Vianajr.com
vianajr@mit.edu

Identifying Exoplanets with AI

Javier Viaña[1,2], Andrew Vanderburg [1,2], Mason Fang [1,2], Zoë de Beurs [1,2], Jennifer Yee [3], Michelle Kunimoto[1,2], Daniel Muthukrishna[1,2]

[1] Massachusetts Institute of Technology
[2] Kavli Institute for Astrophysics and Space Research
[3] Center for Astrophysics | Harvard & Smithsonian

Subscribe

Acknowledgements

All this works were possible thanks to the support of:

- The NASA Extremely Precise Radial Velocity Foundation Science Program (No. 80NSSC22K0848)
- The NASA Astrophysical Data Analysis Program (No. 80NSSC22K1408)
- The Botton Fellowship for Postdoctoral Research at MIT
- 3 UROP MIT Grants for Undergraduate Research

Motivation

The study of exoplanets, or planets outside of our Solar system, is entering the big data era. The most recent Decadal Survey in Astronomy and Astrophysics, which sets the priorities for astronomy research every 10 years, identified the search for habitable planets and life outside the solar system as the highest funding priority for the next decade. Numerous experiments and space missions are in the planning/implementation phase, and the amount of data available in the coming years will skyrocket.

The increasing volume of data acquired through space instruments presents a new processing challenge: Traditional algorithms that have been used in the past, are now insufficient to fully harness the entirety of the available data. In order to keep up with the deluge of new data, we must transition to using machine learning (ML) and artificial intelligence (AI) algorithms, as these methods possess the capability to effectively ingest and process all the observations. This work is at the intersection of AI and the search of exoplanets. We present updates in 4 different ongoing projects:

SWIPES: Sliding Window Inference Pipeline for Exoplanet Search

How common are Earth analog exoplanets?

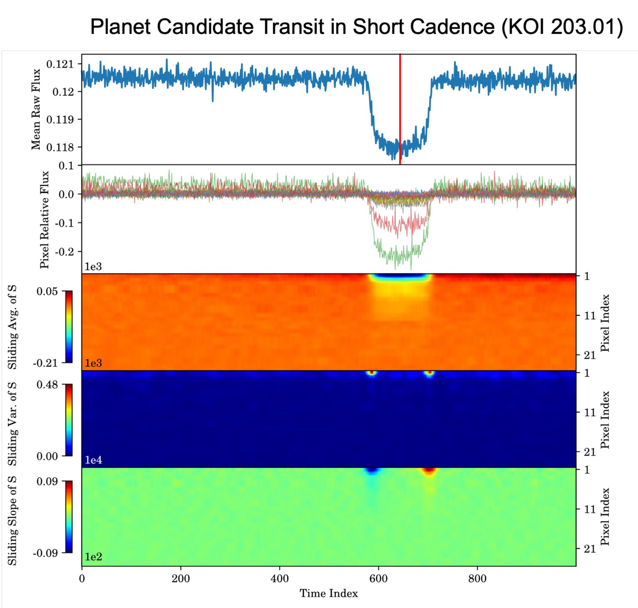
This question is key to understanding habitability and designing future experiments to search for life, but traditional searches of data from the Kepler space telescope have failed to answer it convincingly. To solve this, we are developing SWIPES, a novel explainable AI algorithm that analyzes all the available pixel-level flux data to reveal hidden Earth-like exoplanets that were previously unknown in Kepler data.

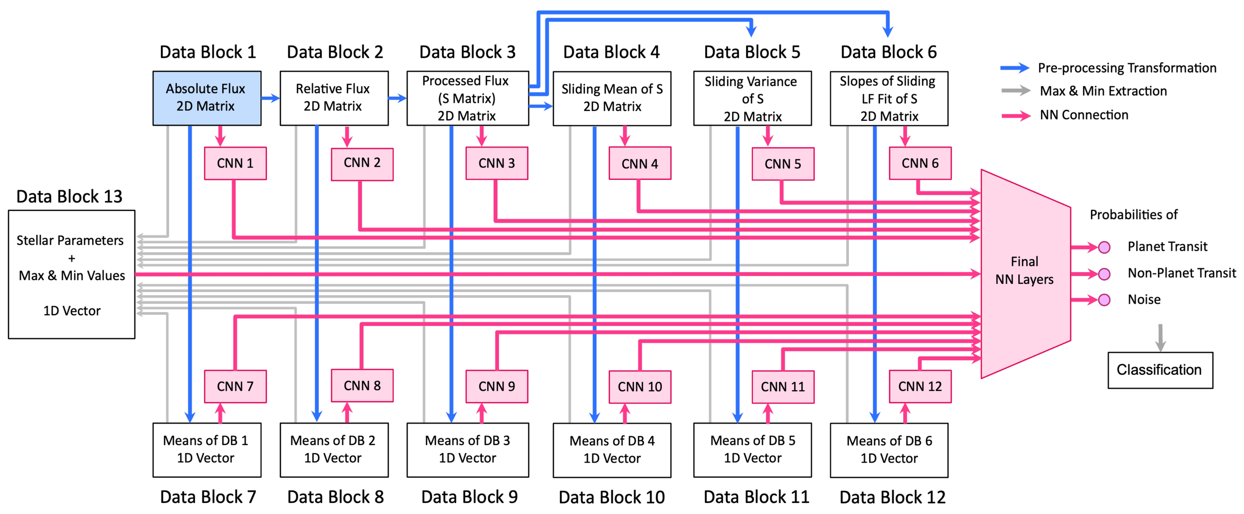
SWIPES:

SWIPES is a convolutional neural network architecture that leverages pixel-level flux information in order to find potential transits as it examines the time series data. This is a design choice that promises to dramatically reduce the instrumental artifacts, and reveal genuine exoplanet transits.

Dataset used:

For the training of the neural network, we generated a database using Kepler's labeled observations. The instances of the dataset contain the pixel-level flux values equivalent to one full day of observations. The time window was chosen so it could fit inside Earth-like transits. Each instance was labeled with one of the following categories: planet candidate transit, false planet transit, or noise.





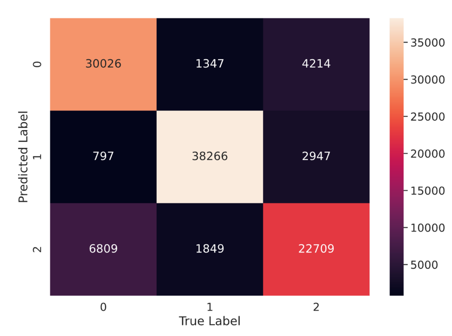
Architecture:

The architecture chosen is a **branched Convolutional Neural Network** (CNN) that processes separately all the different blocks of treated data from the window. The outputs of the CNNs are aggregated using final dense layers that obtain the probabilities of each of the three categories.

Results:

The confusion matrix below refers to the classification of the instances for the test set with 0,1, and 2 labels for the noise, non-planet transits, and real planet transits, respectively.

Confusion Matrix Test Set



SCOOP: Star Classifier for On & Off-target Predictions

Introduction:

From the data collected by TESS, experts must isolate exoplanets from false positives, which may be caused by other stellar phenomena such as eclipsing binaries. However, human vetting of exoplanet candidates is still relatively inefficient and inconsistent considering the large volume of data coming from the satellite every month. Teams of experts spend a few days performing triage by filtering out obvious false positives, then may spend on the order of a week vetting the remaining high-quality candidates. In addition, human vetting is vulnerable to subjective differences, with different vetters having different interpretations of the same light curves (Yu et al. 2019).

On vs Off Target:

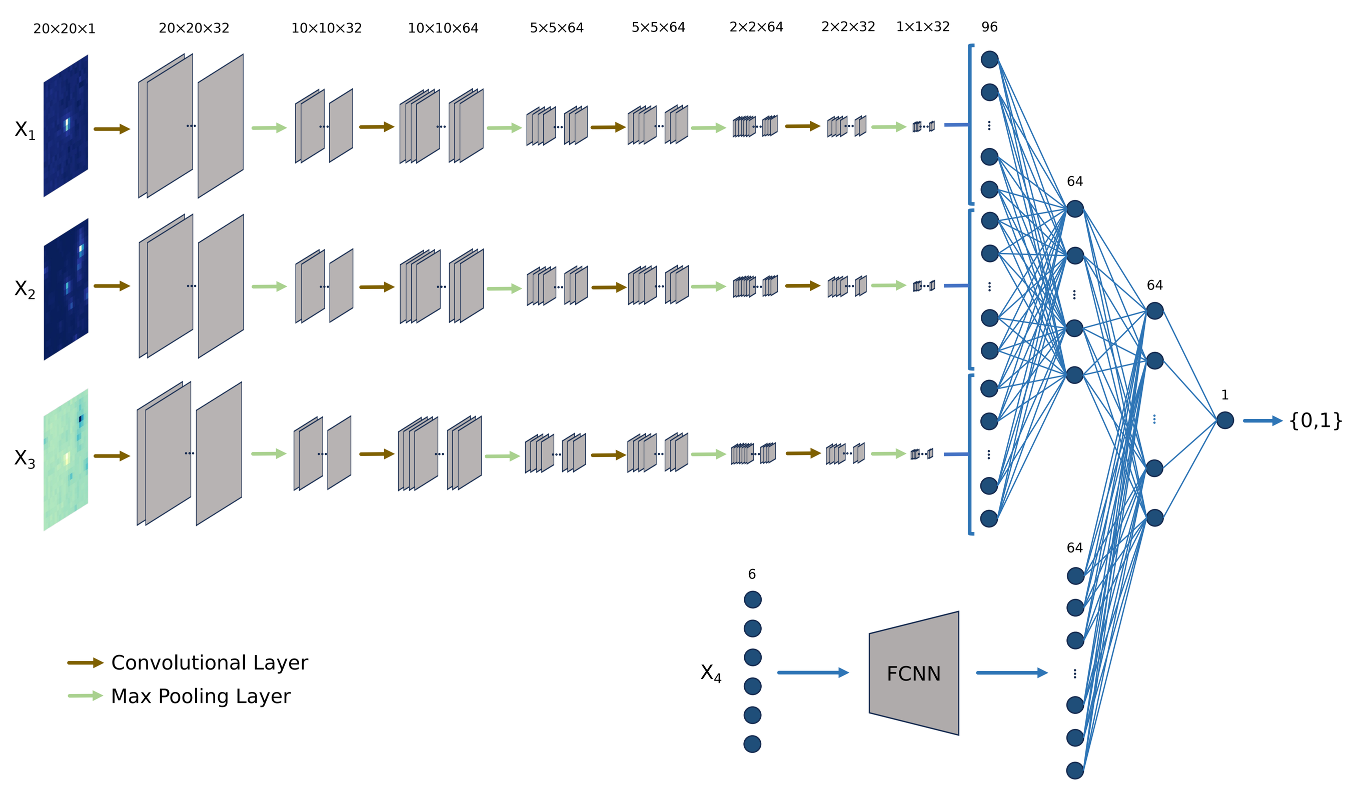
One such area of common disagreement is in identifying whether or not a transit signal is coming from the target star or not (on-target vs. off-target). Due to the fact that flux data from TESS comes in discrete pixels, identifying the location of the source of a transit signal can be extremely difficult and is often up to subjective interpretation.

Star Classifier for On & Off-target Predictions (SCOOP):

SCOOP is a pipeline based on a convolutional neural network that efficiently determines whether or not a candidate is on-target. It uses as inputs both a reference out-of-transit, the on-transit images of the target star, and a feature matrix that mixes the information of these two. These are then encoded by separate 2-d convolutional branches of a network, together with various 1-d features of the star. The outputs of all the branches are then concatenated and passed through a final set of dense layers, producing a single output value that represents the probability of being on target.

Results and Deployment:

Our ensemble of models currently achieves over 88% accuracy on the test set. We are planning to integrate this algorithm in the pipeline of Astronet to enhance its performance. These results will be published in Fang et al. (in prep).



Astronet Updates

Introduction:

Since its launch in 2018, the Transiting Exoplanet Survey Satellite (TESS) has provided countless observations of targeted stars, allowing astronomers to use the data offered to search for exoplanets through the transit method. However, only a fraction of these data contains detectable exoplanet transits. Deep Neural Network pipelines such as Astronet have proven excellent at distinguishing potential exoplanet transit signals from other phenomena like instrumental effects and stellar variability.

Astronet Updates:

In the latest results of Astronet (Tey 2023) our group published a curated dataset, manually reviewed, containing light curves from both the Primary Mission and 1st Extended Mission full-frame images and periodic signals detected via box least-squares. This data set was used to re-train a new network named Astronet-Triage-v2.

Testing results of Astronet-Triage-v2:

- Recall rate: 99.6%
- Precision: 75.7%
- Area under the precision-recall curve: 0.965.
- Improvement over Astronet-Triage (previous version of Astronet): 4%
- The model's generalization was tested on withheld 1st Extended Mission data.

Deployment:

- Astronet-Triage-v2 was deployed in the Quick-Lookup Pipeline for planet candidate triage
- Astronet-Triage-v2 has already identified 3577 potential new planet candidates.
- In the same conditions, the first version of Astronet (Astronet-Triage) missed nearly 200 of these planets.

Summary:

Upgrading to Astronet-Triage-v2 helped us prevent the loss of several potential planet candidates.

Microlensing Detection Algorithm

What is the population of cold exoplanets in our galaxy?

Our team has developed a new machine learning method to enhance exoplanet discovery using microlensing. Current detection methods rely on human inspection, and thus are slow and inconsistent. This new AI system should significantly increase the efficiency and sensitivity of these surveys, while explaining how the inference is made.

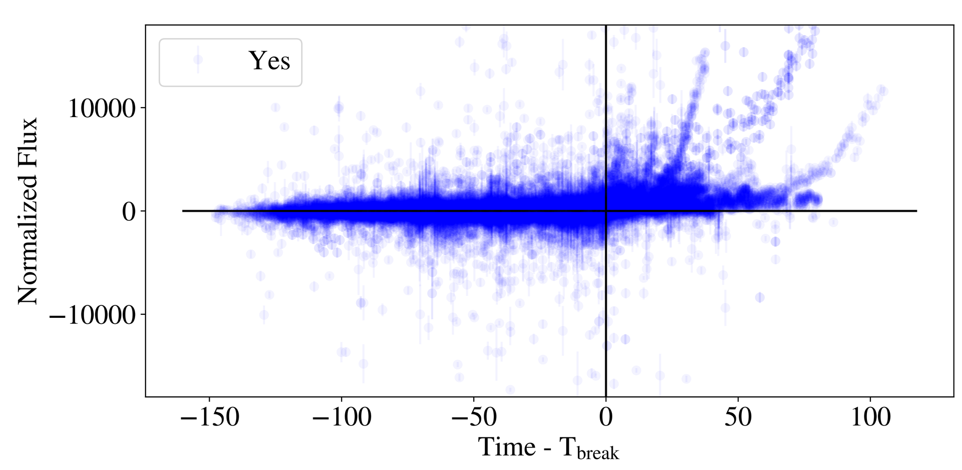
First, we rely on a microlensing-event detection algorithm customized for the multi-observatory setup of the Korea Microlensing Telescope Network (KMTNet). This algorithm is specifically designed to identify 'rising' events, where the brightness increases over time (Hyoun-Woo, K. et al. 2018).

Out of all the alerts found by the system of three telescopes (BLG, SSO, and SAO), our pipeline is able to distinguish potential microlensing alerts and fake signals such as pixel bleeding in the CCD cameras, diffraction effects from nearby stars or other instrumentals.

This last pipeline allows us to disregard a vast amount of data to obtain a selection of promising alerts. It is designed as a Recurrent Neural Networks architecture that not only gets the time-series information of the flux but also the sky background, FWHM, flux error, PSF quality, and air mass from each of the telescopes.

With our current architecture we are getting accuracies above 80% and we expect to improve the results with more data and a more complex algorithm.

Aggregated Training Dataset for the Yes category



References

Borucki, W. J., et al. 2010, Science, 327, 977
Mullally, F., et al., 2018, AJ, 155, 210
Jenkins, J. M. 2002, ApJ, 575, 493
Kovács, G., et al. 2002, A&A, 391, 369
Tey, E., et al. 2023, AJ, 165, 3
Yu, L., et al. 2019, AJ, 158, 25
Hyoun-Woo, K. et al. 2018 arXiv