

AAS-PROVIDED PDF • OPEN ACCESS

reLAISS: A Python Package for Flexible Similarity Searches of Supernovae and Their Host Galaxies

To cite this article: E. Reynolds *et al* 2025 *Res. Notes AAS* **9** 189

Manuscript version: AAS-Provided PDF

This AAS-Provided PDF is © 2025 **The Author(s)**. Published by the American Astronomical Society.



Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence

<https://creativecommons.org/licenses/by/4.0>

Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required.

View the [article online](#) for updates and enhancements.

DRAFT VERSION JULY 14, 2025

Typeset using L^AT_EX R^NAAS style in AAS_{TeX}631**reLAISS: A Python Package for Flexible Similarity Searches of Supernovae and Their Host Galaxies**E. Reynolds,¹ A. Gagliano,^{2,1,3} and V. A. Villar^{1,2}¹*Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA 02138, USA*²*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*³*Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA***ABSTRACT**

Discovery rates of supernovae are expected to surpass one million events annually with the Vera C. Rubin Observatory. With unprecedented sample sizes of both common and rare transient types, photometric classification alone will be insufficient for finding one-in-a-million events and prioritizing the 1% of events for spectroscopic follow-up observations. Here, we present **reLAISS**, a modified framework for similarity searches of supernovae using extracted features of ZTF light curves and Pan-STARRS host galaxy photometry and built on the original LAISS framework. Unlike its predecessor, **reLAISS** couples interpretable light curve morphology features with extinction-corrected host-galaxy colors to probe both explosion physics and associated stellar populations simultaneously. The library allows users to customize the number of neighbors retrieved, the weight of host and light curve features, and the use of Monte Carlo simulations to ensure relevant matches when features are poorly constrained. We release **reLAISS** as a pip-installable package with an accompanying reference set of 20,000 features, and a set of tutorials that demonstrate the code’s expanded functionality. All source code can be found at the project’s GitHub repository: <https://github.com/evan-reynolds/re-laiss>.

Keywords: Light curves(918) — Time domain astronomy(2109)

INTRODUCTION

Wide-field surveys such as the Zwicky Transient Facility (Bellm et al. 2019) and the Young Supernova Experiment (Jones et al. 2021) have cataloged the explosions of stars as supernovae (SNe) at unprecedented rates. The classification of SNe by spectroscopy and photometry has become an indispensable tool for understanding discovered events and coordinating follow-up observations. The suite of available algorithms now employ template matching (Rodney & Tonry 2009), feature extraction (de Soto et al. 2024), and deep learning (Villar et al. 2020; Boone 2021) to distinguish between the explosions of old and young stars and identify energy inputs beyond ⁵⁶Ni.

As sample sizes of both common and rare types continue to swell, there is increasing demand for algorithms that can identify relevant SNe directly from the observed data instead of relying exclusively on our existing taxonomy. This may include finding in-class outliers (e.g., atypically luminous Type Ia SNe) that may signify the extremes of an evolutionary pathway. Though photometry and spectroscopy offer the most direct probe of an SN’s progenitor system and underlying physics, an event’s host galaxy provides broad constraints. The observational properties of a galaxy tightens the prior on the stellar population that produced a transient where event data is minimal, and transient populations obtained from host galaxy selection cuts can facilitate comparative studies of SN sub-types (Burgaz et al. 2025).

In this work, we present **reLAISS** (Reynolds et al. 2025), a library for supernova similarity searches. **reLAISS** expands upon the initial Light curve Anomaly Identification and Similarity Search (LAISS) library constructed by Aleo et al. (2024) and enables flexible and scalable searches of transients of interest. We outline the major modifications to the library from the initial publication and provide examples of its intended use cases in the following section.

METHODOLOGY

reLAISS matches a queried transient by the observational features of its photometry and, optionally, its host galaxy. Given a transient ZTFID, **reLAISS** first queries the ANTARES broker¹ for the alert photometry in ZTF- g and ZTF- r filters (a user-provided photometry file can also be provided). In each filter, the code calculates the peak magnitude, time from first observation to peak, rise and decline times to half-flux, duration above half-flux, and the rolling variance. The $g - r$ color at peak, mean $g - r$ across all detections, and mean change in $g - r$ are also calculated using linear interpolation to align observations to a consistent phase. Finally, the median local curvature of photometry in each band is calculated in the 20 days prior to peak and the 20 days following peak, where the curvature is estimated by the centered second finite difference of the magnitudes:

$$\kappa_i = \frac{m_{i+1} - 2m_i + m_{i-1}}{\left[\frac{1}{2}(t_{i+1} - t_{i-1})\right]^2} \quad (1)$$

The above features, which replace the **light-curve**-calculated features originally used in LAISS (Aleo et al. 2024), strike an optimal balance between interpretability and flexibility.

After extracting light curve features, **reLAISS** identifies the most likely host galaxy of the transient in the Pan-STARRS 3 π catalog using the **Prost** library², which replaces the **GHOST** host association code used by LAISS (Gagliano et al. 2021). **Prost** (Gagliano 2025; Gagliano et al. 2025) estimates the posterior probability of association using a Monte Carlo simulation over observed galaxy properties and user-defined priors for the true host’s redshift, fractional offset, and absolute magnitude. Following association, **reLAISS** calculates the extinction-corrected Kron magnitude of the host galaxy in Pan-STARRS $griz$ and the $g - r$, $r - i$, and $i - z$ colors of the galaxy (which are independently weighted to de-prioritize outliers from catastrophic failures in color estimation).

As with LAISS, **ANNOY**³ is used to build an indexed space of reference features and conduct approximate nearest-neighbors searches via locality-sensitive hashing (details on the algorithms used for indexing and retrieval are given in Aleo et al. 2024). We provide an updated data set of 20,000 archival light curve and host galaxy features for querying. Measured uncertainties in the queried object’s extracted features can be used to re-query and re-rank retrieved neighbors.

A user can choose to find neighbors by light curve properties, host properties, or both feature sets. If all properties are chosen, a weighting term can be input that allows the user to prioritize better matches along one feature set than the other. The user can define the number of neighbors to retrieve; if none are specified, the **kneed** package (Satopaa et al. 2011) is used to determine the inflection point in neighbor distances versus number of neighbors and a number of relevant neighbors is suggested.

reLAISS can be used to find similar events to an observed transient or the closest matches to a theoretical model of interest. A user can also manually specify an observed host galaxy, in cases where a theoretical model is queried and correlations with a particular stellar population are desired. We now fit a Generalized Pareto distribution to the upper-tail of the nearest-neighbor distance distribution of the reference data bank; the real-time feature probabilities associated a queried transient are converted with a sigmoid transformation to an anomaly score between 0 and 100. This replaces the previous random forest algorithm within LAISS for real-time anomaly detection. Future work will be devoted to investigating the robustness of the chosen feature set and anomaly detection method for partial-phase and low-cadence light curves, and extending the features to other photometric passbands and surveys beyond ZTF.

ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, <http://iaifi.org/>).

¹ <https://antares.noirlab.edu/>

² <https://github.com/alexandergagliano/Prost>

³ <https://github.com/spotify/annoy>

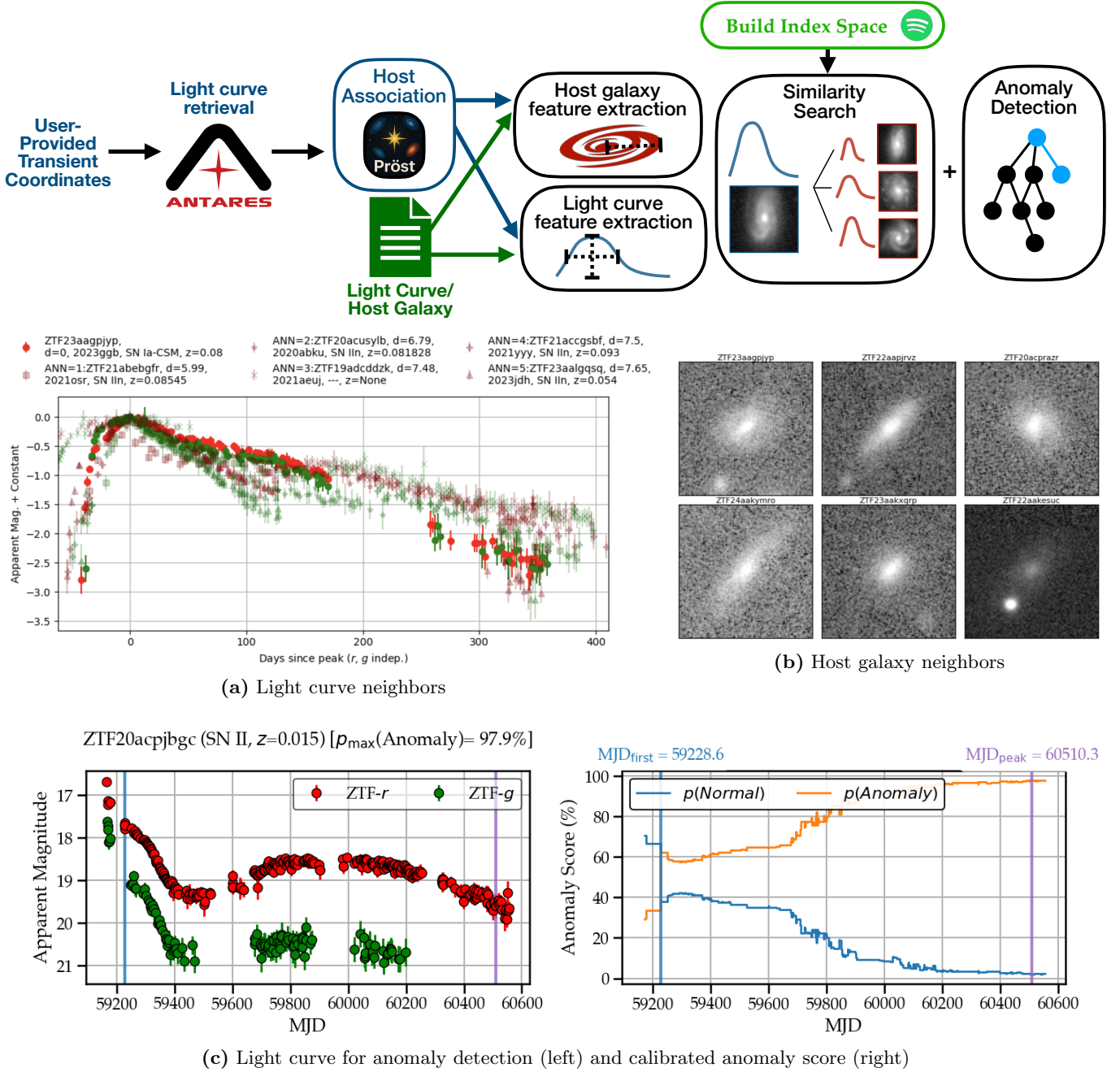


Figure 1. Top: Workflow of the reLAISS library. Starting from user-provided transient coordinates or an optional light curve file, the pipeline (i) identifies the most probable Pan-STARRS host using ProSt, (ii) extracts a suite of interpretable light curve and host-galaxy features, (iii) indexes and queries these features with locality-sensitive hashing via ANNOY, and (iv) returns nearest neighbors and highlights anomalies for follow-up. **Middle:** Example nearest neighbor searches. Panel (a) overlays light curves of neighbors for ZTF23aagpjyp, an SN Ia-CSM, identified using light curve features only. Panel (b) charts neighbor host galaxies of the same transient identified using host galaxy features only. Each search used 20 monte carlo simulations without PCA. **Bottom:** Example anomaly detection for anomalous transient. Panels (c) and (d) show the light curve of transient ZTF20acpjbgc (a long-duration SN II) and its time-evolving anomaly score.

REFERENCES

- Aleo, P. D., Engel, A. W., Narayan, G., et al. 2024, *ApJ*, 974, 172, doi: [10.3847/1538-4357/ad6869](https://doi.org/10.3847/1538-4357/ad6869)
- Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, *PASP*, 131, 018002, doi: [10.1088/1538-3873/aaecbe](https://doi.org/10.1088/1538-3873/aaecbe)
- Boone, K. 2021, *AJ*, 162, 275, doi: [10.3847/1538-3881/ac2a2d](https://doi.org/10.3847/1538-3881/ac2a2d)
- Burgaz, U., Maguire, K., Dimitriadis, G., et al. 2025, *A&A*, 694, A13, doi: [10.1051/0004-6361/202452571](https://doi.org/10.1051/0004-6361/202452571)
- de Soto, K. M., Villar, V. A., Berger, E., et al. 2024, *ApJ*, 974, 169, doi: [10.3847/1538-4357/ad6a4f](https://doi.org/10.3847/1538-4357/ad6a4f)
- Gagliano, A. 2025, in preparation
- Gagliano, A., de Soto, K., Boesky, A., & Manning, T. A. 2025, alexandergagliano/Prost, Version: v1.2.11, Zenodo, doi: [10.5281/zenodo.15397886](https://doi.org/10.5281/zenodo.15397886)
- Gagliano, A., Narayan, G., Engel, A., Carrasco Kind, M., & LSST Dark Energy Science Collaboration. 2021, *ApJ*, 908, 170, doi: [10.3847/1538-4357/abd02b](https://doi.org/10.3847/1538-4357/abd02b)
- Jones, D. O., Foley, R. J., Narayan, G., et al. 2021, *ApJ*, 908, 143, doi: [10.3847/1538-4357/abd7f5](https://doi.org/10.3847/1538-4357/abd7f5)
- Reynolds, E., Gagliano, A., & Villar, V. A. 2025, RE-LAISS: Robust Similarity Search for Supernovae, v1.2.3, Zenodo, doi: [10.5281/zenodo.15874800](https://doi.org/10.5281/zenodo.15874800)
- Rodney, S. A., & Tonry, J. L. 2009, *ApJ*, 707, 1064, doi: [10.1088/0004-637X/707/2/1064](https://doi.org/10.1088/0004-637X/707/2/1064)
- Satopaa, V., Albrecht, J., Irwin, D., & Raghavan, B. 2011, in 2011 31st International Conference on Distributed Computing Systems Workshops (IEEE), 166–171, doi: [10.1109/ICDCSW.2011.20](https://doi.org/10.1109/ICDCSW.2011.20)
- Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020, *The Astrophysical Journal*, 905, 94, doi: [10.3847/1538-4357/abc6fd](https://doi.org/10.3847/1538-4357/abc6fd)