

Lawrence Berkeley National Laboratory

LBL Publications

Title

DeepZipper: A Novel Deep-learning Architecture for Lensed Supernovae Identification

Permalink

<https://escholarship.org/uc/item/5w89j8g6>

Journal

The Astrophysical Journal, 927(1)

ISSN

0004-637X

Authors

Morgan, R
Nord, B
Bechtol, K
[et al.](#)

Publication Date

2022-03-01

DOI

10.3847/1538-4357/ac5178

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



DeepZipper: A Novel Deep-learning Architecture for Lensed Supernovae Identification

R. Morgan^{1,2,3}, B. Nord^{2,4,5}, K. Bechtol^{1,6}, S. J. González¹, E. Buckley-Geer^{2,4}, A. Möller⁷, J. W. Park^{8,9},
 A. G. Kim¹⁰, S. Birrer^{8,9}, M. Aguena¹¹, J. Annis², S. Bocquet¹², D. Brooks¹³, A. Carnero Rosell¹¹,
 M. Carrasco Kind^{14,15}, J. Carretero¹⁶, R. Cawthon¹⁷, L. N. da Costa^{11,18}, T. M. Davis¹⁹, J. De Vicente²⁰, P. Doel¹³,
 I. Ferrero²¹, D. Friedel¹⁴, J. Frieman^{2,5}, J. García-Bellido²², M. Gatti²³, E. Gaztanaga^{24,25}, G. Giannini¹⁶, D. Gruen¹²,
 R. A. Gruendl^{14,15}, G. Gutierrez², D. L. Hollowood²⁶, K. Honscheid^{27,28}, D. J. James²⁹, K. Kuehn^{30,31},
 N. Kuropatkin², M. A. G. Maia^{11,18}, R. Miquel^{16,32}, A. Palmese³³, F. Paz-Chinchón^{14,34}, M. E. S. Pereira^{35,36},
 A. Pieres^{11,18}, A. A. Plazas Malagón³⁷, K. Reil⁹, A. Roodman^{9,38}, E. Sanchez²⁰, M. Smith³⁹, E. Suchyta⁴⁰,
 M. E. C. Swanson³⁹, G. Tarle³⁵, and C. To²⁷

¹ Physics Department, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706, USA; robert.morgan@wisc.edu

² Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA

³ Legacy Survey of Space and Time Corporation Data Science Fellowship Program, USA

⁴ Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA

⁵ Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

⁶ Legacy Survey of Space and Time, 933 North Cherry Avenue, Tucson, AZ 85721, USA

⁷ Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Victoria 3122, Australia

⁸ Kavli Institute for Particle Astrophysics and Cosmology, Department of Physics, Stanford University, Stanford, CA 94305, USA

⁹ SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

¹⁰ Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

¹¹ Laboratório Interinstitucional de e-Astronomia—LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ—20921-400, Brazil

¹² Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstraße 1, D-81679 Munich, Germany

¹³ Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK

¹⁴ Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark Street, Urbana, IL 61801, USA

¹⁵ Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801, USA

¹⁶ Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona) Spain

¹⁷ Physics Department, William Jewell College, Liberty, MO 64068 USA

¹⁸ Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ—20921-400, Brazil

¹⁹ School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia

²⁰ Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

²¹ Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway

²² Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

²³ Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

²⁴ Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain

²⁵ Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain

²⁶ Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA

²⁷ Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA

²⁸ Department of Physics, The Ohio State University, Columbus, OH 43210, USA

²⁹ Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

³⁰ Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia

³¹ Lowell Observatory, 1400 Mars Hill Road, Flagstaff, AZ 86001, USA

³² Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain

³³ Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA

³⁴ Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

³⁵ Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

³⁶ Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, D-21029 Hamburg, Germany

³⁷ Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA

³⁸ Kavli Institute for Particle Astrophysics & Cosmology, Stanford University, P.O. Box 2450, Stanford, CA 94305, USA

³⁹ School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK

⁴⁰ Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Received 2021 December 2; revised 2022 January 27; accepted 2022 February 1; published 2022 March 9

Abstract

Large-scale astronomical surveys have the potential to capture data on large numbers of strongly gravitationally lensed supernovae (LSNe). To facilitate timely analysis and spectroscopic follow-up before the supernova fades, an LSN needs to be identified soon after it begins. To quickly identify LSNe in optical survey data sets, we designed ZipperNet, a multibranch deep neural network that combines convolutional layers (traditionally used for images) with long short-term memory layers (traditionally used for time series). We tested ZipperNet on the task of classifying objects from four categories—no lens, galaxy-galaxy lens, lensed Type-Ia supernova, lensed core-collapse supernova—within high-fidelity simulations of three cosmic survey data sets: the Dark Energy Survey, Rubin Observatory's Legacy Survey of Space and Time (LSST), and a Dark Energy Spectroscopic Instrument (DESI) imaging survey. Among our results, we find that for the LSST-like data set, ZipperNet classifies LSNe with

a receiver operating characteristic area under the curve of 0.97, predicts the spectroscopic type of the lensed supernovae with 79% accuracy, and demonstrates similarly high performance for LSNe 1–2 epochs after first detection. We anticipate that a model like ZipperNet, which simultaneously incorporates spatial and temporal information, can play a significant role in the rapid identification of lensed transient systems in cosmic survey experiments.

Unified Astronomy Thesaurus concepts: [Supernovae \(1668\)](#); [Strong gravitational lensing \(1643\)](#); [Neural networks \(1933\)](#)

1. Introduction

Strong gravitational lensing is an expansive probe of both astrophysics and cosmology. In systems with strong lensing, light from a background object (the source) is deflected by the gravitational potential of an interposed foreground object (the lens), producing characteristic features such as arcs, Einstein rings, and/or multiple images (Treu 2010). A key subclass of lenses, strongly lensed transients, have time-variable brightness in the source galaxy. Due to the cosmological distances involved in strong lensing, the most common transient objects to observe lensed are quasars, which vary in brightness on timescales of several years (Hook et al. 1994; Helfand et al. 2001), and supernovae (SNe), which can reach peak brightness within days and then dim over the course of weeks to months (Mihalas 1963).

One of the principle features of lensed transients is the time delay between the arrival of photons that take different paths around the lenses. Photons from multiple images of source objects travel different distances to Earth and experience different magnitudes of the gravitational potential due to the geometry of the lensing system. Given the constant speed of light, the differences in the path lengths and gravitational potentials traversed by the photons produce an offset in the arrival times for photons that are emitted at the same point in the source light curve. Therefore, sources with time-varying brightness in these systems exhibit an observable time delay between the individual images of the lensed source (Refsdal 1964).

Strongly lensed transients are particularly useful for a wide range of investigations. For example, the magnification of these background sources and their environments can reveal new information about the eruption and prevalence of these objects at earlier times in the universe (Treu & Marshall 2016). Time-delay cosmography (TDC) is a technique that uses the time-varying brightness of compact systems that have undergone strong gravitational lensing (SL) to perform a geometric measurement of H_0 (Refsdal 1964). TDC entails a measurement of this time delay and modeling of the full strong-lensing system. H_0 is then inversely proportional to the time delay between photons. This technique has been used with quasars (persistent variable objects) to measure H_0 to better than 5% precision in time-varying SL systems (Wong et al. 2019; Shajib et al. 2020).

The cosmic expansion rate today, H_0 , is a critical parameter for understanding the evolution of the universe. There are multiple probes of H_0 , including the cosmic distance ladder, extrapolation from the cosmic microwave background (CMB), and strongly lensed variable sources, such as SNe and quasars. There is currently a significant tension among the probes (Freedman 2021), particularly between early-universe probes like the CMB (Planck Collaboration et al. 2020) and late-universe probes like the type-Ia SNe distance ladder anchored on Cepheid variable stars (Riess et al. 2021). TDC, as a probe of H_0 , does not rely on anchoring measurements to late-

universe objects or extrapolating from early-universe physics, so it offers a new perspective on the expansion rate of the universe today.

Aside from quasars, SNe are the other major class of common time-varying sources that are bright enough to be detected at cosmological distances. SNe present an experimental challenge in TDC because they become bright and fade on the scale of months, therefore requiring rapid identification and analysis to obtain a measurement of the time delay. However, the larger variability on shorter timescales compared to quasars offers a competitive advantage in measuring the time delay. Another advantage of LSNe is that a common subclass of SNe (SNe-Ia) can facilitate highly accurate modeling of the lensing gravitational potential—and therefore a more precise H_0 measurement (Kolatt & Bartelmann 1998; Oguri & Kawano 2003; Foxley-Marrable et al. 2018a; Birrer et al. 2021a) as a result of their standardizable brightness (Tripp 1998). To date, only a handful of LSNe have been detected (Amanullah et al. 2011; Quimby et al. 2014; Kelly et al. 2015; Rodney et al. 2021), and only two LSNe-Ia have been discovered (Rodney et al. 2015; Goobar et al. 2017).

Large optical surveys are ideal data sets to search for LSNe because high area coverage and a return to the same field multiple times increase the chances of observing an LSN, although the rarity of LSNe still makes their detection a challenging problem. For example, based on the area covered, imaging depth, and length of observations, only 0.5–2 LSNe are expected to be in DES data (Oguri 2019). Looking forward to the next era of optical survey astronomy, the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) plans to cover the entire southern sky to greater depth and with higher cadence than any predecessor survey, e.g., the Dark Energy Survey (DES; Diehl et al. 2018). Preliminary forecasts indicate that the Rubin Observatory will detect hundreds to thousands of these systems (Goldstein & Nugent 2016; Oguri 2019; Wojtak et al. 2019). Each detected time-varying SL system has the potential to produce an independent measurement of H_0 , meaning that the measured statistical precision on H_0 using this technique will scale with $1/\sqrt{N}$, where N is the number of detected systems. Therefore, one of the main goals in the LSST era is the identification and characterization of as many LSNe-Ia as possible to precisely measure H_0 (The LSST Dark Energy Science Collaboration et al. 2018).

Fast and robust algorithms for detecting LSNe-Ia are essential to keep pace with the data stream of LSST and surveys with comparable data size. Furthermore, because SNe fade after their explosion, they must be identified rapidly to facilitate follow-up observations and more detailed characterization of the system for lensing analyses. One approach to detecting LSNe is to observe known SL systems and wait for SN. This approach leverages existing SNe detection strategies and infrastructure and is expected to detect all SNe in known

SL systems in the southern hemisphere. However, the Rubin Observatory’s LSST will probe deeper than any previous optical survey. It is expected that a large population ($\sim 10\%$) of all SL systems will have a source galaxy that is too faint to have been detected previously, and they will be missing from the list of target systems (Ryczanowski et al. 2020). Another approach leverages the standardizable brightness of SNe-Ia and proposes the search for brighter-than-expected (due to lensing magnification) SNe near elliptical galaxies (Goldstein & Nugent 2016). This magnitude threshold technique is expected to detect ~ 500 LSNe-Ia, but is specific to elliptical lens galaxies and will miss LSNe-Ia whose date of peak brightness does not align with the LSST cadence.

In this work, we introduce a deep neural network architecture designed to identify LSNe-Ia without the requirements of targeting known SL systems, elliptical galaxies, or observing the peak brightness of the SN. Deep-learning algorithms have been highly successful, both in terms of accuracy and speed, in the fields of image-based SL system detection (Jacobs et al. 2019, among several others) and light-curve-based SNe classification (Möller & deBoissière 2019, among several others). A convolutional neural network (CNN; LeCun et al. 1989) is a kind of deep algorithm that slides learnable matrix operators along images to emphasize or de-emphasize characteristic shapes, such as lensing arcs. Recurrent neural networks (RNNs; Elman 1990) can model sequences of data, such as light curves, to make classifications based on how the data vary in time. Within the RNN architecture, long short-term memory (LSTM) networks (Hochreiter & Schmidhuber 1997), which introduce cells with learnable pathways for information to be passed along, have demonstrated improved performance in sequence characterization. In our approach, we combined a CNN and an LSTM as two branches and optimally use the information from both branches—an architecture that leverages the data structures of each input and is tailored to the challenge of immediate LSNe-Ia identification. Other studies, such as Ramanah et al. (2021), have combined convolutional and recurrent layers in different deep-learning architectures to target LSNe identification, although our multibranch approach is unique in that it places the spatial and temporal information on the same footing from the start.

We present this work as follows: In Section 2 we describe the simulations used for training and testing ZipperNet, the data-processing procedure for using both spatial and time-series information, and the architecture of ZipperNet. In Section 3 we describe the results from applying ZipperNet to four simulated data sets that emulate modern survey data products. In Section 4 we discuss the performance of ZipperNet with respect to the data set properties and the network architecture. We conclude in Section 5.

2. Methods

2.1. Data Simulation

We simulated images of astronomical strong-lensing systems with the open-source software package `deeplens` (Morgan et al. 2021b), which is built around `lenstronomy` (Birrer & Amara 2018; Birrer et al. 2021b), a widely used package that performs gravitational lensing calculations, modeling, and simulations in a variety of contexts. `deeplens` provides additional features that are important for the era of large-scale surveys and deep-learning studies

of strong lenses, e.g., image and SN injection, probability distribution sampling, and realistic observing conditions.

2.1.1. Survey Emulation

We simulated four data sets, distinguished by their camera specifications, observing conditions, and cadence, each emulating a distinct modern or next-generation cosmic survey: one wide- and one deep-field data set for DES, called DES-wide and DES-deep, respectively; a wide-field LSST data set (LSST-wide), and a three-day cadence Dark Energy Camera (DECam; Flaugher et al. 2015) data set similar to the Dark Energy Spectroscopic Instrument DECam Observation of Transients (DESI-DOT) program (Palmese & Wang; DECam Proposal 2021A-0148). All data sets use the g , r , i , and z optical filters and include 45-by-45-pixel images. The DES-wide, DES-deep, and DESI-DOT data sets simulate images from the DECam, which connects the pixel size, gain, and read noise across those data sets. The LSST-wide data set simulates LSSTCam (Stalder et al. 2020) images, which have similar but slightly adjusted values for the camera properties.

The DES-wide and DESI-DOT data sets both use the real observing conditions (seeing and sky brightness) from the DES wide-field survey (Abbott et al. 2018). For DES-wide and DESI-DOT, the exposure times are 90 and 60 s, respectively. The DES-deep data set has different seeing, sky brightness, zeropoint, and exposure times chosen for the DES SN program (Abbott et al. 2019). In general, these exposure times are on the order of 200 s, but the acceptable seeing criteria can be worse than for the DES wide-field survey. The LSST-wide observing conditions are estimated from simulations of the first year of the survey and use 30 s exposures (Marshall et al. 2017).

Each data set has a specific and distinct cadence, and the density of observations significantly affects the analysis in this work. The LSST-wide, DESI-DOT, and DES-deep data sets contain a baseline of 14 epochs per band. While the LSST main survey cadence is still being designed at the time of this writing, our fiducial 14 epoch data sequences are sufficiently short that they are obtainable from both the “baseline” and “rolling” cadences that are under consideration for the survey. The DES-wide data set contains seven exposures in each band spread over 5.5 yr to match the real survey. We sampled the observation times of several fields from the DES footprint to generate the `deeplens` simulations. The LSST-wide cadence is estimated using several realizations of an intraband spacing of 12 ± 5 days over a three-month period (Marshall et al. 2017). The DES-deep cadence is estimated using several realizations of an intraband spacing of 6 ± 1 days over a one-month period (Abbott et al. 2019). Last, the DESI-DOT cadence is an exposure in each band every three nights over a one-month period.

We seek to avoid jargon confusion between astronomy and machine learning contexts with respect to the term “epoch.” In this work, “epoch” refers to one astronomical exposure or data-collection period. When discussing neural network training steps, we use the term “training iteration” instead of the traditional “epoch” that is used in machine learning. The data sets are summarized in Table 1. All `deeplens` input files for this analysis are accessible for reproduction of the data sets in Morgan et al. (2021a).

Table 1

Summary of the Instruments and Observational Procedures Emulated in the Simulated Data Sets: Camera Properties, Observing Conditions, and Survey Cadence

Data set	Gain (e ⁻ /count)	Read Noise (e ⁻)	Pixel Size (arcsec)	Exp. Time (seconds)	Seeing (FWHM arcsec)	Cadence (days)	Reference
DES-wide	6.083	7.0	0.263	90	0.91 ± 0.12	220 ± 191	Abbott et al. (2018)
LSST-wide	2.3	10.0	0.2	30	0.71	12 ± 5	Marshall et al. (2017)
DES-deep	6.083	7.0	0.263	200	1.04 ± 0.44	6 ± 1	Abbott et al. (2019)
DESI-DOT	6.083	7.0	0.263	60	0.90 ± 0.12	3	Abbott et al. (2018)

Note. We show only the *i* band for each property because it illustrates the key discerning features between the data sets. The cadence displays the mean and standard deviation of the intraband time separation. The full cadence information and data quality properties are available in the `deeplenstronomy` input files that accompany this work (Morgan et al. 2021a).

2.1.2. Object, System, and Population Simulation

In total, we simulate 17 different types of astronomical systems to reflect the diversity of systems that classifiers are likely to encounter when applied to observed optical survey data: (1) one galaxy behind one foreground galaxy; (2) two galaxies at the same redshift and with small angular separation (1" to 4"); (3) one galaxy (with one SN Ia) behind one foreground galaxy; (4) one galaxy (with one SN-CC) behind one foreground galaxy; (5) one galaxy; (6) two galaxies (with small angular separation and at the same redshift) in front of one background galaxy that is at a higher redshift; (7) one galaxy behind one foreground galaxy that has one star from the Milky Way in the image cutout; (8) two galaxies with small angular separation and at similar redshifts that have one star from the Milky Way in the image cutout; (9) one galaxy (with one SN Ia) behind one foreground galaxy that has one star from the Milky Way in the image cutout; (10) one galaxy (with one SN-CC) behind a foreground galaxy that has one star from the Milky Way in the image cutout; (11) one galaxy that has one star from the Milky Way in the image cutout; (12) two galaxies with small angular separation and at similar redshifts in front of one background galaxy at a higher redshift and one star from the Milky Way in the image cutout; (13) empty sky; (14) one Milky Way star; (15) two Milky Way stars; (16) one galaxy with one SN Ia; and (17) one galaxy with one SN-CC. Figure 1 shows sample images and time series of the 17 systems.

To further enhance realism, the properties of all simulated objects are drawn from real data: all inherent physical correlations of these parameters are included in our data set. First, a galaxy that enters the simulations as the lens has properties drawn from a population of ~ 2000 observed galaxies. The velocity dispersion and spectroscopic redshift were measured by the Sloan Digital Sky Survey (SDSS; York et al. 2000). We obtain a color-independent ellipticity, as well as a band-wise half-light radius, Sérsic profile index, and magnitude from DES Year 1 data (Abbott et al. 2018; Tarsitano et al. 2018). Next, a putative source galaxy, whether used in a lensing or nonlensing system, draws its properties from a population of $\sim 500,000$ galaxies measured by DES (Abbott et al. 2021). The band-wise magnitudes of foreground Milky Way stars were also drawn from DES data (Abbott et al. 2021). Finally, the SNe were injected using public SN spectral energy distributions (Kessler et al. 2010) available in `deeplenstronomy`, which redshifts the distribution and calculates the observed magnitude in each band. The injected SN reaches peak brightness anytime between 20 days before the first observation and 20 days after the final observation, so the data set contains falling light curves, rising light curves, and complete light curves. We do not include the effects of

microlensing in our simulated data set because it is expected to be small compared to the change in brightness observed from an SN (Foxley-Marrable et al. 2018b).

For all four survey emulation data sets, we simulate the same strong-lensing systems: all strong-lensing systems are emulated in all four of the cosmic survey contexts. While our simulated data sets are subject to selection biases from the detection limits of DES and SDSS for source and galaxies, respectively, the data nonetheless contain realistic collections of object properties, which enables the validation of our deep-learning detection method on realistic survey data.

2.2. Data Processing

`deeplenstronomy` emulates observational surveys by producing a time series of images. With the exception of small effects from observing conditions, images in a series will be approximately identical because astronomical objects are approximately stationary on month-long timescales. Even in the case of an SN, the primary difference is the presence of one or more point sources in some of the images in the time series. We condensed the image information to single-image input for ZipperNet by averaging all images in the time series on a pixel-by-pixel basis within each band. This processing reduces noise fluctuations from the observing conditions while preserving the presence of SNe, thus increasing the overall signal-to-noise ratio of the image and making faint objects more visible. After averaging the images, the pixel values of the mean images are scaled to range from 0 to 1 on a per-example basis to preserve color relations.

To concisely characterize the temporal behavior of a time series of images and to avoid relying on source identification or deblending algorithms, we follow a process that reflects a standardized background-subtracted aperture flux measurement in astronomy. We measure the signal (*S*) and background (*B*) to extract a background-subtracted brightness (*S*−*B*) of individual images within a predefined circular aperture at the center of each image. In equation form, this process can be expressed as

$$S = \sum_{i,j}^N W_{i,j} X_{i,j}, \quad (1)$$

$$B = \text{median}_{W_{i,j}=0} [X_{i,j}] \times \sum_{i,j}^N W_{i,j}, \quad (2)$$

$$\text{Brightness} = S - B, \quad (3)$$

where *i* and *j* index the row and column of the image pixels, *N* gives the number of pixels along one dimension of the images, *X* is an image, and *W* is the aperture. *W*_{*i,j*} is zero outside the

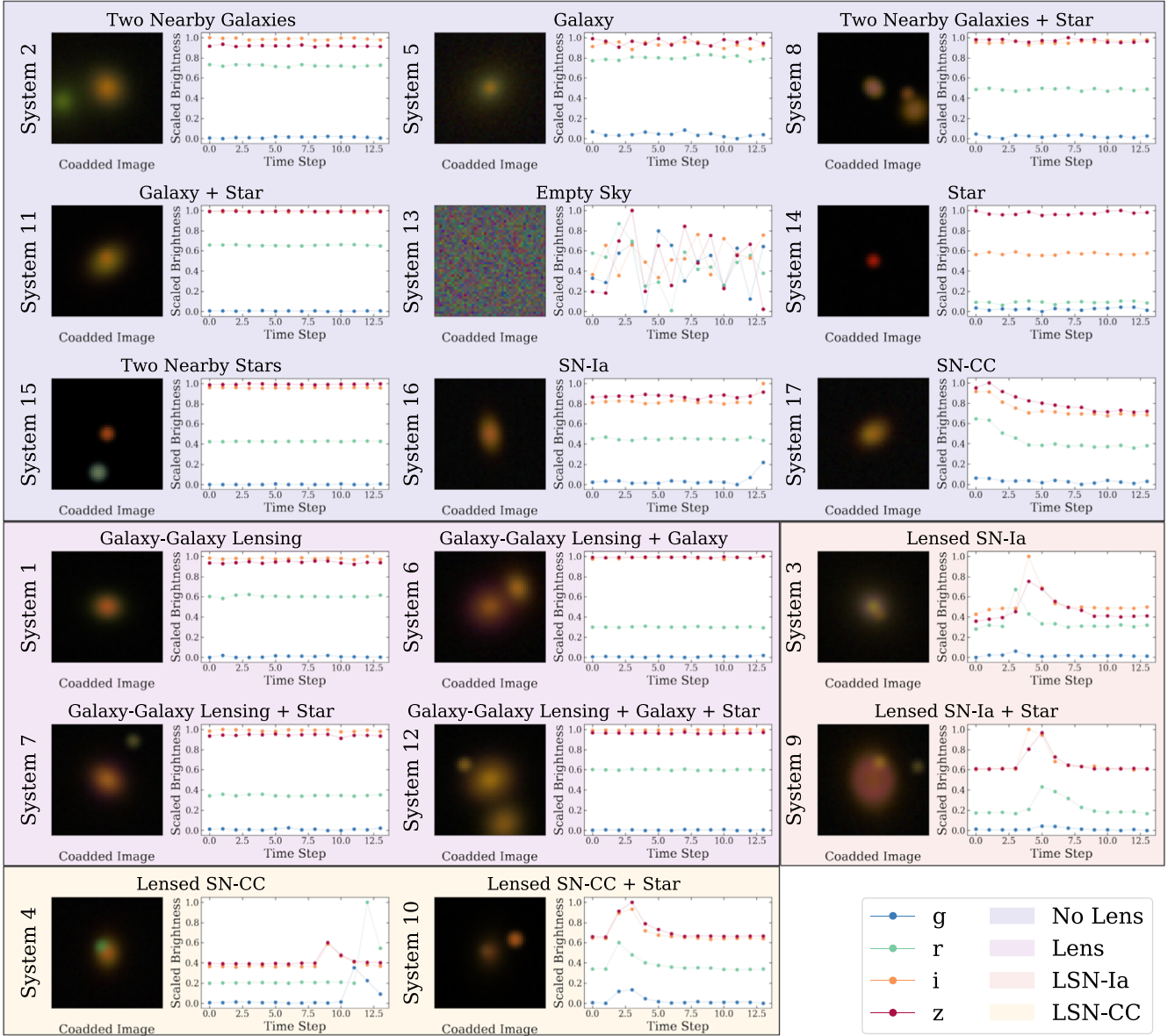


Figure 1. Examples of the 17 simulated systems from the LSST-wide data set grouped into No Lens (blue), Lens (magenta), LSN-Ia (orange), and LSN-CC (yellow) classes. Each example displays a coadded composite image of the gri bands and the extracted light curve from our processing in Section 2.2. Each time step in the LSST-wide data set is approximately 12 days.

aperture and one inside the aperture. In this work, the circular aperture has a 20-pixel radius, which corresponds to $5''.26$ for DECam and $4''.0$ for LSST-Cam—both much larger than typical galaxy-scale lens Einstein radii, which are approximately in the range $[0.5, 1.2]$ arcsec. For processing within the neural networks, we again scale the extracted brightness to between 0 and 1 on a per-example basis.

A byproduct of this process is the significant increase in noise in the photometry measurements; we find, however, that this effect does not hinder the deep-learning methods. After the averaging images and extracting light curves, each example input to ZipperNet is a 45-pixel-by-45-pixel image in each band and a light curve of the extracted brightness at each time step in each band. The operations required to extract the photometry of the systems have little computational cost and can easily be broadcasted, which is a great benefit considering the scale of modern astronomical data sets.

Last, we define a classification scheme for our 17 simulated systems. We construct a four-class problem, where the classes are “No Lens,” “Lens,” “LSNe-Ia,” and “LSNe-CC,” as shown in Figure 1. The No Lens class collects the cases where there is no gravitational lensing present, labeled 2, 5, 8, 11, 13, 14, 15, 16, and 17. The Lens class collects cases where there is gravitational lensing, but no SNe in the background galaxy, labeled 1, 6, 7, and 12. The LSNe-Ia and LSNe-CC classes collect cases with gravitational lensing and an SN in the background galaxy, labeled 3 and 9 and as 4 and 10. Each of the four classes contains 1250 examples, with equal representation of the individual constituent cases. To augment the data sets and increase the size of each class eightfold to 10,000, we rotated and mirrored the images; the light curve was unaffected due to the circular aperture extraction method. When we split the data sets into smaller training and testing data sets, none of the examples in the testing data set is rotated or mirrored

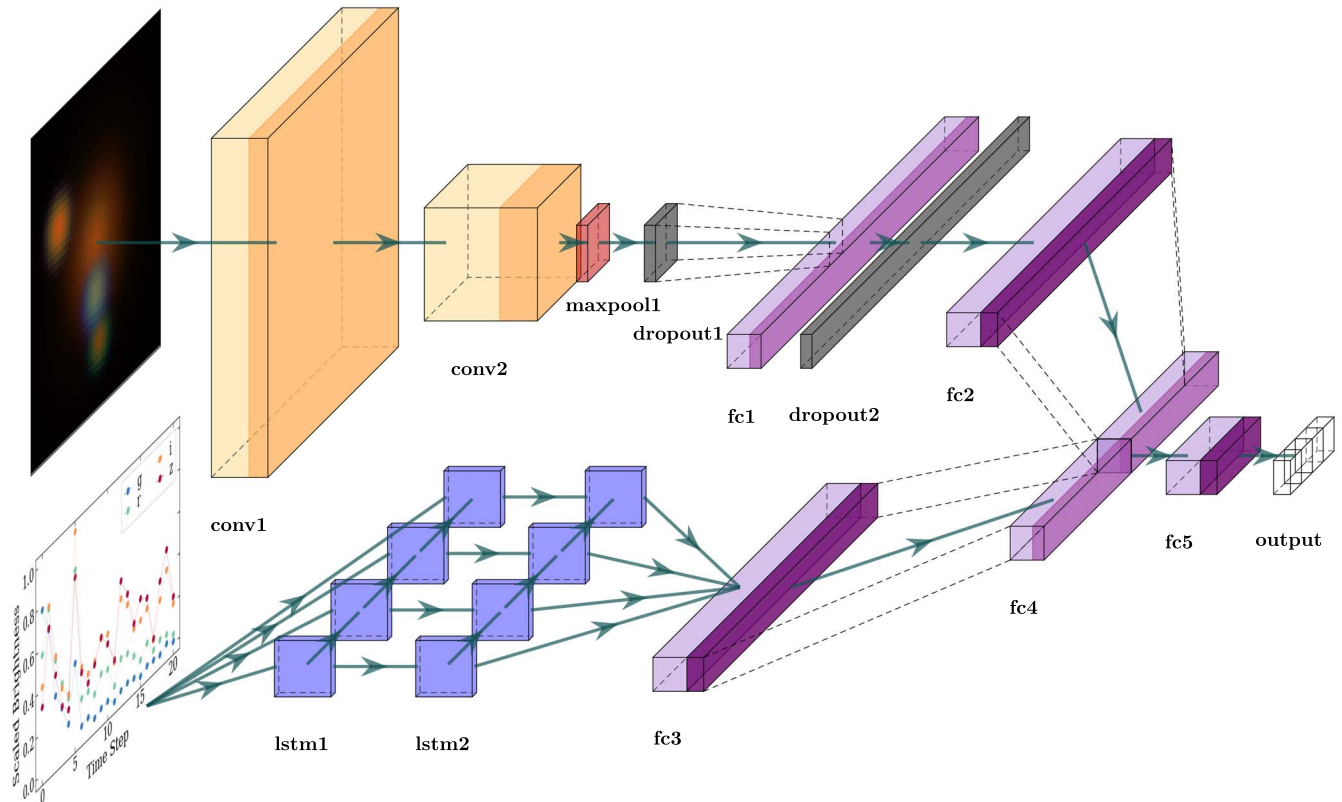


Figure 2. A diagram of the ZipperNet architecture. Two convolutional layers (orange) receive griz images, while two LSTM layers (blue) receive extracted griz light curves. Fully connected layers (purple) process the flattened and concatenated outputs. Nonlinear activation functions are indicated by a darkened band at the right edge of a layer. Table 2 displays the layer specifications. The final output is a length-4 array, with each element representing the score of one of the four classes in the data set. This visualization was made with the `PlotNeuralNet` library (Iqbal 2018).

versions of objects in the training data set: the two data sets are rotated and mirrored independently.

2.3. ZipperNet

Our ZipperNet architecture is designed to treat the image-based information and the light-curve-based information equally. On one branch, the images are passed through convolutional layers, flattened into one-dimensional arrays, and condensed in size. On another branch, the light curves are passed through recurrent layers composed of LSTM cells and flattened into one-dimensional arrays. The flattened images and light curves output by each branch are condensed to equal sizes, concatenated, and then mapped to four output features, one for each class of our problem. We then obtain a single classification by determining which of the four output features has the highest value. By zipping convolutional layers and recurrent layers into one coherent deep-learning architecture (in joining the branches), the training of the network will optimize weights in both types of layers simultaneously. The architecture is illustrated in Figure 2, and the specifications of each layer are presented in Table 2. All deep-learning code in this analysis uses the `PyTorch` (Paszke et al. 2019) library.

For each of the four data sets in our analysis, we trained an individual ZipperNet: we trained on 90% (9000 samples) of the simulated data and used the remaining 10% (1000 samples) for testing. We did not use any of our simulated data as a validation data set for hyperparameter optimization. Rather, we chose the ZipperNet hyperparameter settings based on an independent toy data set composed of images of different shapes (squares versus circles) with different time-varying

Table 2
ZipperNet Layer Specifications. We Use the Following Shorthand: Kernel Size (k), Padding (p), Stride (s), Dropout Fraction (f), and Hidden Units (h)

Layer	Specifications
conv1^a	Conv2D—(k : 15, p : 2, s : 3)—(4 → 48)
conv2^a	Conv2D—(k : 5, p : 2, s : 1)—(48 → 96)
maxpool1	MaxPool2D (k : 2)
dropout1	Dropout2D (f : 0.25)
fc1^a	Fully Connected (3456 → 408)
dropout2	Dropout (f : 0.5)
fc2^b	Fully Connected (408 → 25)
lstm1	LSTM (h : 128)
lstm2	LSTM (h : 128)
fc3^b	Fully Connected (128 → 25)
fc4^a	Fully Connected (50 → 8)
fc5^b	Fully Connected (8 → 4)

Notes. Arrows indicate the change in the size of the data representation as it is passed through the layer.

^a indicates a Rectified Linear Unit (ReLU) activation function.

^b indicates a LogSoftmax activation function. In total, our model contains 1,551,632 trainable parameters.

properties (parabolic versus linear change in total brightness) and fixed the settings for each of the four ZipperNet instances. This choice is motivated by a desire to keep the model constant and prevent the hyperparameter settings from favoring one of the simulated data sets over another. We therefore control confounding variables in our experiment such that we can

Table 3

Training and Testing Accuracy for ZipperNet on Each of the Four Data Sets (Rows) at Training Iterations 10 and 40 (Columns)

Data set	Training Acc.		Testing Acc.	
	Iter. 10	Iter. 40	Iter. 10	Iter. 40
DES-wide	0.496	0.503	0.473	0.487
LSST-wide	0.750	0.838	0.709	0.785
DES-deep	0.550	0.613	0.535	0.573
DESI-DOT	0.737	0.813	0.680	0.735

connect differences in model performance between the four simulated data sets to data set properties.

We chose a batch size of five for the training because LSTM cells generally perform better when processing smaller amounts of information at the same time. We also used categorical cross entropy as the loss function of the network and a learning rate of 0.001 with the Adam (Kingma & Ba 2017) optimizer. As the network trained, we monitored the accuracy (the number of correct classifications divided by the total number of samples) for the training and testing data sets. In each case, the training and testing accuracy plateaued after ~ 10 training iterations, but we allowed the training to continue for 40 training iterations. The fully trained network is chosen as the point during training with the highest testing accuracy. We use the term “training iteration” in place of the traditional “epoch” to avoid confusion with the astronomy term “epoch” that is used in other parts of this analysis. The accuracy for the training and testing sets for each of the four data sets is presented in Table 3.

3. Results

In general, we find that the ZipperNet model is capable of identifying SL systems, identifying the presence of an SN within SL systems, and distinguishing between LSN-Ia LSN-CC. Furthermore, we demonstrate that ZipperNet can perform these classifications on simulated data sets across wide ranges of depth, observing conditions, and cadence. We use a four-class confusion matrix to compare the predicted and true labels resulting from ZipperNet’s execution (see Figure 3). In all the matrices, the strong representation along the main diagonal indicates correct classifications. ZipperNet had the lowest performance on the DES-wide data, and we discuss this result in Section 4.

There are multiple primary sources of confusion. Confusion between the No Lens and Lens classes is likely due to pixel-based features being difficult to distinguish from the images. This confusion is strongest within the DES-wide data set, so we attribute this behavior to the optical depth of the images because the DES-wide data set is the shallowest data set we simulated and faint source galaxies would become more difficult to identify. In the deeper data sets, this confusion is caused by exposures with high seeing or systems with small Einstein radii, where in both cases objects are blurred. The confusion between the LSNe-Ia and LSNe-CC classes is likely due to the difference in cadence and seeing in the surveys. DES-deep, LSST-wide, and DESI-DOT are much higher-cadence data sets than DES-wide (see Table 1), indicating that more densely sampled SN light curves are easier for the LSTM cells to classify. A comparison of the DES-deep and DES-wide confusion matrices indicates that there are situations where cadence can be more important than seeing (specifically noting

the LSNe-Ia versus LSNe-CC confusion) because the DES-wide data set had better seeing than DES-deep, but that these situations require dramatic differences in cadence density. Furthermore, the LSST-wide and DESI-DOT data sets have much better seeing than the DES-deep data set, which shows the importance of being able to resolve spatial features when making classifications. We discuss the importance of the cadence and seeing in more detail in Section 4.

In practice, a general LSN identifier is itself a useful tool: LSNe-CC can be used for time-delay cosmography measurements, although they offer less precision on the final H_0 measurement. If we reframe this classification scheme from a four-class problem to a two-class problem (No Lens and Lens in one class and LSNe-Ia and LSNe-CC in the second class), the performance is boosted. We can obtain two-class problem classifications from our four-class network outputs by selecting the class predicted by the network and sorting it into LSN or not LSN. Figure 4 shows a receiver operating characteristic (ROC) curve for the DES-wide, LSST-wide, DES-deep, and DESI-DOT data sets. ROC curves are standard tools for assessing the predictive power of a classifier by calculating the false-positive rate and true-positive rate at all possible probabilities output by the classifier. A perfect classifier will have an area under curve (AUC) of 1.0, while a classifier that guesses randomly will have an AUC of 0.5. ZipperNet shows high performance when classifying LSNe versus everything else, and this high performance extends across the LSST-wide, DES-deep, and DESI-DOT data sets. This result can also be interpreted from the confusion matrix (Figure 3), where there is little confusion between the LSNe classes and non-LSNe classes. We comment on this high performance in the context of ZipperNet’s architecture in Section 4.

We also estimate the baseline true-positive rate and false-positive rate for LSNe using this technique in the different data sets. When applying the ZipperNet technique to real data, we would expect real data to be used in the training and validation, which would produce more accurate estimates. That being said, we can initially report an LSN true- (false-) positive rate of 90.2 % (29.6 %) for DES-wide, 87.0 % (21.9 %) for DES-deep, 91.5 % (9.8 %) for LSST-wide, and 89.0 % (12.6 %) for DESI-DOT. The true-positive rate of approximately 91.5 % for LSST-wide is the higher than the estimated recovery rates of LSNe for the non-deep-learning approaches mentioned in Section 1. Furthermore, with relatively low false-positive rates, we do not expect the data stream of the LSST to be overwhelmed by other astronomical systems being incorrectly labeled as LSNe.

In the two-class problem, we performed an additional analysis to interpret the features ZipperNet identified for making classifications. Figure 5 displays examples from the LSST-wide data set arranged into groups of correctly classified LSNe (true positives), other astronomical systems correctly classified as not LSNe (true negatives), other astronomical systems erroneously labeled as LSNe (false positives), and LSNe that were missed (false negatives). These examples were found to be representative of the general relation between the properties of objects and the predictions made by ZipperNet. ZipperNet is able to correctly classify LSNe when the Einstein radius is small ($0''.5-1''.0$) and with foreground stars in the image, both of which would trouble a standalone CNN. Upon close inspection, the true-positive images display galaxies with nonuniform light

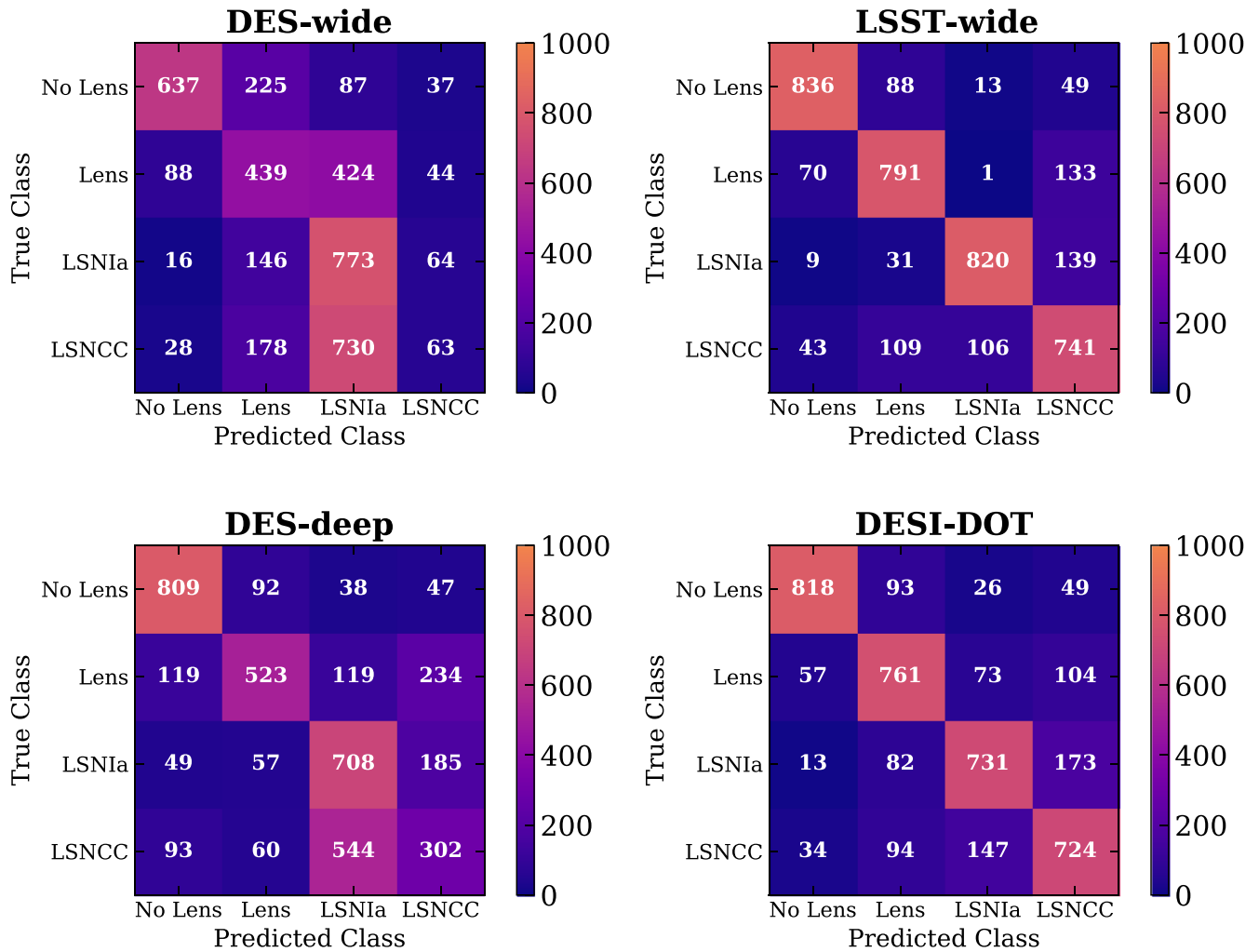


Figure 3. Confusion matrices from applying a trained ZipperNet to the DES-wide, LSST-wide, DES-deep, and DESI-DOT data sets. The matrices display the predictions of ZipperNet as a function of the true classes of the examples for the test data sets. Each row in each matrix has 1000 examples in total, so the matrix counts are representative of normalized percentages.

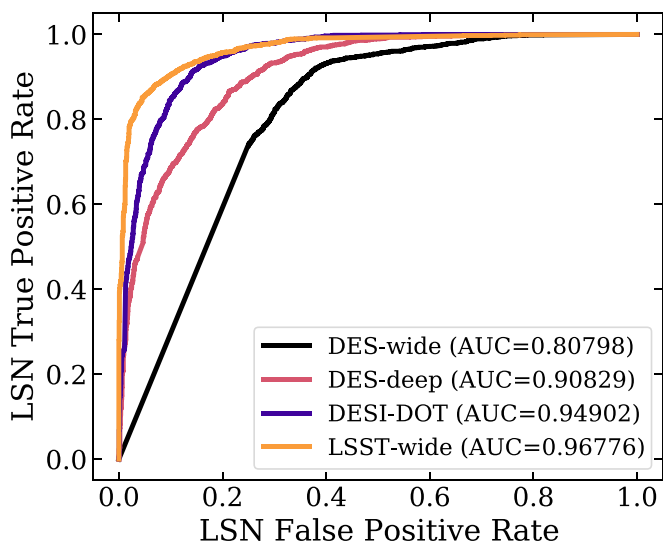


Figure 4. Receiver operating characteristic (ROC) curves for the LSST-wide, DES-deep, and DESI-DOT data sets. The ROC curves are calculated for the two-class problem of LSNIa and LSNCC vs. No Lens and Lens. An area under curve (AUC) of 1.0 indicates perfect performance, while an AUC of 0.5 indicates random guessing.

profiles, hinting at the presence of lensing, but the dominating feature is clearly the large fluctuation in brightness detected by the LSTM. The true negatives show systems with evidence of lensing as well, but this time, there is no temporal behavior to indicate the presence of an SN. The false positives contain images of lensing or crowded fields with small, but non-negligible coherent time-varying behavior; they in general contain both spatial and temporal features similar to the LSNe class. Last, the false negatives are the most important group to understand due to the rareness of LSNe. In some cases, the presence of a star in the aperture used to extract the scaled brightnesses can be bright enough to obscure the change in brightness of the LSNe. Similarly, if the source galaxy is distant and the alignment of the lensing system does not produce sufficient magnification, the imaging may not be deep enough to see a lensed source or a background SNe. Both these cases of false negatives demonstrate difficult to detect systems with LSNe and point toward a physically motivated selection as opposed to inaccurate feature representations learned by ZipperNet. In general, we find that ZipperNet learns features we would a priori expect.

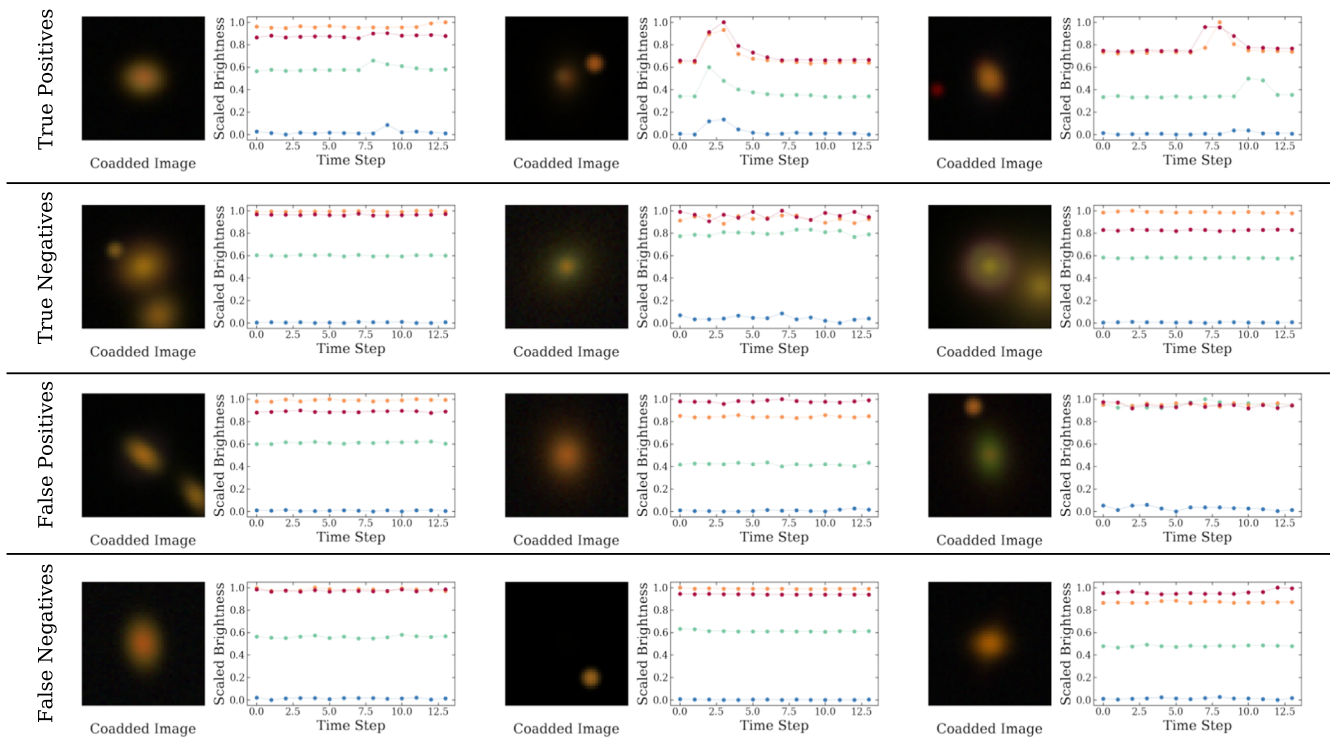


Figure 5. Examples of classifications made by ZipperNet from the LSST-wide data set, grouped by true positives, true negatives, false positives, and false negatives. The color scheme for the bands used in the light curves is the same as Figure 1.

4. Discussion

In general, ZipperNet can identify SL systems, identify LSNe within those systems, and classify the LSNe as LSNe-Ia or LSNe-CC. The variance in the performance across different data sets indicates a correspondence between data quality and LSN identification power. The DES-deep and DESI-DOT data sets have slightly higher cadences (more samples within a time series) than the LSST-wide data set, but the ZipperNet accuracy was considerably higher for the LSST-wide data set, which has higher seeing (lower image quality). The DES-deep data set emulates that of the DES SNIa observing program, in which exposures were collected on nights with slightly poorer (higher) seeing to optimize the seeing of the exposures used for weak-lensing measurements. In DES data processing and SNIa analysis, the difference-imaging (Kessler et al. 2015) and scene-modeling (Brout et al. 2019) techniques can nevertheless detect and measure SNe when seeing is up to $2''$. We bypassed these time-consuming techniques with our circular aperture extraction method for the light-curve brightnesses. Without these techniques, the performance of ZipperNet is degraded because the higher seeing in the DES-deep data set obscures image patterns that would otherwise be detectable in LSST-wide and DESI-DOT data sets.

A second data quality factor in ZipperNet’s predictive power is the cadence of the observations. The DES-wide data set has excellent depth and seeing, but a low-density time sampling: ZipperNet fails to perform at similar levels to the other data sets. With DESI-DOT, which has high cadence and similar depth and seeing as DES-wide, ZipperNet was able to learn the underlying features of the four classes extremely well. Overall, we find that together, a seeing of $\lesssim 1''.2$ (corresponding to typical upper limits on Einstein radii of galaxy-scale lenses) and a cadence with intraband spacing of $\lesssim 15$ nights

(corresponding to roughly the timescale for SNe evolution) can also improve performance.

Nevertheless, even when ZipperNet had confusion between LSN-Ia and LSN-CC (primarily caused by deficient cadence of the data set) or confusion between the No Lens and Lens classes (primarily caused by seeing greater than typical Einstein radii), ZipperNet performs extremely well as an LSNe finder. Reducing the classification to two classes—LSN and everything else—shows high performance for the LSST-wide, DES-deep, and DESI-DOT data sets (Figure 4). The ZipperNet architecture as an LSNe finder in this setting does not suffer from the dependence on seeing observed in the four-class problem and is slightly less dependent on the cadence. By balancing the image and temporal inputs, ZipperNet finds weightings and combinations of the two data products optimal for LSNe detection. We interpret this result as a demonstration of a key strength of the ZipperNet architecture.

Finally, we compare the ZipperNet architecture to a standalone “RNN,” a standalone “CNN,” and a combination of those two standalone networks (“COMBO”) in the context of the two-class problem and the LSST-wide data set. The standalone networks are identical in structure to the corresponding constituents of ZipperNet and trained under the same conditions. The combination classifier does not connect the feature representations of the standalone networks internally: it merely uses the outputs from each of the standalone classifiers, requiring them to individually identify an object as an LSN. Therefore, the COMBO classifier reflects a simplistic approach to LSN identification in astronomical surveys—first a CNN is used to identify an SL system followed by a transient detection algorithm used to search for LSNe. We expect that the standalone CNN will identify systems with lensing, while the standalone RNN will identify systems with SNe: by requiring both a lensing classification and an SN classification, we

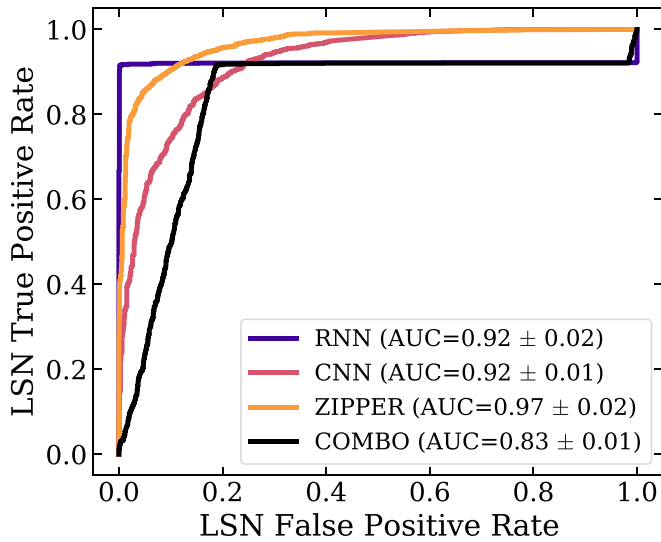


Figure 6. Receiver operating characteristic (ROC) curves for the LSST-wide data sets using different networks. The COMBO label is an optimized weighting of the CNN and RNN network predictions. The ROC curves are calculated for the two-class problem of LSNIa and LSNe vs. No Lens and Lens. An area under curve (AUC) of 1.0 indicates perfect performance, while an AUC of 0.5 indicates random guessing.

establish a baseline for the performance of deep-learning architectures that are not connected internally in ways similar to ZipperNet’s design. ROC curves for the different classifiers are shown in Figure 6. The CNN performs well and mostly identifies all systems with lensing or point sources in galaxies, but this performance can still result in high numbers of false positives (due to the rarity of LSNe) in practical applications. The RNN is effectively an SN identifier in this context: the vast majority of the false positives come from classifying unlensed SNe (labels 16 and 17) as LSNe. While the ROC AUC is high for the RNN, it alone could not be used as an LSN finder due to the much higher volumetric rate of SNe compared to LSNe. As shown by this test, ZipperNet outperforms both constituent networks (CNN and RNN) and more importantly, the simplistic combination of its constituent networks’ outputs (COMBO), indicating that connecting the feature representations of the RNN and CNN internally yields better overall performance and is a worthwhile deep-learning strategy for the problem of LSNe detection. The ZipperNet architecture shows the highest performance.

The motivating challenge for ZipperNet is to facilitate the identification of LSNe during the Rubin Observatory’s LSST, which is expected to have a very high data stream rate compared to previous large-scale surveys. The high data stream rate is not a challenge for ZipperNet because classifications can be parallelized and preprocessing is economical (because we directly extract band-wise light curves from images without the need for deblending or expensive photometric analysis). The role of a tool like ZipperNet in this setting would be to process the data stream and report a list of candidates ordered by a probability of being an LSN. At present, the binary classification produced by ZipperNet with confidence quantified by the measured true- and false-positive rates is our focus. However, a small amount of additional calibration could straightforwardly map the ZipperNet outputs to physical probabilities to facilitate the generation of candidate lists.

The remaining test of ZipperNet for Rubin Observatory main survey operations is detection at various epochs into the

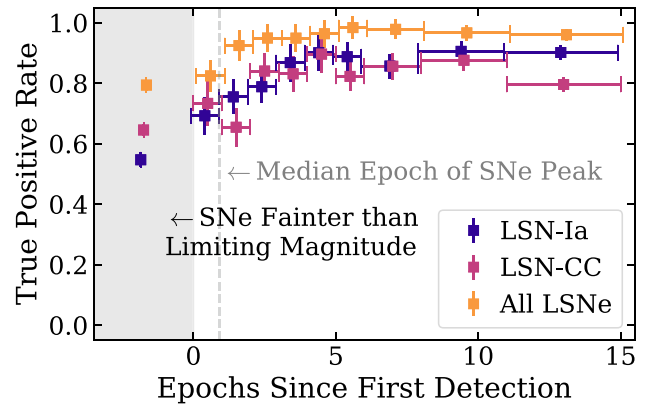


Figure 7. The true-positive rates for LSNe-Ia and LSNe-CC as functions of light-curve phase for the LSST-wide data set. The 5σ limiting magnitude used for the first detection is 24.4 mag in the r band based on the single-epoch survey requirements (The LSST Dark Energy Science Collaboration et al. 2018). The “All LSNe” label refers to the two-class problem of general LSNe detection.

SN light curves for community alerts: How early into the light curve can ZipperNet find an LSN? We present in Figure 7 the LSN true-positive rates as functions of the number of light-curve epochs after the first detection. In this context, an r -band magnitude brighter than 24.4 mag serves as the first detection, which is motivated by the LSST science requirements (The LSST Dark Energy Science Collaboration et al. 2018). We find that even when only one or two epochs are present in the light curve after the first detection, the true-positive rate is $>75\%$, and it improves as more epochs are added. After five post-detection observing epochs in the light curve, the LSN identification has a true-positive rate $>95\%$. Furthermore, even LSNe fainter than the 5σ limiting magnitude are identifiable with a true-positive rate of $\sim 80\%$, indicating high performance where other detection methods would lose sensitivity: both proposed methods discussed in Section 1 rely on SNe being brighter than the detection threshold and realized as individual objects. For these faint LSNe, ZipperNet likely finds success by identifying systems with image-based strong-lensing features and then noticing small changes in brightness. This result supports the expectation that ZipperNet will perform well as a real-time LSN identifier in Rubin Observatory main survey data.

In summary, ZipperNet introduces a new deep-learning architecture for lensed transient detection where spatial and temporal features are treated on the same footing in a single framework. This balanced framework produces an ROC curve AUC of 0.97 when identifying LSNe and 79% accuracy at outright identification of LSNe-Ia in LSST wide-field data even in the early phases of the light curve. Therefore, we expect ZipperNet to play a large role in the rapid identification of LSNe for spectroscopic characterization and time-delay cosmography during the main survey operations of the Rubin Observatory.

5. Conclusion

Detecting LSNe soon after explosion will be an important goal for the Vera C. Rubin Observatory and other high-cadence optical surveys. In this work, we introduced ZipperNet, a deep-learning tool for LSNe identification. ZipperNet combines a convolutional neural network with a recurrent neural network

to simultaneously process spatial and temporal data. We used *deeplens* to simulate four distinct optical survey data sets for testing. ZipperNet performed well when the cadence and seeing were LSST-like or better. Specifically, ZipperNet was able to identify LSNe in LSST-like data with an ROC AUC of 0.97 and distinguish LSNe-Ia from LSNe-CC in LSST-like data with 79% accuracy. With ZipperNet, high-cadence optical surveys can accurately identify both LSNe-CC and LSNe-Ia early into their light curves, and furthermore distinguish between the two transient classes. Thus, we expect ZipperNet to be a powerful tool in identifying lensing systems for time-delay cosmography measurements during the Rubin Observatory main survey operations.

According to the CRediT system, we acknowledge the contributions from authors in detail. R. Morgan designed the analysis and deep-learning methodologies, prepared simulations, trained networks, and prepared this publication. B. Nord and K. Bechtol offered feedback and guidance on the analysis and publication throughout the project. S. J. Gonzalez prepared merged SDSS/DES catalogs essential for realistic data set simulation. L. Buckley-Geer, A. Möller, J. W. Park, and A. G. Kim served as internal reviewers within DES. S. Birrer provided useful comments on the draft and analysis. All other authors contributed to DES infrastructure upon which this project was based.

R. Morgan thanks the Universities Research Association Fermilab Visiting Scholars Program for funding his work on this project. R. Morgan also thanks the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining grant #1829740, the Brinson Foundation, and the Moore Foundation; his participation in the program has benefited this work.

We acknowledge the Deep Skies Lab as a community of multidomain experts and collaborators who have facilitated an environment of open discussion, idea generation, and collaboration. This community was important for the development of this project.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under grant No. 1744555. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory at NSF's NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.


The DES data management system is supported by the National Science Foundation under grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) e-Universe (CNPq grant 465376/2014-2).





































This paper has gone through internal review by the DES collaboration.

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

Software: *astropy* (Astropy Collaboration et al. 2013), *deeplens* (Morgan et al. 2021b), *lens* (Birrer & Amara 2018), *matplotlib* (Hunter 2007), *numpy* (Harris et al. 2020), *pandas* (The Pandas Development Team 2020), *PlotNeuralNet* (Iqbal 2018), *PyTorch* (Paszke et al. 2019), *Scikit-Learn* (Pedregosa et al. 2011), *scipy* (Jones et al. 2001).

ORCID iDs

R. Morgan  <https://orcid.org/0000-0002-7016-5471>
 B. Nord  <https://orcid.org/0000-0001-6706-8972>
 K. Bechtol  <https://orcid.org/0000-0001-8156-0429>
 S. J. González  <https://orcid.org/0000-0001-7282-3864>
 E. Buckley-Geer  <https://orcid.org/0000-0002-3304-0733>
 A. Möller  <https://orcid.org/0000-0001-8211-8608>
 J. W. Park  <https://orcid.org/0000-0002-0692-1092>
 S. Birrer  <https://orcid.org/0000-0003-3195-5507>

M. Agüena  <https://orcid.org/0000-0001-5679-6747>
 J. Annis  <https://orcid.org/0000-0002-0609-3987>
 S. Bocquet  <https://orcid.org/0000-0002-4900-805X>
 D. Brooks  <https://orcid.org/0000-0002-8458-5047>
 A. Carnero Rosell  <https://orcid.org/0000-0003-3044-5150>
 M. Carrasco Kind  <https://orcid.org/0000-0002-4802-3194>
 J. Carretero  <https://orcid.org/0000-0002-3130-0204>
 R. Cawthon  <https://orcid.org/0000-0003-2965-6786>
 L. N. da Costa  <https://orcid.org/0000-0002-7731-277X>
 T. M. Davis  <https://orcid.org/0000-0002-4213-8783>
 J. De Vicente  <https://orcid.org/0000-0001-8318-6813>
 D. Friedel  <https://orcid.org/0000-0002-3632-7668>
 J. Frieman  <https://orcid.org/0000-0003-4079-3263>
 J. García-Bellido  <https://orcid.org/0000-0002-9370-8360>
 E. Gaztanaga  <https://orcid.org/0000-0001-9632-0815>
 G. Giannini  <https://orcid.org/0000-0002-3730-1750>
 D. Gruen  <https://orcid.org/0000-0003-3270-7644>
 R. A. Gruendl  <https://orcid.org/0000-0002-4588-6517>
 G. Gutierrez  <https://orcid.org/0000-0003-0825-0517>
 D. L. Hollowood  <https://orcid.org/0000-0002-9369-4157>
 K. Honscheid  <https://orcid.org/0000-0002-6550-2023>
 D. J. James  <https://orcid.org/0000-0001-5160-4486>
 K. Kuehn  <https://orcid.org/0000-0003-0120-0808>
 N. Kuropatkin  <https://orcid.org/0000-0003-2511-0946>
 M. A. G. Maia  <https://orcid.org/0000-0001-9856-9307>
 R. Miquel  <https://orcid.org/0000-0002-6610-4836>
 A. Palmese  <https://orcid.org/0000-0002-6011-0530>
 F. Paz-Chinchón  <https://orcid.org/0000-0003-1339-2683>
 A. Pieres  <https://orcid.org/0000-0001-9186-6042>
 A. A. Plazas Malagón  <https://orcid.org/0000-0002-2598-0514>
 A. Roodman  <https://orcid.org/0000-0001-5326-3486>
 E. Sanchez  <https://orcid.org/0000-0002-9646-8198>
 M. Smith  <https://orcid.org/0000-0002-3321-1432>
 E. Suchyta  <https://orcid.org/0000-0002-7047-9358>
 G. Tarle  <https://orcid.org/0000-0003-1704-0781>
 C. To  <https://orcid.org/0000-0001-7836-2261>

References

- Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, *ApJS*, 239, 18
 Abbott, T. M. C., Allam, S., Andersen, P., et al. 2019, *ApJL*, 872, L30
 Abbott, T. M. C., Adamow, M., Agüena, M., et al. 2021, *ApJS*, 255, 20
 Amanullah, R., Goobar, A., Clément, B., et al. 2011, *ApJL*, 742, L7
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
 Birrer, S., & Amara, A. 2018, *PDU*, 22, 189
 Birrer, S., Dhawan, S., & Shajib, A. J. 2021a, *ApJ*, 924, 2
 Birrer, S., Shajib, A. J., Gilman, D., et al. 2021b, *JOSS*, 6, 3283
 Brout, D., Sako, M., Scolnic, D., et al. 2019, *ApJ*, 874, 106
 Diehl, H. T., Neilsen, E., Gruendl, R. A., et al. 2018, *Proc. SPIE*, 10704, 107040D
 Elman, J. L. 1990, *Cogn. Sci.*, 14, 179
 Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *AJ*, 150, 150
 Foxley-Marrable, M., Collett, T. E., Vernardos, G., Goldstein, D. A., & Bacon, D. 2018a, *MNRAS*, 478, 5081
 Foxley-Marrable, M., Collett, T. E., Vernardos, G., Goldstein, D. A., & Bacon, D. 2018b, *MNRAS*, 478, 5081
 Freedman, W. L. 2021, *ApJ*, 919, 16
 Goldstein, D. A., & Nugent, P. E. 2016, *ApJL*, 834, L5
 Goobar, A., Amanullah, R., Kulkarni, S. R., et al. 2017, *Sci*, 356, 291
 Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, 585, 357–62
 Helfand, D. J., Stone, R. P. S., Willman, B., et al. 2001, *AJ*, 121, 1872
 Hochreiter, S., & Schmidhuber, J. 1997, *Neural Comput.*, 9, 1735–80
 Hook, I. M., McMahon, R. G., Boyle, B. J., & Irwin, M. J. 1994, *MNRAS*, 268, 305
 Hunter, J. D. 2007, *CSE*, 9, 90
 Iqbal, H. 2018, HarisIqbal88/PlotNeuralNet v1.0.0, Zenodo, doi:10.5281/zenodo.2526396
 Ivezić, Z., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
 Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, *ApJS*, 243, 17
 Jones, E. D., Oliphant, T. E., & Peterson, P. 2001, *NatMe*, 17, 261
 Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015, *Sci*, 347, 1123
 Kessler, R., Bassett, B., Belov, P., et al. 2010, *PASP*, 122, 1415
 Kessler, R., Marriner, J., Childress, M., et al. 2015, *AJ*, 150, 172
 Kingma, D. P., & Ba, J. 2017, arXiv:1412.6980
 Kolatt, T. S., & Bartelmann, M. 1998, *MNRAS*, 296, 763
 LeCun, Y., Boser, B., Denker, J. S., et al. 1989, *Neural Comput.*, 1, 541
 Marshall, P., Clarkson, W., Shemmer, O., et al. 2017, LSST Science Collaborations Observing Strategy White Paper: “Science-driven Optimization of the LSST Observing Strategy” v1.0, Zenodo, doi:10.5281/zenodo.842713
 Mihalas, D. 1963, *PASP*, 75, 256
 Möller, A., & de?Boissière, T. 2019, *MNRAS*, 491, 4277
 Morgan, R., Nord, B., Bechtol, K., & Jonzález, J. 2021a, DeepZipper: Code and Simulation Inputs, Zenodo, doi:10.5281/zenodo.5425727
 Morgan, R., Nord, B., Birrer, S., Lin, J. Y.-Y., & Poh, J. 2021b, *JOSS*, 6, 2854
 Oguri, M. 2019, *RPPH*, 82, 126901
 Oguri, M., & Kawano, Y. 2003, *MNRAS*, 338, L25
 Paszke, A., Gross, S., Massa, F., et al. 2019, in Advances in Neural Information Processing Systems 32, ed. H. Wallach et al. (Red Hook, NY: Curran Associates, Inc), 8024, <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
 Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, 12, 2825, <http://jmlr.org/papers/v12/pedregosa1a.html>
 Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6
 Quimby, R. M., Oguri, M., More, A., et al. 2014, *Sci*, 344, 396
 Ramanah, D. K., Arendse, N., & Wojtak, R. 2021, arXiv:2107.12399
 Refsdal, S. 1964, *MNRAS*, 128, 307
 Riess, A. G., Casertano, S., Yuan, W., et al. 2021, *ApJL*, 908, L6
 Rodney, S. A., Brammer, G. B., Pierel, J. D. R., et al. 2021, *NatAs*, 5, 1118
 Rodney, S. A., Patel, B., Scolnic, D., et al. 2015, *ApJ*, 811, 70
 Ryczanowski, D., Smith, G. P., Bianconi, M., et al. 2020, *MNRAS*, 495, 1666
 Shajib, A. J., Birrer, S., Treu, T., et al. 2020, *MNRAS*, 494, 6072
 Stalder, B., Reil, K., Claver, C., et al. 2020, *Proc. SPIE*, 11447, 114470L
 Tarsitano, F., Hartley, W. G., Amara, A., et al. 2018, *MNRAS*, 481, 2018
 The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al. 2018, arXiv:1809.01669
 The Pandas Development Team, et al. 2020, pandas-dev/pandas: Pandas v1.4.1, Zenodo, doi:10.5281/zenodo.3509134
 Treu, T. 2010, *ARA&A*, 48, 87
 Treu, T., & Marshall, P. J. 2016, *A&ARv*, 24, 11
 Tripp, R. 1998, *A&A*, 331, 815
 Wojtak, R., Hjorth, J., & Gall, C. 2019, *MNRAS*, 487, 3342
 Wong, K. C., Suyu, S. H., Chen, G. C.-F., et al. 2019, *MNRAS*, 498, 1420
 York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, *AJ*, 120, 1579