Seismic Detection Across the Solar System: Spectrogram Analysis Using Machine Learning Models

Daniel Zárate (danzarate902@gmail.com), Karen Mosquera (est.karend.mosquera@unimilitar.edu.co), y Humberto Fierro (betfie1@gmail.com)

Summary - The project aims to explore the detection and classification of seismic events in data collected by NASA's InSight instrument on Mars and by the Apollo mission on the Moon. To achieve this, the seismic signals are converted into spectrogram images, which are then processed and used as input for an object detection model, specifically a YOLO-based algorithm. The use of spectrograms allows the visualization of the energy distribution of signals across different frequencies and times, facilitating the identification of characteristic patterns of earthquakes and related events. By leveraging this approach, the project not only enhances the detection of seismic activity but also provides a comparative analysis between Martian and Lunar data, enabling a deeper understanding of seismic behavior in different planetary environments. This research contributes to the development of tools and techniques for planetary exploration and seismic monitoring beyond Earth.

I. INTRODUCTION

The analysis of planetary seismic data is a fundamental tool for understanding the internal structure and tectonic activity of celestial bodies such as the Moon and Mars. In this project, we aim to develop a program capable of identifying seismic events using data collected by the Apollo missions and the InSight mission. These records, organized into training and testing subsets, allow us to explore the unique characteristics of seismicity on other planets and satellites.

The challenge lies in applying signal processing techniques and machine learning algorithms to analyze seismic data, detecting not only known events but also potential new events that have not been cataloged. By leveraging this approach, we aim to contribute to the study of planetary seismology, providing a deeper understanding of the seismic behavior of extraterrestrial bodies and potentially revealing unknown seismic patterns and phenomena.
.

II. RESPONSE TO THE CHALLENGE

A. Approach

Seismic Data Processing:

- The raw seismic signals from the InSight instrument on Mars are obtained.
- The signals are filtered and segmented into defined time intervals for further analysis.

Espectrogramas Spectrogram Generation:

- The filtered signals are transformed into spectrogram images using the Short-Time Fourier Transform (STFT).
- The spectrogram visually represents how the power of the signal is distributed across frequency and time, allowing the identification of characteristic events.

Application of Percentile and Normalization:

- **99th Percentile Adjustment**: The maximum value (vmax) of the spectrogram is adjusted using the 99th percentile. This ensures that the top 1% of power values, which may correspond to peaks or outliers, do not dominate the visualization. This technique highlights the more common variations in the signal, improving the readability of the spectrogram.

- **Normalization**: After adjusting the vmax, the spectrogram is normalized to a range of 0 to 1. Normalization ensures that all generated spectrograms have a consistent range, regardless of the amplitude of the original signals. This makes the images more suitable for machine learning models by providing a homogeneous input.

- **Control of Outliers:** Using the percentile method prevents high-power peaks from distorting the color scale, which is essential for highlighting the more common power variations in the spectrogram.

- **Data Consistency:** Normalization standardizes the images, ensuring that all spectrograms have the same value range (0-1). This consistency simplifies image processing for the machine learning model, improving classification accuracy.

Spectrogram Conversion to Images:

- The spectrograms are saved as PNG images with exact dimensions of 640x640 pixels, as this is the required size for the YOLO model.

- Borders and labels are removed to ensure that the images only contain relevant frequency and time information.

Classification Model Training (YOLO):

- The generated images are used as input for the YOLO model.
- Labels are created for each image, identifying whether the image corresponds to a seismic event or noise.
- The model is trained with a set of labeled images to learn how to identify characteristic patterns of seismic events.

Evaluation and Validation:

- The trained model is tested with a previously unseen validation dataset.
- The model's performance is evaluated in terms of precision, sensitivity, and its ability to detect seismic events with high reliability.

B. Martian Data.

The training catalog provided includes only two recorded seismic events on Mars: 2022-02-03T08:08:27.00 and 2022-01-02T04:35:30.00, both captured by a single channel, "BHV". Since this is insufficient for robust training, we expanded our analysis to include three channels, incorporating data from channels "BHU" and "BHW". To access this additional information, we queried the InSight data stored in the .Iris platform (Seismological Facility for the Advancement of Geoscience), focusing on "Scientific Data on the Martian Surface" for the ELYSE station.

Station Code	Description
ELYS0	Scientific Data on the Lander Deck
ELYH0	State of Health on the Lander Deck
ELYSE	Scientific data on the Martian surface
ELYHK	State of Health on the Martian surface

Fig. 1. Iris station table with available channels for the ELYSE station on Mars.

The broad-band channels available for this station cover different frequencies. We chose to work with the 20 Hz data associated with Location Code 02, which includes the high-gain channels “BHU”, “BHV”, and “BHW”, as well as the initially provided data.

Frequency	Location Code	Gain	Station Code
2 Hz	02	High	MHU, MHV, MHW
2 Hz	07	Low	MLU, MLV, MLW
10 Hz	03	High	BHU, BHV, BHW
10 Hz	08	Low	BLU, BLV, BLW
20 Hz	02	High	BHU, BHV, BHW
20 Hz	07	Low	BLU, BLV, BLW
100 Hz	00	High	HHU, HHV, HHW
100 Hz	05	Low	HLU, HLV, HLW

Fig. 2. Iris table of frequencies and stations.

By doing so, we increased the number of training examples from two to six, to which we added various types of noise to further enrich the dataset. For the analysis, we used 1-hour time windows that included the specific timing of the events listed in the catalog (Rel. Arrival). The data obtained from the IRIS query requires additional processing: it must first be scaled using the scale factor provided in the response, and the trend of the signal must be removed.

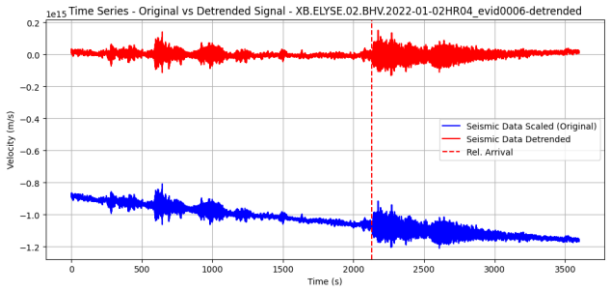


Fig. 3. Comparison of signal with and without trend over time.

These two steps are not strictly necessary, as it is possible to work with the raw data. The linear trend of the signal (a gradual change in amplitude over time) does not alter the frequencies present in the signal, but mainly affects the amplitude in the time domain. When calculating a spectrogram, which analyzes the energy distribution in different frequencies over time and where the data is normalized, the presence of a linear trend has little or no effect on the spectral representation. Therefore, we can forgo trend adjustment and scaling in this particular case.

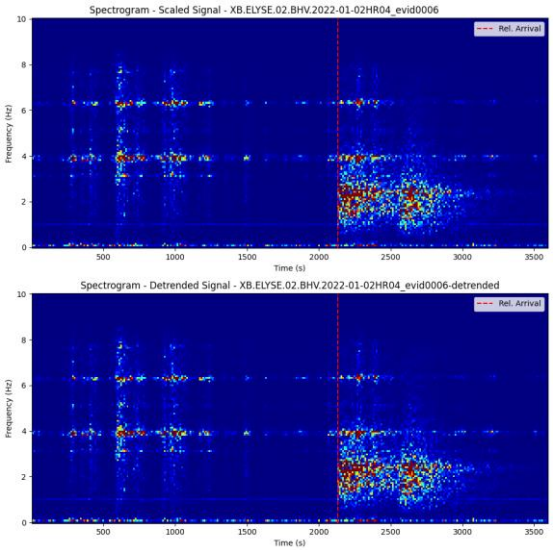


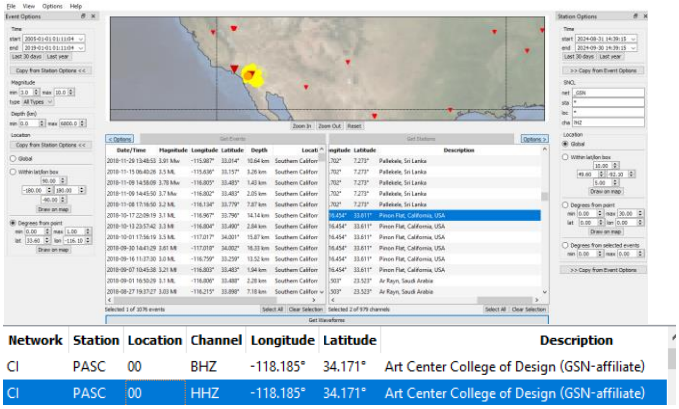
Fig. 4. Comparison of spectrograms with and without signal trend.

C. *Lunar Data*

Unlike the Martian data, lunar records are captured at a lower sampling rate (6.625 Hz vs. 20 Hz), which allows for longer time periods to be covered with the same number of data points. However, this difference in sampling rate is not an issue for our analysis, as we have designed a generalized model that does not differentiate between lunar, Martian, or terrestrial data. By focusing on the frequency content of the data rather than its origin, the model can recognize seismic patterns uniformly, regardless of the sampling rate or duration of the records.

D. *Terrestrial Data*

Unlike the provided test data for the challenge, validation data is used during the training process to evaluate the model's performance with previously unseen information. Since there are no additional confirmed seismic event records from the Moon or Mars, we will use terrestrial earthquake data as a reference for this process. This will help us fine-tune the model and improve its generalization ability before applying it to planetary data.



Network	Station	Location	Channel	Longitude	Latitude	Description
CI	PASC	00	BHZ	-118.185°	34.171°	Art Center College of Design (GSN-affiliate)
CI	PASC	00	HHZ	-118.185°	34.171°	Art Center College of Design (GSN-affiliate)

Fig. 5. PyWeed.

We have 40 examples of terrestrial earthquakes, which we augmented to a total of 120 by applying various types of noise. Additionally, we have 71 lunar examples, increased to 284 using the same technique, and 6 Martian seismic examples, expanded to 24 training data points. In total, we generated 428 examples, of which 15% were reserved for model validation. This strategy allows for a more precise and robust evaluation of the model's performance.

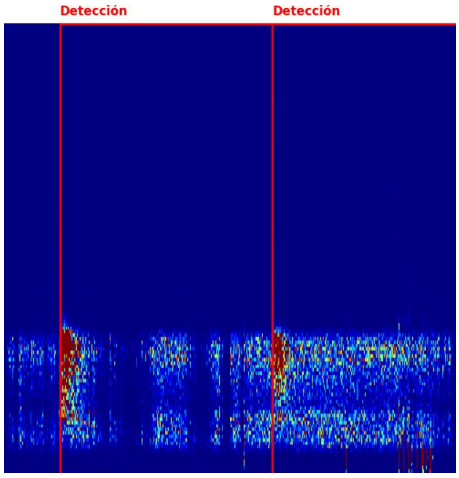


Fig. 6. Dataset sample with detection bounding boxes.

E. *Model Training for Event Detection*

We used the YOLO V10 model as the starting point for seismic event detection. The training was carried out using a dataset specifically labeled to identify a single category called "Event". For this process, the following hyperparameters were configured: training was performed with input image sizes of 640 pixels and 1024 pixels. The optimizer used was AdamW with an initial learning rate of 0.001, a momentum of 0.937, and a weight decay rate of 0.0005.

Various data augmentation techniques were applied to strengthen the model, including rotations of up to 5 degrees, translations of up to 10%, scaling, shearing, saturation variations, and brightness variations (hsv_v=0.2). In addition, strategies such as randaugment and copy_paste=0.1 were implemented to further enrich the dataset.

The training was conducted using a pretrained model configured to detect a single class (single_cls=True). This configuration enabled us to obtain a refined model for seismic event detection, maximizing accuracy and minimizing overfitting to the target class.

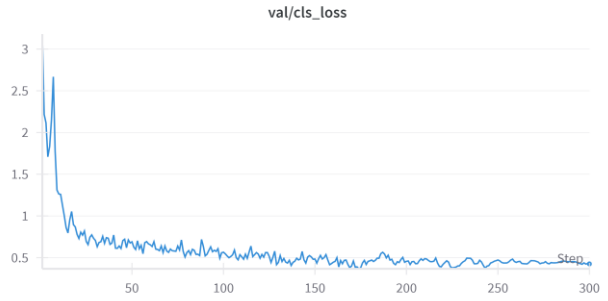


Fig. 7. Cls_loss Train.

III. EVENT DETECTION

The model outputs the coordinates of a bounding box, where the main point of interest is the left vertical line of the box, as it corresponds to the detection of a potential seismic event. Along with this information, the model provides the probability of the detection, indicating the confidence level that the identified region actually corresponds to a seismic event.

The coordinates provided by the model are in pixel units, and their maximum value depends on the dimensions used during the model's training (640 pixels, in our case). To translate this spatial information from the image into a corresponding time instance in the original signal, it is necessary to perform a conversion.

The relationship between the pixel coordinate and the signal's time is calculated using the following formula:

$$\text{Detection Time} = \left(\frac{\text{X Coordinate} \times \text{Signal Duration}}{\text{Input Dimension in Pixels} \times \text{Sampling Rate}} \right)$$

Where:

X Coordinate: Position value in pixels of the left vertical line of the bounding box returned by the model.
Signal Duration: Total length of the signal.
Input Dimension in Pixels: The number of pixels used to train the model (e.g., 640 pixels).
Sampling Rate: The sampling frequency of the original signal in Hz.

This conversion is determined by the Temporal Resolution, which defines the mapping between the duration of the original signal and the model's input dimension in pixels:

$$\text{Temporal Resolution} = \frac{\text{Signal Duration}}{\text{Input Dimension in Pixels}}$$

Where:
Signal Duration: Total length of the signal.
Input Dimension in Pixels: The number of pixels used to train the model (e.g., 640 pixels).

The Temporal Resolution indicates how many seconds of the signal are represented by each pixel in the image. This ensures that the detection in the image is correctly mapped to a time instance in the seismic signal, thereby facilitating the analysis of the detected event within its original temporal context.

Given:
Signal Duration = 80,000 data points (total number of samples in the original signal)
Sampling Rate for Martian Data = 20 Hz
Sampling Rate for Lunar Data = 6.625 Hz
Input Dimension in Pixels = 640 pixels or 1024 pixels

Martian Data:
Signal Duration in Seconds: 4000 seconds

- With an input dimension of **640 pixels**, the Temporal Resolution is **6.25 seconds/pixel**, meaning that each pixel represents 6.25 seconds of the Martian signal.
- With an input dimension of **1024 pixels**, the Temporal Resolution is higher, at **3.906 seconds/pixel**, providing a finer temporal mapping.

Lunar Data:
Signal Duration in Seconds: 12,073.48 seconds
- With an input dimension of **640 pixels**, the Temporal Resolution is **18.86 seconds/pixel**, representing a longer time span per pixel due to the lower sampling rate.
- With an input dimension of **1024 pixels**, the Temporal Resolution is **11.79 seconds/pixel**, allowing for a more detailed mapping of the lunar signal compared to 640 pixels, but still with less precision than the Martian data.

We recommend two alternatives to improve detection: the first is to use short-duration time windows to enhance detection accuracy, which allows the model to focus on smaller segments of the signal, reducing noise and improving the identification of subtle seismic events. The second alternative is to utilize a model trained with higher-resolution input images (a greater number of pixels), which increases the temporal resolution and enables the detection of events that may be overlooked with lower-resolution inputs. Implementing these strategies together can further optimize the model's performance, resulting in more precise and reliable seismic event detection.

IV. RESULTS

Below are the detection results for a seismic event on Mars labeled as "*XB.ELYSE.02.BHV.2021-12-24HR22_evid0007*". We decided to compare our findings with the seismic event detected in the Mars Monitor Dashboard by IRIS, which is illustrated below.

Fig. 9. Seismic Mars Monitor.

In the following image, the green line indicates the start of the seismic event according to the IRIS data, while the red lines represent the outputs of our 640-pixel model, suggesting potential indicators of a seismic event.

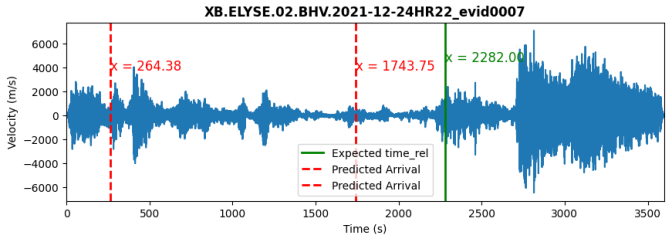


Fig. 10. Detect 640 px input, Output Signal.

The spectrogram would look like this:

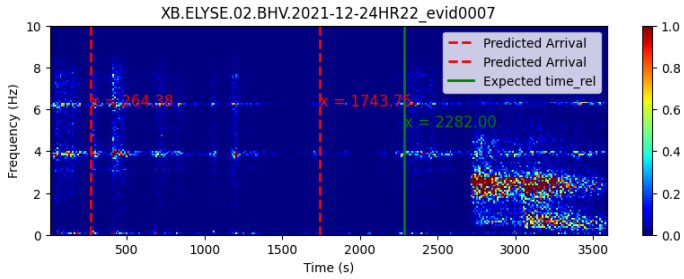


Fig. 11. Detect 640 px input, Spectrogram.

With the 1024-pixel model, we obtain the following results:

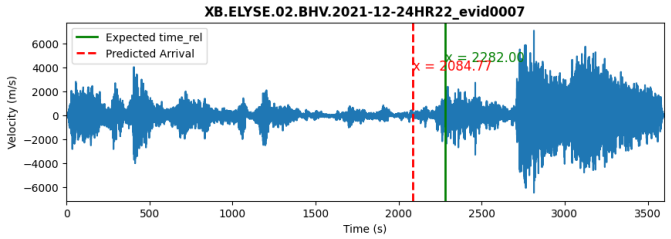


Fig. 12. Detect 1240 px input, Output Signal.

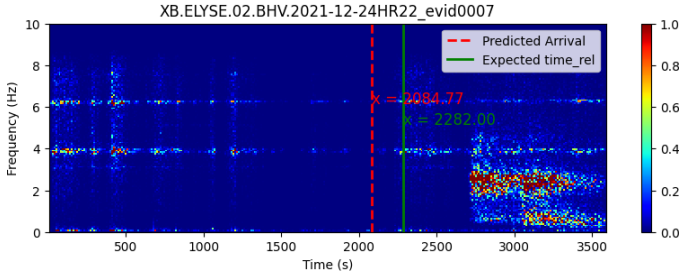


Fig. 13. Detect 1240 px input, Spectrogram.

As shown in the images, the detection results from our model are very close to those proposed by IRIS, demonstrating a high level of agreement and validating the reliability of our approach.

IX. CONCLUSIONS

Successful Detection of Seismic Events: The implementation of a YOLO-based model for seismic event detection on planetary data (Mars and Moon) proved to be effective, demonstrating high accuracy in identifying seismic events across different datasets.

Use of Spectrograms for Enhanced Analysis: Converting seismic signals into spectrogram images allowed for a more visual representation of the data, facilitating the identification of patterns and improving the model's ability to distinguish between seismic events and noise.

Adaptability to Different Planetary Environments: The model was able to generalize well across Martian, Lunar, and terrestrial data, despite variations in sampling rates and signal characteristics. This demonstrates the versatility and robustness of the approach in handling diverse seismic data from multiple celestial bodies.

Data Augmentation Improved Model Robustness: Applying various data augmentation techniques, such as noise addition and scaling, helped overcome the limitations of the small dataset, enhancing the model's training and performance.

Importance of Temporal Resolution: Adjusting temporal resolution through different input dimensions (e.g., 640px vs. 1024px) significantly impacted the model's precision, showing that higher resolutions enable a more detailed analysis of seismic events.

Potential for New Event Discovery: The approach not only detected known events with high reliability but also showed potential for identifying new, previously uncataloged seismic events, contributing valuable insights for planetary seismology research.

Future Applications: The methods and findings from this project can be further applied to other planetary bodies or integrated into future space missions, aiding in the continuous monitoring and exploration of extraterrestrial seismic activities.

To improve the model, more data is required. One alternative is to generate synthetic data, which is why we have implemented a prototype containing a sensor to detect vibrations and thus generate additional data. This same approach could be applied to probes on Earth, such as the replica of InSight.

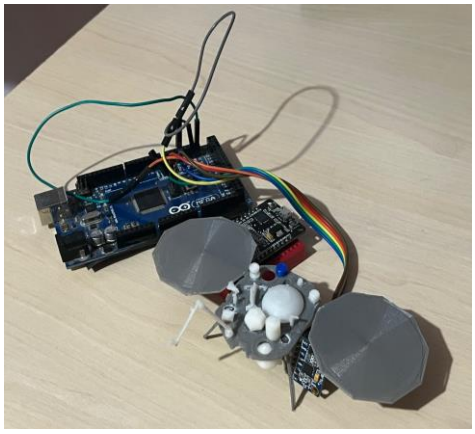


Fig. 14. Insight sensor prototype.

REFERENCES

[1] NASA's PDS, "Mars InSight Seismic Data," available at: https://pds-geosciences.wustl.edu/insight/urn-nasa-pds-insight_seis/data/xb/.

[2] IRIS, "ELYSE - Scientific Data on the Martian Surface," available at: https://www.iris.edu/hq/sis/insight#DMC.

[3] IRIS, "Mars Event Monitor," available at: https://www.iris.edu/app/mars-monitor/events?format=text&nodata=404&minmagnitude=0&orderby=time&starttime=2019-02-28.

[4] IRIS, "SAGE," available at: https://www.iris.edu/hq/.

[5] IRIS, "Earth Data," available at: https://ds.iris.edu/ds/nodes/dmc/software/downloads/pyweed/.

[6] Excelmaster, "Public Repository of the Project Code on the Github Platform," available at: https://github.com/excelmaster/seismicDetector.

[7] Ultralytics, "Ultralytics," available at: https://github.com/ultralytics/ultralytics?tab=readme-ov-file.

[8] Canva, "Infographic," available at: https://www.canva.com/design/DAGSuR66Chg/Ge3AFkk9GGBupHayqQHrwA/edit.