TOWARDS DEEP LEARNING FOR TRANSITING EXOPLANET SEARCH USING SIMULATED TESS DATA. Y. J. Rusticus[1,2,4], B. H. Foing[2-5], A. M. Heras[2], V. Foing[2,4,6], C. J. Hönes[2,4,6] and J. M. Terpstra[2,4,7], [1]University of Amsterdam (yke.rusticus@student.uva.nl), [2]ESA ESTEC, [3]Leiden Observatory, [4]VU Amsterdam, [5]ILEWG EuroMoonMars, [6]University of Amsterdam, [7]Amsterdam University College

Introduction: Space-based observatories such as TESS (Transiting Exoplanet Survey Satellite [1]) produce light curves of hundreds of thousands of stars over their mission lifetimes. A tiny fraction of these stars shows periodic dips in brightness, resulting from one or more orbiting planets crossing in front of the stellar surface in our line of sight. While the physics of transiting planets is well-understood, the search for their signatures in light curves is hindered by stellar and spacecraft induced noise. In order to distinguish transit signatures from noise and underlying complex patterns in the data, Deep Learning (DL) is expected to play an important role, especially in future missions such as PLATO [2]. However, while the abundance of data continues to grow, the development of DL-based methods for this task seems to lag behind. In this work, we set up a framework for developing and comparing methods for detecting transits in simulated light curves. Furthermore, research prospects are given in line of this work, on the exploration of Recurrent Neural Networks (RNNs) and their application to real TESS data containing known transiting exoplanets.

Background: The steps of transit discovery may be categorized as detection, validation and characterization. Often, these steps require some form of preprocessing or detrending of the data. However, commonly used approaches for detrending such as median filtering or the use of Gaussian Processes (GPs) may distort the transit signals [3]. To avoid this, [4] proposed to model transits simultaneously with the background noise. Another approach is to exploit the flexibility of Neural Networks (NNs), which are known for their ability to learn complex patterns from data, often by utilizing labeled training examples. In recent years, Convolutional NNs (CNNs) have been explored for detection by [5-7]. RNNs, which are designed specifically for sequential data, were used for detrending by [8], but they assumed the transits to be known beforehand. Naturally, a next step could be to explore RNNs for transit detection.

Methods: DL-approaches generally require labeled training data, so we opted to use simulated data for the development and comparison of methods. This allows us to compare the 'ground-truth' of a sample with the predictions made by the detection method. In order to mimic light curves as observed by TESS, each sample is constructed using GPs similar to [6]. Different than [6], we describe the process of stellar variability as a mixture of stochastically-driven damped harmonic oscillators, following [9]. Transits are generated from a range of parameters and injected into a portion of the light curves, making use of `batman` [10]. To reduce the amount of unrealistic samples, parameter ranges were chosen to approximately match with those of known stars and exoplanets. These parameters include, for example, orbital period, transit duration and planet radius relative to its host star. However, as we are now in complete control over the data, we may also choose to simulate only, for example, small planets with long periods, or planets in multi-planet systems. Lastly, large gaps and missing values can be varied per sample to simulate imperfections, which need to be taken into account when applying a transit detection method to real-world data. Fig 1 shows a randomly generated light curve compared to real TESS data. Together with auxiliary variables such as stellar parameters, these samples are used to create datasets for the training and testing of DL-methods.

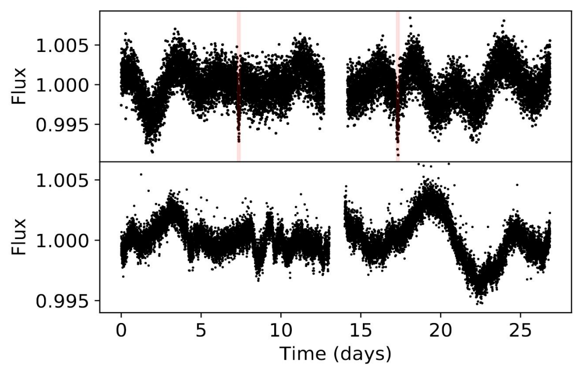

Fig 1. (top) Simulated light curve with 'ground-truth' transits; (bottom) TESS light curve of TIC 243271623.

Experiment and results: As an initial experiment, we applied the box least squares (BLS) algorithm [11] to search for transits in a set of 5k simulated light curves of 2-minute cadence. For detrending, we used a time-windowed median filter with a window of 0.5 days. Each light curve contained at least two transits of a single planet, so the detection method could be evaluated on its ability to retrieve the correct orbital period P and the time of the first transit t_0. Note that the samples in this experiment were chosen such that we could evaluate the behaviour of a simple detection method applied to our simulated data, so the presented results are not in direct correspondence to results that could be obtained with real TESS data. Moreover,

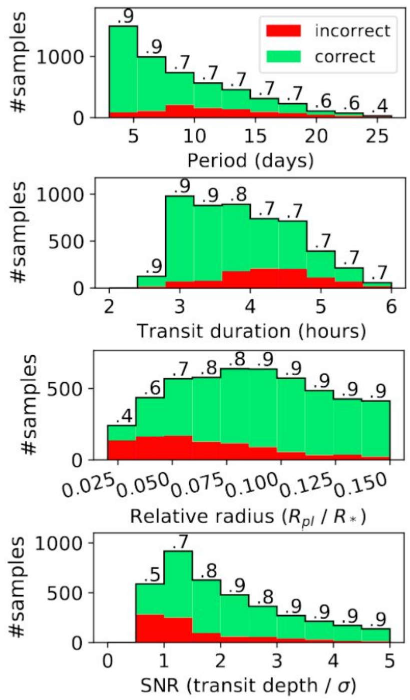

Fig 2. Distributions of correctly and incorrectly detected transits in simulated data for varying parameter values. The fraction of correct detections is given per bin. Results were obtained using a BLS-based detection method. For reference, a relative radius of 0.02 corresponds to a transit depth of about 430 μmag. The bottom figure shows the signal-to-noise ratio (SNR) distribution, where the SNR is calculated as the transit depth divided by the white-noise σ.

transit-less light curves were excluded in this experiment. When developing DL-based detection methods, transit-less light curves will be included to allow precision and recall to be measured and compared. For 81% of the samples (i.e. 4050 samples), both t_0 and P were correctly detected by the BLS-based detection method. Fig 2 and Fig 3 show that most errors were made for the smallest and longest-period planets, and for long duration transits. More transits could probably be detected using a larger number of candidate periods to search over, but the duration at which they can be detected is limited by the detrending method that is being used. The results are as expected, and show that our data are suitable for use in developing a new DL-based detection method.

Further research: Ultimately, our goal is to detect currently unknown transit events in real TESS data. In particular, we aim to search for additional transits in systems with an already known transiting planet. This is because one may expect the orbital planes of planets around the same star to be oriented similarly, thus increasing our chances of finding additional transits. It is known that this is not always the case, but we hope to support the investigation in when this claim holds. Also, finding additional planets in known planetary systems would be of great value for our understanding of planetary system architectures and dynamics. However, before we move towards real-world data, we plan to explore RNNs in the context of transit detection in simulated light curves. Possibly, their ability to learn complex temporal patterns will help in detecting the smallest planets with long periods, and long duration transits that are distorted by stellar or systematic noise.

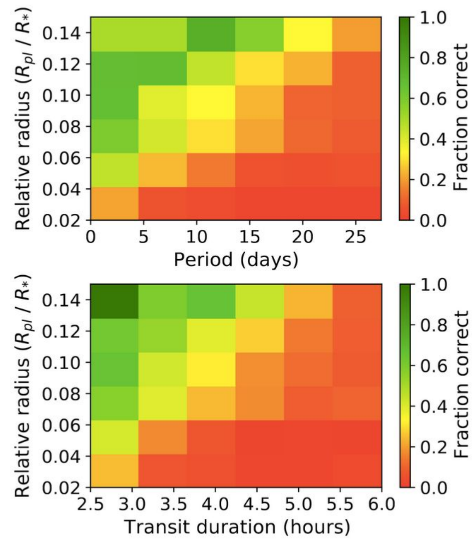

Fig 3. 2D plots of the detection results shown in Fig 2.

Acknowledgments: We thank D. Ruhe for providing helpful insights and supervision on the Machine Learning aspect of this work. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

References: [1] Ricker G. R. et al. (2014) *J. Astron. Telesc. Instrum. Syst.*, 1(1), 014003. [2] Rauer H. et al. (2014) *Exp Astron*, 38, 249-330. [3] Hippke M. et al. (2019) *AJ*, 158, 143. [4] Foreman-Mackey D. et al. (2015) *ApJ*, 806, 215. [5] Pearson K. A. et al. (2018) *Mon. Notices Royal Astron. Soc.,* 474, 1, 478-491. [6] Zucker S. and Giryes R. (2018) *AJ,* 155, 147. [7] Chintarungruangchai P. and Jiang I. (2019) *PASP,* 131, 064502. [8] Morvan M. et al. (2020) *AJ,* 159, 109. [9] Foreman-Mackey D. et al. (2017) *AJ,* 154, 220. [10] Kreidberg L. (2015) *PASP*, 127, 1161. [11] Kovács G. et al. (2002) *A&A*, 391(1), 369-377.