Submission code: 6L9Z0Q7GBA

See this and other abstracts on our website: https://home.sbgf.org.br/Pages/resumos.php

Generate Seismic Synthetic Data for Training Neural Networks

Igor Guimarães Barros (Oxaala Tecnologias), Vitor Pereira Fontes (Federal University of the São Francisco Valley (UNIVASF)), Alysson de Jesus Alves (Federal University of the São Francisco Valley (UNIVASF)), João Pedro Silva Gomes (Federal University of Bahia - GAIA), Gabriel Vergne (Federal University of Bahia - GAIA), Paulo Barros (Oxaala Tecnologias), José Luiz Silva (Oxaala Tecnologias), ALEXSANDRO CERQUEIRA (Federal University of Bahia - GAIA)

Generate Seismic Synthetic Data for Training Neural Networks

Introduction

The development of robust machine learning models for seismic interpretation relies heavily on high-quality training data. However, acquiring and labeling real-world seismic datasets is often costly and time-consuming. Synthetic data generation addresses this challenge by creating computationally simulated datasets that mimic real geological structures and seismic responses. This study presents a modular framework for generating synthetic seismic data tailored to train convolutional neural networks (CNNs) in tasks such as horizon interpolation and fault detection. The system produces configurable datasets with controlled noise levels, diverse geological features, and standardized formats, enabling efficient model training and validation.

Method and/or Theory

The synthetic data generation methodology employs a systematic workflow to create realistic seismic datasets for machine learning applications. First, velocity and density models are constructed to simulate subsurface geology, incorporating adjustable parameters for layer thickness, acoustic impedance, and reflectivity. Seismic wavelets (as Ricker) with configurable central frequencies (10–50 Hz) are convolved with synthetic reflectivity to represent seismic sources. Faults are simulated using algorithms that introduce normal, reverse, and complex deformations, with customizable dip, throw, and strike angles. Controlled noise (Gaussian, speckle) is injected at user-defined signal-to-noise ratios (SNR) to mimic field data artifacts. Finally, the data is normalized to amplitudes between -1 and 1 and exported in standardized formats like NetCDF or SEG-Y, complete with metadata for geometry and coordinates.

The framework supports batch generation of 3D seismic cubes (e.g., 128×128×128 voxels) and binary fault masks, optimized for compatibility with deep learning pipelines. Each synthetic volume undergoes validation through statistical metrics, ensuring alignment with real seismic data characteristics. The system integrates planar and sinusoidal deformations to enhance geological complexity, while configurable noise levels enable robustness testing of neural networks. Outputs include not only seismic amplitudes and fault probabilities but also comprehensive metadata for reproducibility. This modular approach bridges the gap between limited field data and the growing demand for diverse, AI-ready training datasets in geophysics.

Results and Conclusions

The synthetic data generation module produced realistic 3D seismic cubes and fault masks with configurable parameters, including variable layer geometries, fault throws (5–50 ms), and adjustable noise levels (2–20 dB SNR). Validation confirmed the data's statistical alignment with real seismic characteristics, while maintaining computational efficiency. Benchmarks demonstrated a 15–20% improvement in CNN-based fault detection accuracy when augmenting training with synthetic data. The standardized metadata enabled seamless integration with machine learning pipelines and ensured reproducibility. This approach effectively addresses the scarcity of labeled field data for AI training, offering scalable solutions for diverse geological scenarios. Future enhancements will incorporate more complex features like salt bodies to further expand its applicability in seismic interpretation.