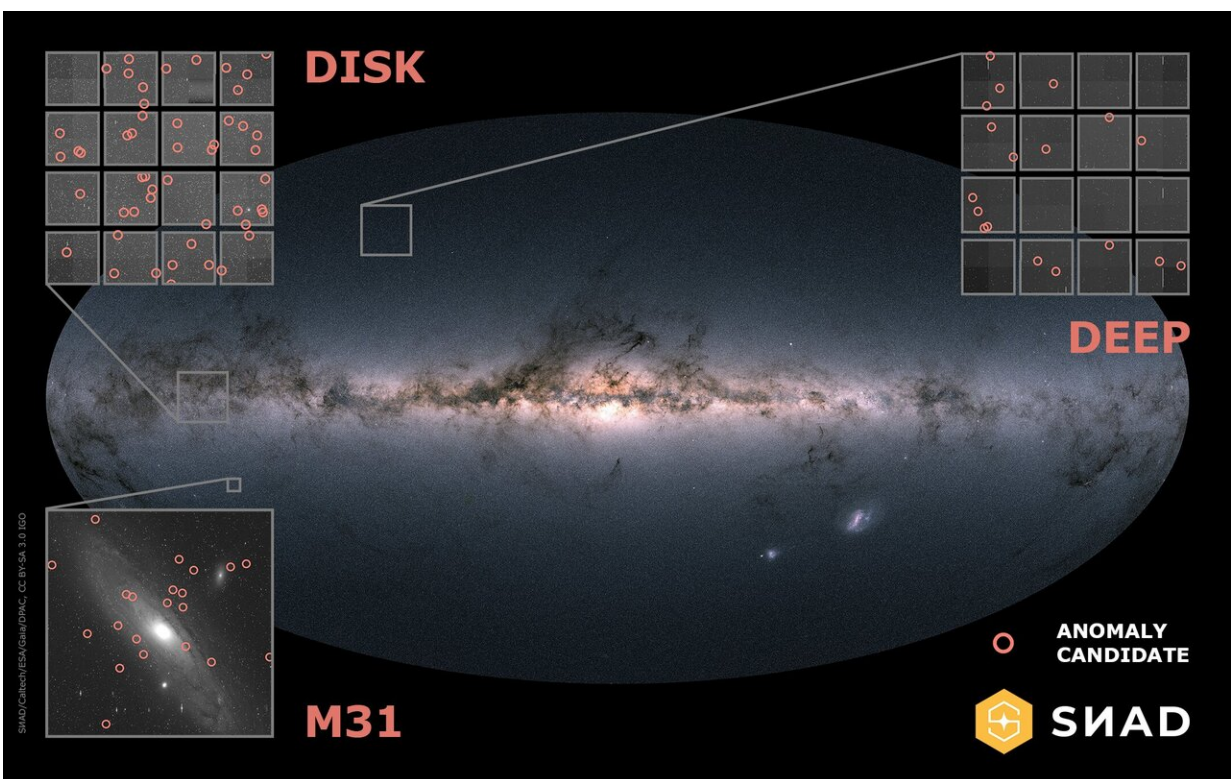


A new anomaly detection pipeline for astronomical discovery and recommendation systems

February 10 2021, by Patrick Aleo



Location of three ZTF fields analysed in this work with marked anomaly candidates. Credit: Maria Pruzhinskaya (2020)

The SNAD team, an international network formed by researchers from Russia, France and the U.S., has developed a pipeline to find rare and

exotic objects among the haystacks of data from astronomical surveys.

Given the ever increasing size of astronomical data sets, even if our telescopes do detect unexpected interesting astronomical phenomena, it is very unlikely that we will be able to recognize them in the middle of millions or even billions of observations. The solution lies in automatic tools specifically designed to recognize unusual behaviors hidden among billions of measurements. Some of these tools already exist and are employed, for example, to identify fraud credit card activities among millions of transactions every day. However, their adaptation to scientific data is not straightforward due to complications risen from the nature of observations in astronomy. The SNAD team has been working for 3 years in the development and adaptations of such solutions to the context of astronomy.

During their [last annual meeting](#), the group focused their efforts on objects whose brightness varies in time. The pipeline combines the strengths of machine learning algorithms and the irreplaceable knowledge from human experts to build a robust anomaly detection tool. The article describes results from applying this framework to the third data release of the Zwicky Transient Facility. Its three-stage process involved feature extraction on light curves (which tracks the brightness of objects over time), search for anomaly candidates using several machine learning algorithms and manually filtering of candidates by a human expert. This last stage also included performing observations with other telescopes whenever possible. In this study, 4 automatic learning algorithms were used to flag 277 anomaly candidates for human investigation—out of an initial data set of 2.25 million objects.

The group also developed a specially designed [web interface](#) which allowed immediate visualization and cross-match of each candidate with existing astronomical catalogs. This was constructed in order to facilitate the work of the experts who need to correlate the anomaly candidates

with any other publicly available information about the sky coordinates under investigation.

From the 277 objects considered as anomalous by the machine, 188 (68%) were found to display unusual features due to non-astrophysical effects (including defects due to ZTF's image subtraction pipeline), 66 (24%) were objects already cataloged before and 23 (8%) were previously unknown objects. The first category includes some amusing curiosities and the two latter cases of scientific interest. For example, one [object](#) flagged as anomaly by the machine was actually the [occultation of a background star by the Barcelona asteroid](#), which from the point of view of an observer from Earth was detected as a variable point source when in reality neither the star nor the asteroid actually changed brightness. The authors also characterized reoccurring and exotic image subtraction artifacts which interfere with light curve analysis and can trick an anomaly detection pipeline into thinking it is a real, anomalous object. In order to help quickly sort the first class from the remaining candidates, they were able to identify a simple bi-dimensional relation which can be used to aid in filtering potentially bogus light curves in future studies.

Among the second and third categories, the authors found four [supernovae](#) candidates, six previously unclassified [eclipsing binaries](#), four [pre-main-sequence](#) candidates, one possible [red dwarf](#) flare, and spectroscopically confirmed a [RS Canum Venaticorum](#) star, among other anomaly candidates.

Quickly and effortlessly separating artifacts from interesting [anomaly](#) candidates are crucial for current and soon-approaching next generation observatories, such as the Vera Rubin Observatory Legacy Survey of Space and Time (LSST). LSST will generate roughly 10 million transient sources per night—sophisticated and robust algorithms will be needed to sift through all that data so unexpected and interesting objects are not

missed, and scientists can better understand these space oddities.

Lead author Konstantin Malanchev, researcher at the University of Illinois at Urbana-Champaign (U.S.) and the Sternberg astronomical institute of the Lomonosov Moscow (Russia), says, "Designing specifically dedicated tools to search for astrophysically interesting anomalies is our only option to ensure the full exploitation of data sets we fought so hard to acquire. The SNAD team is fully committed to help the astronomical community in exploring the full potential of future data sets."

The [article](#) has been accepted for publication in *Monthly Notices of the Royal Astronomical Society* and is also publicly available as a [pre-print](#). The [source code](#) and results, including a complete list of objects with potential scientific application, as well as the pipeline techniques, are open to the public for the benefit of and verification by the astronomical community.

More information: K L Malanchev et al. Anomaly detection in the Zwicky Transient Facility DR3, *Monthly Notices of the Royal Astronomical Society* (2021). [DOI: 10.1093/mnras/stab316](https://doi.org/10.1093/mnras/stab316)

Provided by SNAD research network

Citation: A new anomaly detection pipeline for astronomical discovery and recommendation systems (2021, February 10) retrieved 26 April 2024 from <https://phys.org/news/2021-02-anomaly-pipeline-astronomical-discovery.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--