A Case Study into GAN-Based Data Augmentation: Enhancing Seismic Predictive Models

Paul Wolfe
Department of Computer Science
University of Mary Hardin-Baylor
pwolfe@umhb.edu

Abstract

This case study explores the novel application of Generative Adversarial Networks (GANs) for seismic data augmentation to improve predictive model performance. Focusing on the enhancement of arrival-time picking using PhaseNet, this work demonstrates how synthetically generated seismic data can mitigate the scarcity of significant seismic events. Our research evaluates the effects of GAN-augmented data on various machine learning performance metrics and considers its potential to transform seismological practices.

1. Introduction

Seismology plays a dual role in our understanding of Earth: scientific exploration and societal protection. Traditionally reliant on physical sensors like seismographs [1], the field faces a challenge—significant seismic events are rare and hard to capture consistently. This scarcity limits the data available for training accurate predictive models.

This study investigates whether synthetic seismic data, generated via GANs, can bridge that gap. The objective is to augment real-world seismic datasets and assess their impact on PhaseNet [4], a deep neural network designed for arrival-time picking of P and S waves.

2. Data Collection and Preprocessing

Our dataset originates from the United States Geological Survey (USGS) and is filtered for earthquakes of magnitude 2.5 or higher within the last six months.

Preprocessing Steps:

- Outlier removal and anomaly detection.
- Normalization of numerical attributes.
- Splitting into training, validation, and test sets.

3. GAN Architecture and Data Augmentation

A standard GAN framework is deployed, consisting of a generator and a discriminator in adversarial training. The goal is to produce realistic synthetic seismic events conditioned on the properties of real data.

Outcomes:

- Augmented dataset exhibits improved variance and volume.
- Synthetic data demonstrates plausible waveform characteristics.

4. Integration with PhaseNet

PhaseNet [4] is a neural network trained to detect seismic wave arrival times. We tested PhaseNet on both the original and GAN-augmented datasets to evaluate performance differences.

Key Metrics Used:

- Precision, Recall, and F1-Score.
- Mean Absolute Error (MAE).
- False Alarm Rate and Detection Accuracy.

5. Findings and Observations

GAN Performance:

- Generated data improved model generalization during testing.
- Augmented training yielded higher accuracy on unseen seismic patterns.

PhaseNet Improvements:

- F1-Score increased by up to 7% with synthetic augmentation.
- Lower false positives in noisy environments.

6. Conclusion and Future Work

The preliminary results suggest that GANs can be effective in augmenting seismic datasets, improving the training base for neural predictors like PhaseNet. Future directions include:

- Experimenting with Conditional GANs for more tailored seismic scenarios.
- Applying this framework to early-warning systems in seismic risk zones.
- Expanding testing to global datasets beyond USGS.

References

[1] Shearer, P. M. (2009). *Introduction to seismology*. Cambridge University Press.

[2] Lay, T., & Wallace, T. C. (1995). *Modern global seismology*. Academic Press.

[3] Allen, R. M., Gasparini, P., Kamigaichi, O., & Bose, M. (2009). The status of earthquake early warning around the world: An introductory overview. *Seismological Research Letters*, 80(5), 682–693.

[4] Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1), 261–273.