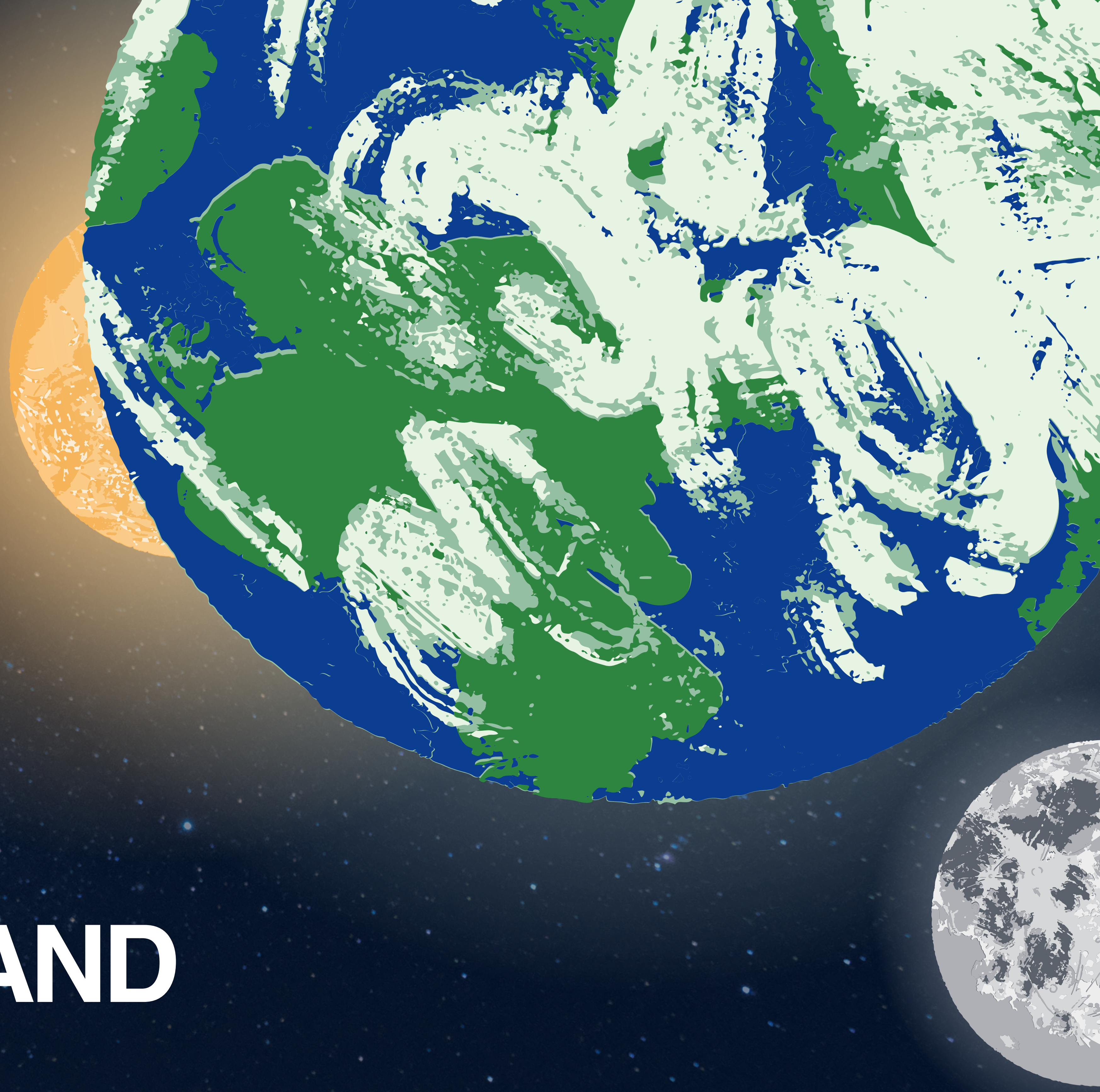
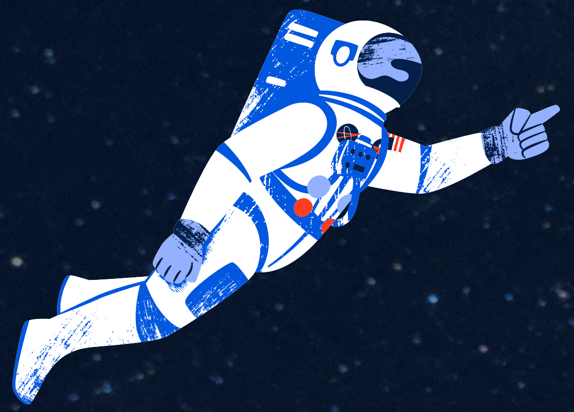


Nasa Space Apps Challenge

INNOVATIVE SEISMIC DETECTION WITH CNN AND ENSEMBLE LEARNING

Team: Ana Carolina Correia, Elisceu Cordeiro de Brito, Rafael da Silva Barros, Victor Amarante e Vitor Negromonte

Introduction



About our project

Our project presents an innovative framework for seismic activity detection using advanced machine learning, specifically utilizing Convolutional Neural Networks (CNNs) to analyze time-series seismic data and detect patterns of seismic quakes. By incorporating Taylor's expansion, the model's ability to predict complex, non-linear patterns is enhanced. Additionally, we use ensemble learning to combine predictions from multiple models, improving accuracy and robustness, making the detection system more reliable across different types of seismic data, from minor tremors to major quakes.

Goals

We aim to develop a seismic detection system that can accurately predict seismic activity both on Earth and other planetary bodies, such as the Moon or Mars. Our ultimate goal is to assist in early earthquake detection, contribute to planetary research, and enable scientists to better understand seismic activity in challenging environments.

How it works?

1

- Seismic time-series data is preprocessed using advanced filtering techniques (such as high-pass filters) and wavelet denoising to reduce noise and enhance signal clarity. The data consists of both ground velocity measurements and related metadata, which are passed into the model.

2

- The preprocessed time-series data is fed into a CNN model designed to detect key features in the seismic waveforms. The CNN layers help extract high-level representations from the raw seismic signals. We developed a custom architecture based in a mix of GoogLeNet Inception blocks and residual relations based on Taylor's expansion and ResNet model.

3

- To further enhance the model's ability to approximate complex waveforms, we apply techniques inspired by Taylor's expansion, allowing us to better model non-linear relationships in the seismic data. This results in improved precision for identifying seismic events.

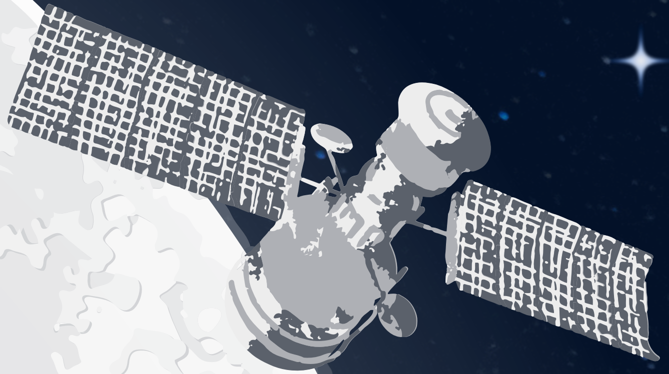
How it works?

4

- Multiple CNN models are trained using varied parameters and architectures. These models are then combined in an ensemble, where the outputs are aggregated to produce a final prediction. This approach improves the robustness of the system, reducing the likelihood of false positives and ensuring accurate detection across a wide range of seismic event types.

5

- After training the models on initial datasets (such as Caltech seismic data), we fine-tune the models using more specialized datasets, such as NASA lunar seismic data, ensuring the model generalizes well to diverse environments.



Benefits

1. **Higher Accuracy:** The combined use of CNNs and ensemble learning improves detection accuracy.

2. **Adaptability**: The models, trained on a large dataset from the California Institute of Technology, can be fine-tuned for various environments.

3. **Noise resilience:** Advanced denoising techniques and Taylor-inspired expansions enable the system to handle noisy real-world data.

Tools

Software: Python, PyTorch, HDF5, Optuna, SciPy, PyWavelets, Matplotlib, Seaborn
Hardware: NVIDIA Tesla T4 16Gb

Use of Artificial Intelligence

For helping in the project structure, we used GPT 4o and Mistral Large 2.
For coding we used GitHub Copilot.

Dataset

ConvNetQuake is trained on data from Oklahoma (USA):

- **streams**: 2.5 years of monthly streams from GSOK029 and GSOK027 in .mseed
- **catalogs:** earthquake catalogs the OGS (years 2014 to 2016) and from Benz et al. 2015 (Feb. to Sept. 2014)
- **6_clusters**: windows used for training and testing of ConvNetQuake with 6 geographic areas
- **50_clusters**: windows used for training and testing of ConvNetQuake with 50 geographic areas
- **known_template:** template T_1 used to generate synthetic data
- **unknown_template:** template T_2 used to generate synthetic data
- **synth:** directory of synthetic data for testing