The Astrophysical Journal, 930:70 (30pp), 2022 May 1 https://doi.org/10.3847/1538-4357/ac5c48

© 2022. The Author(s). Published by the American Astronomical Society.

OPEN ACCESS

Spectroscopic Studies of Type Ia Supernovae Using LSTM Neural Networks

Lei Hu[1,2], Xingzhuo Chen[3], and Lifan Wang[3]
[1] Purple Mountain Observatory, Nanjing 210023, People's Republic of China
[2] School of Astronomy and Space Sciences, University of Science and Technology of China, 230029 Hefei, People's Republic of China
[3] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics & Astronomy, Texas A&M University, Department of Physics and Astronomy, 4242 TAMU, College Station, TX 77843, USA; lifan@tamu.edu

Received 2021 June 30; revised 2022 January 15; accepted 2022 February 4; published 2022 May 4

Abstract

We present a data-driven method based on long short-term memory (LSTM) neural networks to analyze spectral time series of Type Ia supernovae (SNe Ia). The data set includes 3091 spectra from 361 individual SNe Ia. The method allows for accurate reconstruction of the spectral sequence of an SN Ia based on a single observed spectrum around maximum light. The precision of the spectral reconstruction increases with more spectral time coverages, but the significant benefit of multiple epoch data at around optical maximum is only evident for observations separated by more than a week. The method shows great power in extracting the spectral information of SNe Ia and suggests that the most critical information of an SN Ia can be derived from a single spectrum around the optical maximum. The algorithm we have developed is important for the planning of spectroscopic follow-up observations of future SN surveys with the LSST/Rubin and WFIRST/Roman telescopes.

Unified Astronomy Thesaurus concepts: Supernovae (1668); Type Ia supernovae (1728)

Supporting material: machine-readable table

1. Introduction

Type Ia supernovae (SNe Ia) are among the most luminous phenomena in the transient universe that empower accurate measurements of the history of the expansion of the universe. It is believed that SNe Ia result from the thermonuclear explosions of carbon/oxygen (C/O) white dwarfs (WDs; Hoyle & Fowler 1960) in binary systems. They are used as standardizable distance candles to probe the expansion history of the universe (Riess et al. 1998; Perlmutter et al. 1999; Riess 2019) and constrain the properties of the dark energy content of the universe (Knop et al. 2003; Kowalski et al. 2008; Amanullah et al. 2010; Suzuki et al. 2012; Betoule et al. 2014; Scolnic et al. 2018; Abbott et al. 2019).

Although the underlying progenitor systems and the physical pathways toward the explosions remain elusive, the diverse spectroscopic observations seem to epitomize, to a certain extent, intrinsic diversities among SNe Ia (e.g., Branch & Wheeler 2017). The ignition processes and explosion geometries may imprint their signatures on the observed optical spectra (Wang & Wheeler 2008; Cikota et al. 2019; Chen et al. 2020; Yang et al. 2020). The intrinsic brightness of SNe Ia and their magnitude dispersion on the Hubble diagram seem to be correlated to the environment they explode (Wang et al. 1997; Uddin et al. 2020). The viewing angle effect of asymmetric explosions has been proposed to be likely responsible for the diversity of the ejecta velocities measured from spectral lines such as the Si II λ6355 absorption line (Maeda et al. 2010; Maund et al. 2010). The intrinsic diversity of SN Ia phenomena implies that each individual SN Ia bears discrete observational signatures, but the observational data set can be analyzed through data-driven approaches.

Normal SNe Ia can be divided into two subclasses in terms of their expansion velocities at around peak brightness (Benetti et al. 2004; Wang et al. 2009b). Correlations between the Si II velocity and the host galaxy properties were identified in several studies (Wang et al. 2013; Li et al. 2021). A subset of SNe Ia showed excess emission in the first few days after explosion (Cao et al. 2015; Marion et al. 2016; Hosseinzadeh et al. 2017; Jiang et al. 2017; Wee et al. 2018; Dimitriadis et al. 2019; Li et al. 2019; Shappee et al. 2019; Wang et al. 2020; Jiang et al. 2021; Li et al. 2022). An ultraviolet pulse was detected within 4 days of the explosion of iPTF 14atg (Cao et al. 2015), in line with the theoretical predictions by models (Kasen 2010) involving ejecta–companion interaction in a single-degenerate scenario (see Kromer et al. 2016, for a different view). The thermal radiation from the ejecta–companion star interaction can account for the early blackbody-dominated spectrum without prominent absorption features. Another example is MUSSES 1604D, which was modeled in the context of the explosion of a WD triggered by a helium detonation (Jiang et al. 2017). MUSSES 1604D showed early excess in the red, which contradicts the ejecta–companion star interaction but is consistent with the radiation due to surface radioactivity in a helium detonation scenario. The prominent titanium absorption trough around maximum light of MUSSES 1604D also corroborates the existence of radioactive species synthesized by helium detonation. However, these diverse observational characteristics are usually compromised by the lack of extensive spectral and time coverage. New statistical tools need to be developed to accomplish comprehensive quantification of the observed features.

Optical spectra play an important role in SN cosmology (e.g., Saunders et al. 2018). The classic standardization method is to describe the peak luminosity as a function of the light-curve width and the color at maximum light (Phillips 1993; Perlmutter et al. 1997, 1999; Riess et al. 1996; Tripp 1998; Wang et al. 2003; Guy et al. 2005; Jha et al. 2007; Taylor et al. 2021). The intrinsic diversities among SN Ia properties beyond

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

the stretch and color correction can introduce systematic errors to the measurement of cosmological parameters. It motivates additional terms to be included to better control potential systematic errors. For example, the mass of SN host galaxies has been recognized to be correlated to the Hubble residuals (Kelly et al. 2010; Sullivan et al. 2010). Several studies have been devoted to the quantification of spectral features with applications to the standardization of SNe Ia for distance calibrations (e.g., Wagers et al. 2010; Bailey et al. 2009; Wang et al. 2009b; Foley & Kasen 2011; Chotard et al. 2011; Fakhouri et al. 2015; Zheng et al. 2018; Léget et al. 2020). The ejecta velocity can be used as an additional parameter to reduce the scatter of the Hubble residuals (Wang et al. 2009b, 2013; Zheng et al. 2018). Fakhouri et al. (2015) developed an interesting approach to measure distances using spectroscopic twins of SNe Ia. A data-driven approach to SN Ia spectral analysis may also enable identifications of observational features that are most sensitive to the intrinsic luminosity of SNe Ia and identify potential systematic errors in the cosmological applications of SNe Ia.

Optical spectra are needed to perform *k*-corrections to derive the standardized magnitudes for cosmology. The *k*-corrections rely on the spectral energy distributions (SEDs) of the SNe Ia with well-calibrated photometric and spectroscopic observations, which should cover a broad range of subtypes of SNe Ia. Currently, empirical spectral models with low degrees of freedom (Nugent et al. 2002; Guy et al. 2007; Hsiao et al. 2007; Jha et al. 2007; Burns et al. 2011) are employed for *k*-corrections, although more accurate models with more free parameters have been constructed (Saunders et al. 2018). A data-driven approach to SN Ia observations will naturally lead to spectral libraries that can be used for *k*-corrections of SN Ia spectra.

The next-generation SN surveys with LSST (LSST Science Collaboration et al. 2009) can discover a vast number of SNe, which makes the acquirement of spectral time series for these SNe extremely challenging. It is neither realistic nor cost-effective to trigger high-cadence spectroscopic follow-ups for the SNe discovered by the LSST. For the subset of a posteriori SN Ia discoveries identified from photometric light curves, taking spectral data at multiple phases may become impossible for the majority of the SNe. However, in light of the rapidly growing spectroscopic data set of nearby SNe, it is promising that a generative data-driven model for spectral inference via machine-learning techniques can be derived to mitigate these difficulties.

There are some existing studies applying machine learning to transient studies. For example, the spectral types of the SNe can be classified based on their light-curve data (Möller et al. 2016; Muthukrishna et al. 2019a; Takahashi et al. 2020; Villar et al. 2020), and transients can be identified from the astronomical survey images (Goldstein et al. 2015; Mahabal et al. 2019; Gómez et al. 2020). The light curves of SNe Ia can be well modeled by functional principal component analysis (FPCA; He et al. 2018), where it was shown remarkably that a set of FPCA eigenvectors that are independent of the photometric filters can be derived from the observed light curves of SNe Ia. There are a few studies of the application of deep learning neural networks to the spectral data of SNe. For example, Muthukrishna et al. (2019b) used a convolutional neural network (CNN) for automated SN type classification based on SN spectra. Several other works (Chen et al. 2020; Vogl et al. 2020; Kerzendorf et al. 2021) applied a Gaussian process, principal component analysis (PCA), and deep learning neural networks to radiative transfer models of SNe. Sasdelli et al. (2016) used unsupervised learning algorithms to investigate the subtypes of SNe Ia. Stahl et al. (2020) developed neural networks to predict the photometric properties of SNe Ia (phase and Δm_{15}) based on spectroscopic data. Saunders et al. (2018) used PCA to find low dimensional representations of the spectral sequences of 140 well-observed SNe Ia. Chen et al. (2020), in particular, built an artificial intelligence assisted inversion (AIAI) of radiative transfer models and used that to link the observed SN spectra with theoretical models. The AIAI is able to retrieve the elemental abundances and density and temperature profiles from observed SN spectra. The AIAI approach has the potential for quantitatively coupling complex theoretical models with the ever-increasing amount of high-quality observational data.

This paper aims to build a data-driven model of the spectral time evolution of SNe Ia using long short-term memory (LSTM) neural networks (Hochreiter & Schmidhuber 1997). Section 2 presents the data sample. Section 3 shows the preprocessing of the data to bring the data set to a uniform standard for further processing using FPCA (Section 4). The neural network architecture and the assignment of statistical weights of the data are shown in Section 5. Section 6 shows the application of the neural network to the construction of spectral time sequences of SNe Ia. Section 7 presents the application of the neural network based on spectral data taken at a single epoch around optical maximum to the analyses of normal and high velocities (Section 7.1) and the different subtypes of SNe Ia (Section 7.2) and the application of the neural network based on spectral data taken at two epochs around optical maximum (Section 7.3). Moreover, we explore the potential application of our method in predicting spectral phases in Section 7.4. Section 8 gives the discussions and conclusions.

A `Python` implementation of the method proposed in this paper is available on Github.[4] The software allows users to apply the LSTM neural networks to their own observations and provide open access to the data sample presented in Section 2 and the spectral templates of 361 SNe Ia constructed in Section 6.

2. Sample Selection

We use the publicly available SN spectra from WISeREP (Yaron & Gal-Yam 2012), Kaepora (Siebert et al. 2019), and a data set of high-quality Very Large Telescope (VLT) observations of SNe Ia from the SN polarimetry program (Wang & Wheeler 2008; Cikota et al. 2019; Yang et al. 2020). The SNe are selected with the following criteria.

1. The redshift of the host galaxy and the *B*-band maximum are accurately measured.
2. The SN is classified as one of the following five subtypes: Ia-norm, Ia-91T, Ia-91bg, Ia-99aa, and Iax.
3. The spectrum of the SN is between −15 and 33 days relative to the *B*-band maximum, and the wavelength covers 3800–7200 Å in the rest frame.
4. The SN has more than two distinct spectra.

---

[4] https://github.com/thomasvrussell/snail

The above distilling criteria lead to 3091 spectra from 361 SNe Ia. Moreover, we collected the published B and V light curves of these SNe for spectrophotometric recalibration of the spectral data. In Table 1, we show the details of the selected SNe used in this study.

The characteristics of the SNe included in this work are shown in Figure 1. The vast majority of them reached maximum light during 1995 and 2015, with 95% of them originating from $z < 0.05$, and nearly half of the SNe have $\leqslant$five observations. About a quarter of the sample have temporal sampling with >10 observations. A few nearby SNe, such as SN 2011fe, have >50 spectroscopic measurements.

The distribution of the source of the spectra in this compilation and the phase from B-band maximum are shown in Figure 2. The sample consists of observations from four SN programs: the program from the Harvard-Smithsonian Center for Astrophysics (Blondin et al. 2012), the Berkeley SuperNova Ia Program (Silverman et al. 2012), the Carnegie Supernova Program (Folatelli et al. 2013), and the Supernova Polarimetry Program (Wang & Wheeler 2008; Cikota et al. 2019; Yang et al. 2020), thereby yielding >70% spectra observed by the FLWO 1.5 m, Lick 3 m, and LCO duPont telescopes. The VLT and Keck also provide a substantial portion of spectroscopic data, mostly with high signal-to-noise ratios (S/Ns). One can also see in Figure 2 that spectroscopic time coverage peaks around maximum light, and there is a deficiency of SNe observed at the infancy stages (<−10 days).

In addition to the above data set for the construction of a data-driven predictive model, we have also generated an auxiliary data set with the following criteria:

1. The redshift of the host galaxy is accurately measured.
2. The SN is classified as an SN Ia.
3. The spectral wavelength covers 3800–7200 Å in the rest frame.

These less restrictive criteria yield 8501 spectra from 3536 SNe Ia. Throughout this paper, we will refer to the smaller data set as "the data set" and use the term "the extended data set" to denote the larger data set.

3. Preprocessing of the Data

The sources of the spectral data are heterogeneous and often not well calibrated photometrically. We evaluated the noise levels of the spectral data (Section 3.1), which were used to assign statistical weightings to each spectrum. Moreover, the spectral data were processed through the following four steps to ensure the self-consistency and uniformity of the data set: deredshifting (Section 3.2), smoothing and resampling (Section 3.3), removal of telluric lines (Section 3.4), and recalibration of spectral flux levels (Section 3.5).

3.1. Estimation of the Spectral Noise

The data are of diverse S/N, and their statistical weights need to be approximately accounted for in our studies. An approximate error spectrum for each original spectrum in the extended data set is constructed following the method introduced in the Kaepora database (Siebert et al. 2019). We further calculate the S/N of each spectrum in the wavelength range 4800–6200 Å and use this as a measure of the quality of each spectrum (see Section 5 for more details on the assignments of statistical weights).

3.2. Deredshifting

The original spectra in the extended data set were deredshifted from the observer frame to the rest frame using the host galaxy redshifts. In some cases, a correction had already been applied to the data we have downloaded. To avoid double correction, we visually inspected each spectrum to ensure that the wavelengths of the telluric absorption lines at 6867–6884 and 7594–7621 Å were consistent with the values expected in the observer's frame before deredshifting. We also compared the spectra with other spectra of the same SN at similar phases to confirm that the spectra had not been deredshifted in their source database. We found 185 spectra in our data set showing conspicuous redshift inconsistencies, and they were included in our analyses after removing the inconsistencies.

3.3. Smooth and Rebinning

The next step was to apply a two-order Savitzky–Golay filter (Savitzky & Golay 1964) to suppress the ubiquitous random noise in the SN spectra. Here we adopted a smoothing width of 1000 km s^{-1}, which is broader than the high-frequency noise but significantly smaller than the typical spectral absorption features. Each deredshifted spectrum from the extended data set was resampled to a wavelength grid with a bin size of 2 Å from 3800 and 7200 Å by linear interpolation and normalized by dividing its average flux. For clarity, these processed spectra are hereafter referred to as the homogenized spectral data set (HSD), which contains a total of 8501 spectra. Each homogenized spectrum is a 1700 dimensional array with unit mean.

3.4. Removal of Telluric Features and Galaxy Emission Lines

We removed the telluric absorptions at 7605 and 6869 Å and all of the conspicuous hydrogen Balmer lines originating from the host galaxies at 6565 and 4861 Å for each homogenized spectrum. The wavelength regions affected by the absorption and emission lines were filled with values from second-order B-spline interpolations. Note that the interpolation was not directly performed on the homogenized spectra but rather on the homogenized spectra already smoothed by an inverse-variance Gaussian smooth algorithm (Blondin et al. 2006; Siebert et al. 2019) to ensure numerical stability.

3.5. Spectral Flux Recalibration

Both the continuum and spectral line components are important in our study, since our goal is to develop a model that enables predictions of both the spectral features and the overall spectral energy density distribution. However, the spectral fluxes and colors integrated from SN spectra are usually inconsistent with those derived from broadband photometries. Such inconsistency is mainly due to the technical difficulties in spectrophotometry and can be conspicuous in many spectra observed at the same phases by different telescopes.

The flux scales of the preprocessed homogenized spectra were recalibrated to eliminate the flux scale inconsistency by enforcing the $B - V$ colors integrated from the spectra to agree with their corresponding photometric observations. For this to work, we must have a sample of SNe Ia with excellent multicolor light-curve coverage. We went through a

Table 1
Table of SNe in the Spectral Data Set

SN Name	SN Subtype	Redshift	Number of Spectra	First Epoch	Last Epoch	Spectrum Source	Redshift Reference	MJD$_{\mathrm{max}}$ Reference	Photometry Reference
SN 2012fr	Ia-norm	0.005457	74	−13.99	+30.45	WISeREP	1	2	2
SN 2005cf	Ia-norm	0.006461	73	−12.40	+29.29	WISeREP, Kaepora, VLT	1	3	4
SN 2011fe	Ia-norm	0.000804	68	−14.82	+27.51	WISeREP, Kaepora	1	5	6, 7
SN 2002bo	Ia-norm	0.004240	44	−13.58	+29.36	WISeREP, Kaepora	1	3	8
SN 2006X	Ia-norm	0.005240	40	−10.68	+32.07	WISeREP, Kaepora, VLT	1	3	9
SN 1994D	Ia-norm	0.002058	38	−12.47	+29.44	Kaepora	1	3	10, 11
SN 2007le	Ia-norm	0.006721	36	−10.63	+23.33	WISeREP, Kaepora	1	3	9
SN 2003du	Ia-norm	0.006384	35	−12.70	+32.98	Kaepora	1	3	12
SN 2004dt	Ia-norm	0.019730	32	−9.44	+32.72	WISeREP, Kaepora, VLT	1	3	8
SN 2003cg	Ia-norm	0.004130	32	−7.48	+26.67	Kaepora	1	3	8
SN 2001V	Ia-norm	0.015018	27	−13.20	+27.97	WISeREP, Kaepora	1	3	13
SN 2007af	Ia-norm	0.005464	27	−5.60	+32.07	WISeREP, Kaepora	1	3	9
SN 2002er	Ia-norm	0.008569	26	−9.94	+32.66	WISeREP, Kaepora	1	3	8

Note. References: (1) Siebert et al. (2019); (2) Contreras et al. (2018); (3) Blondin et al. (2012); (4) Wang et al. (2009a); (5) Parrent et al. (2012); (6) Stahl et al. (2019); (7) Tsvetkov et al. (2013); (8) Ganeshalingam et al. (2013); (9) Stritzinger et al. (2011); (10) Gomez et al. (1996); (11) Richmond et al. (1995); (12) Stanishev et al. (2007); (13) Hicken et al. (2009); (14) Pan et al. (2015); (15) Jha et al. (2006); (16) Gall et al. (2018); (17) Yaron & Gal-Yam (2012); (18) Maguire et al. (2013); (19) Walker et al. (2015); (20) Brown (2014); (21) Cikota & Pauldrach (2018); (22) Brown et al. (2014); (23) Krisciunas et al. (2017b); (24) Stritzinger et al. (2014); (25) Folatelli et al. (2013); (26) Srivastav et al. (2016); (27) Amanullah et al. (2015); (28) Yamanaka et al. (2015); (29) Hicken et al. (2012); (30) Silverman et al. (2012); (31) Stritzinger et al. (2002); (32) Hachinger et al. (2013); (33) Srivastav et al. (2017); (34) Li et al. (2018); (35) Shappee et al. (2016); (36) Smitka et al. (2016); (37) Krisciunas et al. (2017a); (38) Weyant et al. (2018); (39) Sandage et al. (1995); (40) Ardeberg & de Groot (1973); (41) van Genderen (1975); (42) Lira et al. (1998); (43) Krisciunas et al. (2006); (44) Smitka et al. (2015); (45) Leibundgut et al. (1993); (46) Cikota et al. (2019); (47) Krisciunas et al. (2003); (48) Friedman et al. (2015); (49) Krisciunas et al. (2004); (50) Graham et al. (2017); (51) Riess et al. (2009); (52) Maguire et al. (2011); (53) Danziger et al. (1998); (54) Marion et al. (2016); (55) Yamanaka et al. (2014); (56) Firth et al. (2015); (57) Wee et al. (2018); (58) Cartier et al. (2017).

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

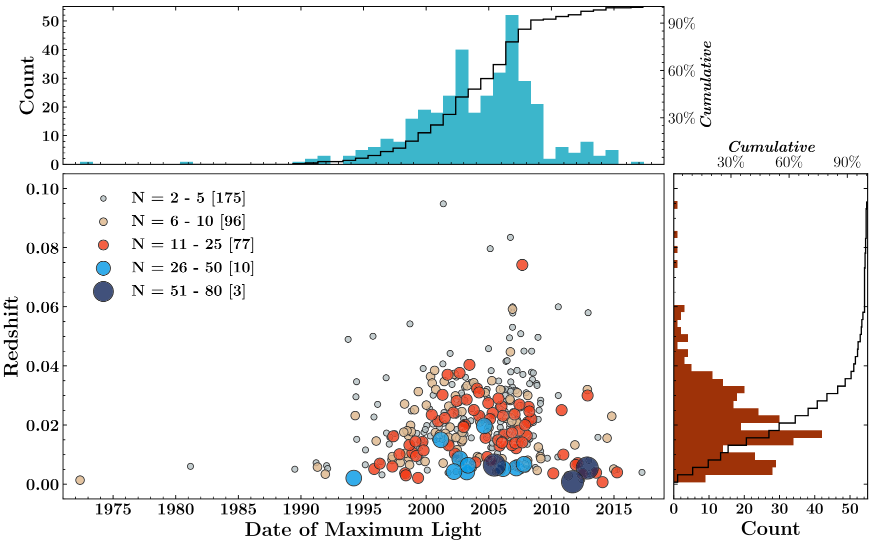


Figure 1. The SN redshifts versus their times of B-band maximum for all SNe Ia in our data set. These objects are divided into five subgroups according to the number of spectroscopic observations, where each subgroup is assigned a specific color and marker size, as shown in the upper left corner. The letter N in the legend stands for the number of spectra, and the numbers in the square brackets refer to the total counts of the subgroups. The attached panels at the top and right are the histograms of the times of maximum and host galaxy redshifts, respectively, with black curves showing their cumulative distributions.

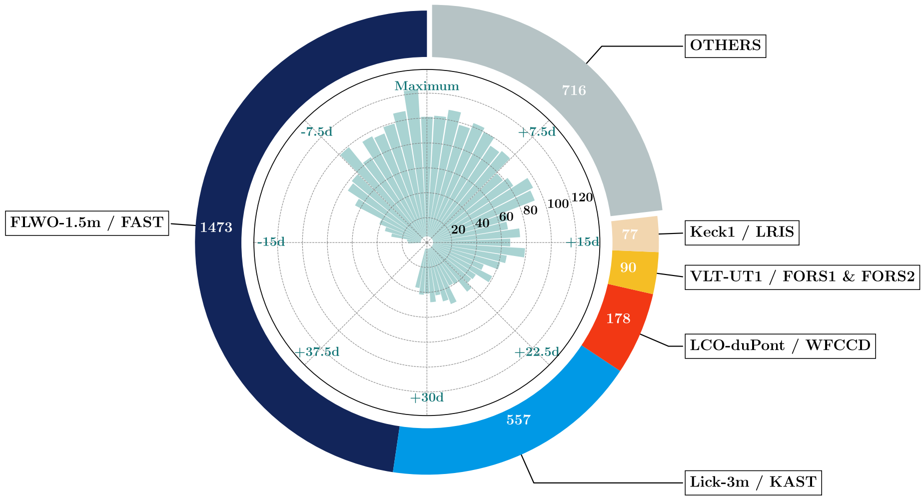


Figure 2. Outer: contributions of each major instrument to the total number of spectra in the data set. Inner: histogram of the number of individual SN spectra at each epoch in the data set.

comprehensive literature search for the photometry of all of these SNe and collected published light curves of all of them, together with the specific filter bands at which they were observed. The sources of the light curves are shown in Table 1. For the data that are published in the natural system, the spectral flux level corrections were made with the appropriate transmission curves downloaded from the sources of the original data.

Fortunately, in general, the SNe with more spectral coverage also had more complete photometric coverage. Over 80% of the SNe with more than four optical spectra were found to have excellent light-curve coverage to allow for detailed template fits (He et al. 2018) or interpolations to their light curves. We used the Gaussian process with a radial-basis function kernel (`scikit-learn`; Pedregosa et al. 2011), a model-independent interpolation method, to fit the B and V light curves and hence derive the $B-V$ colors at the epochs of the spectroscopic observations. Subsequently, we should adjust the original spectrum so that its synthetic $B-V$ color can be in line with the derived photometric $B-V$ color, as the flux levels of the available spectra are usually poorly calibrated. One straightforward approach is to multiply each original spectrum by a monotonic flux scaling function with low degrees of freedom such that the adjusted spectrum will have a $B-V$ color that is consistent with the corresponding photometric measurement at the same epoch. The specific choice of the function can be somewhat arbitrary. In our work, we adopt the functional form of the CCM89 extinction law (Cardelli et al. 1989) to perform the flux scaling, with the parameter R_V fixed to 3.1 but leaving $E(B-V)$ as the only fitting parameter. All of the adjusted spectra thus have $B-V$ colors that are identical to the values on their corresponding photometric color curves.

The homogenized spectra, starting from 3800 Å in the rest frame, do not fully cover the entire effective wavelength range of the B band. The missing data in the homogenized spectra were set to zero first when calculating the B-band magnitudes. This introduces a systematic error to each $B-V$ color integrated from a spectrum that was corrected by employing a spectral template of SNe Ia (Hsiao et al. 2007) at the nearest phase. In doing so, the spectra of Hsiao et al. (2007) were truncated by setting the fluxes outside of 3800–7200 Å to zero. The $B-V$ colors of both the Hsiao template (covering 1000–25,000 Å) and its truncates were calculated, and the differences were taken as their approximate systematic offsets. Note that this offset is not merely a function of the phase but also a function of the redshift of the SN and the specific transmission curves used for the corresponding photometric data.

The fit has zero degrees of freedom and results in a precise match of $B-V$ color between a color-calibrated spectrum and its corresponding photometric data. Such a treatment of the data may introduce systematic uncertainties that are difficult to quantify. We defined a quantity $|\Delta_{B-V}|$, which is the absolute value of the difference between the $B-V$ color measured on an uncalibrated spectrum and its corresponding observed photometric color to account for the amount of the color correction. Presumably, larger values of $|\Delta_{B-V}|$ imply higher levels of uncertainties. We used this quantity to set the statistical weightings of each spectrum in training the neural networks (see Section 5).

There are 801 spectra ($\sim 1/4$) in the data set for which the above $B-V$ calibration cannot be applied due to the lack of sufficient photometric coverage. This could be either due to unavailable photometric data or because the phases of the spectra are beyond the limited photometric data coverage. No photometric corrections were applied to these spectra, but their weightings are lowered in the training of the neural network (see Section 5 for details).

Each spectrum was then renormalized by its average flux across the wavelength range of 3800–7200 Å. This forms a new spectral data set, which we hereafter refer to as the corrected spectral data set (CSD). There are a total of 3091 corrected spectra in the CSD out of a total of 8501 spectra of the extended spectral data set.

4. FPCA Parameterization

The FPCA (Hall et al. 2006) was utilized to reduce the spectral dimensionality of the spectral data. It was applied to construct light-curve templates of SNe Ia and build Hubble diagrams using nearby, well-observed SNe (He et al. 2018), and it was adopted by Kou et al. (2020) to parameterize the spectra of SNe Ia. Here we follow a similar approach and use the `fpca` package in the R language (Peng & Paul 2009) to solve for the optimal set of FPCA solutions.

The FPCA algorithm uses a series of orthogonal functions as the principal components and a linear combination of these principal components to reconstruct the input data set. The function $\psi(\lambda)$ for the reconstruction can be written as

$$\psi_i(\lambda) = \mu(\lambda) + \sum_{n=1}^{N} x_{i,n}\phi_n(\lambda), \tag{1}$$

where $\mu(\lambda)$ is the average function, $\phi_n(\lambda)$ is the nth-order component in the form of a function, and $x_{i,n}$ is the nth-order FPCA score for the ith spectrum. To find the best FPCA score series $x_{i,n}$, the mean squared error (MSE) between the input data $Y_i(\lambda)$ and the fitting function $\psi(\lambda)$ is minimized,

$$\mathrm{MSE}_i = \int_{\lambda_0}^{\lambda_1} (Y_i(\lambda) - \psi_i(\lambda))^2 d\lambda, \tag{2}$$

where λ_0 and λ_1 are the lower and upper limits of the wavelength range. The principal components $\phi_n(\lambda)$ and the average function $\mu(\lambda)$ are solved for a given spectral data set of size N. The average function is

$$\mu(\lambda) = \frac{1}{N}\sum_{n=1}^{N} Y_n(\lambda), \tag{3}$$

and the principal components are solved by maximizing the variance of FPCA scores over the input spectral data set,

$$\phi_i(\lambda) = argmax(\mathrm{Var}(x_i)), \tag{4}$$

where $\mathrm{Var}(x_i)$ is the variance of all of the ith FPCA scores of the input data, $x_{i,1}$, $x_{i,2}$, ..., $x_{i,N}$. The solutions to the principal components are subject to two additional conditions. First, the principal components are orthogonal:

$$\int_{\lambda_0}^{\lambda_1} \phi_i(\lambda)\phi_j(\lambda)d\lambda = \delta_{ij}. \tag{5}$$

Second, the variances are ordered with the order of principal components by

$$\forall\ i,\ \mathrm{Var}(x_i) > \mathrm{Var}(x_{i+1}). \tag{6}$$

The solutions to the FPCA were derived from the more diverse extended data set to achieve maximum generalizability. Given the high computational cost (mostly RAM limitations) of `fpca`, 500 homogenized spectra were randomly drawn from the HSD to solve for the FPCA principal components, and each corrected spectrum in the CSD was decomposed into the resulting basis functions. The spectral data were divided into two sections: a blue section covering the wavelength range 3800–5500 Å and a red section covering 5500–7200 Å. At a spectral resolution of 2 Å pixel^{-1}, each spectral section has a total of 850 pixels. Each spectral section was then subtracted by its own average flux and subsequently divided by its standard deviation. We used 90 FPCA principal components for each spectral section. As a result, the full spectrum with two sections (1700 pixels) could be reconstructed by 180 FPCA scores and four additional factors accounting for the average fluxes and standard deviations of the two spectral sections. We denote the combined array that concatenates 180 FPCA scores and the four additional parameters as $[V_{\mathrm{FPCA}}]$ (hereafter the FPCA-encoded array). The FPCA parameterization for each spectral section was performed separately. This could introduce some artifacts at the wavelength boundary around 5500 Å, but the effect was not significant enough to affect the analyses in this paper. Applying FPCA on the full range (5500–7200 Å) of spectra without division over wavelength can eliminate these artifacts at the boundary. However, the corresponding RAM demand for solving FPCA with a satisfactory reconstruction accuracy exceeds the affordable level of our current computation platform.

The mean absolute percentage error (MAPE) is used to evaluate the difference between a corrected spectrum and its FPCA reconstruction. As shown in Figure 3, where the statistics are drawn from the reconstruction of the CSD, the errors are considerably smaller when more principal components are used. A closer look at the trend (see the lower panel of Figure 3) shows that the improvement of reconstruction accuracy decelerates as the number of components increases. The decrease of the median MAPE is larger than 1% from 5 to 10 components but falls to less than 0.01% from 85 to 90

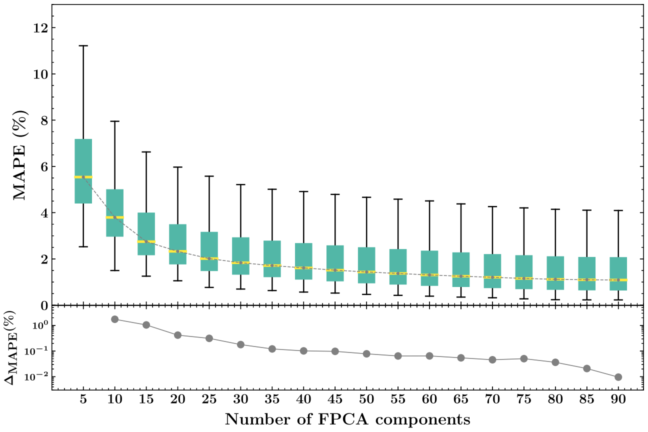


Figure 3. Accuracy of applying FPCA to the corrected spectra in the data set. The upper panel gives the MAPE of FPCA reconstruction as a function of the number of used FPCA components. The box plot shows the distribution of MAPE for all of the corrected spectra. Note that the box is drawn from the first quartile (Q1) to the third quartile (Q3), with a horizontal yellow line to denote the median; in addition, the lower (upper) whisker is at the lowest (highest) datum above Q1 − 1.5 × IQR (below Q3 + 1.5 × IQR), where IQR = Q3 − Q1. The dashed gray curve indicates a monotonically decreasing trend of the median MAPE as more FPCA components are employed. The lower panel shows the decrease of the median-level MAPE (i.e., overall accuracy improvement) for every five additional FPCA components used.

components. The median MAPE of the FPCA reconstruction over the CSD can reach ∼1.1% using 90 components. Note that MAPE is a pixel-by-pixel measurement without taking into account the fluctuations due to observational noise. Thus, a spectrum with low S/N is more likely to have a larger MAPE. This trend is partially responsible for the broad and skewed distribution of MAPE, shown in the box plot of Figure 3. Figure 4 demonstrates the performance of FPCA reconstruction by six representative examples selected from the CSD. These examples show excellent agreement across the prominent spectral features of SNe Ia.

5. Network Architecture and Sample Weights

The proposed model is built upon a multilayer LSTM (Hochreiter & Schmidhuber 1997) neural network, a widely used subclass of recurrent neural network (RNN). Previous works like Stahl et al. (2020) and Chen et al. (2020) adopted CNNs to analyze the spectral data of SNe Ia, where one-dimensional convolution processes are applied to the wavelength axis. In contrast, the LSTM architecture in our work is not convolutional. The spectral data have been compressed into isolated FPCA-encoded arrays, and additional dimensions are needed to store the phase information required for training the neural network. Therefore, the input data cannot maintain a structure with a concept of spatial correlation as in the original spectra, thus becoming incompatible with the CNN approach.

Like all other subtypes of RNN, LSTM has a chain-like structure with a repeating module that allows the algorithm to learn from spectral time sequences. Figure 5 shows the repeating module in our LSTM architecture. We will always use the subscript "*t*" to denote the index of a spectral time sequence but the subscript "tar" to indicate an attribute of the target spectrum. The LSTM is characterized by four gates: an input gate (***i***), a forget gate (***f***), an output gate (***o***), and an input modulation gate (***g***). These gates conduct the following operations:

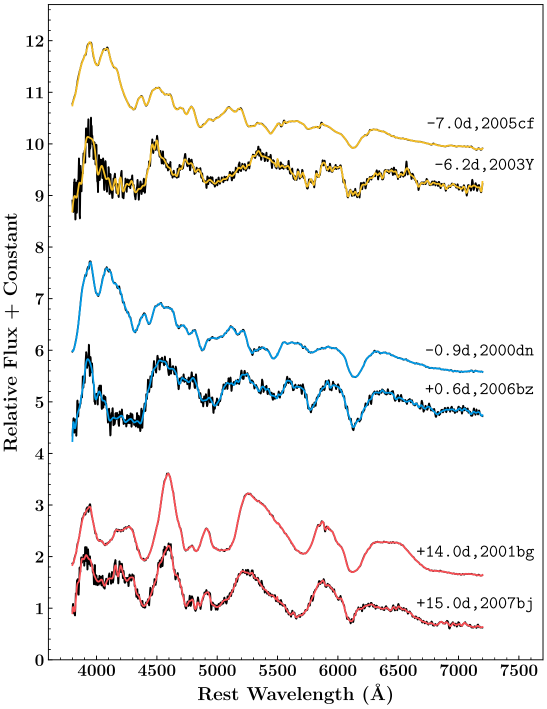


Figure 4. The FPCA reconstruction of six representative spectra of the CSD. The FPCA reconstructions of the corrected spectra (solid black curves) are plotted as colored solid lines. The six spectra are selected from three observational epochs: ∼1 week before maximum, maximum light, and ∼2 weeks past maximum. Note that a relatively noisy spectrum is presented with a spectrum of very high S/N for each epoch. The spectra are arbitrarily shifted in the vertical direction for clarity of display.

$$\begin{aligned} i &= \mathbf{sigm}(h_{t-1}U_i + x_t W_i),\\ f &= \mathbf{sigm}(h_{t-1}U_f + x_t W_f),\\ o &= \mathbf{sigm}(h_{t-1}U_o + x_t W_o),\\ g &= \mathbf{tanh}(h_{t-1}U_g + x_t W_g), \end{aligned} \tag{7}$$

and

$$c_t = f \circ c_{t-1} + i \circ g \qquad h_t = o \circ tanh(c_t), \tag{8}$$

where sigm (tanh) is the sigmoid (hyperbolic tangent) activation function and the circle symbol refers to an element-wise product. Here x_t is the input data delivered into the module, and the internal state c_t (h_t) is known as the cell (hidden) state at time step t. The weight matrices $\boldsymbol{w} = \{W_i, U_i, W_f, U_f, W_o, U_o, W_g, U_g\}$ are repeatedly used by each time step.

Our study aims to predict the spectrum at a specific target phase $p_{\rm tar}$ by feeding a spectral sequence with arbitrary time sampling. However, for typical discrete-time dynamic systems, such as in text classification (Dai & Le 2015) and stock price prediction (Eapen et al. 2019), LSTM was applied to ordered data without time labels or with constant time-sampling rates. Observations of SNe Ia are usually irregularly time-sampled. Consequently, the proposed predictive model needs to include the specific phases of the spectra as part of the input parameters, yet LSTM does not contain a channel in its structure to incorporate the spectral phases. A variant LSTM known as phased LSTM, which is proposed to handle unevenly sampled time series by adding a new time gate (Neil et al. 2016), is a likely choice. Unfortunately, the target phase $p_{\rm tar}$ of

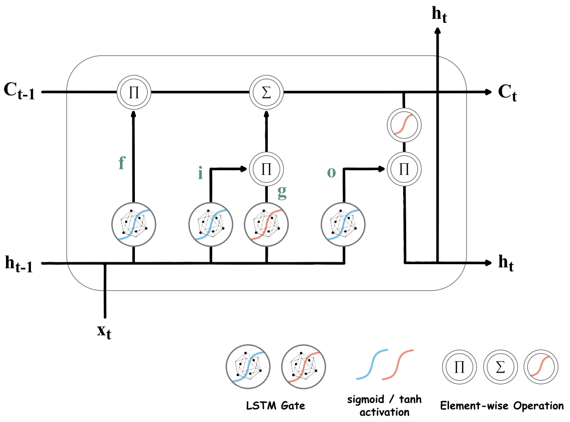


Figure 5. Repeating module in LSTM that contains four interacting layers. The definitions of the nodes in the module are shown at the bottom of the figure. The detailed calculations in the four LSTM gates are given in Equation (7). The element-wise operations are described in Equation (8).

the predicted spectrum is also a variable in our framework, making it infeasible to apply phased LSTM directly.

Our solution to this problem is a straightforward integration of the phase information into the input spectral data. Recall that each corrected spectrum has been compressed and encoded into a much shorter array denoted as $[V_{\rm FPCA}]$. We concatenated the phase of each input spectrum and the target phase for spectral prediction at the beginning of the FPCA-encoded array, namely, the input data of the first LSTM layer at the tth time step $x_t = (p_{\rm tar}, p_t, [V_{\rm FPCA}]_t)$. As a result, the LSTM acquires an input layer with 186 neural processing units.

Figure 6 shows the architecture of the proposed LSTM implemented using the `Python` module `keras` (Chollet et al. 2015). For each SN, the input spectra are a given number of K spectra from a total of L spectra, allowing repetitions but with the phases only in nondecreasing order. The neural network contains three bidirectional LSTM layers (Schuster & Paliwal 1997) and a fully connected output layer. The bidirectional LSTM allows the neural network to learn from the time-sequence data in reverse order and strengthens the robustness of the neural network. Note that the last layer is time-distributed (`keras.TimeDistributed`), thereby yielding an output for each time step. The difference between the output of each time step and the FPCA-encoded array of the target spectrum contributes equally to the loss function during the training process. The Nadam algorithm (`keras.Nadam`) is used to optimize the network, with the loss function being the MSE weighted by the data quality (see Equations (9)–(11) and discussions below for details). The last box shown in Figure 6 reconstructs the spectrum at the target phase during the predicting process; the outputs from different time steps were averaged to generate the predicted spectrum via FPCA reconstruction, followed by a final flux normalization.

To alleviate overfitting in the training process, we adopted the commonly used dropout method (Hinton et al. 2012; Zaremba et al. 2014). The efficient regularization technique suppresses the coadaptations among the neurons by stochastically dropping rows of weight matrices. The dropout can also be activated in the testing step (aka Monte Carlo dropout) to make the network probabilistic (Gal & Ghahramani 2015, 2016). With Monte Carlo dropout, the model is thereby no longer deterministic after the training process. A forward pass can generate a different result by feeding the same input, which is now determined by the applied dropout mask as a realization of a Bernoulli process. In this scenario, one can estimate the model uncertainty through the variance of predictions from multiple forward passes. In our study, we set the kernel dropout rate (namely, for W matrices in Equation (7)) to 0.14 and the recurrent dropout rate (namely, for U matrices in Equation (7)) to 0.16, respectively. The only exception is that we disabled the kernel dropout for the first LSTM layer, and no dropout is performed on the final fully connected layer.

Conceptually, a dynamic number of time sequence K is allowed for RNNs; this is also our initial motivation to adopt LSTM to handle spectral sequences with arbitrary time coverages. However, the complexity of training the neural network increases drastically with K due to the vast amount of spectral combinations for large K. Our current model only supports $K = 2$, i.e., a spectral pair (with nondecreasing phases) as input. In particular, the structure also allows predictions from a single spectrum through one-time duplication. Moreover, following the selection limit on the phases of the SNe in our data set, the target phase is restricted to be in the range from -15 to $+33$ days.

To mitigate the effect of the heterogeneity of the spectral data, each spectrum in the CSD was assigned an approximate statistical weight $w_{\rm spec}$ based on the S/N of that spectrum defined in Section 3.1 and a formula that is a function of $|\Delta_{B-V}|$ defined in Section 3.5. The weight $w_{\rm spec}$ is expected to be positively correlated with the S/N, which approximately measures the pixel-to-pixel fluctuations due to shot noise, and negatively correlated to $|\Delta_{B-V}|$. In the practical implementation of the neutral networks, the weight $w_{\rm spec}$ for each spectrum in the CSD is given as follows:

$$
\begin{aligned}
w_{\rm spec} &= w_{\rm S/N} {}^{*} w_{\rm color}\\
w_{\rm S/N} &= 0.7 + 0.3^{*}\min\{{\rm S/N}/100,\, 1\}\\
w_{\rm color} &= 0.7 + 0.3\\
&\quad \times {}^{*}(10/7 - \min\{\max\{|\Delta_{B-V}|/\sigma_{\rm GP},\, 3\},\, 10\}/7),
\end{aligned} \tag{9}
$$

where $\sigma_{\rm GP}$ denotes the photometric error of the color $B - V$, which is given by $\sqrt{(\sigma_B^2 + \sigma_V^2)}$, with σ_B and σ_V being the photometric errors of the B and V bands, respectively. About three-fourths of the spectra in the CSD had been calibrated by their corresponding photometric colors for which the $|\Delta_{B-V}|/\sigma_{\rm GP}$ values could be computed. For the rest of the spectra, the ratio $|\Delta_{B-V}|/\sigma_{\rm GP}$ was set to 5 artificially. Notice that $w_{\rm S/N}$ and $w_{\rm color}$ were restricted to values in the range from 0.7 to 1.0 such that their weightings do not differ drastically for data with very different S/N and $|\Delta_{B-V}|$.

Finally, the statistical weight for a training sample containing a total of $K + 1$ spectra from the CSD, i.e., K input spectra together with the output spectrum, was constructed by multiplying the $K + 1$ spectral weights,

$$
w_{\rm sample} = \alpha * \prod w_{\rm spec}, \tag{10}
$$

where the factor α was introduced as a penalty factor to account for the heterogeneous instrumental sources of the $K + 1$ spectra, which was set to 1 if all of the $K + 1$ spectra are observed by the same instrument (i.e., no penalty), 0.6 if two instruments are involved, and 0.4 if more than two instruments are used. Similar to the weights used for individual spectra, the penalty factors used herein avoid making extreme sample

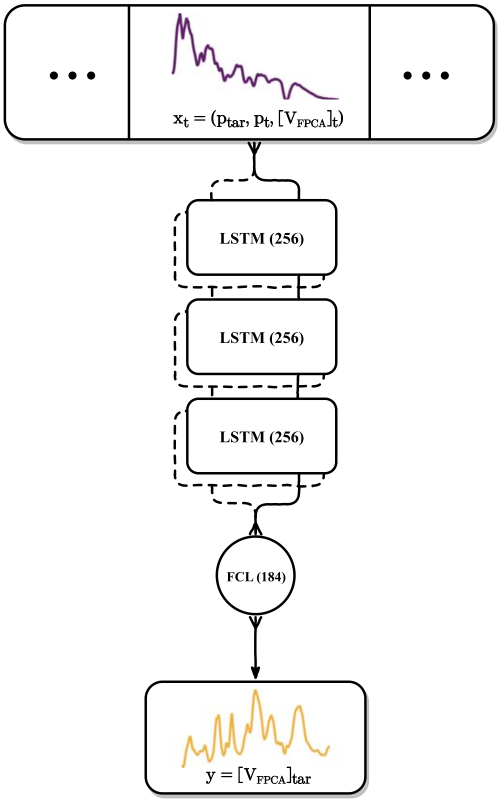


Figure 6. Architecture of the proposed LSTM model. The input data (shown by the boxes at the top) are one or more one-dimensional arrays describing the target phase for which the neural network will make spectral predictions (p_{tar}), the phase of the input spectral data (p_t), and the FPCA-encoded array of the tth spectral data ($[V_{\mathrm{FPCA}}]_t$). The three boxes labeled with LSTM refer to the repeating module described in Figure 5, where each layer in the module comprises 256 neurons. Note that the dashed lines indicate bidirectional operations that make the proposed architecture symmetric to the sequence inversion of the input data. A fully connected layer (FCL) follows the last LSTM module and converts the 256 neural processing units back to the 180 FPCA scores and the four scaling parameters. The last step is a reconstruction of the spectroscopic data for the phase t_{tar} during the predicting process. Note that the architecture itself can accommodate any variable-length spectral time series as input, but our model used herein currently supports a fixed length of 2.

weights. The loss function of the neural networks is computed as

$$\mathcal{L} = w_{\mathrm{sample}} * \sum_{i,t} (Y_t^i - [V_{\mathrm{FPCA}}]_{\mathrm{tar}}^i)^2, \qquad (11)$$

where Y_t is the neural network output at time step t, and $[V_{\mathrm{FPCA}}]_{\mathrm{tar}}$ is the FPCA-encoded array of the target spectrum. The superscript i denotes the dimension index.

6. LSTM Applied to the Construction Spectral Templates

In this section, we apply the proposed neural networks to construct the spectral template for each of the 361 SNe Ia in our data set. Here a spectral template is defined as a spectral time sequence from −15 to +33 days with a constant cadence of 3 hr. The goal of the neural networks is to enable accurate predictions of the time evolution of spectral features based on spectral observations of limited spectral time coverage. In this section, the accuracy of the template construction using LSTM is evaluated using a test spectral set. Its performance is also compared with other published spectral templates of SNe Ia.

6.1. Training the Neural Networks for Spectral Template Constructions

We split the CSD into a training set of 90% (2782) spectra and a testing set of 10% (309) spectra, where the testing set was randomly drawn from the CSD with the following restrictions.

1. The S/N of the spectral data is higher than 15.
2. The phase difference between the selected spectrum and the nearest spectral observation of the same SN in the CSD is larger than two-thirds of a day. This avoids the selection of multiple spectra of the same SN taken on almost the same night.
3. The testing set does not contain all of the observed spectra of any SN in the CSD, as is required by the algorithm of spectral template construction shown in Algorithm 1.

The remaining spectra of the CSD after the selection of the test data set were employed to train an LSTM model for the spectral template constructions using the framework shown in Figure 6. Due to the simplicity of the proposed LSTM architecture (Figure 6) and the limited sample size, no validation set is created to fine-tune the hyperparameters of the LSTM neural networks.

In the algorithm, a training sample consists of $K + 1$ training spectra of a particular SN with their corresponding phases. The first K spectra form a time sequence with phases in nondecreasing order as neural network input, and the last spectrum is the target spectrum for the neural network prediction. Repetitions in the $K + 1$ spectra are allowed. The training samples are generated by exhausting all possible permutations with repetition of the $K + 1$ spectra over the training set.

6.2. Construct the Spectral Templates

The construction of the spectral templates follows the procedures given in Algorithm 1. The LSTM model obtained over the training spectra was used as the input model. A spectral template for each SN in the CSD was constructed by feeding the training spectra of the SN into the trained LSTM model. Throughout this study, K is set to 2, allowing spectral sequences at given target phases to be constructed using two spectra. It also allows for the special cases that the two spectra are identical; i.e., the predictions can be triggered by only one spectrum. The restriction on K makes it infeasible to apply the trained LSTM model directly when an SN has more than two training spectra. Nevertheless, Algorithm 1 provides a general strategy to construct spectral templates for all SNe in the CSD.

Algorithm 1. Spectral Template Construction

```
input: The trained LSTM model
input: The spectral training set: $\mathbf{X}$
define the set of ordered spectral pairs for one SN $I_0 \leftarrow$ all $(\mathbf{x}_l, \mathbf{x}_m)$ with
  $p_{\mathbf{x}_l} \leqslant p_{\mathbf{x}_m}$, where $\mathbf{x}_l, \mathbf{x}_m \in \mathbf{X}$
if the size of set $I_0 \leqslant 128$ then
let $I = I_0$
end
if the size of set $I_0 > 128$ then
    for $(\boldsymbol{x}_l, \boldsymbol{x}_m)_r \in I_0$ do
    initialize $\beta \leftarrow 1$;
initialize $\gamma \leftarrow 1$;
if $\boldsymbol{x}_l$ and $\boldsymbol{x}_m$ were taken by different instruments then
```

(Continued)

```
10         $\beta \leftarrow 0.3$;
11       end
12     if $p_{x_m} - p_{x_l} \leqslant 3/2$ then
13       $\gamma \leftarrow 0.2$;
14   end
15   assign selection probability $\phi_r = \beta * \gamma * w_{\rm spec}|_{\mathbf{x}_l} * w_{\rm spec}|_{\mathbf{x}_m}$;
16   end
17   randomly select $I \subseteq I_0$ of size 128 with selection weights $\phi_r$;
18 end
19 for $(\boldsymbol{x}_l, \boldsymbol{x}_m)_i \in I$ do
20     for $k \in [1, 2, ..., 24]$ do
21       feed $(\mathbf{x}_l, \mathbf{x}_m)_i$ to LSTM model to predict spectral time sequence $y_{i,k}(\lambda, p)$, where $p \in [-15, -14.875, ..., +29.875, +33]$
22   end
23   compute predictive mean: $\mu_i(\lambda, p) = \mathtt{MEAN}(y_{i,1}, y_{i,2}, ..., y_{i,24})$
24   compute predictive uncertainty: $\sigma_i(\lambda, p) = \mathtt{STD}(y_{i,1}, y_{i,2}, ..., y_{i,24})$;
25 end
26 compute the final spectral template: $T(\lambda, p) = \sum_i \sigma_i^{-2}(\lambda, p) * \mu_i(\lambda, p)$;
27 renormalize the final spectral template: $T(\lambda, p) \leftarrow T(\lambda, p) / \sum_\lambda T(\lambda, p)$
```

For SNe with large numbers of observed spectra such as SN 2011fe, which has a total of 65 training spectra, the total number of possible combinations (the size of I_0) is as large as 2145, which makes it very time-consuming to exhaustively calculate all possible input spectral combinations. To construct the input spectral pairs I, a subset of possible combinations is sufficient, as the information from such well-observed SNe is also likely redundant. A weighted selection scheme is adopted by assigning a selection probability for each possible spectral pair (see line 15 of Algorithm 1). A spectral pair with higher spectral weights $w_{\rm spec}$ has a higher probability of being selected in predicting the spectral sequence. We also introduced two penalty factors, β and γ, which preferably select spectral pairs taken by the same instrument and disfavor the spectral pairs observed on almost the same night, respectively.

Each spectral pair in this selection was fed to the neural networks for the spectral sequence prediction that covers spectral phases between −15 to +33 days (see line 21 of Algorithm 1). The activation of dropout can make the predictions probabilistic (see Section 5) and allows for estimates of the uncertainties of the predictions. For every input spectral pair, the prediction process is repeated 24 times with a new stochastic dropout mask applied each time. The final target spectrum prediction is given by the average of these predictions weighted by their predictive uncertainties (see line 26 of Algorithm 1). At last, each spectrum of the spectral template was renormalized by dividing its average flux (see line 27 of Algorithm 1).

As examples, Figure 7 shows the predicted spectral templates of two SNe with no more than five training spectra. Of the two SNe, SN 2004ey is a normal-velocity (NV; Zhao et al. 2015) and SN 1998ec is a high-velocity (HV) object (Blondin et al. 2012). Figure 7 offers a glimpse of SN Ia diversities in their spectral evolution. Around the *B*-band maximum, for instance, SN 1998ec has distinctly broader (relative to SN 2004ey) and stronger Si λ6355, Fe II λ4404, and Mg II λ4481 lines, in line with typical broad-line (BL) SNe Ia (Wang et al. 2013; Parrent et al. 2011) in Branch classes (Branch et al. 2006). As the objects evolve from −10 to +5 days, SN 1998ec shows a steeper velocity gradient across the Si II features than SN 2004ey. Object SN 1998ec was classified as a high-velocity gradient object (Parrent et al. 2011). The rapid line profile evolution is well captured by the neural network predictions. Moreover, the predicted spectral templates reconstruct the smooth spectral evolution without any unexpected discontinuities over the phase dimension. The LSTM neural networks also demonstrate excellent consistency at the boundaries of the blue and red sections at 5500 Å.

6.3. Test the Spectral Template Constructions

The test set was used to evaluate the performance of the neural networks. The observed spectra in the testing set were compared with the spectral template constructed from the neural network for the same SN at the same phase. To precisely match the phases of the testing spectra (hereafter testing phases) to those of the spectral template, the spectral template used for comparison was constructed at the phases that exactly match the phases of observation of an SN instead of the fixed time grid, as shown in line 21 of Algorithm 1. Examples of the spectral comparisons are shown in Figures 8 and 9 for normal SNe (NV and HV) and some peculiar SNe, respectively.

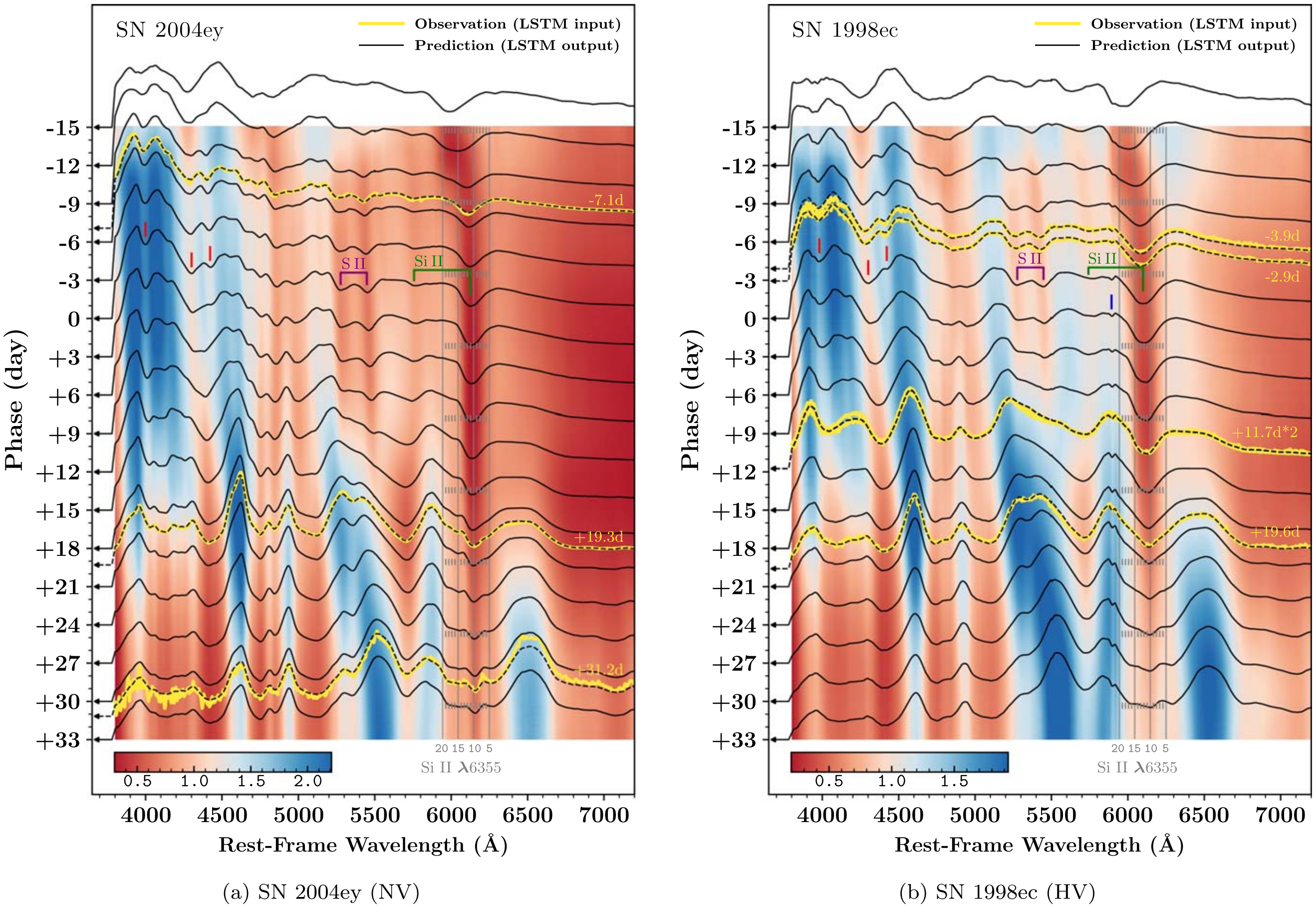
Figure 10 demonstrates the results of the spectral comparison by measuring the MAPE between the testing spectra and the corresponding neural network constructions. We found that 80% of the testing spectra show an MAPE smaller than 7.1%, and the overall median MAPE is 4.4%. The MAPE does not show any significant dependence on the phases of the SNe, which means the performance of the neural network is not statistically biased at any particular phases. We noticed that the MAPEs over the testing spectra without normalizing to the observed photometric colors are systematically larger than the MAPEs over those spectra with photometric color calibrations; 35 of 94 unnormalized testing spectra have an MAPE larger than 7.1%. By contrast, the ratio is 27 out of 215 for the color-calibrated testing spectra. We surmise that the MAPEs for testing spectra without color calibrations are more likely to be dominated by the inaccuracy of the observational colors of the spectra. Alternatively, we note that the median S/Ns over the test spectra without and with color calibrations are 31.9 and 46.7, respectively. The uncalibrated testing spectra generally have lower S/Ns, which could also contribute to their larger MAPEs.

The reliability of the prediction may also depend on the number of available spectra in the training set. Figure 11 shows the MAPEs between the observed spectra in the test set and the neural network–predicted spectra for SNe with different numbers of spectra in the training set. The figure confirms the general trend that the errors decrease as the number of spectra used in the training set increases.

6.4. Comparisons with Other Models

To showcase the fidelity improvement of our method, we compared the spectral templates constructed by LSTM with those generated by two empirical models of SNe Ia, i.e., the template of Hsiao et al. (2007) and the SALT3 model (Kenworthy et al. 2021).

The spectral template built by Hsiao et al. (2007; hereafter the Hsiao template) is a phase-dependent SED model based on a compilation of ∼100 SNe Ia with ∼600 spectra. This uniform template $H(p, \lambda)$, as a function of rest-frame wavelength λ and phase p, was obtained by averaging the observed spectra at different epochs after correcting their spectral colors to align



(a) SN 2004ey (NV)

(b) SN 1998ec (HV)

Figure 7. Predicted spectral templates of SN 2004ey (left) and SN 1998ec (right) using LSTM neural networks. The training spectra (yellow curves) of SN 2004ey (SN 1998ec) are fed into the LSTM model to construct a spectral template (the heat map). To improve the readability of the two-dimensional image, a selected spectral sequence (black curves) that constitutes certain discrete slices of the template spectral surface is overplotted to demonstrate the spectral features of certain phases as shown by the arrows on the left. Some typical absorption features of SNe Ia are marked by short vertical lines, including the Si II lines (at 6355 and 5972 Å; green lines), the W-shaped S II lines (at 5400 and 5600 Å; purple lines), Si II λ4130 (left red line), Fe II λ4404 and Mg II λ4481 (middle red line), and Si III λ4560 (right red line). Moreover, in the case of SN 1998ec, the persistent interstellar absorption Na I D line (blue line) is also reproduced by the spectral template. Note that there are two spectra of SN 1998ec observed at the same epoch (+11.7 days) by two different instruments. The vertical gray lines indicate the positions of the blueshifted Si II λ6355 at different velocities (the velocity unit is 1000 km s^{-1}).

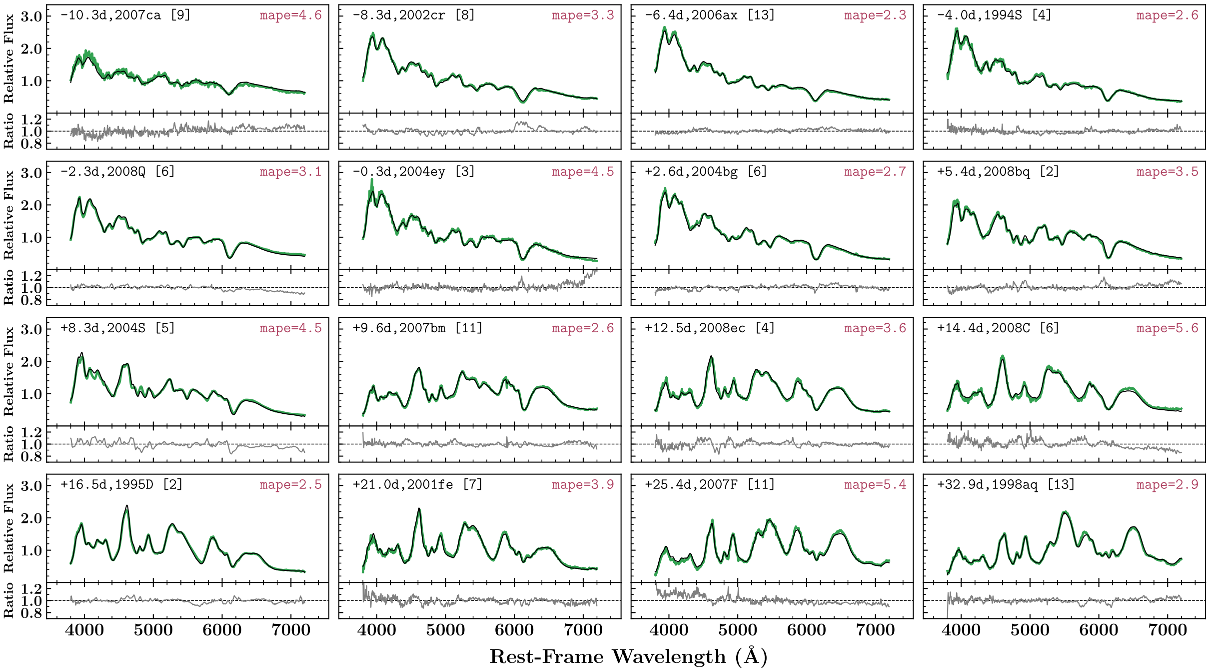
with a typical normal SN Ia with a "stretch" value of 1 (Hsiao et al. 2007; Knop et al. 2003). SALT2 (Guy et al. 2007) typifies the SN Ia spectral templates that are portrayed by a few free parameters. The spectral flux of the SALT2 model is given by

$$F(p, \lambda) = x_0[M_0(p, \lambda) + x_1 M_1(p, \lambda)] \cdot \exp[c \cdot CL(\lambda)], \tag{12}$$

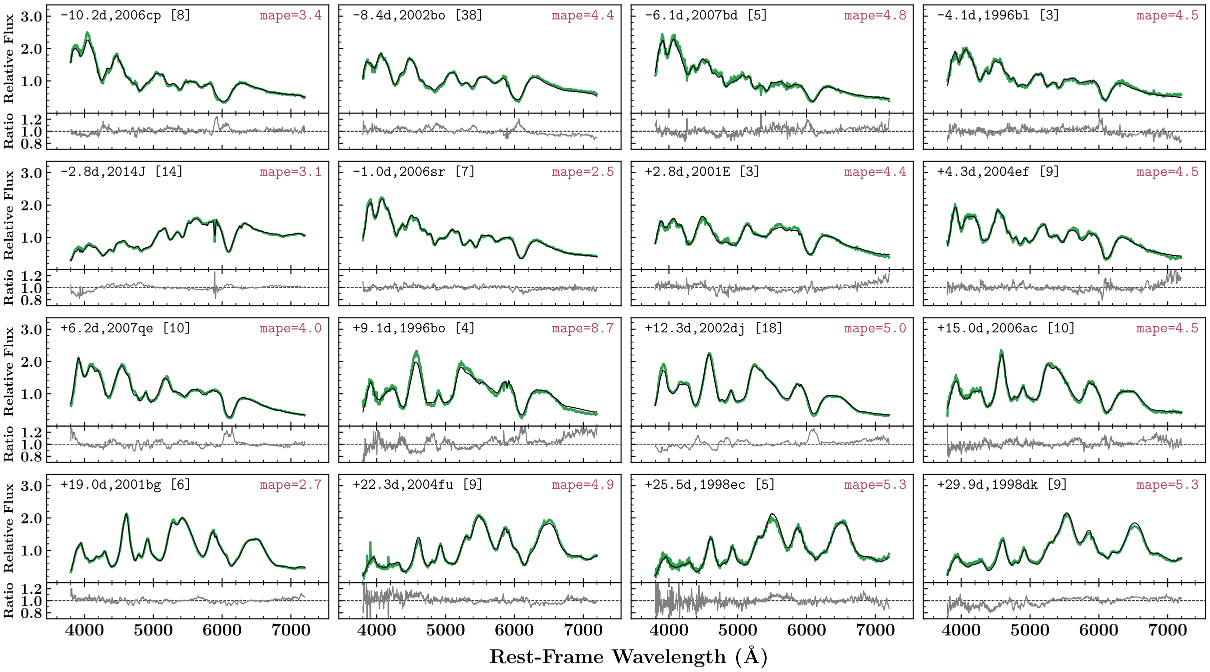
where M_0 and M_1 as principal components are flux surfaces derived from a training spectral sample, and $CL(\lambda)$ represents the average color correction law. A SALT2 template is determined by three parameters: x_0 (the overall flux normalization), x_1 (linked to the light-curve stretch), and c (a color-law coefficient). Beyond giving an average spectral evolution of SNe Ia, SALT2 also models the variations from SN stretch and includes a modulation term encoding the time-invariant color component. SALT3 (Kenworthy et al. 2021) is an improved version of SALT2 using the same framework but has better uncertainty estimation and better disentanglement of color and SN stretch. The SALT3 baseline model used in our work is the one presented in Kenworthy et al. (2021), which was trained on a data set of 1207 spectra from 1083 SNe. Throughout the paper, we used the `Python` implementation of the Hsiao template and SALT3 model in the `sncosmo` library (Barbary et al. 2016) for the baseline comparisons.

In this section, we applied the two Hsiao et al. (2007) and SALT3 models in the CSD. The reddenings by interstellar dust are needed for these models. For SALT3, photometric light curves are required. Only 118 SNe out of the 361 SNe in the CSD have photometric coverages and published host galaxy reddenings that are appropriate for such comparisons. The Milky Way (MW) foreground extinction of each SN is derived using the Schlafly & Finkbeiner (2011) reddening map. For 108 of the 118 SNe, we directly adopted the values of host extinction provided by the Kaepora database (Siebert et al. 2019), which lists A_V derived from MLCS2k2 (Jha et al. 2007) assuming $R_V = 2.5$. The host extinctions for the remaining 10 SNe are not available in the Kaepora database and are adopted from the following publications: SN 2011iv (Ashall et al. 2018), LSQ 12gdj (Scalzo et al. 2014), SN 2010ae (Stritzinger et al. 2014), SN 2014J (Ashall et al. 2014), SN 2012cu (Huang et al. 2017), SN 2012Z (Stritzinger et al. 2015), SN 2010jn (Hachinger et al. 2013), SN 2014ek (Li et al. 2018), ASASSN-14lp (Shappee et al. 2016), and SN 2011 hr (Zhang et al. 2016).

A model by the Hsiao template is uniquely determined by the assumed dust extinctions from the MW and the host galaxy.



(a) NV

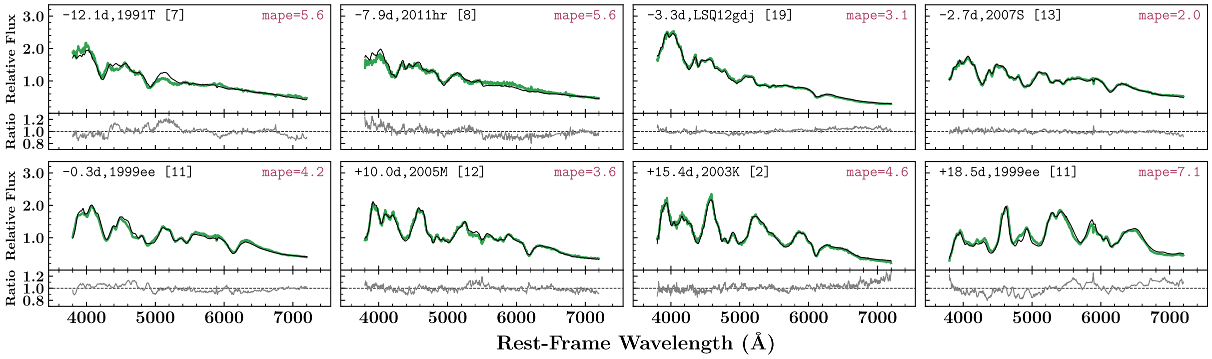


(b) HV

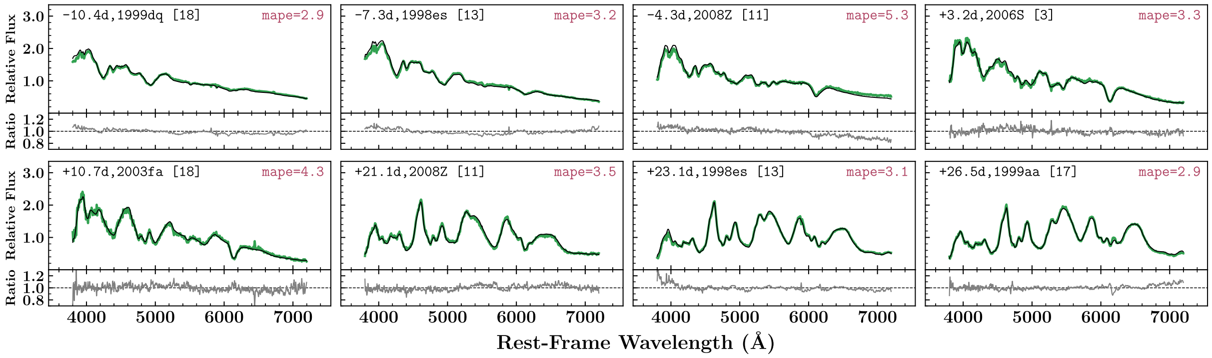
Figure 8. Comparisons of the LSTM-predicted template spectra (black curves) with the observed spectra in the test set (green curves) for 16 NV objects and 16 HV objects. The flux ratios of the predictions to the observations are shown in the lower attached frame for each panel. The number at the top left of each panel shows the phase from *B*-band maximum. The number in square brackets in the top left corner of each panel indicates the number of training spectra of the SN. The MAPEs between the predicted template spectra and the observed testing spectra are shown in the top right corner of each panel.

We generated the model Hsiao templates for the 118 SNe after corrections for both the MW and the host reddening, assuming the Cardelli, Clayton & Mathis (CCM) extinction law (Cardelli et al. 1989). Unlike the Hsiao template, the SALT3 model requires the observed spectra as the input data to fit the parameterized spectral time sequence. The SALT3 template of each SN is derived by fitting the LSTM training spectra corrected by their photometric $B-V$ color.

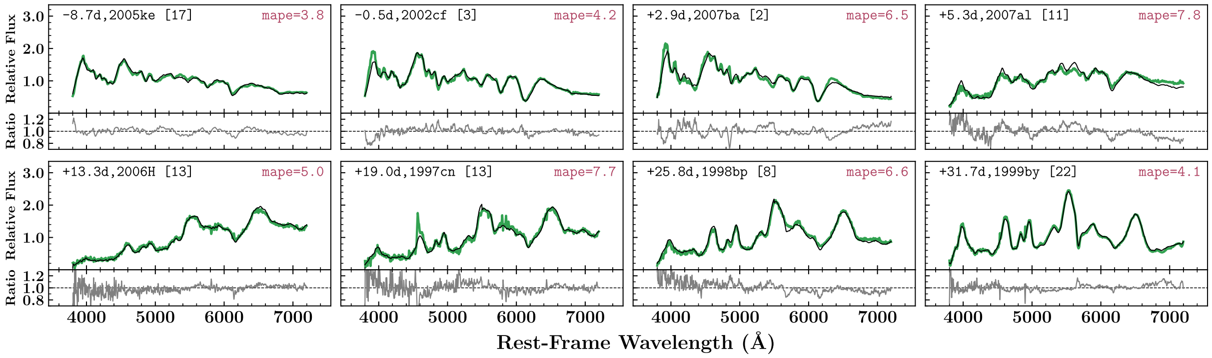
The spectra in the CSD are all flux normalized by their average flux (see Section 3.5), whereas the SALT3 model requires the spectra being fitted to preserve the time evolution of the flux. Therefore, the flux levels of the input spectra are



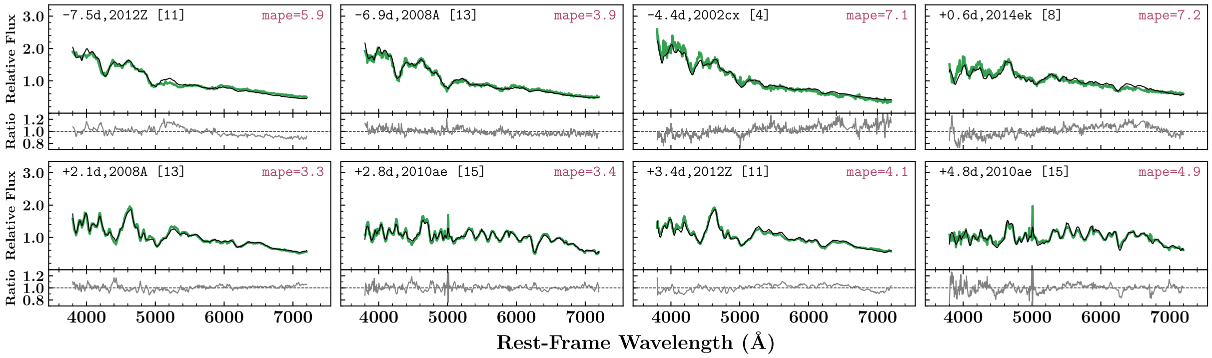
(a) Ia-91T



(b) Ia-99aa



(c) Ia-91bg



(d) Iax

Figure 9. Comparison of the predicted template spectra (black curves) with the observed testing spectra (green curves) for (a) Ia-91T, (b) Ia-99aa, (c) Ia-91bg, and (d) Iax. The panel format is the same as in Figure 8.

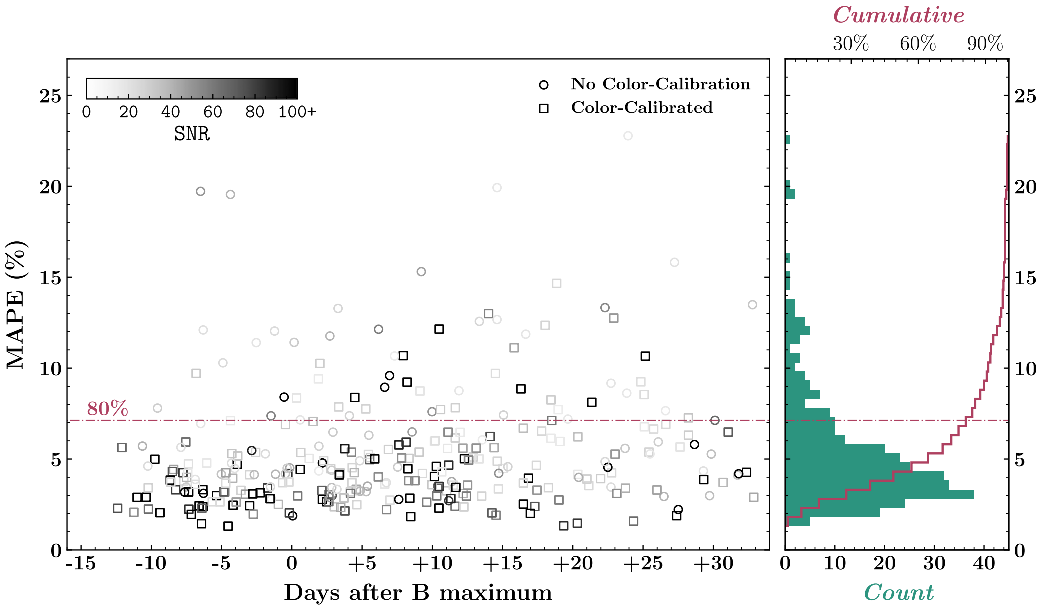


Figure 10. The MAPEs between the observed spectra in the test set and the predicted template spectra as a function of phase. We use squares (circles) to indicate the testing spectra with (without) available color calibration. The gray-scale colors of the data points represent the S/N of the testing spectra. A histogram of total MAPEs with a cumulative curve (red line) is attached to the right side. The horizontal dashed line shows the 80th percentile of the total MAPE distribution.

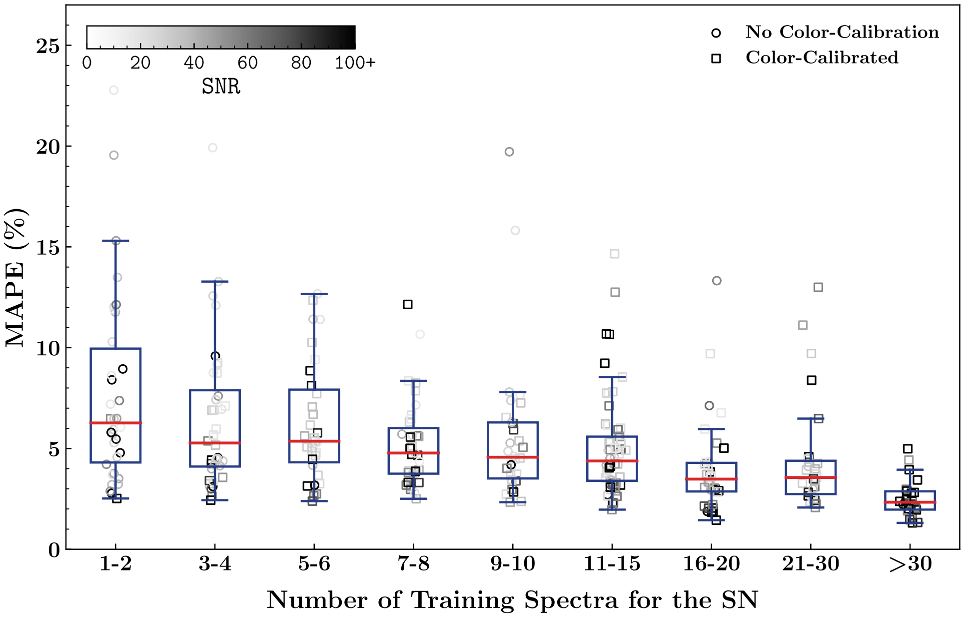


Figure 11. The MAPEs between the observed spectra in the test set and the predicted template spectra versus the number of training spectra for the SN. The symbols are the same as in Figure 10. The data points are arbitrarily shifted with small displacements in the horizontal direction for clarity of display. A box plot is overplotted onto each group of data points (see the definition of a box plot in Figure 3).

renormalized to their corresponding *V*-band magnitudes before SALT3 fitting. The SALT3 model allows for fitting spectra with given spectral uncertainties. Here the uncertainty spectra are deduced from the approximate error spectra described in Section 3.1 by taking into account the flux scaling factors.

Figure 12 shows examples of the template spectra at testing epochs generated by the LSTM, Hsiao et al. (2007), and SALT3 models. The SNe of distinct spectral properties at different phases are shown. In all cases, the LSTM method generally offers a considerable improvement over the other two models in reconstructing the spectral features. The Hsiao templates are unable to capture the diversity of the spectral profiles of most of the SNe. The SALT3 model only shows moderate improvement over the Hsiao template in most cases. The advantages of LSTM become even more obvious for the peculiar SNe, such as Ia-91T, Ia-91bg, and Iax. For these peculiar events, not only are the spectral features poorly fit, but the continuum levels are missed by the Hsiao and SALT3 templates. These models also have severe difficulties matching the spectral features of HV SNe.

Figure 13 shows the MAPEs between the testing spectra from the 118 SNe and their corresponding template spectra versus the template generation methods. The statistics also confirms that the LSTM templates with a median MAPE of

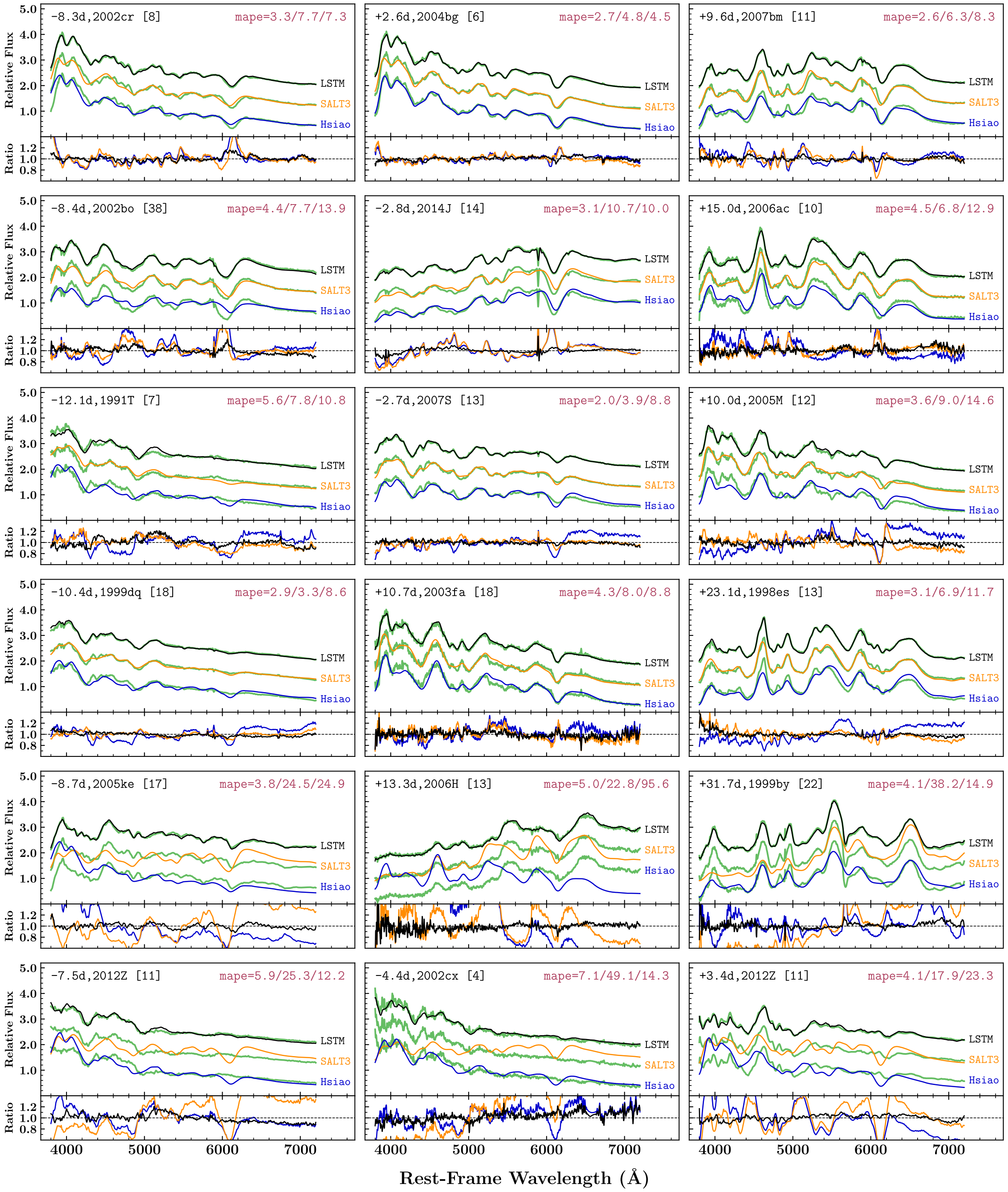


Figure 12. Comparisons of the template spectra generated by different methods (black curves for LSTM; yellow curves for SALT3; blue curves for the Hsiao template) with the observed spectra in the test set (green curves). Each row shows three cases at different phases for a specific subtype, from top to bottom, NV, HV, Ia-91T, Ia-99aa, Ia-91bg, and Iax. The panel format is the same as in Figure 8, but additional comparisons for the baseline models are also presented. In each panel, the template spectra have been arbitrarily shifted in the vertical direction for display clarity, and the same shifts are applied to the observed spectra (green curves). The MAPEs between the template spectra and the observed testing spectra are given in the top right corner of each panel; from left to right, the numbers are for LSTM, SALT3, and the Hsiao template.

4.0% outperform the two other models. The median MAPEs are 8.3% and 10.4% for the SALT3 and Hsiao templates, respectively.

7. LSTM Applied to the Analyses of SN Ia Spectra

Given the deficiency of spectroscopic resources, it is challenging for future transient surveys to acquire multiepoch spectroscopy. It is interesting to see how well a neural network–based algorithm can predict the spectral sequence of an SN based only on one or two spectra. Such a prediction is also a direct assessment of the critical information contained in any individual spectrum that can be employed to derive the intrinsic properties of an SN Ia.

A more detailed quantitative study of SN properties based on the neural networks constructed here will be presented in an upcoming paper. Here we only show examples of the neural network predictions to demonstrate their potential in SN Ia spectral studies.

Algorithm 2. Spectral Sequence Construction from One or Two Spectra

```
input: The trained LSTM model
input: A spectral pair ($\mathbf{x}_l$, $\mathbf{x}_m$) with $p_{\mathbf{x}_l} \leqslant p_{\mathbf{x}_m}$ ($\mathbf{x}_l = \mathbf{x}_m$ is allowed)
1 for $k \in [1, 2, ..., 64]$ do
2     feed ($\mathbf{x}_l$, $\mathbf{x}_m$) to LSTM model to predict spectral time sequence $y_k(\lambda, p)$,
      where $p \in [-15, -14.875, ..., +29.875, +33]$
3 end
4 compute predictive mean: $\mu(\lambda, p) = \mathtt{MEAN}(y_1, y_2, ..., y_{64})$
5 compute predictive uncertainty: $\sigma(\lambda, p) = \mathtt{STD}(y_1, y_2, ..., y_{64})$
```

7.1. The SNe with NV and HV

Two representative examples of NV and HV SNe Ia are SN 2011fe (Zhao et al. 2015) and SN 2002bo (Blondin et al. 2012). Both SNe are well observed, with extensive multiphase spectroscopic coverage. In this section, we introduce a fictitious scenario: an SN similar to SN 2011fe or SN 2002bo is newly discovered with only one available spectrum around its maximum light. We are interested in inferring the entire spectral sequence of them using the neural network trained on the data of all of the SNe Ia, excluding these well-observed SNe.

Two separate LSTM models were constructed specifically for these two SNe. The training sets of these two different models are different in that for each SN, the training set contains all of the spectra in the CSD except the SN being modeled. The training samples for both cases are also generated by exhausting all possible permutations as in Section 6.1.

The spectral sequence is constructed with the procedures described in Section 6.2 but following Algorithm 2. This process is a simplified version of Algorithm 1. In Algorithm 1, there is more than one element in set I (see line 19 in Algorithm 1). However, here we only have one or two spectra; thus, the input spectral pair is uniquely determined, and the combination process (see lines 26 and 27 in Algorithm 1) is no longer necessary. Unlike Section 6.2, the process bears a much lower computational cost, so we adjust the repeating times from 24 to 64 (see line 20 in Algorithm 1 and line 1 in Algorithm 2).

Figure 14 shows the resulting mean spectral sequences of SN 2011fe and SN 2002bo derived from the neural networks using a single spectrum at maximum light following Algorithm 2. The Si II λ6355 at the maximum of SN 2002bo (HV object) is obviously broader and stronger than that of SN 2011fe (NV object). Figure 15 presents the comparisons of the predictive mean sequences shown in Figure 14 with their corresponding uncertainties. The uncertainties seem to have footprints broadly consistent with the mean templates, making them reminiscent of the observational noise dominated by photon shot noise. The uncertainties at the earliest phases are relatively large due to the scarcity of spectroscopic data of young SNe Ia in the training data set. Overall, the uncertainties are slightly lower for SN 2011fe than for SN 2002bo.

Figure 16 shows the spectral comparisons for SN 2011fe and SN 2002bo. The predictions of NV object SN 2011fe at phases >-8 days are in excellent agreement with observations. For earlier phases, an obvious discrepancy across the Si II λ6355 feature emerges as the model shows a broader absorption. Meanwhile, the model performs poorly at the blue end with $\lambda < 5200$Å, where the spectra are dominated by absorption features of Si, Fe, and Mg. Nevertheless, the prominent S II lines at around 4800 Å seem to be well predicted. For HV object SN 2002bo, the model can properly predict the broad and strong Si II λ6355 line and the prominent S II lines. However, the performance is less satisfactory between 4500 and 5000 Å.

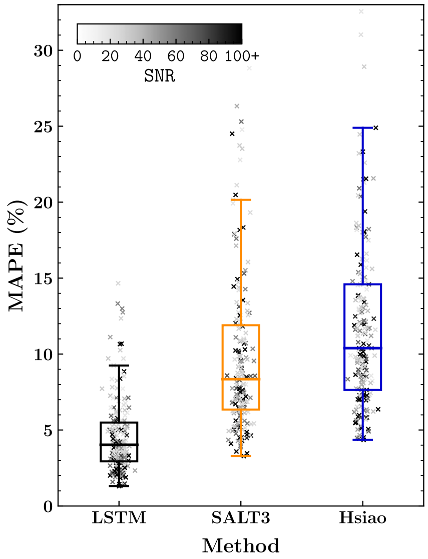
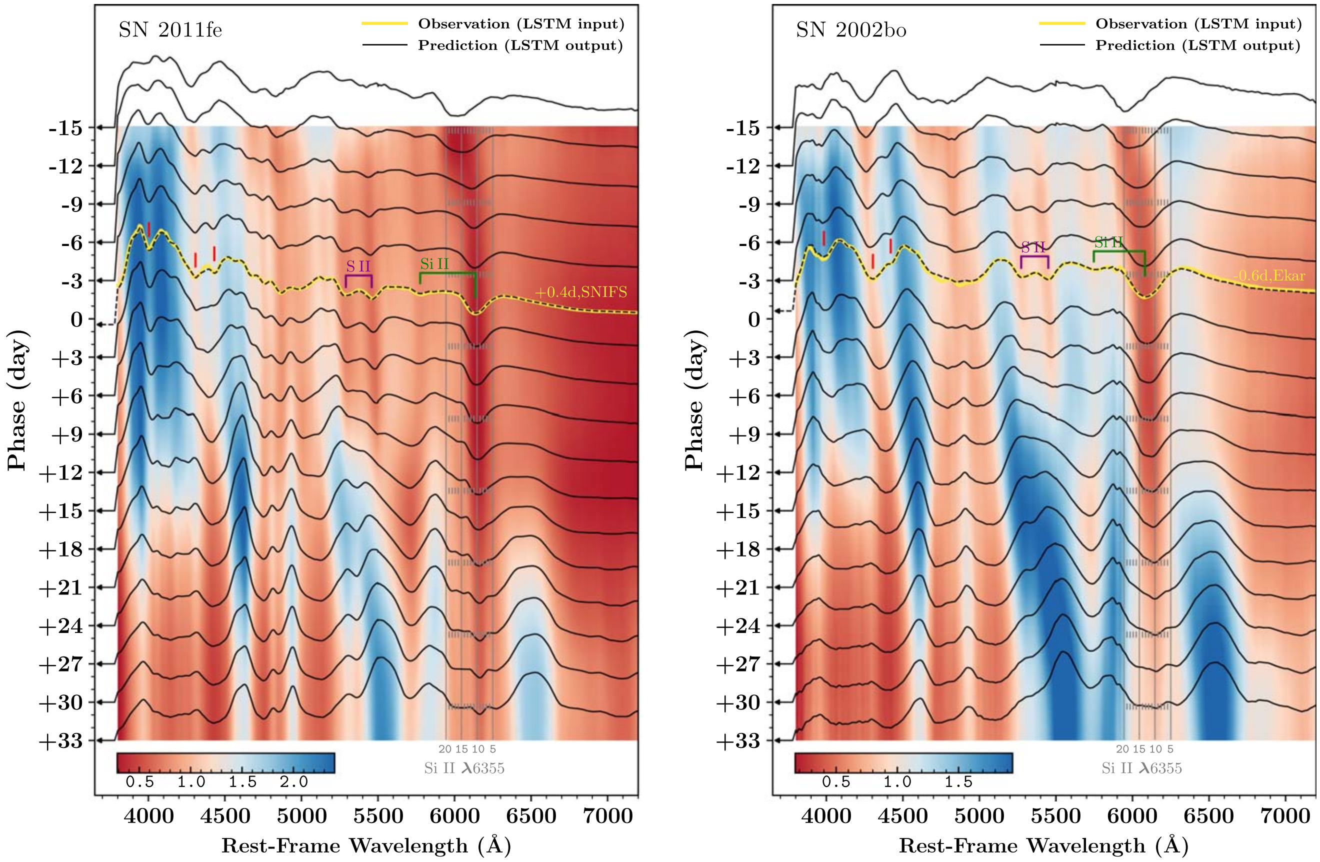


Figure 13. The MAPEs between the testing spectra from the 118 SNe and the template spectra versus the template generation methods. The data points are arbitrarily shifted with small displacements in the horizontal direction for clarity of display. A box plot is overplotted onto each group of data points (see the definition of a box plot in Figure 3).

We also generated the spectral models for SN 2011fe and SN 2002bo using the Hsiao et al. (2007) and SALT3 models for comparison. The Hsiao templates of both SNe are generated as in Section 6.4. For the SALT3 model, we followed the same procedures as described in Section 6.4 to make the spectral predictions, except that the spectral data are fitted only for the spectrum at the maximum light. These models yield less accurate predictions for both SNe, as shown in Figure 17. Although the same prior knowledge (the single spectrum at maximum light) is used in the fits for the LSTM neural networks and the SALT3 model, the LSTM neural networks show significantly better performance.

The diversity among normal SNe Ia can be examined by their photospheric velocities. The expansion of the photosphere may evolve quite differently for different SNe. It is interesting to investigate how the neural networks can learn and capture the spectral evolution of different spectral types of SNe Ia. Figure 18 shows the velocity of the absorption dip of Si II λ6355 Å for SN 2011fe and SN 2002bo. The errors of the neural network predictions are generally less than 200 km s^{-1} from day -10 to 30 for SN 2011fe. The errors for SN 2002bo are slightly larger but typically less than 500 km s^{-1} from day -10 to about day 30. For both SNe, the overall trend of line velocity evolution from about a week before maximum to 4 weeks past maximum is well captured by the neural networks using only a single spectrum around optical maximum as input. In contrast, the line velocity evolutions measured on the predicted spectra from the other two models have much



(a) SN 2011fe (NV)

(b) SN 2002bo (HV)

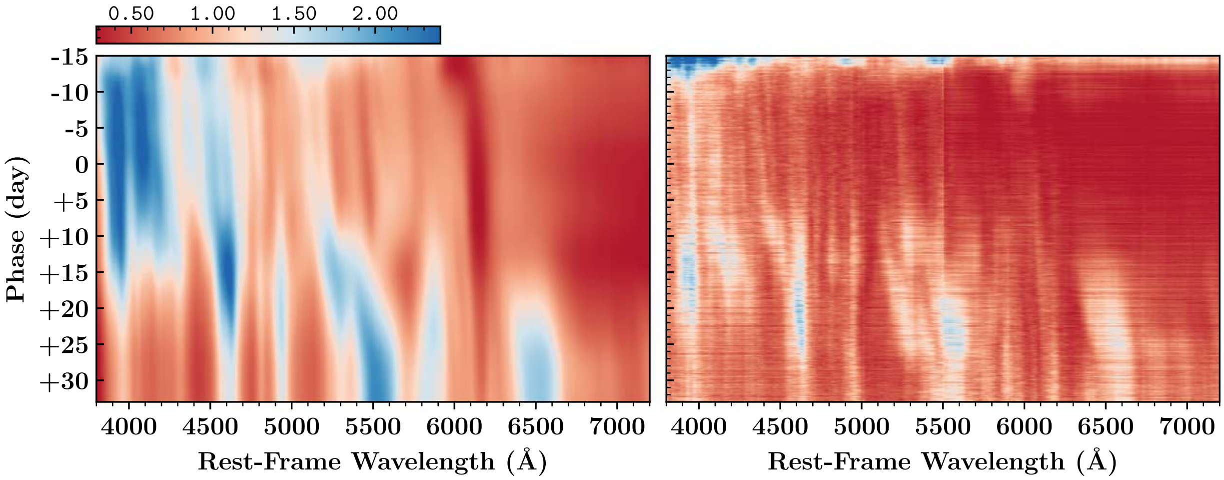
Figure 14. Predictive mean sequences of SN 2011fe (left) and SN 2002bo (right) derived from a single observed spectrum at maximum light using LSTM neural networks. For each SN, the spectrum (yellow curve) observed at maximum light is fed into the LSTM model (trained on the data excluding this SN), then the predictive mean sequence (heat map) is obtained by averaging all results from multiple forward passes. The overplotted discrete spectra (black curves) are drawn from the full sequence. The panel format is the same as in Figure 7.

stronger bias, as shown in Figure 18. The Hsiao template is insensitive to the spectral diversity of SNe Ia. Though SALT3 allows for higher flexibility to model spectral features than the Hsiao template, it still fails to capture the evolution of spectral velocities of any individual SN Ia.

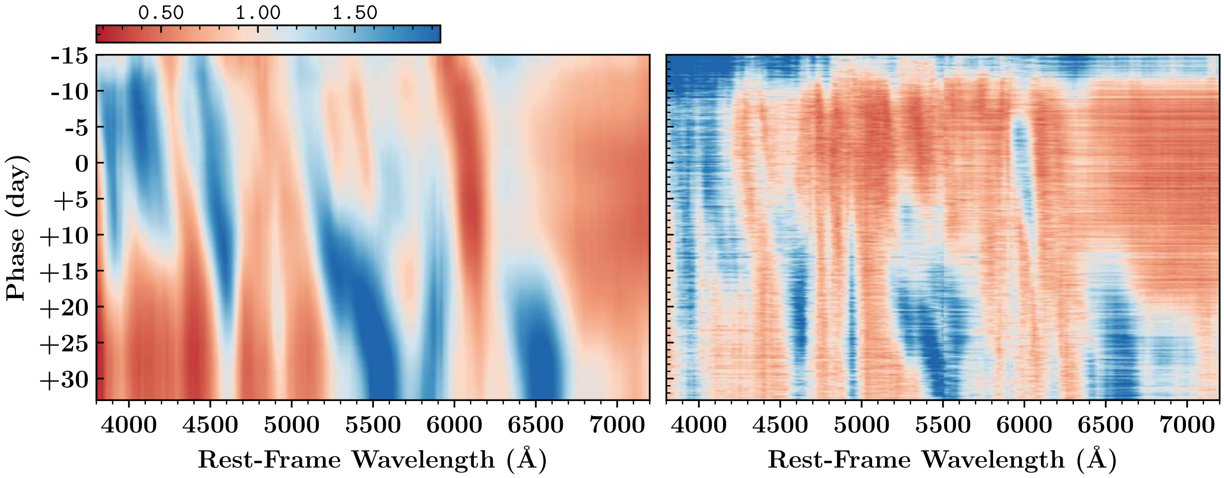
Accurately predicting the underlying continuum component of the spectra is also a goal of the neural networks we have constructed. Figure 19 shows the comparisons of the integrated spectral $B-V$ colors measured on predictions and observations for SN 2011fe and SN 2002bo. The B and V magnitudes were directly integrated over the spectra with standard (Bessell) B and V transmission curves in the rest frame. The missing data are padded by zero before calculating the B-band magnitudes. Recall that no extinction corrections were applied on the spectra of the training set. The $B-V$ colors are not the intrinsic color but a metric to quantify the colors of the predicted spectral sequence. The effect of photometric color calibration during preprocessing (see Section 3.5) is also assessed. The spectral $B-V$ colors of the homogenized spectra are also shown in Figure 19.

In general, we find that the $B-V$ color residuals for SN 2011fe and SN 2002bo are typically less than 0.03 mag in both cases, and the overall trend of the color evolution is very well reproduced by LSTM neural networks. Yet both the Hsiao et al. (2007) and SALT3 models show less satisfactory performance in predicting the color evolution, especially for the HV object SN 2002bo. The color evolution of the Hsiao template is determined by the uniform average trend of its own training set. It does not accommodate any intrinsic color diversity. Its performance for modeling the color evolution appears to be better than that of SALT3. This indicates that using a single spectrum around maximum, the SALT3 model, although it offers more flexibility, is not well constrained to derive a full spectral sequence of SNe Ia, especially the HV SNe Ia.

One may notice in Figure 19 that the predicted $B-V$ colors from LSTM neural networks are slightly more consistent with those of the homogenized spectra, which are not calibrated to the observed photometric colors, than with the corrected spectra. This could be an indication that a small portion of original spectral data, e.g., those from SNIFS, may have already achieved excellent flux calibration, and the color calibration we adopted is based on photometric observations from heterogeneous sources, which may, in fact, be less accurate. The oversimplified color calibration process is essential in constructing a uniform data set but may also inherit the errors of the input photometric data. The SN 2002bo data shown in Figure 19 illustrate a more common situation: original spectra can exhibit wrong colors and appear as outliers on the $B-V$ evolution curve. Overall, Figure 19 demonstrates that the color evolution of an SN Ia can be reliably predicted using only one spectrum taken around optical maximum using LSTM neural networks.



(a) SN 2011fe (NV)



(b) SN 2002bo (HV)

Figure 15. Comparison of the predictive mean sequences with their corresponding uncertainty sequences of SN 2011fe (top) and SN 2002bo (bottom) derived from a single observed spectrum at maximum light. The left heat maps are identical to those shown in Figure 14 with adjusted sizes. The right panels show the 1σ uncertainties amplified by a factor of 50.

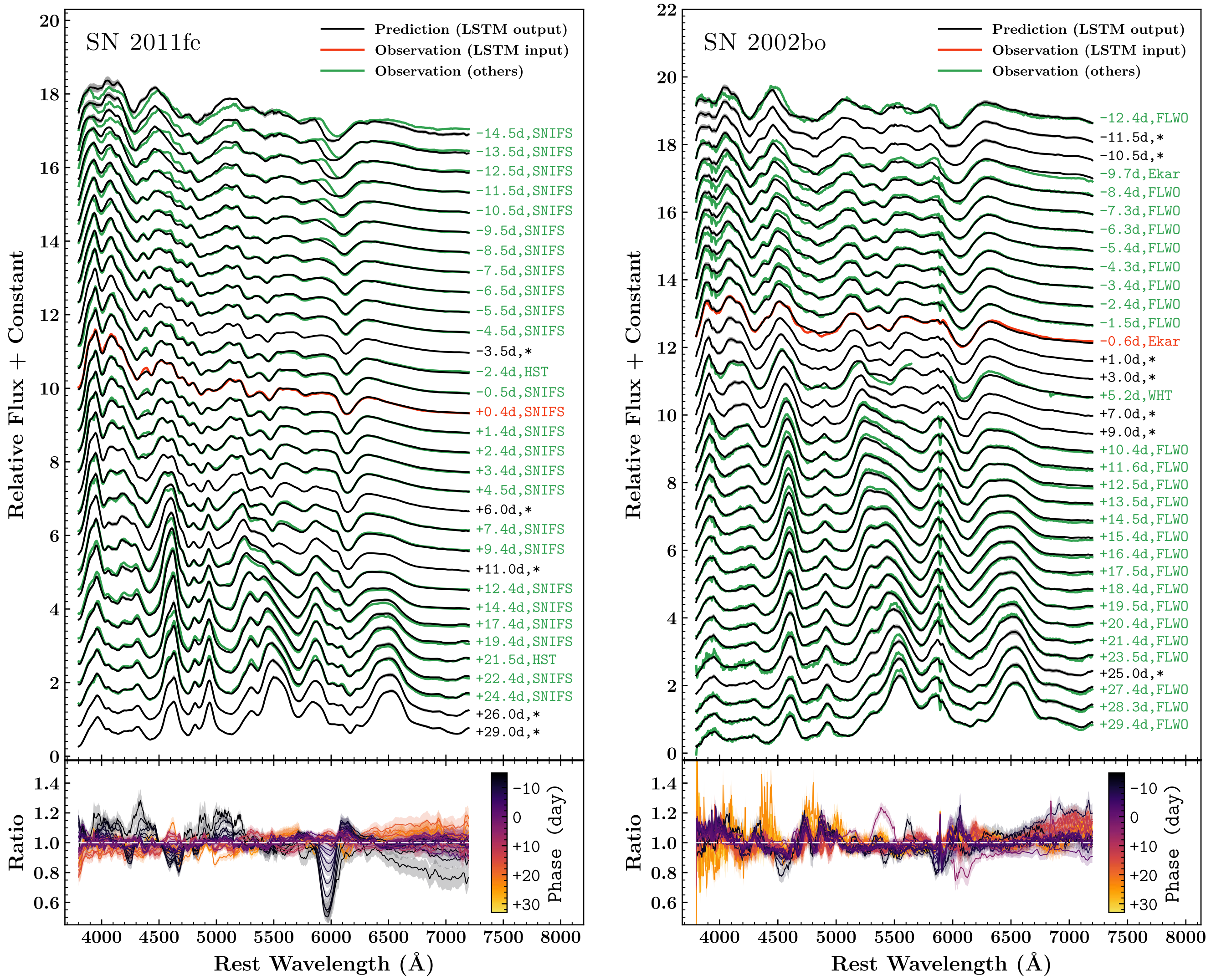
The success in predicting the color evolution of an SN Ia based on spectral data of one or two epochs is a remarkable achievement that may lead to new observational strategies for future SN cosmology.

7.2. The Diverse SN Ia Spectral Family

A wide range of spectral diversities are demonstrated by SNe Ia. A critical question is whether the neural networks can capture such diversities. In this section, we use four representative SNe (LSQ 12gdj, SN 2008Z, SN 2005ke, and SN 2012Z) to examine the performance of the neural networks on the diverse subtypes of SNe Ia. For each representative SN, a separate LSTM model is trained over the samples generated from all of the SNe Ia excluding the representative SN under study, as in Section 7.1. The spectral sequence is constructed similarly following Algorithm 2.

Figure 20(a) shows the spectral sequence predicted from one spectrum at maximum light for LSQ 12gdj (González-Gaitán et al. 2014), which is a Ia-91T SN (Filippenko et al. 1992; Phillips et al. 1992) with a very shallow Si II 6355 Å line before optical maximum. The spectral sequences constructed by using a spectrum taken at +0.7 days after *B* maximum are shown together with the observed spectra. At the time of optical maximum, the Si II 6355 Å line is well developed. The spectral features match well throughout the period covered by the observations. In particular, the neural network is able to reproduce the extremely shallow Si 6355 Å feature at epochs around 1 week before maximum. This suggests that the spectra around the optical maximum carry enough information to define a Ia-91T event. Identifying Ia-91T events from normal SNe Ia is important for SN cosmology, as it is shown by recent studies that Ia-91T SNe are potential sources of systematic errors of SN cosmology (Jiawen Yang et al., in preparation).

Object SN 1999aa represents another subtype of peculiar SN Ia, which is similar to Ia-91T SNe but with the subtle differences of having weak signatures of Ca II H and K and Si II absorption prior to maximum light (Garavini et al. 2004). Figure 20(b) shows the model sequence calculated with a single spectrum +0.3 days from maximum together with the observations for SN 2008Z (Silverman et al. 2012), which is a



(a) SN 2011fe (NV)

(b) SN 2002bo (HV)

Figure 16. Spectral sequence predicted from one spectrum at maximum light for SN 2011fe (left) and SN 2002bo (right) using LSTM neural networks. The solid black lines show the predictive mean spectra, and the gray shaded areas indicate 2σ standard deviations, representing 95% confidence. The corresponding observations (photometric color-corrected) are plotted as green curves, except for the input spectrum around the peak highlighted in red. All of these spectra have been arbitrarily shifted in the vertical direction for clarity of display. They are labeled with phase and instrumental source, where the asterisk is a placeholder for the cases without corresponding spectroscopic observations. The flux ratios of predictions to observations for the spectral sequence are shown in the lower panel, where the lighter colors represent 95% confidence caused by the predictive uncertainty. In the case of SN 2002bo, the spectrum contributed by WHT at +5.2 days is a homogenized spectrum from the extended data set, and notice that the segment from 5700 to 6000 Å of this WHT spectrum is not available.

Ia-99aa SN. In this case, the Si II feature is stronger a week before maximum than Ia-91T and again well reproduced by the neural network. This demonstrates that the neural network can effectively distinguish Ia-91T and Ia-99aa SNe based on spectroscopy around optical maximum.

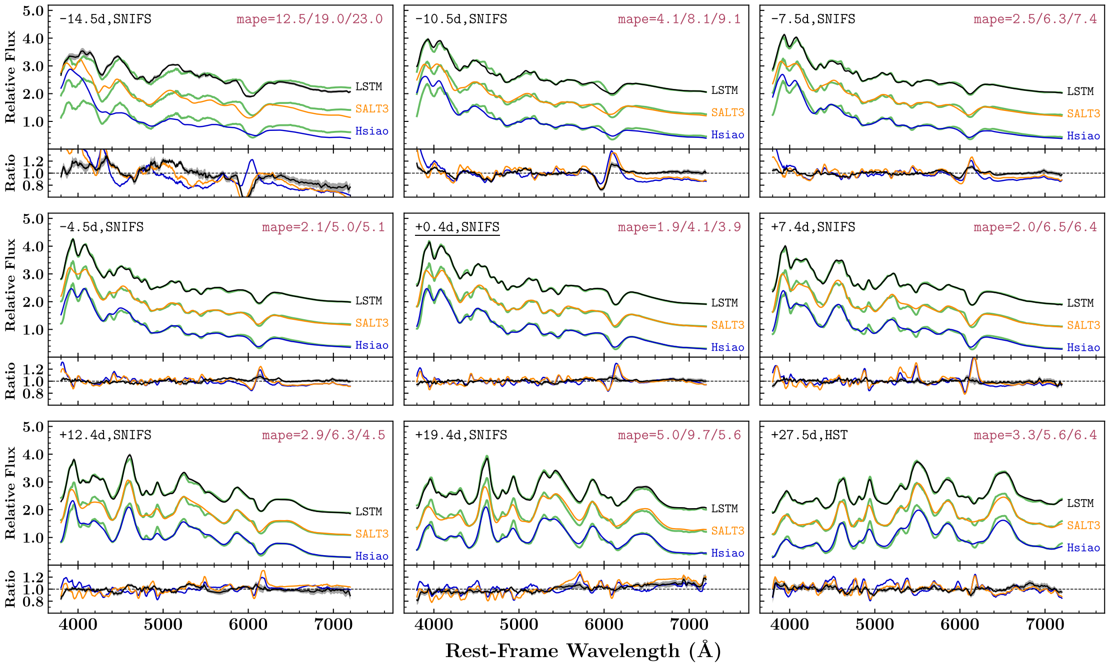
Object SN 2005ke (Krisciunas et al. 2017b) is a Ia-91bg (Filippenko 1997) subluminous SN. The SN shows rapid spectral evolution and much redder overall spectral color. Figure 20(c) shows the spectral sequence computed from a spectrum of SN 2005ke at +0.3 days from optical maximum and the observed spectra. The rapid evolution of the spectral features is well produced from more than a week before maximum to about a month after maximum.

Object SN 2012Z (Stahl et al. 2019) belongs to the peculiar type Iax (Foley et al. 2013) SNe. The spectral data of this subtype are sparse, but nonetheless, the neural network captures all of the major spectral features of the observations, as shown in Figure 20(d).

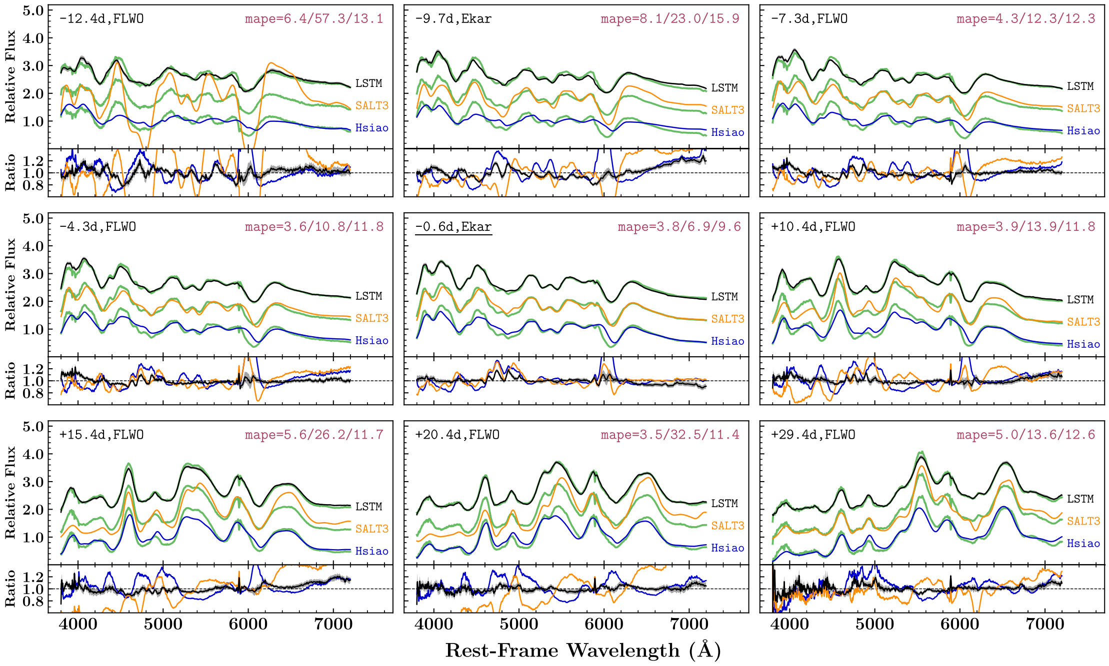
Like in Section 6.4, we compared the results of the LSTM neural networks with the other two template models, as shown in Figures 21 and 22. We found again that the Hsiao et al. (2007) and SALT3 models using only one spectrum around maximum cannot accurately reconstruct the spectral evolution of any of these objects. The LSTM models can robustly reproduce the spectral sequences of these SNe from the date of explosion to about a month past optical maximum, except for the Iax SN 2012Z, which shows increasing deviations at around 2 weeks past maximum.

7.3. Spectral Sequences from Two Spectra

One important question in scheduling spectroscopic observations of SNe Ia is what the optimal time gaps between



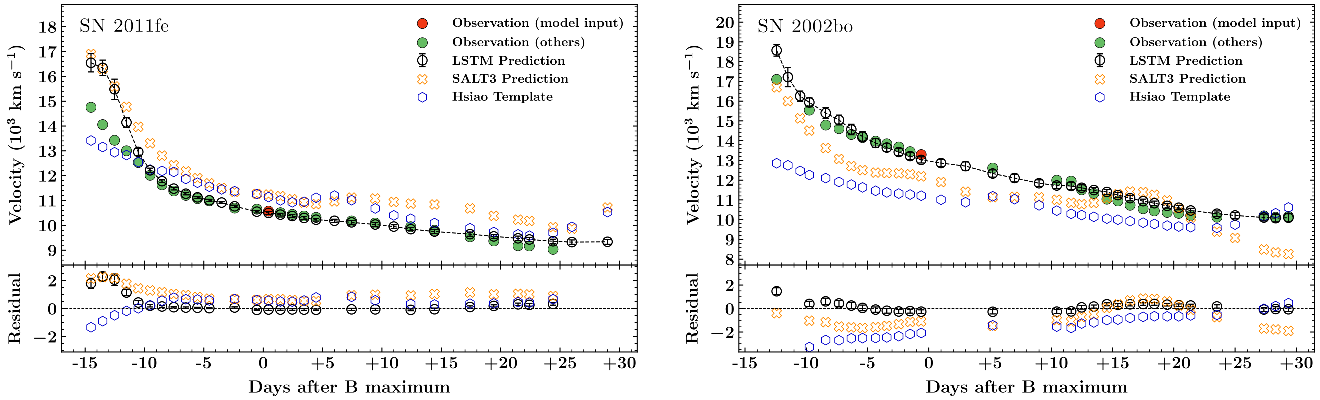
(a) SN 2011fe (NV)



(b) SN 2002bo (HV)

Figure 17. Comparison of the predicted spectra generated by different methods (black curves for LSTM; yellow curves for SALT3; blue curves for the Hsiao template) with the observed spectra (green curves) for SN 2011fe (top) and SN 2002bo (bottom). For each SN, the predicted spectra from LSTM and SALT3 are obtained by fitting on the single spectrum at maximum light, while the Hsiao template is the one already constructed in Section 6.4. Each panel shows a comparison at a different epoch, and the panel format is the same as in Figure 12, but the SN name is replaced by the instrument source in the top left corner of each panel. The underlined text in the central panel of each SN highlights the epoch of the spectra employed in the fits.

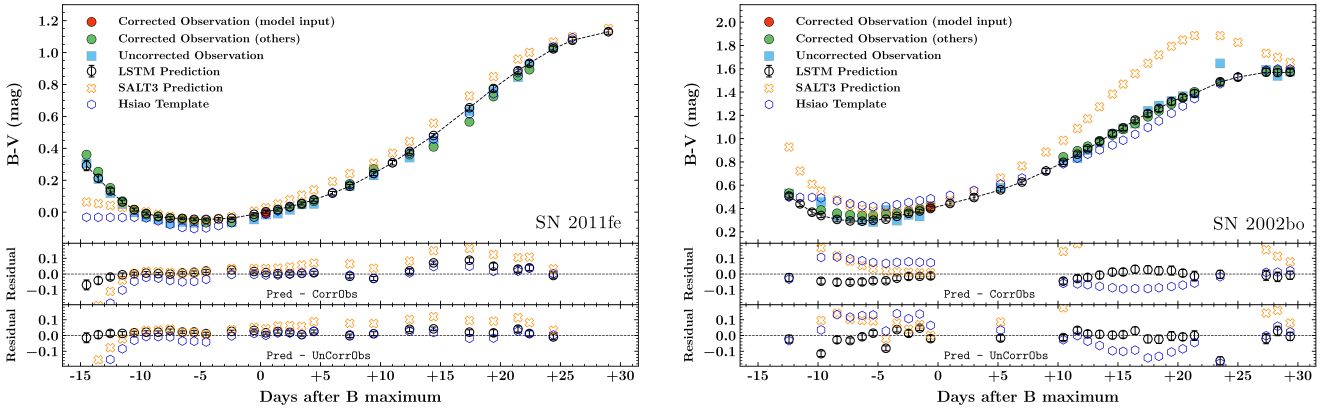
observations would be if multiple observations could be acquired. With the neural networks, such a question is related to the spectral sequence prediction using multiple spectra as input data. Figures 23(a)–(d) show the predictions with two spectra of SN 2011fe separated by ∼2, 4, 8, and 16 days in spectral phase, each at [−0.5, +0.4], [−2.4, 1.4], [−4.5, +3.4], and [−8.5, +7.4] days, respectively. The corresponding measurements of Si II λ6355 velocity and spectral $B-V$ colors are given by Figures 24(a)–(d) and 25(a)–(d). The overall spectral fits at early phases improve significantly, as a spectrum at phases well before maximum is used as the input. The fits to the data around optical maximum do not show any



(a) SN 2011fe (NV)

(b) SN 2002bo (HV)

Figure 18. The Si ii λ6355 velocity measured on the spectral sequence predicted from a single spectrum at maximum light using LSTM neural networks (black) and the corresponding observed spectra (green) for SN 2011fe (left) and SN 2002bo (right). The LSTM-predicted spectra and the observed spectra are the same as those shown in Figure 16, and the measurements from the input observed spectra at maximum light have been highlighted in red. Here the error bars are 1σ uncertainties of the measurements over the LSTM-predicted spectra from different forward passes. For comparison, the velocities measured on the SALT3-predicted spectra (yellow) and Hsiao template spectra (blue) at the epochs of LSTM predictions are overplotted. As the broken WHT spectrum of SN 2002bo has well covered the silicon feature, we also present its Si λ6355 velocity measurement in the figure.



(a) SN 2011fe (NV)

(b) SN 2002bo (HV)

Figure 19. Synthetic $B-V$ color measured from the spectral sequence predicted from a single spectrum at maximum light using LSTM neural networks (black circles) and the corresponding observational data (green circles and blue squares) for (a) SN 2011fe and (b) SN 2002bo. The LSTM-predicted spectra and the observed spectra are those already shown in Figure 16, and the measurements from the input observed spectra at maximum light have been highlighted in red. The error bars are 1σ uncertainties of the measurements over the LSTM-predicted spectra from different forward passes. To analyze the effect of photometric color calibration of the spectra, the measurements on both the homogenized (blue squares, labeled as uncorrected) and corrected (green circles) spectra are shown in the figure. For comparison, the synthetic $B-V$ colors measured on the SALT3-predicted spectra (yellow crosses) and Hsiao template spectra (blue hexagons) at the epochs of LSTM predictions are also plotted. Note that the spectral $B-V$ color is not measured on the broken WHT spectrum of SN 2002bo. Moreover, the last two corrected spectra of SN 2011fe and the first corrected spectrum of SN 2002bo did not go through any color calibration due to the deficiency of photometric coverage. Hence, the measurements from the homogenized and corrected spectra are consistent for them.

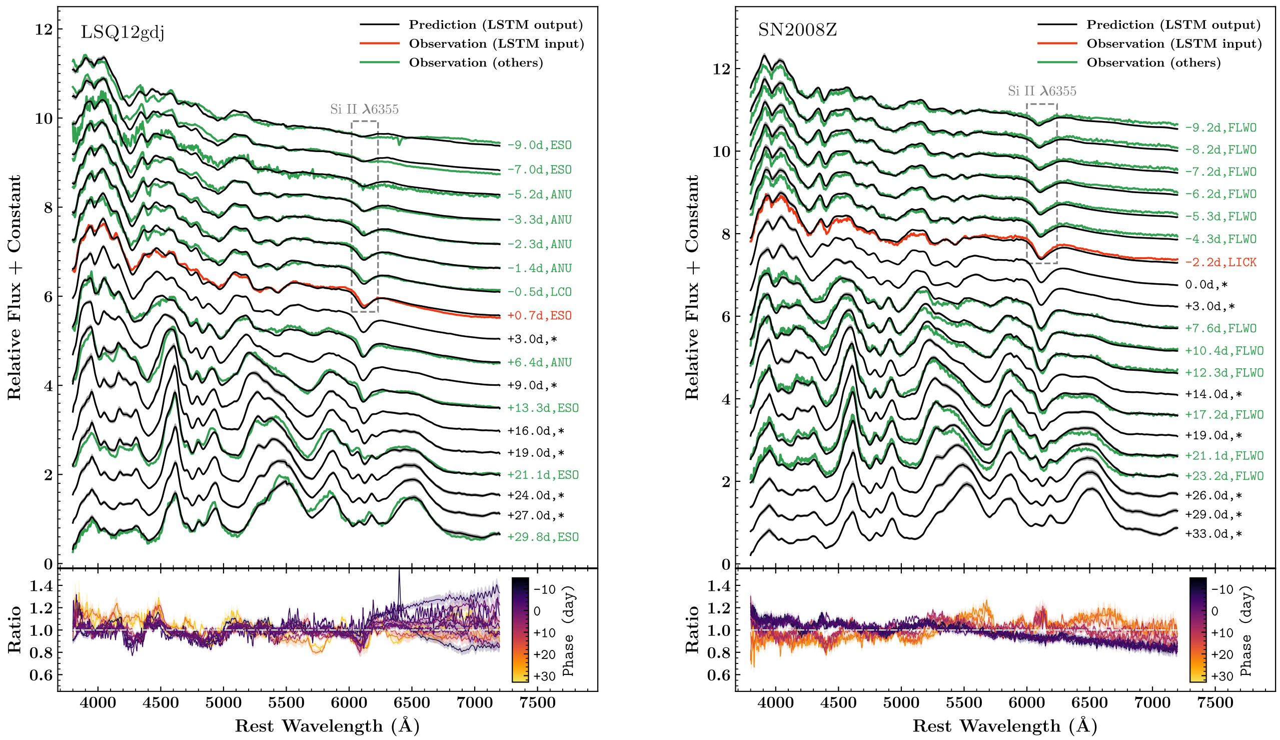
obvious deterioration, even when both input spectra are more than 1 week from optical maximum. The same is true for HV SN 2002bo, as shown in Figures 24(e)–(h) and 25(e)–(h).

Based on these neural network predictions, we may conclude that spectral data separated by more than 8–16 days around optical maximum can provide the maximum constraining power on the intrinsic properties of an SN Ia.

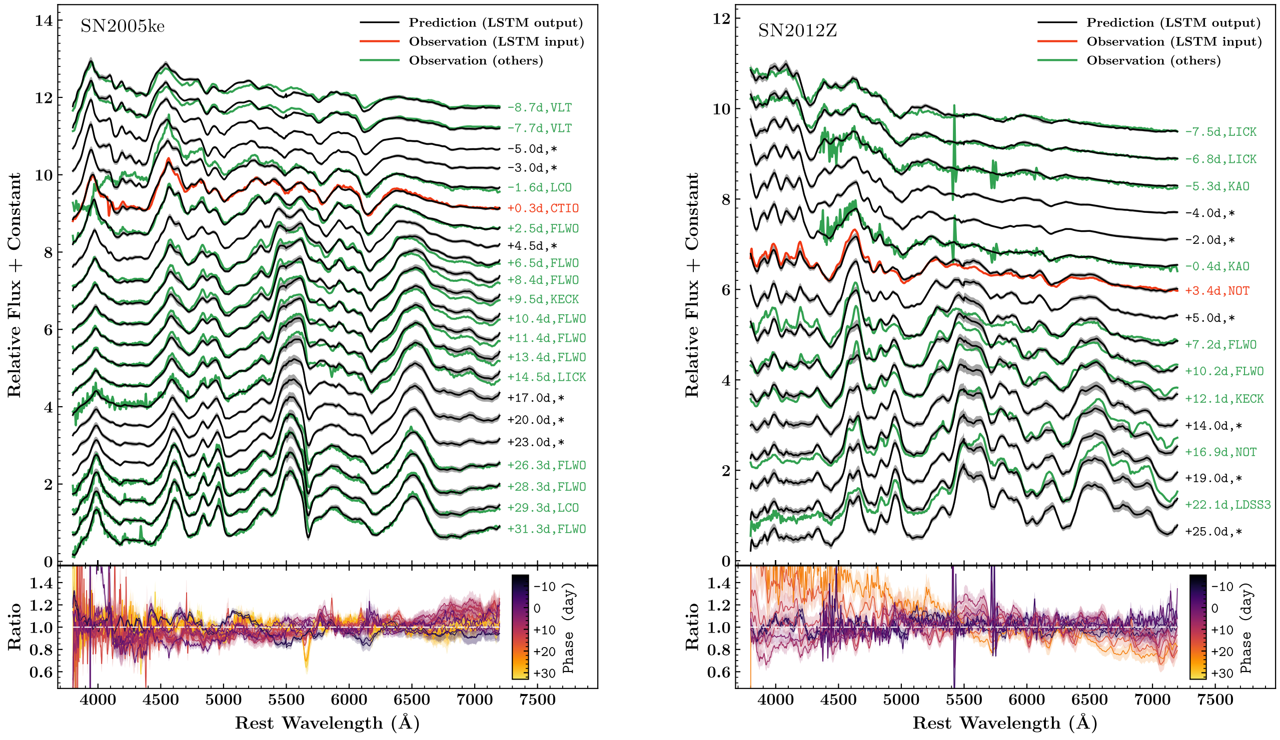
7.4. When the Spectral Phase Is Unknown

In our framework, the spectral phase is set as prior knowledge, whereas the time of maximum light may not always be available, especially when a newly discovered SN is still being actively monitored. A CNN can be used to provide a phase estimate based on an SN spectrum `deepSIP` (Stahl et al. 2020)). The success of the LSTM neural networks allows for an alternative approach to derive the phase of an SN spectrum without light-curve data. The basic concept is that a wrong spectral phase fed into the LSTM neural networks is likely to degrade the resulting predictions; that is to say, the correct spectral phase stands the highest chance of maximizing the predictive performance.

We tested this concept with six representative SNe Ia (SN 2011fe, SN 2002bo, LSQ 12gdj, SN 2008Z, SN 2005ke, and SN 2012Z). Their spectra are shown in Figures 16 and 20 (except the WHT spectrum of SN 2002bo with data missing). Each spectrum is assigned a sequence of phases from −15 to +33 days and fed into the LSTM model (trained without the same SN). The predictive MAPEs at the time of the input



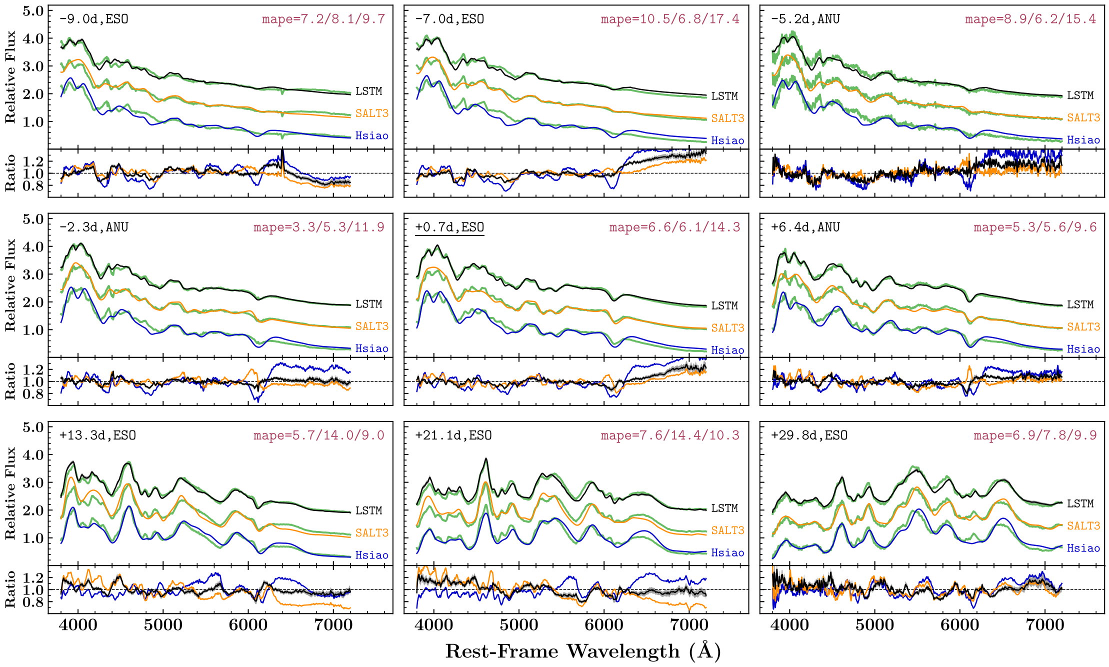
(a) LSQ12gdj (Ia-91T)   (b) SN 2008Z (Ia-99aa)



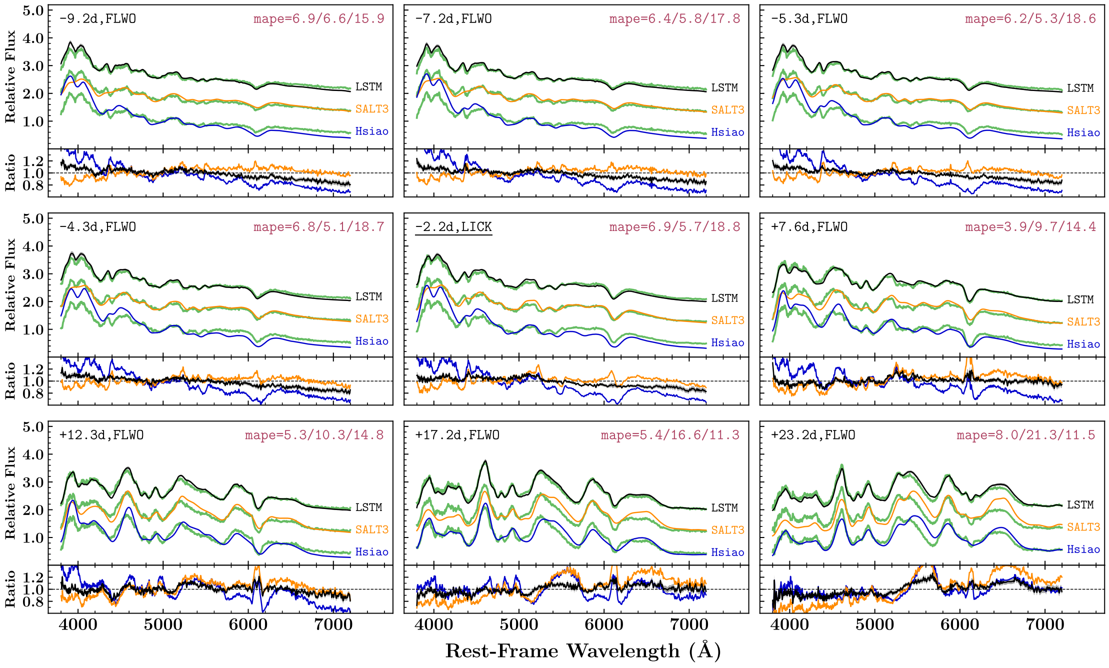
(c) SN 2005ke (Ia-91bg)   (d) SN 2012Z (Iax)

Figure 20. Same as Figure 16 but for the spectral sequence predicted from one spectrum at maximum light using LSTM neural networks for (a) Ia-91T object LSQ 12gdj, (b) Ia-99aa object SN 2008Z, (c) Ia-91bg object SN2005ke, and (d) Iax object SN 2012Z. The absorption features at Si II λ6355 are highlighted by a gray dashed box for Ia-91T object LSQ 12gdj and Ia-99aa object SN 2008Z.

spectrum are used as the loss function to estimate the optimal phase. As shown in Figure 26, for these six SNe, the MAPEs are at their minimum only when the input phases are close to the true phases. Figure 27 compares the predicted and observed phases. The root mean squared error (RMSE) of this method is ∼1.6 days for normal SNe Ia and ∼2.4 days for other subtypes.



(a) LSQ12gdj (Ia-91T)



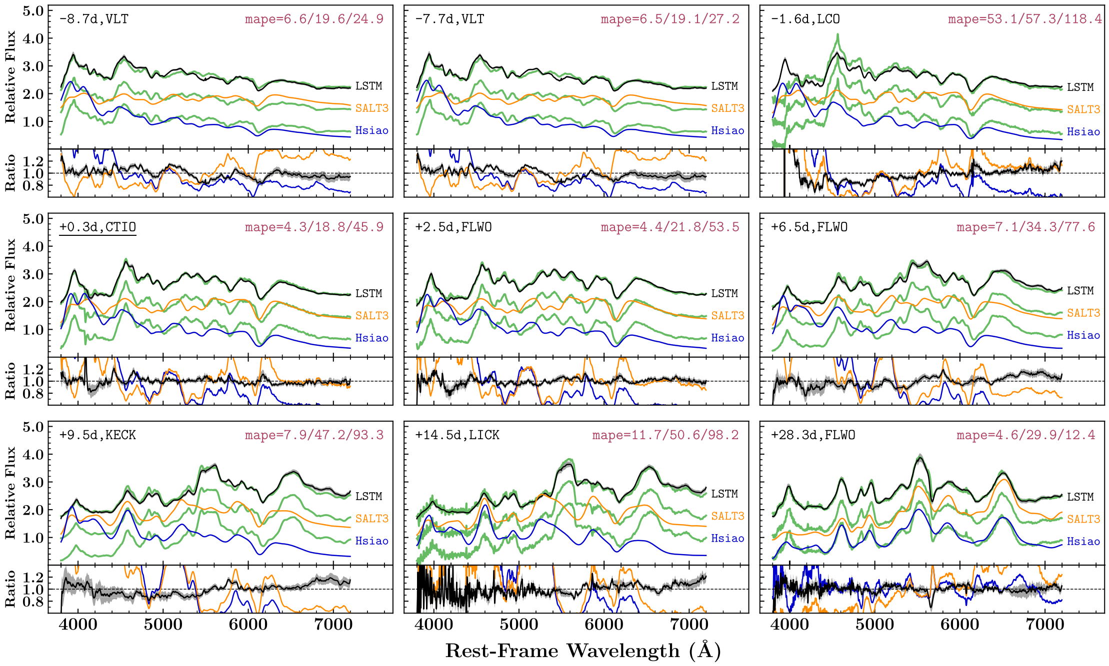
(b) SN 2008Z (Ia-99aa)

Figure 21. Comparison of the predicted spectra generated by different methods (black curves for LSTM; yellow curves for SALT3; blue curves for the Hsiao template) with the observed spectra (green curves) for (a) Ia-91T object LSQ 12gdj and (b) Ia-99aa object SN 2008Z. The predicted spectra from LSTM (black) and SALT3 (yellow) were obtained by fitting on the single spectrum at maximum light. The Hsiao template (blue) for LSQ 12gdj is constructed as described in Section 6.4. The panel format is the same as in Figure 17.

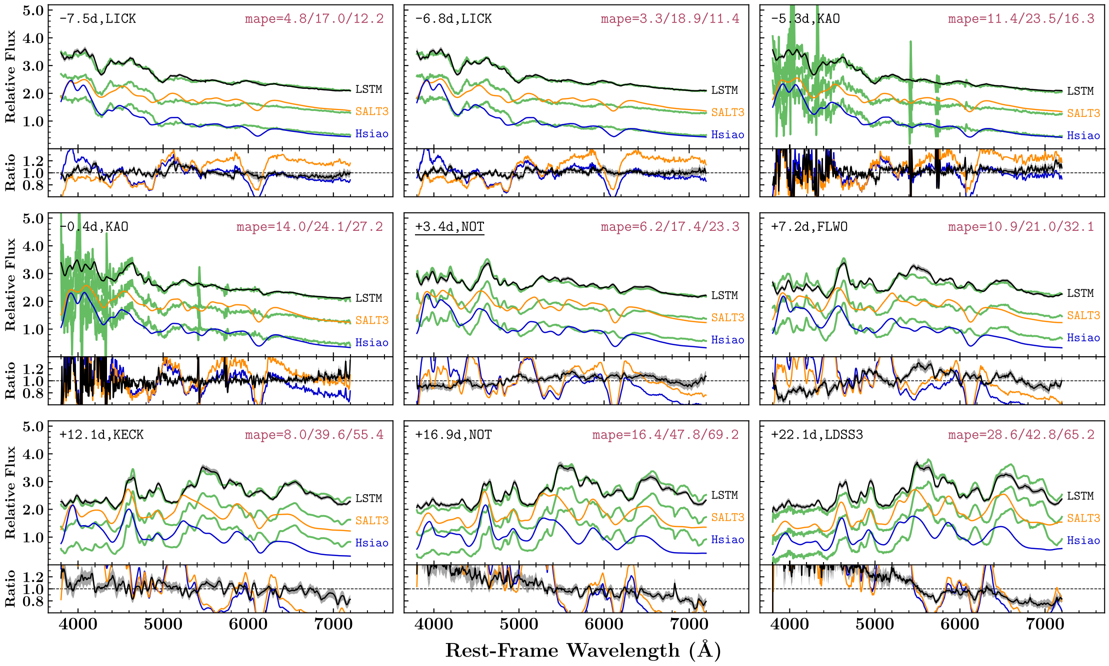
In addition, our publicly available software also accepts two spectra with unknown phases as input.

8. Discussions and Conclusions

We constructed a neural network–based algorithm to predict the spectral time series of SNe Ia from sparsely time-sampled spectroscopy. In order to train and test the models, we have compiled a spectral database of 3091 optical spectra from 361 SNe Ia. Given the heterogeneous nature of the spectroscopic observations, we homogenized the spectra to obtain uniform wavelength sampling, then performed flux recalibration through photometric observations. The spectral data were reprojected into a lower-dimensional space by FPCA parameterization. At the heart of the proposed method is the



(a) SN 2005ke (Ia-91bg)



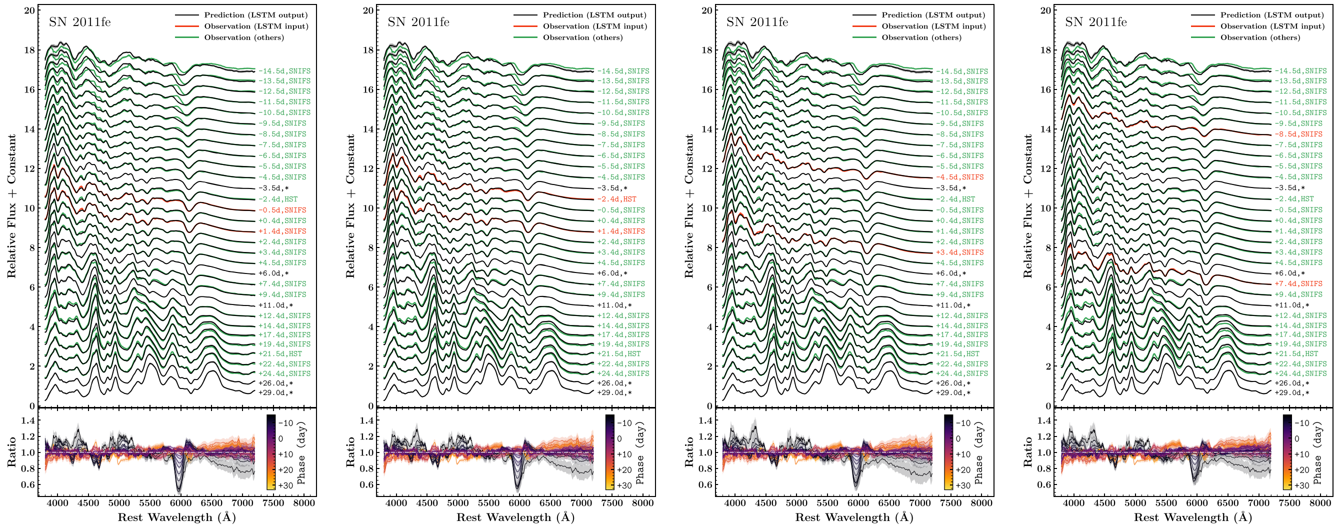
(b) SN 2012Z (Iax)

Figure 22. Same as Figure 21 but for (a) Ia-91bg object SN 2005ke and (b) Iax object SN 2012Z.

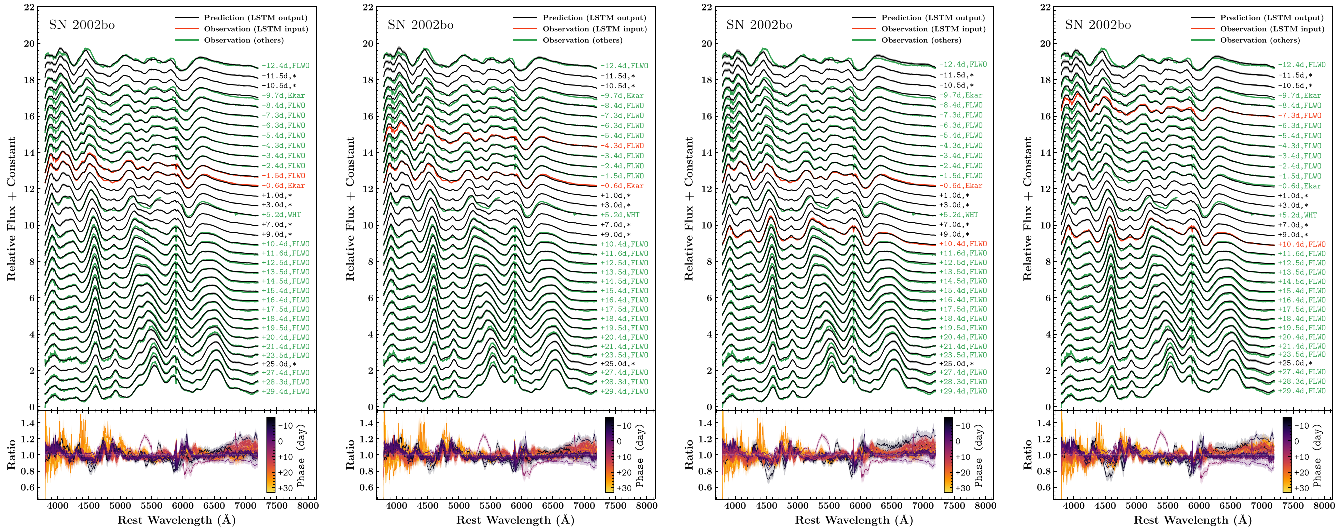
multilayer LSTM network. It allows predictions of spectra at any specific phase by using a sequence of observed spectra of an SN as input. The model thus enables the construction of spectral sequences from observations with limited time coverage.

With this method, we have constructed 361 spectral templates for the 361 SNe Ia in our data set, where each template is a spectral time series from −15 to 33 days relative to maximum light covering 3800–7200 Å in the rest frame. Running the testing procedures of the method with the test set, which was not involved in the template construction, has confirmed that our model can reliably build spectral sequences up to a median MAPE of 4.4%. No obvious bias appears in the distribution of the reconstruction accuracy over the evolution stages of the SNe. Although normal SNe Ia dominate the spectral data set, the neural network seems to be able to show reasonable performance on other less represented subtypes, such as Ia-91T, Ia-91bg, Ia-99aa, and Iax.

We further verified that the method can work well when only one observed spectrum is available. We used SN 2011fe and



(a) SN 2011fe, $\Delta P = 2$d (b) SN 2011fe, $\Delta P = 4$d (c) SN 2011fe, $\Delta P = 8$d (d) SN 2011fe, $\Delta P = 16$d



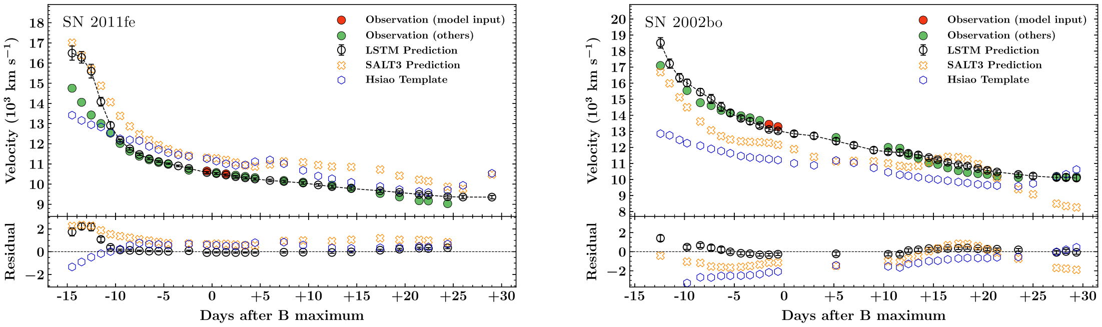
(e) SN 2002bo, $\Delta P = 2$d (f) SN 2002bo, $\Delta P = 5$d (g) SN 2002bo, $\Delta P = 11$d (h) SN 2002bo, $\Delta P = 18$d

Figure 23. Spectral sequence predicted from two spectra with phase difference ΔP using LSTM neural networks for SN 2011fe (top row) and SN 2002bo (*bottom row*). The format of each panel is the same as in Figure 16.

SN 2002bo as the representative cases of NV and HV objects, respectively, and found that their spectral sequences can be accurately predicted by using a single spectrum around maximum light. We also confirmed the model accuracy by measuring the Si II $\lambda6355$ velocity and spectral $B-V$ colors.

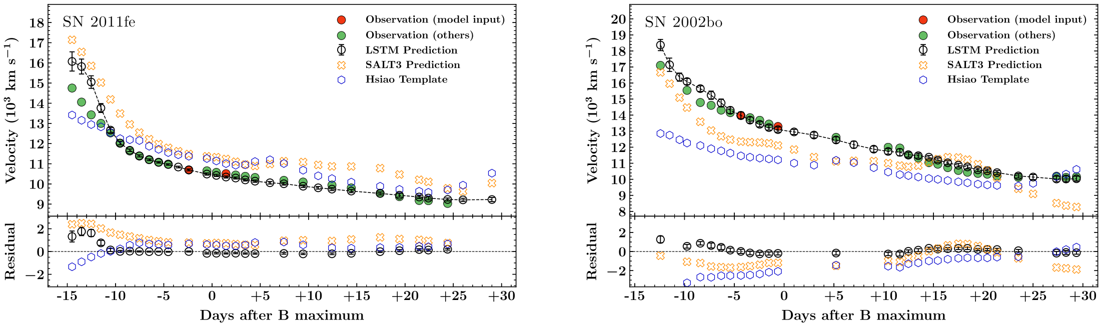
The difficulties in acquiring spectroscopic data have been the biggest challenge in SN studies and for future time-domain surveys. In SN cosmology, spectroscopy of SNe Ia is normally only attempted for a single epoch around the optical maximum. In the upcoming era of LSST/Rubin and the WFIRST/Roman survey, a large number of transients will be discovered whereas detailed spectroscopy will be impossible for the majority of the transients. The spectral follow-ups of the transients will need to be built with knowledge of the existing data set. The trend motivated us to develop this data-driven method for spectral inference. This is more than an interpolation tool. It allows reconstruction of complete spectral time series from limited available observations, as has been applied to the analyses of the Kepler-observed SN 2018agk in a recent study (Wang et al. 2021).

Our method can be used to investigate the spectral properties and reveal the intrinsic diversities among SNe Ia. Therefore, it may give new insight into the error budget of cosmological parameters. The direct applications of the proposed method in SN cosmology also include *k*-correction and searching for spectroscopic twin SNe. With spectroscopy from a single epoch, one may reconstruct the entire spectral time sequence to derive more reliable *k*-corrections. The method may also project SN observations from different epochs to the same epoch for direct comparisons to search for spectral twins (Fakhouri et al. 2015). For SN surveys, our method might be useful to optimize the spectroscopic follow-up strategies. The LSTM neural network allows for the phases of SNe to be estimated during an observing campaign



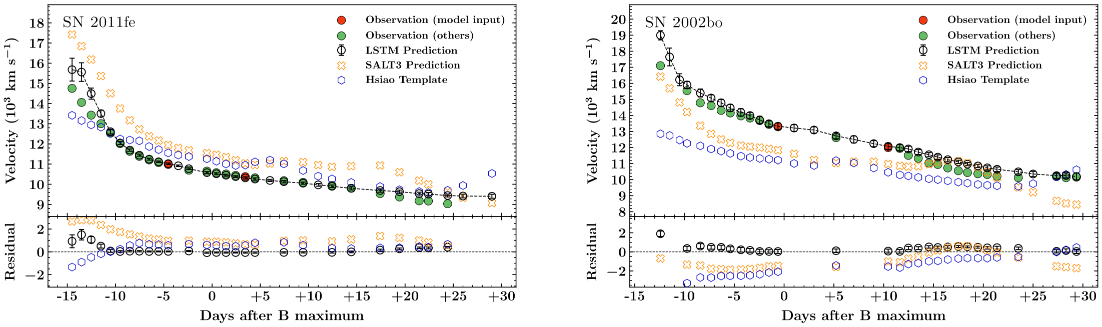
(a) SN 2011fe, $\Delta P = 2$d

(e) SN 2002bo, $\Delta P = 2$d



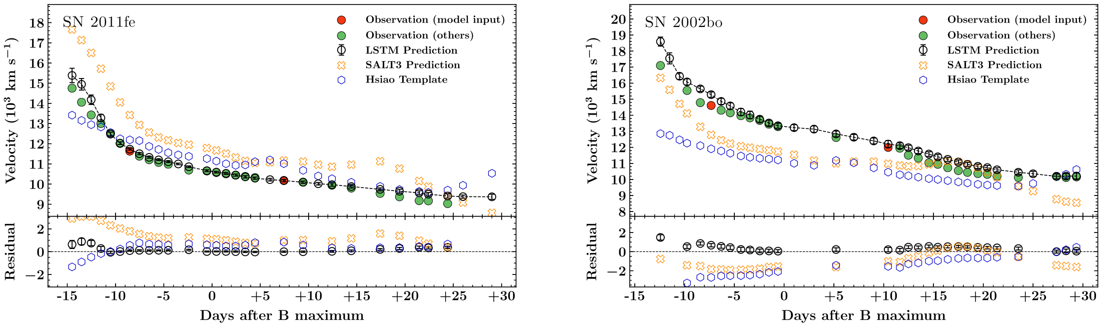
(b) SN 2011fe, $\Delta P = 4$d

(f) SN 2002bo, $\Delta P = 5$d



(c) SN 2011fe, $\Delta P = 8$d

(g) SN 2002bo, $\Delta P = 11$d



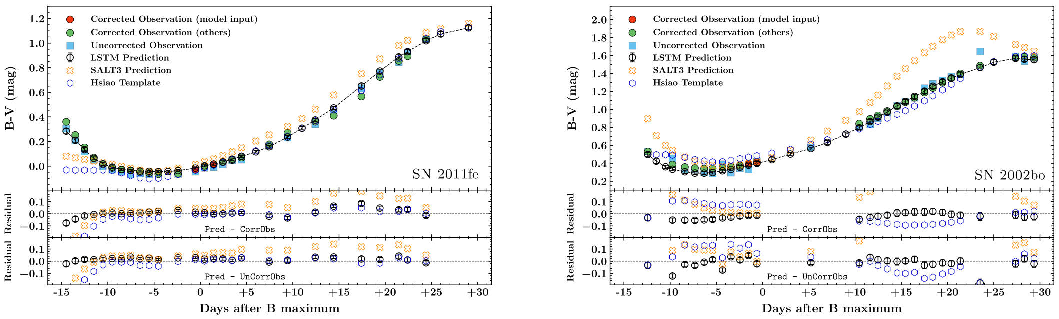
(d) SN 2011fe, $\Delta P = 16$d

(h) SN 2002bo, $\Delta P = 18$d

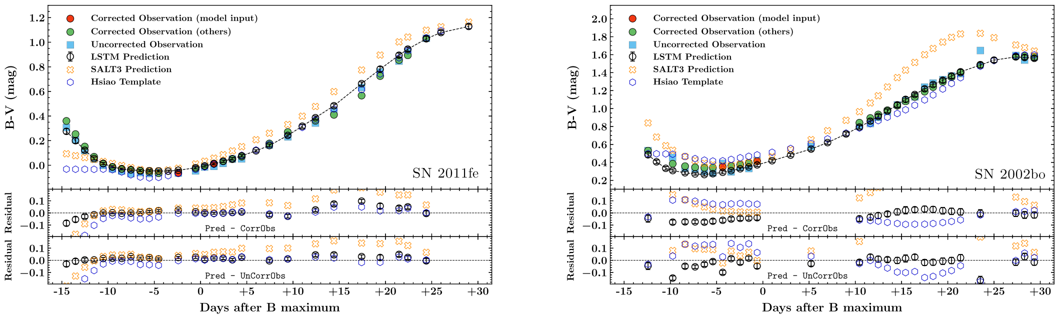
Figure 24. The Si II λ6355 velocity measured on the spectral sequence predicted from two spectra with phase difference ΔP using LSTM neural networks (black) and the corresponding observed spectra (green) for SN 2011fe (left column) and SN 2002bo (right column). The format of each panel is the same as in Figure 18.

and spectral follow-ups to be compared with LSTM predictions in real time.

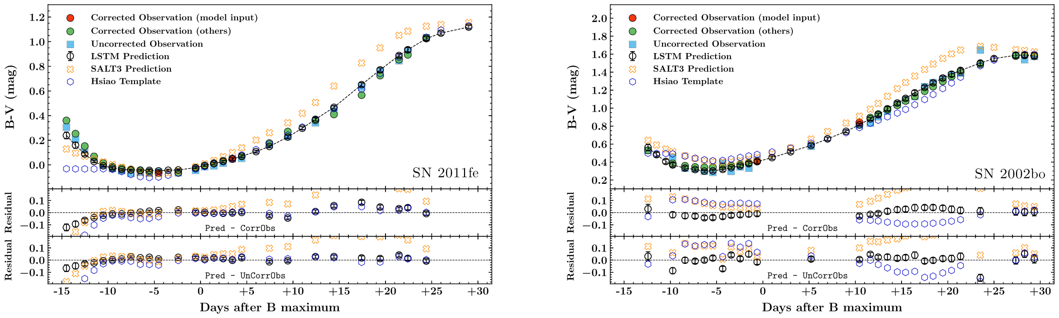
For future work, the number of input spectra K, which is fixed to 2 in this paper, may be increased to allow for direct spectral reconstruction with more input spectra. Our choice of wavelength coverage from 3800 to 7200 Å, which was chosen to include as many spectra as possible, may be extended when more data become available. The loss function MSE is a



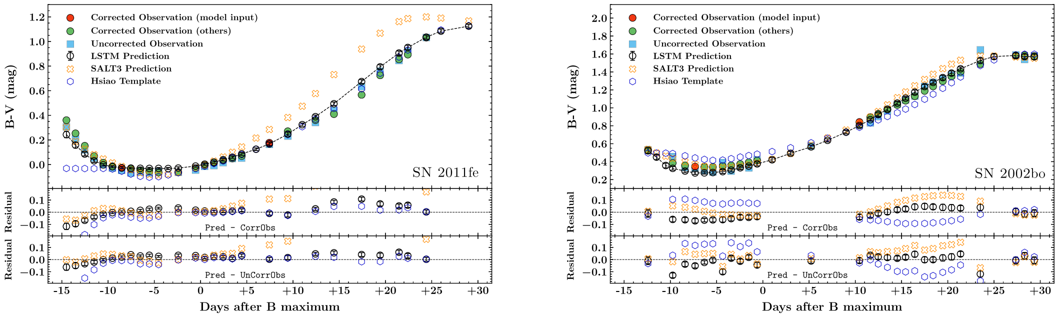
(a) SN 2011fe, $\Delta P = 2\text{d}$ (e) SN 2002bo, $\Delta P = 2\text{d}$



(b) SN 2011fe, $\Delta P = 4\text{d}$ (f) SN 2002bo, $\Delta P = 5\text{d}$



(c) SN 2011fe, $\Delta P = 8\text{d}$ (g) SN 2002bo, $\Delta P = 11\text{d}$



(d) SN 2011fe, $\Delta P = 16\text{d}$ (h) SN 2002bo, $\Delta P = 18\text{d}$

Figure 25. Synthetic $B - V$ color measured from the spectral sequence predicted from two spectra with phase difference ΔP using LSTM (black circles) and the corresponding observational data (green circles and blue squares) for SN 2011fe (left column) and SN 2002bo (right column). The format of each panel is the same as in Figure 19.

generic metric and prone to spectra with incorrect colors. We may develop more sophisticated loss functions that weigh more heavily on specific spectral features. The neural network uncertainties used in this work are still oversimplified. We certainly need to better understand the predictive uncertainty. This may include dividing it into epistemic uncertainty and aleatoric uncertainty (Kiureghian & Ditlevsen 2009; Gal 2016; Hortúa et al. 2020). The epistemic uncertainty is reducible by

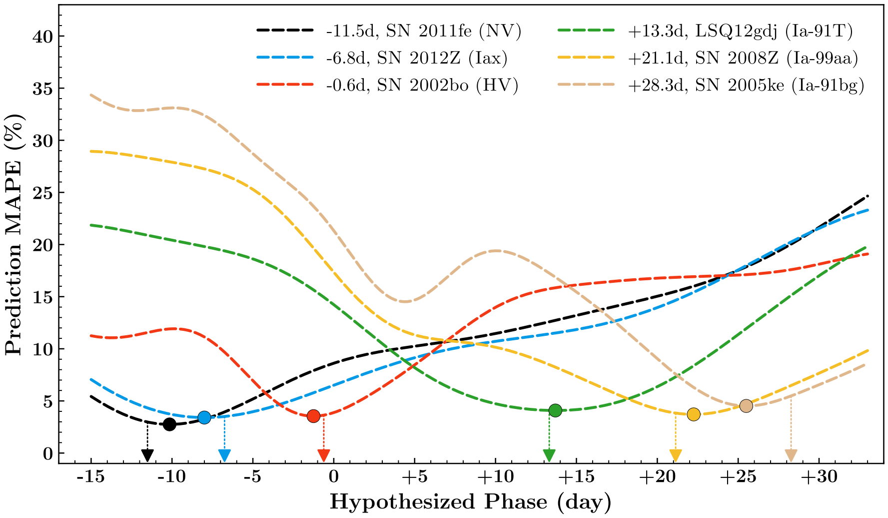


Figure 26. The MAPEs between the observed and predicted spectra using LSTM neural networks as a function of the hypothesized phase. This figure presents the results for six representative spectra with different subtypes and phases. The curves show the MAPEs for SN 2011fe (black), SN 2012Z (light blue), SN 2020bo (red), LSQ 12gdj (green), SN 2008Z (yellow), and SN 2005ke (brown). The filled circles show the locations of the minimum MAPEs, and the arrows point to the true phases as derived from photometric light curves.

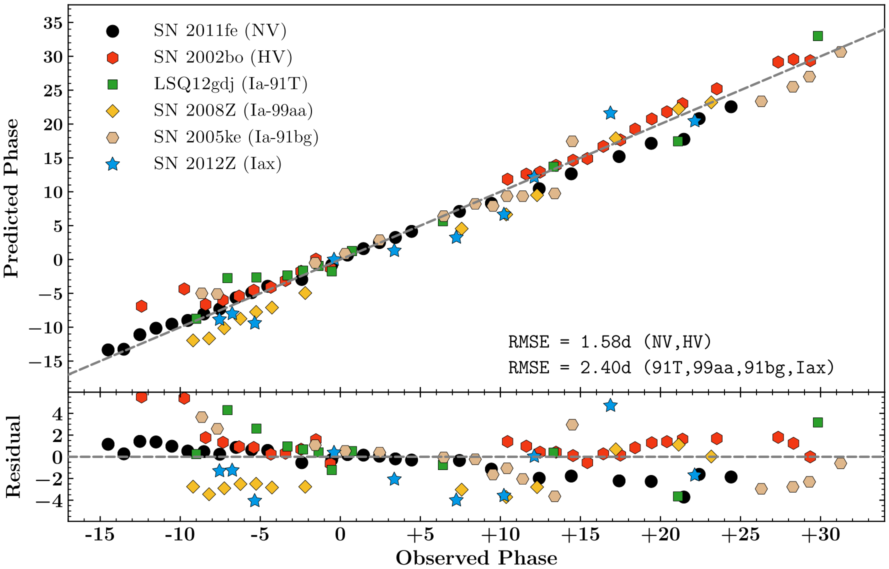


Figure 27. Predicted phase versus observed phase for six representative SNe Ia discussed in Section 7. The predicted phase is determined by minimizing the MAPE, as demonstrated in Figure 26. The legend shows the symbols used for each SN. The upper panel shows the correlation between the predicted and observed phases. The dashed gray line shows the relation when the predicted and observed phases are equal. The residuals of the predictions are shown in the lower panel. The RMSEs of the phase predictions are given in the bottom right corner of the upper panel for the SNe representing NV, HV, Ia-91T, Ia-99aa, Ia-91bg, and Iax.

increasing the number of observations, as the limited training set can be insufficient for the entire feature space, while the aleatoric uncertainty captures the noise of intrinsic randomness (such as photon noise) and cannot be reduced by collecting more data.

L.H. acknowledges support by the National Natural Science Foundation of China (11761141001) and the Major Science and Technology Project of Qinghai Province (2019-ZJ-A10). L.W. and X.C. acknowledge support from NSF grant AST-1817099. We acknowledge the Supernova Polarimetry Project for providing the high-S/N VLT data set. We thank Jiawen Yang for helpful discussions on the package developments. The authors are also grateful to an anonymous referee whose report has significantly improved this manuscript.

ORCID iDs

Lei Hu https://orcid.org/0000-0001-7201-1938
Xingzhuo Chen https://orcid.org/0000-0003-3021-4897
Lifan Wang https://orcid.org/0000-0001-7092-9374

References

Abbott, T. M. C., Allam, S., Andersen, P., et al. 2019, ApJL, 872, L30
Amanullah, R., Lidman, C., Rubin, D., et al. 2010, ApJ, 716, 712
Amanullah, R., Johansson, J., Goobar, A., et al. 2015, MNRAS, 453, 3300
Ardeberg, A., & de Groot, M. 1973, A&A, 28, 295
Ashall, C., Mazzali, P., Bersier, D., et al. 2014, MNRAS, 445, 4427
Ashall, C., Mazzali, P. A., Stritzinger, M. D., et al. 2018, MNRAS, 477, 153
Bailey, S., Aldering, G., Antilogus, P., et al. 2009, A&A, 500, L17
Barbary, K., Barclay, T., Biswas, R., et al. 2016, SNCosmo: Python Library for Supernova Cosmology, Astrophysics Source Code Library, ascl:1611.017
Benetti, S., Meikle, P., Stehle, M., et al. 2004, MNRAS, 348, 261

Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, A22

Blondin, S., Dessart, L., Leibundgut, B., et al. 2006, AJ, 131, 1648

Blondin, S., Matheson, T., Kirshner, R. P., et al. 2012, AJ, 143, 126

Branch, D., & Wheeler, J. C. 2017, Supernova Explosions (Berlin: Springer)

Branch, D., Dang, L. C., Hall, N., et al. 2006, PASP, 118, 560

Brown, P. 2014, in Proc. Swift: 10 Years of Discovery, Vol. 233 (Trieste: PoS), 125

Brown, P. J., Breeveld, A. A., Holland, S., Kuin, P., & Pritchard, T. 2014, Ap&SS, 354, 89

Burns, C. R., Stritzinger, M., Phillips, M. M., et al. 2011, AJ, 141, 19

Cao, Y., Kulkarni, S. R., Howell, D. A., et al. 2015, Natur, 521, 328

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Cartier, R., Sullivan, M., Firth, R. E., et al. 2017, MNRAS, 464, 4476

Chen, X., Hu, L., & Wang, L. 2020, ApJS, 250, 12

Chollet, F., et al. 2015, Keras, https://keras.io

Chotard, N., Gangler, E., Aldering, G., et al. 2011, A&A, 529, L4

Cikota, A., & Pauldrach, A. 2018, PhD Thesis, Ludwig-Maximilians University of Munich

Cikota, A., Patat, F., Wang, L., et al. 2019, MNRAS, 490, 578

Contreras, C., Phillips, M. M., Burns, C. R., et al. 2018, ApJ, 859, 24

Dai, A. M., & Le, Q. V. 2015, in Advances in Neural Information Processing Systems, ed. C. Cortes et al., Vol. 28 (Red Hook, NY: Curran Associates, Inc.), 3079

Danziger, I. J., della Valle, M., Kissler-Patig, M., & Storm, J. 1998, MmSAI, 69, 245

Dimitriadis, G., Foley, R. J., Rest, A., et al. 2019, ApJL, 870, L1

Eapen, J., Bein, D., & Verma, A. 2019, in 2019 IEEE 9th Annual Computing and Communication Workshop and Conf. (CCWC) (Piscataway, NJ: IEEE), 0264

Fakhouri, H. K., Boone, K., Aldering, G., et al. 2015, ApJ, 815, 58

Filippenko, A. V. 1997, ARA&A, 35, 309

Filippenko, A. V., Richmond, M. W., Matheson, T., et al. 1992, ApJL, 384, L15

Firth, R. E., Sullivan, M., Gal-Yam, A., et al. 2015, MNRAS, 446, 3895

Folatelli, G., Morrell, N., Phillips, M. M., et al. 2013, ApJ, 773, 53

Foley, R. J., & Kasen, D. 2011, ApJ, 729, 55

Foley, R. J., Challis, P. J., Chornock, R., et al. 2013, ApJ, 767, 57

Friedman, A. S., Wood-Vasey, W. M., Marion, G. H., et al. 2015, ApJS, 220, 9

Gal, Y. 2016, PhD Thesis, University of Cambridge

Gal, Y., & Ghahramani, Z. 2015, in Proc. of the 33rd Int. Conf. on Machine Learning, Vol. 48 (Brookline, MA: JMLR), 1050

Gal, Y., & Ghahramani, Z. 2016, in Advances in Neural Information Processing Systems 29 (Red Hook, NY: Curran Associates, Inc.), 1019

Gall, C., Stritzinger, M. D., Ashall, C., et al. 2018, A&A, 611, A58

Ganeshalingam, M., Li, W., & Filippenko, A. V. 2013, MNRAS, 433, 2240

Garavini, G., Folatelli, G., Goobar, A., et al. 2004, AJ, 128, 387

Gómez, C., Neira, M., Hernández Hoyos, M., Arbeláez, P., & Forero-Romero, J. E. 2020, MNRAS, 499, 3130

Goldstein, D. A., D'Andrea, C. B., Fischer, J. A., et al. 2015, AJ, 150, 82

Gomez, G., Lopez, R., & Sanchez, F. 1996, AJ, 112, 2094

González-Gaitán, S., Hsiao, E. Y., Pignata, G., et al. 2014, ApJ, 795, 142

Graham, M. L., Kumar, S., Hosseinzadeh, G., et al. 2017, MNRAS, 472, 3437

Guy, J., Astier, P., Nobili, S., Regnault, N., & Pain, R. 2005, A&A, 443, 781

Guy, J., Astier, P., Baumont, S., et al. 2007, A&A, 466, 11

Hachinger, S., Mazzali, P. A., Sullivan, M., et al. 2013, MNRAS, 429, 2228

Hall, P., Müller, H.-G., & Wang, J.-L. 2006, AnSta, 34, 1493

He, S., Wang, L., & Huang, J. Z. 2018, ApJ, 857, 110

Hicken, M., Challis, P., Jha, S., et al. 2009, ApJ, 700, 331

Hicken, M., Challis, P., Kirshner, R. P., et al. 2012, ApJS, 200, 12

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2012, arXiv:1207.0580

Hochreiter, S., & Schmidhuber, J. 1997, Neural Comput., 9, 1735

Hortúa, H. J., Volpi, R., Marinelli, D., & Malagò, L. 2020, PhRvD, 102, 103509

Hosseinzadeh, G., Sand, D. J., Valenti, S., et al. 2017, ApJL, 845, L11

Hoyle, F., & Fowler, W. A. 1960, ApJ, 132, 565

Hsiao, E. Y., Conley, A., Howell, D. A., et al. 2007, ApJ, 663, 1187

Huang, X., Raha, Z., Aldering, G., et al. 2017, ApJ, 836, 157

Jha, S., Riess, A. G., & Kirshner, R. P. 2007, ApJ, 659, 122

Jha, S., Kirshner, R. P., Challis, P., et al. 2006, AJ, 131, 527

Jiang, J.-A., Doi, M., Maeda, K., et al. 2017, Natur, 550, 80

Jiang, J.-a., Maeda, K., Kawabata, M., et al. 2021, ApJL, 923, L8

Kasen, D. 2010, ApJ, 708, 1025

Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, ApJ, 715, 743

Kenworthy, W. D., Jones, D. O., Dai, M., et al. 2021, ApJ, 923, 265

Kerzendorf, W. E., Vogl, C., Buchner, J., et al. 2021, ApJL, 910, L23

Kiureghian, A. D., & Ditlevsen, O. 2009, Struct. Saf., 31, 105

Knop, R. A., Aldering, G., Amanullah, R., et al. 2003, ApJ, 598, 102

Kou, S., Chen, X., & Liu, X. 2020, ApJ, 890, 177

Kowalski, M., Rubin, D., Aldering, G., et al. 2008, ApJ, 686, 749

Krisciunas, K., Prieto, J. L., Garnavich, P. M., et al. 2006, AJ, 131, 1639

Krisciunas, K., Suntzeff, N. B., Espinoza, J., et al. 2017a, RNAAS, 1, 36

Krisciunas, K., Contreras, C., Burns, C. R., et al. 2017b, AJ, 154, 211

Krisciunas, K., Suntzeff, N. B., Candia, P., et al. 2003, AJ, 125, 166

Krisciunas, K., Suntzeff, N. B., Phillips, M. M., et al. 2004, AJ, 128, 3034

Kromer, M., Fremling, C., Pakmor, R., et al. 2016, MNRAS, 459, 4428

Léget, P. F., Gangler, E., Mondon, F., et al. 2020, A&A, 636, A46

Leibundgut, B., Kirshner, R. P., Phillips, M. M., et al. 1993, AJ, 105, 301

Li, L., Zhang, J., Dai, B., et al. 2022, ApJ, 924, 35

Li, W., Wang, X., Bulla, M., et al. 2021, ApJ, 906, 99

Li, L., Wang, X., Zhang, J., et al. 2018, MNRAS, 478, 4575

Li, W., Wang, X., Vinkó, J., et al. 2019, ApJ, 870, 12

Lira, P., Suntzeff, N. B., Phillips, M. M., et al. 1998, AJ, 115, 234

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201

Maeda, K., Benetti, S., Stritzinger, M., et al. 2010, Natur, 466, 82

Maguire, K., Sullivan, M., Thomas, R. C., et al. 2011, MNRAS, 418, 747

Maguire, K., Sullivan, M., Patat, F., et al. 2013, MNRAS, 436, 222

Mahabal, A., Rebbapragada, U., Walters, R., et al. 2019, PASP, 131, 038002

Marion, G. H., Brown, P. J., Vinkó, J., et al. 2016, ApJ, 820, 92

Maund, J. R., Höflich, P., Patat, F., et al. 2010, ApJL, 725, L167

Möller, A., Ruhlmann-Kleider, V., Leloup, C., et al. 2016, JCAP, 2016, 008

Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019a, PASP, 131, 118002

Muthukrishna, D., Parkinson, D., & Tucker, B. E. 2019b, ApJ, 885, 85

Neil, D., Pfeiffer, M., & Liu, S.-C. 2016, in Proc. 30th Int. Conf. on Neural Information Processing Systems, NIPS'16 (Red Hook, NY: Curran Associates Inc.), 3889

Nugent, P., Kim, A., & Perlmutter, S. 2002, PASP, 114, 803

Pan, Y. C., Foley, R. J., Kromer, M., et al. 2015, MNRAS, 452, 4307

Parrent, J. T., Thomas, R. C., Fesen, R. A., et al. 2011, ApJ, 732, 30

Parrent, J. T., Howell, D. A., Friesen, B., et al. 2012, ApJL, 752, L26

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825

Peng, J., & Paul, D. 2009, J. Comput. Graph. Stat., 18, 995

Perlmutter, S., Gabi, S., Goldhaber, G., et al. 1997, ApJ, 483, 565

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Phillips, M. M. 1993, ApJL, 413, L105

Phillips, M. M., Wells, L. A., Suntzeff, N. B., et al. 1992, AJ, 103, 1632

Richmond, M. W., Treffers, R. R., Filippenko, A. V., et al. 1995, AJ, 109, 2121

Riess, A. G. 2019, NatRP, 2, 10

Riess, A. G., Press, W. H., & Kirshner, R. P. 1996, ApJ, 473, 88

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Riess, A. G., Macri, L., Li, W., et al. 2009, ApJS, 183, 109

Sandage, A. R., Kron, R. G., Longair, M. S., Binggeli, B., & Buser, R. 1995, The Deep Universe (Berlin: Springer)

Sasdelli, M., Ishida, E. E. O., Vilalta, R., et al. 2016, MNRAS, 461, 2044

Saunders, C., Aldering, G., Antilogus, P., et al. 2018, ApJ, 869, 167

Savitzky, A., & Golay, M. J. E. 1964, AnaCh, 36, 1627

Scalzo, R. A., Childress, M., Tucker, B., et al. 2014, MNRAS, 445, 30

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103

Schuster, M., & Paliwal, K. K. 1997, ITSP, 45, 2673

Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, ApJ, 859, 101

Shappee, B. J., Piro, A. L., Holoien, T. W. S., et al. 2016, ApJ, 826, 144

Shappee, B. J., Holoien, T. W. S., Drout, M. R., et al. 2019, ApJ, 870, 13

Siebert, M. R., Foley, R. J., Jones, D. O., et al. 2019, MNRAS, 486, 5785

Silverman, J. M., Foley, R. J., Filippenko, A. V., et al. 2012, MNRAS, 425, 1789

Smitka, M. T., Brown, P. J., Kuin, P., & Suntzeff, N. B. 2016, PASP, 128, 034501

Smitka, M. T., Brown, P. J., Suntzeff, N. B., et al. 2015, ApJ, 813, 30

Srivastav, S., Anupama, G. C., Sahu, D. K., & Ravikumar, C. D. 2017, MNRAS, 466, 2436

Srivastav, S., Ninan, J. P., Kumar, B., et al. 2016, MNRAS, 457, 1000

Stahl, B. E., Martínez-Palomera, J., Zheng, W., et al. 2020, MNRAS, 496, 3553

Stahl, B. E., Zheng, W., de Jaeger, T., et al. 2019, MNRAS, 490, 3882

Stanishev, V., Goobar, A., Benetti, S., et al. 2007, A&A, 469, 645

Stritzinger, M. D., Hamuy, M., Suntzeff, N. B., et al. 2002, AAS Meeting, 200, 95.03

Stritzinger, M. D., Phillips, M. M., Boldt, L. N., et al. 2011, AJ, 142, 156

Stritzinger, M. D., Hsiao, E., Valenti, S., et al. 2014, A&A, 561, A146

Stritzinger, M. D., Valenti, S., Hoeflich, P., et al. 2015, A&A, 573, A2
Sullivan, M., Conley, A., Howell, D. A., et al. 2010, MNRAS, 406, 782
Suzuki, N., Rubin, D., Lidman, C., et al. 2012, ApJ, 746, 85
Takahashi, I., Suzuki, N., Yasuda, N., et al. 2020, PASJ, 72, 89
Taylor, G., Lidman, C., Tucker, B. E., et al. 2021, MNRAS, 504, 4111
Tripp, R. 1998, A&A, 331, 815
Tsvetkov, D. Y., Shugarov, S. Y., Volkov, I. M., et al. 2013, CoSka, 43, 94
Uddin, S. A., Burns, C. R., Phillips, M. M., et al. 2020, ApJ, 901, 143
van Genderen, A. M. 1975, A&A, 45, 429
Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020, ApJ, 905, 94
Vogl, C., Kerzendorf, W. E., Sim, S. A., et al. 2020, A&A, 633, A88
Wagers, A., Wang, L., & Asztalos, S. 2010, ApJ, 711, 711
Walker, E. S., Baltay, C., Campillay, A., et al. 2015, ApJS, 219, 13
Wang, L., Goldhaber, G., Aldering, G., & Perlmutter, S. 2003, ApJ, 590, 944
Wang, L., Höflich, P., & Wheeler, J. C. 1997, ApJL, 483, L29
Wang, L., & Wheeler, J. C. 2008, ARA&A, 46, 433
Wang, L., Contreras, C., Hu, M., et al. 2020, ApJ, 904, 14
Wang, Q., Rest, A., Zenati, Y., et al. 2021, ApJ, 923, 167
Wang, X., Li, W., & Filippenko, A. V. 2009a, AAS Meeting, 213, 312.04
Wang, X., Filippenko, A. V., Ganeshalingam, M., et al. 2009b, ApJL, 699, L139
Wang, X., Wang, L., Filippenko, A. V., Zhang, T., & Zhao, X. 2013, Sci, 340, 170
Wee, J., Chakraborty, N., Wang, J., & Penprase, B. E. 2018, ApJ, 863, 90
Weyant, A., Wood-Vasey, W. M., Joyce, R., et al. 2018, AJ, 155, 201
Yamanaka, M., Maeda, K., Kawabata, M., et al. 2014, ApJL, 782, L35
Yamanaka, M., Maeda, K., Kawabata, K. S., et al. 2015, ApJ, 806, 191
Yang, Y., Hoeflich, P., Baade, D., et al. 2020, ApJ, 902, 46
Yaron, O., & Gal-Yam, A. 2012, PASP, 124, 668
Zaremba, W., Sutskever, I., & Vinyals, O. 2014, arXiv:1409.2329
Zhang, J.-J., Wang, X.-F., Sasdelli, M., et al. 2016, ApJ, 817, 114
Zhao, X., Wang, X., Maeda, K., et al. 2015, ApJS, 220, 20
Zheng, W., Kelly, P. L., & Filippenko, A. V. 2018, ApJ, 858, 104