Maven: A Multimodal Foundation Model for Supernova Science

Gemma Zhang *
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
Department of Physics, Harvard University, Cambridge, MA 02138, USA

Thomas Helfer *
Institute for Advanced Computational Science, Stony Brook University
Stony Brook, NY 11794 USA

Alexander T. Gagliano
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, MS-16, Cambridge, MA 02138, USA

Siddharth Mishra-Sharma †
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
Department of Physics, Harvard University, Cambridge, MA 02138, USA

V. Ashley Villar
Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, MS-16, Cambridge, MA 02138, USA
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions

Abstract

A common setting in astronomy is the availability of a small number of high-quality observations, and larger amounts of either lower-quality observations or synthetic data from simplified models. Time-domain astrophysics is a canonical example of this imbalance, with the number of supernovae observed photometrically outpacing the number observed spectroscopically by multiple orders of magnitude. At the same time, no data-driven models exist to understand these photometric and spectroscopic observables in a common context. Contrastive learning objectives, which have grown in popularity for aligning distinct data modalities in a shared embedding space, provide a potential solution to extract information from these modalities. We present Maven, the first foundation model for supernova science. To construct Maven, we first pre-train our model to align photometry and spectroscopy from 0.5M synthetic supernovae using a contrastive objective. We then fine-tune the model on 4,702 observed supernovae from the Zwicky Transient Facility. Maven reaches state-of-the-art performance on both classification and redshift estimation, despite the embeddings not being explicitly optimized for these tasks. Through ablation studies, we show that pre-training with synthetic data improves overall performance. In the upcoming era of the Vera C. Rubin Observatory, Maven serves

*Equal contribution.
†Currently at Anthropic; work performed while at MIT/IAIFI.

Foundation Models for Science Workshop,38th Conference on Neural Information Processing Systems (NeurIPS 2024).

as a Rosetta Stone for leveraging large, unlabeled and multimodal time-domain datasets.

1 Introduction

The discovery rate of supernovae (SNe) has grown exponentially over the past four decades, thanks in large part to wide-field, untargeted optical surveys (e.g., All Sky Automated Survey for SuperNovae (ASAS-SN; [1]), ATLAS ([2]), the Zwicky Transient Facility (ZTF; [3]) and the Young Supernova Experiment (YSE; [4]). Today, well over ten-thousand SNe are discovered annually. The upcoming Legacy Survey of Space and Time (LSST; [5]), conducted by the Vera C. Rubin Observatory, will enable the photometric discovery of over one million SNe annually, in addition to millions of other non-SN variable phenomena. While photometry is easily obtained, spectroscopy is significantly more time-consuming to acquire (long integration times are needed to build up sufficient signal across a spectrograph). This challenge has catalyzed research into techniques to infer the underlying physics of an explosion directly from photometric observations, including the classification of SN types [e.g., 6–12] and inference of SN redshifts [13, 14]. In this context, supervised machine learning has dominated the training of models for the classification of SN types and the estimation of SN redshift. The labels used in the supervised training scenario must be first extracted from spectra, demanding large spectroscopic datasets for sufficient model performance. To overcome this issue, researchers have begun to explore self-supervised learning to leverage the structure of unlabeled photometric datasets, by training a feature extraction network and generating a low-dimensional latent space [15, 16]. The learned latent space can then be used to classify events using supervised methods.

Self-supervised representation learning for time-domain astrophysics is appealing for multiple reasons. Pre-trained models have been shown to produce latent data representations that are more robust against distribution shifts than their supervised counterparts [17, 18]. Distribution shift is a common obstacle when applying models trained on bright, spectroscopically-confirmed low-redshift transients to fainter, more distant phenomena that are underrepresented in the training data. Self-supervised learning may also be less sensitive to the class imbalances observed in transient science [19]: labeled SN samples are dominated by type Ia SNe due to their high luminosities relative to other classes. The generalizability of learned representations [20, 21] also offers the potential for using a pre-trained model for multiple inference tasks and across diverse time-domain surveys, with only minimal fine-tuning.

Contrastive learning has emerged as an effective pre-training objective for combining multiple data modalities. Radford et al. [22] present an embedding scheme called Contrastive Language–Image Pre-training (CLIP) for aligning natural language and associated images in a shared latent space. Inspired by CLIP, we present Maven, the first multimodal foundation model for SNe. In contrast to previous models for SN classification and redshift inference, our model is constructed using spectroscopic and photometric information simultaneously. Motivated by previous work in synthetic pre-training, we first train Maven by aligning simulated light curve-spectrum pairs via contrastive learning, and fine-tune it on a small sample of observed data using the same approach. Our final model achieves state-of-the-art performance on multiple downstream tasks. We also train a model with only observed data, called Maven-lite, to quantify the impact of synthetic pre-training. Though we limit our analysis to classification and redshift (two popular inference tasks in SN science), the model is a milestone toward general-purpose training for a range of downstream tasks.

2 Datasets and Simulations

In this study, we utilize two datasets: a simulated dataset for pre-training and a dataset of observations for subsequent fine-tuning and validation[3]. We give an overview of the datasets and provide more details in Appendix A.1.

For pre-training, we simulate observations of the Zwicky Transient Facility [3] using the SNANA simulation code [23] and the framework described in [24], which approximately matches the redshift distribution of the SNe in our observed sample (described in A.1.2). We simulate 500,000 total events evenly split between five different SN classes, using SED models from the Photometric LSST

[3]All data are available at `https://huggingface.co/datasets/anonymous/anonymous`

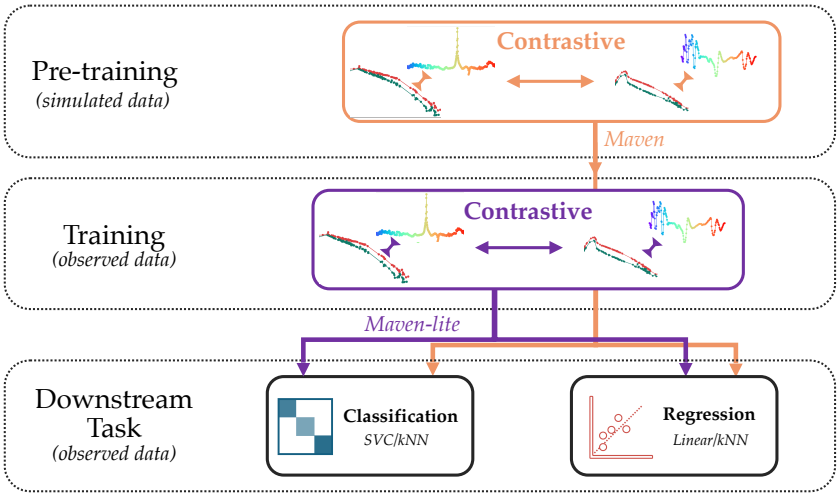

Figure 1: **Overview of our training workflows**. We first pre-train on a large simulated data set using contrastive methods (using light curves and spectra). We follow up by training on the ZTF observations and then use a simple model to translate these embedding to downstream tasks. Different colors indicate different first training steps and their arrows indicate subsequent training steps.

Astronomical Time-Series Classification Challenge [25]: SNe Ia, SNe Ib/c, SLSNe-I, and SNe II (which includes both SNe IIP/IIL), and SNe IIn.

For our observation dataset, we obtain metadata for 4,702 spectroscopically-classified SNe from the ZTF Bright Transient Survey [26].We consolidate our resulting sample to only include events spectroscopically classified as "normal": SN Ia, SN Ib/c, SN II, SLSN-I, and SN IIn. In each training iteration, we augment our training data by applying Gaussian noise to the photometric and spectroscopic observations with mean zero and standard deviation equal to the magnitude of the reported observational errors.

3 Methodology

Here, our goal is to use contrastive learning to build a shared representation space using photometric and spectroscopic data from the same event, and to explore the predictive properties of these representations for downstream tasks.

3.1 Modality Encoders

The encoders $f : I \to \mathbb{R}^{d_{\text{emb}}}$ and $g : T \to \mathbb{R}^{d_{\text{emb}}}$ are designed to efficiently extract information from high-dimensional data for the two considered modalities. Both light curve and spectrum encoders are based on the transformer architecture [27].

The light curve encoder processes magnitude-time pairs $X = ((m_1, t_1), ..., (m_n, t_n))$, where t_i is defined as the number of days from the first observation. The normalized magnitudes are initially linearly projected to the d_{model}-dimensional embedding space of the transformer. Each transformer layer applies multi-head self-attention (with n_{heads} heads acting separately) followed by a 2-layer feedforward network: $\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2$. Layer normalization and residual connections are applied after attention as well as the feedforward layer. To account for the temporal nature of light curves, we use sinusoidal time encodings to project the times t_i to a higher-dimensional space,

$$\text{TE}(t_i, j) = \begin{cases} \sin(t_i / n_t^{2j/d_{\text{model}}}) & \text{if } i \text{ is even} \\ \cos(t_i / n_t^{2j/d_{\text{model}}}) & \text{if } i \text{ is odd} \end{cases}, \tag{1}$$

where j is the time embedding index, t_i are the input times, and n_t is a hyperparameter governing the periodicity of the time encodings. This encoding allows the model to capture both absolute and

relative timing of observations across a wide range of timescales. In addition, to incorporate light curve measurements from multiple photometric filters, we concatenate all measurements for each SN and add an additional band encoding. Different bands are one-hot encoded with integers and then added to light curve magnitude embeddings before being passed into the transformer encoder.

The spectrum encoder utilizes a similar transformer-based architecture to that of the light curve encoder, but interprets the input sequence as $((f_1, \lambda_1), ..., (f_n, \lambda_n))$, where f_i represents the flux at observer-frame wavelength λ_i. The positional encoding for wavelengths follows the same sinusoidal pattern as the light curve encoder, but with λ in place of t.

For both the light curve and spectrum encoders, in addition to deterministic aggregate e.g., mean or max pooling, we consider attention-based learnable aggregation to convert the per-sequence representation to a 1-D representation vector. This enables the model to learn a data-dependent aggregation scheme, potentially better capturing correlations in the data. We initialize a learnable query vector $Q_{\text{learned}} \in \mathbb{R}^{d_{\text{emb}}}$, where d_{emb} is the embedding dimension. A projection of the encoded sequence after the final transformer layer gives the keys and values for the attention mechanism. We use a multi-head attention architecture with two heads to then get $x_{\text{agg}} = \text{Attention}(Q_{\text{learned}}, K_{\text{final}}, V_{\text{final}}) \in \mathbb{R}^{d_{\text{emb}}}$. We treat the aggregation method as a hyperparameter: in the hyperparameter tuning process, we consider both mean and attention-based aggregation.

3.2 Training

After pre-training some of our models on simulations, we fine-tune on the small set of ZTF BTS observations. We explore two different transfer learning approaches. First, we begin with the pre-trained model and continue training *all* of its weights using the observed data. In the second approach, we again begin with the pre-trained model, but instead allow only the weights in the first transformer block to be learnable and freeze all other weights during fine-tuning. We find that the first approach leads to better performance on downstream tasks compared to the second approach. Therefore, we only show results from the first approach hereafter. We define our hyperparameter-optimized pre-trained model as 'Maven', and our CLIP model without pre-training as 'Maven-lite' (see Fig. 1).

For both pre-training and fine-tuning, we use the standard softmax-based bidirectional variant of the InfoNCE [28] contrastive loss function. Given a minibatch $\mathcal{B}$ of $|\mathcal{B}|$ associated pairs $\{(X_i, Y_i)\}_{i=1}^{|\mathcal{B}|}$ (light curves and spectra in this work), our goal is to align the learned representations of corresponding (positive) pairs (X_i, Y_i) (here, the spectrum and light curve of a single SN) while repelling the representations of unaligned (negative) pairs $(X_i, Y_{j\neq i})$:

$$\mathcal{L}(\mathcal{B}) = -\frac{1}{2|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \left(\log \frac{e^{x_i \cdot y_i/\tau}}{\sum_{j=1}^{|\mathcal{B}|} e^{x_i \cdot y_j/\tau}} + \log \frac{e^{x_i \cdot y_i/\tau}}{\sum_{j=1}^{|\mathcal{B}|} e^{x_j \cdot y_i/\tau}} \right) \tag{2}$$

where $x_i = f(X_i)/\|f(X_i)\|$ and $y_i = g(Y_i)/\|g(Y_i)\|$ are the normalized representations of the i-th data pairs associated with each other, and τ is a learnable hyperparameter.

We perform a stratified 5-fold cross-validation on the ZTF observations to quantify model uncertainties. We show results for the mean and standard deviation from these runs. To avoid added computational overhead, we do not perform it on the simulation-based pre-training.

To determine hyperparameter values for model architecture and training, we perform a hyperparameter search for our models using Weights & Biases [29]. A list of parameter values in our search are provided in configuration files in our public code repository.[4]

3.3 Downstream Tasks

We evaluate the performance of Maven and Maven-lite on two primary downstream tasks: classification and regression.

Classification of SNe from photometry *alone* has been an area of active study due to the long integration times required to build up sufficient signal-to-noise with spectroscopy and the subsequent rise of wide-field photometric surveys. SN classes are highly imbalanced in observed samples, due to

[4] `https://github.com/anonymous/anonymous`

a combination of different intrinsic volumetric rates and a steep selection function toward brighter classes (SNe Ia). We present results for a three-way classification task (SN Ia, SN II, SN Ib/c).

In addition, we attempt to predict the redshift of each SN (which we call our "regression task"). Redshift estimation using spectroscopic and photometric SNe Ia is a fundamental tool for cosmological analyses. Although non-Ia classes are significantly more observationally diverse [e.g., 30], estimating SN redshift remains critical for estimating the intrinsic properties of an explosion.

To transform our contrastive-trained light curve embeddings into classification predictions, we explore both support vector classification (SVC) and k-Nearest Neighbors classification (kNN). For redshift regression, we explore both linear regression and kNN regression. In the following sections, we only quote results from kNN as it produces the best performance on downstream tasks.

Lastly, we train supervised baseline models directly on the observational ZTF dataset. For the classification baseline model, we optimize for the multi-class cross-entropy loss and take the class with highest pseudo-probability score as the prediction for each event. The regression baseline model outputs a single value and is optimized using the mean squared error (MSE) loss.

4 Results

4.1 t-SNE Visualization of Latent Spaces

To explore the impact of CLIP-style pre-training on the latent space of our Maven models, we first visualize a sample of embedded light curves. We compute Maven and Maven-lite embeddings of our five dominant classes for both the synthetic and observed samples: SNe Ia, SNe II, SNe Ib/c, SLSNe I, and SNe IIn. We further reduce the dimensionality of our latent space using principal component analysis from the encoder output of 128 features to 50 features for computational efficiency, confirming that the subsequent 50 features retain $>$99.999% of the variance in the original embeddings. Finally, we produce two-dimensional representations of these embeddings using the t-distributed stochastic neighbor embedding tool (t-SNE; [31]). Our results are presented in Fig. 2 for Maven-lite (left column) and Maven (right column), where the embeddings are colored by class in the top row and shaded by redshift in the bottom row.

Significant differences are visible between the two latent spaces. Considering the Maven-lite embeddings, only the synthetic SLSN-I light curves (blue) are well-separated from the other classes; the core-collapse (SN Ib/c, II, IIn) and thermonuclear (Ia) events show significant overlap. Observed Ia and II light curves (outlined in black) show similar embeddings independent of class, and little consistency with the synthetic embeddings: the majority of observed SN Ia and SN II lie at the boundary between synthetic SLSN-I and SN II/SN IIn embeddings.

In our Maven embeddings, we observe both clear separation of classes and consistent redshift gradients across our embedded light curves. The simulated SNe Ia appear best-organized by redshift, consistent with their photometric homogeneity. The redshift gradient across observed SNe Ia is also well-aligned with that of the synthetic sample, whereas a similar distribution is not observed in the Maven-lite embeddings. Synthetic SNe Ib/c appear strongly mixed with both SNe Ia and SNe II, indicative of the photometric degeneracies between these classes.

Interestingly, although observed SN Ia and SN II embeddings lie closest to the synthetic events of the same class, the overlap between synthetic and observed data remains low. We attribute this to a distributional shift between synthetic and observed data. Observed events are prioritized for spectroscopic confirmation if they are brighter than (or expected to brighten above) $m < 18.5$th magnitude, and additional quality and purity cuts are imposed (see Section A.1 for details). While a detailed comparison between synthetic and observed events is beyond the scope of this work, this separation may also reflect the simplistic nature of our simulations relative to reality, and emphasizes the need for significantly larger observed SN samples for effective pre-training.

4.2 Classification Performance

Our results are visualized using a set of confusion matrices for our three-way classification task in Fig. 3. We show the confusion matrices for precision (normalized by predicted class) and recall (normalized by true class) for our models. We note higher recall by Maven on the two dominant classes in our sample: 0.79 for SNe II and 0.99 for SNe Ia, compared to 0.74 for SNe II and 0.91 for

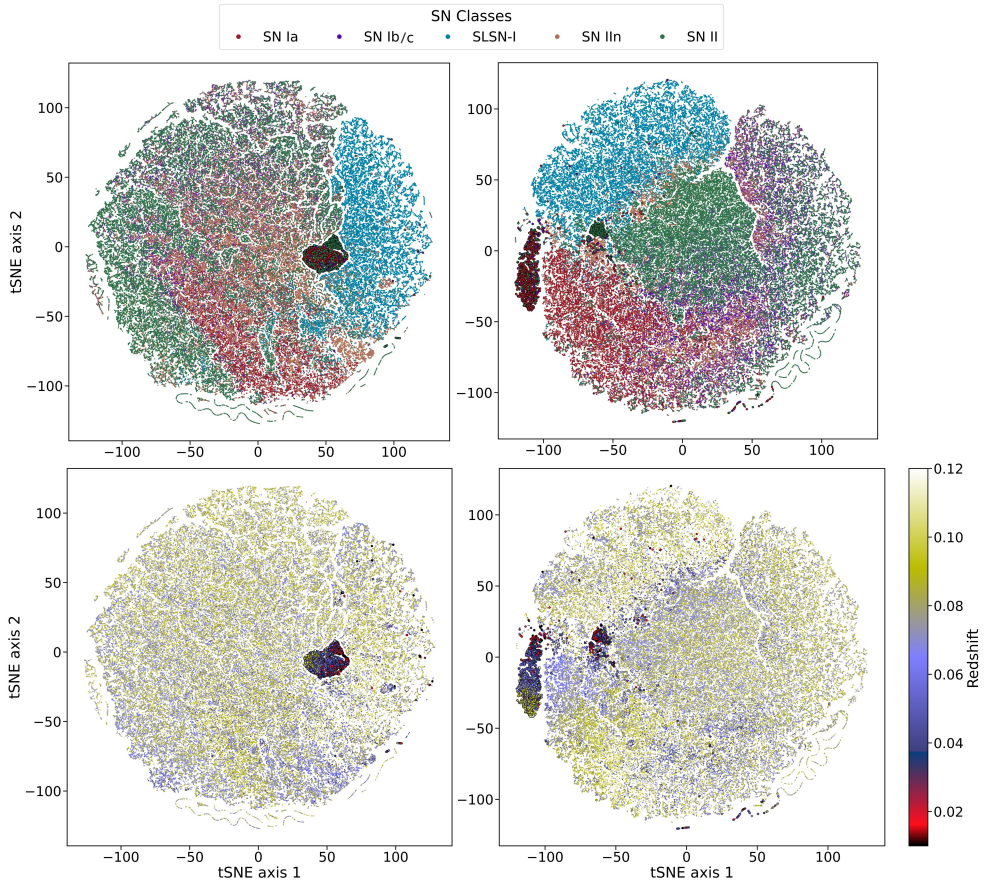

Figure 2: **Visualization of synthetic and observed light curves embedded by Maven-lite (left column) and Maven (right column).** Points in top row are colored by SN class, and points in bottom row are shaded by spectroscopic redshift. Observed data are outlined in black.

SNe Ia with the baseline model. We observe poorer recall with the minority SN Ib/c class, which comprises $\sim$5% of the observed sample: 0.18 with simulated pre-training compared to 0.61 for the baseline. We predict that the baseline model is better able to outline the decision boundaries for this class.

We observe the opposite results on the minority class when considering class precision. Our two-stage pre-training model achieves comparable precision to the baseline for SNe II and SNe Ia but substantially higher precision for SNe Ib/c, 0.58 compared to 0.28. We note that, with substantially higher discovery rates of rare classes anticipated with the Vera C. Rubin Observatory, classification precision is essential for obtaining spectroscopic follow-up observations of relevant events. We have explored the misassociation rate as a function of event peak brightness, but identify no obvious correlations.

A common metric in classification tasks is the F_1 score, which for a class C is defined as the harmonic mean between the class's recall r and precision p: $F_{1,C} := 2p_C r_C(p_C + r_C)$. We calculate for each model both the micro-averaged F_1 score, which averages performance across all events irrespective of class; and the macro-averaged F_1 score, which averages the F_1 score computed independently for each class. The macro-averaged F_1 score is a valuable indicator for our use case given the significant class imbalance, as the micro-F_1 can approach unity when all events are labeled as the dominant class. We present these results, along with the macro-averaged precision and recall ('mac-p' and 'mac-r') in Table 1. We further show the macro-F_1 score of each model as a bar plot in Fig. 4.

We observe macro-F_1 scores within 1-σ of the baseline model for the majority of our pre-trained kNN classifiers, from a score of 0.6874 ± 0.0342 for Maven compared to a baseline of 0.7011 ± 0.0303.

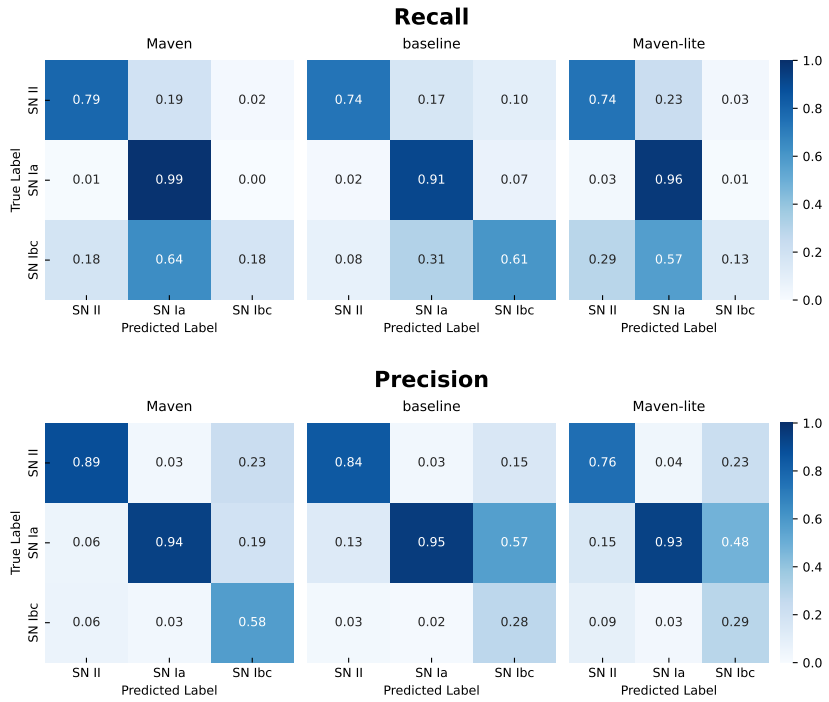

Figure 3: **Normalized precision and recall confusion matrices** for supernova classification across different models and modalities. The models compared are: (a) CLIP with simulation pre-training (Maven), (b) baseline using a supervised model approach, and (c) CLIP without simulation pre-training (Maven-lite). The classes included are SN II, SN Ia, and SN Ib/c.

Table 1: **Overview of classification model performance.** We present three classification models: the baseline only trained on the ZTF dataset, Maven-lite without synthetic pre-training, and Maven with synthetic pretraining and observed fine-tuning. A more comprehensive overview over the runs performed in this paper can be found in Table 3.

Name	Pre-trained	mac-F_1	mic-F_1	mac-p	mac-r
baseline	no	**0.701 ± 0.030**	0.873 ± 0.021	0.693 ± 0.036	**0.753 ± 0.025**
Maven	CLIP	0.687 ± 0.034	**0.925 ± 0.007**	**0.804 ± 0.083**	0.652 ± 0.022
Maven-lite	no	0.627 ± 0.023	0.894 ± 0.011	0.667 ± 0.053	0.612 ± 0.012

The scores for these models are systematically higher than both Maven-lite and the majority of CLIP kNN classifiers without pre-training: the average F_1 score is 0.68 for all pre-trained kNN classifiers compared with an average of 0.63 for the kNN classifiers trained with only observed data.

We have also calculated the performance of our models for the five-way classification task, which additionally considers the rarer classes SN IIn and SLSN I. Here, we observe a marginally higher average F_1 score for Maven than the baseline, though the results are consistent to within one standard deviation (0.50 ± 0.03 for the best model compared to 0.49 ± 0.04).

4.3 Comparison to Transformer-Based SN Classifiers

Next, we compare our multimodal model to photometric classifiers in the literature with transformer-based architectures. Cabrera-Vives et al. [32] apply a custom transformer model (ATAT) to synthetic photometry and metadata from the Extended LSST Astronomical Time-Series Classification Challenge (ELAsTiCC[5]). The ATAT model consists of separate transformers, one which encodes light

[5] `https://portal.nersc.gov/cfs/lsst/DESC_TD_PUBLIC/ELASTICC/`

Table 2: **Overview of regression model performance.** A more comprehensive overview over the runs performed in this paper can be found in Table 4.

Name	R^2	L1	L2	OLF
Maven	**0.6496 ± 0.0398**	**0.0095 ± 0.0004**	**0.0152 ± 0.0014**	0.0002 ± 0.0005
baseline	0.6129 ± 0.0245	0.0104 ± 0.0004	0.0160 ± 0.0010	0.0002 ± 0.0005
Maven-lite	0.6078 ± 0.0408	0.0103 ± 0.0006	0.0161 ± 0.0014	0.0002 ± 0.0005

curves with a temporal encoding based on Fourier series and a quantile tokenizer for extracted photometric features (including the number and phases of non-detections and the flux characteristics of detections). While the model was trained and validated on synthetic photometry for 20 transient and variable astronomical classes, we can generally compare the performance by averaging the reported precision, recall, and F_1 scores of SNe Ia, II, and Ib/c from their Table. Their final model achieves an average macro-F_1 score of 0.67 across the three classes, compared with 0.70 for our end-to-end baseline and 0.69 for our best-performing light curve and spectra-aligned models. They report an average recall of 0.71 for these classes, compared to 0.75 for our end-to-end baseline and 0.65 for Maven; and an average precision of 0.63, compared to 0.69 for our baseline and 0.80 for Maven. We caution that these datasets are distinct, limiting further comparison.

The results of [33] are more directly comparable to this work. Pimentel et al. [33] present a transformer model for ZTF photometry in which the time of each observation is encoded as the phase from first detection using a Fourier decomposition-based temporal modulation, with noise added to the values in training to prevent overfitting. In a two-stage pre-training scheme with both synthetic and observed events, the optimization problem is defined with reconstruction and cross-entropy regularization terms to preserve class-specific information in the encoded light curves. The resulting 'S-TimeModAttn' model is trained and validated on g and r-band light curves from the Zwicky Transient Facility, with presumably substantial overlap with the observational dataset considered in this work. Pimentel et al. [33] report a macro-F_1 of 0.614 ± 0.036 in the task of four-way classification (Ia, II, Ib/c, and SLSN), compared to our 0.6874 ± 0.0342 for three-way classification; and a macro-precision of 0.598 ± 0.030 compared to our 0.804 ± 0.083. A macro-recall (also referred to as completeness) score of 0.72 for three-way classification can be inferred from their confusion matrices, compared to our lower 0.6516 ± 0.0216 from Fig. 3. Class-specific F_1 scores and precisions (also referred to as purity) are not reported.

4.4 Regression Performance

We next consider the task of redshift estimation. We quantify the performance of our models with the coefficient of determination R^2, the L1 and L2 error, and the outlier fraction 'OLF', defined as $|z_{\rm pred} - z_{\rm true}|/(1 + z_{\rm true}) > 0.15$. We report these values for contrastive pre-trained models in Table 2. We also present a bar plot of the R^2 values in Fig. 4. We calculate an R^2 value of $R^2 = 0.6496 \pm 0.0398$ for Maven compared to the end-to-end baseline performance of $R^2 = 0.6129 \pm 0.0245$. The L1 and L2 errors are also lower on average for Maven than for our regression baseline, while the outlier fraction is comparable. We conclude that, on average, Maven outperforms the baseline regression model. Maven-lite, our model without pre-training, achieves an R^2 value of 0.6078 ± 0.0408, lower than both Maven and the baseline model.

Though a comparable photometric redshift model for low-redshift ZTF SNe does not exist in literature, an outlier fraction of 0.004 is reported for 289 photometric SNe Ia in the Supernova Legacy Survey (SNLS), nearly an order of magnitude higher than our best model but with a substantially higher maximum redshift $z < 1.0$ [34]. Another analytic photometric redshift estimator proposed by [35] for SNe Ia discovered by LSST finds an outlier fraction of 0.0023 over $z < 1.0$, compared to our 0.0002.

5 Discussion

We have explored the value of constrastive pre-training in constructing a foundational model for SN science. By first training with synthetic events and fine-tuning with observed events, we have con-

structed a model, Maven, whose performance on the downstream tasks of photometric classification and redshift is competitive with models optimized end-to-end for these tasks. Maven outperforms our classification baseline model, with a micro-averaged-$F1$ score of 0.92. Similarly, Maven outperforms our baseline for redshift regression, with an L2-loss of 0.015 and minimal outlier fraction. While we have limited our study to ZTF data, adapting Maven to incorporate additional photometric filters and classes of astronomical transients would allow us to repurpose it for diverse time-domain studies with the Vera C. Rubin Observatory using fewer computational resources than building multiple specialized models.

CLIP-style pre-training has been proposed as a simple and effective mechanism for extracting information from multiple modalities in a single model. The following conditions need to be met for multimodal constrastive learning to be effective: that significant information content is shared across these modalities; that the mutual information is relevant for the downstream tasks; and that the shared information is the maximal information in each modality relevant for the downstream tasks. Recently, [36] formalized this picture by defining 'multi-view redundancy' as a necessary condition for effective pre-training using traditional constrastive learning. In our case, we know spectra to be highly informative for both classification (the taxonomy is *defined* by spectra obtained early after a SN's explosion, with the temporal evolution of the explosion rarely considered) and redshift inference, which is achieved primarily through the identification of spectral lines. Supernova photometry, although containing some spectral information, is significantly more lossy: the collection of photons through a broadband photometric filter destroys valuable information about a supernova's underlying spectral energy distribution that might otherwise be valuable for these tasks. For these reasons, we can characterize supernova light curves as an 'information-poor' modality and spectra as an 'information-rich' modality for our tasks. CLIP-style pre-training, in this case, is unable to bring significant performance gains beyond end-to-end optimized models. This behavior persists despite aligning these modalities directly with the relevant downstream information (metadata of an event's spectroscopic classification and redshift, as discussed in Appendix A.2): *the least-informative modality sets an upper limit on the mutual information that can be extracted.* We therefore advise caution in the use of multimodal constrastive pre-training, which should be specialized for the input modalities and the anticipated downstream tasks. In our case, additional improvements may be possible with a pre-training scheme designed to preserve both mutual and unique information content relevant for multiple diownstream tasks (as is proposed in [36]).

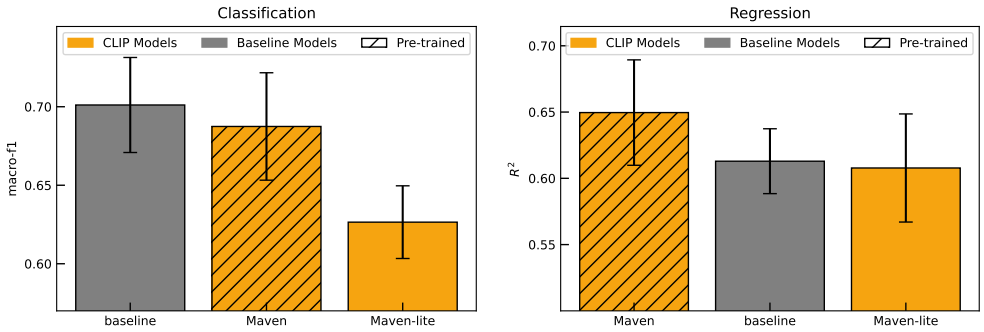

Figure 4: **Final performance metrics** for Maven, Maven-lite, and baseline models for on downstream classification and regression tasks.

6 Conclusion

We have presented Maven, the first model trained with supernova data for multi-task inference. We summarize our key findings below:

1. We train Maven through self-supervised contrastive learning on SN spectra and light curves. Maven is able to achieve state-of-the-art performance on both redshift estimation and SN classification.

2. We find that pre-training on a large simulated dataset significantly improves Maven's performance on downstream tasks over a contrastively-trained model on solely the observed data.
3. Maven does not dramatically outperform supervised models optimized directly for each downstream task. We hypothesize that this is due to the light curve being an information-poor modality, which limits the amount of information our unsupervised objective is able to extract.

Traditional multimodal models have considered complementary representations of the same astronomical source (in this case, photometry and spectroscopy of a SN). When neither spectroscopic *or* photometric coverage of a transient event is available, however, broad physical properties can be inferred using data from the event's host galaxy [37–40]. Early efforts have emphasized the value of host-galaxy photometry for classification of SNe [41, 42, 10, 43]. LSST data will contain photometry for tens of billions of galaxies, millions of which will be spectroscopically-confirmed through the Dark Energy Spectroscopic Instrument [DESI; 44] or 4MOST [45]. Additional work should be dedicated to exploring the linking of modalities spanning distinct lengthscales, which would allow for both SN and host-galaxy information to be consolidated in a single pre-training scheme.

Acknowledgments and Disclosure of Funding

This work was initiated at the IAIFI AstroML Hackathon, held at MIT in January 2024. This work is supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, `http://iaifi.org/`) and AST-2108676. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics of U.S. Department of Energy under grant Contract Number DE-SC0012567. This work was performed in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452. Research reported in this publication was supported by a Postdoctoral Fellowship at the Institute for Advanced Computational Science, Stony Brook University. This work was carried out at the Advanced Research Computing at Hopkins (ARCH) core facility (rockfish.jhu.edu), which is supported by the National Science Foundation (NSF) grant number OAC1920103, as well as the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

References

[1] Benjamin Shappee, J Prieto, KZ Stanek, CS Kochanek, T Holoien, J Jencson, U Basu, JF Beacom, D Szczygiel, G Pojmanski, et al. All sky automated survey for supernovae (asas-sn or" assassin"). In *American Astronomical Society Meeting Abstracts# 223*, volume 223, pages 236–03, 2014.

[2] JL Tonry, L Denneau, AN Heinze, B Stalder, KW Smith, SJ Smartt, CW Stubbs, HJ Weiland, and A Rest. Atlas: a high-cadence all-sky survey system. *Publications of the Astronomical Society of the Pacific*, 130(988):064505, 2018.

[3] 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, et al. The zwicky transient facility: system overview, performance, and first results. *Publications of the Astronomical Society of the Pacific*, 131(995):018002, 2018.

[4] DO Jones, RJ Foley, G Narayan, Jens Hjorth, ME Huber, PD Aleo, KD Alexander, CR Angus, Katie Auchettl, VF Baldassare, et al. The young supernova experiment: survey goals, overview, and operations. *The Astrophysical Journal*, 908(2):143, 2021.

[5] 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. *ApJ*, 873(2):111, March 2019. doi: 10.3847/1538-4357/ab042c.

[6] Daniel Muthukrishna, Gautham Narayan, Kaisey S. Mandel, Rahul Biswas, and Renée Hložek. RAPID: Early Classification of Explosive Transients Using Deep Learning. *PASP*, 131(1005): 118002, November 2019. doi: 10.1088/1538-3873/ab1609.

[7] VA Villar, E Berger, G Miller, R Chornock, A Rest, DO Jones, MR Drout, RJ Foley, R Kirshner, Ragnhild Lunnan, et al. Supernova photometric classification pipelines trained on spectroscopically classified supernovae from the pan-starrs1 medium-deep survey. *The Astrophysical Journal*, 884(1):83, 2019.

[8] A. Möller and T. de Boissière. SuperNNova: an open-source framework for Bayesian, neural network-based supernova classification. *MNRAS*, 491(3):4277–4293, January 2020. doi: 10.1093/mnras/stz3312.

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

[10] Alexander Gagliano, Gabriella Contardo, Daniel Foreman-Mackey, Alex I. Malz, and Patrick D. Aleo. First Impressions: Early-time Classification of Supernovae Using Host-galaxy Information and Shallow Learning. *ApJ*, 954(1):6, September 2023. doi: 10.3847/1538-4357/ace326.

[11] Nabeel Rehemtulla, Adam A. Miller, Theophile Jegou Du Laz, Michael W. Coughlin, Christoffer Fremling, Daniel A. Perley, Yu-Jing Qin, Jesper Sollerman, Ashish A. Mahabal, Russ R. Laher, Reed Riddle, Ben Rusholme, and Shrinivas R. Kulkarni. The Zwicky Transient Facility Bright Transient Survey. III. `BTSbot`: Automated Identification and Follow-up of Bright Transients with Deep Learning. *arXiv e-prints*, art. arXiv:2401.15167, January 2024. doi: 10.48550/arXiv.2401.15167.

[12] Kaylee M. de Soto, 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+: Realtime Fitting and Classification of Supernova Light Curves. *arXiv e-prints*, art. arXiv:2403.07975, March 2024. doi: 10.48550/arXiv.2403.07975.

[13] Ayan Mitra, Richard Kessler, Surhud More, Renee Hlozek, and LSST Dark Energy Science Collaboration. Using Host Galaxy Photometric Redshifts to Improve Cosmological Constraints with Type Ia Supernovae in the LSST Era. *ApJ*, 944(2):212, February 2023. doi: 10.3847/1538-4357/acb057.

[14] Helen Qu and Masao Sako. Photo-zSNthesis: Converting Type Ia Supernova Lightcurves to Redshift Estimates via Deep Learning. *ApJ*, 954(2):201, September 2023. doi: 10.3847/1538-4357/aceafa.

[15] Joseph W. Richards, Darren Homrighausen, Peter E. Freeman, Chad M. Schafer, and Dovi Poznanski. Semi-supervised learning for photometric supernova classification. *MNRAS*, 419(2): 1121–1135, January 2012. doi: 10.1111/j.1365-2966.2011.19768.x.

[16] 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. *ApJ*, 905(2):94, December 2020. doi: 10.3847/1538-4357/abc6fd.

[17] Yuge Shi, Imant Daunhawer, Julia E Vogt, Philip Torr, and Amartya Sanyal. How robust are pre-trained models to distribution shift? In *ICML 2022: Workshop on Spurious Correlations, Invariance and Stability*, 2022.

[18] Priya Goyal, Quentin Duval, Isaac Seessel, Mathilde Caron, Ishan Misra, Levent Sagun, Armand Joulin, and Piotr Bojanowski. Vision Models Are More Robust And Fair When Pretrained On Uncurated Images Without Supervision. *arXiv e-prints*, art. arXiv:2202.08360, February 2022. doi: 10.48550/arXiv.2202.08360.

[19] Yuzhe Yang and Zhi Xu. Rethinking the Value of Labels for Improving Class-Imbalanced Learning. *arXiv e-prints*, art. arXiv:2006.07529, June 2020. doi: 10.48550/arXiv.2006.07529.

[20] Daehee Kim, Seunghyun Park, Jinkyu Kim, and Jaekoo Lee. SelfReg: Self-supervised Contrastive Regularization for Domain Generalization. *arXiv e-prints*, art. arXiv:2104.09841, April 2021. doi: 10.48550/arXiv.2104.09841.

[21] Linus Ericsson, Henry Gouk, Chen Change Loy, and Timothy M. Hospedales. Self-Supervised Representation Learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, 39(3):42–62, May 2022. doi: 10.1109/MSP.2021.3134634.

[22] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[23] Richard Kessler, Joseph P. Bernstein, David Cinabro, Benjamin Dilday, Joshua A. Frieman, Saurabh Jha, Stephen Kuhlmann, Gajus Miknaitis, Masao Sako, Matt Taylor, and Jake Vanderplas. SNANA: A Public Software Package for Supernova Analysis. *PASP*, 121(883):1028, September 2009. doi: 10.1086/605984.

[24] 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. *ApJS*, 266(1):9, May 2023. doi: 10.3847/1538-4365/acbfba.

[25] R. Kessler, G. Narayan, A. Avelino, E. Bachelet, R. Biswas, P. J. Brown, D. F. Chernoff, A. J. Connolly, M. Dai, S. Daniel, R. Di Stefano, M. R. Drout, L. Galbany, S. González-Gaitán, M. L. Graham, R. Hložek, E. E. O. Ishida, J. Guillochon, S. W. Jha, D. O. Jones, K. S. Mandel, D. Muthukrishna, A. O'Grady, C. M. Peters, J. R. Pierel, K. A. Ponder, A. Prša, S. Rodney, V. A. Villar, LSST Dark Energy Science Collaboration, and Transient and Variable Stars Science Collaboration. Models and Simulations for the Photometric LSST Astronomical Time Series Classification Challenge (PLAsTiCC). *PASP*, 131(1003):094501, September 2019. doi: 10.1088/1538-3873/ab26f1.

[26] C. Fremling, A. A. Miller, Y. Sharma, A. Dugas, D. A. Perley, K. Taggart, J. Sollerman, A. Goobar, M. L. Graham, J. D. Neill, J. Nordin, M. Rigault, R. Walters, I. Andreoni, A. Bagdasaryan, J. Belicki, C. Cannella, E. C. Bellm, S. B. Cenko, K. De, R. Dekany, S. Frederick, V. Z. Golkhou, M. J. Graham, G. Helou, A. Y. Q. Ho, M. M. Kasliwal, T. Kupfer, R. R. Laher, A. Mahabal, F. J. Masci, R. Riddle, B. Rusholme, S. Schulze, D. L. Shupe, R. M. Smith, S. van Velzen, Lin Yan, Y. Yao, Z. Zhuang, and S. R. Kulkarni. The Zwicky Transient Facility Bright Transient Survey. I. Spectroscopic Classification and the Redshift Completeness of Local Galaxy Catalogs. *ApJ*, 895(1):32, May 2020. doi: 10.3847/1538-4357/ab8943.

[27] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *arXiv e-prints*, art. arXiv:1706.03762, June 2017. doi: 10.48550/arXiv.1706.03762.

[28] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[29] Lukas Biewald. Experiment tracking with weights and biases, 2020. URL `https://www.wandb.com/`. Software available from wandb.com.

[30] Maryam Modjaz, Claudia P. Gutiérrez, and Iair Arcavi. New regimes in the observation of core-collapse supernovae. *Nature Astronomy*, 3:717–724, August 2019. doi: 10.1038/s41550-019-0856-2.

[31] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[32] G. Cabrera-Vives, D. Moreno-Cartagena, N. Astorga, I. Reyes-Jainaga, F. Förster, P. Huijse, J. Arredondo, A. M. Muñoz Arancibia, A. Bayo, M. Catelan, P. A. Estévez, P. Sánchez-Sáez, A. Álvarez, P. Castellanos, P. Gallardo, A. Moya, and D. Rodriguez-Mancini. ATAT: Astronomical Transformer for time series And Tabular data. *arXiv e-prints*, art. arXiv:2405.03078, May 2024. doi: 10.48550/arXiv.2405.03078.

[33] Óscar Pimentel, Pablo A. Estévez, and Francisco Förster. Deep Attention-based Supernovae Classification of Multiband Light Curves. *AJ*, 165(1):18, January 2023. doi: 10.3847/1538-3881/ac9ab4.

[34] N. Palanque-Delabrouille, V. Ruhlmann-Kleider, S. Pascal, J. Rich, J. Guy, G. Bazin, P. Astier, C. Balland, S. Basa, R. G. Carlberg, A. Conley, D. Fouchez, D. Hardin, I. M. Hook, D. A. Howell, R. Pain, K. Perrett, C. J. Pritchet, N. Regnault, and M. Sullivan. Photometric redshifts for type Ia supernovae in the supernova legacy survey. *A&A*, 514:A63, May 2010. doi: 10.1051/0004-6361/200913283.

[35] Yun Wang, E. Gjergo, and S. Kuhlmann. Analytic photometric redshift estimator for Type Ia supernovae from the Large Synoptic Survey Telescope. *MNRAS*, 451(2):1955–1963, August 2015. doi: 10.1093/mnras/stv1090.

[36] Paul Pu Liang, Zihao Deng, Martin Ma, James Zou, Louis-Philippe Morency, and Ruslan Salakhutdinov. Factorized Contrastive Learning: Going Beyond Multi-view Redundancy. *arXiv e-prints*, art. arXiv:2306.05268, June 2023. doi: 10.48550/arXiv.2306.05268.

[37] A. A. Hakobyan, V. Zh. Adibekyan, L. S. Aramyan, A. R. Petrosian, J. M. Gomes, G. A. Mamon, D. Kunth, and M. Turatto. Supernovae and their host galaxies. I. The SDSS DR8 database and statistics. *A&A*, 544:A81, August 2012. doi: 10.1051/0004-6361/201219541.

[38] Yijung Kang, Young-Wook Lee, Young-Lo Kim, Chul Chung, and Chang Hee Ree. Early-type Host Galaxies of Type Ia Supernovae. II. Evidence for Luminosity Evolution in Supernova Cosmology. *ApJ*, 889(1):8, January 2020. doi: 10.3847/1538-4357/ab5afc.

[39] Steve Schulze, Ofer Yaron, Jesper Sollerman, Giorgos Leloudas, Amit Gal, Angus H. Wright, Ragnhild Lunnan, Avishay Gal-Yam, Eran O. Ofek, Daniel A. Perley, Alexei V. Filippenko, Mansi M. Kasliwal, Shrinivas R. Kulkarni, James D. Neill, Peter E. Nugent, Robert M. Quimby, Mark Sullivan, Nora Linn Strotjohann, Iair Arcavi, Sagi Ben-Ami, Federica Bianco, Joshua S. Bloom, Kishalay De, Morgan Fraser, Christoffer U. Fremling, Assaf Horesh, Joel Johansson, Patrick L. Kelly, Nikola Knežević, Sladjana Knežević, Kate Maguire, Anders Nyholm, Seméli Papadogiannakis, Tanja Petrushevska, Adam Rubin, Lin Yan, Yi Yang, Scott M. Adams, Filomena Bufano, Kelsey I. Clubb, Ryan J. Foley, Yoav Green, Jussi Harmanen, Anna Y. Q. Ho, Isobel M. Hook, Griffin Hosseinzadeh, D. Andrew Howell, Albert K. H. Kong, Rubina Kotak, Thomas Matheson, Curtis McCully, Dan Milisavljevic, Yen-Chen Pan, Dovi Poznanski, Isaac Shivvers, Sjoert van Velzen, and Kars K. Verbeek. The Palomar Transient Factory Core-collapse Supernova Host-galaxy Sample. I. Host-galaxy Distribution Functions and Environment Dependence of Core-collapse Supernovae. *ApJS*, 255(2):29, August 2021. doi: 10.3847/1538-4365/abff5e.

[40] Sudeshna Chakraborty, Benjamin Sadler, Peter Hoeflich, Eric Y. Hsiao, M. M. Phillips, C. R. Burns, T. Diamond, I. Dominguez, L. Galbany, S. A. Uddin, C. Ashall, K. Krisciunas, S. Kumar, T. B. Mera, N. Morrell, E. Baron, C. Contreras, M. D. Stritzinger, and N. B. Suntzeff. Type Ia Supernova Progenitor Properties and their Host Galaxies. *ApJ*, 969(2):80, July 2024. doi: 10.3847/1538-4357/ad4702.

[41] Sebastian Gomez, Edo Berger, Peter K. Blanchard, Griffin Hosseinzadeh, Matt Nicholl, V. Ashley Villar, and Yao Yin. FLEET: A Redshift-agnostic Machine Learning Pipeline to Rapidly Identify Hydrogen-poor Superluminous Supernovae. *ApJ*, 904(1):74, November 2020. doi: 10.3847/1538-4357/abbf49.

[42] R. Carrasco-Davis, E. Reyes, C. Valenzuela, F. Förster, P. A. Estévez, G. Pignata, F. E. Bauer, I. Reyes, P. Sánchez-Sáez, G. Cabrera-Vives, S. Eyheramendy, M. Catelan, J. Arredondo, E. Castillo-Navarrete, D. Rodríguez-Mancini, D. Ruz-Mieres, A. Moya, L. Sabatini-Gacitúa, C. Sepúlveda-Cobo, A. A. Mahabal, J. Silva-Farfán, E. Camacho-Iñiguez, and L. Galbany. Alert Classification for the ALeRCE Broker System: The Real-time Stamp Classifier. *AJ*, 162 (6):231, December 2021. doi: 10.3847/1538-3881/ac0ef1.

[43] Xinyue Sheng, Matt Nicholl, Ken W. Smith, David R. Young, Roy D. Williams, Heloise F. Stevance, Stephen J. Smartt, Shubham Srivastav, and Thomas Moore. NEural Engine for Discovering Luminous Events (NEEDLE): identifying rare transient candidates in real time from host galaxy images. *MNRAS*, 531(2):2474–2492, June 2024. doi: 10.1093/mnras/stae1253.

[44] Michael Levi, Lori E. Allen, Anand Raichoor, Charles Baltay, Segev BenZvi, Florian Beutler, Adam Bolton, Francisco J. Castander, Chia-Hsun Chuang, Andrew Cooper, Jean-Gabriel Cuby, Arjun Dey, Daniel Eisenstein, Xiaohui Fan, Brenna Flaugher, Carlos Frenk, Alma X. Gonzalez-Morales, Or Graur, Julien Guy, Salman Habib, Klaus Honscheid, Stephanie Juneau, Jean-Paul Kneib, Ofer Lahav, Dustin Lang, Alexie Leauthaud, Betta Lusso, Axel de la Macorra, Marc Manera, Paul Martini, Shude Mao, Jeffrey A. Newman, Nathalie Palanque-Delabrouille, Will J. Percival, Carlos Allende Prieto, Constance M. Rockosi, Vanina Ruhlmann-Kleider, David Schlegel, Hee-Jong Seo, Yong-Seon Song, Greg Tarle, Risa Wechsler, David Weinberg, Christophe Yeche, and Ying Zu. The Dark Energy Spectroscopic Instrument (DESI). In *Bulletin of the American Astronomical Society*, volume 51, page 57, September 2019. doi: 10.48550/arXiv.1907.10688.

[45] J. Dumayne, I. M. Hook, S. C. Williams, G. A. Lowes, D. Head, A. Fritz, O. Graur, B. Holwerda, A. Humphrey, A. Milligan, M. Nicholl, B. F. Roukema, and P. Wiseman. Using 4MOST to refine the measurement of galaxy properties: a case study of supernova hosts. *RAS Techniques and Instruments*, 2(1):453–469, January 2023. doi: 10.1093/rasti/rzad036.

[46] J. Guy, P. Astier, S. Baumont, D. Hardin, R. Pain, N. Regnault, S. Basa, R. G. Carlberg, A. Conley, S. Fabbro, D. Fouchez, I. M. Hook, D. A. Howell, K. Perrett, C. J. Pritchet, J. Rich, M. Sullivan, P. Antilogus, E. Aubourg, G. Bazin, J. Bronder, M. Filiol, N. Palanque-Delabrouille, P. Ripoche, and V. Ruhlmann-Kleider. SALT2: using distant supernovae to improve the use of type Ia supernovae as distance indicators. *A&A*, 466(1):11–21, April 2007. doi: 10.1051/0004-6361:20066930.

[47] Richard Kessler, Bruce Bassett, Pavel Belov, Vasudha Bhatnagar, Heather Campbell, Alex Conley, Joshua A Frieman, Alexandre Glazov, Santiago González-Gaitán, Renée Hlozek, et al. Results from the supernova photometric classification challenge. *Publications of the Astronomical Society of the Pacific*, 122(898):1415, 2010.

[48] V Ashley Villar, Edo Berger, Brian D Metzger, and James Guillochon. Theoretical models of optical transients. i. a broad exploration of the duration–luminosity phase space. *The Astrophysical Journal*, 849(1):70, 2017.

[49] Louis-Gregory Strolger, Tomas Dahlen, Steven A. Rodney, Or Graur, Adam G. Riess, Curtis McCully, Swara Ravindranath, Bahram Mobasher, and A. Kristin Shahady. The Rate of Core Collapse Supernovae to Redshift 2.5 from the CANDELS and CLASH Supernova Surveys. *ApJ*, 813(2):93, November 2015. doi: 10.1088/0004-637X/813/2/93.

[50] Isaac Shivvers, Maryam Modjaz, WeiKang Zheng, Yuqian Liu, Alexei V. Filippenko, Jeffrey M. Silverman, Thomas Matheson, Andrea Pastorello, Or Graur, Ryan J. Foley, Ryan Chornock, Nathan Smith, Jesse Leaman, and Stefano Benetti. Revisiting the Lick Observatory Supernova Search Volume-limited Sample: Updated Classifications and Revised Stripped-envelope Supernova Fractions. *PASP*, 129(975):054201, May 2017. doi: 10.1088/1538-3873/aa54a6.

[51] R. Hounsell, D. Scolnic, R. J. Foley, R. Kessler, V. Miranda, A. Avelino, R. C. Bohlin, A. V. Filippenko, J. Frieman, S. W. Jha, P. L. Kelly, R. P. Kirshner, K. Mandel, A. Rest, A. G. Riess, S. A. Rodney, and L. Strolger. Simulations of the WFIRST Supernova Survey and Forecasts of Cosmological Constraints. *ApJ*, 867(1):23, November 2018. doi: 10.3847/1538-4357/aac08b.

[52] Piero Madau and Mark Dickinson. Cosmic Star-Formation History. *ARA&A*, 52:415–486, August 2014. doi: 10.1146/annurev-astro-081811-125615.

[53] Nadejda Blagorodnova, James D. Neill, Richard Walters, Shrinivas R. Kulkarni, Christoffer Fremling, Sagi Ben-Ami, Richard G. Dekany, Jason R. Fucik, Nick Konidaris, Reston Nash, Chow-Choong Ngeow, Eran O. Ofek, Donal O' Sullivan, Robert Quimby, Andreas Ritter, and Karl E. Vyhmeister. The SED Machine: A Robotic Spectrograph for Fast Transient Classification. *PASP*, 130(985):035003, March 2018. doi: 10.1088/1538-3873/aaa53f.

[54] Jason A. Cardelli, Geoffrey C. Clayton, and John S. Mathis. The Relationship between Infrared, Optical, and Ultraviolet Extinction. *ApJ*, 345:245, October 1989. doi: 10.1086/167900.

[55] Daniel Muthukrishna, David Parkinson, and Brad E. Tucker. DASH: Deep Learning for the Automated Spectral Classification of Supernovae and Their Hosts. *ApJ*, 885(1):85, November 2019. doi: 10.3847/1538-4357/ab48f4.

[56] Stéphane Blondin and John L. Tonry. Determining the Type, Redshift, and Age of a Supernova Spectrum. *ApJ*, 666(2):1024–1047, September 2007. doi: 10.1086/520494.

[57] Thomas Matheson, Carl Stubens, Nicholas Wolf, Chien-Hsiu Lee, Gautham Narayan, Abhijit Saha, Adam Scott, Monika Soraisam, Adam S. Bolton, Benjamin Hauger, David R. Silva, John Kececioglu, Carlos Scheidegger, Richard Snodgrass, Patrick D. Aleo, Eric Evans-Jacquez, Navdeep Singh, Zhe Wang, Shuo Yang, and Zhenge Zhao. The ANTARES Astronomical Time-domain Event Broker. *AJ*, 161(3):107, March 2021. doi: 10.3847/1538-3881/abd703.

[58] Sagi Ben-Ami, Nick Konidaris, Robert Quimby, Jack TC Davis, Chow Choong Ngeow, Andreas Ritter, and Alexander Rudy. The sed machine: a dedicated transient ifu spectrograph. In *Ground-based and Airborne Instrumentation for Astronomy IV*, volume 8446, pages 1044–1052. SPIE, 2012.

[59] Nadejda Blagorodnova, James D Neill, Richard Walters, Shrinivas R Kulkarni, Christoffer Fremling, Sagi Ben-Ami, Richard G Dekany, Jason R Fucik, Nick Konidaris, Reston Nash, et al. The sed machine: a robotic spectrograph for fast transient classification. *Publications of the Astronomical Society of the Pacific*, 130(985):035003, 2018.

[60] M Rigault, JD Neill, N Blagorodnova, A Dugas, M Feeney, R Walters, V Brinnel, Y Copin, C Fremling, J Nordin, et al. Fully automated integral field spectrograph pipeline for the sedmachine: pysedm. *Astronomy & Astrophysics*, 627:A115, 2019.

[61] S Bradley Cenko, Derek B Fox, Dae-Sik Moon, Fiona A Harrison, SR Kulkarni, John R Henning, C Dani Guzman, Marco Bonati, Roger M Smith, Robert P Thicksten, et al. The automated palomar 60 inch telescope. *Publications of the Astronomical Society of the Pacific*, 118(848):1396, 2006.

[62] Ofer Yaron and Avishay Gal-Yam. WISeREP—An Interactive Supernova Data Repository. *PASP*, 124(917):668, July 2012. doi: 10.1086/666656.

[63] Daniel A. Perley, Christoffer Fremling, Jesper Sollerman, Adam A. Miller, Aishwarya S. Dahiwale, Yashvi Sharma, Eric C. Bellm, Rahul Biswas, Thomas G. Brink, Rachel J. Bruch, Kishalay De, Richard Dekany, Andrew J. Drake, Dmitry A. Duev, Alexei V. Filippenko, Avishay Gal-Yam, Ariel Goobar, Matthew J. Graham, Melissa L. Graham, Anna Y. Q. Ho, Ido Irani, Mansi M. Kasliwal, Young-Lo Kim, S. R. Kulkarni, Ashish Mahabal, Frank J. Masci, Shaunak Modak, James D. Neill, Jakob Nordin, Reed L. Riddle, Maayane T. Soumagnac, Nora L. Strotjohann, Steve Schulze, Kirsty Taggart, Anastasios Tzanidakis, Richard S. Walters, and Lin Yan. The Zwicky Transient Facility Bright Transient Survey. II. A Public Statistical Sample for Exploring Supernova Demographics. *ApJ*, 904(1):35, November 2020. doi: 10.3847/1538-4357/abbd98.

A Appendix / supplemental material

A.1 Data

Here, we provide more details about the SNe simulation datasets used for pre-training and the ZTF dataset used for fine-tuning and inference.

A.1.1 Simulating Supernovae with the SNANA Simulation Code

We generate synthetic SN samples using the SNANA simulation code. SNANA mimics the observing process beginning from a rest-frame spectral energy distribution (SED) of an astrophysical transient. A volumetric rate is chosen and the sky is populated at random with transients. A survey strategy, detection efficiency, and the survey's estimated noise properties (zeropoint and sky noise) are provided to construct synthetic observations. Our 500,000 simulated events are evenly split between five different SN classes: SNe Ia (using the `SALT2` model; [46]); SNe Ib/c (`SNIbc-Templates`; [47]); SLSNe-I (using the model `SLSNI-MOSFIT`; [48]); and SNe II (`SNII-Templates`; [47]), which includes both SNe IIP/IIL; and SNe IIn (`SNIIn-MOSFIT`; [48])

To produce our simuations, we use the same volumetric rates for SNE II, SNe IIn, and SNe Ib/c as in the PLAsTiCC challenge [49], re-scaled to match the fractional rates presented in [50]. The volumetric rate for SNe Ia is taken from [51], and that for SLSNe-I traces the star-formation history parameterized in [52]. Our simulations mimic the ZTF survey strategy, filter transmissions, and reported sky noise. This results in a similar selection function favoring low-redshift ($z < 0.1$) SNe as our observed sample, although we do not explicitly define a brightness threshold for photometry as is done with the BTS sample [26] and our sole quality cut is removing events with fewer than 4 total photometric observations. As a result, our simulated events are intrinsically fainter and lower-quality than our observed events.

In addition to the previously-developed simulations, we define a spectrograph object in SNANA with wavelength bins corresponding to the wavelength coverage of the ZTF SED machine [53], with which the vast majority of our observed SNe were classified. To mimic the stochasticity inherent to SN classification in practice, we allow synthetic spectra to be obtained randomly from explosion to peak light, and with sufficient exposure time to achieve S/N of 5 within an arbitrary wavelength window. Galactic extinction is applied to both modalities at the simulated SN location following the extinction law from [54]. We then pre-process all spectra in the same manner as in [55]. we apply low-pass median filtering to remove high-frequency noise, re-bin the data to log-wavelength space, and estimate the flux continuum using a polynomial fit and divide it out. While this continuum-division step removes color information, it has been shown that it has a negligible impact on redshift estimation [56]. The spectra are kept in the observer frame (not redshift-corrected).

A.1.2 The Zwicky Transient Facility Bright Transient Survey

Since 2019, the Zwicky Transient Facility (ZTF; [3]) has conducted a wide-field public survey consisting of photometry obtained with the Palomar 48-inch Schmidt telescope at a cadence of roughly 2 nights. The telescope observes in three photometric filters: ZTF-g, ZTF-r, and ZTF-i. Photometry is promptly reduced and streamed to alert brokers including ANTARES [the Arizona-NOIRLab Temporal Analysis and Response to Events System; 57]. For non-Galactic transients observed at or expected to peak brighter than an apparent magnitude of $\sim$18.5, a classification spectrum is automatically obtained using the Spectral Energy Distribution Machine (SEDM; [58–60]), a low-resolution spectrograph mounted on the Palomar 60-inch telescope [61]. SEDM spectra are then uploaded to the Transient Name Server and the Weizmann Interactive Supernova Data Repository [WISeREP; 62]. 5,377 SNe have been spectroscopically confirmed at the time of writing as part of this Bright Transient Survey.

We obtain metadata for 4,702 spectroscopically-classified SNe on June 18th, 2024 from the ZTF Bright Transient Survey [26] after applying all quality and purity cuts available on the ZTF BTS webpage[6] (described in detail in [63]). The subsequent SNe have photometric coverage before and after peak light, good visibility throughout the photospheric phase, an uncontaminated reference image, and occurred in low extinction fields.

[6] `https://sites.astro.caltech.edu/ztf/bts/bts.php`

Table 3: **Classification performance for three classes by model configuration** : This table presents the classification performance of various models using light curve data from the ZTF dataset. The models are categorized based on whether they utilized simulation pre-training ('pre-trained'), the type of last layer added to embedding models ('last-layer'). The modalities taken into account when training on the real ZTF dataset are indicated in 'real-pre' (lc - light curve, sp - spectrum, m - metadata) as well as whether a SVC or kNN. Performance metrics include macro-F1 (mac-f1), micro-F1 (mic-f1), macro-precision (mac-p), and macro-recall (mac-r). The results are presented as mean ± standard deviation, calculated over five folds. Baseline models, trained in an end-to-end supervised fashion using only the ZTF data, are included for comparison.

pre-trained	last-layer	real-pre	mac-f1	mac-p	mac-r
no	end-to-end baseline		0.7011 ± 0.0303	0.6934 ± 0.0360	0.7527 ± 0.0247
clip	kNN	lc-m	0.6920 ± 0.0217	0.7286 ± 0.0377	0.6721 ± 0.0183
clip	kNN	lc-sp	0.6874 ± 0.0342	0.8041 ± 0.0833	0.6516 ± 0.0216
clip	kNN	lc-sp-m	0.6849 ± 0.0194	0.7280 ± 0.0334	0.6643 ± 0.0161
clip	SVC	lc-m	0.6747 ± 0.0297	0.8026 ± 0.0257	0.6435 ± 0.0257
clip	SVC	lc-sp-m	0.6522 ± 0.0237	0.7892 ± 0.0975	0.6247 ± 0.0215
no	kNN	lc-sp-m	0.6268 ± 0.0251	0.7204 ± 0.0701	0.6000 ± 0.0199
no	kNN	lc-sp	0.6265 ± 0.0231	0.6670 ± 0.0532	0.6119 ± 0.0121
no	kNN	lc-m	0.6249 ± 0.0228	0.7309 ± 0.0661	0.6035 ± 0.0184
clip	SVC	lc-sp	0.6195 ± 0.0190	0.7753 ± 0.0994	0.6056 ± 0.0172
no	SVC	lc-m	0.5971 ± 0.0220	0.7871 ± 0.1858	0.5842 ± 0.0163
no	SVC	lc-sp-m	0.5938 ± 0.0156	0.7892 ± 0.1873	0.5802 ± 0.0077
no	SVC	lc-sp	0.5749 ± 0.0099	0.5857 ± 0.0126	0.5686 ± 0.0102

Next, we use the `Python` client of the `antares` alert broker [57] to consolidate difference photometry for all SNe in ZTF-g and ZTF-r [ZTF-i observations are mainly private, comprising $\sim$10% of all observations; 24], and download their associated SEDM spectra from the Transient Name Server[7] and WISEReP[8] [9]. We pre-process the observed spectra following the same procedure as our synthetic ones.

A.2 Metadata CLIP

In addition to SN spectrum and light curve measurements, we also considered SN metadata as an additional modality for training a CLIP model. The metadata modality used in our training includes supernovae redshifts and class labels. We encode each class label with a learnable embedding vector. The metadata encoder consists of a multilayer perceptron (MLP) that takes in the concatenated vector of class embedding and redshift and outputs the final embedding. The number of hidden layers and the hidden layer dimension in the MLP were chosen from a hyperparameter search.

The models which directly align event photometry with relevant metadata (redshift and class) in pre-training do not significantly outperform the models in which photometry and spectroscopy alone are aligned. Considering only pre-trained models for the task of classification, we observe comparable three-way macro-F_1 scores when aligning light curves with metadata (0.692 ± 0.022), light curves with spectra (0.687 ± 0.034), and light curves with both spectra and metadata (0.685 ± 0.019). Each of our CLIP objectives featured photometry as a modality, and we predict that this more information-poor modality is driving the observed performance across each of these models, as we discuss in additional detail in section 6.

[7] https://www.wis-tns.org/

[8] https://www.wiserep.org/

[9] Despite spectroscopic classifications being available on the ZTF website for all listed SNe, SEDM spectra could not be found for a few objects. When an SEDM spectrum is not available, we instead use the first reported spectrum. A positional encoding is used for the wavelengths of each spectrum, so in principle our spectrum encoder has the capacity to generalize to multiple spectrographs.

Table 4: **Regression Performance by Model Configuration**: This table presents the regression performance of various models using light curve data from the ZTF dataset. The models are categorized based on whether they utilized simulation pre-training ('pre-trained'), the type of last layer added to embedding models ('last-layer'). The modalities taken into account when training on the real ZTF dataset is indicated in 'real-pre' (lc - light curve, sp - spectrum, m - metadata) as well weather we use a linear or kNN layer to translate our embedding to a redshift prediction ('last-layer'). Performance metrics include the coefficient of determination (R^2), L1 loss, and L2 loss. The results are presented as mean ± standard deviation, calculated over five folds. Baseline models, trained in an end-to-end supervised fashion using only the ZTF data, are included for comparison.

pre-trained	**last-layer**	**real-pre**	$\mathbf{R}^2$	**L1**	**L2**
clip	kNN	lc-m	0.6543 ± 0.0280	0.0094 ± 0.0005	0.0152 ± 0.0010
clip	Linear	lc-sp-m	0.6513 ± 0.0440	0.0096 ± 0.0005	0.0152 ± 0.0016
clip	kNN	lc-sp	0.6496 ± 0.0398	0.0095 ± 0.0004	0.0152 ± 0.0014
clip	kNN	lc-sp-m	0.6470 ± 0.0422	0.0094 ± 0.0006	0.0152 ± 0.0012
clip	Linear	lc-sp	0.6386 ± 0.0447	0.0099 ± 0.0003	0.0155 ± 0.0016
clip	Linear	lc-m	0.6345 ± 0.0444	0.0100 ± 0.0006	0.0156 ± 0.0014
no	kNN	lc-m	0.6150 ± 0.0294	0.0103 ± 0.0003	0.0160 ± 0.0012
no	end-to-end baseline		0.6129 ± 0.0245	0.0104 ± 0.0004	0.0160 ± 0.0010
no	kNN	lc-sp-m	0.6090 ± 0.0464	0.0102 ± 0.0005	0.0161 ± 0.0015
no	kNN	lc-sp	0.6078 ± 0.0408	0.0103 ± 0.0006	0.0161 ± 0.0014
no	Linear	lc-sp	0.5948 ± 0.0402	0.0107 ± 0.0007	0.0164 ± 0.0015
no	Linear	lc-sp-m	0.5938 ± 0.0450	0.0108 ± 0.0004	0.0164 ± 0.0016
no	Linear	lc-m	0.5927 ± 0.0399	0.0107 ± 0.0004	0.0165 ± 0.0015