Searching for Anomalies in the ZTF Catalog of Periodic Variable Stars

Ho-Sang Chan[1,2], Siu-Hei Cheung[1,2], Ashley Villar[3], Shirley Ho[2]

[1]The Chinese University of Hong Kong
[2]Center for Computational Astrophysics, Flatiron Institute
[3]The Pennsylvania State University

Introduction

Wide-field, untargeted surveys continue to exponentially increase our discovery rates of periodic variable stars (PVSs). It is reasonable to expect anomalous PVSs which defy expectations. They could potentially contain information about new physics. Here, we provide an anomaly detection algorithm to effectively search for anomalous periodic variables detected with Zwicky Transient Facility (ZTF). ZTF contains numerous publicly available data which serve as a good training set for big data problems in astronomy.

Methodology

We extracted light curves of PVSs from the ZTF catalog of periodic variable stars (ZTF CPVS). The data are given in two bands (g-band and r-band) of magnitude. We phase-folded the light curves by their joint period, and chose to interpolate them using the method of **multivariate Gaussian process regression** (MGPR):

$$K = C\exp\left(-\frac{|\vec{\phi}-\vec{\phi}'|}{l_{\phi}^{2}} - -\frac{|\vec{\lambda}-\vec{\lambda}'|}{l_{\lambda}^{2}}\right) + \begin{cases} \delta \text{ if } \vec{r} = \vec{r}' \\ 0 \text{ otherwise} \end{cases}$$

Here, $\vec{r} = (\vec{\phi}, \vec{\lambda})$ is a high dimensional vector. K is the covariance function, ϕ is the phase, λ is the wavelength of the band filters, C is a constant, and δ measures white noises.

We then generated 160 evenly spaced data points along with the phase direction for both bands. We stacked them horizontally to form an "image" of size 2×160. They will be encoded through a **convolutional variational autoencoder** to generate their latent features.

Conclusion

We present a convolutional autoencoder-based pipeline for detecting anomalous PVSs in the ZTF CPVS. Detailed follow-up studies are essential in order to fully understand these anomalies.

The Top Anomalies

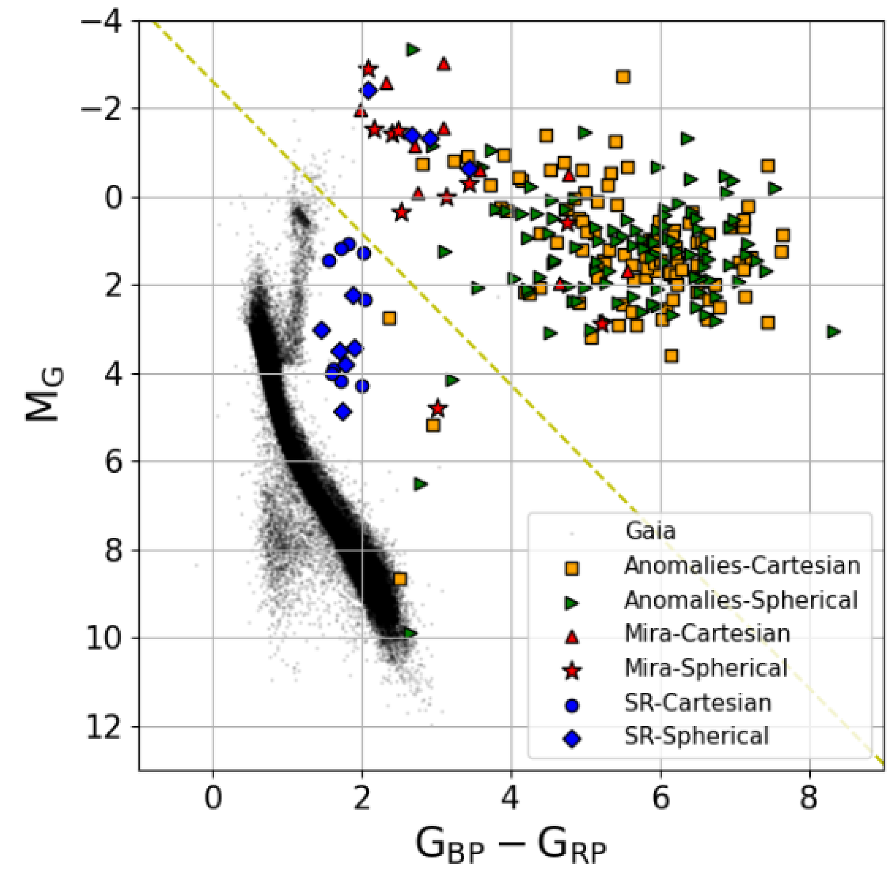

Figure 2. The distribution of anomalies in the Gaia HR-Diagram against the normal SR (blue) and Mira (red) variables, and main-sequence stars (black). We omit the description of the yellow straight line.

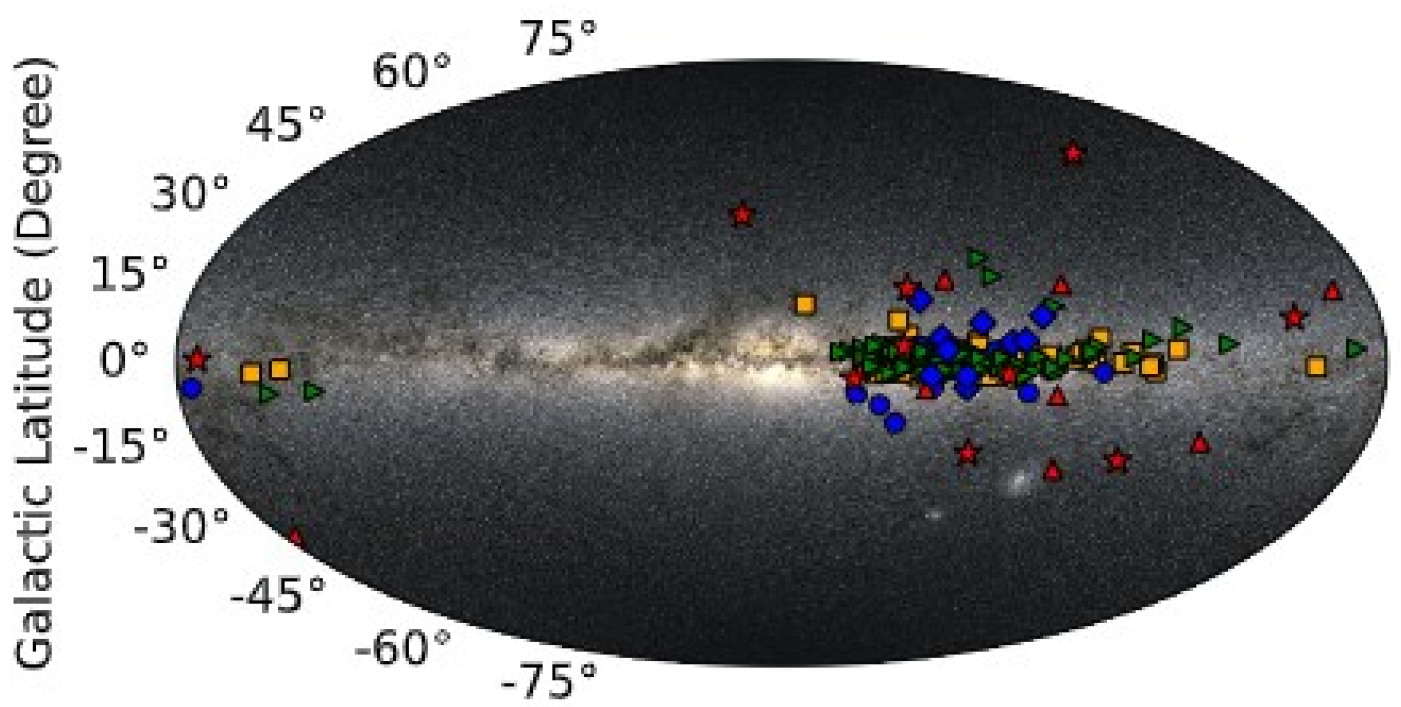

Figure 3. The distribution of anomalies in the Milky Way galactic coordinate system. Galactic longitude is omitted. Legends are the same as Figure 2. Image credit: ESA/Gaia/DPAC.

We find that our learned latent space exhibits an **annular structure**, which inspires us to transform the latent space using **spherical coordinates**. We search for anomalies in both spaces using an isolation forest and compare results. We identify most of the anomalies to be **Semi-regular** (SR) and **Mira** variables. They are having high-variability and irregular light curves. Their position in the Gaia-HR Diagram (Figure 2) indicates that they are **bright but cooler** than typical Mira and SR variables. They are also tightly distributed **near the Galactic plane** (Figure 3).

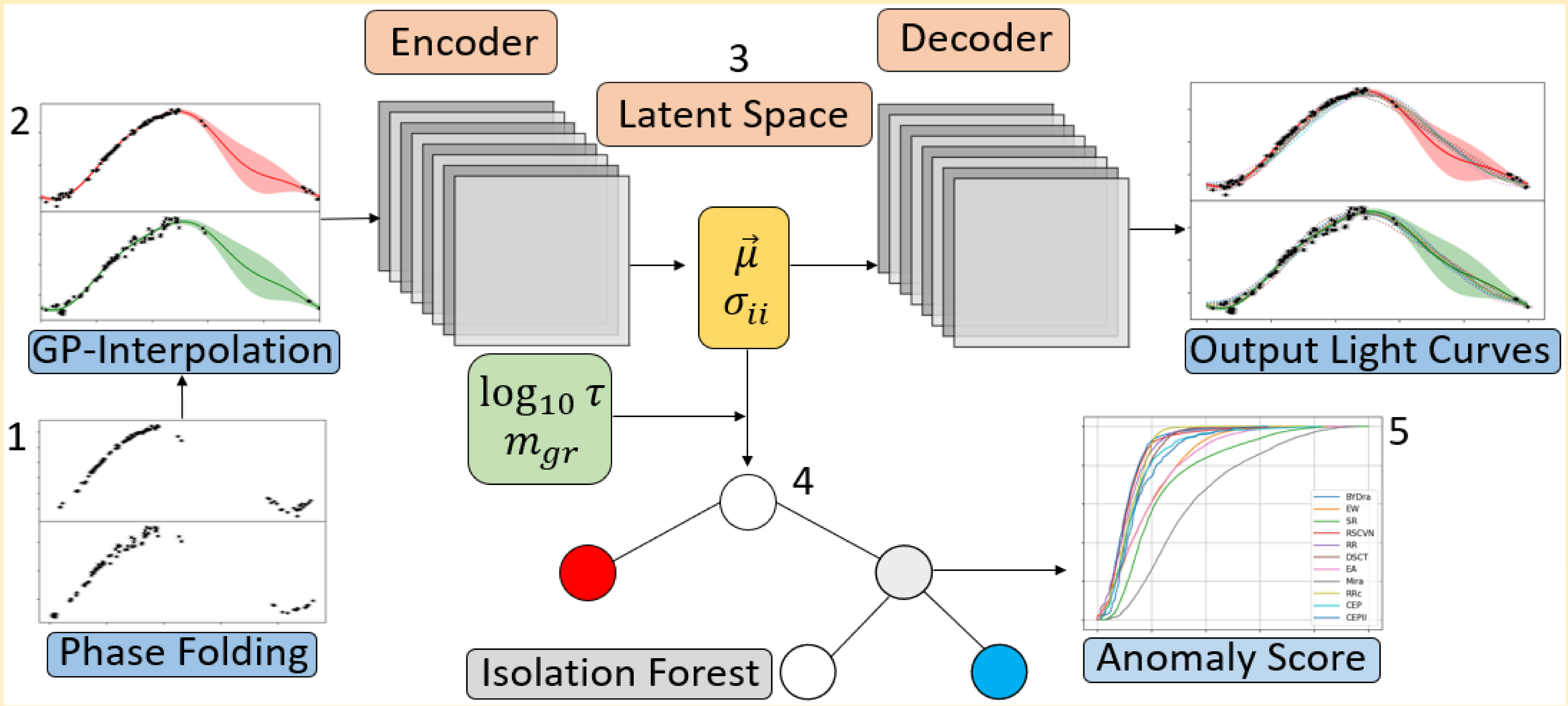

Figure 1. The anomaly detection pipeline: 1. Phase-folding the raw detection data. 2. Interpolation using the Multivariate Gaussian Process. 3. Encoding to generate latent features. 4. Append additional hand-engineered features. 5. Isolation forest and ranking the anomalies.

Spectral Information

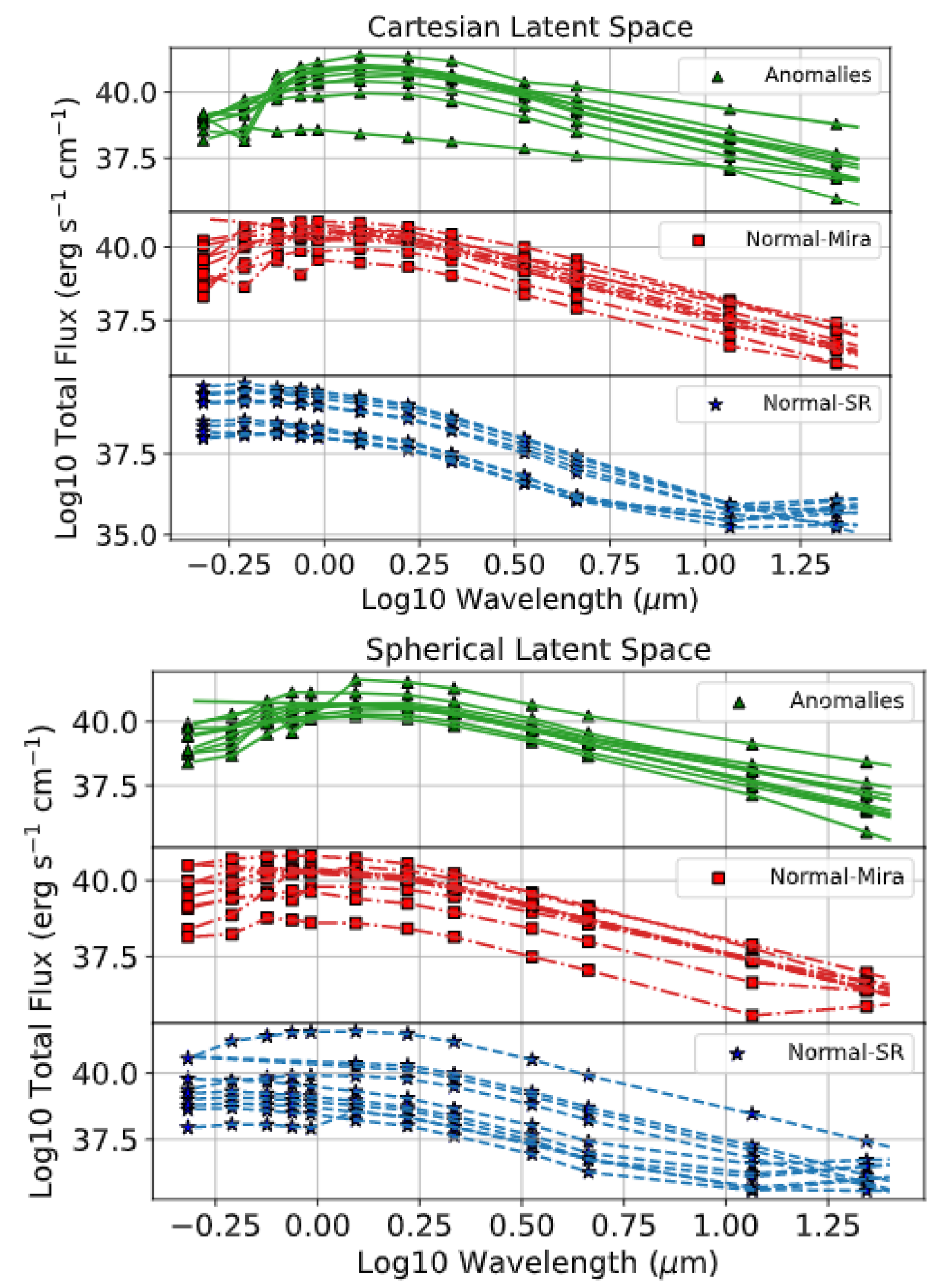

Figure 4. The spectral energy distribution (SED) of the anomalies, normal SR and Mira variables in both coordinates.

We extracted spectral information (interstellar extinction corrected) from the 2MASS, URAT, and Pan-STARRS1 surveys. We find that the anomalies are **more reddish** (Figure 4), which might indicate that they are **typically dusty**.

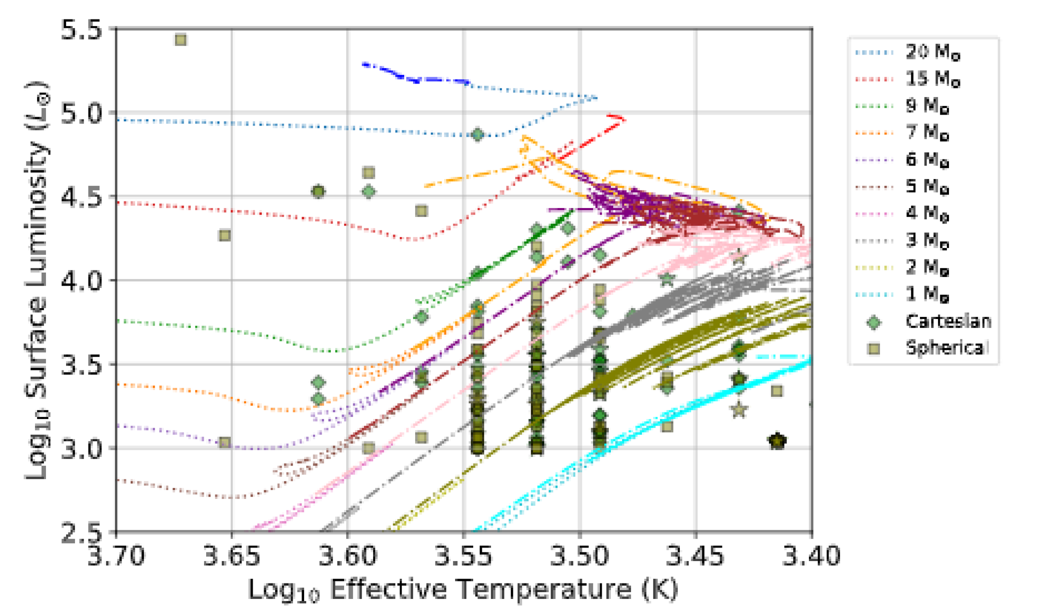

Figure 5. The luminosity against surface temperature distribution as obtained from fitting SEDs of the anomalies, tabulated with MESA stellar evolution track.

We fit their SEDs using the open-source code DESK. The fitting results (Figure 5) show that they are mostly consistent with **low-mass RGB/AGB stars.**