The Young Supernova Experiment Data Release 1 (YSE DR1): Light Curves and Photometric Classification of 1975 Supernovae

P. D. Aleo[1,2,36], K. Malanchev[1,3], S. Sharief[1], D. O. Jones[4,5,37], G. Narayan[1,2], R. J. Foley[4], V. A. Villar[6,7,8],
C. R. Angus[9], V. F. Baldassare[10], M. J. Bustamante-Rosell[4], D. Chatterjee[1,2,11], C. Cold[9], D. A. Coulter[4],
K. W. Davis[4], S. Dhawan[12], M. R. Drout[13], A. Engel[1], K. D. French[1,2], A. Gagliano[1,2,35], C. Gall[9], J. Hjorth[9],
M. E. Huber[14], W. V. Jacobson-Galán[15], C. D. Kilpatrick[4,16], D. Langeroodi[9], P. Macias[4], K. S. Mandel[12],
R. Margutti[15], F. Matasić[1], P. McGill[4], J. D. R. Pierel[17], E. Ramirez-Ruiz[4], C. L. Ransome[6,7], C. Rojas-Bravo[4],
M. R. Siebert[4,17], K. W. Smith[18], K. M. de Soto[6,7,8], M. C. Stroh[16], S. Tinyanont[4], K. Taggart[4], S. M. Ward[12],
R. Wojtak[9], K. Auchettl[4,19,20], P. K. Blanchard[16], T. J. L. de Boer[14], B. M. Boyd[12], C. M. Carroll[10,34],
K. C. Chambers[14], L. DeMarchi[16], G. Dimitriadis[4,21], S. A. Dodd[4], N. Earl[1], D. Farias[9], H. Gao[14], S. Gomez[17],
M. Grayling[12], C. Grillo[22], E. E. Hayes[12], T. Hung[4], L. Izzo[9], N. Khetan[9], A. N. Kolborg[4,9],
J. A. P. Law-Smith[23], N. LeBaron[15], C.-C. Lin[14], Y. Luo[1,24], E. A. Magnier[14], D. Matthews[15], B. Mockler[25,26],
A. J. G. O'Grady[13,27], Y.-C. Pan[28], C. A. Politsch[12], S. I. Raimundo[9,25,29], A. Rest[17,30], R. Ridden-Harper[31],
A. Sarangi[9], S. L. Schrøder[9], S. J. Smartt[18,32], G. Terreran[16], S. Thorp[12], J. Vazquez[1], R. J. Wainscoat[14],
Q. Wang[30], A. R. Wasserman[1,2,36], S. K. Yadavalli[6,7], R. Yarza[4,33], and Y. Zenati[30,38]
(Young Supernova Experiment)

[1] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA; paleo2@illinois.edu
[2] Center for AstroPhysical Surveys, National Center for Supercomputing Applications, Urbana, IL 61801, USA
[3] Sternberg Astronomical Institute, Lomonosov Moscow State University, 13 Universitetsky pr., Moscow 119234, Russia
[4] Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA
[5] Gemini Observatory, NSF's NOIRLab, 670 N. A'ohoku Place, Hilo, HI 96720, USA
[6] Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA
[7] Institute for Computational & Data Sciences, The Pennsylvania State University, University Park, PA 16802, USA
[8] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
[9] DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 128, DK-2200 Copenhagen, Denmark
[10] Department of Physics & Astronomy, Washington State University, Pullman, WA 99164, USA
[11] LIGO Laboratory and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 185 Albany Street, Cambridge, MA 02139, USA
[12] Institute of Astronomy and Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK
[13] David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario, M5S 3H4 Canada
[14] Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
[15] Department of Astronomy, University of California, Berkeley, CA 94720, USA
[16] Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA
[17] Space Telescope Science Institute, Baltimore, MD 21218, USA
[18] Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK
[19] School of Physics, The University of Melbourne, VIC 3010, Australia
[20] ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
[21] School of Physics, Trinity College Dublin, The University of Dublin, Dublin 2, Ireland
[22] Dipartimento di Fisica, Università degli Studi di Milano, via Celoria 16, I-20133 Milano, Italy
[23] Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA 02138, USA
[24] Department of Physics & Astronomy, University of Wyoming, Laramie, WY 82070, USA
[25] Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA
[26] Observatories of the Carnegie Institute for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA
[27] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON, M5S 3H4, Canada
[28] Graduate Institute of Astronomy, National Central University, 300 Zhongda Road, Zhongli, Taoyuan 32001, Taiwan
[29] Department of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK
[30] Physics and Astronomy Department, Johns Hopkins University, Baltimore, MD 21218, USA
[31] School of Physical and Chemical SciencesTe Kura Matu, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
[32] Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK
[33] Texas Advanced Computing Center, University of Texas, Austin, TX 78759, USA

---

[34] LSSTC Catalyst Fellow.
[35] NSF Graduate Fellow.
[36] Center for AstroPhysical Surveys (CAPS) Fellow.
[37] NASA Einstein Fellow.
[38] ISEF Postdoc Fellow.

Abstract

We present the Young Supernova Experiment Data Release 1 (YSE DR1), comprised of processed multicolor PanSTARRS1 *griz* and Zwicky Transient Facility (ZTF) *gr* photometry of 1975 transients with host–galaxy associations, redshifts, spectroscopic and/or photometric classifications, and additional data products from 2019 November 24 to 2021 December 20. YSE DR1 spans discoveries and observations from young and fast-rising supernovae (SNe) to transients that persist for over a year, with a redshift distribution reaching $z \approx 0.5$. We present relative SN rates from YSE's magnitude- and volume-limited surveys, which are consistent with previously published values within estimated uncertainties for untargeted surveys. We combine YSE and ZTF data, and create multisurvey SN simulations to train the ParSNIP and SuperRAENN photometric classification algorithms; when validating our ParSNIP classifier on 472 spectroscopically classified YSE DR1 SNe, we achieve 82% accuracy across three SN classes (SNe Ia, II, Ib/Ic) and 90% accuracy across two SN classes (SNe Ia, core-collapse SNe). Our classifier performs particularly well on SNe Ia, with high (>90%) individual completeness and purity, which will help build an anchor photometric SNe Ia sample for cosmology. We then use our photometric classifier to characterize our photometric sample of 1483 SNe, labeling 1048 (∼71%) SNe Ia, 339 (∼23%) SNe II, and 96 (∼6%) SNe Ib/Ic. YSE DR1 provides a training ground for building discovery, anomaly detection, and classification algorithms, performing cosmological analyses, understanding the nature of red and rare transients, exploring tidal disruption events and nuclear variability, and preparing for the forthcoming Vera C. Rubin Observatory Legacy Survey of Space and Time.

Unified Astronomy Thesaurus concepts: Supernovae (1668); Surveys (1671); Astronomy databases (83); Light curves (918); Light curve classification (1954)

Supporting material: machine-readable tables

1. Introduction

In the past decade, time-domain astrophysics has spearheaded astronomy's big data revolution due to the rousing success of wide-field surveys. Such efforts have enabled the community to discover a combined $\sim 10^4$ new supernovae (SNe) and other optical transients yearly, and the community has used these discoveries to better understand transient and variable phenomena on short timescales. The discovery rate will soon accelerate dramatically—the Vera C. Rubin Observatory's (Rubin) Legacy Survey of Space and Time (LSST) will discover $\sim 10^6$ transients yearly (LSST Science Collaboration et al. 2009; Ivezić et al. 2019), and will usher in a new wave of study of transient physics and progenitor discovery. In anticipation of Rubin, there has been active development in two key areas: (1) smaller-scale time-domain multiwavelength surveys or fast-cadence transient searches, and (2) photometric classifiers—algorithms that classify transients based on their observed light curves and contextual information, such as redshift.

Systematic searches for SNe began in the late 1990s, but the announcement of LSST in 2009[39] (LSST Science Collaboration et al. 2009) became the driver for increased interest in time-domain surveys. Such surveys are broadly distinguished by systematic searches for low-redshift SNe (Catalina Real-Time Transient Survey; Drake et al. 2009; Palomar Transient Factory survey, hereafter PTF; Law et al. 2009; CHASE; Pignata et al. 2009; MASTER; Lipunov et al. 2010; ATLAS; Tonry 2011; Tonry et al. 2018; La Silla QUEST; Baltay et al. 2013; ASAS-SN; Shappee et al. 2014; PSST; Huber et al. 2015; DLT40; Valenti et al. 2017; Zwicky Transient Facility, hereafter ZTF; Bellm et al. 2019; Young Supernova Experiment, hereafter YSE; Jones et al. 2021a) and high-redshift SNe (Dark Energy Survey; Bernstein et al. 2012; Dark Energy Survey Collaboration et al. 2016; Abbott et al. 2019; Brout et al. 2019; Pan-STARRS Medium-Deep Survey,

hereafter PS1-MDS; Rest et al. 2014; Jones et al. 2018; Villar et al. 2020; Hubble Space Telescope, hereafter HST, surveys CANDELS, CLASH, and the Frontier Fields; Graur et al. 2014; Rodney et al. 2014; Kelly et al. 2015; Subaru Hyper Suprime-Cam Transient Survey; Tanaka et al. 2016). Other identifying factors are fast-cadence searches, including the ZTF 1 day survey, Kepler (K2; Howell et al. 2014), the Evryscope (Law et al. 2015), the Korea Microlensing Telescope Network (Kim et al. 2016), DLT40 (Valenti et al. 2017), TESS (Fausnaugh et al. 2021), and wavelength regimes beyond the optical (near-infrared via the Vista Infrared Extragalactic Legacy Survey, Hönig et al. (2017); and ultraviolet via the Galaxy Evolution Explorer (GALEX) time-domain survey (Gezari et al. 2013). Data from these surveys helped in our understanding of SN explosion mechanisms and local rates, SN siblings, host galaxy properties, black holes, stellar evolution, and the expansion of our universe (Nomoto et al. 2013; Riess et al. 2016; Abbott et al. 2017; Scolnic et al. 2018; Fremling et al. 2020; Gagliano et al. 2021; Graham et al. 2022).

Despite such progress, key questions remain unanswered due to the limitations of coverage, area, cadence, depth, or photometric calibration in current surveys. YSE was created in an effort to answer such questions regarding young, red, or rare transients, cosmological parameters, and transient phenomena in the local universe. YSE is a 3+ yr transient survey, which began operations 2019 November 24 using a 7% time allocation on Pan-STARRS1 (PS1) to survey $\sim 750 \, \mathrm{deg}^2$ of sky (with full survey operations to observe 1512 deg^2) with a planned 3 day cadence to a depth of $gri \approx 21.5$ mag, and $z \approx 20.5$ mag. YSE survey strategy emphasizes increased redder wavelength coverage (*iz*), combined with excellent photometric calibration of PS1 and improved depth, to demarcate our discovery demographics from other active time-domain surveys. When possible, YSE interleaves observations with those of ZTF for an improved effective cadence to identify young and fast-evolving transients. YSE has access to several spectroscopic facilities for prompt follow-up studies.

---

[39] https://www.lsst.org/sites/default/files/docs/sciencebook/SB_Whole.pdf

With currently limited spectroscopic resources, it is projected that ∼0.1% of transients will be spectroscopically classified in the LSST era (Hsu et al. 2022). The dearth of spectroscopic resources—and subsequently known classification labels—has been the driving force to develop and apply photometric classifiers to active and future surveys: including ZTF (Muthukrishna et al. 2019; Nordin et al. 2019), PS1-MDS (Villar et al. 2019; Hosseinzadeh et al. 2020; Villar et al. 2020), YSE (this work), and LSST via the Photometric Astronomical Time Series Classification Challenge (PLAsTiCC, Hložek et al. 2020), the Extended LSST Astronomical Time-series Classification Challenge (ELAsTiCC),[40] and The Simulated Catalogue of Optical Transients and Correlated Hosts (Lokken et al. 2023).

Although the overwhelming majority of photometric classification efforts invoke machine-learning algorithms and deep-learning architectures, there is a variety of approaches for common tasks. When extracting light-curve features, some fit an empirical functional form (Bazin et al. 2009; Karpenka et al. 2012; Villar et al. 2019; Sánchez-Sáez et al. 2021); some estimate a smooth approximation to the light curve using Gaussian process interpolation (Lochner et al. 2016; Boone 2019; Alves et al. 2022), multilayer perceptrons (MLPs; Demianenko et al. 2022), normalizing flows (Demianenko et al. 2022); some apply neural networks such as temporal convolutional networks (Muthukrishna et al. 2019), recurrent neural networks (Charnock & Moss 2017; Möller et al. 2021; Gagliano et al. 2022a), convolutional neural networks (Pasquet et al. 2019b; Qu et al. 2021; Burhanudin & Maund 2023), Bayesian neural networks (Demianenko et al. 2022), variational autoencoders (VAEs; Villar et al. 2020; Boone 2021); or finally, some use a mix of the above. In this work, we use VAEs (Kingma & Welling 2013) from the literature; a VAE model approximates the input's posterior distribution over the (low-dimensional) latent space using variational inference called an "encoder," from which a generative model ("decoder") reconstructs the input given a position in latent space. Training is performed by comparing a given input to the reconstructed one via the encoding and decoding process.

Despite the advancement of sophisticated algorithms, extracted light-curve features are highly dependent on observational properties that affect the light-curve profile. One such dependency is redshift, which becomes problematic for photometric classification tasks because labeled data sets (including YSE DR1) are often biased toward bright, low-redshift transients (Lochner et al. 2016; Boone 2019). Moreover, with a lack of SNe or host galaxy spectroscopic redshifts (spec-z), photometric classifiers are reliant on photometric redshift estimates (photo-z), which remains an area of active research (see, e.g., Beck et al. 2016; Pasquet et al. 2019a; Salvato et al. 2019; Tarrío & Zarattini 2020; Beck et al. 2021; Schuldt et al. 2021, and references therein). We require photometric classifiers that perform well using photo-z estimates and are robust to fainter, poorly sampled events. Thus, we adopt a photometric classifier that attempts to disentangle the intrinsic properties of the transients from any extrinsic observational effects like redshift, dust along the line of sight, varying cadences, and observations in different passbands across multiple telescopes: the parameterization of supernova intrinsic properties ("ParSNIP") classifier from

Boone (2021). ParSNIP is a hybrid physics–VAE architecture that models the intrinsic time varying spectra of transients combined with an explicit physical model for observational effects.

We compare the performance of ParSNIP to that of another state-of-the-art autoencoder-based architecture called Super-RAENN (Villar et al. 2020). Both networks are exclusively trained on simulated YSE SNe and interleaving simulated ZTF data (rather than trained on observed SNe) with mock photo-zs. In keeping with the literature convention, our classifier is optimized for general SNe classification across the three most common classes: thermonuclear SNe Ia, core-collapse (CC) types SNe II, and stripped envelope SNe Ib/Ic (which due to their small sample sizes we consider as one holistic SNe Ibc class). We provide performance results on the test set of 472 spectroscopically confirmed SNe. Moreover, we characterize the 1975 YSE DR1 SN-like transients. We publish the processed multicolor PS1-$griz$ and ZTF-gr photometry of 1975 transients with host galaxy associations, redshifts, spectroscopic and/or photometric classifications, and additional data products from 2019 November 24 to 2021 December 20 for community use on Zenodo (Aleo et al. 2022; doi:10.5281/zenodo.7317476).

This manuscript is structured as follows. In Section 2, we overview the YSE survey strategy. In Section 3, we remark on the complementary ZTF survey observations. In Section 4, we describe the data products and transient demographics of YSE DR1. In Section 5, we describe the photometric classification methodology from data reduction, host association, simulation generation, ParSNIP architecture, and classifier training. In Section 6, we examine the results of our classifier performance. In Section 7, we discuss the impact of our classifier on YSE DR1 and YSE science drivers, including future work. We conclude in Section 8.

2. YSE Survey Overview and Strategy

Here we give an overview of the Pan-STARRS telescopes and photometric system, and outline YSE survey characteristics.

2.1. YSE Instrumentation

YSE is a 3+ yr optical time-domain survey observing with the 1.8 m PS1 and Pan-STARRS2 (PS2) telescopes. Both are equipped with 1.4 gigapixel cameras (GPC1 and GPC2; Kaiser et al. 2002) to provide a 7 deg^2 field of view (FOV). PS1 has imaged over 3π sr of the sky since formal survey operation began in 2010 May (Chambers et al. 2016). One of the strengths of the Pan-STARRS telescopes is their excellent relative and absolute photometric calibration; Schlafly et al. (2012) achieved a relative precision of <10 mmag in gri_{P1} and ∼10 mmag in zy_{P1}, Scolnic et al. (2015) further improved PS1's absolute calibration by comparing secondary standard stars across various SN samples' photometric systems (PS1, Supernova Legacy Survey, Sloan Digital Sky Survey, Carnegie Supernova Project, and Center for Astrophysics Redshift Surveys (1)–(4)), and Brout et al. (2022) leveraged the Pan-STARRS stellar photometry catalog to cross-calibrate against tertiary standards for recalibrating photometric systems used in the Pantheon+ SNe Ia sample. This impressive calibration is imperative for transient science and precise SN cosmology, a key science goal for YSE and an area of current research.

Currently, YSE has 7% total observing time allocation on PS1 and PS2 (but only PS1 data is in YSE DR1) to scan 750 deg^2 of sky with 3 day's cadence in four[41] broadband filters $griz_{P1}$ (hereafter PS1-$griz$). YSE has flat PS1-$griz$ exposure times of 27 s. Although YSE recently began observing with PS2 starting January 2022, this date is after the data collection cutoff for this work.[42] Thus, all YSE DR1 data presented here are from PS1 in the period 2019 November 24–2021 December 20. For a more in-depth discussion of the YSE survey and the Pan-STARRS telescopes, see Jones et al. (2021a) and Chambers et al. (2016), respectively.

2.2. Filter Strategy

YSE observes in two passbands per epoch. This allows for a large survey area while retaining transient color information (Jones et al. 2021a). During dark time (Moon illumination between 0% and 33%), YSE alternates between PS1-gr and PS1-gi observations. During bright time (Moon illumination between 66% and 100%), YSE alternates between PS1-ri and PS1-rz observations due to the ∼1 mag shallower PS1-g depth. During brief periods of gray time (Moon illumination between 33% and 66%), YSE alternates between PS1-gi and PS1-gz observations. This strategy guarantees at least one PS1-g or PS1-r observation per epoch while prioritizing PS1-iz filters. The former is important for helping measure the rise of young SNe and comparing to ZTF-gr data without making explicit assumptions on the color. The latter is crucial to the discovery of red and intrinsically faint transients typically missed by blue-sensitive surveys.

YSE was estimated to achieve approximate single-visit depths of ∼21.5 mag in PS1-gri and ∼20.5 mag in PS1-z (Jones et al. 2021a). Empirically, we find deeper approximate single-visit depths to a limit of ∼22.2 mag in PS1-g, ∼22.1 mag in PS1-r, ∼22.0 mag in PS1-i, and ∼21.6 mag in PS1-z. Predicted to be ∼0.4–0.8 mag deeper than ZTF-gr in dark time (Jones et al. 2021a), we find YSE observations are closer to ∼0.4–0.5 mag deeper than single-visit depth limits of ZTF: ∼21.8 mag in ZTF-g and ∼21.6 mag in ZTF-r. See Table 2 of Jones et al. (2021a) for a direct comparison of YSE survey characteristics with other active time-domain surveys such as ATLAS, ASAS-SN, PSST, and ZTF.

2.3. Field Selection

YSE's field selection criteria is explained in detail in Section 3 of Jones et al. (2021a). To summarize, YSE prioritizes fields with the following: high Galactic latitude, low Milky Way extinction, substantial archival data, an advantageous position near or on the equatorial plane (as equatorial fields can be observed with follow-up facilities from both hemispheres), decl. >30°, overlapping ZTF fields, rising and/or scientifically interesting transients from other surveys that would benefit greatly from YSE observations, and a larger number of nearby galaxies within 150 Mpc (particularly galaxies at <10 Mpc).

In addition to the general field selection criteria, YSE also dedicates two Pan-STARRS pointings to the Virgo Cluster when Virgo is observable at airmass <1.5. The patch of sky in and around the Virgo Cluster has proven to be a treasure trove of SNe over the past decade plus; the community has spectroscopically classified over 30 SNe within the radii of the two YSE Virgo pointings our team has adopted (not accounting for detector masking) in the past ∼15 yr. Moreover, YSE is capable of detecting preexplosion outbursts for Virgo transients to an approximate absolute magnitude of $M_{peak} \sim -10$. In YSE DR1, there are 117 transients in or adjacent to the Virgo fields.[43] Two well-observed examples are the nearby Virgo-adjacent SNe Ia 2020ue (Tinyanont et al. 2021) and 2020nlb (Sand et al. 2021), non-YSE discoveries with excellent YSE photometric coverage.

Targeted YSE observations and the criteria for moving an existing YSE survey field to target a new SN are outlined in Section 3.4 of Jones et al. (2021a). In general, YSE keeps roughly 50% of its fields static for long-term monitoring, and the other 50% flexible in order to follow interesting transients that meet YSE's science goals. In YSE DR1, we have 31 unique targeted objects: 20 normal SNe Ia, 1 SN Ia-91T–like, 1 SN Ib, 1 SN Ibn, 1 SN Ic, 3 SNe II, 2 SNe IIb, and 1 luminous blue variable (LBV).

2.4. Data Processing and YSE Forced Photometry

All YSE data from Pan-STARRS undergo basic data processing by the University of Hawaii's PS1 Image Processing Pipeline (IPP; Magnier et al. 2020a, 2020b). First, all Pan-STARRS images are ingested, processed, and archived by IPP; then undergo template image convolution and subtraction. PS1 template images are created from stacked exposures (primarily from the PS1 3π survey), from which new nightly images are resampled and astrometrically aligned to match a skycell in the PS1 sky tessellation. A nightly image zero-point is calculated by comparing PS1 stellar catalogs (Chambers et al. 2016) to the point-spread function (PSF) photometry. Template images are convolved with nightly images and matched to their PSF via a three-Gaussian kernel before being subtracted with HOTPANTS (Becker 2015).

Such data products are further processed by the Transient Science Server at Queens University Belfast (QUB; Section 4, Smith et al. 2020), which applies a combination of machine learning (ML) and catalog crossmatching to isolate new transient events in nightly images. Transient candidates have their photometry produced via the Photpipe (Rest et al. 2005, 2014) forced-photometry pipeline. For each epoch, a flux-weighted centroid is forced to be at the transient candidate position, and a nightly zero-point is applied to calculate the source's brightness. In tandem with the Transient Science Server and visual inspection, possible transients are sent to the TNS and are later ingested in the YSE collaboration's transient survey management platform, YSE-PZ[44] (Coulter et al. 2022).

---

[41] Some events have additional y_{P1} data that are not included in the data release, and we do not use these data for classification. Additional y_{P2} data have been taken in 2022 with Pan-STARRS2 operations, but this is past the data cutoff. Additionally, YSE does not observe in the w_{P1} passband.

[42] We note that the survey strategy remains largely the same, except that we have swapped PS2-z observations in favor of PS2-y observations, and we have added a new set of daily fields on part of the SDSS Stripe 82 region.

[43] For this calculation, we find all transients within 182° < R.A. < 191°, −4° < decl. < 20°.

[44] YSE-PZ ingests every transient reported to TNS, combines YSE transients with external data like ZTF and ATLAS, and stores follow-up data obtained by our team. YSE-PZ also enables the YSE survey to plan observations and exposes a powerful query engine to find, track, and follow-up targets of interest, as well as select scientific samples like YSE DR1. The code base is publicly available on GitHub at https://github.com/davecoulter/YSE_PZ, and we encourage collaboration and new contributors.

2.5. Magnitude- and Volume-limited Survey Strategy

The Lick Observatory Supernova Search (LOSS; Li et al. 2011) is a widely used reference for local rate measurements; however, LOSS performed a galaxy-targeted survey. Targeting massive galaxies inherently includes more passive elliptical galaxies than in an untargeted survey, which likely biased against finding CC SNe (Taubenberger 2017). Performing untargeted galaxy surveys to include the faint tail of the luminosity function (LF) has been proposed as a solution (e.g., Perley et al. 2020d), and the topic of rates and LFs biased by SNe–galaxy environment correlations has been abundantly discussed (Smith et al. 2007; Quimby et al. 2011; Sanders et al. 2012; Taggart & Perley 2021). Such biases that induce rate uncertainty include but are not limited to (1) highly reddened transients that are missed by blue-sensitive surveys; (2) SNe with short-lived progenitor stars that preferentially occur in dusty environments (Kelly & Kirshner 2012).

The observed SNe LF can be derived from either a magnitude- or volume-limited search if certain criteria are met (Li et al. 2011). For a volume-limited search, one needs the completed type, luminosity, and light-curve information for all SNe that constitute the sample (to fit a family of light curves and constrain the peak magnitudes). Magnitude-limited searches require similar information, but additionally need to observe deep enough to sample the faint end of the LF to correct for survey volumes containing SNe of varying brightnesses (e.g., Bazin et al. 2009).

YSE's criterion for the untargeted magnitude-limited sample is any transient that exceeds a peak apparent r magnitude brighter than 18.5 mag.[45] Meanwhile, YSE's volume-limited sample criterion is any transient that has a distance $D < 250$ Mpc. We share magnitude- and volume-limited SN fractions from YSE DR1 in Section 4.1.3. There, we compare to the magnitude-limited survey results of LOSS (Li et al. 2011), ASAS-SN (Holoien et al. 2019), and the ZTF Bright Transient Survey (ZTF BTS; Fremling et al. 2020), and the volume-limited survey results of LOSS (Li et al. 2011) and ZTF CLU (De et al. 2020). The construction of the untargeted YSE magnitude- and volume-limited samples provides another opportunity to help constrain observed rates and LFs of SNe classes without relying on galaxy associations, redshift catalogs, and assumed distances. Because YSE is a survey with greater sensitivity at redder wavelengths than other active surveys like ZTF, ATLAS, and ASAS-SN due to its PS1-iz coverage, there is an increased likelihood of discovering red or rare SNe and better represent them in the measured rates and LFs.

3. Complementing ZTF Observations

Here we briefly outline the instrumentation and data processing of ZTF observations of YSE DR1 transients.

3.1. ZTF Instrumentation

The ZTF survey[46] (Bellm et al. 2019) is the successor to the PTF (Law et al. 2009). ZTF is housed at the Palomar 48 inch Schmidt telescope, equipped with a 47 deg^2 FOV camera and an 8 s readout time, observing in three passbands: ZTF-g, ZTF-r, and ZTF-i. The Infrared Processing and Analysis Center (IPAC) provides ZTF image reduction and object identification in near real-time, producing transient alerts from raw images in ~4 minutes to be available to the community in the ZTF public alert stream via alert brokers such as Arizona-NOIRLab Temporal Analysis and Response to Events System[47] (Matheson et al. 2021), Automatic Learning for the Rapid Classification of Events (ALeRCE[48]; Förster et al. 2021), the Las Cumbres Observatory's Make Alerts Really Simple project,[49] Fink[50] (Möller et al. 2021), and Lasair.[51]

ZTF phase (I) started in 2018, transitioning into its phase (II) operations in 2020 December. ZTF phase (II) allocates 50% of camera time and 50% of Spectral Energy Distribution Machine (SEDM) spectrograph time to a 2 night cadence public survey of the entire northern sky in ZTF-g and ZTF-r bands. SEDM spectra are uploaded daily to the Transient Name Server (TNS),[52] and the forced PSF-fit photometry ("forced photometry") on ZTF difference images is now included in the alert packets (Patterson et al. 2019), and can also be requested directly through their ZTF forced-photometry service[53] (Masci et al. 2019). In this work, we augment the YSE forced photometry with ZTF-gr forced photometry when available. We do not use ZTF-i observations because of poor coverage (~10% of all ZTF observations) and the 18 month grace period for private survey data before public release.

3.2. ZTF Forced Photometry

In general, YSE attempts to organize its field selection observing schedule to precede ZTF observations by one calendar day in order to maximize discoveries of young SNe and increase the effective cadence for many shared YSE and ZTF transients. Planning such overlapping observations is possible due to the (now public) International Virtual Observatory Alliance Observation Locator Table Access Protocol.[54] Thus, with YSE acting as a precursor and complement to ZTF observations, we utilize ZTF forced photometry in photometric classification, and effectively create a combined YSE+ZTF survey.

For each YSE SN in YSE DR1, we request any available ZTF forced-photometry observations at the target position using the ZTF forced-photometry service (Masci et al. 2019). We supply the SN's R.A., decl., as well as the beginning Julian date (JD) corresponding to 5 months before the YSE survey (JD = 2458627.5; 2019 May 24) to conservatively catch any currently active SN at the start of the YSE survey (JD = 2458811.5; 2019 November 24), and the ending JD corresponding to the cutoff of YSE DR1 (JD = 2459568.5; 2021 December 20). Of the 1975 objects in the YSE DR1 sample, only 53 (~3%) were not observed by ZTF, and an additional 318 (~16%) have no ZTF forced-photometry data after quality cuts (e.g., S/N $\geqslant$ 4).

---

[45] For this work, if there is no qualifying bright PS1-r magnitude, but there are qualifying ZTF-r observations with concurrent YSE coverage, we include these objects as well. However, follow-up observations are prioritized by transients with PS1-$r < 18.5$ mag.
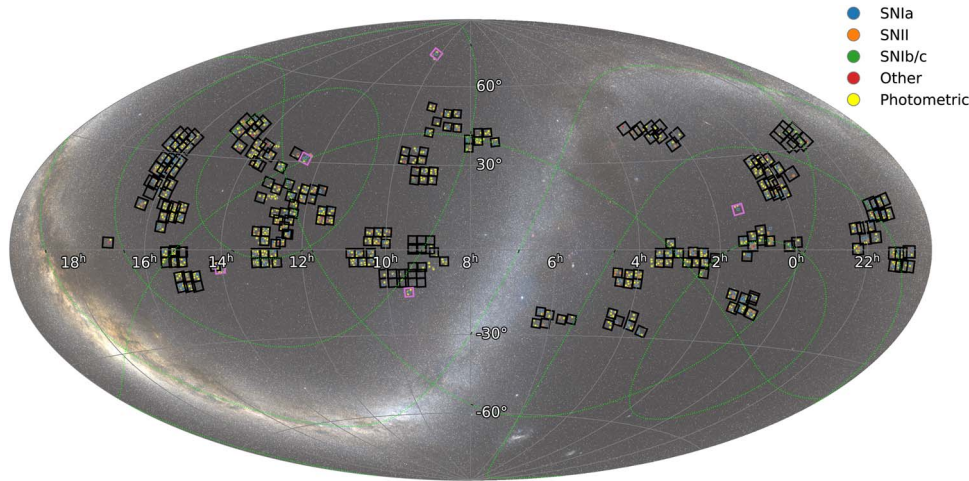[46] http://ztf.caltech.edu

[47] http://antares.noirlab.edu
[48] http://alerce.science
[49] https://Mars.lco.global
[50] https://fink-broker.org
[51] http://lasair.roe.ac.uk/
[52] https://www.wis-tns.org
[53] https://irsa.ipac.caltech.edu/data/ZTF/docs/forcedphot.pdf
[54] https://www.ivoa.net/documents/ObsLocTAP/index.html

Figure 1. An equatorial skymap of all 1975 YSE DR1 transients as of 2021 December 20, each marked as a circle with its color denoting the classification (spectroscopic or photometric). YSE fields chosen prior to 2021 December 1 are outlined as black squares, with pink squares highlighting five Cepheid calibrator SNe (2020jgl, 2020uxz, 2021J, 2021hpr; Ward et al. 2022; 2021pfs) not located in regular field pointings. YSE favors fields above 20° from the ecliptic plane. The survey does not observe below $\delta \approx -30$, because reference sky templates for difference imaging do not yet exist at these declinations. See Figure 1 of Jones et al. (2021a) for the distribution of YSE fields chosen prior to 2020 October 1.

4. YSE DR1 Properties and Statistics

In this section, we outline the instruments, properties, and statistics of YSE DR1, including the combination of YSE and ZTF data.

4.1. Overview of the Contents of YSE DR1

Here we briefly discuss the now publicly available materials collectively known as YSE DR1. For an in-depth explanation of the data processing, see Section 2.4. For the photometric classification methodology as a whole, see Section 5.

The full sample. YSE DR1 data files contain forced-photometry light curves in PS1-*griz* filters, observational properties, and metadata (e.g., R.A., decl., Milky Way extinction, redshift) for 1975 SN-like transients with YSE forced photometry. SN-like events are defined to have at least three observations for which the signal-to-noise ratio (S/N) > 4 in any passbands, and no previous detection within the survey following Jones et al. (2017), Villar et al. (2020). We applied Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) color selection criteria for active galactic nuclei (AGNs) from Jarrett et al. (2011), Stern et al. (2012), and removed likely AGN after inspection of the light curve and ancillary data. When available, we provide additional ZTF-*g*, ZTF-*r* forced-photometry observations (Masci et al. 2019), and host galaxy metadata including the host redshift (host-*z*; see Section 5.2) or manually validated SN spec-*z*. Otherwise, we provide a best-estimate photo-*z* (see Section 5.4). Host galaxy associations and additional host galaxy properties are from GHOST (Gagliano et al. 2021) and/or Sherlock[55] (Smith et al. 2020), which were subsequently vetted by eye. We also provide the SN class prediction and confidence scores from our photometric classifier (see Section 6.2). We do not provide any further YSE DR1 spectra, because all YSE classification spectra are uploaded to TNS on a nightly basis. The data files can be found on Zenodo (Aleo et al. 2022).

The distribution of these transients in equatorial coordinates is depicted in Figure 1. Of the 1975 YSE DR1 objects, 953

---

[55] https://github.com/thespacedoctor/sherlock

Table 1
YSE DR1 TNS Reporting Group Statistics

TNS Reporting Group	YSE DR1 (1975 Objects)	Spec. Sample (492 Objects)
YSE	953	119
ALeRCE	389	142
Pan-STARRS	273	27
ATLAS	146	93
ZTF	126	67
SGLF	58	19
GaiaAlerts	10	7
AMPEL	7	7
None[a]	5	4
ASAS-SN	3	3
GOTO	1	1
SNHunt	1	1
SIRAH	1	1
Fink	1	1
TAROT	1	0

Note. [a] No listed reporting group.

were first discovered by YSE and reported to TNS, or ∼48%. Other TNS reporting group statistics can be found in Table 1.

The photometric sample. 1483 objects comprise the photometric sample (75% YSE DR1). As expected, it has fewer observations on average, is dimmer, and is of higher redshift than the spectroscopic sample. In this work, we present the photometric classifications of these objects, which breaks down into 1048 (∼71%) SNe Ia, 339 (∼23%) SNe II, and 96 (∼6%) SNe Ib/Ic.

The spectroscopic sample. The remaining 492 SN-like objects constitute the spectroscopic sample, defined as having spectra providing both spectroscopic classification and spectroscopic redshift (or in rare cases, only the classification without an SN spectroscopic redshift, in which case we use the host–galaxy redshift or photo-*z* estimate). These spectra include all those aggregated from publicly available sources such as TNS, WISeREP (Yaron & Gal-Yam 2012), the literature, results from YSE's spectroscopic follow-up programs (which are
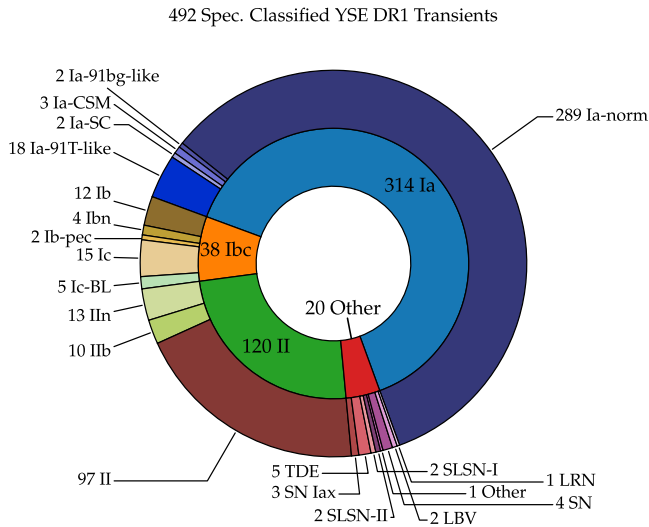
Figure 2. By-type breakdown of the 492 spectroscopically classified YSE DR1 transients. The inner circle represents the main class types SN Ia ($\sim$64%), SN II ($\sim$24.5%), and SN Ibc ($\sim$7.5%), while the outer circle further breaks this general classification into subtypes when appropriate (e.g., SN IIn within the umbrella SN II classification). The unlabeled inner red slice is a catchall for an Other class ($\sim$4.0%) of SN-like objects, which are spectroscopically identified but do not fall in one of the three validation set classes, which are TDE (5 objects), SN Iax (3 objects), SLSN-II (2 objects), SLSN-I (2 objects), SN with no distinction of SN I or SN II (4 objects), LBV (2 objects), LRN (1 object), and a remaining Other classification (1 object, AT 2021seu), which is a suspected Bowen fluorescence flare.

posted to TNS). We use SNID (Blondin & Tonry 2007) for a few new and updated classifications (Table 7). The spectroscopic class labels and redshifts are provided by the YSE team, other active systematic spectroscopic initiatives, or other active time-domain surveys. Of the 492 spectroscopic sample objects, 119 were first discovered by YSE and reported to TNS, or 24%. The by-type breakdown of the spectroscopic sample is shown in Figure 2.

Spectroscopic and/or photometric classification table. We provide the spectroscopic (when available as published on TNS or in the literature) or photometric classifications for the 1975 YSE DR1 SN-like transients. The complete, machine-readable version of this table is provided on Zenodo, and a shortened version of this text is shown in Appendix C.

4.1.1. The Spectroscopic Sample in Detail

For this work, we limit the spectroscopic sample to three potential general SN classes for validation of our photometric classifier:

1. SN Ia (including subtypes *Super-Chandrasekhar* SN Ia, hereafter SN Ia-SC, SN Ia interacting with a dense circumstellar medium, hereafter SN Ia-CSM, SN Ia-91T–like, SN Ia-91bg–like)—white dwarf thermonuclear explosions (314 objects);
2. SN II (including subtypes SN IIn, SN IIb)—CC explosions from red supergiant and LBV progenitors (120 objects), and because SN IIP and SN IIL are believed to originate from the same progenitor population (Sanders et al. 2015), we aggregate these into a singular SN "II" label;
3. SN Ibc (including subtypes SN Ib-pec, SN Ic, SN Ic-BL, SN Ib, SN Ibn)—CC explosions from massive stars
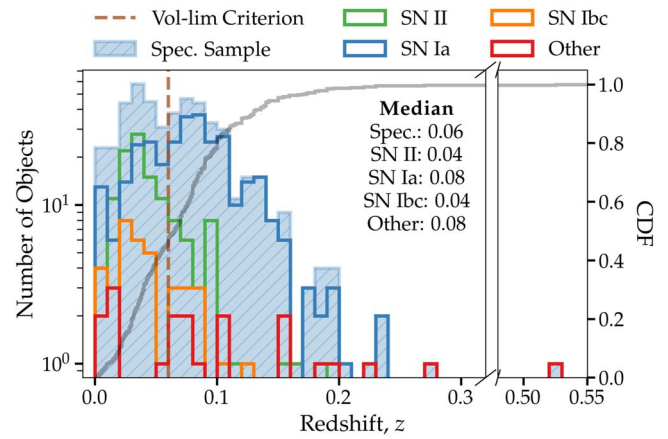
Figure 3. The redshift distribution of all objects in the spectroscopic sample (light blue). The cumulative SN distribution is shown as a solid gray line. A dashed red line indicates the volume-limited sample criterion ($D \leqslant 250$ Mpc; or $z \leqslant 0.06$). The most populous redshift bin is $z \in [0.03, 0.04]$, the approximate peak and median values for SN II (green) and SN Ibc (orange). The z peak for SN Ia (blue) occurs at $z \sim 0.08$. The sample is overwhelmingly dominated by SNe Ia past $z \sim 0.1$, except for a few SLSNe. The limiting redshift for YSE DR1 is $z \sim 0.3$, as predicted by Jones et al. (2021a). The only exception is SN 2021uwx, an SLSN-I at $z \sim 0.525$.

having lost their hydrogen (Ib) and helium (Ic) envelopes (38 objects).

We include a catchall fourth class, Other, for objects that do not fall into the three major classes; this class consists of the following groups: tidal disruption event (TDE), hydrogen-rich superluminous SN (SLSN-II), hydrogen-poor superluminous SN (SLSN-I), type Iax (SNe Iax), rare SN imposters such as an LBV outburst and a luminous red nova (LRN), and a remaining other[56] classification. These Other events (20 objects) are neither simulated nor used to evaluate our classifier due to their small sample size, but are included in YSE DR1 for completeness. Subsequent analyses of such objects are being performed, and the light-curve data can be provided upon submitting a "small data policy" request on YSE's official website.[57] Note that we do not include spectroscopically classified stellar or non-SN-like sources in this YSE DR1, such as cataclysmic variables and known AGNs, for example. Selected highlights from the spectroscopically classified YSE DR1 transients are shown in Table C1.

Approximately 25% of the full sample was spectroscopically observed. A majority of this selection was chosen to satisfy our magnitude- and volume-limited survey strategies (see Section 2.5). For those outside that purview, the young, fast-rising, and redder or interesting transients were promoted for quick follow-up observations. The redshift distribution of spectroscopically observed objects shown in Figure 3 extends to $z \approx 0.3$ (with the exception SLSN-I SN 2021uwx at $z \sim 0.525$), with an apparent double peak at redshift bin $z \in [0.03, 0.04]$ (the approximate peaks for CC SN II and SN Ibc) and $z \in [0.08, 0.09]$ (the approximate peak and median value for SN Ia). Past $z \approx 0.1$, the spectroscopic sample is mostly SNe Ia, due to their high intrinsic luminosity. Additionally, fast rising or young light curves, particularly those that exhibit SN Ia profiles, are prioritized for follow-up

---

[56] AT 2021seu is suspected to be a possible Bowen fluorescence flare: https://www.wis-tns.org/astronotes/astronote/2021-195.
[57] https://yse.ucsc.edu
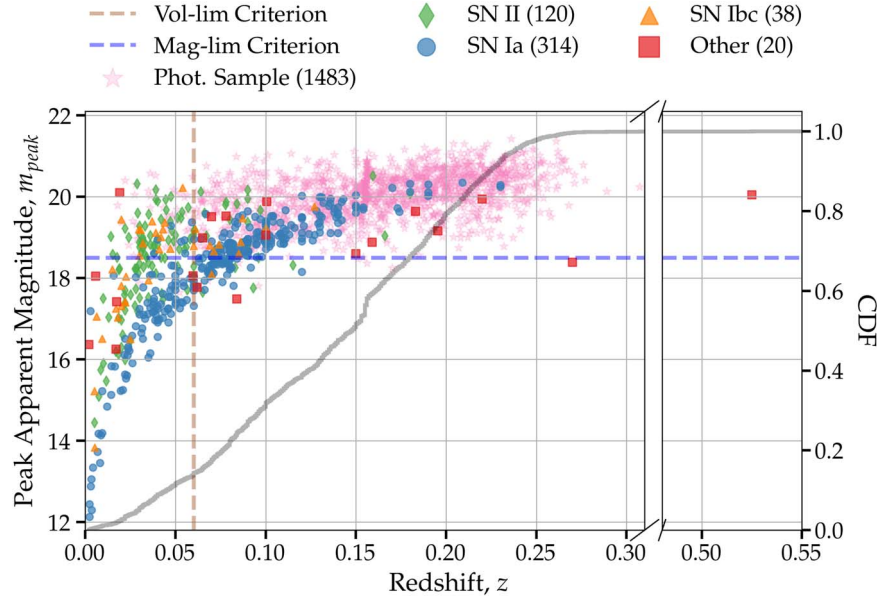
Figure 4. A scatter plot of the peak apparent magnitude m_{peak} vs. redshift for YSE DR1. Of the spectroscopic sample, SN Ia are blue circles, SN II are green diamonds, SN Ibc are orange triangles, and red squares represent the Other classification. Pink stars represent members of the photometric sample, which uses the spectroscopic host-z (if available, see Section 5.2) or photo-z (Section 5.4) for the redshift. Dashed lines delineate the magnitude- and volume-limited thresholds (purple and brown, respectively). The cumulative distribution function of YSE DR1 SNe is shown as a solid gray line. The two highest-redshifted objects in the sample are AT 2020aeid, a photometric SN II with a spectroscopic host-z of 0.307, and SN 2021uwx, an SLSN-I at $z \sim 0.525$.

observations. This high-z sample will be critical for measuring the dark energy equation-of-state parameter w, testing general relativity, and helping quantify CC SNe contamination in future cosmology analyses. For YSE DR1, we find the median redshift of 293 untargeted spectroscopic SN Ia to be $z = 0.08$, an underestimation of the prediction $z \approx 0.12$ from Jones et al. (2021a; see their Appendix A). A YSE SNe Ia sample and cosmological analysis is forthcoming.

To further investigate how the redshift distribution correlates with the observed peak apparent magnitude m_{peak}, we plot the scatter of the spectroscopic and photometric samples in Figure 4. Assuming SNe photometric redshifts are correct, the median redshift for photometric objects alone is $z \sim 0.16$, whereas from Figure 3 we know that the median redshift of the spectroscopic sample is $z \sim 0.065$ (near the threshold of the volume-limited sample at $z \sim 0.06$). Overall, this translates into a median redshift of $z \sim 0.14$ for the entire YSE DR1 sample. Moreover, we find only $\sim 26\%$ of transients above the median predicted survey redshift of $z = 0.19$ from simulations (Jones et al. 2021a). This finding cements the early survey yields from Jones et al. (2021a), who originally reported only 26% of transients were above this median predicted survey redshift threshold. Reasons for this may include remnant biases in the photo-z determinations from `Easy PhotoZ` (as reliable redshifts are more difficult to estimate for faint or undetected host galaxies), noise in the template images (which simulations assume to be negligible), and potential efficiency losses in our "real/bogus" algorithm. Simulations from Jones et al. (2021a) did correctly predict that YSE's depth allows us to find SN up to $z \sim 0.3$.

Using redshift measurements, we can transform apparent magnitudes into absolute magnitudes using a flat ΛCDM cosmology with $H_0 = 70 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$, and $\Omega_M = 0.3$. We investigate the distribution per SN class of the peak absolute magnitude (M_{peak}, uncorrected for dust extinction) for objects with an observed peak in Figure 5. Here, we use the brightest
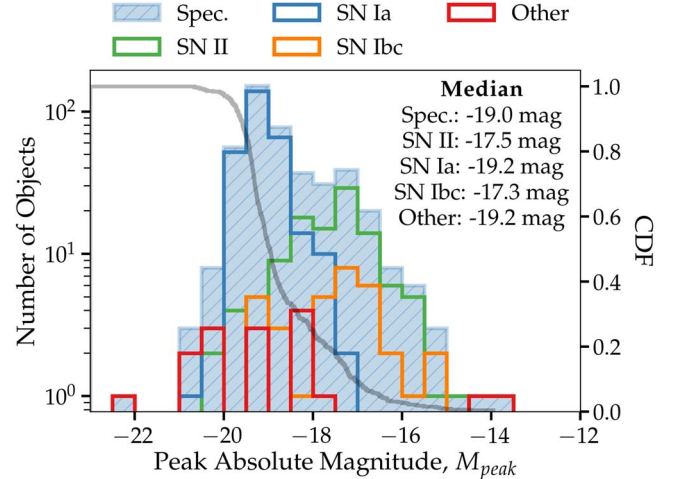
Figure 5. The peak absolute magnitudes (uncorrected for dust extinction) of all objects in the spectroscopic sample with an observed peak (445 objects), using the brightest single detection as a proxy for M_{peak}. The cumulative SN distribution (from least intrinsically bright to most bright) is shown as a solid gray line. The most populated core-collapse (SN II, SN Ibc) M_{peak} bin is $M_{peak} \in [-17, -17.5]$. As expected, we find an average (median) peak absolute magnitude $M_{peak} \sim -19$ mag for SN Ia (blue), with a few brighter and fainter objects due to rarer subtypes, discovery of the object post-peak, or heavy dust extinction. Due to SN Ia events comprising the majority of the spectroscopic sample, the median M_{peak} distribution is $M_{peak} \sim -19.0$ mag.

single detection as a proxy for M_{peak}, including objects that are discovered post-peak. We find an average (median) value of $M_{peak} = -19.0$ mag for the entire spectroscopic distribution, consistent with the typical $M_{peak} = -19$ mag value of SNe Ia that dominate the sample. Additionally, we see the presence of SNe Ia below $M_{peak} \sim -18$ mag primarily due to heavy dust extinction. SNe II objects display the widest range of M_{peak}, due to the use of a holistic SN II label and the diversity of these events. The SNe Ibc population has a relatively flat M_{peak}
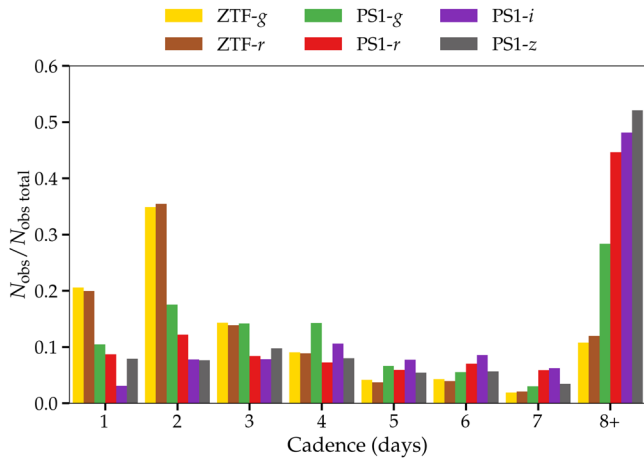
Figure 6. Cadence distribution for YSE DR1 per passband, truncated at 7 days, and rounded to the nearest integer day. $N_{obs}/N_{obs\ total}$ is the fraction of observations at a specific cadence compared to the total number of cadence observations. More cadence statistics can be found in Table 2.

Table 2
YSE DR1 Cadence Statistics

Passband	$N_{obs\ total}$	Median Cadence (days)
PS1-g	8856	4.00
PS1-r	10,133	6.90
PS1-i	9120	7.01
PS1-z	3322	8.01
ZTF-g	15,431	2.05
ZTF-r	21,831	2.04
YSE DR1 Total		
PS1-$griz$	31,431	3.98
PS1-$griz$, ZTF-gr	67,138	1.98

Note. For observations only after quality cuts (e.g., S/N $\geqslant$ 4).

distribution, likely due to the small sample size. Both SNe II and SNe Ibc, the CC objects, share a M_{peak} median value of $M_{peak} \sim -17.5$ mag. The few Other objects brighter than $M_{peak} \sim -21$ mag are the SLSNe.

4.1.2. Cadence

YSE observes each field with a 3 days cadence while monitoring the ZTF observing strategy, resulting in well-sampled light curves (particularly for fast-evolving or short-lived transients). The cadence distribution per passband of YSE DR1 (not accounting for telescope maintenance and/or downtime and Moon avoidance) is shown in Figure 6, with additional cadence statistics in Table 2. Approximately 40% and 30% of observations were carried out at the planned 3 days cadence in PS1-g or PS1-r filters, respectively (at least one of which is required per epoch; see Section 2.2), and ~70% (~50%) of revisits occurred within $\leqslant$ 7 days for PS1-g (PS1-r). Here, we only consider epochs with S/N $\geqslant$ 4.

As previously mentioned, YSE schedules its observations to precede ZTF observations by one calendar day in an effort to increase the combined cadence and maximize discoveries of young SNe. The resulting effect is magnified in dark time, when YSE achieves ~0.4–0.5 mag deeper limits than ZTF. During the lifespan of YSE DR1, ZTF made their nightly
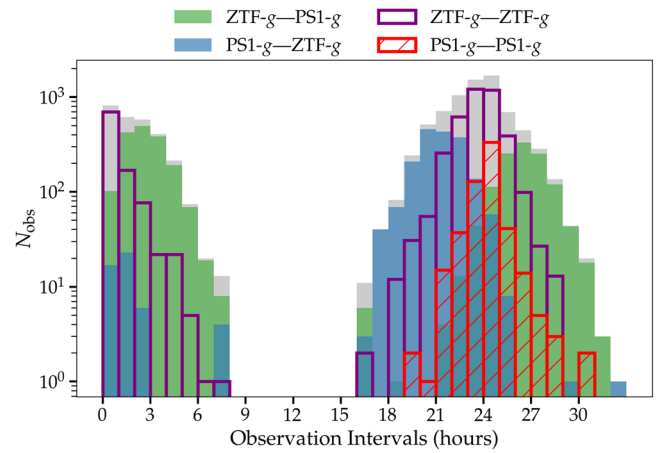
Figure 7. Distributions of time intervals between observations of the same field in the g band for YSE DR1, truncated at 33 hr. We have removed all observations with a separation of <0.5 hr because these are often repeat observations when one observation is poor, and most transients should not evolve significantly on that timescale. We separate intervals from single telescopes and across telescopes to highlight the advantage of having telescopes at multiple longitudes. Intra-telescope intervals are shown as hatched red and empty purple histograms for PS1 and ZTF, respectively. Inter-telescope intervals are shown as green and blue, with the former being observations where ZTF observations precede PS1, and the latter is reversed. The combination from all telescopes is shown in light gray.

observing plans public. This has allowed our team to significantly improve our ability to plan overlapping observations.[58] With the 2 days cadence of ZTF phase (II), and the interleaving observations of shared fields, YSE DR1 achieves an improved cadence over YSE observations alone; the median observed effective cadence is 3.98 days without ZTF observations (compared to 3.9 days median cadence reported in Jones et al. 2021a), and 1.98 days with ZTF observations.

An advantage of combining PS1 and ZTF observations is that the differences in observatory longitudes of ~40° produce a difference in the hour angle for a given target of ~3 hr at any given time. Therefore, if PS1 observes the same field on the same night as ZTF, we expect a typical temporal separation of about 3 hr. If PS1 observations precede ZTF observations by one night, we expect a typical separation of about 21 hr. Observing at different hour angles at a given observatory can reduce or extend any gap.

Figure 7 displays histograms of the temporal separation between observations in the g band. Examining only temporal coverage by a single telescope (ZTF or PS1), we find a peak in the distribution at ~24 hr with an FWHM of ~3 hr. However, the time between a PS1 observation followed by a ZTF observation peaks at 21.1 hr, as expected. The time between a PS1 observation preceded by a ZTF observation has peaks at both 2.6 and 27.7 hr, also as expected. Overall, approximately 33% of total g-band observations are intersurvey.

The combination of PS1 and ZTF data results in a broader distribution of timescales probed near 24 hr, with an FWHM of ~4 hr. Critically, it increases the number of image pairs with a difference of 3–7 hr by a factor of 13.0 over a single telescope alone. Such timescales are especially important for fast-evolving transients.

---

[58] Rubin will adapt the same protocol for their observing schedule, which will enable us to supplement the LSST cadence in 2024.

Table 3
YSE DR1 Statistics of Observed SNe

	$N_{transients}$
SN-like transients, YSE DR1 sample	1975
$r \lesssim 18.5$ mag	213
$D \lesssim 250$ Mpc	294
$r \lesssim 18.5$ mag and $D \lesssim 250$ Mpc	158
Targeted YSE SNe	31
SNe Ia with phase < -10 days	151
Spec. SNe (untargeted)	492 (461)
Spec. SNe and $r \lesssim 18.5$ mag (untargeted)	207 (181)
Spec. SNe and $D \lesssim 250$ Mpc (untargeted)	236 (207)
Spec. SNe and $r \lesssim 18.5$ mag and $D \lesssim 250$ Mpc (untargeted)	158 (133)

Note. Transient statistics from YSE DR1, covering approximately the first 2 yr of the YSE survey (2019 November 24 to 2021 December 12). We have been at 50% of our full observing allocation since early 2020 January, but have lost approximately 4 months due to weather and telescope malfunctions.

Leveraging the combined cadence with deep, multicolor photometry, there are more opportunities to observe the light-curve rise, often at least several days before peak light. For YSE DR1, we report the median phase of the first $S/N > 4$ observation for YSE DR1 transients (based on estimates of the time of maximum light) is -11.7 days.

4.1.3. Magnitude- and Volume-limited Census Results

YSE set the goal to spectroscopically classify every transient with peak $r \leqslant 18.5$ mag, $D < 250$ Mpc, or a detection within 2 days of explosion. A total of 207 out of 213 magnitude-limited sample qualifying-objects have spectroscopic coverage and classification ($\sim$97%). This is a further increase from the 91% spectroscopic classification completeness reported in Jones et al. (2021a). Similarly, out of 294 objects in YSE DR1 that qualify for the volume-limited sample, YSE achieves $\sim$80% spectroscopic classification, or 236 objects. A summary table of our full YSE DR1 statistics is found in Table 3.

A few predominant reasons we are not reaching 100% are a lack of spectroscopic resources because of COVID-19 shutdowns, and discovering and/or observing a magnitude- or volume-limited qualifying-transient either shortly before it sets or before we move fields. Despite these setbacks, some transients falling into YSE's magnitude- or volume-limited samples have been classified by external teams, including but not limited to SNe in the near-infrared avec Hubble (SIRAH; HST-GO 15889, PI: Saurabh Jha), ZTF, and ePESSTO+ (Smartt et al. 2015).

The by-type breakdown of the magnitude- and volume-limited samples are found in Figure 8. Of the untargeted magnitude-limited sample of 181 objects, $\sim$66.5% (120 objects) are SN Ia, $\sim$23% (42 objects) are SN II, $\sim$7% (13 objects) are SN Ibc, and the remaining $\sim$3.5% (6 objects) are the others (3 TDE, 1 SN Iax, 1 SLSN-II, 1 Other). The YSE DR1 sample of 181 magnitude-limited objects in a span of $\sim$2 yr across $\sim$750 deg^2 is outperforming the > 100 magnitude-limited SNe yr^1 projection in the full YSE survey area of 1500 deg^2 from Jones et al. (2021a). Meanwhile, of the untargeted volume-limited sample of 207 objects, $\sim$43% (89 objects) are SN Ia, $\sim$43% (89 objects) are SN II, and $\sim$12% (25 objects) are SN Ibc, and the remaining $\sim$2% (4 objects) are Other (1 TDE, 1 SN Iax, 1 LBV, 1 Other).

The untargeted magnitude-limited sample has good agreement with the relative SN fractions of ZTF BTS and ASAS-SN surveys within uncertainties, as shown in Table 4. The relative rates ($\mathcal{R}$) of $\mathcal{R}$(Ia) $= 0.682 \pm^{0.083}_{0.073}$, $\mathcal{R}$(II) $= 0.239 \pm^{0.064}_{0.079}$, $\mathcal{R}$(Ibc) $= 0.074 \pm^{0.033}_{0.057}$, and $\mathcal{R}$(SLSN) $= 0.006 \pm^{0.005}_{0.005}$ across 181 objects are consistent with the ZTF BTS rates assuming a 90% confidence interval across 761 objects at the same magnitude-limiting threshold ($r < 18.5$ mag) and follow-up strategy (spectroscopically classify every object in this threshold, and include SNe discovered as well as recovered by the survey), and similar cadence. Our results are tenuously in agreement with the ASAS-SN rates (Holoien et al. 2019) across 818 objects. Although we have a similar follow-up strategy, the ASAS-SN magnitude-limiting threshold is at a brighter limit of their $V < 17$ mag and at a faster cadence (1 day). Our relatively higher observed rates of SNe II and SNe Ibc than both ZTF BTS and ASAS-SN could be due to YSE's sensitivity in the redder passbands, picking up a greater number of lower intrinsic luminosity events compared to brighter SNe Ia in more heavily dust-extincted regions.

According to Fremling et al. (2020), the ZTF BTS and ASAS-SN rates use observations drawn from a multinomial distribution with estimated 90% confidence intervals on the true rate via the Goodman approximate method (Goodman 1965) as implemented with the MultinomCI function in the R package DescTools (Andri et al. 2022). We use a similar implementation via the *multinomial_proportions_confint* function from the statsmodels Python package (Seabold & Perktold 2010), and recalculate the relative rates for all surveys considered here using the same 90% confidence interval and Goodman approximate method. Note that we achieve similar but different relative rates from Table 2 of Fremling et al. (2020). At 97% complete, we match the spectroscopic completeness to ZTF BTS within a few percent (95%, Fremling et al. (2020), and a greater completion than ASAS-SN ($\sim$70% complete for $m_{peak} \leqslant 17.0$ in ASAS-SN V and g bands, Holoien et al. 2019[59]). The true rates and their uncertainties require a detailed estimate of the completeness besides the spectroscopic classification completeness, which is beyond the scope of this paper. We will address this in future work.

Although we make no detailed attempt to estimate subtype SN fraction, we note that, of the SN Ia in the galaxy untargeted magnitude-limited sample, we find 113 normal SNe Ia, 4 SNe 91T–like SNe, and 0 SNe 91bg–like SNe among 120 total SNe Ia. The relative SN Ia subtype fraction of $\sim$3% 91T-like and 0% 91bg-like is in agreement to that found by ASAS-SN ($\sim$6% 91T-like and $\sim$1% 91bg-like SNe; Holoien et al. 2017a, 2017b, 2017c, 2019) and ZTF BTS ($\sim$6% 91T-like and $\sim$1% 91bg-like SNe; Fremling et al. 2020), but in more contention with the LOSS results for a magnitude-limited survey ($= 17.7\% \pm^{10.8}_{9.3}$ 91T-like and $= 3.3\% \pm^{2.0}_{1.5}$ 91bg-like SNe; see Table 7 for 1-d cadence in Li et al. 2011). These particular LOSS rates are based on an assumed LF and Monte Carlo simulations, whereas those reported in Holoien et al. (2017a, 2017b, 2017c, 2019), Fremling et al. (2020) use all discovered and recovered SNe from their respective surveys (as we do here). Moreover,

---

[59] We note that, in the updated 2018–2020 ASAS-SN sample from Neumann et al. (2023), the survey changed their limiting threshold to $g \leqslant 18$ mag, and subsequently have a higher spectroscopic completeness measure of 90% complete for $m_{peak} \leqslant 17.0$ mag.

181 Spec. Classified Untargeted YSE DR1 Transients with $r < 18.5$ mag
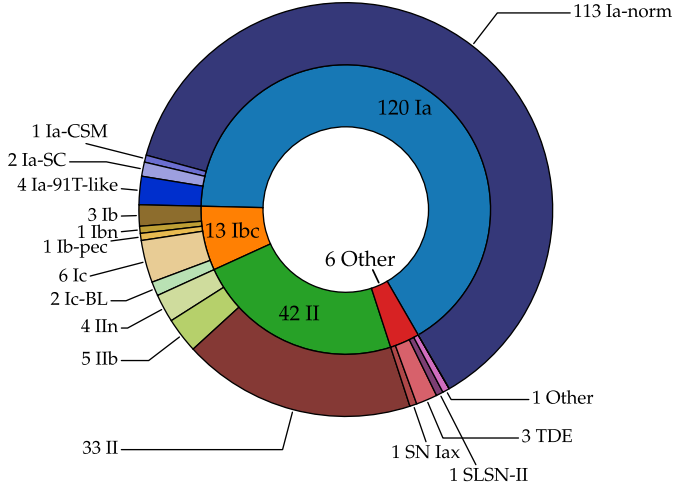
207 Spec. Classified Untargeted YSE DR1 Transients with $D < 250$ Mpc
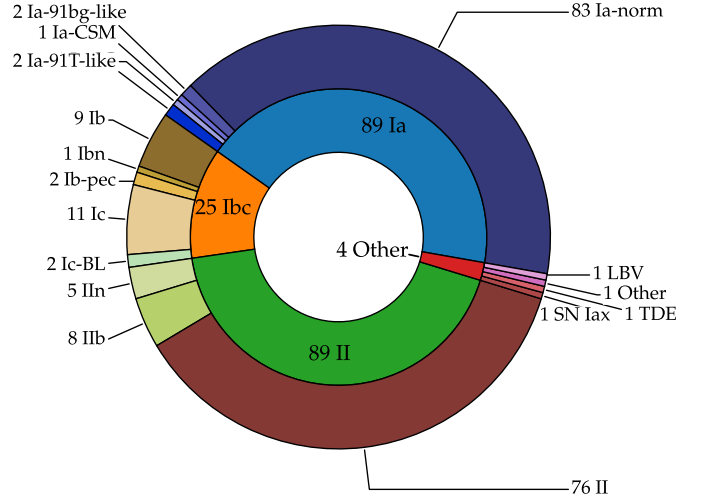
Figure 8. Left: same as right panel of Figure 2, but of the magnitude-limited sample ($r < 18.5$ mag). Approximately ∼97% of all magnitude-limited YSE DR1-qualifying objects have a spectroscopic classification. The relative rates from the magnitude-limited sample have marginal agreement with other magnitude-limited surveys (LOSS, ASAS-SN, ZTF BTS), and their comparison is found in Table 4. Right: same as right panel of Figure 2, but of the volume-limited sample ($D < 250$ Mpc). Approximately 80% of all volume-limited YSE DR1-qualifying objects have a spectroscopic classification. The relative rates from the volume-limited sample compared to other volume-limited surveys (LOSS, ZTF CLU) are found in Table 5. There are 133 members of both the untargeted magnitude- and volume-limited samples.

LOSS may have overestimated the relative rate of SNe Ia-91T if such events preferentially occur in late-type galaxies or are associated with younger stellar populations. This would be a consequence of a galaxy-targeted strategy of massive and high star formation galaxies (Taubenberger 2017).

As we compare the volume-limited SN fractions to those reported in the literature, we must keep in mind the specific distance thresholds. For example, Li et al. (2011) reported that, in a volume-limited sample of 175 SNe within a cutoff distance of 60 Mpc, the relative SN fractions are 57% SN II, 24% SN Ia, 19% SN Ibc. Meanwhile, our cutoff distance (250 Mpc) is over 4× that of Li et al. (2011), and our volume-limited survey finds higher relative fractions of SN Ia, and fewer SN II and SN Ibc. The higher intrinsic luminosity of SNe Ia compared to SNe II and SNe Ibc enables us to discover more SNe Ia in a larger volume (farther distance) given a constant rate. Moreover, our volume-limited survey is only ∼80% complete. Regardless, our results are consistent with the ZTF CLU (De et al. 2020) catalog, which as of 2020 October 1 has logged 1128 SNe with fractions 40% SN Ia, 45% SN II, 13% SN Ibc, and 2% Other out to $D < 200$ Mpc. A direct comparison of our relative rates compared to LOSS and ZTF CLU with estimated uncertainties is found in Table 5.

Our untargeted volume-limited sample has excellent agreement with the relative SN fractions of the ZTF CLU survey within uncertainties. Our relative rates ($\mathcal{R}$) of $\mathcal{R}(\text{Ia}) = 0.438 \pm^{0.072}_{0.075}$, $\mathcal{R}(\text{II}) = 0.438 \pm^{0.072}_{0.075}$, and $\mathcal{R}(\text{Ibc}) = 0.123 \pm^{0.041}_{0.057}$, across 207 objects are consistent with the ZTF CLU rates assuming a 90% confidence interval across 1109 objects at a similar but more local volume-limiting threshold.

Lastly, there are 133 members of both the untargeted magnitude- and volume-limited samples, or ∼7% of all objects in YSE DR1. If we approximate the length of observation of YSE DR1 to 2 yr (roughly accounting for downtime), this is

nearly twice the rate of the 59 SNe yr^{-1} projection from the full YSE survey (∼1500 deg^2) in Jones et al. (2021a) when considering the YSE DR1 survey area of ∼750 deg^2. When accounting for members of the magnitude-limited or volume-limited or both, we observe a rate of ∼260 SNe yr^{-1}.

4.2. Comparison of YSE Spectroscopic and Photometric Samples

We highlight the similarities and differences between the YSE spectroscopic and photometric samples as histograms in Figure 9. On average, the spectroscopic sample has approximately 2.5× the total number of observations ($N_{\text{obs total}}$) than the photometric sample. The bulk difference of the two samples stems from the number of observations after the light-curve peak ($N_{\text{obs after peak}}$). The spectroscopic sample has, on average, more than 3× $N_{\text{obs after peak}}$ than that of the photometric sample, due to spectroscopic objects often being brighter and more closely monitored and/or followed (i.e., targeted YSE objects; see Section 2.3), sometimes for the purpose of capturing multiple spectra. Nearly all YSE DR1 transients with $N_{\text{obs total}} > 100$ have a spectroscopic classification. However, of the 13 photometric objects with $N_{\text{obs total}} > 100$, most are long-lived, but relatively faint $m_{\text{peak}} \leqslant 19.5$ mag. We suspect that, by the light-curve evolution (duration and color), photo-z estimate, and host variability, two were missed SLSN candidates (AT 2020abgb, AT 2020unn).

Another pronounced difference originates from the distributions of the peak apparent magnitude, m_{peak}, per passband, shown in Figure 10. Note that these histograms are from the singular m_{peak} value per object for the passband it applies, as opposed to one per passband. On average (median), the photometric sample is dimmer at m_{peak} by ∼1.5 mag, at a value of ∼20 mag per passband, spanning $m_{\text{peak}} \in [17.5, 21.5]$ mag. Meanwhile, the spectroscopic sample has an average $m_{\text{peak}} \sim 18.5$ mag, spanning $m_{\text{peak}} \in [12, 21]$ mag. The brightest magnitude observations for both samples typically

Table 4
Relative SN Rates from Magnitude-limited Surveys

Survey	Criteria	$\mathcal{R}$(Ia)	N_{Ia}	$\mathcal{R}$(II)	N_{II}	$\mathcal{R}$(Ibc)	N_{Ibc}	$\mathcal{R}$(SLSN)	N_{SLSN}
LOSS	⋯	$0.792\pm^{0.044}_{0.055}$	⋯	$0.166\pm^{0.050}_{0.039}$	⋯	$0.041\pm^{0.016}_{0.013}$	⋯	⋯	⋯
ASAS-SN	$V < 17$ mag	$0.742\pm^{0.036}_{0.033}$	607	$0.211\pm^{0.030}_{0.034}$	173	$0.043\pm^{0.013}_{0.019}$	35	$0.004\pm^{0.003}_{0.009}$	3
ZTF BTS	ZTF-$gr < 18.5$ mag	$0.719\pm^{0.038}_{0.035}$	547	$0.204\pm^{0.031}_{0.035}$	155	$0.053\pm^{0.015}_{0.021}$	40	$0.025\pm^{0.010}_{0.016}$	19
YSE[a]	PS1-$r < 18.5$ mag	$0.682\pm^{0.083}_{0.073}$	120	$0.239\pm^{0.064}_{0.079}$	42	$0.074\pm^{0.033}_{0.057}$	13	$0.006\pm^{0.005}_{0.005}$	1

Note. Relative rates $\mathcal{R}$ of SNe Ia, II, Ibc, and SLSNe from a magnitude-limited search for the LOSS (Li et al. 2011), ASAS-SN (Holoien et al. 2017a, 2017b, 2017c, 2019), ZTF BTS (Fremling et al. 2020), and YSE surveys (this work). The LOSS values are taken directly from Li et al. (2011; their Table 7, using a 1 day cadence), which use an assumed luminosity function and Monte Carlo simulations (thus, the number of SN per class is irrelevant and not reported). We list ASAS-SN rates using both SNe discoveries and SNe recovered, as we do the same for our analysis of YSE DR1. Thus, the numbers shown here are not what is shown in Figure 1 of Holoien et al. (2019), which does not consider SNe recovered. Here, we draw observations from a multinomial distribution with 90% confidence intervals. See text for details.
[a] The untargeted magnitude-limited subsurvey of YSE only. Not the entirety of YSE DR1.

Table 5
Relative SN Rates from Volume-limited Surveys

Survey	Distance (D)	$\mathcal{R}$(Ia)	N_{Ia}	$\mathcal{R}$(II)	N_{II}	$\mathcal{R}$(Ibc)	N_{Ibc}
LOSS	$D < 60$ Mpc	$0.241\pm^{0.037}_{0.035}$	⋯	$0.572\pm^{0.043}_{0.041}$	⋯	$0.187\pm^{0.035}_{0.033}$	⋯
ZTF CLU[b]	$D < 200$ Mpc	$0.410\pm^{0.031}_{0.032}$	454	$0.458\pm^{0.032}_{0.032}$	507	$0.131\pm^{0.020}_{0.023}$	148
YSE[a]	$D < 250$ Mpc	$0.438\pm^{0.072}_{0.075}$	89	$0.438\pm^{0.072}_{0.075}$	89	$0.123\pm^{0.041}_{0.057}$	25

Notes. Relative rates $\mathcal{R}$ of SNe Ia, II, and Ibc from a volume-limited search for the LOSS (Li et al. 2011), ZTF CLU, and YSE surveys (this work). The LOSS values are taken directly from Li et al. (2011; their Table 7), which uses a corrected completeness of each SN in the LF within the considered volume (thus, the number of SN per class is not reported). Although the original volume-limited sample of 175 SNe has a cutoff distance of 80 Mpc for SNe Ia, and 60 Mpc for SNe II and SNe Ibc, the SNe Ia within 60 Mpc are considered together with the CC SNe in the LF sample to derive their relative fractions. To calculate uncertainties for ZTF CLU and YSE, we draw observations from a multinomial distribution with 90% confidence intervals. See text for details.
[a] ZTF CLU (De et al. 2020) results as of 2020 October 1. See https://sites.astro.caltech.edu/ztf/ZTFII_ReverseVisit/Talks/Tzanidakis_CLU.pdf.
[b] The untargeted volume-limited subsurvey of YSE only. Not the entirety of YSE DR1.

come from the g and r bands of either YSE or ZTF. This is expected due to YSE's filter sequence strategy—require one g- or r-band observation per epoch.

We report four SN-like transients (AT 2020ebc, AT 2020fci, AT 2020rkp, AT 2020tkw) from the photometric sample that exceed $m_{peak, PS1-r} \leqslant 18.5$ mag but were missed for spectroscopic follow-up observations. In most cases, YSE detected the transient at peak and near our magnitude threshold, soon declining to a point where classification was particularly difficult. In other cases (e.g., AT 2020rkp), we detected the transient soon before we stopped observing the field. Follow-up observations of other transients just below the magnitude-limited criterion like AT 2020kld were also attempted but failed. We additionally note two transients (AT 2021kwh, AT 2021pgm) in the photometric sample with ZTF-r observations brighter than 18.5 mag, but during a time weeks before YSE began observations of their fields (at which point the transient had faded well below the threshold). Of 213 objects that satisfy the r-band magnitude-limited criterion, 207 are included in the spectroscopic sample, placing our magnitude-limited spectroscopic completion at ∼97%. These spectroscopic completion values exceed by a few percent that of ZTF BTS for the same criterion (∼95%, Fremling et al. 2020).

4.3. SN Offset

With our vetted host associations (Section 5.2), we calculate the offsets between the SN and its host galaxy center. Here, we define the SN offset as the angular separation between the reported SN coordinates and its vetted host's NASA/IPAC Extragalactic Database (NED) coordinates. To understand the physical offset of these SN, we transform the angular separation into a physical distance using the angular diameter distance d_A and a standard ΛCDM cosmology with $H_0 = 70.0$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.30$, and $\Omega_\Lambda = 0.70$. A histogram of these SN offsets (kiloparsecs) for the full YSE DR1 sample, photometric sample, and spectroscopic sample is shown in Figure 11.

Physical SN offsets of YSE DR1 range from 0.038 to 55.63 kpc,[60] with a median of 3.75 kpc. Of the 75 objects with an offset $\geqslant$ 20 kpc (∼4% of YSE DR1), there are 13 SN Ia, 2 SN II, and 1 SN Ib, and 57 are in the photometric sample. The object with the highest SN offset is SN 2020acbc, an SN Ia at 55.6 kpc, or 23″93 at $z \sim 0.13$. When considering only YSE discoveries, the highest physical offset is AT 2020kof at 43.4 kpc, or 14″65 at a host-$z = 0.174$.

When comparing to ASAS-SN discoveries alone (2014–2017, Holoien et al. 2019), YSE discoveries have a smaller median angular separation (1″4 versus 4″5), but have a slightly larger median physical offset (3.87 versus 2.4 kpc). This holds true for the newest release of ASAS-SN discoveries (2018–2020, Neumann et al. 2023), which have median offsets of 3″9 and 2″5 kpc. This is likely due to the YSE's greater depth and smaller pixel scale.

---

[60] Which is about twice that of the ASAS-SN sample: offset $\in (0, 25]$ kpc (Holoien et al. 2019).
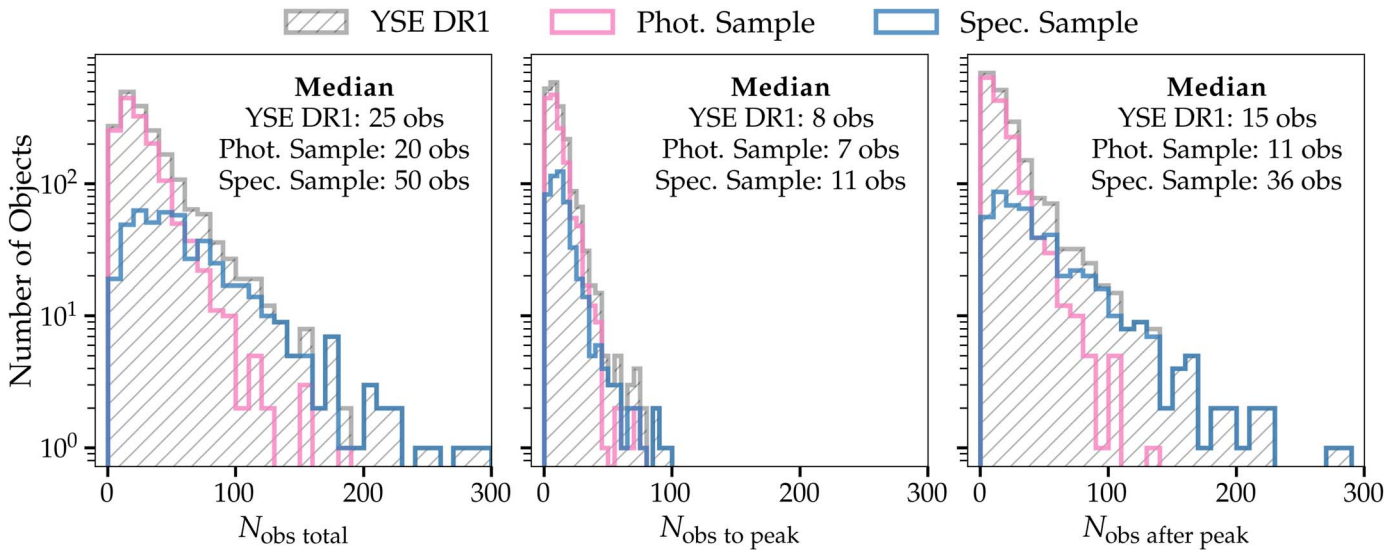
Figure 9. Histograms of the number of observations (N_{obs}) of the full light curve ($N_{obs\ total}$, left), to the epoch of peak apparent magnitude m_{peak} per object ($N_{obs\ to\ peak}$, center) and the number of observations post-peak magnitude per object ($N_{obs\ after\ peak}$, right) across all passbands for YSE DR1 (gray, hatched), spectroscopic (blue), and photometric (pink) samples. In both cases, the photometric sample on average has fewer observations on the rise of the transient's light curve preceding the peak and less observations post-peak when compared to the spectroscopic sample. The most populated $N_{obs\ to\ peak}$ bin is $N_{obs\ to\ peak} \in [5, 10]$ compared to $N_{obs\ to\ peak} \in [10, 15]$ for the spectroscopic sample. Similarly for $N_{obs\ after\ peak}$, the most populated bin is $N_{obs\ after\ peak} \in [0, 5]$ compared to $N_{obs\ after\ peak} \in [5, 10]$ for the spectroscopic sample. Nearly all YSE DR1 transients with $N_{obs\ total} > 100$ have a spectroscopic classification.

5. Photometric Classification

Methodology. In this section, we describe the process of preparing the YSE and ZTF forced-photometry light curves, transient host galaxy association, photo-z estimates, and YSE +ZTF survey simulations with training set composition and comparison to YSE DR1. Then we discuss ParSNIP, the architecture for our photometric classifier, and its training process. If the reader is not interested in such details, we suggest skipping to Section 6 and returning here later if needed.

5.1. Pipeline Overview

To better understand the flow of information and interdependent components of this work, we visualize a streamlined workflow as a schematic in Figure E1. The *A* process describes the process of generating realistic YSE DR1 simulations for training set generation, which is explained in detail in Section 5.6. The *B* process describes the YSE DR1 data processing pipeline, from SN observation to light-curve generation with associated metadata (host association, redshift). The *B* process is described in detail in Sections 2.4–5.2. Lastly, the *C* process summarizes the adapted ParSNIP architecture (Boone 2021), from training to feature selection and photometric classification. This is explored in Section 5.8.

5.2. Host Association

As part of the data release, we provide the best-matched host galaxy for each transient event in YSE DR1. We take the (R.A., decl.) coordinates of each object and associate the object with a host galaxy by cross-referencing the results of two host–galaxy association codes, GHOST (Gagliano et al. 2021) and Sher-lock[61] (Smith et al. 2020), followed by visual inspection for final confirmation. This process is explained as follows.

GHOST is a database of 16,000 Pan-STARRS1 spectroscopic SNe and the catalog-level properties of their host galaxies. It

also contains analysis tools for associating new transients. Using the GHOSTpackage,[62] we provide the (R.A., decl.) coordinates and the astronomical transient (AT) name to the software. GHOST first performs a search for a matching SN name and/or coordinates within its preassociated database. If no match is found, it conducts a 30″ cone search in Pan-STARRS1 and removes stars to construct a list of candidate host galaxies.[63] A final association is made using a combination of the directional light-radius (DLR) method at the catalog level and a gradient ascent (GA) method at the postage stamp level.[64] The code is triaged, such that the DLR method is preferred, and GA is conducted only where no galaxies are found within four DLR radii of the transient or if any candidate galaxy within the 30″ candidate radius is missing size estimates in PS1. In the GA algorithm, a tracer starts at the transient location and updates its location following the gradients in the image. This method requires the selection of a step size, which is chosen whether the true host is presumed to be *large*, *medium*, or *small* based on the image intensity in the pixels surrounding the transient. It also uses this information to inform its final association:

1. if *large*, GHOST picks the closest NED-identified galaxy within 20″ of the final location from GA;
2. if *medium*, GHOST picks the closest NED-identified galaxy within 5″ of the final location from GA;
3. if *small*, GHOST picks the closest NED-identified source explicity not identified as a star within 5″ of the final location from GA.

In this way, the GA method is better able to locate the true host galaxy center even when HII regions and other galaxy substructures dominate the field.

---

[61] https://github.com/thespacedoctor/sherlock

[62] https://pypi.org/project/astro-ghost/
[63] For this work, we used the starcut = 'gentle'', ascentMatch = True arguments.
[64] https://ps1images.stsci.edu

Photometric Sample / Spectroscopic Sample
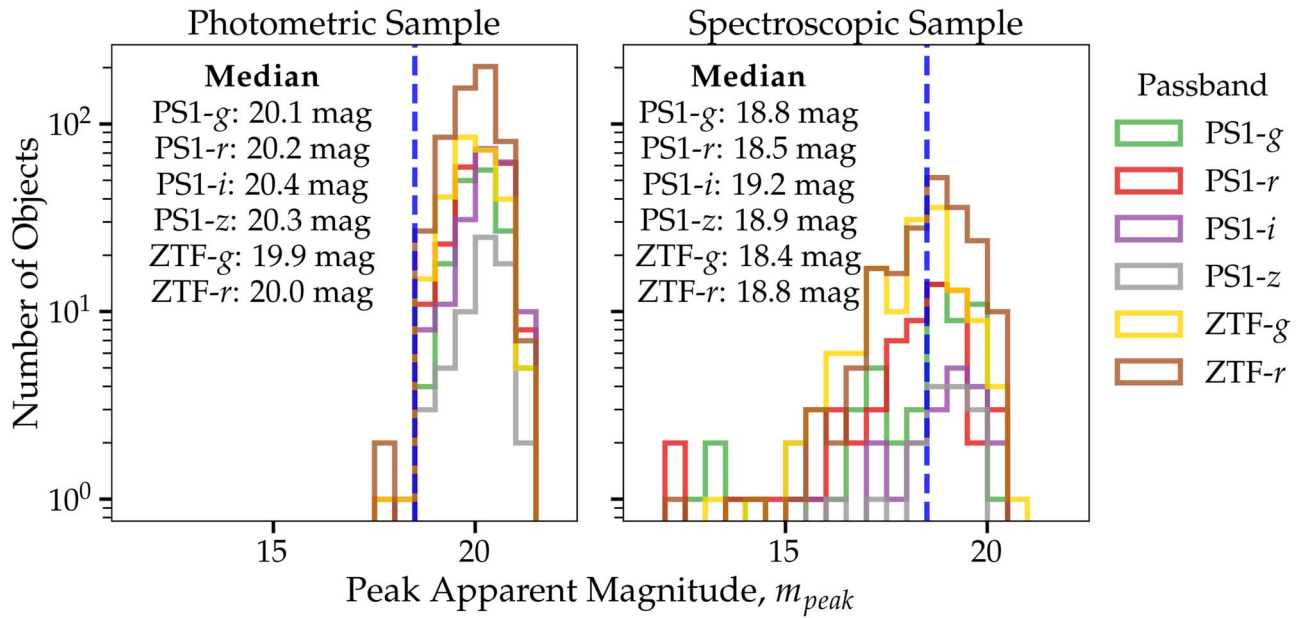
Figure 10. Histograms of the peak apparent magnitude m_{peak} of the photometric sample (left) and spectroscopic sample (right), with one m_{peak} value per object across all passbands. The dashed blue line indicates the r-band magnitude-limited survey threshold. We report four SN-like transients (AT 2020ebc, AT 2020fci, AT 2020rkp, AT 2020tkw) from the photometric sample that exceed $m_{peak,PS1-r} \leqslant 18.5$ mag but were missed for spectroscopic follow-up observations. See text for further details.

Simultaneously, we run the same 1975 YSE DR1 objects through `Sherlock`. `Sherlock` is the QUB transient classifier for Lasair, and provides a massive catalog crossmatch with star, galaxy, AGN, X-ray, and radio catalogs for transient classification via parameters of matched sources and contextual information. `Sherlock` is a boosted decision tree algorithm that calculates angular and physical separations of cross-matched objects, which we leverage for its host association capabilities.

After associating hosts independently, we then cross-reference their associations to remove any discrepancies (i.e., two different potential hosts $> 2''$ for the same transient event). To do so, we visually inspect the host association result of each transient using PS1 postage stamps. If both `GHOST` and `Sherlock` host associations are within $2''$ of each other, and within $2''$ of the transient with a detectable host (and without another nearby viable host), we deem these host associations as a match. From the match, we assign the `GHOST` host contextual information (R.A., decl., PS1 Object ID, etc.) to this transient. If either `GHOST` or `Sherlock` identifies a host $> 2''$ from the other, we resolve the discrepancy and assign a final host by visual inspection of the PS1 postage stamp through additional services such as the DESI Legacy Survey Imaging Surveys[65] and the SDSS SkyServer tool.[66] Moreover, for host associations where the suspected host galaxy center is $> 2''$ from the transient location (typically large angular size, bright, nearby galaxies), we inspect a larger FOV postage stamp to confirm the association. Lastly, through the aforementioned vetting tools, we investigate any cases of an apparent hostless transient (neither algorithm associates a host) to discern whether it is an algorithmic failure due to an artifact, a very low surface brightness host, or whether the transient may truly be hostless.
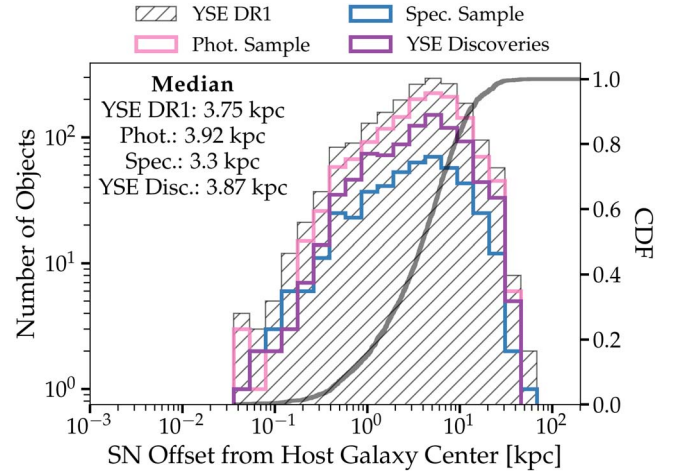
Figure 11. SN offsets (in kiloparsecs) from their vetted host galaxy centers for YSE DR1. Those not shown (62 objects) either are hostless or have an ambiguous host association. The cumulative distribution is shown as a solid gray line.

Examples of PS1 postage-stamps with `GHOST` (blue) and `Sherlock` (orange) host association results are shown in Figure 12.

Of the 1975 YSE DR1 objects, `GHOST` and `Sherlock` agree on 1428 (72%) host associations, 324 (16.5%) have a `GHOST`-only association, 97 (5.0%) have a `Sherlock`-only association, 64 (3.3%) are likely hosts found in vetting that were missed by both algorithms, 58 (3.0%) have no visible host, and 4 (0.2%) have an ambiguous association (i.e., could be one of several hosts). From this list, we take the vetted host galaxy coordinates and query the literature for host galaxy spectroscopic redshifts.

We query the vetted host galaxy coordinates with `astro-query` (Ginsburg et al. 2019) using the NED database. The

---
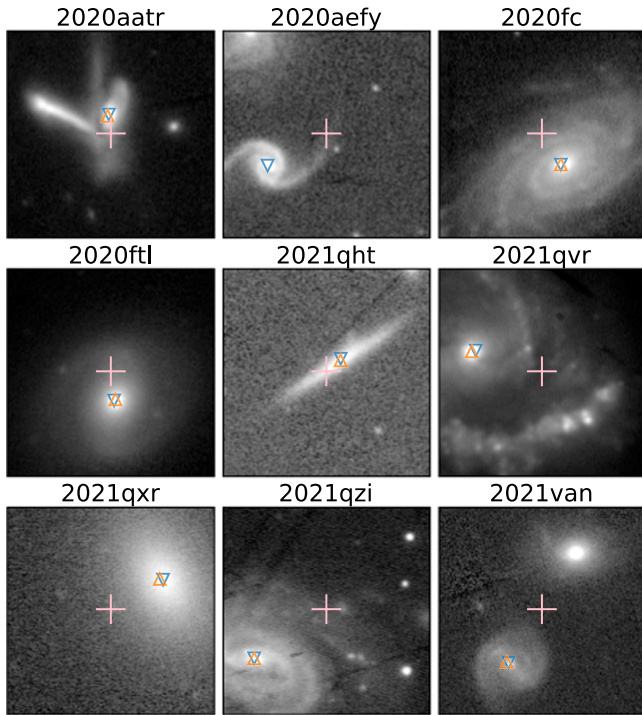
Figure 12. PS1 postage-stamps marking the location of nine transients (pink crosshair "+") and the best-guess host galaxy center from GHOST (blue triangle) and Sherlock (orange triangle) for a variety of galactic environments (merging, spiral, elliptical) and orientation (face-on, edge-on). We visually vet each association, particularly for cases where one association is missing (2020aefy), or for potentially ambiguous associations (2021van).

two main sources for obtaining host galaxy spectroscopic redshifts are SDSS and WISE All-Sky (WISEA) data. We query with the coordinates output from GHOST and/or Sherlock[67] after the vetting process, using an annulus of 6″. For the associated redshift, we pick the nearest candidate returned by the search if the candidate is not a classified SN, because we do not want to contaminate the sample with less robust spectroscopic SN redshifts. Moreover, if the nearest candidate provided only a photometric redshift estimate, we take the associated host galaxy data but favor our photo-z estimate (Section 5.4) instead.

5.3. The Easy PhotoZ Redshift Estimator

Our photometric classifier, based on the ParSNIP architecture, requires a value of redshift for K-corrections and to calculate a pseudoluminosity L (see Section 5.8.1 for details). Of the 1975 objects in YSE DR1, 492 have a spec-z measurement of the SN. Of the remaining, 301 have only a spectroscopic host-z from NED ($\sim$15% YSE DR1). This leaves 1180 objects that require a photometric redshift. We remind the reader that, although the YSE DR1 data files contain any spec-z of the SN or host galaxy, the ParSNIP classifier is trained, tested, and validated, and makes predictions using the photo-z value only. We utilize photo-zs from Easy PhotoZ as implemented into GHOST (Gagliano et al. 2021) for K-corrections inside our classification algorithm because (1) the vast majority of all transients now and especially during the

LSST era will only have a photo-z estimate; (2) we want to simplify the method and use a homogeneous redshift source; and (3) we used a mock redshift value when training the ParSNIP model to validate its performance against redshift errors and biases from current photo-z estimators.

In spite of other large redshift surveys of the nearby universe, popular choices, like the GALEX (Gil de Paz et al. 2007), WISE (Wright et al. 2010), Two Micron All Sky Survey Redshift Survey (2MRS, Huchra et al. 2012), and the 2dF galaxy Redshift Survey (2dFGRS, Colless et al. 2001), either are not deep enough or do not cover a significant enough portion of the sky for the YSE DR1 sample. For example, the 2MRS covers 91% of the sky at a 90th percentile depth of $z \sim 0.05$, and 2dFGRS covers 8% of the sky at a 90th percentile depth of $z \sim 0.19$ (Huchra et al. 2012). The only substantive source of sky coverage with depth akin to YSE DR1, and significant spectroscopic overlap with spectra or photometry of Pan-STARRS-cataloged galaxies is the Sloan Digital Sky Survey (SDSS, Margon 1999). For these reasons, we developed an empirical photo-z estimator for YSE trained on Pan-STARRS galaxy photometry and its crossmatched SDSS galactic spectroscopic redshift data called Easy PhotoZ.

YSE's current photo-z pipeline consists of two internal photo-z methods: a multilinear regression model trained on SDSS DR16 (Ahumada et al. 2020) data from Beck et al. (2016), and an MLP model trained on Pan-STARRS1 data inspired by Beck et al. (2021). The primary reason we do not use the PS1 3π Data Release 1 photo-z catalog from Beck et al. (2021) is that we additionally estimate a posterior redshift distribution along with a point estimate with uncertainty. Easy PhotoZ is trained on Pan-STARRS Data Release 2 Kron, PSF, aperture (R5, R6, R7) mean fluxes, and mean aperature fluxes across PS1-$grizy$ passbands with their associated uncertainties and dust extinction along the line of sight using the dustmaps of Schlegel et al. (1998).

Spectroscopic data were collected from SDSS DR16 (Ahumada et al. 2020), the DEEP2/3 survey (Cooper et al. 2011; Zhou et al. 2019), the Galaxy and Mass Assembly project (Baldry et al. 2018), the VIMos Public Extragalactic Redshift Survey (VIPERS; Guzzo et al. 2014; Scodeggio et al. 2018), the zCOSMOS survey (Lilly et al. 2009), the Six-degree Field Galaxy Survey (Jones et al. 2009), the WiggleZ Redshift Survey (Drinkwater et al. 2018), and the 3D-HST Survey (Brammer et al. 2012). Table 6 describes the contribution of usable redshifts from each survey after matching onto Pan-STARRS-cataloged galaxies with the methods described below.

We combine the data from these surveys to increase the training set size and reduce any bias introduced by the selection criteria from any one experiment. We note that the Easy PhotoZ training, testing, and validation data sets do not represent a random sample from the local population of galaxies. This choice and its bias on photometric redshift estimation are still a topic under active research.

After assembling the spectroscopic redshift data for training, we queried MAST[68] for all objects in Pan-STARRS within 2″ of any astrometric pointings from the spectroscopic surveys. For those objects within 2″, we then searched for objects satisfying $B > 10000$, where B is the "Bayes factor"

---

[67] If neither GHOST nor Sherlock found the right host via manual inspection, we took the coordinates from the Legacy Survey.

[68] https://archive.stsci.edu

Table 6
Contribution of Usable Redshifts from Crossmatched Surveys for `Easy PhotoZ`

Survey Name	No. Before Cuts	No. After Cuts
SDSS	2,840,216	624,039
VIPERS	88,185	398
DEEP	19,775	57
3D-HST	5078	47
6dF	124,647	322
WiggleZ	148,563	1773

Note. From each spectroscopic survey, we require a position and redshift of the object. The number of raw samples is those that passed the quality checks from the original surveys before quality cuts. The samples used are those after quality cuts: crossmatching onto Pan-STARRS, removing duplicates, within our redshift range, and after downsampling to a more uniform redshift distribution. There were zero usable samples for GAMA and zCOSMO after cuts, so these surveys were omitted.

(Budavári & Szalay 2008, Equation (16)):

$$B = \frac{2}{\sigma_1^2 + \sigma_2^2} \exp\left(-\frac{\psi^2}{2(\sigma_1^2 + \sigma_2^2)}\right). \qquad (1)$$

Here, σ_i is the total astrometric error from survey i, and ψ is the angular distance between objects. With multiple observations through many instruments of potentially varying astrometric accuracies, the Bayes factor is a measure of how likely observations from many surveys are from the same source. If two objects simultaneously satisfied $B > 10,000$, we selected the object with the higher score. We then removed any objects that had multiple spectra matched to its Pan-STARRS1 object identifier.[69]

Because many orders of power are detrimental to neural networks, we convert the flux values to inverse hyperbolic sine magnitudes m ("luptitudes," Lupton et al. 1999; Stoughton et al. 2002) for continuous flux scaling in fainter magnitude regimes using the equation

$$m = \frac{-2.5}{\ln 10}\left[\text{asinh}\left(\frac{f/f_0}{2b}\right) + \ln b\right], \qquad (2)$$

for filter zero-point f_0 and softening parameter b. In this manner, we achieve definite values with finite errors as the flux goes to zero. This approach differs from the treatment given in Beck et al. (2021).

We chose to downsample the `Easy PhotoZ` training set to create a roughly even number of samples between redshifts $0.03 < z < 0.38$, which mimics the result of downsampling from Zhou et al. (2021) and encompasses the redshift limits of YSE. After downsampling, we record how many samples are in each class and invert the value to calculate a relative weight, and set the maximum to 20. This is done to minimize the bias from individual surveys' targeting guidelines. For example, given that the overwhelming majority (99.5%) of the training samples are taken from SDSS, it is likely that red, luminous galaxies are overrepresented in the training set. As a final processing step, we

subtract the median and scale by the interquartile range of the training set to normalize the features. This leaves us with 626,636 training examples at $z < 1$. Given the final processed data set, we employ a standard train–validation–test split of 70%/15%/15%, respectively. The training was performed on an NVIDIA GTX 1660 Ti and finished in approximately 1 hr. Details on the MLP architecture and further supplementary materials can be found in Appendix A.

5.4. Photo-z

The Pan-STARRS1 5-layer MLP photo-z estimator is integrated into GHOST (Gagliano et al. 2021). It also goes beyond providing the standard point estimate redshift value by simultaneously estimating the redshift posterior density, $P(z)$. We use $P(z)$ and an independent absolute magnitude-informed probability $Q(z)$ to calculate a final photo-z.

First, we query GHOST for all vetted host galaxy-associated PS1 object ID matches. We provide these 1862 PS1 host galaxy IDs into `Easy PhotoZ`, and using the *calc_photo-z* function from GHOST, return $P(z)$ for $0 < z < 1.0$. For the remaining non-GHOST PS1-associated objects (e.g., no visible host), we assume a simple redshift posterior density to be a uniform distribution $P(z) \sim U(0, 1)$.

Separately, we apply an independent absolute magnitude-informed probability $Q(z)$ for $0 < z < 0.3$. Because the redshift limits for SN detection of YSE and ZTF are effectively $z \approx 0.3$, we restrict the estimated redshift posterior density results to this range. $Q(z)$ is a limiting z-range corresponding to a luminosity distance modulus $\mu_{z, M} = m_{\text{peak}} - M_{-13,-22}$, where $M_{-13,-22}$ is the absolute magnitude in $M \in [-13, -22]$, and m_{peak} is the SN's peak apparent magnitude.[70] Although $M \in [-15, -21]$ is an acceptable range for SNe absolute magnitudes excluding SLSNe (Richardson et al. 2014), we extend this range to $M \in [-13, -22]$ to safely account for small but unquantified errors regarding a K-correction in the calculation of $\mu_{z,M}$ if m_{peak} and $M_{-13,-22}$ are in the same band, variation in m_{peak} (due to discovering an object post-peak), and any other errors due to m_{peak} occurring in different passbands across the sample. For example, if an SN has $m_{\text{peak}} = 18$, then we use the `astropy` *z_at_value*[71] function to retrieve a possible redshift range $Q(z) \sim U(0.003, 0.203)$. In this calculation, we use the same flat ΛCDM cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and $\Omega_M = 0.3$. With the redshift posterior density $P(z)$ and independent absolute magnitude probability $Q(z)$ for each SN, we can update our redshift posterior density $P^*(z)$ via the normalized product of these probabilities. Then, we can determine a new point estimate to use as the photo-z by calculating the expectation value of the updated redshift posterior density. It follows that the photo-z error is the standard deviation, calculated as the square root of the variance of the updated redshift posterior density.

We calculate a photo-z and photo-z error for each object in YSE DR1. To determine if our photo-z values are reliable, we compare to known SN spec-z in Figure 13. Here we show a scatterplot and rms values comparing the agreement of spec-z and photo-z for the spectroscopic sample, with marginal

---

[69] In the future, a criterion for accepting an object with multiple spectral matches should be developed. Moreover, while this matching scheme is based on a probabilistic model, we did not investigate the validity of that model or its assumptions. We expect some pairings are not true matches, but rather contribute to label noise. We leave this value unquantified and leave it for future work.

[70] In keeping with our classifier's final prediction scheme, we assume all objects after cuts in YSE DR1 are SN-like, and fall within SN Ia, SN II, SN Ibc classifications. Thus, the assumed possible absolute magnitude-range $M \in [-13, -22]$ conservatively incorporates these SNe classes.
[71] https://docs.astropy.org/en/stable/api/astropy.cosmology.z_at_value.html
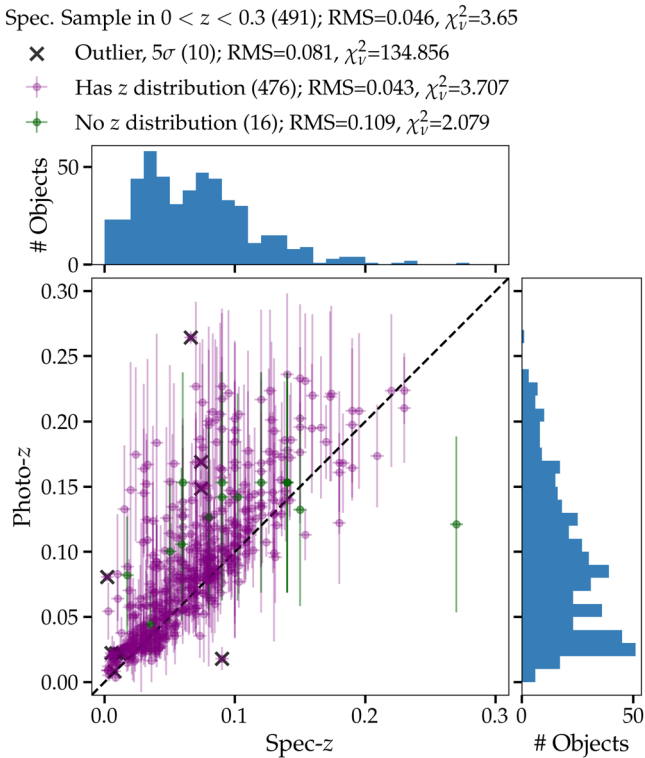
Figure 13. A scatter plot comparing the agreement between spec-z and photo-z for the spectroscopic sample, with the marginal distributions shown as histograms. The redshift posterior density values are from `Easy PhotoZ` as implemented via `GHOST` (Gagliano et al. 2021). Photo-zs and photo-z errors are calculated after leveraging an additional absolute magnitude-weighted z range applied to the redshift posterior density posterior (see text for details). An object with a PS1 host association redshift posterior density from `Easy PhotoZ` is in purple, and an object without such a redshift posterior density (for which we assume a uniform distribution $U(0, 1)$) is in green. Because the redshift limits of YSE and ZTF are only $z \approx 0.3$, we restrict the estimated redshift posterior density results to this range (thus, SLSN-I 2021uwx at $z = 0.525$ is omitted for clarity). Outliers are marked by an "×." The root-mean-square and reduced chi-squared (χ_ν^2) values are individually calculated for these categories in addition to the entire spectroscopic sample.

redshift distributions shown as histograms. The color indicates the presence of a redshift posterior density as determined from a matched PS1 object ID via `Easy PhotoZ`, or an assumed uniform redshift posterior density for unassociated PS1 object catalog objects.

Overall, there is strong agreement between spec-z and photo-z, particularly at low z ($z < 0.1$), and in the PS1 catalog host-associated transients. There is less strong agreement in the high-z ($z > 0.1$) regime and in non-PS1 catalog host-associated transients, as expected, where there is tendency in the model to overestimate the photo-z.[72] The relatively low rms value for the entire spectroscopic sample (rms = 0.046) is vital for both obtaining realistic K-corrections for the photometric classifier, and providing confidence that the photo-z values for the photometric sample will be reasonable. For objects with a PS1 host association redshift distribution from `Easy PhotoZ` (purple), we find an average rms = 0.043.[73] This small offset is

---

[72] This behavior is also observed in Beck et al. (2021); see their Figure 3.
[73] Although not strictly a 1–1 match, the vast majority (~99.9%) of objects with a successful PS1 catalog host association has a returned redshift posterior density from `Easy PhotoZ`. The only object with a successful PS1 catalog host association but without a returned redshift posterior density is AT 2021xbd.

encouraging, because 1862 out of 1975 YSE DR1 transients (~95%) have a PS1 host association redshift distribution from `Easy PhotoZ`. For those without a PS1 host association within the spectroscopic sample, we find rms = 0.109. Both rms values are acceptable for photometric classification with ParSNIP. Moreover, we find that all nonoutlier objects collectively have a low reduced chi-squared (χ_ν^2) value on the order of $\chi_\nu^2 \sim 1$, implying that the residuals on average are on the order of the combined spec-z error and photo-z error.

To investigate the photo-z outlier fraction, we deem photo-z outliers as follows:

$$|\text{spec-}z - \text{photo-}z| < 5 \times \text{photo-}z \text{ error.} \qquad (3)$$

Of the 10 outliers (black "×"), the majority are in the lowest spec-z regime ($z < 0.01$) where the photo-z error is unrealistically small (photo-z error ~ 0.005) and is of the value of the spec-z error (~ 0.005).[74] This is represented by the high reduced chi-squared (χ_ν^2) value of $\chi_\nu^2 \sim 100$, indicating that the photo-z model produces a few catastrophic out-of-distribution samples (particularly at high z), as well as a few underestimated outlier errors at very low z ($z \leqslant 0.01$). A possible reason for this is the lack of representative samples below $z \approx 0.03$, the threshold at which `Easy PhotoZ` is downsampled in training. Thus, there are likely fewer examples of extremely local hosts when training the MLP. Because YSE's volume-limited survey extends to $z \approx 0.06$, in practice such local SNe will have a spec-z. The few outliers remaining are in the high-z range, and also have underestimated errors. Overall, the agreement between photo-z and spec-z value is excellent, and is consistent with the 1–1 relation of spec-z and photo-z values.

5.5. Vetting Public Spectroscopic Classifications

To ensure a consistent classification methodology of the spectroscopic sample, we reclassify all YSE DR1 objects for which there is at least one spectrum with the Supernova Identification package (`SNID`; Blondin & Tonry 2007). Such spectra were obtained through either YSE follow-up observations or public spectra posted to TNS from other observing groups. We have over 1100 total spectra across the spectroscopic sample of 492 objects. Of these, over 25% have multiple spectra. There are two objects (SN 2020ej, SN 2020lrr) that have a spectroscopic label on TNS but no public spectra. We adopt such labels.

We use the 5.0 version of `SNID` for classification, with additional template sets from the Berkeley Supernova Ia Program (Silverman et al. 2012; Liu & Modjaz 2014; Modjaz et al. 2014; Liu et al. 2016; Modjaz et al. 2016; Gutiérrez et al. 2017; Williamson et al. 2019). Thus, our implementation consists of 6145 spectra from 811 templates. Lastly, we use the forcez argument for any object that has a known host-z, as described in Section 5.2.

Our `SNID` classifications agree with the public TNS classifications on the vast majority of the spectroscopic sample. For the remaining, we use the `SNID` reclassifications and not the TNS label. The reclassifications can be found in Table 7.

---

[74] Additionally, we note that any effect from peculiar velocities is within the spec-z error (Davis et al. 2019).

17

Table 7
SNID Reclassifications

SN	TNS Classification	New Classification	Remarks
2020able	Ic[1]	Ibn	See Hiramatsu et al. (2020).
2020acct	IIn[2]	Ic[b]	⋯
2020apw	⋯	SN	Spectrum dominated by noise. No SN redshift determination.[a]
2020awu	⋯	SN	Blue featureless continuum.
2020bwr	II[3]	IIn	Presence of narrow Hα.
2020epi	Ic[4]	⋯	Spectrum quality is too poor to render a classification.
2020esm	IIP[5]	Ia-SC	See Dimitriadis et al. (2022).
2020ghq	⋯	II	Coincident with SN 2021mnj (Dahiwale & Fremling 2021a).
2020qlq	Ib/c[6]	Ic	In the nebular phase. Match to SN 2011bm (Valenti et al. 2012).
2020rdu	⋯	IIn	Hα has narrow, intermediate, and broad components.
2020tlf	IIn[7]	II	See Jacobson-Galán et al. (2022b).
2021aaxi	⋯	Ia	Matches to SN 2002fk (Cartier et al. 2014), SN 2001N (Jha et al. 2001).
2021acjv	⋯	II	Matches to SN 2011fu (Kumar et al. 2013), SN 2006iw (Morrell et al. 2006) at phase $\sim$−10 days.
2021aeuw	⋯	Ic[b]	ΔI_{X-Y} score agreement (Quimby et al. 2018).
2021bug	II[8]	LBV	Similar to AT 2009ip (Foley et al. 2011) outbursts.
2021dib	I[9]	Ic-BL[b]	Match to PTF10qts (Walker et al. 2014). ΔI_{X-Y} score agreement (Quimby et al. 2018).
2021gcv	⋯	Ia[b]	Galaxy dominated, but match to SN Ia at phase $\sim$+50 days.
2021gno	Ib[10]	Ib-pec	See Jacobson-Galán et al. (2022a).
2021joz	⋯	II	Match to SN 2006bp (Quimby et al. 2007) at phase $\sim$+4 days.
2021kqp	⋯	SN	Dominated by early flash ionization lines at $z = 0.1$.
2021lfv	⋯	Ia	Matches to SN 2001ic (Chornock & Filippenko 2002), SN 1996X (Salvo et al. 2001; Hicken et al. 2017) at or around peak brightness.
2021lzg	⋯	Ia-91T	Matches to SN 2002hu (Scalzo et al. 2019), SN 1999aw (Strolger et al. 2002).
2021qzp	Ib/c[11]	Ic-BL	Matches to SN 2003jd (Valenti et al. 2008), SN 2007ru (Sahu et al. 2009).
2021sje	⋯	Ib	Matches to SN 2005bf (Tominaga et al. 2005; Folatelli et al. 2006), SN 2007Y (Stritzinger et al. 2009).
2021uiq	Ia-91T[12]	Ia-CSM	Broad Hα emission, long-lived light curve indicating CSM interaction.
2021utd	⋯	Ic	Strong match to SN 1994I (Clocchiatti et al. 1996; Millard et al. 1999) at or after peak brightness.
2021uwx	⋯	SLSN-I	Fe II, Mg II, and Mg I lines at $z = 0.525$. $M_{\rm peak} \sim$ -22 mag.
2021xbg	Ib/c[13]	Ic	Matches to SN 2017ein (Van Dyk et al. 2018), 2007cl (Foley et al. 2007) before peak brightness.
2021xvu	⋯	Ia	Match to SN 2006cc (Ponticello et al. 2006).

Notes. Updated classifications of our spectroscopic sample. For objects not listed here, we use the public TNS label.
References. (1) Hiramatsu et al. (2020); (2) Siebert et al. (2020e); (3) Do et al. (2020); (4) Leadbeater (2020); (5) Tucker et al. (2020); (6) Dimitriadis et al. (2020b, 2020c); (7) Balcon (2020a); (8) Dahiwale & Fremling (2021b); (9) Moran et al. (2021); (10) Dahiwale & Fremling (2021c); (11) Dimitriadis et al. (2021); (12) Taggart et al. (2021a, 2021b); (13) Csoernyei et al. (2021.
[a] Classification is uncertain.
[b] We use our photo-z value instead.

5.6. YSE and ZTF Simulations

Here we outline the process used to generate the photometric classifier training sample: SNe simulations in the YSE and ZTF surveys. We detail creating individual simulation cadence libraries (*SIMLIBs*) from forced-photometry data of both YSE and ZTF surveys as well as discuss the training sample of generated SN Ia, SN II, and SN Ibc to mimic real observed events.

5.6.1. SIMLIB Generation

We generate YSE and ZTF survey simulations with the SuperNova ANALysis software (SNANA[75]; Kessler et al. 2009).

SNANA requires a SIMLIB file that describes the seeing, sky-noise, zero-points, and cadence of a survey for each pointing. It is the reference with which SNANA generates survey-specific simulations. Note that the SIMLIB uses information generated from survey images (e.g., scales flux errors) but does not use pixels or images directly. SIMLIBs can be generated from either a library of observations containing PSF FWHM, sky-noise, zero-point, gain, and filter, or directly from a data

sample. For this work, we follow the latter case, and create the SIMLIB from a subset of the YSE DR1 data sample.

For the SIMLIB, the subset started with the transient light curves and metadata, which includes the YSE+ZTF forced photometry, cadence information, coordinates, and redshift. Because the SNANA SIMLIB generator requires a redshift for each entry, we used the spectroscopic redshift when available; otherwise, we used the photometric redshift using the methodology outlined in Section 5.4. For the SIMLIB generation only, we simplify the methodology and use photo-zs from only the GHOST host galaxy associations (because this step was performed before final host galaxy vetting of Section 5.2). This is appropriate because SNANA generates a unique redshift per simulated light curve that varies from the template used to generate it, and as will be explained in Section 5.6.3, we slightly augment the simulated light curves' redshifts randomly. This produces a simulated redshift distribution that both recreates the overall shape of the real data redshift distribution and emphasizes a higher randomized concentration at low redshifts ($z < 0.1$)—which follows that of YSE DR1 objects—to be used in training. In principle, ParSNIP should be agnostic to the redshift of the transient,

---
[75] https://github.com/RickKessler/SNANA
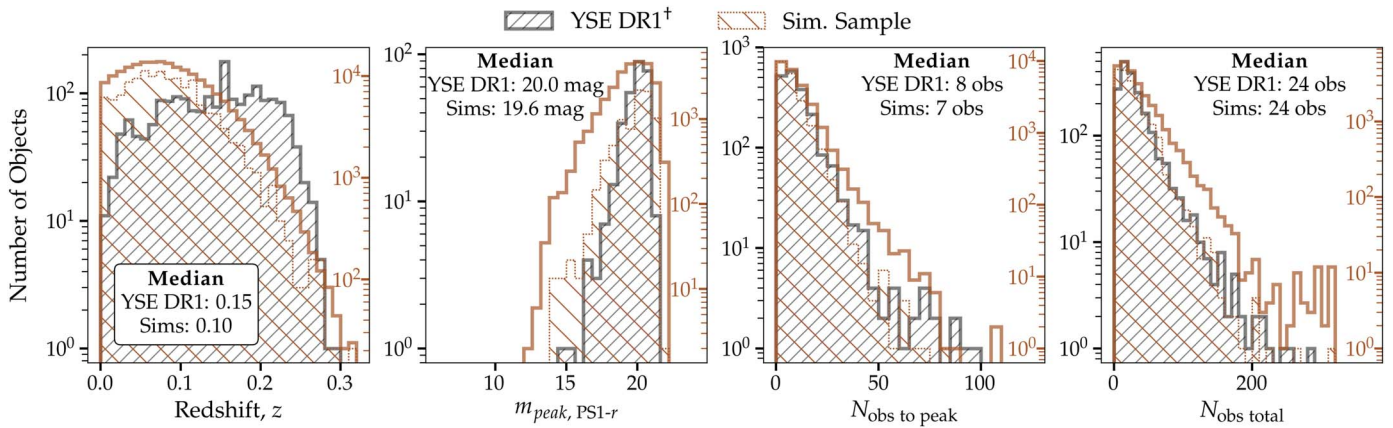
Figure 14. Histograms comparing the distributions of untargeted YSE DR1 objects (gray, hatched) and the same number of randomly sampled objects from the simulated sample (brown, hatched) across four parameters: redshift, (z; left), peak apparent magnitude for PS1-r passband ($m_{peak, PS1-r}$; center left), the number of observations until the light-curve peak (in magnitudes, any passband) ($N_{obs\ to\ peak}$; center right), and ($N_{obs\ total}$; right). The full, unnormalized number of simulations for each parameter is shown as a thick brown line and marked by the secondary y-axis with inward-facing tick marks. Despite any discrepancy, we display the number of simulated samples used to train ParSNIP on the right axis to show we have sufficient coverage at all parameter values for training.

but the choice was motivated to faithfully recreate statistical properties of YSE DR1.

Next, we keep only transient events with observations spanning all passbands (a requirement of SNANA when generating a SIMLIB from a data sample): PS1-*griz*, and ZTF-*gr*. Finally, a SIMLIB is made using SNANA's snana.exe script, which writes out a row for each epoch with a matched cadence, passband, Milky Way extinction, redshift, and metadata like PSF, sky-noise, and zero-point values for unique (R.A., decl.) pointings of each object in the data sample. To do so, the PSF is fixed to $1''$ FWHM, and the zero-point for each observation is computed in order to match the measured S/N in the data. For this script, we additionally supply a file with appropriate filter transmissions, native magnitude for each filter, spectral energy distribution (SED) of the primary reference, and other information for PS1-*griz*, ZTF-*gr* passbands. The remaining YSE transients events are used to populate the SIMLIB. We then scale the flux uncertainties as a function of S/N to better match the uncertainties of the data.[76]

Ultimately, we have 785 unique transient event realistic observing conditions with complementary YSE and ZTF photometry in one combined SIMLIB file. It is this SIMLIB that we used to create the full training sample.

5.6.2. Simulation Selection Effects

SNANA requires efficiency and detection logic files to best capture the nuance in subtle survey observational strategies. Simply, the efficiency file is used to model a survey's search efficiency to evaluate the chance of a detection per passband per a range of S/N values; in our case, to characterize the YSE+ZTF image subtraction pipelines via S/N versus magnitude values.

For all passbands, we adopt modified efficiency properties used in Jones et al. (2021a): a simple 0% efficiency for S/N < 4 and a 100% efficiency for S/N $\geqslant$ 4 (instead of 0% efficiency for S/N < 5 and a 100% efficiency for S/N $\geqslant$ 5), which we found reasonably replicated the magnitude limits, the distribution of S/N observations when compared to YSE DR1
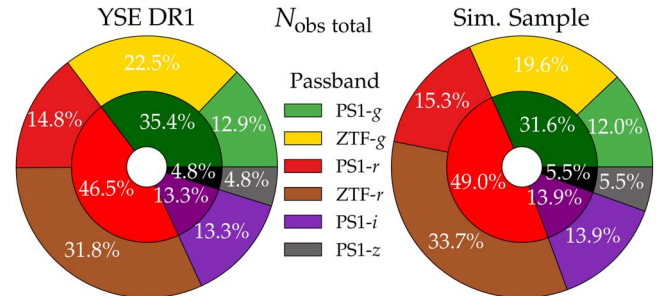
Figure 15. Nested pie charts of the total number of observations $N_{obs\ total}$ per YSE and ZTF passband (displayed as a percentage) for YSE DR1 (left) and a random subset of the entire simulated sample (SN Ia, SN II, SN Ibc; right) such that both pies have 1975 objects. The inner wedges are the result of combining similar optical YSE and ZTF passbands—PS1-g+ZTF-g and PS1-r+ZTF-r. (PS1-g, green; PS1-r, red; PS1-i, purple; PS1-z gray). We apply the same cuts on the simulated sample as we do on YSE DR1 (e.g., S/N > 4), only using observations, and do not include nondetections. Pie charts of the total number of observations ($N_{obs\ tot}$) per YSE passband only are found in Figure B3.

(see Figure 14), and the total number of observations per passband (see Figure 15).

Although the search efficiency information provides the probability of a single-epoch detection across passbands, it does not specify whether the SN would be observed or *discovered*. Thus, this second selection effect is known as the discovery or trigger logic. For the combined YSE+ZTF simulations, we use a trigger logic requiring at least two epochs using any combination of the YSE passbands. We do not require a trigger logic associated with the ZTF passbands, because we only interleave ZTF observations when available, which is dependent on sky position, field selection, weather, telescope downtime, etc. That being said, the vast majority of the simulations naturally are complemented with ZTF observations, as is the case with YSE DR1. Thus, the combination of search efficiency and logic files determines which simulated SNe are discovered (here, two S/N $\geqslant$ 5 detections in any YSE passband constitutes a discovery).

All simulated events with redshifts are true *spectroscopic* redshifts, without host galaxy photo-z. However, we do create a mock host-z for training (see Section 5.6.3).

---

[76] For further explanation and implementation of these legacy noise corrections, see the SNANA manual Section 4.13.5 (2022 March 7 version).

Table 8
YSE DR1 SNANA Simulation Breakdown for the Training Set

Model Type	Template	N_{sim} Generated	Training	Testing	Validation	Citation
SN Ia	SALT3 SED	62,596	37,501	12,511	12,583	Kenworthy et al. (2021)
SN II	NMF	65,127	9858	3249	3168	Kessler et al. (2019)
SN Ibc	Spectral time-series (Spectrophotometric)	64,203	2588	889	899	Vincenzi et al. (2019)

Note. Simulated training, testing, and validation sets are rebalanced for the random forest classifier step based on the ZTF BTS supernovae class fraction during 2019 November 24 to 2021 December 12 (see Figure 22).

5.6.3. Training Sample

We used SNANA (Kessler et al. 2009) to then create our training set assuming a flat ΛCDM cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $w = -1$. Due to SNANA's capability to produce catalog-based simulations with survey-specific noise properties, SN rates, and detection efficiencies, well-constructed simulations can closely reflect a survey's observed SNe with accurate statistics and proper cadence. Moreover, we chose to use SNANA for its built-in suite of transient SED models originally developed (and continuously updated; see Modelers 2021) for the PLAsTiCC SN identification challenge (Kessler et al. 2019; Hložek et al. 2020), and for the upcoming ELAsTiCC challenge (The ELAsTiCC team; LSST Dark Energy Science Collaboration). These SN templates, paired with the survey properties via the SIMLIB, are used to generate our YSE+ZTF training set simulations.

We simulated ∼60,000 of each of SN Ia, SN II, SN Ibc. For the SN Ia model, we use the recent SALT3 noSED model template from Kenworthy et al. (2021). It is an improvement over the SALT2 SN Ia model used in the PLAsTiCC challenge when considering color separation, light-curve stretch, and having publicly available training code. For CC SNe, we simulate several hydrogen-rich (SN II) models for the purpose of selecting the choice that would give the best performance. Of these, the first model type we simulated was a nonnegative matrix factorization (NMF)-based model from Santiago Gonzalez-Gaitan and Lluis Galbany. We also simulated an older model from Jones et al. (2017) with corrected SED magnitude offsets and magnitude smearing weights, combined with the addition of four SN IIb templates to the existing suite of SNANA SN II SED templates at the time. Finally, we simulated a more recent spectral–time-series-based template from Vincenzi et al. (2019). The SN II models from Vincenzi et al. (2019) can be separated by three subtypes (IIP/IIL, IIn, IIb), but we generated simulations spanning all subtypes without distinction to form a broad SN II class. Similarly, for hydrogen-stripped CC SNe (SN Ibc), we simulated the combined results of the spectral–time-series-based SN Ib and SN Ic classes from Vincenzi et al. (2019) to provide a singular SN Ibc class. After training on many permutations of SN II models, we achieved the best accuracy across the three SN classes while maximizing the completeness and purity of SN Ia using a training set generated from the SALT3 SN Ia model (Kenworthy et al. 2021), the NMF-based SN II model, and the SED-based SN Ibc model (Vincenzi et al. 2019). The breakdown of model type, number of simulations, and model template details for the training set are listed in Table 8.

Lastly, we randomly assign a mock host-z in place of the original redshift value generated for all simulated training set light curves using $z \sim N(\text{host-}z, 0.05^2)$. This mock-z acts as a photo-z value for the simulated light curves. We want to balance a strong classifier performance while not biasing our classifier with precise, spec-zs (generated from SNANA) when the majority of the YSE DR1 and future YSE objects will be classified using photo-zs.

Lastly, note that the simulated sample does not mimic the targeted field selection and its specific SN demographics (e.g., the 31 targeted YSE objects as described in Section 2.3 and listed in Table C1). The targeted YSE sample is biased toward particularly bright, young, and (very) local objects, and those that exhibit unusual spectral features. These objects are disproportionately favored for photometric and spectroscopic follow-up, and are not representative of the rest of the YSE DR1 sample. Bright, young, and (very) local objects are still generated for the simulated sample, but are only represented in an unbiased, untargeted manner.

We simulated the training sample in a wall time of 3.855 hr across 10 Intel Xeon "Haswell" processor nodes on the Cray XC40 "Cori" system at the National Energy Research Scientific Computing (NERSC).

5.7. Simulations and YSE DR1 Comparisons

To demonstrate the realistic nature of the simulations used in the training sample for YSE DR1, we perform several comparative tests between the simulated training sample and the untargeted YSE DR1 objects. In Figure 14, we compare normalized histograms between distributions of untargeted YSE DR1 objects and the simulated sample across four parameters: redshift, peak apparent magnitude for PS1-r passband ($m_{peak,PS1-r}$), the number of observations until the light-curve peak (any passband), and the total number of observations. For redshift, the simulated sample slightly overestimates the presence of nearby SN in low z ($z < 0.1$) as well as slightly underestimates the far reaches of the survey ($z \gtrsim 0.2$), with the latter discrepancy originating from small number statistics. Note that the YSE DR1 redshift histogram is the combined spec-z and photo-z values from the individual spectroscopic and photometric samples, respectively. Thus, the unknown true redshifts of photometric objects are an additional source of discrepancy. Despite any discrepancy, we display the number of simulated samples used to train ParSNIP on the right axis to show we have sufficient coverage at all redshifts for training.

For $m_{peak,PS1} - r$, the simulations very nearly match YSE DR1, with only a slight overestimation with regard to the brightest objects in YSE DR1 at $m_{peak,PS1-r} < 18$, which we ascribe to small number statistics. The median of the simulated sample is approximately half a magnitude brighter than its YSE DR1 counterpart. Due to the few bright simulated objects skewing the distribution brighter, this discrepancy is not particularly significant, especially when simulations do follow well at the faintest $m_{peak,PS1-r}$ values (and conservatively go half a magnitude fainter for training), which is where the bulk

of YSE DR1 lies. Lastly, simulations also match YSE DR1 excellently in $N_{\rm obs\ to\ peak}$ and $N_{\rm obs\ total}$ distributions and median values, meaning that the simulations replicate the correct number of total observations in the rise of the light curve up to its peak observation and the decline after peak, ranging from poorly sampled and likely faint events ($N_{\rm obs\ to\ peak} < 5$) to well-sampled and likely bright events ($N_{\rm obs\ total} > 100$). These discrepancies at $N_{\rm obs\ to\ peak} \gtrsim 50$ and $N_{\rm obs\ total} \gtrsim 200$ where the simulations underpredict YSE DR1 are also not particularly significant due to small number statistics (see Figure 9, center plot).

In Figure 15, we show the percentage of the total number of observations $N_{\rm obs\ total}$ per passband (YSE+ZTF) as nested pie charts between YSE DR1 (1975 objects) and a randomly selected subset of 1975 SN from the entire simulated sample. The inner wedges represent the aggregate optical passbands *griz*, where similar PS1-*g* and ZTF-*g* results are combined, as are PS1-*r* and ZTF-*r*. When only considering these four aggregate passbands, the agreement between the simulated and observed samples is excellent (all within 4%). When expanding to all six unique passbands, the agreement improves, where there is only between 0.6% and 2.8% difference in $N_{\rm obs\ total}$ per passband. This means that the simulated sample has the same relative proportion of total observations per passband as that of YSE DR1. This favorable match is due to simulations being generated from a SIMLIB drawn from the real data sample.

A third comparison test involves fitting the parametric analytical model from Villar et al. (2019), hereafter "Villar Fit," to both the simulated training sample and YSE DR1 light curves, and comparing the fitted parameter distributions. The agreement of these distributions would further evidence the simulated SNe are representative of YSE DR1 SNe.

The Villar Fit is a seven-free-parameter SN light-curve fitter, initially presented in Bazin et al. (2009; five free parameters), expanded upon by Karpenka et al. (2012; six free parameters), and modified again (Villar et al. 2019). It can adapt to a wide range of SN light-curve morphologies such as non-SN Ia and is reproduced below (Equation (1), Villar et al. 2019):

$$F = \begin{cases} \dfrac{A + \beta(t - t_0)}{1 + e^{-(t - t_0)/\tau_{\rm rise}}} & t < t_1 \\ \dfrac{(A + \beta(t_1 - t_0))e^{-(t - t_1)/\tau_{\rm fall}}}{1 + e^{-(t - t_0)/\tau_{\rm rise}}} & t \geqslant t_1 \end{cases}. \tag{4}$$

Because we use these parameters for a nearest-neighbors search, we reparameterize Equation (4) via $t_1 = t_0 + \gamma$ where γ is plateau duration. Additionally, we introduce a background baseline flux c:

$$f(t) = c + J \begin{cases} A + \beta(t - t_0), & t < t_0 + \gamma \\ (A + \beta\gamma)\exp\dfrac{-(t - t_0 - \gamma)}{\tau_{\rm fall}}, & t \geqslant t_0 + \gamma \end{cases}, \tag{5}$$

where

$$J = \frac{1}{1 + \exp\frac{-(t - t_0)}{\tau_{\rm rise}}}. \tag{6}$$

Here, $\tau_{\rm rise}$ is the rise time in days, $\tau_{\rm fall}$ is the decline time in days, t_0 is the predicted day of explosion (MJD), A represents the amplitude, β is plateau slope in flux per day, c represents the baseline flux, and γ is the new parameter plateau duration in

days. Villar et al. (2019) notes that the fit does not explicitly model the second peak in the *i* band of SNe Ia 1 month post explosion; however it is able to fit and parameterize SN Ia light-curve morphology well enough for our purposes.

To understand realistic free parameter values, we apply the Villar Fit to spectroscopically confirmed, well-sampled SN Ia from YSE DR1 across ZTF-*g* and ZTF-*r* bands. We note that the parameterized Villar Fit implementation[77] via the light-curve version 0.4 Python package[78] (Malanchev et al. 2021) requires at minimum eight observations per passband to fit the light curve. Because ZTF observations dominate the total observations due to its faster cadence (see Figure 15), and to keep the number of free parameters to a minimum, we only use the ZTF passbands. We do likewise for the SN Ia simulations that satisfy this requirement (19,005 objects). The fitted parameter distributions across all 14 variables are shown in Figure B5. Such distributions between the spectroscopic and simulated SN Ia sample have good agreement.

To further support the claim that the simulated SN Ia light curves well match the properties of the observed SN Ia light curves, we find the closest simulated SN Ia in the training sample to a spectroscopic SN Ia (2021mwb) via a nearest-neighbors search. First, we use principal component analysis (Jolliffe 2002) to reduce the 14D parameter space (seven free parameters in each ZTF-*g* and ZTF-*r*) to an 8D principal component space comprising 83% of the variance. Then we run a *k*D tree (Bentley 1975) to find the nearest neighbors in principal component space to 2021mwb.[79]

We plot the light curves of SN Ia 2021mwb and its closest simulated match (ID 4791133) in Figure 16. Although many parameter distributions span several orders of magnitude, the Villar Fit values for 2021mwb and its closest match are nearly identical. For example, if we define σ to be the number of standard deviations from the mean of the combined YSE DR1 and simulated training set parameter distributions, we find that ZTF-*r* $A = 0.0875\sigma$ (2021mwb), and ZTF-*r* $A = 0.0854\sigma$ (ID 4791133). See Figure B5 for the comparison across all parameters, as well as the next two closest matching simulations from the *k*D tree search (ID 910471, ID 1984218). With the simulated sample characteristic of YSE DR1, we proceed to training our ParSNIP photometric classifier.

5.8. ParSNIP

5.8.1. Hyperparameters and Training

Boone (2021) introduced a hybrid physics–VAE model called ParSNIP to characterize AT photometric light-curve profiles and their intrinsic time-varying spectra. This is done via a 3D latent representation robust to effects that affect the observed light-curve profile (e.g., redshift). The architecture of ParSNIP can be broken down as follows (see Figure E1, *C* process): a transient light curve is input to the encoder, which attempts to predict the posterior distribution in the form of latent variables. These latent variables can be assigned into two groups: explicit and intrinsic. The former represents the known

---

[77] https://docs.rs/light-curve-feature/0.3.3/light_curve_feature/features/struct.VillarFit.html
[78] https://github.com/light-curve/light-curve-python
[79] This process was inspired by Aleo et al. (2022), who used a similar method to find the closest match of simulated SNe light curves to real SNe light curves via a *k*D tree in light-curve feature parameter space.

observing *symmetries* of the model—observations that will change the light curve but not the underlying physics (e.g., propagation effects such as redshift, dust along the line of sight, varying cadences, different passbands across multiple telescopes). These extrinsic latent variables have known explicit functional forms ascribed $s_e = \{A, c, t_0\}$, where A is the amplitude of the light curve, c is the color (capturing dust reddening of the light curve), and t_0 is the reference time for the light curve. Meanwhile, the intrinsic latent variables represent the underlying diversity of transients that are insensitive to the aforementioned observing effects, whose functional form is unknown, denoted as $s_i = \{s_1, s_2, s_3\}$. Such intrinsic latent variables are then used as input to an intrinsic decoder model, which predicts the full time-varying spectrum of the transient. From there, the physics layer leverages the explicit latent variables and observational metadata to model the photometry of the recreated spectrum. Lastly, the trained ParSNIP model performs inference to estimate the latent representations for all YSE DR1 objects. The full set of VAE feature parameters include the following: the latent parameters and their errors s_1, $s_{1,\mathrm{err}}$; s_2, $s_{2,\mathrm{err}}$; s_3, $s_{3,\mathrm{err}}$; color and color error; reference time error; pseudoluminosity L and its error L_{err}. These features are used as input into a classification scheme. Boone (2021) uses a gradient boosted decision tree, but we find a random forest classifier performs best on our data (see Section 6.2.1).

Note that redshift is used to calculate this pseudoluminosity L, but it is not used as an explicit feature. Because the VAE model is redshift-invariant, it is well-equipped for photometric classification, particularly with strongly biased data sets. This VAE model is able to fit out-of-sample multiband photometric light curves of transients with low model uncertainties. Moreover, the VAE model generates a time-varying spectral prediction despite only being trained on photometry. Comparisons between generated spectra and observed spectra are discussed in Section 7.4. Meanwhile, the classifier component is able to use the decoded latent space embedding to photometrically classify the SN. The original ParSNIP model from Boone (2021) was trained on simulated (PLAsTiCC) and real (PS1) data, but for this work, we train a new ParSNIP model with additional modifications outlined below.

We train the ParSNIP VAE exclusively on YSE and ZTF simulated SNANA light curves in PS1-*griz*, ZTF-*gr* passbands across three SNe classes: SN Ia, SN II, SN Ibc (see Section 5.6 for details). We found that training on a SALT3 SN Ia model, NMF SN II model, and SED SN Ibc model with a mock host-$z \sim N(\mathrm{host}\text{-}z, 0.05^2)$ gave the best results when tested on real data. Thus, we used the entire simulated sample (see column (3) of Table 8) to train the VAE (and *not* to train the random forest classifier, for which we split into training, testing, and validation sets). We use the default hyperparameter configurations in Table 1 of Boone (2021) except we use a different learning rate of 10^{-5}. We trained our ParSNIP model for $\sim$24 hr across 30 IBM Power9 CPU cores and one NVIDIA Tesla V100 GPU on the Hardware-Accelerated Learning cluster at the University of Illinois at Urbana-Champaign (Kindratenko et al. 2020).

We perform inference using the ParSNIP model to estimate the latent representations for all simulated SNe. Additionally, we calculate a pseudoluminosity L from the model's measured amplitude using the cosmological parameters from Planck Collaboration et al. (2020) with the following formula:

$$L = -2.5 \log_{10}(A) - \mu_{\mathrm{Planck20}}(z) + 27.5 \qquad (7)$$

where A is the model amplitude of the light curve, μ is the distance moduli from the corresponding redshift using cosmological parameters from Planck Collaboration et al. (2020), and z is the mock host-z redshift. Note that the input fluxes for both the simulated training set and YSE DR1 are normalized to a zero-point of 27.5 on the AB system, which results in the last term.

Once the YSE ParSNIP VAE model was successfully trained on simulations, we applied the model to the full YSE DR1 sample of 1975 SN-like light curves to calculate the VAE features. From there, we separately train a random forest classifier on the VAE features extracted from the simulated set, using a separate 60% training, 20% test, and 20% validation split (see columns (4), (5), and (6) of Table 8). Once the classifier component was trained, we tested its performance on real data by comparing the predicted SN class labels to those of the spectroscopic sample, which fell into the general SN Ia, SN II, SN Ibc description (472 objects). We illustrate this process in Figure E1(*C* process). Our final results and more details on the VAE and random forest classifier are outlined in Section 6.2.

6. Results

The simplest overview metrics to understand the performance of a classifier are completeness, purity, accuracy, and F1 score. These metrics are defined for a single class as follows:

$$\mathrm{Completeness} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}};$$
$$\mathrm{Purity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}};$$
$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{S};$$
$$F_1 = 2 \times \frac{(\mathrm{Purity} \times \mathrm{Completeness})}{(\mathrm{Purity} + \mathrm{Completeness})}$$
$$= \frac{\mathrm{TP}}{\mathrm{TP} + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})} \qquad (8)$$

where TP (FP) is the number of true (false) positives, TN (FN) is the number of true (false) negatives, and S is the total sample size.

Completeness (*recall*) quantifies the percentage of a true spectroscopic type that is correctly classified. Purity (*precision*) quantifies the percentage of a predicted photometric type that is correctly assigned the true spectroscopic type. Accuracy is the overall fraction of events that was correctly classified. Lastly, the F_1 score is the harmonic mean of the precision and recall.

6.1. ParSNIP's Pretrained PS1 Classifier

As a benchmark test, we first use ParSNIP's pretrained photometric classifier using a data set from the PS1-MDS (Chambers et al. 2016), provided by Villar et al. (2020, 2020). This data set is composed of 2885 SN-like light curves with robust host–galaxy redshifts, of which 557 are spectroscopically classified. Similar to YSE light curves, the data use PS1-*griz* passbands. The main difference is that the exposure time of PS1-MDS is 113 s for PS1-*g*, PS1-*r* passbands and 240 s for PS1-*i*, PS1-*z* passbands whereas for YSE the exposure times are 27 s for all PS1-*g*, PS1-*r*, PS1-*i*, and PS1-*z* passbands. Thus, the PS1-MDS has a deeper 5σ depth of $\sim$23 mag and on average lower flux errors in comparison to YSE, which has a 5σ depth of $\sim$21 mag (unstacked; see Jones et al. 2021a for details). Moreover, PS1-MDS observed the majority of transients out to $z \sim 0.7$ (compared to YSE, $z \sim 0.3$).
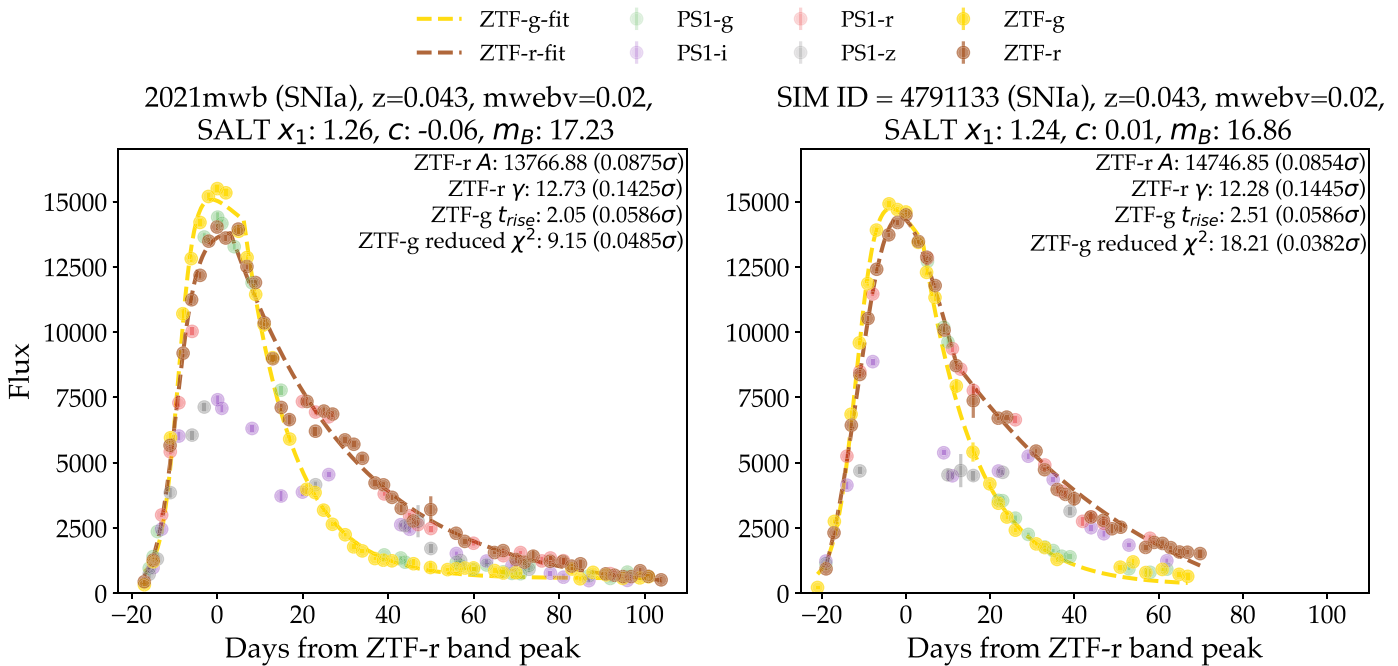
Figure 16. Comparison light curves with a ZTF-*g*, ZTF-*r* parametric model fit from Villar et al. (2019) between a spectroscopically confirmed SN Ia 2021mwb (left panel) and the closest matching SN Ia simulation (ID 4791133) from the simulated sample as determined by a nearest-neighbors search (right panel). Note that the SN Ia simulation has identical redshift and Milky Way extinction values ($z = 0.043$, MWEBV = 0.02). SALT3 fit values are shown for additional comparison. Observations are shown as circles with associated observational errors, and the dashed lines are the Villar Fit to the light curve's ZTF-*gr* passbands (see text for details). A few of the Villar Fit parameter values are shown, with their distance in standard deviations σ to the mean value of the combined YSE DR1, and simulated Villar Fit parameter distributions shown in parentheses. Overall, we can see from the light-curve evolution and model fit that the simulated light curves recreate the observed supernova very well.

Despite the differences, we first test the pretrained PS1-MDS ParSNIP classifier on our spectroscopic YSE DR1 data (excluding ZTF observations) as a benchmark performance. For the binary SN Ia and CC SNe classification, the pretrained PS1-MDS classifier achieves an overall accuracy of 84% with a 90% (86%) SN Ia completeness (purity) and 72% (79%) CC SNe completeness (purity). This performance is better than expected for such different data properties between PS1-MDS and YSE, but speaks to ParSNIP's ability to characterize transient properties disentangled from observing effects. As we will see in Section 6.2, a uniquely trained YSE ParSNIP model achieves not only better results with higher accuracy, completeness, and purity but also is more conservative to not overpredict SN Ia.

6.2. YSE DR1 Performance and Classification

In this subsection, we demonstrate the performance of our trained ParSNIP photometric classifier on the simulated validation set and YSE DR1 spectroscopic test set using only the extracted light-curve features and photo-z estimates. Then, we investigate predictions on the YSE DR1 photometric set.

6.2.1. Simulated Validation Set

The validation sample comprises 20% of our simulated sample, which we withhold from training. We optimize our classifier based on its performance on the validation set, as opposed to optimizing performance based on the spectroscopic test set to avoid biasing the classifier on properties of the spectroscopic set (brighter, more well sampled, etc.). Nevertheless, we strive for optimizing validation set performance while still achieving strong performance on the spectroscopic

sample, with realistic SN Ia ratio prediction on our photometric sample consistent with the literature.

The classifier component performs a custom classification routine based on a combination of ParSNIP's VAE feature parameters. We perform many iterations of varying classification algorithms while varying hyperparameters and the VAE feature set. Some of these tests included an XGB classifier, C-support vector classifier, an MLP, extra-trees classifier, and a random forest classifier. Ultimately, we achieved the best and most robust performance on the simulated validation set via the `sklearn` implementation of a random forest classifier[80] with 1000 trees, using all 11 features: the latent parameters and their errors s_1, $s_{1,\mathrm{err}}$; s_2, $s_{2,\mathrm{err}}$; s_3, $s_{3,\mathrm{err}}$; color and color error; reference time error; pseudoluminosity L and its error L_{err}. Our results can be reproduced by setting the random seed to 0. The performance of this classifier on the simulated YSE DR1 validation set is shown in Figure 17. Note that we rebalance the simulated set based on the ZTF BTS SNe class fraction during 2021 December 20–2019 November 24: ∼74% SN Ia, ∼19% SN II, ∼5% SN Ibc, ∼2% Other (see ZTF BTS results (orange) in Figure 22).

For tertiary classification (SN Ia, SN II, SN Ibc) on the validation set, we achieve an overall accuracy of 97%. The weakest performance is SN Ibc (73% complete, 89% pure), of which there is some confusion with SN Ia. For binary SN Ia versus non-SN Ia CC classification (SN CC), we also achieve an overall accuracy of 97%, and near-perfect SN Ia completeness and purity (99% complete, 98% pure).

---

[80] `sklearn.ensemble.RandomForestClassifier`; see https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.
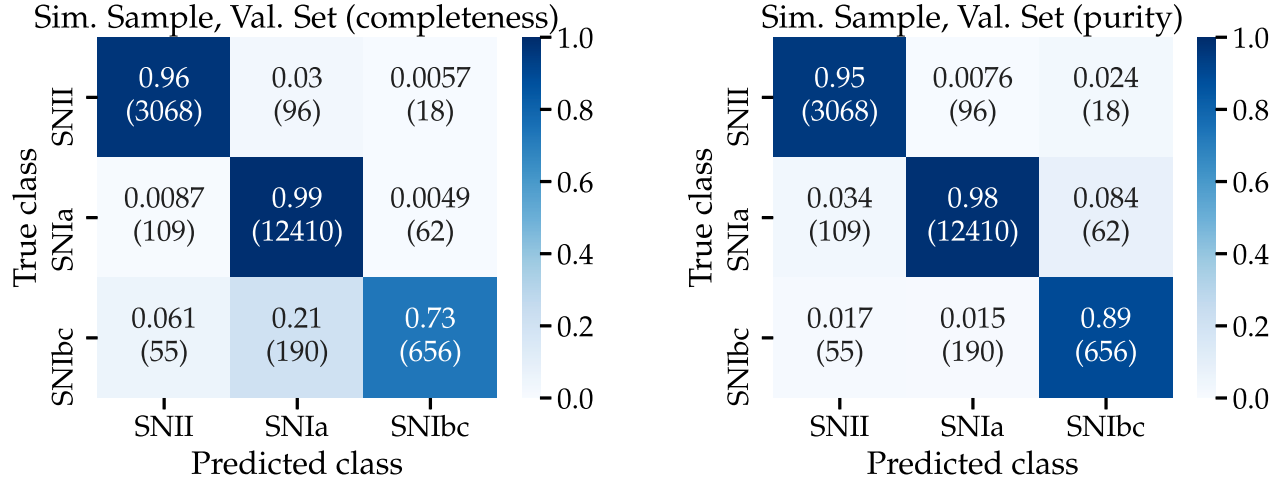
Figure 17. Completeness and purity confusion matrices of our simulated sample validation set for three-type (SN Ia, SN II, SN Ibc; left panel) classification performance. Completeness confusion matrices normalize each row to equal 1, and quantify the percentage of a true spectroscopic type that is correctly classified. Purity confusion matrices normalize each column to equal 1, and quantifies the percentage of a predicted photometric type that is correctly assigned the true spectroscopic type. The accuracy value for both the three-type classification and binary SN Ia vs. non-Ia classification (SN Ia, SN CC) is 97%.
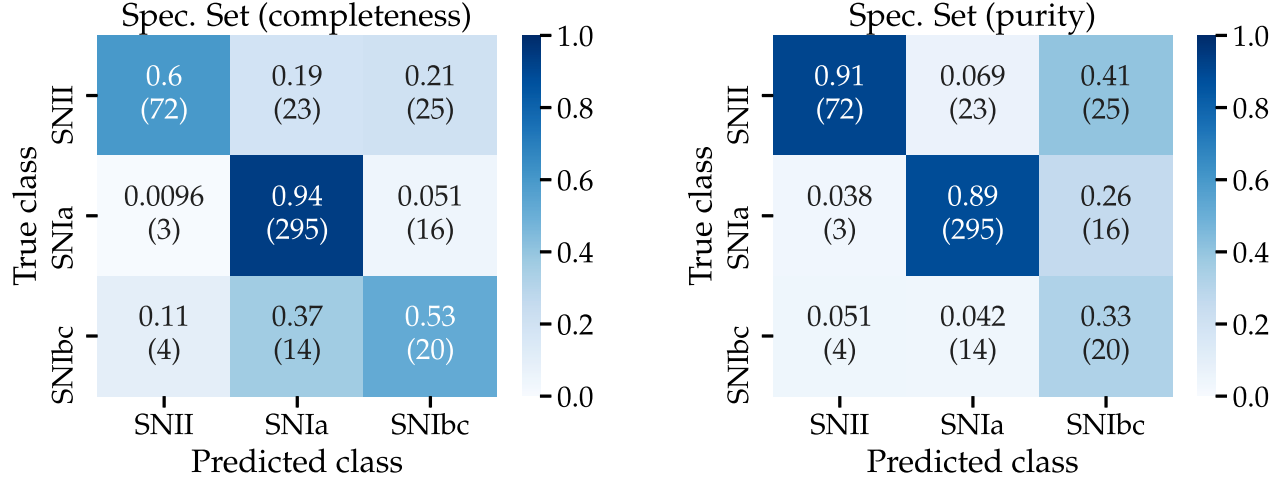
Figure 18. Confusion matrices showing completeness (left panel) and purity (right panel) for three-type (SN Ia, SN II, SN Ibc) classification of our spectroscopic test set (472 objects). We exclude the 20 Other objects, which do not fall into our classifier categories for validating our classifier performance, but we do classify them and discuss the results in Section 7.1. The SN type with the highest completeness and purity is SN Ia. There is confusion between the two core-collapse SNe types, but a very high individual purity of SN II.

6.2.2. Spectroscopic (Test) Set

Despite the strong performance of our classifier on simulated data, we investigate the effectiveness of classifying observed events. Overall, for the three-type classification (SN Ia, SN II, SN Ibc), ParSNIP performs well, achieving a classification accuracy of 82%: SN Ia (94% complete, 89% pure), SN II (61% complete, 91% pure), SN Ibc (53% complete, 33% pure). The "macro-average completeness" value, calculated as the mean of the diagonal terms in the confusion matrix (Villar et al. 2020), is 69% for completeness and 71% for purity. The *weighted* class-averaged (to account for class imbalance) F_1 score is 82%. These results are shown in Figure 18.

When split into the binary SN Ia versus non-SN Ia CC classification (SN CC), the performance is further improved in most categories, achieving an overall accuracy of 90%: 93% completeness and 92% purity for SN Ia, and 84% completeness and 86% purity for CC SN. The completeness and purity

confusion matrices of these binary classifications are found in Figure 19. The binary macro-average value is 89% for completeness and 89% for purity. The weighted class-averaged binary F_1 score is 90%. In effect, the performance on SNe CC is greatly improved at the expense of slightly decreasing performance on SN Ia.

Our full results with the object name, spectroscopic classification, and ParSNIP prediction with per class confidence percentages are compiled in Table C2. We examine in detail the efficacy of a few individual classifications (correct and incorrect), as well as compare performance results to another state-of-the-art classifier, SuperRAENN (Villar et al. 2020), in the Discussion (Section 7).

6.2.3. Photometric Sample

We classify the remaining 1483 photometric objects. For the three-type classification, we predict 1048 (∼71%) SNe Ia, 339
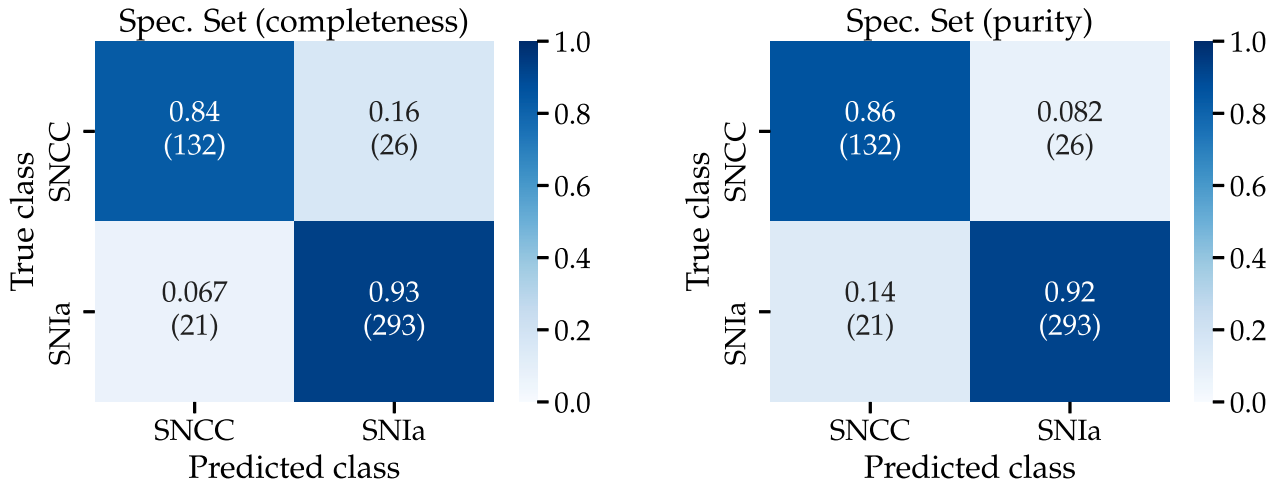
Figure 19. Same as Figure 18, but for binary SN Ia vs. non-Ia (SN Ia, SN CC) classification (472 objects). Again, the SN type with the highest completeness and purity is SN Ia.

($\sim$23%) SNe II, and 96 ($\sim$6%) SNe Ib/Ic. For binary classification, we predict 1004 ($\sim$68%) SN Ia, 479 ($\sim$32%) SN CC. Moreover, these observed class fractions are consistent with the magnitude-limited YSE survey results reported in Table 4, despite the photometric set of YSE DR1 not being magnitude-limited. This is likely due to the random forest classifier being trained on a rebalanced simulated feature set using ZTF BTS magnitude-limited rates (which themselves are in agreement with magnitude-limited YSE survey results). Moreover, they are holistically consistent with observed rates from ASAS-SN (Holoien et al. 2019): 69% SN Ia, 25% SN II, 6% SN Ibc (ASAS-SN discoveries and nondiscoveries totalling 964 objects; see their Figure 1), which lends indirect support and confidence in our photometric classifications. Although they are also consistent with the magnitude-limited survey ZTF-BTS, they are only marginally in agreement with the ASAS-SN magnitude-limited survey. Our results strongly disagree with the predicted LOSS rates (see Table 4).

We investigate how our predictions change as a function of cumulative light-curve observations in Figure 20. Here, we trace the predicted photometric and observed spectroscopic SN Ia fractions in addition to the spectroscopic sample accuracy. As expected, the accuracy of the spectroscopic sample improves with increasing light-curve observations, and remains constant past $N_{obs} \sim 20$ for tertiary and binary classification. Likewise, the predicted photometric and observed spectroscopic SN Ia fractions remain stable past $N_{obs} = 20$ for tertiary classification. For binary classification, this threshold is $N_{obs} = 10$. For light curves with fewer observations, the prediction is overwhelmingly SN Ia. This is likely due to several reasons: higher SN Ia rates, shorter duration/timescale of SN Ia events, and the intrinsic latent space distributions between SN classes for poorly sampled light curves are not well separated (biasing results toward one class prediction rather than several).

7. Discussion

In this section, we classify spectroscopic transients that are beyond our three class schema, and compare the results of our ParSNIP classifier to SuperRAENN (Villar et al. 2020). Then, we discuss the YSE DR1 classification breakdown, including an analysis of the ParSNIP latent space, the quality of the VAE model light-curve fit, and the recreated spectra. Lastly, we discuss future work enabled by this data release.

7.1. Classification of Nonstandard Transients

Our photometric classification schema assumes every transient belongs in one of three classes (SNe Ia, SNe II, SNe Ibc). This is sufficient for classifying the overwhelming majority of SNe in YSE DR1 ($\sim$96% of our spectroscopic sample falls into these three classes), but not appropriate for a few known extragalactic transients whose spectroscopic classification falls outside these labels. Regardless, we investigate how our classifier assigns labels to these events in Table C3.

There are five TDEs in YSE DR1 not discovered by YSE but with accompanying YSE data: AT 2020neh (Angus et al. 2022), AT 2020nov (W. Earl et al. 2022, in preparation), AT 2020opy, AT 2021ehb, and AT 2021qxv. One (AT 2020nov) is classified as an SN Ibc, with confidence $p \sim 0.4$. This classification makes the most sense among the few potential classes, because SNe Ibc can be long-lived and evolve more gradually in color than SNe Ia or SNe II. However, SNe Ibc do not reach the high intrinsic brightness of TDEs. AT 2021qxv has the plateau-like light curve that SNe IIP exhibit, and thus its SN II classification (with confidence $p \sim 0.8$) makes intuitive sense. AT 2021ehb after an initial peak also has a prolonged sustained light curve in the intrinsic brightness range of SNe II, so that classification (with confidence $p \sim 0.4$) is reasonable. The remaining AT 2020neh and AT 2020opy are both classified as SN Ia, with respective confidences $p \sim 0.5$ and $p \sim 0.7$. Both events have a peak absolute magnitude $M_{peak} \sim -19$ mag, which is consistent with the average intrinsic brightness of SNe Ia. Although the TDE sample size is small, it might prove fruitful to search for TDEs in our photometric sample for transients occurring in their galaxy cores, regardless of predicted classification.

Two of the four SLSNe are classified as SNe Ibc (SN 2021aadc, SN 2021nxq), and two are classified as SNe II (SN 2020xsy, SN 2021uwx). The light curves for SN 2021aadc and SN 2021nxq display a long rise until observations abruptly stop before their peak, atypical for SN Ibc. These classifications
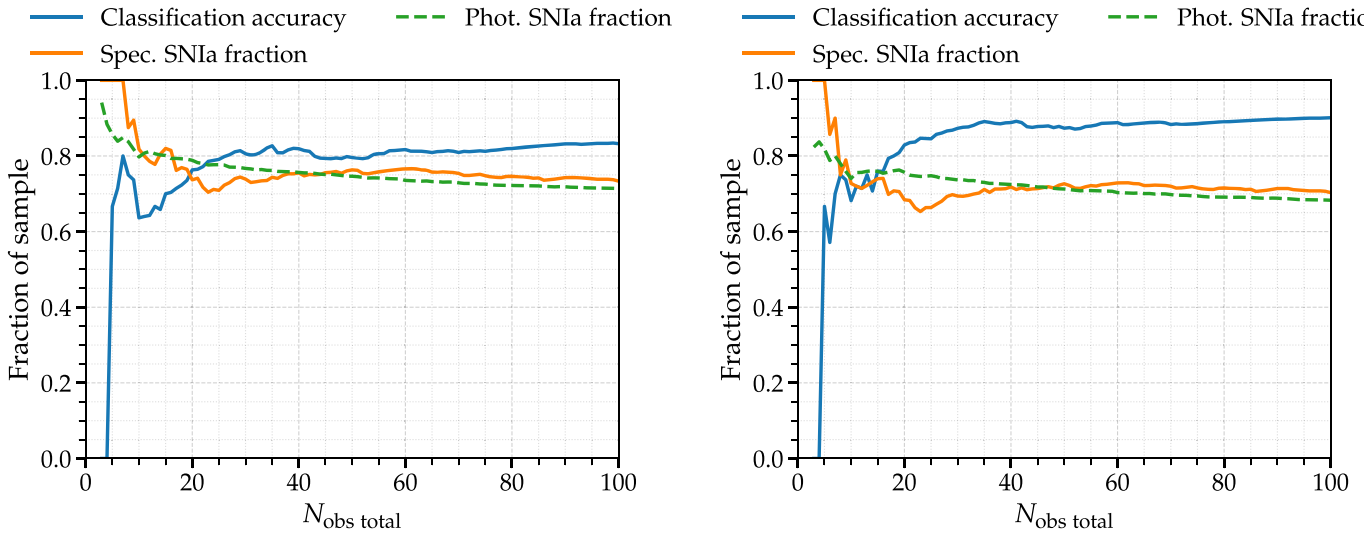
Figure 20. The predicted photometric SN Ia fraction (green), observed spectroscopic SN Ia fraction (orange), and photometric classifier accuracy of the spectroscopic sample (blue) as a cumulative function of the light curve N_{obs} across tertiary (left panel) and binary (right panel) SN classification. The photometric classification for poorly sampled light curves is preferentially SN Ia. Past $N_{obs} = 20$ ($N_{obs} = 10$) for tertiary classification (binary classification); the predicted photometric and observed spectroscopic SN Ia fractions remain stable.

are lower confidence ($p \sim 0.7$ and $p \sim 0.4$, respectively), and it is evident the classifier was not trained on simulated SLSNe light curves and ultimately the out-of-distribution nature of SN 2021aadc and SN 2021nxq resulted in a poor-quality classification. Meanwhile, the SN 2020xsy and SN 2021uwx classifications ($p \sim 0.5$ and $p \sim 0.5$, respectively) as SN II make intuitive sense: for SN 2020xsy, the post-peak profile exhibits a linear decline (in magnitudes), closely resembling SNe IIL. For SN 2021uwx, the pleateau-like light curve for $\sim$70 days resembles that of SNe IIP. Overall, none of the three available classes reach the intrinsic brightness of SLSNe events. Of these SLSNe, SN 2021aadc and SN 2021uwx are YSE discoveries, and the other two have corresponding YSE data.

The remaining Other classifications, including those for our SN, LBV, LRN, and the Other type, are found in Table C3.

7.2. SuperRAENN

Although we cannot determine the correctness of individual photometric classifications, we can see how often our trained ParSNIP classifier agrees with other state-of-the-art photometric classifiers applied to the same test data set. Here, we perform this comparison with the SuperRAENN photometric classifier from Villar et al. (2020).

In summary, SuperRAENN is a semisupervised classification approach, originally developed to photometrically classify the PS1-MDS sample. SuperRAENN is based on a recurrent autoencoder neural network (RAENN), followed by a random forest classifier applied to extracted features for SN classification. One strength is the ability to leverage information from both the labeled and unlabeled subsets.[81] Villar et al. (2020) used this to their advantage and trained SuperRAENN on both the spectroscopically labeled and unlabeled PS1-MDS subsets of the training set.

Although Villar et al. (2020) used real PS1-MDS light curves for training, we retrain SuperRAENN on the same simulated set as outlined in Table 8 in $\sim$24 hr on one CPU machine. We apply SuperRAENN to our labeled test set, and compare its performance to that of ParSNIP. Note that, besides the training set, we leave the overarching SuperRAENN architecture and pipeline unchanged: input light curves are preprocessed via Gaussian process interpolation into nine values per timestep (one value indicating the time relative to maximum; four magnitude values ($griz$)[82] and their associated uncertainties). From there, the encoder embeds them into an encoding vector, which is copied for each new emendation of the next time value. At each new time value, an output light curve is predicted, and compared to the original (the training process). Finally, the encoding (8 values) plus an additional 36 properties derived from Gaussian process-interpolated light curves are the features used as input into an unsupervised random forest classifier. The final SN classification is performed via 350 trees, the Gini information criterion, and leave-one-out cross-validation of the training set. For more details on SuperRAENN and its classification methodology, see Section 3 of Villar et al. (2020).

After retraining SuperRAENN, we apply it to the spectroscopic sample. In preprocessing, eight were dropped because either no rise or fall information was determined. Thus, we present the performance on 464 spectroscopic SN in Figure B1 (tertiary classification), and Figure B2 (binary classification). The overall performance, while not as accurate as ParSNIP (75% versus 82%), is still strong. SuperRAENN, like ParSNIP, has strong completeness and purity of SNe Ia ($\sim$90%), but suffers confusion among the CC SNe. In tertiary classification, SuperRAENN has lower SN Ia completeness, similar SN Ia purity, and lower SN II purity. In binary classification, the completeness and purity of SNe Ia and CC SNe are $\sim$5%–10% percent lower than that of ParSNIP.

With SuperRAENN and ParSNIP treated in the same manner, we compare their performance on the shared test set (464 objects). Under the assumption that these classifiers act

---

[81] It has been shown that SN classes may be clustered in duration, luminosity, and other physically motivated features (e.g., Kasliwal 2012; Villar et al. 2017), which can be used to extract defining characteristics without an explicit class label.

[82] ZTF-gr and PS1-gr filters are treated as the same g and r filters.
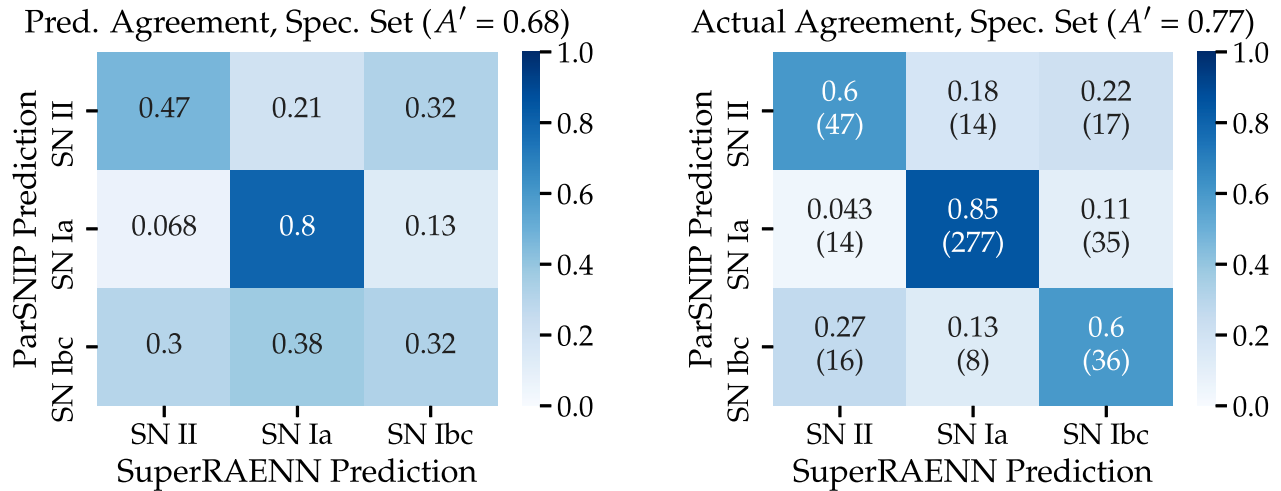
Figure 21. Left: the predicted agreement matrix of the ParSNIP and SuperRAENN classifiers on the spectroscopic sample. See text for details. Right: the actual agreement matrix of the ParSNIP and SuperRAENN classifiers on the spectroscopic sample. We agree on a slightly higher fraction of classifications than expected (77% vs. 68%), including a larger than expected fraction of SNe II (60%) and SNe Ibc (60%). The predicted and actual agreement on SNe Ia are within 5%. Note that this matrix does not tell us if these classifications are correct, but whether classifications from the two classifiers agree.

independently on a given transient, we cannot expect their agreement matrix to be of significantly stronger performance than the product of their test set confusion matrices (see Appendix B, Hosseinzadeh et al. 2020 for a full derivation). We calculate the predicted and actual agreement matrices of the ParSNIP and SuperRAENN classifiers on the shared spectroscopic test set (tertiary classification) in Figure 21. This is calculated via $A = P^T C'$, where A is the theoretical agreement matrix, P is the ParSNIP purity matrix, and C' is the SuperRAENN completeness matrix (see derivation in Hosseinzadeh et al. 2020, Appendix B). We expect relatively low agreement on SNe II and SNe Ibc, with which both classifiers exhibit confusion in the tertiary classification setting.

The actual agreement of SNe II and SNe Ibc is roughly 20%–30% higher than what we expect, whereas SNe Ia is within 5% of expectation. However, this agreement is simply if the two classifiers agree, rather than if the classifications are correct. In practice, of the 464 spectroscopically classified SNe, 320 (69%) objects are correctly classified by ParSNIP and Super-RAENN. Interestingly, this percentage is very close to predicted agreement (68%). ParSNIP edges out SuperRAENN in performance, as it uniquely classifies 60 (13%) SNe correctly, whereas SuperRAENN uniquely classifies 28 (6%) SNe correctly. Both classifiers are incorrect for 56 (12%) SNe. The dominant reason for SuperRAENN misclassification is the overprediction of SN Ibc. Further analysis of ParSNIP and SuperRAENN misclassified light curves is given in Section 7.3.

7.3. Full Sample

The full SNe class breakdown (except the Other class) percentage across all samples considered in this work is shown in Figure 22. The tertiary classification is in the left panel, and the binary classification is in the right panel. This analysis shows confidence that our photometric sample is correctly labeled, as ParSNIP's prediction of the spectroscopic (brown) and photometric samples (green) is in excellent agreement with that of the YSE DR1 spectroscopic (blue) sample, and to a lesser extent with that of the ZTF BTS (orange) sample.

ParSNIP and SuperRAENN (purple) tend to overpredict SN Ia and SN Ibc, yet underpredict SN II of the spectroscopic sample. Because ParSNIP's class prediction of the photometric sample nearly follows the observed percentages of the spectroscopic sample, the entire class prediction of YSE DR1 (pink) closely aligns with the spectroscopic sample rates. Moreover, because the spectroscopic sample is dominated by both magnitude- and volume-limited selection functions, the observed SNe class rates differ slightly from the purely magnitude-limited YSE survey results and magnitude-limited ZTF BTS survey (see Table 4). Our observed and predicted SN fractions agree well with that of the entire ASAS-SN SN sample (gray, Holoien et al. 2019), compiled of ASAS-SN discoveries and non-discoveries between 2017 January 1 and 2017 December 31 that have peak magnitudes of $m_{peak} \leqslant 17$.

To better understand how our classifier's biases (e.g., nearly $\sim 1/4$ spectroscopic SNe II are classified as SNe Ibc) impact the final photometric prediction and SN fractions, we follow the method of Villar et al. (2020) to "correct" the photometric sample breakdown (red). We take the dot product of the purity matrix (right panel, Figure 18) and our original photometric class breakdown (green). After applying this correction, a significant portion of predicted SNe Ibc is now predicted SNe II, while the predicted SNe Ia percentage experiences a slight decrease. When comparing to the ZTF BTS breakdown, our SN Ibc percentages match within $\sim 5\%$, but the correction increases the differences in SN Ia and SN II rates by an additional few percent. Likewise, we notice an intensified discrepancy with the SuperRAENN results, which favors SN Ibc over SN II in its prediction. We stress that this photometric correction should not be used to rigorously study the observational breakdown of SN classes, but rather to understand how biases encoded in the confusion matrices may inform and impact our final classifications. The fact that our original photometric prediction is in agreement with the entire ASAS-SN SN sample (Holoien et al. 2019), which like YSE DR1 is a compilation of the survey's discoveries and nondiscoveries, gives some credence to the correctness of our photometric labels.
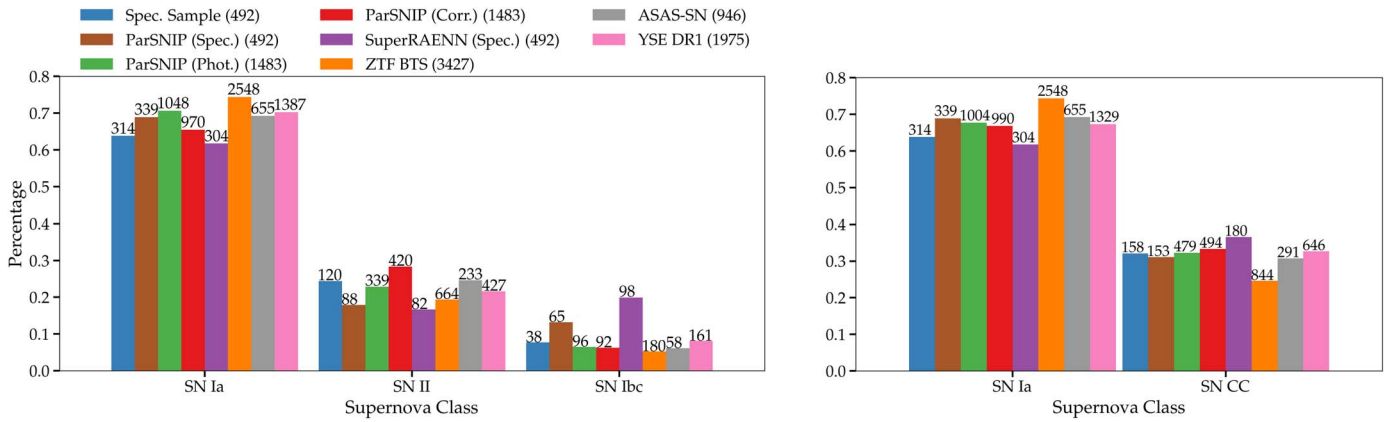
Figure 22. Breakdown of SN subclasses, displayed as percentage for a 3-type (SN Ia, SN II, SN Ibc) classification (left panel), and binary SN Ia vs. CC SNe classification (right panel). Note that we do not plot the few instances of the Other class, but do reflect the total counts in parenthesis. The true YSE DR1 spectroscopic breakdown is blue, our ParSNIP model's prediction of the spectroscopic sample is brown, our ParSNIP model's prediction of the photometric sample is green, our ParSNIP model's *corrected* prediction of the photometric sample is red (see details in text), our SuperRAENN model's prediction of the spectroscopic sample is purple, and the ZTF BTS sample (from 2019 November 24 to 2021 December 20, the same period as YSE DR1) is in orange. For additional comparison, the entire ASAS-SN SN sample (see Figure 1 of Holoien et al. 2019) is in gray, which has the most similar class breakdown to YSE DR1. We see that, when considering the spectroscopic sample only, ParSNIP slightly overpredicts the known fraction of SN Ia and SN Ibc and underpredicts the known fraction of SN II. Further, we have slightly underpredicted the true fraction of SN II in favor of SN Ibc, which is remedied via the binary classification. Both findings are supported by Figure 18. On the whole, ParSNIP's prediction of the spectroscopic sample (orange) follows the true spectroscopic distribution (blue) fairly well. Moreover, for binary classification, ParSNIP's prediction of the photometric sample (green) breakdown is similar to that of its prediction of the spectroscopic sample (brown).

However, there is more information than the label itself: the class prediction confidence score. To investigate the correctness of our labels as a function of confidence, Figure 23 shows cumulative fractions of the classification confidence for the spectroscopic test set. As expected, the majority of misclassifications stem from events with lower confidence and lower number of observations (particularly for SNe Ia). However, there are a few cases of high confidence or well-observed misclassified SN that are not simply due to SN II/SN Ibc confusion (see Figure B4 in Appendix B). Of seven total SNe Ibc misclassified as SNe Ia with high confidence ($> 80\%$), five are of a rare SN Ibc subtype: three SNe Ibn, one SN Ic-BL, and one SN Ib-peculiar. In addition, one object (SN 2020acct) exhibits a peculiar double-peaked light curve.

Visual inspection of misclassified light curves reveals some common patterns. The misclassified SNe Ia from ParSNIP tend to be light curves with the following:

1. significant ($\sim$100 days) gaps, or only observed well after peak (SN 2020uc, SN 2020zmi, SN 2021van, SN 2021vwx);
2. very red (SALT3 $c > 0.3$) SN Ia, which require a large correction for extinction (e.g., $m_{\text{peak}} - \mu \approx -17.5$ mag; SN 2020pki, SN 2020zfn, SN 2021aamo);
3. rare subtype (91T-like/91bg-like) properties (SN 2021bmu, SN 2021ctn);
4. long-lived CSM interaction (e.g., SN Ia–CSM; SN 2020aekp, SN 2020kre, SN 2021uiq);
5. or some combination of the above.

On the other hand, ParSNIP correctly classifies the brightest SNe Ia (in apparent magnitudes), whereas SuperRAENN sometimes misclassifies these events as SNe Ibc. This result is somewhat surprising, because the misclassified SNe II from both ParSNIP and SuperRAENN are often confused with SNe Ibc (particularly the SNe IIb subtype). The few misclassified SNe Ibc from both classifiers are often of a rare subtype (Ibn, Ic-BL), and preferentially assigned an incorrect SN Ia classification. Overall, the dominant source of misclassification for either algorithm is the presence of rare photometric or

spectral features (i.e., a rare subtype). This fact is unsurprising; because of the rare nature of these events, they are underrepresented in SN templates that were used to generate simulations for training ParSNIP and SuperRAENN. An effort to incorporate more rare SNe into SNANA templates (e.g., SN Ia-SC 2020esm, linear SN Ia 2021qvo) is ongoing.

Lastly, the performance of our ParSNIP classifier is relatively unaffected by the physical SN offset in classification tasks, as there is no significant difference between the distributions of objects correctly classified and those incorrectly classified. This speaks to the robustness of our algorithm as well as the oversampled PS1 PSF, which allows for improved image subtraction (and thus data reduction) near the galaxy center. In this manner, transients at the galaxy cores are less affected by data reduction errors, resulting in a less noisy light curve, which would otherwise likely disproportionately affect classification.

7.4. Exploring the ParSNIP Latent Space

To get a better insight into correctly classified and misclassified objects, we can look at the learned intrinsic latent space of the simulations and YSE DR1. A 3D visualization and two 2D slices of the phase-space spanned by ParSNIP's three intrinsic latent parameters (s_1, s_2, s_3) are shown in Figure 24.

Here we observe three distinct groupings from the simulated sample used to train the VAE (faint circles) and the observed spectroscopic SNe (right-side up triangles). The blue cluster is of SN Ia, which are bound in the tightest grouping, with only a few outliers far away from the core distribution. As is the case for all distinct SNe types considered in this work, observed SNe placed far away from their true classification grouping are more likely to be misclassified, best shown in Figure 25.

An example SN Ia mapped to the core of the SN II distribution at $(s_1, s_3) = (0.17, -0.89)$ is SN 2020aekp, an SN Ia-CSM with a long-lived light curve showing interaction. Another example is the SN 2020ybn, an SN IIn misclassified as an SN Ia at $(s_1, s_3) = (0.17, -0.89)$, far outside its latent SN II
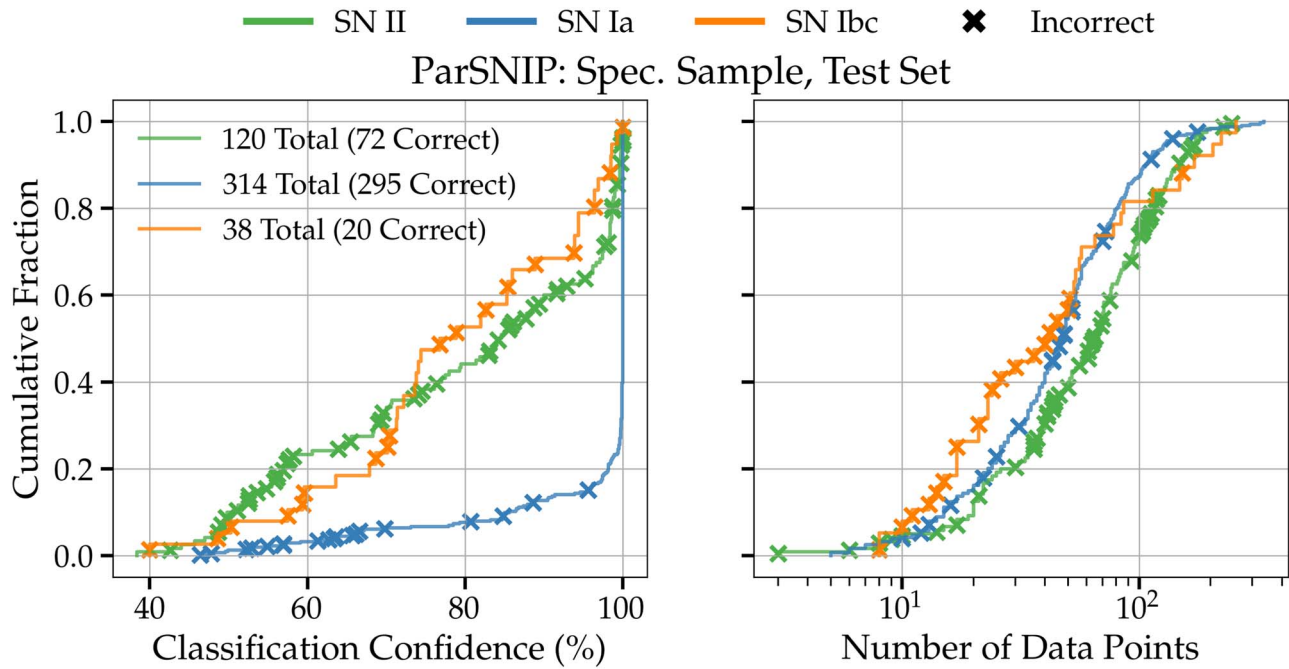
Figure 23. Cumulative fraction of our full spectroscopic test set (472 objects) as a function of classification confidence (left panel) and number of observations (right panel) grouped by spectroscopic class (SN II, green; SN Ia, blue; SN Ibc, orange). Misclassifications are marked with an "**X**." Most of the misclassifications for the best performing classification type, SN Ia, have relatively low classification confidence scores ($< 60\%$). However, a few highly confident predictions are incorrect, most notably in SN II, which we learn from Figure B4 is due to the common issue of SN II/SN Ibc confusion. Thus, for tertiary classification, the higher the classification confidence score and the greater number of observations, the more likely the classifier is correct for SNe Ia only. But for binary classification, this trend holds for both SNe Ia and CC SNe (Figure B4).

core distribution. The incorrect latent space characterization likely stems from the rare subtype light-curve profiles, which are not represented in the training set.

Between the SN Ia cluster and the SN II cluster in this 3D phase-space is the SN Ibc cluster (orange). Perhaps this arrangement alludes to the nature of such events: SN Ibc lack hydrogen in their spectra similar to SN Ia, but are of a different explosion mechanism (CC) like SN II. However, this speculation requires a more thorough investigation into the latent space and connection to fundamental physics. This will be explored in future work.

Empirically, the SN Ibc cluster is the most dispersed; it lacks a true center, and often partially overlaps into the adjacent SN Ia and SN II clusters. This contributes to misclassifications, particularly with SNe II where class blending is most common. Moreover, our phase-space representation may hint toward some deficiency in simulations of SN II diversity. In the region ($s_1 = [-2.5, -1.5]$, $s_2 = [-0.5, 2.5]$), for example, there are few SN II simulations yet an overwhelming majority of SN Ibc simulations. Because the ParSNIP VAE model is trained exclusively on simulations, ParSNIP predominantly predicts events in this region as SN Ibc, despite most spectroscopically confirmed instances being SN II. This issue could be solved with more diverse SN II simulations, or if ParSNIP was trained in part on spectroscopic SNe.

The benefit to using ParSNIP is that it encodes an intrinsic 3D latent parameter representation of each light curve independent of observing symmetries, from which a generative model can reconstruct the input light curve with high fidelity. If the training simulations are realistic and capture the intrinsic diversity of the particular SN class within YSE DR1, then the learned latent representation should be similar to the real YSE DR1 light curves' embeddings. This would likely result in accurate light-

curve interpolation and classification performance. We find this hypothesis to be true via Figure 26. There is good agreement in the distribution of latent parameters of the simulated sample (weighted to match the ZTF BTS SNe types sample fraction; see Section 6.2.1), spectroscopic sample, and photometric sample, particularly for s_1 and s_2 (less so for s_3). However, the s_1 simulated distribution is slightly skewed right ($s_1 \in [0, 0.1]$) to the mean value bin ($s_1 \in [-0.1, 0]$) of the photometric and spectroscopic samples. Overall, the general shape of a broad peak and long, extended tails holds for all three samples. For s_2, the simulated distribution excellently captures that of the spectroscopic and photometric sample. For s_3, the simulated distribution more closely follows the spectroscopic than the photometric sample, but all three samples exhibit a peak at $s_3 = 1.5$ with an extended tail and signs of a secondary peak around $s_3 = -0.9$. Although the exact physical interpretation of the s_3 latent parameter is uncertain, it could be a proxy for intrinsic brightness—the spectroscopic sample is on average brighter and more well sampled than the photometric sample due to the spectroscopic followup selection criteria (with a particular emphasis on SN Ia to provide a low-redshift cosmological anchor for LSST), and the simulated sample slightly favors brighter and lower-redshifted objects (see Figure 14), heavily weighted by the high observed ZTF BTS SN Ia fraction ($\sim$75%). Moreover, an s_3 value of 1.5 cuts through the heart of the SN Ia distribution (see Figures 24, 25), which is on average the most intrinsically bright of the simulated SNe types: normal Type II and Type Ib/c.

Overall the embedding distributions are highly similar between the spectroscopic and photometric samples, which we know from Figure 10 are vastly different and biased data sets (due to brighter, closer, and rarer objects preferentially targeted for spectroscopic follow-up). This is further evidence
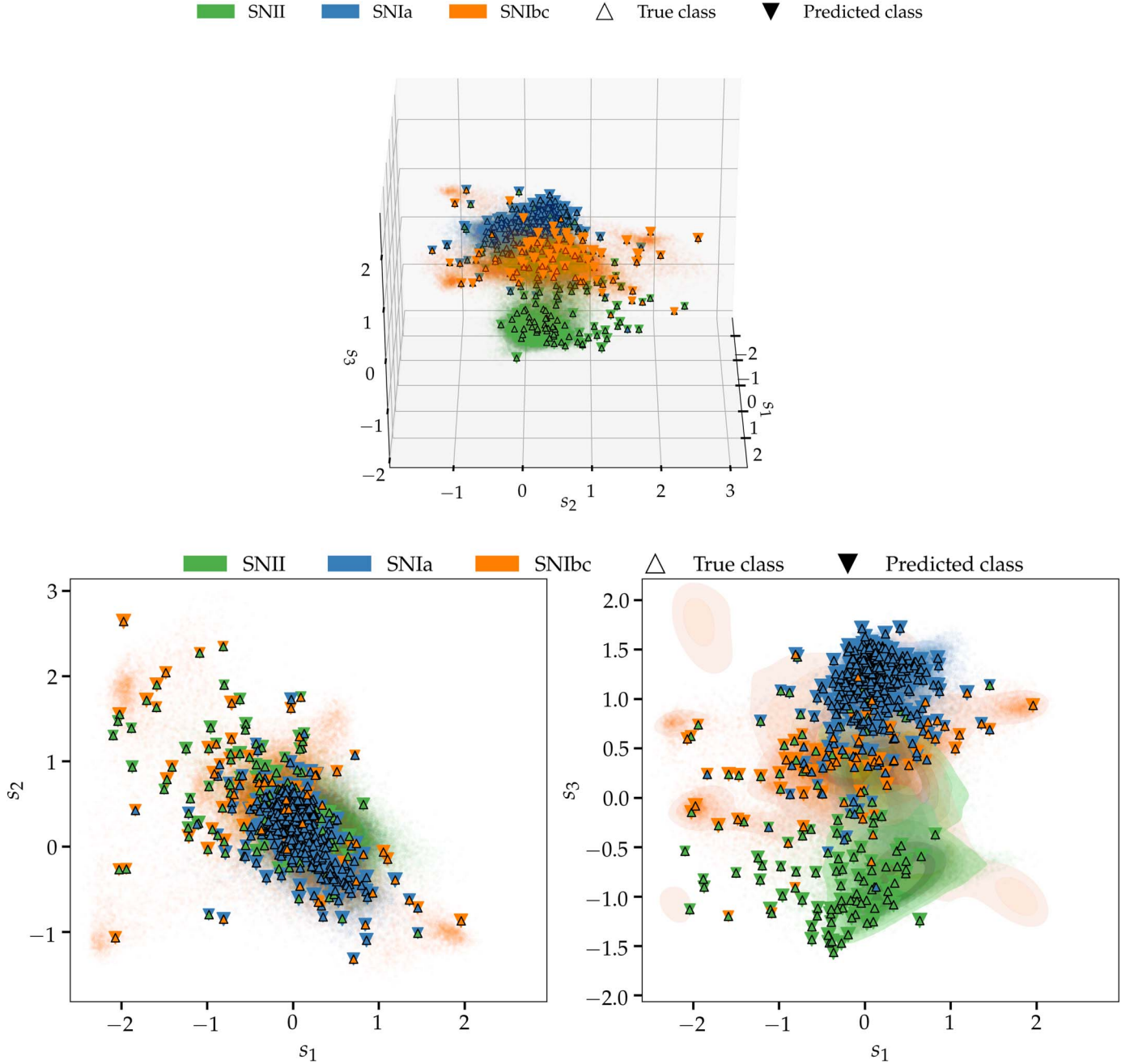
Figure 24. 3D and 2D visualizations of our ParSNIP model's learned intrinsic latent phase-spaces (s_1–s_2–s_3, upper panel; s_1–s_2, lower left panel; s_1–s_3, lower right panel). The model is trained on our simulated sample and tested on our spectroscopic sample. Right-side up triangles and a corresponding color (SN Ia, blue; SN Ibc, orange; SN II, green) denote the true SN class, and an upside-down triangle with the same colors denotes the predicted SN class. The latent representation of simulations used to train the model is shown in faint circles in five contour levels representing isoproportions of the density—10%, 20%, 40%, 60%, 80%. It is evident that the learned intrinsic representation from the simulations cluster in three distinct groups—indicative of the three considered SN classes—and most of the observed supernovae fall in or near these clusters.

of ParSNIP's claimed invariance to observational effects. From these analyses, the simulations used to train our model broadly capture the native diversity of our SN class population of YSE DR1 well enough for this work.

Using ParSNIP's generative model (decoder), we can predict the time-varying spectra and resultant light curve of the transient from its latent representation. We show a few examples of predicted light curves from the spectroscopic test set of YSE DR1 with 1σ uncertainty for each considered SN class in Figure 27. We find that the model generalizes well to real data, despite only being trained on simulations, for the

three transient classes considered. Most of the observations are fit by the mean model prediction with 1σ uncertainty. Note that these light curves are all examples where ParSNIP correctly predicts the SN class.

For a correctly classified SN Ia 2021hpr, we show our ParSNIP model's light-curve fit with 1σ uncertainty in the left panel of Figure 28, and the predicted time-varying spectra in the right panel at phases $\{-12.4, -5.4, +0.6, +21.6, +30.6, +53.6, +76.6\}$ days relative to the model predicted light-curve peak. Overall, throughout the ~90 days of its light-curve spectral evolution, ParSNIP is only able to recreate some broad
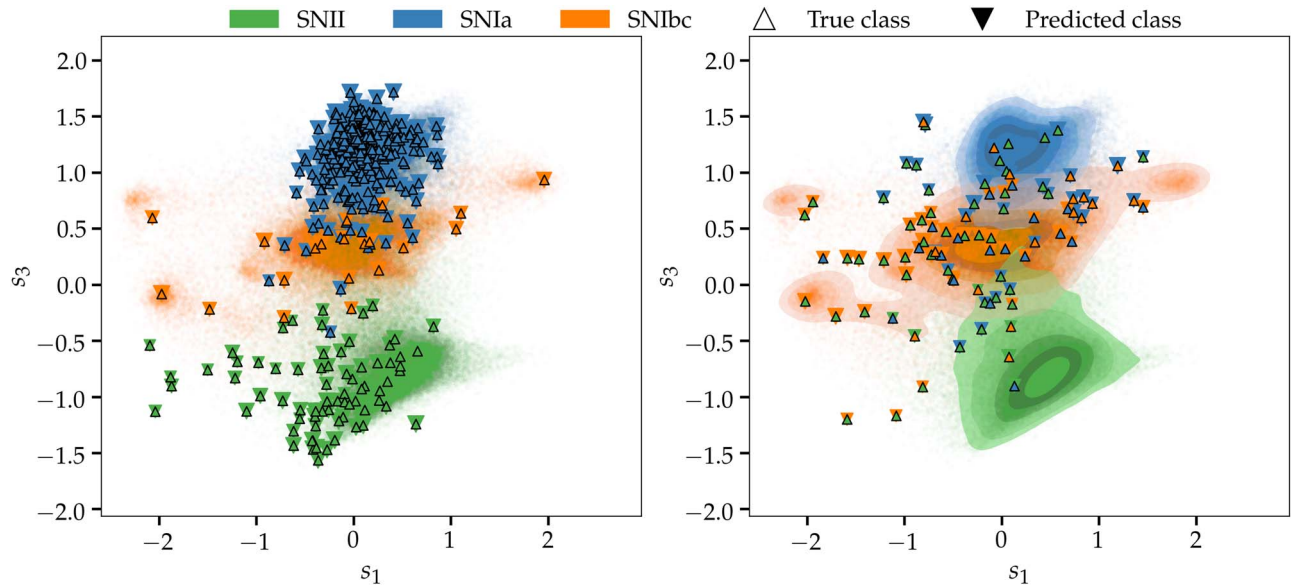
Figure 25. Same as the bottom right panel of Figure 24, except we demarcate the correctly classified members of the spectroscopic sample (left panel), and the misclassifed members (right panel) in the same latent space (s_1–s_3). Observed SN events that are placed far away from their natural grouping in the latent phase-space are most likely to be misclassified.
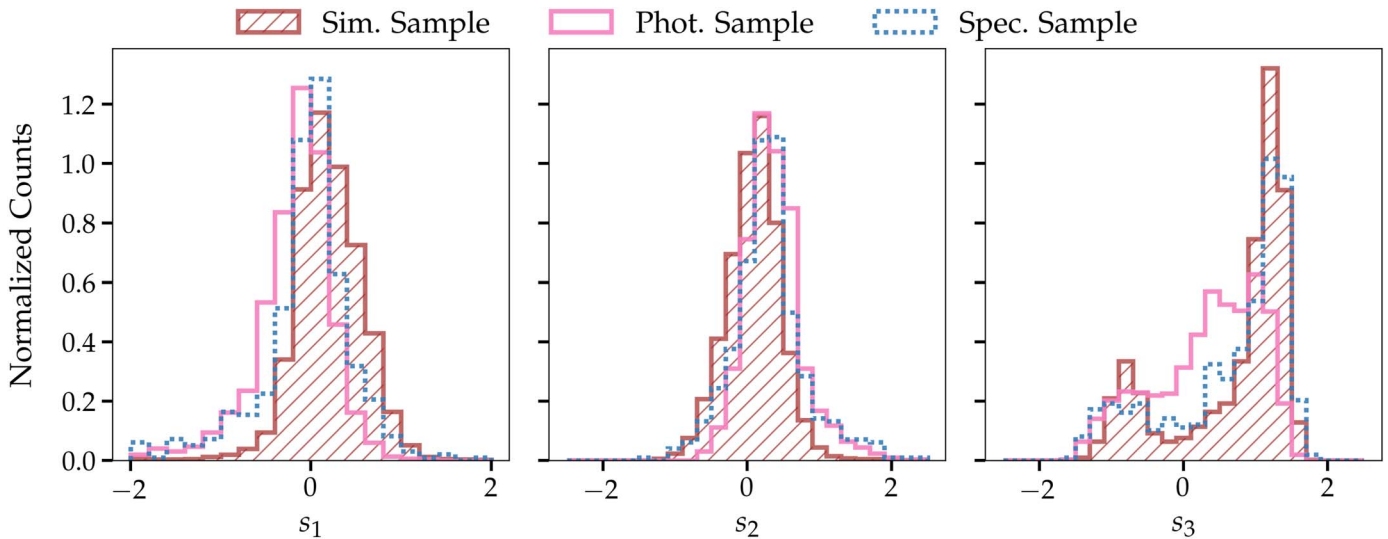
Figure 26. Normalized histograms of our trained ParSNIP model's intrinsic latent parameter distributions s_1, s_2, s_3 for the simulated training sample weighted to match the observed ZTF BTS SNe types sample fraction (brown; see Section 6.2.1), photometric sample (pink), and the spectroscopic test sample (blue).

features of the spectra. Even if we ignore any potentially poor wavelength-dependent flux calibration effects, and simply look at the spectral features, the agreement is marginal. The ParSNIP spectral model displays some nonphysical behavior in the UV wavelengths where PS1 passbands have limited coverage, rendering its interpretation unclear. Moreover, it misses the early O I (∼7500 Å) feature but does recreate the late-time S II (∼5500 Å) and Si II (∼6150 Å) absorption features, albeit appearing slightly redshifted. Despite the tight model fit to the light curve and highly confident correct SN Ia prediction, the recreated spectra are somewhat poor. It is unsurprising that narrow spectral features cannot be cleanly resolved from interpolated wide-band photometry. On the other hand, the ParSNIP model learned SN Ia spectra from deconvolving photometric observations varying redshifts, and no spectra were included in the simulated training data set. We remind the

reader that we do not use the recreated spectra in our photometric classification.

Conversely, we show an example misclassified SN Ia event (2021aamo) with its model fit and spectra prediction in Figure 29. In this case, both the predicted spectra and model fit to the real spectrum and photometry are poor, so it is unsurprising that the event was misclassified as an SN Ibc. Even with the photo-z estimate ($z = 0.062$) being nearly identical to the spec-z ($z = 0.059$), the model fit tends to underpredict the normalized flux. A possible reason for this is that SN 2021aamo is a very red SN Ia, with SALT3 $c > 0.3$, which is beyond the allowed SALT3 fit parameters (Kenworthy et al. 2021). Thus, at $m_{peak} - \mu \approx -17.5$ mag, after a significant correction for extinction, we find that the intrinsic brightness does agree with that of a normal SN Ia. Moreover, there is weak evidence of a secondary red peak. However, the
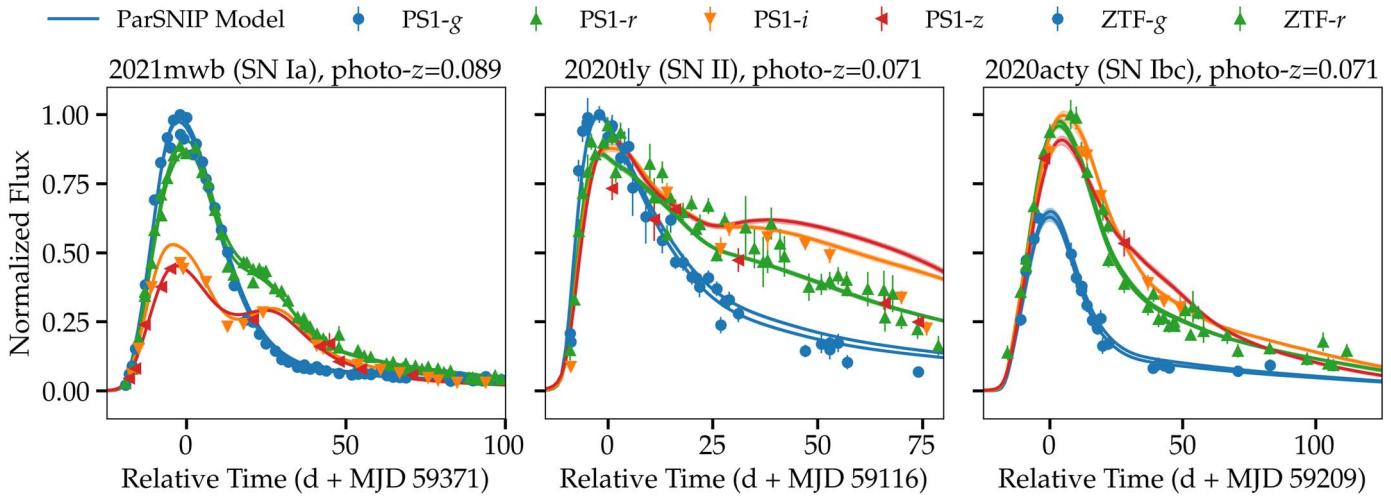
Figure 27. Example light curves of our trained ParSNIP model on the YSE DR1 spectroscopic test set, with each panel showing a different SNe classification type: SN Ia (SN 2021mwb, left panel), SN II (SN 2020tly, center panel), SN Ibc (SN 2020acty, right panel). All these objects are correctly classified. Observations are shown as individual points with their associated uncertainties, colors represent individual passbands, and the mean model predictions are shown as a solid line, with 1σ model fit uncertainties displayed as shaded contours (hard to see because such uncertainties are small). Note that these light curves have never been seen previously by our classifier, as we train exclusively on simulated light curves and test exclusively on real observed light curves.

true classification is undoubtedly SN Ia, as the observed spectrum clearly demonstrates strong Si II absorption at 6150 Å. Although the spectrum is noisy, there is also neither Mg II nor He I absorption, as the recreated ParSNIP model spectra predicts (indicative of an SN Ib).

To help quantify the quality of model fit, we calculate the model residuals for all simulated training set and YSE DR1 observations. For the simulated training set, we find that the model residual dispersions are ~0.05 mag (SN Ia), ~0.04 mag (SN II), ~0.07 mag (SN Ibc), which are approximately within a factor of the statistical uncertainties (~0.03 mag). This could suggest that we may have slightly overfitted the training set. For the full YSE DR1 sample, we see that the distribution of the residuals has a dispersion of ~0.103 mag when accounting for statistical uncertainties (~0.03 mag). This residual dispersion dominates the model error (~0.099 mag).[83] We note that our trained model has a lower residual fit to the simulated light-curve observations than that of the YSE DR1 observations: when considering all observations with a statistical uncertainty less than 0.05 mag, we find that ~83.0% (~95.0%) of YSE DR1 observations have residuals < 0.2 mag (< 0.5 mag); meanwhile, 97.7% (99.7%; SN Ia), 99.1% (99.9%; SN II), 93.2% (99.2%; SN Ibc) of simulated observations have residuals < 0.2 mag (< 0.5 mag). This means the simulations underestimate the number of observations with large errors when compared to YSE DR1, which could help explain why ParSNIP's performance on our simulated validation set (Figure 17) is stronger than on our observed spectroscopic test set (Figures 18, 19).

7.5. Future Work

The curation of YSE DR1 and development of a robust photometric classifier are the first steps to addressing key open questions and challenges in transient astrophysics through YSE. The next steps are to use these data sets and algorithms to enrich our understanding across multiple disciplines: developing new SED template models for SN simulations; observations

of young SNe to learn about their progenitors and environment; building a census of faint, fast, and red transients; anomaly detection and the study of rare or unique SNe (K. Malanchev et al. 2023, in preparation); curating a low-z anchor sample for SN Ia cosmology (G. Narayan et al. 2023, in preparation); and preparing for the imminent Vera C. Rubin Observatory. Many of these topics are in active development.

8. Conclusion

In anticipation of LSST, YSE is an integral piece of the smaller-scale time-domain multiwavelength survey landscape. YSE focuses on discovering fast-rising SNe within a few hours to days of explosion, and, as the only active four-band time-domain survey, provides a unique opportunity to study the earliest epochs of stellar explosions. YSE is a precursor to these next-generation time-domain surveys, as it will produce data sets in similar filters to those of LSST and Nancy Grace Roman Space Telescope. This work constitutes the first official release of YSE data.

Our conclusions and key takeaways are as follows:

1. We present the first data release of the Young Supernova Experiment (YSE DR1), spanning approximately the first 2 yr of the survey (2019 November 24–2021 December 20).

2. YSE DR1 comprises light curves and metadata for 1975 supernova-like sources, of which 492 transients are spectroscopically classified. Light-curve data include YSE observations from the Pan-STARRS1 telescope and complementing public ZTF observations (if available). Metadata include but are not limited to vetted PS1 host galaxy associations, photo-zs, host-zs, ParSNIP and SuperRAENN classifications (spectroscopic or photometric) across three classes (SN Ia, SN II, SN Ibc) and confidence scores, and ParSNIP latent embeddings.

3. We present preliminary relative SN rates from our magnitude- and volume-limited surveys, which are consistent with the literature within estimated uncertainties. Our magnitude-limited ($r < 18.5$ mag) relative SN rates are as follows: $\mathcal{R}(\mathrm{Ia}) = 0.682 \pm^{0.083}_{0.073}$, $\mathcal{R}(\mathrm{II}) = 0.239 \pm^{0.064}_{0.079}$,

---

[83] The model error is $e_m = \sqrt{e_r^2 - e_{\mathrm{stat}}^2}$ where e_m is the model error, e_r is the residual error, and e_{stat} is the statistical error.
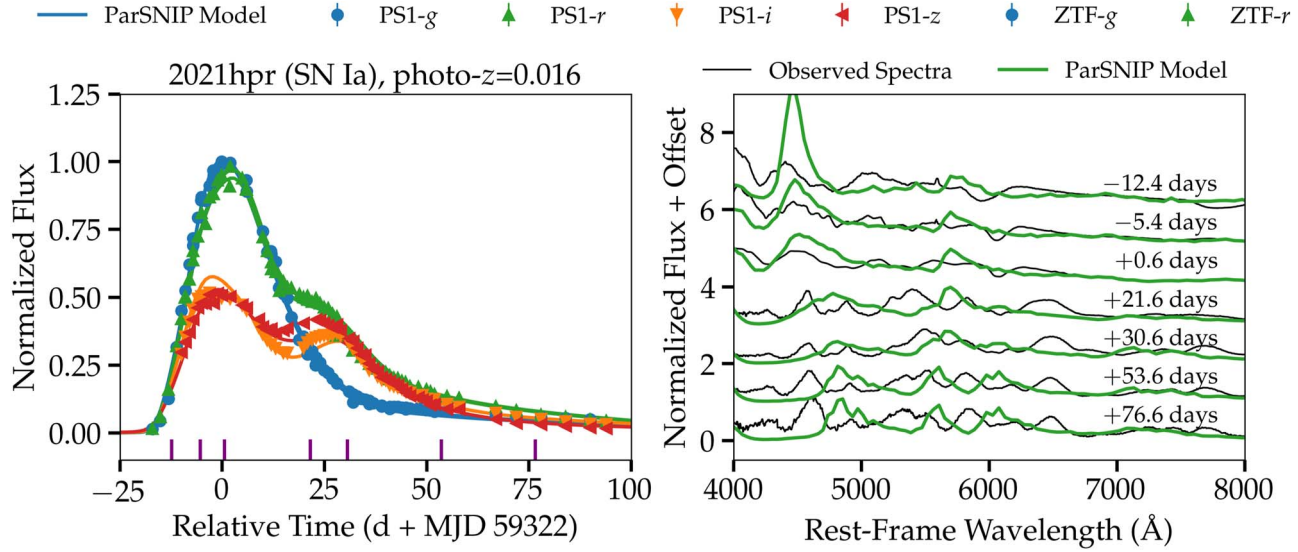
Figure 28. Comparison of the ParSNIP model to observed spectra of a correctly classified SN Ia, 2021hpr. The photometry and model fit of the light curve is shown in the left panel. In the right panel is the observed spectra from the Kast double spectrograph at Lick Observatory and Alhambra Faint Object Spectrograph and Camera (ALFOSC), overplotted with spectra from the ParSNIP model evaluated at the same phases (which is reflected in the left panel as solid purple line segments). We normalize the spectra to the flux at 6000 Å. It is evident that the broad structure of the ParSNIP predicted spectra compared to the observed spectra is marginal to poor (see text).
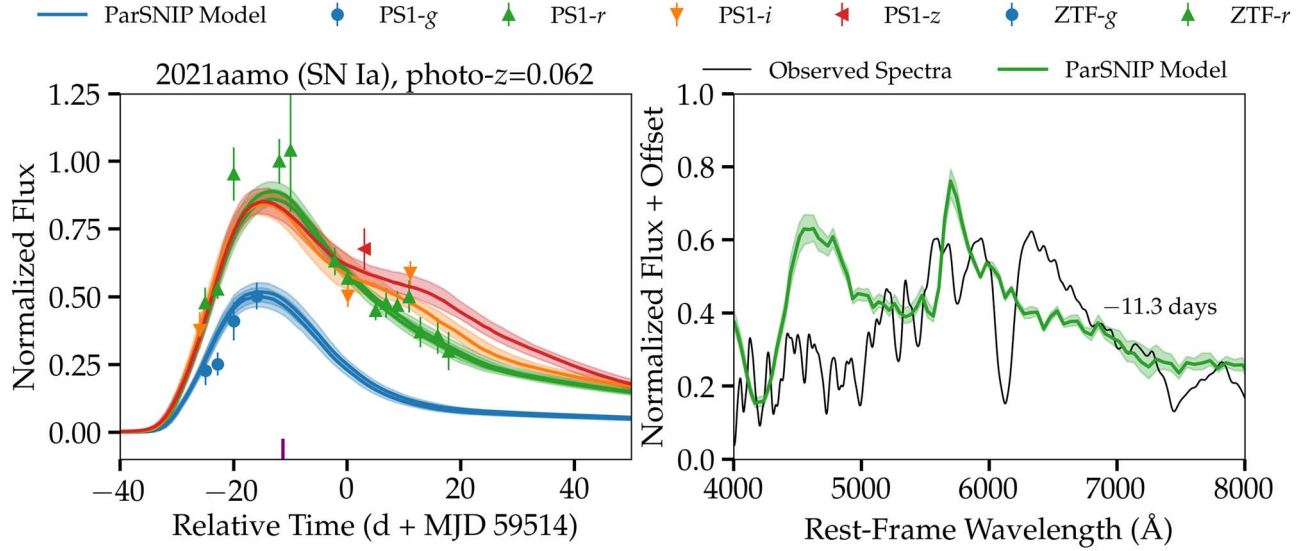
Figure 29. Comparison of the ParSNIP model to observed spectrum of a misclassified SN Ia, 2021aamo. The photometry and model fit of the light curve is shown in the left panel. In the right panel is the observed spectrum from the Kast double spectrograph, overplotted with the predicted ParSNIP spectrum evaluated at the same phase (which is reflected in the left panel as a solid purple line segment). We normalize the spectrum to the flux at 6000 Å. It is evident that there is little agreement between the predicted and observed spectrum. A likely reason for the misclassification is that SN 2021aamo is a very red SN Ia, with SALT3 $c > 0.3$, which is beyond the allowed SALT3 fit parameters (Kenworthy et al. 2021). Thus, at $m_{\mathrm{peak}} - \mu \approx -17.5$ mag and correcting for extinction, we find the intrinsic brightness does agree with that of a normal SN Ia. In the recreated spectrum, ParSNIP predicts the presence of Mg II and He I absorption (indicative of SN Ib), which is not observed. What is observed (and not predicted) is the strong Si II absorption at 6150 Å, the definition for an SN Ia event.

$\mathcal{R}(\mathrm{Ibc}) = 0.074 \pm^{0.033}_{0.057}$, and $\mathcal{R}(\mathrm{SLSN}) = 0.006 \pm^{0.005}_{0.005}$, across 181 objects. Our volume-limited ($D < 250$ Mpc) relative SN rates are as follows: $\mathcal{R}(\mathrm{Ia}) = 0.438 \pm^{0.072}_{0.075}$, $\mathcal{R}(\mathrm{II}) = 0.438 \pm^{0.072}_{0.075}$, and $\mathcal{R}(\mathrm{Ibc}) = 0.123 \pm^{0.041}_{0.057}$, across 207 objects.

4. We generate multisurvey (YSE, ZTF) SNe simulations with SNANA (Kessler et al. 2009) to train the ParSNIP (Boone 2021) classifier for photometric classification tasks. Simulations are now sufficient to exclusively train current photometric classification methods without heavily compromising performance on real data.

5. When validating our final ParSNIP photometric classifier on spectroscopically classified YSE SNe, we achieve 82% accuracy across three SN classes (SN Ia, SN II, SN Ibc) and 90% accuracy across two SN classes (SN Ia, CC SNe). We also report high individual completeness and purity of SN Ia ($> 90\%$), which will be critical for YSE SN Ia cosmology (G. Narayan et al. 2023, in preparation).

6. We use our ParSNIP photometric classifier to characterize the spectroscopically unclassified sample of 1483 YSE SNe: we predict 1048 ($\sim$71%) SNe Ia, 339 ($\sim$23%)

SNe II, and 96 (∼6%) SNe Ib/Ic for tertiary classification, and 1004 (∼68%) SN Ia, 479 (∼32%) CC SNe for binary classification.

7. ParSNIP has particular difficulty in characterizing transients exhibiting rare photometric or spectral features (i.e., of a rare subtype) still absent in simulation models. A common source of misclassification is the SN position in latent space (embedding) on the outer fringes of its true class core distribution.

8. In preparation for the forthcoming Rubin era, multicolor and multisurvey data sets such as YSE DR1 will be an important component of building discovery, anomaly detection, and classification algorithms, performing cosmological analyses, understanding the nature of red and rare transients, exploring TDEs and nuclear variability, and more.

Finally, we remark on the future of the YSE survey. Starting in January 2022, YSE commenced observations with Pan-STARRS2, which will be released in a future data release. It is likely that YSE will continue through 2024 for the anticipated start of LSST.

We will continue to pursue YSE's science goals, which include building a nearby universe census of transients, charting a new discovery space for faint, red, and rare transients, understanding black hole variability and TDEs, and assembling a legacy high-cadence, low-z anchor SN Ia cosmology sample. Using YSE DR1, we will have a sample of several hundred cosmologically useful spectroscopic and photometric SNe Ia with ∼3 mmag photometric calibration.

In 2020-2021, YSE has discovered or observed more than 2500 transients, and reported ∼5% of the total transient candidates reported to the International Astronomical Union. YSE light-curve data and spectroscopy results are yielding new insights into transient physics, the progenitor environment, classification tasks, and more. Ultimately, YSE DR1 and future YSE data releases will help to improve our collective knowledge of the time-domain universe, and prepare the community for Rubin's imminent LSST.

We welcome external collaborators; our external scientist policy, together with a guide to the application, can be found at https://yse.ucsc.edu/collaborate/.

Author contributions are listed below.

P.D.A. as project lead contributed to data preparation (YSE DR1); statistical and data analysis; relative rates; YSE DR1 simulation generation; classifier (ParSNIP) training and performance; host galaxy vetting and photo-z estimation; spectroscopic classification; lead writing and lead editing; figures; and P.D.A. also contributed as YSE collaboration meeting co-lead.

K.M. contributed to statistical analysis; classifier (ParSNIP) training; helpful discussions; writing; and figures.

S.N.S. contributed to data preparation (ZTF); Villar Fit analysis; writing; figures; helpful discussions.

D.O.J. as project scientist contributed to oversight; survey design; field selection; data reduction and preparation (YSE); scheduling observations; YSE collaboration meeting leadership; writing; editing; and helpful discussions.

G.N. as YSE Executive Committee member and project contact contributed to lead oversight; field selection; host galaxy vetting; editing; helpful discussions; figures; PS1 operations; and sniffing and/or sorting and/or flagging transients.

R.J.F. as YSE Executive Committee member contributed to survey design; oversight; spectroscopic reduction and classification; writing and editing; observations; and sniffing and/or sorting and/or flagging transients; and helpful discussions.

V.A.V. as YSE Executive Committee member contributed to classifier (SuperRAENN) training; editing; and helpful discussions.

C.R.A. contributed to follow-up observations (ALFOSC/ Nordic Optical Telescope); spectroscopic reduction and classification; and helpful discussions.

V.F.B. as YSE Executive Committee member contributed to helpful discussions.

M.J.B.-R. as a junior review panelist contributed to observations.

D.C. contributed to YSE DR1 simulation generation; as YSE collaboration meeting co-lead contributed to scheduling observations; and as contributor to `YSE-PZ` contributed to helpful discussions.

C.C. contributed to spectroscopic classification (SNe IIn); and sniffing and/or sorting and/or flagging transients.

D.A.C. contributed to `YSE-PZ`.

K.W.D. contributed to data reduction; spectroscopic reduction and classification; and observations.

S.D. contributed to writing; host galaxy vetting; sniffing and/or sorting and/or flagging transients.

M.R.D. as YSE Executive Committee member contributed to helpful discussions.

A.E. contributed to `Easy PhotoZ` software; helpful discussions; and figures.

K.D.F as YSE Executive Committee member contributed to draft review.

A.G. contributed to `GHOST` and `Easy PhotoZ` software; host galaxy vetting; draft review; and helpful discussions.

C.G. contributed to draft review; and sniffing and/or sorting and/or flagging transients.

J.H. as YSE Executive Committee member contributed to oversight; draft review; spectroscopic classification; and helpful discussions.

M.E.H. contributed to PS1 operations.

W.V.J.-G. contributed to data reduction; and helpful discussions.

C.D.K. contributed to draft review; and observations.

D.L. as a junior review panelist contributed to data reduction; observations; and sniffing and/or sorting and/or flagging transients.

P.M. contributed to observations.

K.S.M. as YSE Executive Committee member contributed to helpful discussions.

R.M. as YSE Executive Committee member contributed to helpful discussions.

F.M. contributed to statistical analysis; helpful discussions; and figures.

P.M. contributed to draft review; and `YSE-PZ` development and maintenance.

J.D.R.P. contibured as a junior review panelist.

E.R.-R. contributed as a YSE Executive Committee member.

C.L.R. contributed as a junior review panelist.

C.R.-B. contributed as a junior review panelist sniffing and/ or sorting and/or flagging transients.

Aleo et al.

Facilities: PS1, ADS, TNS, NED (NASA/IPAC Extragalactic Database (NED) 2019).

Software: GHOST (Gagliano et al. 2021), Astropy (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018), Easy PhotoZ (this work), Matplotlib (Hunter 2007), numpy (van der Walt et al. 2011), Pandas (pandas development team 2020), ParSNIP (Boone 2021), Photpipe (Rest et al. 2005, 2014), Scikit-Learn (Pedregosa et al. 2011), SNANA (Kessler et al. 2009), SNID (Blondin & Tonry 2007), sncosmo (Barbary et al. 2022), SuperRAENN (Villar et al. 2020), YSE-PZ (Coulter et al. 2022).

Appendix A
Easy PhotoZ

We employ a fully connected, five-layer MLP architecture with hidden layer widths [256,1024,1024,1024]. The model terminates in an output layer with 334 neurons with a softmax activation function, which we interpret as an approximation to the discrete posterior density function: $P(z|\text{Data})$. It can be shown under ideal conditions that such a network trained with categorical cross-entropy loss approximates the probability of redshift given the data (Richard & Lippmann 1991; Rojas 1996). We use the leaky rectified linear unit nonlinearity between each hidden layer. Moreover, we use the ADAM optimizer with a base learning rate 10^{-3}, and a learning rate decay of factor 0.5 for every three epochs of stagnant validation loss with a 5% dropout. The total number of model parameters are 2,712,910, and are used to minimize the loss function L

Test Point Performance
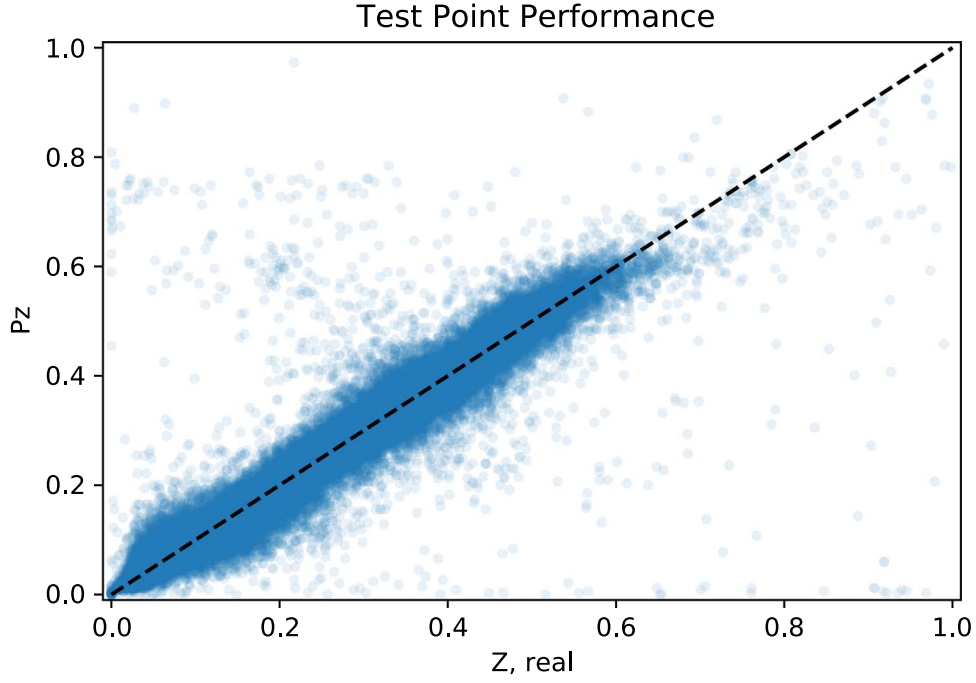
Figure A1. The point estimates of the posterior of `Easy PhotoZ` vs. the spectroscopic host galaxy redshifts (blue points). Despite a few catastrophic estimates, the overwhelming majority of point estimates agrees with the true redshift in a 1–1 relation (dashed black line) with a reasonable uncertainty of ~ 0.03.

with $\gamma = 0.15$:

$$L(\Delta z) = 1 - \frac{1}{1 + \left(\frac{\Delta z}{\gamma}\right)^2}. \tag{A1}$$

To compute `Easy PhotoZ`'s point estimates from the probability distribution, we use the risk minimization technique described in Tanaka et al. (2018) where the risk r is shown as

$$r(z_{\mathrm{phot}}) = \int dz P(z) L\left(\frac{z_{\mathrm{phot}} - z}{1 + z}\right). \tag{A2}$$

For a vector posterior $P(z)$ and vector z of photometric redshift bin centers, we calculate the risk at each z_{phot} and choose the lowest risk as our initial point estimate (before applying any additional magnitude-informed probabilities to the posterior for an updated point estimate for this work).

We plot the initial point estimate of the posterior against the true spectroscopic host galaxy redshifts in Figure A1.

Appendix B
Figures

Additional figures for this work are presented in Figures B1–B5.
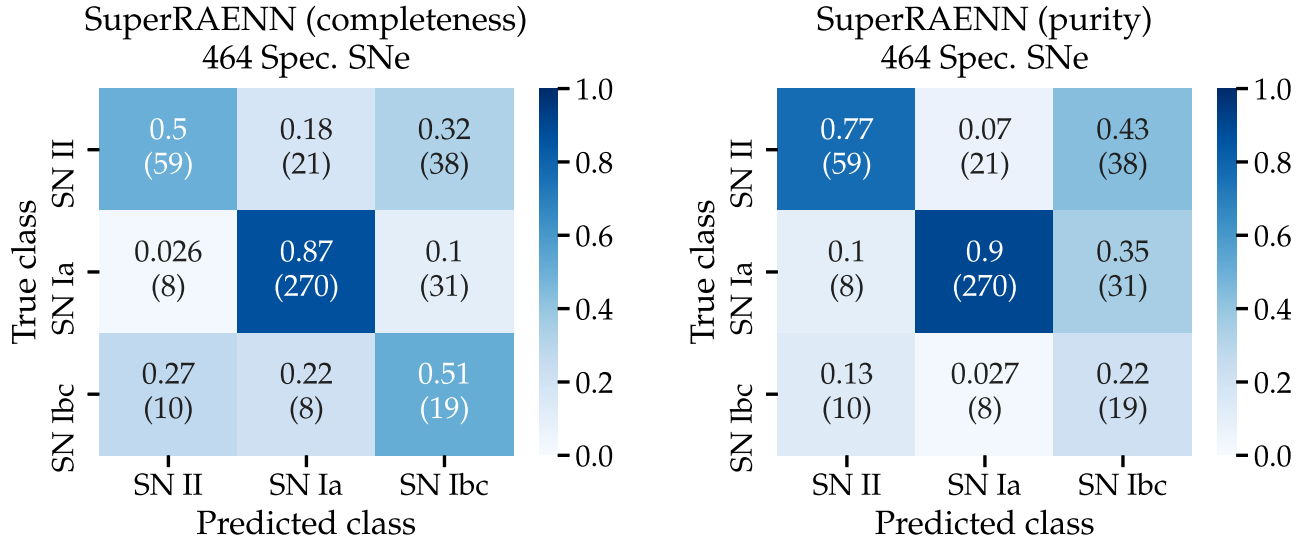
Figure B1. SuperRAENN confusion matrices showing completeness (left panel) and purity (right panel) for 3-type (SN Ia, SN II, SN Ibc) classification of 464 objects out of 472 objects in our spectroscopic test set (8 were removed due to lack of rise and/or decline information due to SuperRAENN's preprocessing pipeline). We exclude the 20 Other objects, which do not fall into our classifier categories for validating the SuperRAENN classifier performance. Like ParSNIP, the SN type with the highest completeness and purity is SN Ia. There is confusion between the two core-collapse SNe types, but a moderate individual purity of SN II.
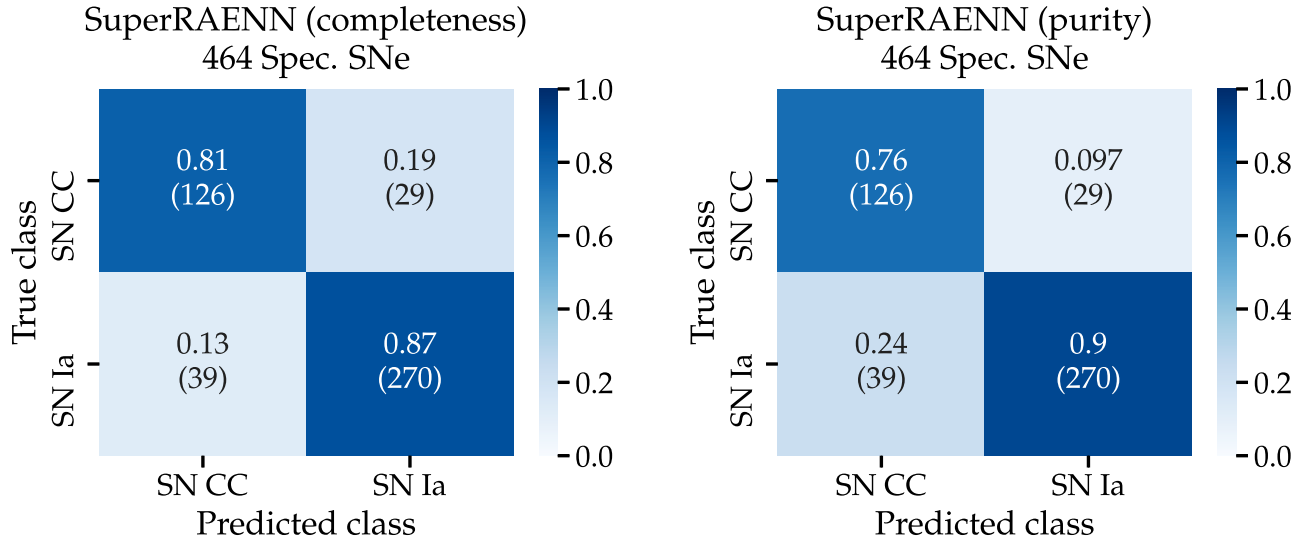
Figure B2. Same as Figure B1, but for binary SN Ia vs. non-Ia (SN Ia, SN CC) classification (464 objects). Again, the SN type with the highest completeness and purity is SN Ia. The completeness and purity of SNe Ia and CC SNe are ∼5%–10% lower than that of ParSNIP.
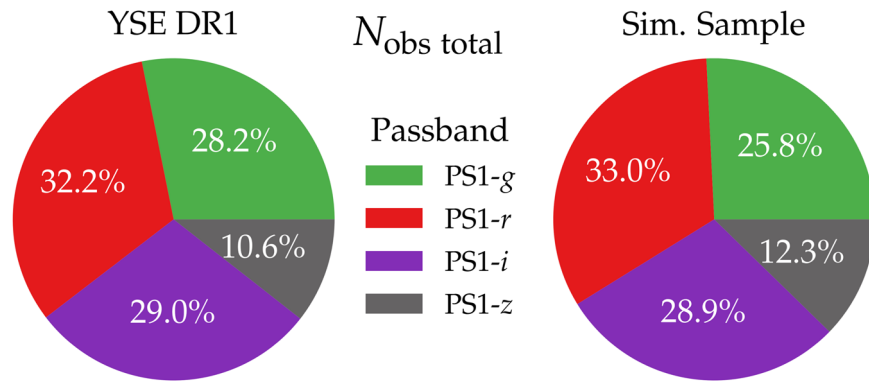
Figure B3. Pie charts of the total number of observations $N_{obs\ tot}$ per YSE passband only (PS1-g, green; PS1-r, red; PS1-i, purple; PS1-z, gray) displayed as a percentage for YSE DR1 (left) and a random subset of the entire simulated sample (SN Ia, SN II, SN Ibc; right) such that both pies have 1975 objects. We apply the same cuts on the simulated sample as we do on YSE DR1 (e.g., S/N > 4), only use observations, and do not include nondetections.
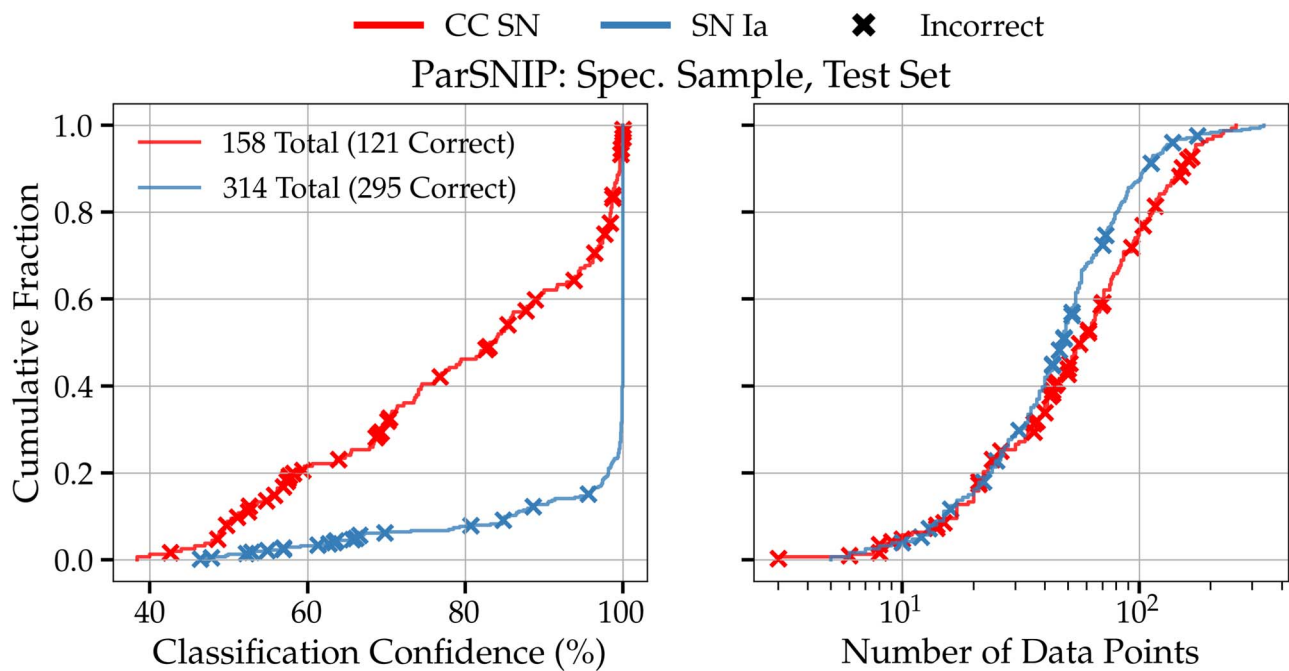
Figure B4. Same as Figure 23, but for binary SN Ia vs. non-Ia (SN Ia, SN CC) classification of our full spectroscopic test set (472 objects). In general, for binary classification, the higher the classification confidence score and the greater number of observations, the more likely the classifier is correct.
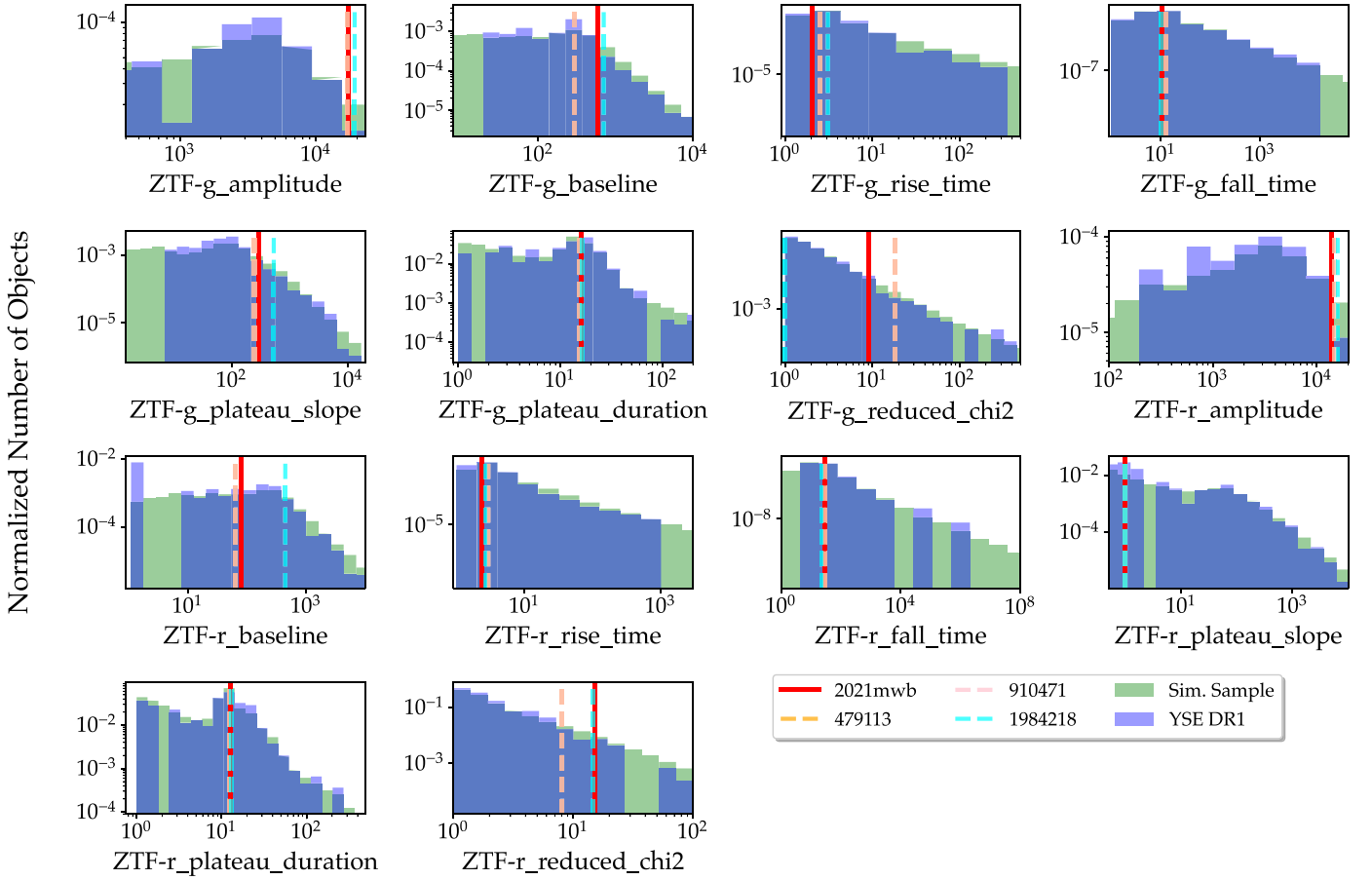
Figure B5. Histograms of SNe Ia Villar Fit parameter distributions for YSE DR1 (green) and the simulated sample (blue) in ZTF-*g* and ZTF-*r* passbands. Overlaid lines represent SN 2021mwb (red, thick line) compared to the three closest matching SN Ia simulations (dashed lines) to SN 2021mwb as determined by our nearest-neighbors search. Note the log scale of the *x*-axis. In all considered Villar Fit parameter distributions, the simulated sample well encompasses the profile and ranges of YSE DR1 as a whole, and is able to match parameter values to a known SN Ia (SN 2020mwb) on an individual basis.

Appendix C
Tables

Additional tables for this work are presented in Tables C1–C3.

Table C1
Selected Highlights from the Spectroscopic YSE DR1 Sample

Category (Counts)	IAU Name	Spec. Class	References	Remarks
Targeted YSE Objects (31)	2019yvr	SN Ib	1	20 SNe Ia, 1 Ia-91T-like, 1 Ib, 1 Ibn,
	2020duv	SN Ia-91T–like	2	1 Ic, 3 II, 2 IIb, 1 LBV, 1 LRN.
	2020dwg	SN Ia	3	
	2020dyc	SN Ia	4	
	2020eci	SN Ia	5	
	2020ftl	SN Ia	6	
	2020ikq	SN IIb	7	
	2020jgl	SN Ia	8	
	2020lbf	SN Ia	9	
	2020lfi	SN II	10	
	2020nlb	SN Ia	11	
	2020nxt	SN Ibn	12	
	2020oi	SN Ic	13	
	2020ppe	SN Ia	14	
	2020pst	SN Ia	15	
	2020rmg	SN Ia	16	
	2020sjo	SN Ia	17	
	2020svn	SN II	18	
	2020tfb	SN II	19	
	2020tjd	SN IIb	20	
	2020ue	SN Ia	21	
	2020uxz	SN Ia	22	
	2020zj	SN Ia	23	
	2021J	SN Ia	24	
	2021biy	LRN	25	
	2021bug	LBV	This work	
	2021dnm	SN Ia	26	
	2021hpr	SN Ia	27	
	2021low	SN Ia	28	
	2021mim	SN Ia	29	
	2021pfs	SN Ia	30	
SNe IIn (13)	2019uit	SN IIn	31	See C. Cold et al. (2023, in preparation).
	2020bwr	SN IIn	This work	
	2020jhs	SN IIn	32	
	2020noz	SN IIn	33	
	2020qmj	SN IIn	34	
	2020rdu	SN IIn	This work	
	2020tan	SN IIn	35	
	2020uaq	SN IIn	36	
	2020utm	SN IIn	37	
	2020ybn	SN IIn	38	
	2021aapa	SN IIn	39	
	2021bmv	SN IIn	40	
	2021xre	SN IIn	41	
Flash ionization features (8)	2020pni	SN II	42	See W. Jacobson-Galán et al. (2023, in preparation).
	2020tlf	SN II	43	
	2020abjq	SN II	44	
	2020svn	SN II	45	
	2020xua	SN II	46	
	2021aaqn	SN II	47	
	2021dbg	SN II	48	
	2021qvr	SN II	49	
SNe II, CSM interaction (6)	2020abjq	SN II	50	Does not include YSE SNe IIn.

Table C1
(Continued)

Category (Counts)	IAU Name	Spec. Class	References	Remarks
	2020svn	SN II	51	See W. Jacobson-Galán et al. (2023, in preparation).
	2020xua	SN II	52	
	2021aaqn	SN II	53	
	2021dbg	SN II	54	
	2021qvr	SN II	55	
TDE class (5)	2020neh	TDE	56	
	2020nov	TDE	57	
	2020opy	TDE	58	
	2021ehb	TDE	59	
	2021qxv	TDE	60	
SNe Ic-BL class (5)	2020fhj	SN Ic-BL	61	
	2021dib	SN Ic-BL	This work	
	2021jw	SN Ic-BL	62	
	2021qzp	SN Ic-BL	This work	
	2021too	SN Ic-BL	63	
SNe Ia-CSM class (3)	2020aekp	SN Ia-CSM	64	
	2020kre	SN Ia-CSM	65	
	2021uiq	SN Ia-CSM	This work	
Ca-strong transients (2)	2021gno	SN Ib-pec	66	
	2021inl	SN Ib-pec	67	
SNe Ia-SC class (2)	2020esm	SN Ia-SC	68	
	2021aagz	SN Ia-SC	This work	

References. (1) Kilpatrick et al. 2021b; (2) Perley et al. (2020a); (3) Perley et al. (2020b); (4) Pineda et al. (2020); (5) Dahiwale & Fremling (2020a); (6) Balcon (2020b); (7) Ho et al. (2021); (8) Galbany et al. (2020); (9) Dahiwale & Fremling (2020b); (10) Siebert et al. (2020a); (11) Sand (2019); (12) Srivastav et al. (2020); (13) Gagliano et al. (2022b); (14) Dahiwale & Fremling (2020c); (15) Jha et al. (2020); (16) Tinyanont et al. (2020); (17) Dahiwale & Fremling (2020d); (18) Dahiwale & Fremling (2020e); (19) Weil & Milisavljevic (2020); (20) Angus (2020); (21) Tinyanont et al. (2021); (22) Burke et al. (2020b); (23) Dahiwale & Fremling (2020f); (24) Gallego-Cano et al. (2022) ; (25) Cai et al. (2022); (26) Dahiwale & Fremling (2021d); (27) Ward et al. (2022); (28) Burke et al. (2021); (29) Perez-Fournon et al. (2021); (30) Wyatt (2021); (31) Siebert et al. (2020b); (32) Dahiwale & Fremling (2020g); (33) Siebert et al. (2020c ; (34) Perley et al. (2020c); (35) Siebert et al. (2020e) ; (36) Siebert et al. (2021c); (37) Siebert et al. (2020g); (38) Hung et al. (2020); (39) Davis (2021); (40) Angus (2021); (41) Davis et al. (2021); (42) Terreran et al. (2022); (43) Jacobson-Galán et al. (2022b); (44) Burke et al. (2020a); (45) Dahiwale & Fremling (2020e); (46) Terreran et al. (2020); (47) Taggart et al. (2021c); (48) Zhang et al. (2021); (49) Kilpatrick et al. (2021a); (50) Burke et al. (2020a); (51) Dahiwale & Fremling (2020e); (52) Terreran et al. (2020); (53) Taggart et al. (2021c); (54) Zhang et al. (2021); (55) Kilpatrick et al. (2021a); (56) Angus et al. (2022); (57) Dahiwale & Fremling (2020h); (58) Goodwin et al. (2023); (59) Gezari et al. (2021) ; (60) Siebert et al. (2021a); (61) Izzo et al. (2020); (62) Siebert et al. (2021b); (63) Pessi et al. (2021); (64) Perley et al. (2021); (65) Dimitriadis et al. (2020a); (66) Jacobson-Galán et al. (2022a); (67) Jacobson-Galán et al. (2022a); (68) Dimitriadis et al. (2022).

Table C2
ParSNIP Results for YSE DR1

Object	R.A.	Decl.	Spec. Class	Redshift, z	Prediction	Confidence	p_SNII	p_SNIa	p_SNIbc
2019lbi	190.088004	1.273998	SNII	0.038	SNIa	0.497	0.278	0.497	0.225
2019pmd	49.599161	−1.930453	SNIa-norm	0.026	SNIa	0.972	0.002	0.972	0.026
2019ppi	133.897475	49.160259	SNII	0.135	SNII	0.495	0.495	0.173	0.332
2019szh	147.176208	−8.734392	SNIa-norm	0.053	SNIa	1.0	0.0	1.0	0.0
2019tvv	177.772596	21.767258	SNIa-norm	0.059	SNIa	1.0	0.0	1.0	0.0
2019ucc	49.642135	−4.429429	NA	0.066	SNIbc	0.87	0.005	0.125	0.87
2019uev	142.706471	30.871843	NA	0.06	SNIbc	0.779	0.019	0.202	0.779
2019uez	51.172538	−0.681409	SNII	0.022	SNII	0.78	0.78	0.053	0.167
2019uit	192.567042	21.337646	SNIIn	0.086	SNII	0.457	0.457	0.101	0.442
2019ulo	133.278015	−6.329666	NA	0.041	SNII	0.986	0.986	0.004	0.01
2019unp	142.397468	35.289745	NA	0.145	SNIa	0.996	0.001	0.996	0.003
2019vuz	112.309986	42.076944	NA	0.21	SNIa	0.466	0.218	0.466	0.316
2019wbv	134.689142	−4.56562	NA	0.067	SNIbc	0.598	0.217	0.185	0.598
2019wbw	150.67996	−8.641407	NA	0.138	SNIa	0.957	0.003	0.957	0.04
2019wca	145.192026	−8.4785	NA	0.202	SNII	0.526	0.526	0.33	0.144

Note. A complete, machine-readable version of this table is available on Zenodo.

(This table is available in its entirety in machine-readable form.)

Table C3
ParSNIP Results on Other YSE DR1 Transients

Object	R.A.	Decl.	Spec. Class	Redshift, z	Prediction	Confidence	p_SNII	p_SNIa	p_SNIbc
2020apw	141.012066	28.814982	SN	0.224	SNIa	0.998	0.002	0.998	0.0
2020awu	171.348033	9.983899	SN	0.087	SNII	0.975	0.975	0.018	0.007
2020inp	226.622583	10.266889	SNIax	0.227	SNIa	0.993	0.007	0.993	0.0
2020neh	230.333667	14.069628	TDE	0.065	SNIa	0.533	0.002	0.533	0.465
2020nov	254.554085	2.117537	TDE	0.078	SNIbc	0.413	0.377	0.21	0.413
2020opy	239.107211	23.372504	TDE	0.195	SNIa	0.765	0.0	0.765	0.235
2020sck	17.645135	2.113941	SNIax	0.015	SNIa	0.768	0.001	0.768	0.231
2020xsy	163.766692	−1.539883	SLSN-II	0.121	SNII	0.492	0.492	0.185	0.323
2020zmn	17.014616	1.144884	LBV	0.057	SNII	0.971	0.971	0.022	0.007
2021aadc	1.136091	19.761508	SLSN-II	0.208	SNIbc	0.607	0.038	0.355	0.607
2021biy	190.516756	32.535522	LRN	0.081	SNII	0.681	0.681	0.011	0.308
2021bug	188.59485	2.317294	LBV	0.021	SNIbc	0.68	0.052	0.268	0.68
2021cob	187.605664	2.151898	SNIax	0.151	SNIa	0.791	0.018	0.791	0.191
2021ehb	46.949208	40.311269	TDE	0.082	SNII	0.444	0.444	0.222	0.334
2021kqp	223.794754	−6.985927	SN	0.095	SNIa	0.981	0.0	0.981	0.019
2021kyv	216.980358	33.00208	SN	0.105	SNII	0.997	0.997	0.002	0.001
2021nxq	216.336179	37.762421	SLSN-I	0.132	SNIbc	0.415	0.345	0.24	0.415
2021qxv	229.747063	−3.195855	TDE	0.168	SNII	0.842	0.842	0.018	0.14
2021seu	215.394287	37.90965	Other	0.077	SNII	0.378	0.378	0.274	0.348
2021uwx	333.35587	6.782652	SLSN-I	0.155	SNII	0.542	0.542	0.302	0.156

(This table is available in machine-readable form.)

Appendix D
Spectra

Here we describe the observations and data reduction for the spectra presented in Figures 28 and 29. A full analysis of all spectra for YSE DR1 objects will be presented in subsequent publications.

SN 2021hpr was discovered in NGC 3147 on 2021 April 2.48 by Itagaki (2021) and classified as an SN Ia on 2021 April 3.51 (Tomasella et al. 2021). We obtained a series of spectra between 2021 April 5 and April 19 with the Kast spectrograph on the Lick Shane telescope and the ALFOSC spectrograph on the Nordic Optical Telescope. In Ward et al. (2022), we describe this SN in detail and present one of the

spectra presented here. Zhang et al. (2022) also present data for SN 2021hpr.

SN 2021aamo was discovered in LEDA 1605946 by us on 2021 October 4.77 (Jones et al. 2021b) and classified by us as an SN Ia on 2022 October 16.68 (Davis 2021). The classification spectrum, obtained with the Kast spectrograph on the Lick Shane telescope, is presented in Figure 29.

To reduce the Kast data, we used the UCSC Spectral Pipeline[85] (Siebert et al. 2020d), a custom data-reduction pipeline based on procedures outlined by Foley et al. (2003), Silverman et al. (2012), and references therein. The 2D spectra were bias-corrected, flat-field corrected, adjusted for varying

---

[85] https://github.com/msiebert1/UCSC_spectral_pipeline

gains across different chips and amplifiers, and trimmed. 1D spectra were extracted using the optimal algorithm (Horne 1986). The spectra were wavelength-calibrated using internal comparison-lamp spectra with linear shifts applied by cross-correlating the observed night-sky lines in each spectrum to a master night-sky spectrum. Flux calibration and telluric correction were performed using standard stars at a similar airmass to that of the science exposures. We combine the sides by scaling one spectrum to match the flux of the other in the overlap region and use their error spectra to correctly weight the spectra when combining. More details of this process are discussed elsewhere (Foley et al. 2003; Silverman et al. 2012; Siebert et al. 2020d; Davis et al. 2022).

Data obtained with ALFOSC were reduced using standard techniques, which included correction for bias, overscan, and flat-field. Spectra of comparison lamps and standard stars acquired during the same night and with the same instrumental setting have been used for the wavelength and flux calibrations, respectively. We employed standard IRAF commands to extract all spectra.

Appendix E
Methodology Schematic

The methodology schematic of this work is presented in Figure E1.
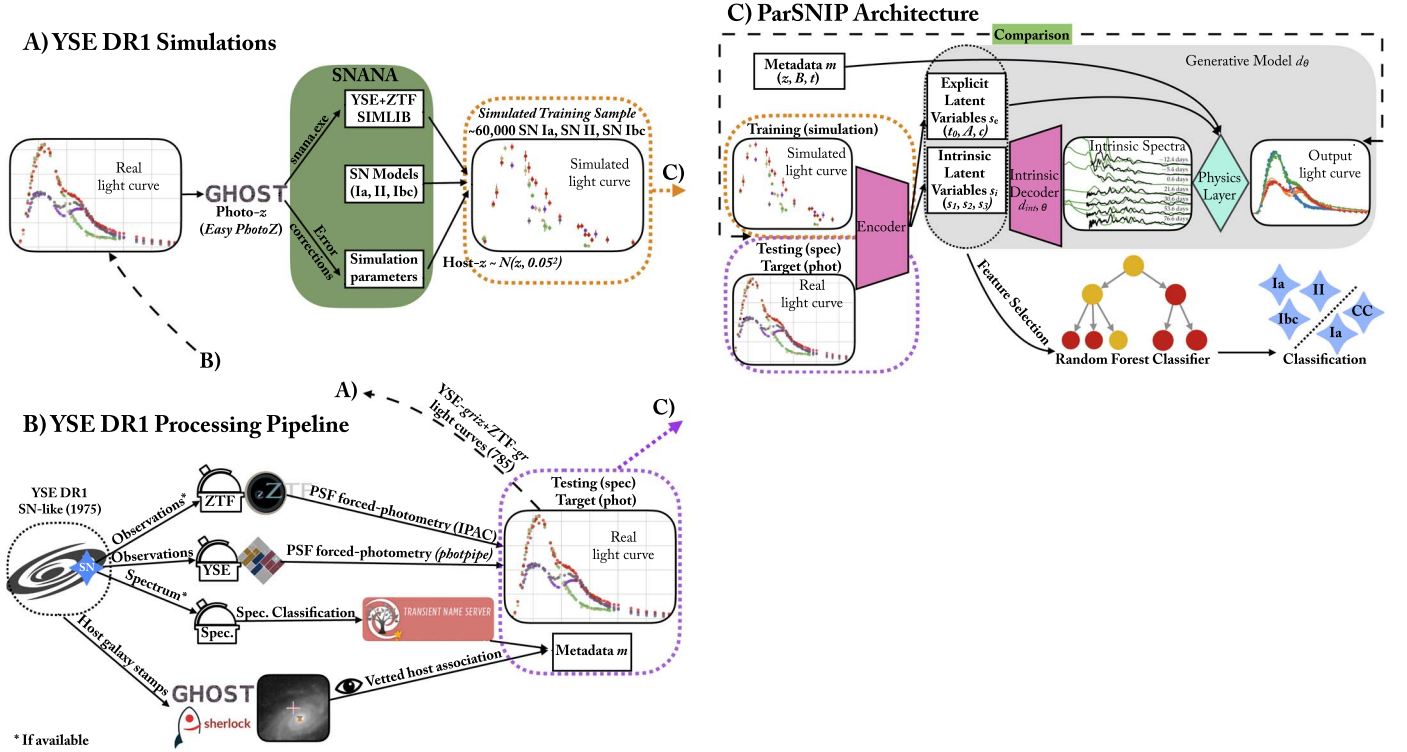
Figure E1. A streamlined workflow schematic of this work. The *A* process describes the method of generating realistic YSE DR1 simulations for training set generation, which is explained in detail in Section 5.6. The *B* process describes the YSE DR1 data processing pipeline, from SN observation (via YSE, ZTF) to light-curve generation with associated metadata (host association, redshift). The *B* process is described in detail in Sections 2.4–5.2. Lastly, the *C* process summarizes the adapted ParSNIP architecture, from training to feature selection and photometric classification. This is explored in Section 5.8.

ORCID iDs

P. D. Aleo ⓘ https://orcid.org/0000-0002-6298-1663
K. Malanchev ⓘ https://orcid.org/0000-0001-7179-7406
S. Sharief ⓘ https://orcid.org/0000-0002-0869-8760
D. O. Jones ⓘ https://orcid.org/0000-0002-6230-0151
G. Narayan ⓘ https://orcid.org/0000-0001-6022-0484
R. J. Foley ⓘ https://orcid.org/0000-0002-2445-5275
V. A. Villar ⓘ https://orcid.org/0000-0002-5814-4061
C. R. Angus ⓘ https://orcid.org/0000-0002-4269-7999
V. F. Baldassare ⓘ https://orcid.org/0000-0003-4703-7276
M. J. Bustamante-Rosell ⓘ https://orcid.org/0000-0003-0416-9818
D. Chatterjee ⓘ https://orcid.org/0000-0003-0038-5468
C. Cold ⓘ https://orcid.org/0000-0001-7666-1874
D. A. Coulter ⓘ https://orcid.org/0000-0003-4263-2228
K. W. Davis ⓘ https://orcid.org/0000-0002-5680-4660
S. Dhawan ⓘ https://orcid.org/0000-0002-2376-6979
M. R. Drout ⓘ https://orcid.org/0000-0001-7081-0082
A. Engel ⓘ https://orcid.org/0000-0003-2348-483X
K. D. French ⓘ https://orcid.org/0000-0002-4235-7337
A. Gagliano ⓘ https://orcid.org/0000-0003-4906-8447
C. Gall ⓘ https://orcid.org/0000-0002-8526-3963
J. Hjorth ⓘ https://orcid.org/0000-0002-4571-2306
M. E. Huber ⓘ https://orcid.org/0000-0003-1059-9603
W. V. Jacobson-Galán ⓘ https://orcid.org/0000-0002-3934-2644
C. D. Kilpatrick ⓘ https://orcid.org/0000-0002-5740-7747
D. Langeroodi ⓘ https://orcid.org/0000-0001-5710-8395
P. Macias ⓘ https://orcid.org/0000-0002-9946-4635
K. S. Mandel ⓘ https://orcid.org/0000-0001-9846-4417
R. Margutti ⓘ https://orcid.org/0000-0003-4768-7586
F. Matasić ⓘ https://orcid.org/0000-0001-5306-1948
P. McGill ⓘ https://orcid.org/0000-0002-1052-6749
J. D. R. Pierel ⓘ https://orcid.org/0000-0002-2361-7201
E. Ramirez-Ruiz ⓘ https://orcid.org/0000-0003-2558-3102
C. L. Ransome ⓘ https://orcid.org/0000-0003-4175-4960
C. Rojas-Bravo ⓘ https://orcid.org/0000-0002-7559-315X
M. R. Siebert ⓘ https://orcid.org/0000-0003-2445-3891
K. W. Smith ⓘ https://orcid.org/0000-0001-9535-3199
K. M. de Soto ⓘ https://orcid.org/0000-0002-9886-2834
M. C. Stroh ⓘ https://orcid.org/0000-0002-3019-4577
S. Tinyanont ⓘ https://orcid.org/0000-0002-1481-4676
K. Taggart ⓘ https://orcid.org/0000-0002-5748-4558
S. M. Ward ⓘ https://orcid.org/0000-0002-1763-2720
R. Wojtak ⓘ https://orcid.org/0000-0001-9666-3164
K. Auchettl ⓘ https://orcid.org/0000-0002-4449-9152
P. K. Blanchard ⓘ https://orcid.org/0000-0003-0526-2248
T. J. L. de Boer ⓘ https://orcid.org/0000-0001-5486-2747
B. M. Boyd ⓘ https://orcid.org/0000-0002-0622-1117
C. M. Carroll ⓘ https://orcid.org/0000-0003-3574-2963
K. C. Chambers ⓘ https://orcid.org/0000-0001-6965-7789
L. DeMarchi ⓘ https://orcid.org/0000-0003-4587-2366
G. Dimitriadis ⓘ https://orcid.org/0000-0001-9494-179X
S. A. Dodd ⓘ https://orcid.org/0000-0002-3696-8035
N. Earl ⓘ https://orcid.org/0000-0003-1714-7415
D. Farias ⓘ https://orcid.org/0000-0002-6886-269X
H. Gao ⓘ https://orcid.org/0000-0003-1015-5367
S. Gomez ⓘ https://orcid.org/0000-0001-6395-6702
M. Grayling ⓘ https://orcid.org/0000-0002-6741-983X
C. Grillo ⓘ https://orcid.org/0000-0002-5926-7143
E. E. Hayes ⓘ https://orcid.org/0000-0003-3847-0780
T. Hung ⓘ https://orcid.org/0000-0002-9878-7889
L. Izzo ⓘ https://orcid.org/0000-0001-9695-8472
N. Khetan ⓘ https://orcid.org/0000-0003-2720-8904
A. N. Kolborg ⓘ https://orcid.org/0000-0001-7364-4964
J. A. P. Law-Smith ⓘ https://orcid.org/0000-0001-8825-4790
C.-C. Lin ⓘ https://orcid.org/0000-0002-7272-5129
Y. Luo ⓘ https://orcid.org/0000-0002-4623-0683
E. A. Magnier ⓘ https://orcid.org/0000-0002-7965-2815
D. Matthews ⓘ https://orcid.org/0000-0002-4513-3849
B. Mockler ⓘ https://orcid.org/0000-0001-6350-8168
A. J. G. O'Grady ⓘ https://orcid.org/0000-0002-7296-6547
Y.-C. Pan ⓘ https://orcid.org/0000-0001-8415-6720
C. A. Politsch ⓘ https://orcid.org/0000-0003-3727-9167
S. I. Raimundo ⓘ https://orcid.org/0000-0002-6248-398X
A. Rest ⓘ https://orcid.org/0000-0002-4410-5387
R. Ridden-Harper ⓘ https://orcid.org/0000-0003-1724-2885
A. Sarangi ⓘ https://orcid.org/0000-0002-9820-679X
S. L. Schrøder ⓘ https://orcid.org/0000-0003-1735-8263
S. J. Smartt ⓘ https://orcid.org/0000-0002-8229-1731
G. Terreran ⓘ https://orcid.org/0000-0003-0794-5982
J. Vazquez ⓘ https://orcid.org/0000-0003-1576-0830
R. J. Wainscoat ⓘ https://orcid.org/0000-0002-1341-0952
Q. Wang ⓘ https://orcid.org/0000-0001-5233-6989
A. R. Wasserman ⓘ https://orcid.org/0000-0002-4186-6164
S. K. Yadavalli ⓘ https://orcid.org/0000-0002-0840-6940
R. Yarza ⓘ https://orcid.org/0000-0003-0381-1039
Y. Zenati ⓘ https://orcid.org/0000-0002-0632-8897

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, PhRvL, 119, 161101
Abbott, T. M. C., Allam, S., Andersen, P., et al. 2019, ApJL, 872, L30
Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Aleo, P. D., Malanchev, K., Sharief, S. N., et al. 2022, The Young Supernova Experiment Data Release 1 (YSE DR1) Light Curves, v1.0.0, Zenodo, doi:10.5281/zenodo.7317476
Aleo, P. D., Malanchev, K. L., Pruzhinskaya, M. V., et al. 2022, NewA, 96, 101846
Alves, C. S., Peiris, H. V., Lochner, M., et al. 2022, ApJS, 258, 23
Andri, S., et al. 2022, DescTools: Tools for Descriptive Statistics, https://cran.r-project.org/web/packages/DescTools/index.html
Angus, C. 2020, TNSCR, 2020-3109
Angus, C. 2021, TNSCR, 2021-649
Angus, C. R., Baldassare, V. F., Mockler, B., et al. 2022, NatAs, 6, 1452
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Balcon, C. 2020a, TNSCR, 2020-2839
Balcon, C. 2020b, TNSCR, 2020-1001
Baldry, I. K., Liske, J., Brown, M. J. I., et al. 2018, MNRAS, 474, 3875
Baltay, C., Rabinowitz, D., Hadjiyska, E., et al. 2013, PASP, 125, 683
Barbary, K., Bailey, S., Barentsen, G., et al. 2022, SNCosmo, v2.8.0, Zenodo, doi:10.5281/zenodo.6363879
Bazin, G., Palanque-Delabrouille, N., Rich, J., et al. 2009, A&A, 499, 653
Beck, R., Dobos, L., Budavári, T., Szalay, A. S., & Csabai, I. 2016, MNRAS, 460, 1371
Beck, R., Szapudi, I., Flewelling, H., et al. 2021, MNRAS, 500, 1633
Becker, A. 2015, HOTPANTS: High Order Transform of PSF ANd Template Subtraction, Astrophysics Source Code Library, ascl:1504.004
Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002
Bentley, J. L. 1975, Commun. ACM, 18, 509
Bernstein, J. P., Kessler, R., Kuhlmann, S., et al. 2012, ApJ, 753, 152
Blondin, S., & Tonry, J. L. 2007, ApJ, 666, 1024
Boone, K. 2019, AJ, 158, 257
Boone, K. 2021, AJ, 162, 275
Brammer, G. B., Sánchez-Janssen, R., Labbé, I., et al. 2012, ApJL, 758, L17
Brout, D., Sako, M., Scolnic, D., et al. 2019, ApJ, 874, 106
Brout, D., Taylor, G., Scolnic, D., et al. 2022, ApJ, 938, 111
Budavári, T., & Szalay, A. S. 2008, ApJ, 679, 301
Burhanudin, U. F., & Maund, J. R. 2023, MNRAS, 521, 1601
Burke, J., Hiramatsu, D., Howell, D. A., et al. 2020a, TNSCR, 2020-3650
Burke, J., Sand, D., Hiramatsu, D., et al. 2020b, TNSCR, 2020-3032

Burke, J., Pellegrino, C., Hiramatsu, D., et al. 2021, TNSCR, 2021-1543
Cai, Y. Z., Pastorello, A., Fraser, M., et al. 2022, A&A, 667, A4
Cartier, R., Hamuy, M., Pignata, G., et al. 2014, ApJ, 789, 89
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Charnock, T., & Moss, A. 2017, ApJL, 837, L28
Chornock, R., & Filippenko, A. V. 2002, IAU Circ., 7783, 3
Clocchiatti, A., Wheeler, J. C., Brotherton, M. S., et al. 1996, ApJ, 462, 462
Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039
Cooper, M. C., Aird, J. A., Coil, A., et al. 2011, ApJS, 193, 14
Coulter, D. A., Jones, D. O., McGill, P., et al. 2022, YSE-PZ: An Open-source Target and Observation Management System, v0.3.0, Zenodo, doi:10.5281/zenodo.7278430
Csoernyei, G., Taubenberger, S., Vogl, C., et al. 2021, TNSCR, 2021-2990
Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
Dahiwale, A., & Fremling, C. 2020a, TNSCR, 2020-753
Dahiwale, A., & Fremling, C. 2020b, TNSCR, 2020-1656
Dahiwale, A., & Fremling, C. 2020c, TNSCR, 2020-2260
Dahiwale, A., & Fremling, C. 2020d, TNSCR, 2020-2724
Dahiwale, A., & Fremling, C. 2020e, TNSCR, 2020-2885
Dahiwale, A., & Fremling, C. 2020f, TNSCR, 2020-152
Dahiwale, A., & Fremling, C. 2020g, TNSCR, 2020-1573
Dahiwale, A., & Fremling, C. 2020h, TNSCR, 2020-3800
Dahiwale, A., & Fremling, C. 2021a, TNSCR, 2021-1721
Dahiwale, A., & Fremling, C. 2021b, TNSCR, 2021-358
Dahiwale, A., & Fremling, C. 2021c, TNSCR, 2021-1008
Dahiwale, A., & Fremling, C. 2021d, TNSCR, 2021-603
Davis, K. 2021, TNSCR, 2021-3592
Davis, K. W., Siebert, M. R., Rojas-Bravo, C., et al. 2021, TNSCR, 2021-3046
Davis, K. W., Taggart, K., Tinyanont, S., et al. 2022, arXiv:2211.05134
Davis, T. M., Hinton, S. R., Howlett, C., & Calcino, J. 2019, MNRAS, 490, 2948
De, K., Kasliwal, M. M., Tzanidakis, A., et al. 2020, ApJ, 905, 58
Demianenko, M., Malanchev, K., Samorodova, E., et al. 2022, arXiv:2209.07542
Dimitriadis, G., Foley, R. J., Arendse, N., et al. 2022, ApJ, 927, 78
Dimitriadis, G., Foley, R. J., Terreran, G., & Angus, C. R. 2021, TNSCR, 2021-2293
Dimitriadis, G., Siebert, M. R., & Foley, R. J. 2020a, TNSCR, 2020-2258
Dimitriadis, G., Siebert, M. R., Taggart, K., Tinyanont, S., & Foley, R. J. 2020b, TNSCR, 2020-2840
Dimitriadis, G., Siebert, M. R., Taggart, K., Tinyanont, S., & Foley, R. J. 2020c, ATel, 14024, 1
Do, A., Tucker, M. A., Payne, A. V., et al. 2020, TNSCR, 2020-607
Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, ApJ, 696, 870
Drinkwater, M. J., Byrne, Z. J., Blake, C., et al. 2018, MNRAS, 474, 4151
Fausnaugh, M. M., Vallely, P. J., Kochanek, C. S., et al. 2021, ApJ, 908, 51
Folatelli, G., Contreras, C., Phillips, M. M., et al. 2006, ApJ, 641, 1039
Foley, R. J., Berger, E., Fox, O., et al. 2011, ApJ, 732, 32
Foley, R. J., Desroches, L. B., Wong, D. S., Moore, M. R., & Filippenko, A. V. 2007, CBET, 974, 1
Foley, R. J., Papenkova, M. S., Swift, B. J., et al. 2003, PASP, 115, 1220
Förster, F., Cabrera-Vives, G., Castillo-Navarrete, E., et al. 2021, AJ, 161, 242
Fremling, C., Miller, A. A., Sharma, Y., et al. 2020, ApJ, 895, 32
Gagliano, A., Contardo, G., Foreman-Mackey, D., Malz, A., & Aleo, P. 2022a, ApJ, submitted
Gagliano, A., Izzo, L., Kilpatrick, C. D., et al. 2022b, ApJ, 924, 55
Gagliano, A., Narayan, G., Engel, A., Carrasco Kind, M. & LSST Dark Energy Science Collaboration 2021, ApJ, 908, 170
Galbany, L., Lavers, A. L. C., Foley, R., et al. 2020, TNSCR, 2020-1270
Gallego-Cano, E., Izzo, L., Dominguez-Tagle, C., et al. 2022, A&A, 666, A13
Gezari, S., Hammerstein, E., Yao, Y., et al. 2021, TNSAN, 103, 1
Gezari, S., Martin, D. C., Forster, K., et al. 2013, ApJ, 766, 60
Gil de Paz, A., Boissier, S., Madore, B. F., et al. 2007, ApJS, 173, 185
Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98
Goodman, L. A. 1965, Technometrics, 7, 247
Goodwin, A. J., Miller-Jones, J. C. A., van Velzen, S., et al. 2023, MNRAS, 518, 847
Graham, M. L., Fremling, C., Perley, D. A., et al. 2022, MNRAS, 511, 241
Graur, O., Rodney, S. A., Maoz, D., et al. 2014, ApJ, 783, 28
Gutiérrez, C. P., Anderson, J. P., Hamuy, M., et al. 2017, ApJ, 850, 89
Guzzo, L., Scodeggio, M., Garilli, B., et al. 2014, A&A, 566, A108
Hicken, M., Friedman, A. S., Blondin, S., et al. 2017, ApJS, 233, 6
Hiramatsu, D., Hosseinzadeh, G., Burke, J., et al. 2020, TNSCR, 2020–3728
Hložek, R., Ponder, K. A., Malz, A. I., et al. 2020, arXiv:2012.12392
Ho, A. Y. Q., Perley, D. A., Gal-Yam, A., et al. 2021, arXiv:2105.08811

Holoien, T. W. S., Brown, J. S., Vallely, P. J., et al. 2019, MNRAS, 484, 1899
Holoien, T. W. S., Stanek, K. Z., Kochanek, C. S., et al. 2017a, MNRAS, 464, 2672
Holoien, T. W. S., Brown, J. S., Stanek, K. Z., et al. 2017b, MNRAS, 467, 1098
Holoien, T. W. S., Brown, J. S., Stanek, K. Z., et al. 2017c, MNRAS, 471, 4966
Hönig, S. F., Watson, D., Kishimoto, M., et al. 2017, MNRAS, 464, 1693
Horne, K. 1986, PASP, 98, 609
Hosseinzadeh, G., Dauphin, F., Villar, V. A., et al. 2020, ApJ, 905, 93
Howell, S. B., Sobeck, C., Haas, M., et al. 2014, PASP, 126, 398
Hsu, B., Hosseinzadeh, G., Villar, V. A., & Berger, E. 2022, ApJ, 937, 13
Huber, M., Chambers, K. C., Flewelling, H., et al. 2015, ATel, 7153
Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, ApJS, 199, 26
Hung, T., Taggart, K., & Foley, R. J. 2020, ATel, 14167, 1
Hunter, J. D. 2007, CSE, 9, 90
Itagaki, K. 2021, TNSTR, 2021-998
Izzo, L., Angus, C., Bruun, S., et al. 2020, TNSAN, 75, 1
Jacobson-Galán, W., Venkatraman, P., Margutti, R., et al. 2022a, ApJ, 932, 58
Jacobson-Galán, W. V., Dessart, L., Jones, D. O., et al. 2022b, ApJ, 924, 15
Jarrett, T. H., Cohen, M., Masci, F., et al. 2011, ApJ, 735, 112
Jha, S., Matheson, T., Challis, P., Kirshner, R., & Calkins, M. 2001, IAU Circ., 7569, 2
Jha, S. W., Dai, M., Perez-Fournon, I., et al. 2020, TNSCR, 2020-2192
Jolliffe, I. 2002, Principal Component Analysis (New York: Springer Verlag)
Jones, D. H., Read, M. A., Saunders, W., et al. 2009, MNRAS, 399, 683
Jones, D. O., Foley, R. J., Narayan, G., et al. 2021a, ApJ, 908, 143
Jones, D. O., French, K. D., Agnello, A., et al. 2021b, TNSTR, 2021-3402
Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2017, ApJ, 843, 6
Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2018, ApJ, 857, 51
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, Proc. SPIE, 4836, 154
Karpenka, N. V., Feroz, F., & Hobson, M. P. 2012, MNRAS, 429, 1278
Kasliwal, M. M. 2012, PASA, 29, 482
Kelly, P. L., Filippenko, A. V., Burke, D. L., et al. 2015, Sci, 347, 1459
Kelly, P. L., & Kirshner, R. P. 2012, ApJ, 759, 107
Kenworthy, W. D., Jones, D. O., Dai, M., et al. 2021, ApJ, 923, 265
Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, PASP, 121, 1028
Kessler, R., Narayan, G., Avelino, A., et al. 2019, PASP, 131, 094501
Kilpatrick, C., Jacobson-Galan, W. V., & Tinyanont, S. 2021a, HST Proposal, 16874
Kilpatrick, C. D., Drout, M. R., Auchettl, K., et al. 2021b, MNRAS, 504, 2073
Kim, S.-L., Lee, C.-U., Park, B.-G., et al. 2016, JKAS, 49, 37
Kindratenko, V., Mu, D., Zhan, Y., et al. 2020, HAL: Computer System for Scalable Deep Learning (New York: Association for Computing Machinery), 41
Kingma, D. P., & Welling, M. 2013, arXiv:1312.6114
Kumar, B., Pandey, S. B., Sahu, D. K., et al. 2013, MNRAS, 431, 308
Law, N. M., Fors, O., Ratzloff, J., et al. 2015, PASP, 127, 234
Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, PASP, 121, 1395
Leadbeater, R. 2020, TNSCR, 2020-926
Li, W., Leaman, J., Chornock, R., et al. 2011, MNRAS, 412, 1441
Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, ApJS, 184, 218
Lipunov, V., Kornilov, V., Gorbovskoy, E., et al. 2010, AdAst, 2010, 349171
Liu, Y., & Modjaz, M. 2014, arXiv:1405.1437
Liu, Y.-Q., Modjaz, M., Bianco, F. B., & Graur, O. 2016, ApJ, 827, 90
Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, ApJS, 225, 31
Lokken, M., Gagliano, A., Narayan, G., et al. 2023, MNRAS, 520, 2887
LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, AJ, 118, 1406
Magnier, E. A., Chambers, K. C., Flewelling, H. A., et al. 2020a, ApJS, 251, 3
Magnier, E. A., Sweeney, W. E., Chambers, K. C., et al. 2020b, ApJS, 251, 5
Malanchev, K. L., Pruzhinskaya, M. V., Korolev, V. S., et al. 2021, MNRAS, 502, 5147
Margon, B. 1999, RSPTA, 357, 93
Masci, F. J., Laher, R. R., Rusholme, B., et al. 2019, PASP, 131, 018003
Matheson, T., Stubens, C., Wolf, N., et al. 2021, AJ, 161, 107
Millard, J., Branch, D., Baron, E., et al. 1999, ApJ, 527, 746
Modelers, P. 2021, Libraries & Recommended Citations for using PLAsTiCC Models, v2, Zenodo, doi:10.5281/zenodo.4419884
Modjaz, M., Blondin, S., Kirshner, R. P., et al. 2014, AJ, 147, 99
Modjaz, M., Liu, Y. Q., Bianco, F. B., & Graur, O. 2016, ApJ, 832, 108
Möller, A., Peloton, J., Ishida, E. E. O., et al. 2021, MNRAS, 501, 3272

Moran, S., Gonzalez-Gaitan, S., Silvestre, J., et al. 2021, TNSCR, 2021-679
Morrell, N., Folatelli, G., & Gonzalez, S. 2006, CBET, 669, 1
Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, PASP, 131, 118002
NASA/IPAC Extragalactic Database (NED) 2019, NASA/IPAC Extragalactic Database (NED), IPAC, doi:10.26132/NED1
Neumann, K. D., Holoien, T. W. S., Kochanek, C. S., et al. 2023, MNRAS, 520, 4356
Nomoto, K., Kobayashi, C., & Tominaga, N. 2013, ARA&A, 51, 457
Nordin, J., Brinnel, V., van Santen, J., et al. 2019, A&A, 631, A147
pandas development team, T 2020, pandas-dev/pandas: Pandas, v1.1.2, Zenodo, doi:10.5281/zenodo.3509134
Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., & Fouchez, D. 2019a, A&A, 621, A26
Pasquet, J., Pasquet, J., Chaumont, M., & Fouchez, D. 2019b, A&A, 627, A21
Patterson, M. T., Bellm, E. C., Rusholme, B., et al. 2019, PASP, 131, 018001
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Perez-Fournon, I., Poidevin, F., Angel, C. J., et al. 2021, TNSCR, 2021-1676
Perley, D. A., Fremling, C., Sollerman, J., et al. 2020d, ApJ, 904, 35
Perley, D. A., Schulze, S., Fremling, C., et al. 2021, TNSAN, 156, 1
Perley, D. A., Taggart, K., Dahiwale, A., & Fremling, C. 2020a, TNSCR, 2020-754
Perley, D. A., Taggart, K., Dahiwale, A., & Fremling, C. 2020b, TNSCR, 2020-784
Perley, D. A., Taggart, K., Dahiwale, A., & Fremling, C. 2020c, TNSCR, 2020-2383
Pessi, P. J., Gromadzki, M., & Strotjohann, N. L. 2021, TNSCR, 2021-2659
Pignata, G., Maza, J., Antezana, R., et al. 2009, in AIP Conf. Proc. 1111, Probing Stellar Populations out to the Distant Universe: CEFALU 2008, ed. G. Giobbi et al. (Melville, NY: AIP), 551
Pineda, J., Gonzalez-Gaitan, S., Galbany, L., et al. 2020, TNSCR, 2020-802
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Ponticello, N. J., Khandrika, H., Madison, D. R., et al. 2006, IAU Circ., 8709, 1
Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, AJ, 156, 123
Qu, H., Sako, M., Möller, A., & Doux, C. 2021, AJ, 162, 67
Quimby, R. M., De Cia, A., Gal-Yam, A., et al. 2018, ApJ, 855, 2
Quimby, R. M., Kulkarni, S. R., Kasliwal, M. M., et al. 2011, Natur, 474, 487
Quimby, R. M., Wheeler, J. C., Höflich, P., et al. 2007, ApJ, 666, 1093
Rest, A., Scolnic, D., Foley, R. J., et al. 2014, ApJ, 795, 44
Rest, A., Stubbs, C., Becker, A. C., et al. 2005, ApJ, 634, 1103
Richard, M. D., & Lippmann, R. P. 1991, Neural Comput., 3, 461
Richardson, D., Jenkins, Robert, L., Wright, I., & Maddox, L, J. 2014, AJ, 147, 118
Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56
Rodney, S. A., Riess, A. G., Strolger, L.-G., et al. 2014, AJ, 148, 13
Rojas, R. 1996, Neural Comput., 8, 41
Sahu, D. K., Tanaka, M., Anupama, G. C., Gurugubelli, U. K., & Nomoto, K. 2009, ApJ, 697, 676
Salvato, M., Ilbert, O., & Hoyle, B. 2019, NatAs, 3, 212
Salvo, M. E., Cappellaro, E., Mazzali, P. A., et al. 2001, MNRAS, 321, 254
Sánchez-Sáez, P., Reyes, I., Valenzuela, C., et al. 2021, AJ, 161, 141
Sand, D. 2019, Chandra Proposal, 21508740
Sand, D. J., Sarbadhicary, S. K., Pellegrino, C., et al. 2021, ApJ, 922, 21
Sanders, N. E., Soderberg, A. M., Gezari, S., et al. 2015, ApJ, 799, 208
Sanders, N. E., Soderberg, A. M., Levesque, E. M., et al. 2012, ApJ, 758, 132
Scalzo, R. A., Parent, E., Burns, C., et al. 2019, MNRAS, 483, 628
Schlafly, E. F., Finkbeiner, D. P., Jurić, M., et al. 2012, ApJ, 756, 158
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Schuldt, S., Suyu, S. H., Cañameras, R., et al. 2021, A&A, 651, A55
Scodeggio, M., Guzzo, L., Garilli, B., et al. 2018, A&A, 609, A84
Scolnic, D., Casertano, S., Riess, A., et al. 2015, ApJ, 815, 117
Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, ApJ, 859, 101
Seabold, S., & Perktold, J. 2010, Proc. 9th Python in Science Conf., ed. S. van der Walt & J. Millman, 92
Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, ApJ, 788, 48

Siebert, M. R., Davis, K., Tinyanont, S., Foley, R. J., & Strasburger, E. 2021a, TNSCR, 2021-2383
Siebert, M. R., Dimitriadis, G., & Foley, R. J. 2020a, TNSCR, 2020-1847
Siebert, M. R., Dimitriadis, G., & Foley, R. J. 2020b, TNSCR, 2020-544
Siebert, M. R., Dimitriadis, G., & Foley, R. J. 2020c, TNSCR, 2020-2271
Siebert, M. R., Dimitriadis, G., & Foley, R. J. 2021b, TNSCR, 2021-133
Siebert, M. R., Dimitriadis, G., Polin, A., & Foley, R. J. 2020d, ApJL, 900, L27
Siebert, M. R., Taggart, K., Dimitriadis, G., et al. 2020e, TNSCR, 2020-3770
Siebert, M. R., Taggart, K., & Foley, R. J. 2021c, TNSCR, 2021-449
Siebert, M. R., Tinyanont, S., Taggart, K., Dimitriadis, G., & Foley, R. J. 2020g, TNSCR, 2020-3121
Silverman, J. M., Foley, R. J., Filippenko, A. V., et al. 2012, MNRAS, 425, 1789
Smartt, S. J., Valenti, S., Fraser, M., et al. 2015, A&A, 579, A40
Smith, K. W., Smartt, S. J., Young, D. R., et al. 2020, PASP, 132, 085002
Smith, N., Li, W., Foley, R. J., et al. 2007, ApJ, 666, 1116
Srivastav, S., Smartt, S. J., McBrien, O., et al. 2020, TNSCR, 2020-2148
Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30
Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485
Stritzinger, M., Mazzali, P., Phillips, M. M., et al. 2009, ApJ, 696, 713
Strolger, L. G., Smith, R. C., Suntzeff, N. B., et al. 2002, AJ, 124, 2905
Taggart, K., Johnson, J. L., & Foley, R. J. 2021a, TNSCR, 2021-2747
Taggart, K., Johnson, J. L., & Foley, R. J. 2021b, ATel, 14841, 1
Taggart, K., & Perley, D. A. 2021, MNRAS, 503, 3931
Taggart, K., Tinyanont, S., Foley, R. J., & Gagliano, A. 2021c, ATel, 14959, 1
Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, PASJ, 70, S9
Tanaka, M., Tominaga, N., Morokuma, T., et al. 2016, ApJ, 819, 5
Tarrío, P., & Zarattini, S. 2020, A&A, 642, A102
Taubenberger, S. 2017, in Handbook of Supernovae, ed. A. W. Alsabti & P. Murdin (Cham: Springer), 317
Terreran, G., Jacobson-Galan, W., & Blanchard, P. K. 2020, ATel, 14115, 1
Terreran, G., Jacobson-Galan, W. V., Groh, J. H., et al. 2022, ApJ, 926, 20
Tinyanont, S., Millar-Blanchaer, M., Kasliwal, M. M., et al. 2021, NatAs, 5, 544
Tinyanont, S., Siebert, M. R., Dimitriadis, G., & Foley, R. J. 2020, TNSCR, 2020-2588
Tomasella, L., Benetti, S., Cappellaro, E., & Pastorello, A. 2021, TNSAN, 107, 1
Tominaga, N., Tanaka, M., Nomoto, K., et al. 2005, ApJL, 633, L97
Tonry, J. L. 2011, PASP, 123, 58
Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, PASP, 130, 064505
Tucker, M. A., Payne, A. V., Hinkle, J., et al. 2020, TNSCR, 2020-861
Valenti, S., Benetti, S., Cappellaro, E., et al. 2008, MNRAS, 383, 1485
Valenti, S., Sand, D. J., Yang, S., et al. 2017, ApJL, 848, L24
Valenti, S., Taubenberger, S., Pastorello, A., et al. 2012, ApJL, 749, L28
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Van Dyk, S. D., Zheng, W., Brink, T. G., et al. 2018, ApJ, 860, 90
Villar, V. A., Berger, E., Metzger, B. D., & Guillochon, J. 2017, ApJ, 849, 70
Villar, V. A., Berger, E., Miller, G., et al. 2019, ApJ, 884, 83
Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020, ApJ, 905, 94
Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020, Light Curves of Pan-STARRS1 SN-like Transients, v1, Zenodo, doi:10.5281/zenodo.3974950
Vincenzi, M., Sullivan, M., Firth, R. E., et al. 2019, MNRAS, 489, 5802
Walker, E. S., Mazzali, P. A., Pian, E., et al. 2014, MNRAS, 442, 2768
Ward, S. M., Thorp, S., Mandel, K. S., et al. 2022, arXiv:2209.10558
Weil, K. E., & Milisavljevic, D. 2020, TNSCR, 2020-3664
Williamson, M., Modjaz, M., & Bianco, F. B. 2019, ApJL, 880, L22
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Wyatt, S. 2021, TNSCR, 2021-2003
Yaron, O., & Gal-Yam, A. 2012, PASP, 124, 668
Zhang, J., Zhai, Q., & Wang, X. 2021, ATel, 14398, 1
Zhang, Y., Zhang, T., Danzengluobu, et al. 2022, PASP, 134, 074201
Zhou, R., Cooper, M. C., Newman, J. A., et al. 2019, MNRAS, 488, 4565
Zhou, R., Newman, J. A., Mao, Y.-Y., et al. 2021, MNRAS, 501, 3309