StellarFold

Introduction

This is a scientific tool that detects exoplanet existence in orbit of a star. It is meant to automatically identify exoplanet existence in preexisting data of stars with zero confirmed exoplanets so far. It can be used either on newly obtained data or on re-evaluating pre-existing data to point out stars of interests to scientists.

Detection is done using a pre-trained machine learning models using differential geometry (manifolds). In contrast to popular models, like neural networks or transformers, these models can learn the geometrical structures of data, which is fitting for an astronomy application.

The datasets

Exoplanet Data

There are various datasets available online from NASA missions collecting data on exoplanets. Each exoplanet has info about its star(s) and is labeled as “Confirmed”, “Candidate” and “False Positive”.

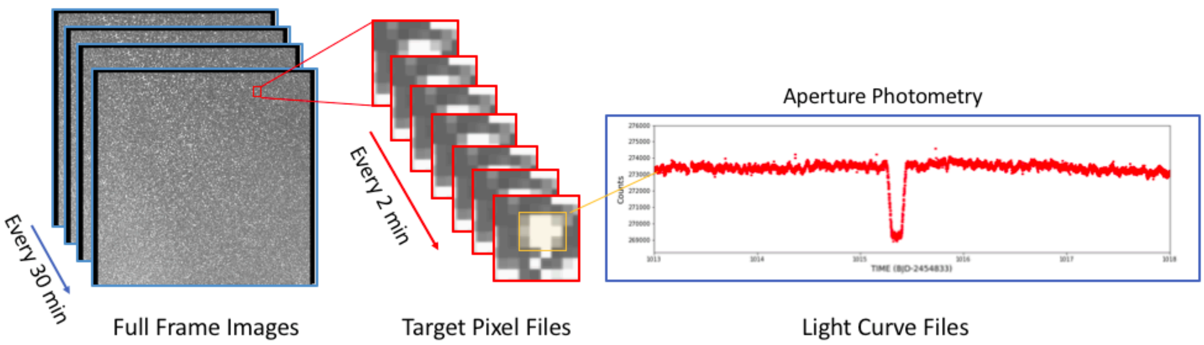
Exoplanet metadata:

- **Kepler Objects of Interest (KOI)**
- **TESS Objects of Interest (TOI)**
- **K2 Planets and Candidates**

Light Curve Data

Each mission captures Full Frame Images (FFIs) every 30 minutes and Targeted Pixel Files (TPFs) every 2 minutes of target stars. Light Curve Files (LCFs) are created from the light flux density calculated in a specified areas of a series of TPFs.

Star photometric data: https://heasarc.gsfc.nasa.gov/docs/tess/data-products.html



Model Training

Extracting SPD features

Features are picked for each TPF (mean, std, kurtosis, skewness, etc.) to create vectors. The Covariance matrix of these vectors is used as a descriptor of each star for the model.

What does the model learn?

The model learns a set of cluster centroids that represent four classes of stars that have:

- at least one confirmed exoplanet
- no confirmed exoplanets and at least one candidate exoplanet
- no confirmed or candidate exoplanets and at least one False Positive exoplanet
- no exoplanet info so far

These classes have been created based on the annotations on the available exoplanet data.

After training, the model will calculate an exoplanet existence indicator that can be used for scientists to pick the star most likely to have an exoplanet.

How does it learn the centroids?

The cluster model is learned by using Riemannian Conjugate Gradient a classification objective function like:

$$\min_{\mathbb{C}} f(x, y, z) = \sum_{i=1}^{N} y_i * max(0, D_{ne} - D_{he}) + (1 - y_i) * max(0, D_{he} - D_{ne}),$$

$$\text{where } D_{he} = w_{conf}\, \mathrm{D}(X_i, C_{conf}) + w_{cand}\, \mathrm{D}(X_i, C_{cand}) + w_{ref}\, \mathrm{D}(X_i, C_{ref})$$
$$D_{ne} = w_{fp}\, \mathrm{D}(X_i, C_{fp})$$

Demo

A public demo is included in the repository (`main.mlx`), implemented in **MATLAB** using the **Manopt** optimization toolbox.

1. https://www.mathworks.com/?s_tid=gn_logo
2. https://www.manopt.org/

Challenge Submission

Summary

We propose a machine learning model to identify candidate stars that may host previously undetected exoplanets within existing observational data.

The model applies concepts from differential geometry, representing statistical features of Target Pixel Files (TPFs) on manifolds. Unlike conventional models such as neural networks or transformers, this approach preserves the intrinsic geometric structure of stellar flux variations, enabling more accurate modeling of complex, nonlinear temporal patterns.

The system is designed to re-analyze both new and archival stellar data, flagging stars previously classified as planet-free but potentially containing overlooked exoplanet signals, thereby aiding scientific re-evaluation efforts.

Project Details

Provide details about your project. Consider the following questions: What does it do or how does it work? What benefits does it have? What is the intended impact of the project? What tools, coding languages, hardware, or software did you use to develop your project? How is your project creative? What factors did your team consider?

StellarFold is an machine learning model on Symmetric Positive Definite (SPD) manifolds. It can detect stars that have at least one exoplanet, which would help to find stars that were assumed to have no exoplanet indication so far.

How does it work?

This is done by learning the intrinsic geometry of temporal variations in stellar flux in Targeted Pixel Files (TPFs) obtained through satellite missions like Keppler, TESS and K2. The information is captured in SPD matrices that work as cluster centroids that represent stars that have:

a. ≤1 confirmed exoplanets

b. 0 confirmed exoplanets, ≤1 candidate exoplanets

c. 0 confirmed or candidate exoplanets, ≤1 refuted exoplanets

d. 0 confirmed, candidate or refuted exoplanets, ≤1 false positive exoplanets

The groups were picked based on available exoplanet labels from datasets obtained by the missions. Categories a, b and c are assumed to indicate exoplanet existence, so they are used to created a weighted indicator of exoplanet existence while category d is used to indicate the lacking of exoplanets.

Intended Impact

The expected result is to have the model scan all new and previously obtained stellar flux data and point out stars that were assumed to have no exoplanets so far while in reality they were misjudged. The model indicator can be used for sorting stars of interest. It cannot be a simple classifier or exoplanet existence as stars would never be labels to have zero exoplanets.

Project Tools & Creativity

Sample code for the project was coded in Matlab and is available in our public Github repository. It uses the manifold optimization (manopt) toolbox to accurately learn the model.

The matlab live-script on the project demonstrates how the StellarFold model would be trained based on the Kepler NASA Mission data. It is incomplete and works as a sample with comments and some completed functions to show its functionalities.

In a complete state the project should check a group of stars with no identified exoplanets and create a ranking of interest. Scientists will be able to set a threshold for the indicator which would show when the ranked stars are really important to check again.

Use of Artificial Intelligence (AI)

We used AI tools to search generic information about discovering exoplanets on stars as well as scientific tools and data available as we are not proficient in this scientific field.

References

1. Kepler Objects of Interest (KOI)
2. TESS Objects of Interest (TOI)
3. K2 Planets and Candidates
4. Database Table Search
5. A Study of Light Intensity of Stars for Exoplanet Detection using Machine Learning
6. Application of Manifold Corrections in Tidal Evolution of Exoplanetary Systems