacked autoencoders based machine learning for noise reduction and signal reconstruction in geophysical data. *Neural Computing and Applications (under review)*, 2019.

[3] SEG Open data. http://software.seg.org/datasets/2d/index.html.

[4] G. Ganssle. https://github.com/gganssle/cseg-imXlate.

[5] G. Ganssle. Denoising seismic records with image translational networks. *CSEG RECORDER*, 43(1):32–34, 2018.

[6] A. P. Valentine, L. M. Kalnins, and J. Trampert. Discovery and analysis of topographic features using learning algorithms: A seamount case study. *Geophys. Res. Lett.*, 40(12):3048–3054, 2013.

[7] Y. Wu, Y. Lin, and Z. Zhou. InversionNet: Accurate and efficient seismic waveform inversion with convolutional neural networks. In *SEG Technical Program Expanded Abstracts*, pages 2096–2100, 2018.

[8] Y. Xiong and R. Zuo. Recognition of geochemical anomalies using a deep autoencoder network. *Comput. Geosci.*, 86:75–82, 2016.

[9] T. Zhao and P. Mukopadhyay. A fault detection workflow using deep learning and image processing. In *SEG Technical Program Expanded Abstracts*, pages 1966–1970, 2018.