Sharing Space: A Survey-agnostic Variational Autoencoder for Supernova Science

Kaylee de Soto
Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA
kaylee.de_soto@cfa.harvard.edu

Ana Sofia Uzsoy
Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA

V. Ashley Villar
Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions

Abstract

The next decade of large-scale astronomical surveys will facilitate the detection and characterization of millions of supernovae across multiple frequency domains. However, this photometry cannot easily be combined across astronomical surveys with different filter profiles, observing patterns, and systematics. Here, we present a survey-agnostic variational autoencoder that can encode supernova light curves into a shared latent space irrespective of the observing instrument. We show that encouraging a filter-invariant latent space through pre-training and contrastive learning (1) yields reconstructions that evolve smoothly over filter wavelength and (2) improves classification of encodings from sparser surveys.

Introduction

Supernovae (SNe), or the explosive deaths of stars, are discovered by the thousands annually thanks to various ground-based astronomical surveys like the Zwicky Transient Facility (ZTF; [1]) and the Young Supernova Experiment (YSE; [2]). Although SNe have been historically classified based on their spectroscopic properties (see e.g., [3] for a review), the abundance of photometric survey data has driven the development of new data-driven methodologies for photometric classification (i.e., using a time series of flux measurements). There are many examples of photometric supernova classification pipelines that perform well [4, 5, 6, 7, 8, and others]; however, they are all trained to be used with data from a *single* survey. In anticipation of the Vera C. Rubin Legacy Survey of Space and Time (LSST; [9],[10]), scalable and generalizable architectures are critical for properly analyzing large amounts of data while minimizing computational cost [11].

Here, we build on recent efforts to use variational autoencoders (VAEs) to learn a smooth latent representation of SN observations [12, 13, 14]. We present the first survey-agnostic VAE for SN science that can leverage multiple, disparate datasets to learn a more informed latent representation. As one downstream application, we show that a SN classifier trained on survey-agnostic encodings outperforms one trained on encodings from a baseline single-survey VAE. Our code is publicly available via Github[1].

[1] https://github.com/kdesoto-astro/survey-agnostic-sn-vae

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

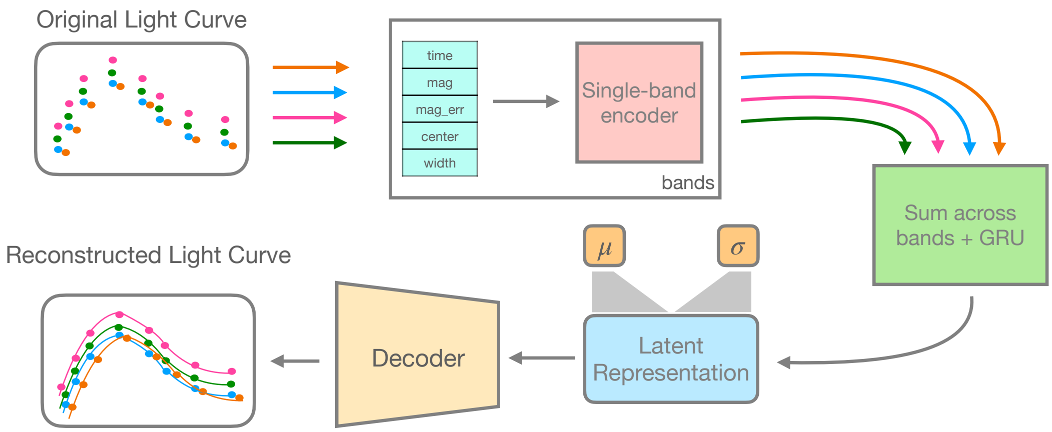

Figure 1: Overview of model architecture. Photometry is divided into single-filter light curves and passed through a time- and filter-distributed multilayer perceptron (MLP). The per-filter outputs are averaged and passed into a gated recurrent unit (GRU) and subsequently encoded in a four-dimensional latent space. A point in latent space can then be decoded back through an MLP into a reconstructed light curve.

Methods

Data: We pre-train our network using multiband photometry generated by the "default" model of `MOSFiT`, a standard one-zone SN model powered by radioactive decay [15]. We simulate light curves across multiple bands from Rubin, ZTF, Pan-STARRS [16], SWIFT [17], and 2MASS [18]. We set survey observing details according to Appendix A. We further augment the simulated events by resampling magnitudes from uncertainties, subsampling observed phases, and/or truncating the light curve within randomly selected phase ranges.

We fine-tune the pre-trained VAE using photometry from ZTF, downloaded from the Transient Name Server (TNS;[19]) and ALeRCE data broker [20], and Pan-STARRS, downloaded from the YSE Data Release 1 (DR1; [21]). These are well-suited for our study here: while ZTF typically has longer baselines and higher-cadence data compared to YSE, YSE has higher wavelength coverage (here we include ZTF's g- and r-bands, and Pan-STARRS's g-, r-, i-, and z-bands for YSE). Additionally, the YSE dataset has sufficiently few events and observations per event to benefit from blending with the larger ZTF dataset. For both the simulated and observed ZTF datasets, we exclude bands with fewer than five SNR ≥ 4 detections (we keep sparser Pan-STARRS light curves). We then remove events with fewer than two remaining bands. This leaves 4,622 events in the augmented pre-training dataset and 8,666 events in the combined ZTF and Pan-STARRS dataset. 1,576 samples have both ZTF and Pan-STARRS observations, 6,700 have only ZTF observations, and 390 have only Pan-STARRS observations.

For each sample, single-filter light curves are duplicated until there are six total bands to maintain uniform input dimensionality, and a Gaussian process is used to interpolate missing filters for each observed phase. All interpolated light curves are either truncated or padded to 32 time steps. The model input for each event is then a series of phases with associated multiband absolute magnitudes, magnitude uncertainties, filter central wavelengths, and filter widths.

Model Architecture: A schematic of our pipeline is shown in Figure 1. To ensure the encodings are invariant to the input filter order (which is typically fixed in the literature), the inputs are separated by filter and passed into a time- and filter-distributed encoder: here, a multi-layer perceptron (MLP) with two dense layers. Because permutation invariance corresponds to an arithmetic mean across filters [22], we average the outputs of the single-filter network across filters before feeding into a gated recurrent unit (GRU). We pass the final state of the GRU into two dense layers that output a four-dimensional latent mean and latent log variance. We sample from these defined latent distributions using the reparametrization trick [23], and concatenate these samples with the original phases, central filter wavelengths, and filter widths. These inputs pass through a decoder of time distributed dense layers, yielding reconstructed absolute magnitudes. All hidden layers have sixteen nodes and are divided by leaky ReLU activations. We note that this architecture is more appropriate than, e.g., one-hot encoding survey as an appended integer (as then the system must be retrained for each new survey, [24]). Transfer VAEs (in which an observation is "conditioned" on a survey by again appending a relevant latent variable), similarly, requires retraining for novel surveys.

Loss function: Our loss function is a sum of reconstruction loss, Kullback-Leibler (KL) divergence, and contrastive loss. The KL divergence enforces continuity and regularity in the latent space by pushing latent distributions closer to unit Gaussians [25]. The reconstruction loss measures similarity between the original and reconstructed light curves:

$$L_{\mathrm{recon}} = \frac{1}{N_{\mathrm{tot}}} \sum_{j}^{N_{\mathrm{LC}}} \sum_{i}^{N_{t,j}} \frac{(M_{\mathrm{pred},(i,j)} - M_{(i,j)})^2}{\sigma^2_{M,(i,j)}} \tag{1}$$

where $M_{\mathrm{pred},(i,j)}$ is the absolute magnitude at the ith time step of the jth single-filter light curve. N_{LC} is the number of such light curves in the dataset, $N_{t,j}$ is the number of time steps for light curve j; and $N_{\mathrm{tot}} = \sum_j^{N_{\mathrm{LC}}} N_{t,j}$ is the total number of time steps across all light curves. We do not include interpolated or padded magnitudes in this calculation.

Finally, the contrastive loss ensures that observations from the same event observed in multiple surveys will be co-located in the learned latent space:

$$L_{\mathrm{contrast}} = \frac{1}{N^2} \sum_{i}^{N} \log \frac{\sum_{j \in \phi(i)} \exp\left[-d(z_i, z_j)/\tau\right]}{\sum_{j \neq i} \exp\left[-d(z_i, z_j)/\tau\right]} \tag{2}$$

where N is the number of events and τ is a temperature parameter we set to 1. The mapping $\phi(i)$ returns all sample indices not equal to i that correspond to the same underlying event, and $d(z_i, z_j)$ is the standard Mahalanobis distance between the latent encodings z_i and z_j [26].

Training loop: We split both the pre-training and observed datasets into a train and validation set (9:1). Each epoch, we randomly divide each sample into two sets of photometry by subsampling three filters for each. This creates training pairs for contrastive learning. We train the aforementioned architecture with batch size of 128 and learning rate of 0.001. We first pre-train with the simulated dataset and 250 epochs of only the contrastive and KL loss (no reconstruction) to ensure that the encodings are grouped by unique events. Then, reconstruction loss is added and training continues for 1,150 epochs (until validation losses plateau). We note that, due to limited samples, we do *not* include a separate test set; such a set would validate "true" performance on unseen data and will be explored when this model is pushed into production.

After pre-training, we continue training five variants of the model using the ZTF and Pan-STARRS datasets to explore the importance of various components of our pipeline. The first two variants use the full dataset and are trained with the contrastive loss. The first variant "freezes" all layers excluding the bottleneck mean/log variance layers and first decoder layer. Previous works [27] have explored the benefits of partially frozen pre-trained networks in transfer learning to smaller or sparser datasets. The third variant does not utilize the contrastive loss. Finally, the last two are trained without pre-training and only on either the ZTF or Pan-STARRS light curves. These serve as our baseline models and most closely align with previous works. We note that we do not train a model on the combined ZTF/Pan-STARRS dataset without pre-training; while this work focuses on the impact of contrastive learning on VAE performance, future work will better isolate the effect of pre-training. For each variant, we train for an additional 200 epochs (following pre-training) at the same learning rate of 0.001. Training plateaus by this epoch in all variants.

Results & Discussion

We see from Table 1 that all validation set loss terms are minimized by the unfrozen model that is trained on both Pan-STARRS and ZTF light curves and uses contrastive learning (the "optimal" model). Without contrastive learning, there is greater overfitting and divergence between the train

Table 1: Final Validation Metrics for Model Variants

Dataset	Contrastive	Frozen	$\log_{10} L_{total}$	$\log_{10} L_{recon}$	$\log_{10} L_{KL}$	$\log_{10} L_{contrast}$
Full	yes	yes	1.34 ± 0.01	1.20 ± 0.01	0.42 ± 0.01	0.56 ± 0.02
Full	**yes**	**no**	$\mathbf{1.01 \pm 0.03}$	$\mathbf{0.73 \pm 0.04}$	$\mathbf{0.27 \pm 0.01}$	$\mathbf{0.48 \pm 0.03}$
Full	no	no	1.22 ± 0.07	1.20 ± 0.08	-0.02 ± 0.02	N/A
ZTF	no	no	1.21 ± 0.07	1.18 ± 0.08	0.02 ± 0.02	N/A
Pan-STARRS	no	no	1.05 ± 0.01	0.99 ± 0.02	0.18 ± 0.01	N/A

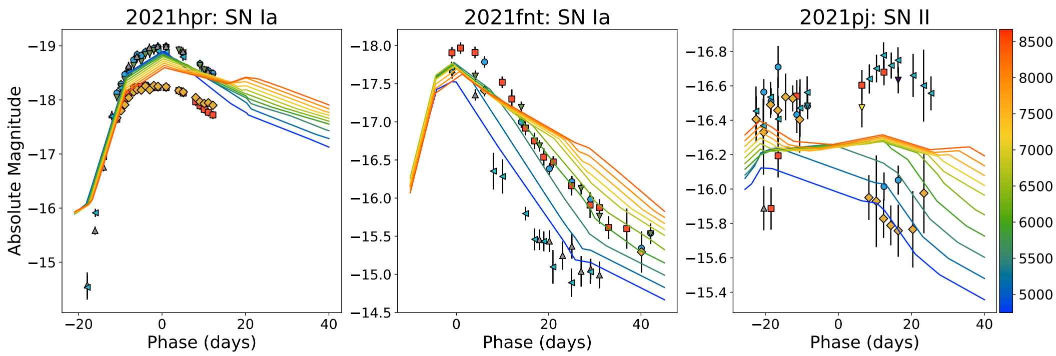

Figure 2: Reconstructions of example YSE light curves for two Type Ia supernovae (left and center), and a Type II (right) supernova using the optimal model with colors corresponding to linearly spaced wavelengths in Angstroms. The observed datapoints are overlaid to highlight the features captured by and missing from the reconstructions.

and validation losses. When trained only on Pan-STARRS data, we see lower losses simply because the YSE Pan-STARRS light curves are sparser than those from ZTF.

Using the optimal model, we encode example photometry from our observed dataset, and reconstruct light curves for an evenly sampled range of wavelengths and a constant filter width of 1300 Å. Figure 2 shows that the decodings correctly recreate the flatter plateaus of SNe II and secondary bumps among longer wavelengths for SNe Ia, but fail to capture the magnitude disparity between high and low wavelengths. A higher-dimensional latent space could potentially remedy this, but for our downstream classification task, reconstruction contrast between SN classes is more important than reconstruction fidelity within classes.

Classification: Finally, we show that shared learning across surveys leads to higher classification accuracy. We use events labeled spectroscopically as SNe Ia, SNe II, SNe IIn, SLSNe-I, and SNe Ib/c to train a random forest (RF) for three-way (Ia, II, Ib/c), four-way (plus IIn), and five-way (plus SLSN-I) classification. Filtering events with missing or miscellaneous labels leaves 7,053 samples (6,693 from ZTF and 360 from Pan-STARRS). The RF takes in eight input features: the four-dimensional latent means and log variances. For this task, we compare performance using encodings from the optimal model, the full dataset model without contrastive learning, and the baseline Pan-STARRS model. For each model, we compare RF performance on Pan-STARRS encodings when we include or exclude ZTF encodings during training.

Table 2: Pan-STARRS F_1-score (Accuracy) Including/Excluding ZTF Encodings

Model	no ZTF			with ZTF		
	3-way	4-way	5-way	3-way	4-way	5-way
Optimal	0.60 (0.77)	0.53 (0.75)	0.42 (0.74)	0.59 (0.77)	0.47 (0.73)	0.54 (0.74)
No Contrastive	0.57 (0.78)	0.48 (0.74)	0.39 (0.74)	0.60 (0.78)	0.51 (0.75)	0.46 (0.76)
Baseline	0.55 (0.73)	0.44 (0.67)	0.33 (0.68)	N/A	N/A	N/A

The resulting classification accuracies and macro-averaged F_1 scores are shown in Table 2. We see significant improvement from baseline for both pre-trained models, suggesting that a latent space pre-trained with a variety of surveys and filters better captures intrinsic differences between SN classes. When not incorporating ZTF encodings, the optimal model has the best performance; the Pan-STARRS encodings are more closely aligned with the ZTF encodings and therefore are better organized in the latent space. When ZTF encodings are included during RF training, the two full dataset models yield similar performance.

We also consider RF performance across ZTF encodings. The resulting Table 5 is in Appendix B; there is negligible change when Pan-STARRS encodings are included in the RF training, and only a slight performance increase from the baseline ZTF model. This is expected, as there are ~ 20 times fewer labeled Pan-STARRS events than ZTF events.

Conclusion

We present a framework for jointly encoding SN events observed across multiple (here, two) surveys. We show that via a combination of (1) a symmetry-informed embedding, (2) pre-training across a large range of wavelengths, and (3) enforcing a shared embedding space with contrastive learning, we can increase accuracy of downstream tasks (e.g., classification) across unique surveys within one learned representation space.

Future work will expand our training sets to a greater number of surveys (e.g., ATLAS; [28]) and present a "production" quality version of this pipeline on Rubin Alert Brokers. We plan to further quantify reconstruction fidelity as a function of the input and reconstructed wavelengths, such that future pre-training can target wavelength regimes of poor reconstruction and reduce the need for extrapolation.

References

[1] Eric C. Bellm, Shrinivas R. Kulkarni, Matthew J. Graham, Richard Dekany, Roger M. Smith, Reed Riddle, Frank J. Masci, George Helou, Thomas A. Prince, Scott M. Adams, C. Barbarino, Tom Barlow, James Bauer, Ron Beck, Justin Belicki, Rahul Biswas, Nadejda Blagorodnova, Dennis Bodewits, Bryce Bolin, Valery Brinnel, Tim Brooke, Brian Bue, Mattia Bulla, Rick Burruss, S. Bradley Cenko, Chan-Kao Chang, Andrew Connolly, Michael Coughlin, John Cromer, Virginia Cunningham, Kishalay De, Alex Delacroix, Vandana Desai, Dmitry A. Duev, Gwendolyn Eadie, Tony L. Farnham, Michael Feeney, Ulrich Feindt, David Flynn, Anna Franckowiak, S. Frederick, C. Fremling, Avishay Gal-Yam, Suvi Gezari, Matteo Giomi, Daniel A. Goldstein, V. Zach Golkhou, Ariel Goobar, Steven Groom, Eugean Hacopians, David Hale, John Henning, Anna Y. Q. Ho, David Hover, Justin Howell, Tiara Hung, Daniela Huppenkothen, David Imel, Wing-Huen Ip, Željko Ivezić, Edward Jackson, Lynne Jones, Mario Juric, Mansi M. Kasliwal, S. Kaspi, Stephen Kaye, Michael S. P. Kelley, Marek Kowalski, Emily Kramer, Thomas Kupfer, Walter Landry, Russ R. Laher, Chien-De Lee, Hsing Wen Lin, Zhong-Yi Lin, Ragnhild Lunnan, Matteo Giomi, Ashish Mahabal, Peter Mao, Adam A. Miller, Serge Monkewitz, Patrick Murphy, Chow-Choong Ngeow, Jakob Nordin, Peter Nugent, Eran Ofek, Maria T. Patterson, Bryan Penprase, Michael Porter, Ludwig Rauch, Umaa Rebbapragada, Dan Reiley, Mickael Rigault, Hector Rodriguez, Jan van Roestel, Ben Rusholme, Jakob van Santen, S. Schulze, David L. Shupe, Leo P. Singer, Maayane T. Soumagnac, Robert Stein, Jason Surace, Jesper Sollerman, Paula Szkody, F. Taddia, Scott Terek, Angela Van Sistine, Sjoert van Velzen, W. Thomas Vestrand, Richard Walters, Charlotte Ward, Quan-Zhi Ye, Po-Chieh Yu, Lin Yan, and Jeffry Zolkower. The Zwicky Transient Facility: System Overview, Performance, and First Results. *Publications of the Astronomical Society of the Pacific*, 131(995):018002, January 2019. doi: 10.1088/1538-3873/aaecbe.

[2] D. O. Jones, R. J. Foley, G. Narayan, J. Hjorth, M. E. Huber, P. D. Aleo, K. D. Alexander, C. R. Angus, K. Auchettl, V. F. Baldassare, S. H. Bruun, K. C. Chambers, D. Chatterjee, D. L. Coppejans, D. A. Coulter, L. DeMarchi, G. Dimitriadis, M. R. Drout, A. Engel, K. D. French, A. Gagliano, C. Gall, T. Hung, L. Izzo, W. V. Jacobson-Galán, C. D. Kilpatrick, H. Korhonen, R. Margutti, S. I. Raimundo, E. Ramirez-Ruiz, A. Rest, C. Rojas-Bravo, M. R. Siebert, S. J. Smartt, K. W. Smith, G. Terreran, Q. Wang, R. Wojtak, A. Agnello, Z. Ansari, N. Arendse, A. Baldeschi, P. K. Blanchard, D. Brethauer, J. S. Bright, J. S. Brown, T. J. L. de Boer, S. A.

Dodd, J. R. Fairlamb, C. Grillo, A. Hajela, C. Hede, A. N. Kolborg, J. A. P. Law-Smith, C. C. Lin, E. A. Magnier, K. Malanchev, D. Matthews, B. Mockler, D. Muthukrishna, Y. C. Pan, H. Pfister, D. K. Ramanah, S. Rest, A. Sarangi, S. L. Schrøder, C. Stauffer, M. C. Stroh, K. L. Taggart, S. Tinyanont, R. J. Wainscoat, and Young Supernova Experiment. The Young Supernova Experiment: Survey Goals, Overview, and Operations. *The Astrophysical Journal*, 908(2):143, February 2021. doi: 10.3847/1538-4357/abd7f5.

[3] Alexei V Filippenko. Optical spectra of supernovae. *Annual Review of Astronomy and Astrophysics*, 35(1):309–355, 1997.

[4] Griffin Hosseinzadeh, Frederick Dauphin, V. Ashley Villar, Edo Berger, David O. Jones, Peter Challis, Ryan Chornock, Maria R. Drout, Ryan J. Foley, Robert P. Kirshner, Ragnhild Lunnan, Raffaella Margutti, Dan Milisavljevic, Yen-Chen Pan, Armin Rest, Daniel M. Scolnic, Eugene Magnier, Nigel Metcalfe, Richard Wainscoat, and Christopher Waters. Photometric Classification of 2315 Pan-STARRS1 Supernovae with Superphot. *The Astrophysical Journal*, 905:93, December 2020. ISSN 0004-637X. doi: 10.3847/1538-4357/abc42b. URL `https://ui.adsabs.harvard.edu/abs/2020ApJ...905...93H`. Publisher: IOP ADS Bibcode: 2020ApJ...905...93H.

[5] V Ashley Villar, Griffin Hosseinzadeh, Edo Berger, Michelle Ntampaka, David O Jones, Peter Challis, Ryan Chornock, Maria R Drout, Ryan J Foley, Robert P Kirshner, et al. Superraenn: a semisupervised supernova photometric classification pipeline trained on pan-starrs1 medium-deep survey supernovae. *The Astrophysical Journal*, 905(2):94, 2020.

[6] Kyle Boone. Parsnip: Generative models of transient light curves with physics-enabled deep learning. *The Astronomical Journal*, 162(6):275, 2021.

[7] P Sánchez-Sáez, I Reyes, C Valenzuela, F Förster, S Eyheramendy, F Elorrieta, FE Bauer, G Cabrera-Vives, PA Estévez, M Catelan, et al. Alert classification for the alerce broker system: The light curve classifier. *The Astronomical Journal*, 161(3):141, 2021.

[8] Kaylee M. de Soto, V. Ashley Villar, Edo Berger, Sebastian Gomez, Griffin Hosseinzadeh, Doug Branton, Sandro Campos, Melissa DeLucchi, Jeremy Kubica, Olivia Lynn, Konstantin Malanchev, and Alex I. Malz. Superphot+: Real-time Fitting and Classification of Supernova Light Curves. *The Astrophysical Journal*, 974(2):169, October 2024. doi: 10.3847/1538-4357/ad6a4f.

[9] Željko Ivezić, Steven M. Kahn, J. Anthony Tyson, Bob Abel, Emily Acosta, Robyn Allsman, David Alonso, Yusra AlSayyad, Scott F. Anderson, John Andrew, James Roger P. Angel, George Z. Angeli, Reza Ansari, Pierre Antilogus, Constanza Araujo, Robert Armstrong, Kirk T. Arndt, Pierre Astier, Éric Aubourg, Nicole Auza, Tim S. Axelrod, Deborah J. Bard, Jeff D. Barr, Aurelian Barrau, James G. Bartlett, Amanda E. Bauer, Brian J. Bauman, Sylvain Baumont, Ellen Bechtol, Keith Bechtol, Andrew C. Becker, Jacek Becla, Cristina Beldica, Steve Bellavia, Federica B. Bianco, Rahul Biswas, Guillaume Blanc, Jonathan Blazek, Roger D. Blandford, Josh S. Bloom, Joanne Bogart, Tim W. Bond, Michael T. Booth, Anders W. Borgland, Kirk Borne, James F. Bosch, Dominique Boutigny, Craig A. Brackett, Andrew Bradshaw, William Nielsen Brandt, Michael E. Brown, James S. Bullock, Patricia Burchat, David L. Burke, Gianpietro Cagnoli, Daniel Calabrese, Shawn Callahan, Alice L. Callen, Jeffrey L. Carlin, Erin L. Carlson, Srinivasan Chandrasekharan, Glenaver Charles-Emerson, Steve Chesley, Elliott C. Cheu, Hsin-Fang Chiang, James Chiang, Carol Chirino, Derek Chow, David R. Ciardi, Charles F. Claver, Johann Cohen-Tanugi, Joseph J. Cockrum, Rebecca Coles, Andrew J. Connolly, Kem H. Cook, Asantha Cooray, Kevin R. Covey, Chris Cribbs, Wei Cui, Roc Cutri, Philip N. Daly, Scott F. Daniel, Felipe Daruich, Guillaume Daubard, Greg Daues, William Dawson, Francisco Delgado, Alfred Dellapenna, Robert de Peyster, Miguel de Val-Borro, Seth W. Digel, Peter Doherty, Richard Dubois, Gregory P. Dubois-Felsmann, Josef Durech, Frossie Economou, Tim Eifler, Michael Eracleous, Benjamin L. Emmons, Angelo Fausti Neto, Henry Ferguson, Enrique Figueroa, Merlin Fisher-Levine, Warren Focke, Michael D. Foss, James Frank, Michael D. Freemon, Emmanuel Gangler, Eric Gawiser, John C. Geary, Perry Gee, Marla Geha, Charles J. B. Gessner, Robert R. Gibson, D. Kirk Gilmore, Thomas Glanzman, William Glick, Tatiana Goldina, Daniel A. Goldstein, Iain Goodenow, Melissa L. Graham, William J. Gressler, Philippe Gris, Leanne P. Guy, Augustin Guyonnet, Gunther Haller, Ron Harris, Patrick A. Hascall,

Justine Haupt, Fabio Hernandez, Sven Herrmann, Edward Hileman, Joshua Hoblitt, John A. Hodgson, Craig Hogan, James D. Howard, Dajun Huang, Michael E. Huffer, Patrick Ingraham, Walter R. Innes, Suzanne H. Jacoby, Bhuvnesh Jain, Fabrice Jammes, M. James Jee, Tim Jenness, Garrett Jernigan, Darko Jevremović, Kenneth Johns, Anthony S. Johnson, Margaret W. G. Johnson, R. Lynne Jones, Claire Juramy-Gilles, Mario Jurić, Jason S. Kalirai, Nitya J. Kallivayalil, Bryce Kalmbach, Jeffrey P. Kantor, Pierre Karst, Mansi M. Kasliwal, Heather Kelly, Richard Kessler, Veronica Kinnison, David Kirkby, Lloyd Knox, Ivan V. Kotov, Victor L. Krabbendam, K. Simon Krughoff, Petr Kubánek, John Kuczewski, Shri Kulkarni, John Ku, Nadine R. Kurita, Craig S. Lage, Ron Lambert, Travis Lange, J. Brian Langton, Laurent Le Guillou, Deborah Levine, Ming Liang, Kian-Tat Lim, Chris J. Lintott, Kevin E. Long, Margaux Lopez, Paul J. Lotz, Robert H. Lupton, Nate B. Lust, Lauren A. MacArthur, Ashish Mahabal, Rachel Mandelbaum, Thomas W. Markiewicz, Darren S. Marsh, Philip J. Marshall, Stuart Marshall, Morgan May, Robert McKercher, Michelle McQueen, Joshua Meyers, Myriam Migliore, Michelle Miller, David J. Mills, Connor Miraval, Joachim Moeyens, Fred E. Moolekamp, David G. Monet, Marc Moniez, Serge Monkewitz, Christopher Montgomery, Christopher B. Morrison, Fritz Mueller, Gary P. Muller, Freddy Muñoz Arancibia, Douglas R. Neill, Scott P. Newbry, Jean-Yves Nief, Andrei Nomerotski, Martin Nordby, Paul O'Connor, John Oliver, Scot S. Olivier, Knut Olsen, William O'Mullane, Sandra Ortiz, Shawn Osier, Russell E. Owen, Reynald Pain, Paul E. Palecek, John K. Parejko, James B. Parsons, Nathan M. Pease, J. Matt Peterson, John R. Peterson, Donald L. Petravick, M. E. Libby Petrick, Cathy E. Petry, Francesco Pierfederici, Stephen Pietrowicz, Rob Pike, Philip A. Pinto, Raymond Plante, Stephen Plate, Joel P. Plutchak, Paul A. Price, Michael Prouza, Veljko Radeka, Jayadev Rajagopal, Andrew P. Rasmussen, Nicolas Regnault, Kevin A. Reil, David J. Reiss, Michael A. Reuter, Stephen T. Ridgway, Vincent J. Riot, Steve Ritz, Sean Robinson, William Roby, Aaron Roodman, Wayne Rosing, Cecille Roucelle, Matthew R. Rumore, Stefano Russo, Abhijit Saha, Benoit Sassolas, Terry L. Schalk, Pim Schellart, Rafe H. Schindler, Samuel Schmidt, Donald P. Schneider, Michael D. Schneider, William Schoening, German Schumacher, Megan E. Schwamb, Jacques Sebag, Brian Selvy, Glenn H. Sembroski, Lynn G. Seppala, Andrew Serio, Eduardo Serrano, Richard A. Shaw, Ian Shipsey, Jonathan Sick, Nicole Silvestri, Colin T. Slater, J. Allyn Smith, R. Chris Smith, Shahram Sobhani, Christine Soldahl, Lisa Storrie-Lombardi, Edward Stover, Michael A. Strauss, Rachel A. Street, Christopher W. Stubbs, Ian S. Sullivan, Donald Sweeney, John D. Swinbank, Alexander Szalay, Peter Takacs, Stephen A. Tether, Jon J. Thaler, John Gregg Thayer, Sandrine Thomas, Adam J. Thornton, Vaikunth Thukral, Jeffrey Tice, David E. Trilling, Max Turri, Richard Van Berg, Daniel Vanden Berk, Kurt Vetter, Francoise Virieux, Tomislav Vucina, William Wahl, Lucianne Walkowicz, Brian Walsh, Christopher W. Walter, Daniel L. Wang, Shin-Yawn Wang, Michael Warner, Oliver Wiecha, Beth Willman, Scott E. Winters, David Wittman, Sidney C. Wolff, W. Michael Wood-Vasey, Xiuqin Wu, Bo Xin, Peter Yoachim, and Hu Zhan. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *The Astrophysical Journal*, 873(2):111, March 2019. doi: 10.3847/1538-4357/ab042c.

[10] J. Anthony Tyson. Large Synoptic Survey Telescope: Overview. In J. Anthony Tyson and Sidney Wolff, editors, *Survey and Other Telescope Technologies and Discoveries*, volume 4836 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 10–20, December 2002. doi: 10.1117/12.456772.

[11] Melissa L. Graham, Andrew J. Connolly, Željko Ivezić, Samuel J. Schmidt, R. Lynne Jones, Mario Jurić, Scott F. Daniel, and Peter Yoachim. Photometric Redshifts with the LSST: Evaluating Survey Observing Strategies. *The Astronomical Journal*, 155(1):1, January 2018. doi: 10.3847/1538-3881/aa99d4.

[12] V. Ashley Villar, Griffin Hosseinzadeh, Edo Berger, Michelle Ntampaka, David O. Jones, Peter Challis, Ryan Chornock, Maria R. Drout, Ryan J. Foley, Robert P. Kirshner, Ragnhild Lunnan, Raffaella Margutti, Dan Milisavljevic, Nathan Sanders, Yen-Chen Pan, Armin Rest, Daniel M. Scolnic, Eugene Magnier, Nigel Metcalfe, Richard Wainscoat, and Christopher Waters. SuperRAENN: A Semisupervised Supernova Photometric Classification Pipeline Trained on Pan-STARRS1 Medium-Deep Survey Supernovae. *The Astrophysical Journal*, 905 (2):94, December 2020. doi: 10.3847/1538-4357/abc6fd.

[13] Kyle Boone. ParSNIP: Generative Models of Transient Light Curves with Physics-enabled Deep Learning. *The Astronomical Journal*, 162(6):275, December 2021. doi: 10.3847/1538-3881/ac2a2d.

[14] George Stein, Uroš Seljak, Vanessa Böhm, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, S. Bongard, K. Boone, C. Buton, Y. Copin, S. Dixon, D. Fouchez, E. Gangler, R. Gupta, B. Hayden, W. Hillebrandt, M. Karmen, A. G. Kim, M. Kowalski, D. Küsters, P. F. Léget, F. Mondon, J. Nordin, R. Pain, E. Pecontal, R. Pereira, S. Perlmutter, K. A. Ponder, D. Rabinowitz, M. Rigault, D. Rubin, K. Runge, C. Saunders, G. Smadja, N. Suzuki, C. Tao, S. Taubenberger, R. C. Thomas, M. Vincenzi, and Nearby Supernova Factory. A Probabilistic Autoencoder for Type Ia Supernova Spectral Time Series. *The Astrophysical Journal*, 935(1):5, August 2022. doi: 10.3847/1538-4357/ac7c08.

[15] James Guillochon, Matt Nicholl, V. Ashley Villar, Brenna Mockler, Gautham Narayan, Kaisey S. Mandel, Edo Berger, and Peter K. G. Williams. MOSFiT: Modular Open Source Fitter for Transients. *The Astrophysical Journal Supplement Series*, 236(1):6, May 2018. doi: 10.3847/1538-4365/aab761.

[16] K. C. Chambers, E. A. Magnier, N. Metcalfe, H. A. Flewelling, M. E. Huber, C. Z. Waters, L. Denneau, P. W. Draper, D. Farrow, D. P. Finkbeiner, C. Holmberg, J. Koppenhoefer, P. A. Price, A. Rest, R. P. Saglia, E. F. Schlafly, S. J. Smartt, W. Sweeney, R. J. Wainscoat, W. S. Burgett, S. Chastel, T. Grav, J. N. Heasley, K. W. Hodapp, R. Jedicke, N. Kaiser, R. P. Kudritzki, G. A. Luppino, R. H. Lupton, D. G. Monet, J. S. Morgan, P. M. Onaka, B. Shiao, C. W. Stubbs, J. L. Tonry, R. White, E. Bañados, E. F. Bell, R. Bender, E. J. Bernard, M. Boegner, F. Boffi, M. T. Botticella, A. Calamida, S. Casertano, W. P. Chen, X. Chen, S. Cole, N. Deacon, C. Frenk, A. Fitzsimmons, S. Gezari, V. Gibbs, C. Goessl, T. Goggia, R. Gourgue, B. Goldman, P. Grant, E. K. Grebel, N. C. Hambly, G. Hasinger, A. F. Heavens, T. M. Heckman, R. Henderson, T. Henning, M. Holman, U. Hopp, W. H. Ip, S. Isani, M. Jackson, C. D. Keyes, A. M. Koekemoer, R. Kotak, D. Le, D. Liska, K. S. Long, J. R. Lucey, M. Liu, N. F. Martin, G. Masci, B. McLean, E. Mindel, P. Misra, E. Morganson, D. N. A. Murphy, A. Obaika, G. Narayan, M. A. Nieto-Santisteban, P. Norberg, J. A. Peacock, E. A. Pier, M. Postman, N. Primak, C. Rae, A. Rai, A. Riess, A. Riffeser, H. W. Rix, S. Röser, R. Russel, L. Rutz, E. Schilbach, A. S. B. Schultz, D. Scolnic, L. Strolger, A. Szalay, S. Seitz, E. Small, K. W. Smith, D. R. Soderblom, P. Taylor, R. Thomson, A. N. Taylor, A. R. Thakar, J. Thiel, D. Thilker, D. Unger, Y. Urata, J. Valenti, J. Wagner, T. Walder, F. Walter, S. P. Watters, S. Werner, W. M. Wood-Vasey, and R. Wyse. The Pan-STARRS1 Surveys. *arXiv e-prints*, art. arXiv:1612.05560, December 2016. doi: 10.48550/arXiv.1612.05560.

[17] N. Gehrels, G. Chincarini, P. Giommi, K. O. Mason, J. A. Nousek, A. A. Wells, N. E. White, S. D. Barthelmy, D. N. Burrows, L. R. Cominsky, K. C. Hurley, F. E. Marshall, P. Mészáros, P. W. A. Roming, L. Angelini, L. M. Barbier, T. Belloni, S. Campana, P. A. Caraveo, M. M. Chester, O. Citterio, T. L. Cline, M. S. Cropper, J. R. Cummings, A. J. Dean, E. D. Feigelson, E. E. Fenimore, D. A. Frail, A. S. Fruchter, G. P. Garmire, K. Gendreau, G. Ghisellini, J. Greiner, J. E. Hill, S. D. Hunsberger, H. A. Krimm, S. R. Kulkarni, P. Kumar, F. Lebrun, N. M. Lloyd-Ronning, C. B. Markwardt, B. J. Mattson, R. F. Mushotzky, J. P. Norris, J. Osborne, B. Paczynski, D. M. Palmer, H. S. Park, A. M. Parsons, J. Paul, M. J. Rees, C. S. Reynolds, J. E. Rhoads, T. P. Sasseen, B. E. Schaefer, A. T. Short, A. P. Smale, I. A. Smith, L. Stella, G. Tagliaferri, T. Takahashi, M. Tashiro, L. K. Townsley, J. Tueller, M. J. L. Turner, M. Vietri, W. Voges, M. J. Ward, R. Willingale, F. M. Zerbi, and W. W. Zhang. The Swift Gamma-Ray Burst Mission. *The Astrophysical Journal*, 611(2):1005–1020, August 2004. doi: 10.1086/422091.

[18] M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard, T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. Van Dyk, and S. Wheelock. The Two Micron All Sky Survey (2MASS). *The Astronomical Journal*, 131 (2):1163–1183, February 2006. doi: 10.1086/498708.

[19] A. Gal-Yam. The TNS alert system. In *American Astronomical Society Meeting Abstracts*, volume 237 of *American Astronomical Society Meeting Abstracts*, page 423.05, January 2021.

[20] P. Sánchez-Sáez, I. Reyes, C. Valenzuela, F. Förster, S. Eyheramendy, F. Elorrieta, F. E. Bauer, G. Cabrera-Vives, P. A. Estévez, M. Catelan, G. Pignata, P. Huijse, D. De Cicco, P. Arévalo, R. Carrasco-Davis, J. Abril, R. Kurtev, J. Borissova, J. Arredondo, E. Castillo-Navarrete, D. Rodriguez, D. Ruz-Mieres, A. Moya, L. Sabatini-Gacitúa, C. Sepúlveda-Cobo,

and E. Camacho-Iñiguez. Alert Classification for the ALeRCE Broker System: The Light Curve Classifier. *The Astronomical Journal*, 161(3):141, March 2021. doi: 10.3847/1538-3881/abd5c1.

[21] P. D. Aleo, K. Malanchev, S. Sharief, D. O. Jones, G. Narayan, R. J. Foley, V. A. Villar, C. R. Angus, V. F. Baldassare, M. J. Bustamante-Rosell, D. Chatterjee, C. Cold, D. A. Coulter, K. W. Davis, S. Dhawan, M. R. Drout, A. Engel, K. D. French, A. Gagliano, C. Gall, J. Hjorth, M. E. Huber, W. V. Jacobson-Galán, C. D. Kilpatrick, D. Langeroodi, P. Macias, K. S. Mandel, R. Margutti, F. Matasić, P. McGill, J. D. R. Pierel, E. Ramirez-Ruiz, C. L. Ransome, C. Rojas-Bravo, M. R. Siebert, K. W. Smith, K. M. de Soto, M. C. Stroh, S. Tinyanont, K. Taggart, S. M. Ward, R. Wojtak, K. Auchettl, P. K. Blanchard, T. J. L. de Boer, B. M. Boyd, C. M. Carroll, K. C. Chambers, L. DeMarchi, G. Dimitriadis, S. A. Dodd, N. Earl, D. Farias, H. Gao, S. Gomez, M. Grayling, C. Grillo, E. E. Hayes, T. Hung, L. Izzo, N. Khetan, A. N. Kolborg, J. A. P. Law-Smith, N. LeBaron, C. C. Lin, Y. Luo, E. A. Magnier, D. Matthews, B. Mockler, A. J. G. O'Grady, Y. C. Pan, C. A. Politsch, S. I. Raimundo, A. Rest, R. Ridden-Harper, A. Sarangi, S. L. Schrøder, S. J. Smartt, G. Terreran, S. Thorp, J. Vazquez, R. J. Wainscoat, Q. Wang, A. R. Wasserman, S. K. Yadavalli, R. Yarza, Y. Zenati, and Young Supernova Experiment. The Young Supernova Experiment Data Release 1 (YSE DR1): Light Curves and Photometric Classification of 1975 Supernovae. *The Astrophysical Journal Supplement Series*, 266(1):9, May 2023. doi: 10.3847/1538-4365/acbfba.

[22] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper_files/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf`.

[23] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *CoRR*, abs/1312.6114, 2014.

[24] Mohammad Lotfollahi, Mohsen Naghipourfar, Fabian J Theis, and F Alexander Wolf. Conditional out-of-distribution generation for unpaired data using transfer vae. *Bioinformatics*, 36 (Supplement_2):i610–i617, 2020.

[25] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv e-prints*, art. arXiv:1312.6114, December 2013. doi: 10.48550/arXiv.1312.6114.

[26] Kilian Q Weinberger, John Blitzer, and Lawrence Saul. Distance metric learning for large margin nearest neighbor classification. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems*, volume 18. MIT Press, 2005. URL `https://proceedings.neurips.cc/paper_files/paper/2005/file/a7f592cef8b130a6967a90617db5681b-Paper.pdf`.

[27] Jason Yosinski, Jeff Clune, Y. Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems (NIPS)*, 27, 11 2014.

[28] J. L. Tonry, L. Denneau, A. N. Heinze, B. Stalder, K. W. Smith, S. J. Smartt, C. W. Stubbs, H. J. Weiland, and A. Rest. ATLAS: A High-cadence All-sky Survey System. *Publications of the Astronomical Society of the Pacific*, 130(988):064505, June 2018. doi: 10.1088/1538-3873/aabadf.

Table 3: MOSFIT Model Constraints

Parameter	Value
kappa	0.1
kappagamma	1000
temperature	3000
codeltatime	0.001
doeltalambda	0.1
redshift	0.02

Table 4: MOSFIT Survey Properties

Survey	Bands	Mean σ_M	Cadence (d)
ZTF	g, r	0.2	2.0
Pan-STARRS	g, r, i, z	0.12	3.0
LSST	u, g, r, i, z, Y	0.1	4.0
SWIFT	B, UVM2, UVW1, UVW2, U, V	0.14	5.0
2MASS	H, J, Ks	0.4	6.0

A MOSFIT Details

To simulate SN photometry, we use the "default" MOSFIT model [15] with parameters set according to Table 3. Observed photometry is generated for the survey bands and limiting magnitudes detailed in Table 4. Magnitude uncertainties are drawn from Gaussians with means from the table and standard deviation equal to the means divided by five.

B Supplemental Table

Here we show classification metrics for a random forest trained on ZTF and, optionally, Pan-STARRS encodings. Including Pan-STARRS encodings during training negligibly affects RF performance.

Table 5: ZTF F_1-score (Accuracy) Including/Excluding Pan-STARRS Encodings

Model	no PS1			with PS1		
	3-way	4-way	5-way	3-way	4-way	5-way
Optimal	0.70 (0.90)	0.67 (0.88)	0.63 (0.87)	0.71 (0.91)	0.67 (0.89)	0.64 (0.88)
No Contrastive	0.70 (0.91)	0.67 (0.88)	0.64 (0.87)	0.70 (0.90)	0.67 (0.88)	0.65 (0.87)
Baseline	0.68 (0.90)	0.64 (0.87)	0.63 (0.87)	N/A	N/A	N/A