The Multimodal Universe:
Enabling Large-Scale Machine Learning with 100 TB of Astronomical Scientific Data

The Multimodal Universe Collaboration

Eirini Angeloudi[1,2], Jeroen Audenaert[3], Micah Bowles[4,5],
Benjamin M. Boyd[6], David Chemaly[6], Brian Cherinka[7], Ioana Ciucă[8,9,10],
Miles Cranmer[6,5] Aaron Do[6], Matthew Grayling[6], Erin E. Hayes[6],
Tom Hehir[6,5] Shirley Ho[11,12,13,5], Marc Huertas-Company[1,2,9],
Kartheik G. Iyer[14,11,9], Maja Jablonska[10,9] Francois Lanusse[11,5,15],
Henry W. Leung[16], Kaisey Mandel[6], Juan Rafael Martínez-Galarza[17,18]
Peter Melchior[13], Lucas Meyer[11,5], Liam H. Parker[11,5,19], Helen Qu[20]
Jeff Shen[13], Michael J. Smith[21,9], Connor Stone[22,23,24], Mike Walmsley[16],
John F. Wu[7,25]

[1]Instituto de Astrofisica de Canarias [2]Universidad de La Laguna
[3]Massachusetts Institute of Technology [4]University of Oxford [5]Polymathic AI
[6]University of Cambridge [7]Space Telescope Science Institute
[8]Stanford University [9]UniverseTBD [10]Australian National University
[11]Flatiron Institute [12]New York University [13]Princeton University
[14]Columbia University [15]Université Paris-Saclay, Université Paris Cité, CEA,
CNRS, AIM [16]University of Toronto
[17]Center for Astrophysics, Harvard & Smithsonian [18]AstroAI
[19]University of California, Berkeley [20]University of Pennsylvania [21]Aspia Space
[22]Université de Montréal [23]Ciela Institute [24]Mila [25]Johns Hopkins University

Abstract

We present the MULTIMODAL UNIVERSE, a large-scale multimodal dataset of scientific astronomical data, compiled specifically to facilitate machine learning research. Overall, the MULTIMODAL UNIVERSE contains hundreds of millions of astronomical observations, constituting 100 TB of multi-channel and hyper-spectral images, spectra, multivariate time series, as well as a wide variety of associated scientific measurements and "metadata". In addition, we include a range of benchmark tasks representative of standard practices for machine learning methods in astrophysics. This massive dataset will enable the development of large multi-modal models specifically targeted towards scientific applications. All codes used to compile the MULTIMODAL UNIVERSE and a description of how to access the data is available at https://github.com/MultimodalUniverse/MultimodalUniverse

1 Introduction

Web-scale datasets containing billions of text and image examples have been instrumental in the development and scaling of large foundation models for language and vision [125, 60]. Surprisingly, unified datasets of comparable scale have not yet been developed for many scientific domains, hindering the progress of large, advanced machine learning (ML) models for
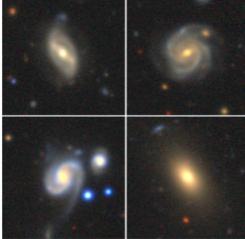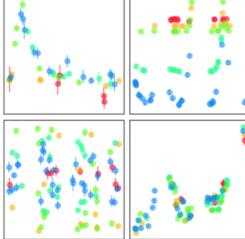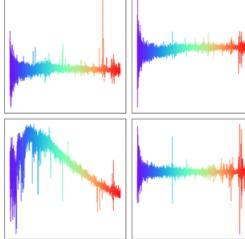
	Images	Time-Series	Spectra
# examples	140M	4.5M	225M
Description	images in a variety of wavelength ranges, including optical and infrared	multivariate time-series of flux + uncertainty in different wavelength ranges	flux as a function of wavelength
Tasks	galaxy classification, physical property estimation	time-series classification, redshift estimation	physical property estimation
Examples			

Figure 1: Illustration of the main modalities included in the MULTIMODAL UNIVERSE, along with typical associated machine learning tasks. In addition, the MULTIMODAL UNIVERSE also includes a small amount of hyperspectral images and tabular data, not shown here.

science [130]. Yet, the integration of large-scale, standardized, and comprehensive scientific datasets spanning multiple data modalities is essential for the development of ML models capable of fully leveraging the diverse spectrum of scientific data beyond traditional text and images.

To that end, we introduce the MULTIMODAL UNIVERSE (Figure 1), a large-scale, multimodal dataset of scientific data designed not only for ML research in astronomy but also to enable the broader development of scientific foundation models. The MULTIMODAL UNIVERSE capitalizes on the wide variety of astronomical observations available from a diversity of ground- and space-based telescopes to aggregate a dataset of hundreds of millions of astrophysical objects and phenomena. Altogether, this constitutes over 100 TB of open-access and copyright-free data spanning multiple observational modalities, including multi-channel and hyperspectral images, optical spectra, multivariate time-series, and an extensive array of associated scientific measurements. This diverse dataset provides an unprecedented resource for the development of sophisticated ML models for astrophysics and scientific data in general.

In keeping with its scientific goals, the MULTIMODAL UNIVERSE also emphasizes the importance of relevant scientific "metadata". This includes relevant contextual information for each observation (e.g. observational noise, pixel scale, instrumental response, etc.) that not only helps preserve the dataset's utility for scientific ML but also enables the development of models that integrate information beyond the raw data level. In addition to the dataset, we also describe a range of benchmark tasks and baseline deep learning models that reflect current best practices in astrophysics. These benchmarks provide a foundation for evaluating new models, enabling researchers to compare their approaches against established standards and push the boundaries of what is possible with ML in astronomy.

Finally, extensibility and accessibility are core to the MULTIMODAL UNIVERSE project. All data subsets are accessible as standardized Hugging Face datasets. They can be downloaded locally in full or accessed as streaming datasets directly from the Hugging Face Hub, see section 6. The code used to compile each component of the MULTIMODAL UNIVERSE from its respective official data source is hosted on the project's GitHub repository for transparency and reproducibility, and utilities to generate cross-matched (and multi-modal) versions of the dataset are also included. The architecture of the MULTIMODAL UNIVERSE source code is consistent, maintainable, and extensible. This is in contrast to the common practice in the field, which is for independent researchers, groups, or collaborations to decide their data formats in isolation of one another, creating a significant barrier to entry for multimodal research in astronomy. By making MULTIMODAL UNIVERSE publicly available, we aim to

catalyze innovation and collaboration among the astrophysics and ML communities. We believe that this dataset will not only advance our understanding of the universe but also contribute to the broader development of multimodal and metadata-aware ML methodologies.

2   Related Work

Scientific Machine Learning Datasets in Other Fields   Large curated and open datasets are gaining traction outside of the textual domain. In remote sensing, Major-TOM [54] is a collated set of ESA Copernicus mission data that aims to standardise earth observation imagery into a common, ML-friendly format. Currently, Major-TOM comprises a combined 62 TB of imagery. Like the MULTIMODAL UNIVERSE, Major-TOM is run as an open source project that the community can collectively contribute to. In a similar vein, the MOMENT project [62] introduces a large time-series dataset, as a publicly-available collection of 1.2B timestamps taken from 13 domains. These domains include—but are not limited to—medical information, economic indicators, power consumption, IoT weather data, speech, and general instrumental monitoring.

Large-Scale Machine Learning Datasets in Astrophysics   While astrophysics is inherently data-rich, most datasets have been compiled with traditional analyses in mind. This has resulted in data archives that are experiment-specific, non-uniformly stored, and not optimized for ML applications. Indeed, only a few exceptions exist where large datasets have been compiled for ML. For instance, [134, 132] introduced a substantial dataset of 76M images, enabling one of the first instances of self-supervised learning in astrophysics. Additionally, the PLAsTiCC [138] light curve classification challenge is the largest collection of astronomical time-series data for ML research, with 3.5 million simulated light curves. Finally, the Galaxy Zoo project has provided large galaxy morphology classification datasets for ML [10, 145] that have been instrumental in various astronomy specific ML studies. This includes early astronomy applications of CNNs [45], Bayesian deep learning [148], and modern neural scaling law analyses [144, 131]. However, none of these examples provide multimodal data, nor reach the scale of the data collected for the MULTIMODAL UNIVERSE.

3   Creating a Large-Scale Dataset of Diverse Astronomical Data

Modern astronomy has entered an era of large-scale, systematic surveys, supported by a range of space-based and ground-based instruments. These surveys generate unprecedented volumes of science-ready data products that are publicly available through open-access databases. However, these products are typically designed for conventional scientific workflows and not well-suited for ML applications; for example, astronomical imaging surveys typically provide full-frame mosaics from which individual uniform galaxy images must be cut out. As such, using these datasets effectively requires a high level of expertise and can be time-consuming. In addition, data structures and data access vary significantly among surveys, making it difficult to navigate diverse systems and interfaces to gather the necessary data. Therefore, to make use of astrophysical data, researchers must understand the intricacies of each separate survey and corresponding archive. This steep learning curve is a significant barrier to entry for ML applications in astrophysics, highlighting the need for a more streamlined and standardized approach to data access and preparation.

3.1   Methodology

Data Curation Strategy   To overcome the barriers between different subfields and datasets imposed by the significant domain-specific scientific expertise required, we build the MULTIMODAL UNIVERSE by drawing on a community of diverse domain scientists. To facilitate their engagement, we have structured the data curation process to have a relatively low barrier to entry. Specifically, we enable domain scientists to contribute by simply adding scripts that download and process data from the official archives they know how to access into a standardized MULTIMODAL UNIVERSE format. While we do not impose stringent code quality requirements at this stage, it is essential that these scripts are reproducible and clearly document any selection or cuts applied to the original data. This approach
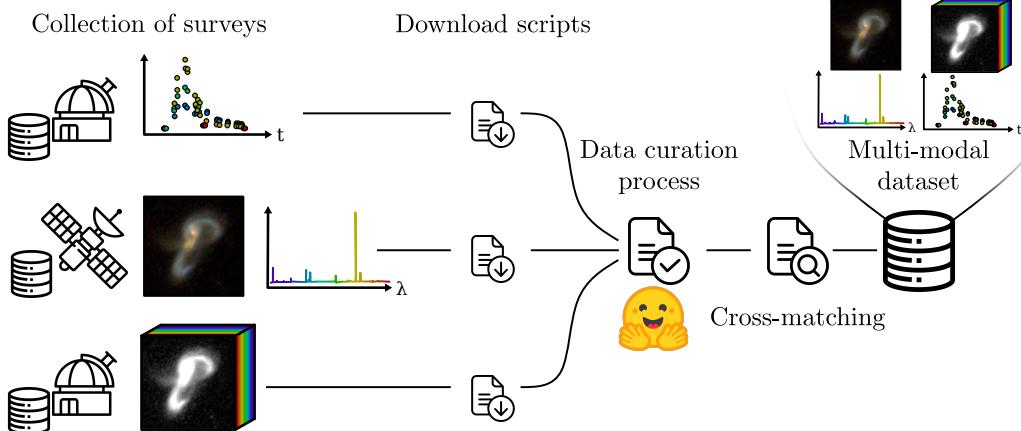
Figure 2: Illustration of the methodology behind the MULTIMODAL UNIVERSE. Domain scientists with expertise in a given astronomical survey provide data download and formatting scripts through Pull Requests on the project repository https://github.com/MultimodalUniverse/MultimodalUniverse. All datasets are then downloaded from their original source and made available as Hugging Face datasets sharing a common data schema for each modality and associated metadata. End-users can then generate any combination of MULTIMODAL UNIVERSE subsets using provided cross-matching utilities to generate multimodal datasets.

allows researchers to contribute without extensive software engineering skills, as long as data collection is transparent and well-documented. All data collection scripts are publicly available to ensure the data curation process is reproducible. Once data is downloaded from its official repository, our data curation strategy involves adopting the HuggingFace Datasets framework and imposing a standardized format for each modality, ensuring compatibility across different surveys and ease of use for ML applications.

Multimodal Cross-Matching Multimodality is a crucial aspect of the MULTIMODAL UNIVERSE, but it presents several challenges. Specifically, each astronomical survey provides data in a single modality; therefore, to build multimodal examples, it is necessary to cross-match observations from several sources. Fortunately, a fundamental property of every survey is the provision of sky coordinates for each observation, enabling cross-survey matching via coordinate matching. However, the intersection between two or more surveys is often relatively small. This limitation inherently prevents the collection of dataset pairs on the scale of projects like Laion-5B [125], primarily because some modalities are much more expensive to acquire than others. Consequently, it is essential to provide access to large-scale unpaired datasets, along with the tools to easily generate various multimodal dataset pairings. The MULTIMODAL UNIVERSE addresses this need by offering a set of common cross-matched datasets and utilities that enable end-users to generate custom cross-matches tailored to their specific scientific applications. This approach ensures flexibility and maximizes the dataset's utility across different research needs.

Figure 2 summarizes the steps of our data curation and data access methodology.

3.2 Overview of the Dataset

Table 1 provides a breakdown of the different scientific modalities included in the MULTIMODAL UNIVERSE, along with a description of their origin and their scientific relevance. Overall, the MULTIMODAL UNIVERSE represents 100 TB of data, largely dominated by 88TB of multi-band imaging.

We provide below further description of these modalities and summarize the standardized fields and metadata defined for the main modalities of the MULTIMODAL UNIVERSE in Table 2. A complete description of the schema for each modality in the dataset is further provided in Appendix A.

Table 1: Summary of samples included in the MULTIMODAL UNIVERSE, details of all samples are provided in Appendix A. N_c indicates the number of channels of a given observation. [1] These are represented as 110 basis coefficients that can be resampled to an arbitrary wavelength grid. [2] Simulated dataset.

Modality	Source Survey	N_c	Shape	Number of samples	Main science
Images	Legacy Surveys DR10 [43]	4	160×160	124M	Galaxies
	Legacy Surveys North [43, 134]	3	152×152	15M	Galaxies
	HSC [5, 3]	5	160×160	477K	Galaxies
	BTS [56, 114, 120]	3	63×63	400K	Supernovae
	JWST [13, 14, 50]	6-7	96×96	300K	Galaxies
Spectra	Gaia BP/RP [59]	-	110[1]	220M	Stars
	SDSS-II [1]	-	Variable	4M	Galaxies, Stars
	DESI [41]	-	7081	1M	Galaxies
	APOGEE SDSS-III [6]	-	7514	716k	Stars
	GALAH [28]	-	Variable	325k	Stars
	Chandra [51]	-	Variable	129K	Galaxies, Stars
	VIPERS [126]	-	557	91K	Galaxies
Hyperspectral Image	MaNGA SDSS-IV [2]	4563	96×96	12k	Galaxies
Time Series	PLAsTiCC[2][138]	6	Variable	3.5M	Time-varying objects
	TESS [121, 33]	1	Variable	1M	Exoplanets, Stars
	CfA Sample [68, 69, 18, 70]	5-11	Variable	1K	Supernovae
	YSE [7]	6	Variable	2K	Supernovae
	PS1 SNe Ia [127]	4	Variable	369	Supernovae
	DES Y3 SNe Ia [24]	4	Variable	248	Supernovae
	SNLS [63]	4	Variable	239	Supernovae
	Foundation [53, 81]	4	Variable	180	Supernovae
	CSP SNe Ia [36, 135, 86]	9	Variable	134	Supernovae
	Swift SNe Ia[26]	6	Variable	117	Supernovae
Tabular	Gaia [59]	-	-	220M	Stars
	PROVABGS [65]	-	-	221K	Galaxy
	Galaxy10 DECaLS [147, 92]	-	-	15K	Galaxy

Optical and Infrared Imaging Astronomical images are obtained via cameras installed at the focal plane of both space and ground-based telescopes, that capture photon emission from distant astronomical sources. An observed image I can be described as: $I = S * \Pi + n$ where S is the intrinsic source emission on the sky, n is the measurement noise, $*$ is the convolution operation, and Π is the instrumental response of the telescope, also known as the Point Spread Function (PSF). The PSF is at minimum characterized by its full width half maximum (fwhm), which sets the spatial resolution of the image. In addition, these images are typically acquired in a number of specific broad wavelength ranges, referred to as bands or channels. Knowledge of the bands, PSF, noise properties, and pixel scale, are generally required to fully characterize an observation and take into account the specificities of a particular instrument or observatory. Compared to natural images, astronomical images are typically noisier, and exhibit a large dynamic range, generally spanning several orders of magnitude between the bright objects and fainter sources.

Spectroscopy Similarly to imaging data, light from distant objects is gathered and focused first by the telescope, but instead of being directly imaged with a camera, the light from cosmic objects is dispersed into various wavelengths or energies using a spectrograph. A spectrum is therefore a 1-d signal representing the decomposition of the light from an object as function of wavelength or energy of the photons. Similarly to images, a measured spectrum y is the result of the following observational process: $y = L * x + n$ where L is the instrument's impulse response (also known as Line Spread Function; LSF) and n is a measurement noise. Contrary to images, where the absolute position of each pixel on the sky is not directly needed to interpret the signal, for spectroscopy knowledge of absolute wavelength corresponding to each pixel is crucial, which is why each spectrum not only provides a array of measured flux y but also the corresponding wavelength array λ.

Hyperspectral imaging Hyperspectral imaging is a data modality that combines imaging and spectroscopy to produce spatially resolved spectra in a two-dimensional region. It uses an Integral Field Unit (IFU) to produce a three-dimensional "datacube" consisting of 2d array of spatial pixel elements, i.e "spaxels", each containing a spectrum in a specific wavelength

Table 2: Description of standardized fields and metadata provided for the main modalities. These fields represent necessary and near-sufficient information to allow for the consistent interpretation of observations from multiple surveys or instruments.

Modality	Field	Description
Images	flux	Array of flux measurements of the image
	ivar	Inverse variance of noise in the image
	band	Key indicating the wavelength range of the image
	psf_fwhm	Size of the instrumental response (Point Spread Function)
	scale	Scale of pixels on the sky
Spectra	flux	Measured flux as a function of wavelength
	ivar	Inverse variance of noise on measured flux
	lsf_sigma	Size of the instrumental response (Line Spread Function)
	lambda	Wavelength of each flux measurement
Time Series	flux	Measurements of flux as a function of time
	flux_err	Uncertainty on flux measurement
	band	Key indicating the wavelength range of the measurement
	time	Time of observation

range. IFUs provide a wealth of information. For example, by analyzing the spectrum within each spaxel, one can produce "images" of spectral features and derived parameters, such as emission/absorption lines and the velocities of gas and stars in the galaxy. By aggregating spectra together within specific spatial regions, one can compare the properties (such as star forming rates) of different physical regions of the galaxy. By extension of both images and spectra, hyperspectral images are accompanied by an instrumental response in the form of both Point Spread Function and Line Spread Function information.

Time-Series Time series data, often called light curves (brightness over time), are commonly examined when astronomical objects or events vary in brightness in time. These include relatively short-lived, explosive transients such as supernovae (exploding stars) and tidal disruption events, long-lived sources that show stochastic variation over time such as active galactic nuclei and quasars, and (quasi-)periodic sources such as pulsating stars and exoplanets. These time series are a sequence of flux (i.e. intensity) measurements. Because of their unique properties, astronomical time-series pose a challenge to traditional ML architectures. For one, most light curves are sparsely and irregularly sampled in both time and wavelength due to the unique survey strategy of each telescope. Flux measurements are also often plagued with large uncertainties, especially those from faint sources. Finally, data from different telescopes are highly heterogeneous and thus difficult to combine (e.g., to build a larger training set for ML applications). Together, these factors make our time-series data products a compelling and challenging dataset for ML researchers.

Tabular data The MULTIMODAL UNIVERSE also aggregates large catalogs of tabular data from different sources, corresponding either to particular measurements (e.g. measured flux for a particular object), or the result of the processing of these observations by an analysis pipeline (e.g. calculations of galaxy properties or reporting of galaxy morphology)[1].

4 Examples of Machine Learning Tasks on Different Modalities

The data accumulated in the MULTIMODAL UNIVERSE covers a wide range of scientific uses-cases, spanning sub-fields of astrophysics from stellar physics to galaxy formation. We include in this section a few conventional examples of ML tasks in astrophysics as a way to illustrate how to interact with the MULTIMODAL UNIVERSE dataset and to provide illustrations of scientific uses cases. We note that these examples are far from exhaustive.

---

[1]We do not enforce standard definitions for tabular data and instead refer the user to the original survey description.

Pretraining	Model	Top-1 Accuracy
	EfficientNetB0	**80.9** ±0.1 %
No pretraining	ConvNext-nano [144]	75.6 ±1.8 %
	ResNet18	73.9 ±0.9 %
	DenseNet121	73.5 ±2.4 %
Galazy Zoo	ConvNext-nano [144]	**89.3** ±0.1 %
ImageNet-12k	ConvNext-nano [144]	83.9 ±0.3 %

Table 3: Top-1 Accuracy on the Galaxy10 DECaLS morphology classification dataset.

4.1 Characterisation of Galaxies

Morphology Classification   Galaxy shape or morphology is a first-order tracer of the history of a galaxy. As such, classifying images of galaxies based on their apparent morphologies and visual features is a common task. Labeling these features is usually done by visual inspection from experts or using citizen science, such as [10, 145]. Here, we use the Galaxy10 DECaLS catalog of human-labeled annotations spanning ten broad morphological classes paired with the MULTIMODAL UNIVERSE's RGB-converted images from the Legacy Surveys and report the top-1 accuracy of supervised baselines on the held-out test set in Table 3. We include state of the art results from the literature from [144] as a point of reference and compare different common architectures for this task (EfficientNetB0[136], ResNet18[66], and DenseNet121[75]).

Physical Property Inference   Astrophysicists often seek to predict fundamental properties of galaxies from observational data. Accurate predictions of these properties provide key insights into galaxy formation and evolution. Here, we consider the following properties:

- Redshift (Z_{HP}): The extent to which the light from a galaxy has been stretched by the expansion of the universe, helping to determine the galaxy's distance from earth.
- Stellar Mass ($\log M_*$): The total mass of stars in a galaxy in units of solar masses.
- Age ($t_{age,MW}$): The age of stars in a galaxy, weighted by galaxy stellar mass.
- Specific Star-Formation Rate ($sSFR$): The rate at which stars in the galaxy are formed normalized by the galaxy stellar mass.
- Metallicity (Z_{MW}): The abundance of elements heavier than hydrogen and helium in the galaxy, typically weighted by the galaxy mass.

To implement this task, we use the MULTIMODAL UNIVERSE's built-in cross-matching functionality to combine the PROVABGS catalog of physical properties, with both the Legacy Survey North and South imaging sample and the DESI spectroscopic sample. We report the R^2 value on the held-out test set in Table 4.

Table 4: Model R^2 performance comparison for predicting galaxy properties from different observational data modalities.

Modality	Source Survey	Model	Z_{HP}	$\log M_*$	Z_{MW}	$t_{age,MW}$	$sSFR$
		ResNet18	0.771	0.725	0.381	0.210	0.405
Image	Legacy Surveys	DenseNet121	0.774	0.734	0.414	0.267	0.446
		EfficientNetB0	0.697	0.645	0.395	0.260	0.421
Spectrum	DESI	Conv+Att [104]	0.982	0.871	0.659	0.488	0.679
Photometry	PROVABGS	MLP	0.696	0.681	0.383	0.308	0.343

4.2 Characterisation of Transient events

Candidate Identification   Given their short-lived nature, early identification of transients is of great value to allow for further observations which study their evolution. The publicly-available BTSbot model [119, 120] - a multimodal convolutional neural network - was developed to automate identification of new astrophysical transient candidates in series of images and associated metadata. We report the performance of BTSbot on the binary classification of real astrophysical transients and bogus or uninteresting detections using a dataset from the Zwicky Transient Facility [ZTF; 15, 16] Bright Transient Survey [BTS; 56, 114].

Table 5: Results for astronomical time-series tasks. The performance of each model on the associated metric for the task is shown in the `Performance` column. Avocado is a Random Forest architecture while Connect Later is a transformer architecture.

Source Survey	Task	Metric	Model	Performance
BTS	Transient Candidate Identification	AUC	BTSbot [119, 120]	0.985
YSE	SN Ia classification	AUC	Random forest	0.90
PLAsTiCC	14-way classification	Accuracy	Avocado [20]	77.4
	14-way classification	Accuracy	Connect Later [117]	79.9
	Redshift estimation	RMSE	Connect Later [117]	0.247

Light-curve Classification and Regression   Classifying astrophysical sources can be challenging due to the sparsity of time-series data, limited filter coverage, or simply similarity between time-series of different object types. Spectroscopic data are required for unambiguous classification, but these follow-up observations will be prohibitively expensive for large surveys. Accurate photometric classification of astrophysical transients is thus highly desirable. This is a supervised learning problem, and has been attempted with both simulated [e.g. 83, 101, 107, 116, 71] and real training data [e.g. 143, 46, 30, 31, 91]. We provide baseline photometric classification results on a real dataset from the Young Supernova Experiment (YSE) as well as a simulated dataset from the Photometric LSST Astronomical Classification Challenge (PLAsTiCC). While we report state-of-the-art results from the literature on PLAsTiCC, YSE represents a specific challenge due to its limited size, and as such specifically provides an implementation of a Random Forest classifier, which is a conventional approach for handling limited dataset sizes. Details of these methods are included in Appendix B. Finally, as a related task, we also include an example of estimating redshift from light curves on the PLAsTiCC dataset. All metric performances are stated in Table 5.

4.3   Bridging the Modality Gap: Contrastive Image-Spectrum Pretraining

As a concrete example of the types of models that can be created thanks to the access to a large repository of multimodal data, we present a reproduction of the recent `AstroCLIP` method [112]. This variant of Contrastive Language-Image Pretraining (CLIP) [118] aims to build in a self-supervised representations of multimodal data by contrastive alignment between image and spectra modalities. `AstroCLIP` embeddings have been shown to perform in-line with supervised baselines on tasks like redshift prediction, galaxy property prediction, and morphology classification while also allowing for powerful in-modality and cross-modal retrieval. We demonstrate how MULTIMODAL UNIVERSE makes it possible to easily build such models by generating a cross-matched dataset of images from the Legacy Survey sample and optical spectra from the DESI sample. From this dataset we can train by contrastive learning image and spectra embeddings. We show in Table 6 k-Nearest Neighbour zero-shot regression results for redshift and galaxy property prediction, and show better performance compared to the supervised results reported in subsection 4.1.

Table 6: R^2 performance of zero-shot prediction of galaxy properties from image and spectrum `AstroCLIP` embeddings following the strategy described in [112].

Modality	Source Survey	Z_{HP}	$\log M_*$	Z_{MW}	$t_{age,MW}$	$sSFR$
Image	Legacy Surveys	0.801	0.737	0.432	0.240	0.435
Spectrum	DESI	0.986	0.879	0.584	0.441	0.643

5   Discussion

Paving the road towards Foundational Research in Scientific Machine Learning Beyond its multimodal and large-scale nature, the MULTIMODAL UNIVERSE also serves as a valuable testbed for addressing some of the practical challenges encountered in the use of machine learning within application domains such as astrophysics and other scientific disciplines. Although machine learning has been applied in astronomy since at least the 1990s, and been extensively adopted in astrophysics research over the past decade, most

applications are tailored to specific datasets, requiring domain expertise, with models often being trained from scratch for similar purposes. Consequently, machine learning applications face issues such as distribution shifts [47, 155], uncertainty quantification [106], and the proper calibration of predictive models [78]. The MULTIMODAL UNIVERSE is designed to facilitate the development of next-generation machine learning models in astrophysics, enabling research into each of these critical facets of applied machine learning in astronomy. We believe that the complexity, multimodal nature, scale, and the culture of publicly released data rooted in the astronomical community will enable the machine learning community at large to tackle some of the key challenges in large-scale machine learning development.

Limitations While the MULTIMODAL UNIVERSE includes a significant collection of multimodal scientific data, a notable omission is the lack of associated textual data. Providing image-text or observation-text pairs is not currently the norm for astronomical observations. Therefore, despite promising work on astronomy fine-tuned LLMs [109, 139], LLM evaluation [152], observation proposal based CLIP models [105], and NLP for improved language in astronomy [22, 23], the MULTIMODAL UNIVERSE does not currently contain any text-based captions or corpora. An additional limitation is that the MULTIMODAL UNIVERSE does not guarantee fully cross-matched samples (i.e. all modalities for each source). This is a consequence of observing strategies and science targets of the respective surveys. However, we note that the current outlook of astronomy surveys is towards 'all sky surveys'. As such, we expect that any future version of the MULTIMODAL UNIVERSE will necessarily contain more cross-matched samples than are currently available.

6   Availability and Maintenance

The dataset is hosted in full at the Flatiron Institute and accessible both through HTTPS[2] or through GLOBUS[3] while associated resources and all source code necessary to reproduce the dataset compilation are hosted on GitHub. In addition, a preview of each dataset is hosted on the Hugging Face Hub in parquet format to allow for online streaming and easy prototyping. To ensure the continued evolution and relevance of MULTIMODAL UNIVERSE, the project is organized as an open and collaborative effort. Contributions from researchers are actively encouraged, with a well-defined process for submitting new data, benchmarks, and improvements via GitHub. A dedicated team of maintainers oversees the integration of these contributions, ensuring that the dataset remains up-to-date and continues to meet the needs of the scientific and machine learning communities. This framework will enable regular updates to incorporate new data from ongoing and future astronomical surveys, as well as periodic evaluations and improvements based on feedback from the research community. Additionally, our open collaboration hosts forums via GitHub and a dedicated slack to engage with both ML and astrophysics communities, and help the MULTIMODAL UNIVERSE adapt and evolve to the respective dynamic research environments.

---

[2] https://users.flatironinstitute.org/~polymathic/data/MultimodalUniverse/
[3] https://app.globus.org/file-manager?origin_id=57136152-fc1d-418e-b74e-75ca52bddd21

For data specific acknowledgments see each individual dataset in Appendix A.

Contributors

The contribution categories listed below are broadly indicative of the ways in which authors contributed to the project, although we note these categories are not exhaustive.

Coordination Team: Micah Bowles, Marc Huertas-Company, Francois Lanusse, Liam H. Parker, Helen Qu, Michael J. Smith.

Datasets: Erini Angeloudi, Jeroen Audenaert, Micah Bowles, Benjamin M. Boyd, David Chemaly, Brian Cherinka, Aaron Do, Matthew Grayling, Erin E. Hayes, Tom Hehir, Marc Huertas-Company, Kartheik G. Iyer, Maja Jablonska, Francois Lanusse, Henry W. Leung, Kaisey Mandel, Juan Rafael Martínez-Galarza, Peter Melchior, Liam H. Parker, Helen Qu, Jeff Shen, Michael J. Smith, Connor Stone, Mike Walmsley, John F. Wu.

Baselines: Eirini Angeloudi, Jeroen Audenaert, Micah Bowles, Benjamin M. Boyd, David Chemaly, Aaron Do, Matthew Grayling, Erin Hayes, Tom Hehir, Marc Huertas-Company, Kaisey Mandel, Liam H. Parker, Helen Qu, Michael J. Smith.

Overall/Infrastructure: Micah Bowles, Ioana Ciucă, Miles Cranmer, Shirley Ho, Marc Huertas-Company, Francois Lanusse, Lucas Meyer, Liam H. Parker, Michael J. Smith, Mike Walmsley.

References

[1] Kevork N. Abazajian, Jennifer K. Adelman-McCarthy, Marcel A. Agüeros, Sahar S. Allam, Carlos Allende Prieto, Deokkeun An, Kurt S. J. Anderson, Scott F. Anderson, James Annis, Neta A. Bahcall, C. A. L. Bailer-Jones, J. C. Barentine, Bruce A. Bassett, Andrew C. Becker, Timothy C. Beers, Eric F. Bell, Vasily Belokurov, Andreas A. Berlind, Eileen F. Berman, Mariangela Bernardi, Steven J. Bickerton, Dmitry Bizyaev, John P. Blakeslee, Michael R. Blanton, John J. Bochanski, William N. Boroski, Howard J. Brewington, Jarle Brinchmann, J. Brinkmann, Robert J. Brunner, Tamás Budavári, Larry N. Carey, Samuel Carliles, Michael A. Carr, Francisco J. Castander, David Cinabro, A. J. Connolly, István Csabai, Carlos E. Cunha, Paul C. Czarapata, James R. A. Davenport, Ernst de Haas, Ben Dilday, Mamoru Doi, Daniel J. Eisenstein, Michael L. Evans, N. W. Evans, Xiaohui Fan, Scott D. Friedman, Joshua A. Frieman, Masataka Fukugita, Boris T. Gänsicke, Evalyn Gates, Bruce Gillespie, G. Gilmore, Belinda Gonzalez, Carlos F. Gonzalez, Eva K. Grebel, James E. Gunn, Zsuzsanna Györy, Patrick B. Hall, Paul Harding, Frederick H. Harris, Michael Harvanek, Suzanne L. Hawley, Jeffrey J. E. Hayes, Timothy M. Heckman, John S. Hendry, Gregory S. Hennessy, Robert B. Hindsley, J. Hoblitt, Craig J. Hogan, David W. Hogg, Jon A. Holtzman, Joseph B. Hyde, Shin-ichi Ichikawa, Takashi Ichikawa, Myungshin Im, Željko Ivezić, Sebastian Jester, Linhua Jiang, Jennifer A. Johnson, Anders M. Jorgensen, Mario Jurić, Stephen M. Kent, R. Kessler, S. J. Kleinman, G. R. Knapp, Kohki Konishi, Richard G. Kron, Jurek Krzesinski, Nikolay Kuropatkin, Hubert Lampeitl, Svetlana Lebedeva, Myung Gyoon Lee, Young Sun Lee, R. French Leger, Sébastien Lépine, Nolan Li, Marcos Lima, Huan Lin, Daniel C. Long, Craig P. Loomis, Jon Loveday, Robert H. Lupton, Eugene Magnier, Olena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Bruce Margon, John P. Marriner, David Martínez-Delgado, Takahiko Matsubara, Peregrine M. McGehee, Timothy A. McKay, Avery Meiksin, Heather L. Morrison, Fergal Mullally, Jeffrey A. Munn, Tara Murphy, Thomas Nash, Ada Nebot, Jr. Neilsen, Eric H., Heidi Jo Newberg, Peter R. Newman, Robert C. Nichol, Tom Nicinski, Maria Nieto-Santisteban, Atsuko Nitta, Sadanori Okamura, Daniel J. Oravetz, Jeremiah P. Ostriker, Russell Owen, Nikhil Padmanabhan, Kaike

Pan, Changbom Park, George Pauls, Jr. Peoples, John, Will J. Percival, Jeffrey R. Pier, Adrian C. Pope, Dimitri Pourbaix, Paul A. Price, Norbert Purger, Thomas Quinn, M. Jordan Raddick, Paola Re Fiorentin, Gordon T. Richards, Michael W. Richmond, Adam G. Riess, Hans-Walter Rix, Constance M. Rockosi, Masao Sako, David J. Schlegel, Donald P. Schneider, Ralf-Dieter Scholz, Matthias R. Schreiber, Axel D. Schwope, Uroš Seljak, Branimir Sesar, Erin Sheldon, Kazu Shimasaku, Valena C. Sibley, A. E. Simmons, Thirupathi Sivarani, J. Allyn Smith, Martin C. Smith, Vernesa Smolčić, Stephanie A. Snedden, Albert Stebbins, Matthias Steinmetz, Chris Stoughton, Michael A. Strauss, Mark SubbaRao, Yasushi Suto, Alexander S. Szalay, István Szapudi, Paula Szkody, Masayuki Tanaka, Max Tegmark, Luis F. A. Teodoro, Aniruddha R. Thakar, Christy A. Tremonti, Douglas L. Tucker, Alan Uomoto, Daniel E. Vanden Berk, Jan Vandenberg, S. Vidrih, Michael S. Vogeley, Wolfgang Voges, Nicole P. Vogt, Yogesh Wadadekar, Shannon Watters, David H. Weinberg, Andrew A. West, Simon D. M. White, Brian C. Wilhite, Alainna C. Wonders, Brian Yanny, D. R. Yocum, Donald G. York, Idit Zehavi, Stefano Zibetti, and Daniel B. Zucker. The Seventh Data Release of the Sloan Digital Sky Survey. ApJS, 182(2):543–558, June 2009.

[2] Abdurro'uf, Katherine Accetta, Conny Aerts, Víctor Silva Aguirre, Romina Ahumada, Nikhil Ajgaonkar, N. Filiz Ak, Shadab Alam, Carlos Allende Prieto, Andrés Almeida, Friedrich Anders, Scott F. Anderson, Brett H. Andrews, Borja Anguiano, Erik Aquino-Ortíz, Alfonso Aragón-Salamanca, Maria Argudo-Fernández, Metin Ata, Marie Aubert, Vladimir Avila-Reese, Carles Badenes, Rodolfo H. Barbá, Kat Barger, Jorge K. Barrera-Ballesteros, Rachael L. Beaton, Timothy C. Beers, Francesco Belfiore, Chad F. Bender, Mariangela Bernardi, Matthew A. Bershady, Florian Beutler, Christian Moni Bidin, Jonathan C. Bird, Dmitry Bizyaev, Guillermo A. Blanc, Michael R. Blanton, Nicholas Fraser Boardman, Adam S. Bolton, Médéric Boquien, Jura Borissova, Jo Bovy, W. N. Brandt, Jordan Brown, Joel R. Brownstein, Marcella Brusa, Johannes Buchner, Kevin Bundy, Joseph N. Burchett, Martin Bureau, Adam Burgasser, Tuesday K. Cabang, Stephanie Campbell, Michele Cappellari, Joleen K. Carlberg, Fábio Carneiro Wanderley, Ricardo Carrera, Jennifer Cash, Yan-Ping Chen, Wei-Huai Chen, Brian Cherinka, Cristina Chiappini, Peter Doohyun Choi, S. Drew Chojnowski, Haeun Chung, Nicolas Clerc, Roger E. Cohen, Julia M. Comerford, Johan Comparat, Luiz da Costa, Kevin Covey, Jeffrey D. Crane, Irene Cruz-Gonzalez, Connor Culhane, Katia Cunha, Y. Sophia Dai, Guillermo Damke, Jeremy Darling, Jr. Davidson, James W., Roger Davies, Kyle Dawson, Nathan De Lee, Aleksandar M. Diamond-Stanic, Mariana Cano-Díaz, Helena Domínguez Sánchez, John Donor, Chris Duckworth, Tom Dwelly, Daniel J. Eisenstein, Yvonne P. Elsworth, Eric Emsellem, Mike Eracleous, Stephanie Escoffier, Xiaohui Fan, Emily Farr, Shuai Feng, José G. Fernández-Trincado, Diane Feuillet, Andreas Filipp, Sean P. Fillingham, Peter M. Frinchaboy, Sebastien Fromenteau, Lluís Galbany, Rafael A. García, D. A. García-Hernández, Junqiang Ge, Doug Geisler, Joseph Gelfand, Tobias Géron, Benjamin J. Gibson, Julian Goddy, Diego Godoy-Rivera, Kathleen Grabowski, Paul J. Green, Michael Greener, Catherine J. Grier, Emily Griffith, Hong Guo, Julien Guy, Massinissa Hadjara, Paul Harding, Sten Hasselquist, Christian R. Hayes, Fred Hearty, Jesús Hernández, Lewis Hill, David W. Hogg, Jon A. Holtzman, Danny Horta, Bau-Ching Hsieh, Chin-Hao Hsu, Yun-Hsin Hsu, Daniel Huber, Marc Huertas-Company, Brian Hutchinson, Ho Seong Hwang, Héctor J. Ibarra-Medel, Jacob Ider Chitham, Gabriele S. Ilha, Julie Imig, Will Jaekle, Tharindu Jayasinghe, Xihan Ji, Jennifer A. Johnson, Amy Jones, Henrik Jönsson, Ivan Katkov, Dr. Khalatyan, Arman, Karen Kinemuchi, Shobhit Kisku, Johan H. Knapen, Jean-Paul Kneib, Juna A. Kollmeier, Miranda Kong, Marina Kounkel, Kathryn Kreckel, Dhanesh Krishnarao, Ivan Lacerna, Richard R. Lane, Rachel Langgin, Ramon Lavender, David R. Law, Daniel Lazarz, Henry W. Leung, Ho-Hin Leung, Hannah M. Lewis, Cheng Li, Ran Li, Jianhui Lian, Fu-Heng Liang, Lihwai Lin, Yen-Ting Lin, Sicheng Lin, Chris Lintott, Dan Long, Penélope Longa-Peña, Carlos López-Cobá, Shengdong Lu, Britt F. Lundgren, Yuanze Luo, J. Ted Mackereth, Axel de la Macorra, Suvrath Mahadevan, Steven R. Majewski, Arturo Manchado, Travis Mandeville, Claudia Maraston, Berta Margalef-Bentabol, Thomas Masseron, Karen L. Masters, Savita Mathur, Richard M. McDermid, Myles Mckay, Andrea Merloni, Michael Merrifield, Szabolcs Meszaros, Andrea Miglio, Francesco Di Mille, Dante Minniti, Rebecca Minsley, Antonela Monachesi,

Jeongin Moon, Benoit Mosser, John Mulchaey, Demitri Muna, Ricardo R. Muñoz, Adam D. Myers, Natalie Myers, Seshadri Nadathur, Preethi Nair, Kirpal Nandra, Justus Neumann, Jeffrey A. Newman, David L. Nidever, Farnik Nikakhtar, Christian Nitschelm, Julia E. O'Connell, Luis Garma-Oehmichen, Gabriel Luan Souza de Oliveira, Richard Olney, Daniel Oravetz, Mario Ortigoza-Urdaneta, Yeisson Osorio, Justin Otter, Zachary J. Pace, Nelson Padilla, Kaike Pan, Hsi-An Pan, Taniya Parikh, James Parker, Sebastien Peirani, Karla Peña Ramírez, Samantha Penny, Will J. Percival, Ismael Perez-Fournon, Marc Pinsonneault, Frédérick Poidevin, Vijith Jacob Poovelil, Adrian M. Price-Whelan, Anna Bárbara de Andrade Queiroz, M. Jordan Raddick, Amy Ray, Sandro Barboza Rembold, Nicole Riddle, Rogemar A. Riffel, Rogério Riffel, Hans-Walter Rix, Annie C. Robin, Aldo Rodríguez-Puebla, Alexandre Roman-Lopes, Carlos Román-Zúñiga, Benjamin Rose, Ashley J. Ross, Graziano Rossi, Kate H. R. Rubin, Mara Salvato, Sebástian F. Sánchez, José R. Sánchez-Gallego, Robyn Sanderson, Felipe Antonio Santana Rojas, Edgar Sarceno, Regina Sarmiento, Conor Sayres, Elizaveta Sazonova, Adam L. Schaefer, Ricardo Schiavon, David J. Schlegel, Donald P. Schneider, Mathias Schultheis, Axel Schwope, Aldo Serenelli, Javier Serna, Zhengyi Shao, Griffin Shapiro, Anubhav Sharma, Yue Shen, Matthew Shetrone, Yiping Shu, Joshua D. Simon, M. F. Skrutskie, Rebecca Smethurst, Verne Smith, Jennifer Sobeck, Taylor Spoo, Dani Sprague, David V. Stark, Keivan G. Stassun, Matthias Steinmetz, Dennis Stello, Alexander Stone-Martinez, Thaisa Storchi-Bergmann, Guy S. Stringfellow, Amelia Stutz, Yung-Chau Su, Manuchehr Taghizadeh-Popp, Michael S. Talbot, Jamie Tayar, Eduardo Telles, Johanna Teske, Ani Thakar, Christopher Theissen, Andrew Tkachenko, Daniel Thomas, Rita Tojeiro, Hector Hernandez Toledo, Nicholas W. Troup, Jonathan R. Trump, James Trussler, Jacqueline Turner, Sarah Tuttle, Eduardo Unda-Sanzana, José Antonio Vázquez-Mata, Marica Valentini, Octavio Valenzuela, Jaime Vargas-González, Mariana Vargas-Magaña, Pablo Vera Alfaro, Sandro Villanova, Fiorenzo Vincenzo, David Wake, Jack T. Warfield, Jessica Diane Washington, Benjamin Alan Weaver, Anne-Marie Weijmans, David H. Weinberg, Achim Weiss, Kyle B. Westfall, Vivienne Wild, Matthew C. Wilde, John C. Wilson, Robert F. Wilson, Mikayla Wilson, Julien Wolf, W. M. Wood-Vasey, Renbin Yan, Olga Zamora, Gail Zasowski, Kai Zhang, Cheng Zhao, Zheng Zheng, Zheng Zheng, and Kai Zhu. The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data. ApJS, 259(2):35, April 2022.

[3] Hiroaki Aihara, Yusra AlSayyad, Makoto Ando, Robert Armstrong, James Bosch, Eiichi Egami, Hisanori Furusawa, Junko Furusawa, Sumiko Harasawa, Yuichi Harikane, Bau-Ching Hsieh, Hiroyuki Ikeda, Kei Ito, Ikuru Iwata, Tadayuki Kodama, Michitaro Koike, Mitsuru Kokubo, Yutaka Komiyama, Xiangchong Li, Yongming Liang, Yen-Ting Lin, Robert H. Lupton, Nate B. Lust, Lauren A. MacArthur, Ken Mawatari, Sogo Mineo, Hironao Miyatake, Satoshi Miyazaki, Surhud More, Takahiro Morishima, Hitoshi Murayama, Kimihiko Nakajima, Fumiaki Nakata, Atsushi J. Nishizawa, Masamune Oguri, Nobuhiro Okabe, Yuki Okura, Yoshiaki Ono, Ken Osato, Masami Ouchi, Yen-Chen Pan, Andrés A. Plazas Malagón, Paul A. Price, Sophie L. Reed, Eli S. Rykoff, Takatoshi Shibuya, Mirko Simunovic, Michael A. Strauss, Kanako Sugimori, Yasushi Suto, Nao Suzuki, Masahiro Takada, Yuhei Takagi, Tadafumi Takata, Satoshi Takita, Masayuki Tanaka, Shenli Tang, Dan S. Taranu, Tsuyoshi Terai, Yoshiki Toba, Edwin L. Turner, Hisakazu Uchiyama, Bovornpratch Vijarnwannaluk, Christopher Z. Waters, Yoshihiko Yamada, Naoaki Yamamoto, and Takuji Yamashita. Third data release of the Hyper Suprime-Cam Subaru Strategic Program. PASJ, 74(2):247–272, April 2022.

[4] Hiroaki Aihara, Yusra AlSayyad, Makoto Ando, Robert Armstrong, James Bosch, Eiichi Egami, Hisanori Furusawa, Junko Furusawa, Sumiko Harasawa, Yuichi Harikane, Bau-Ching Hsieh, Hiroyuki Ikeda, Kei Ito, Ikuru Iwata, Tadayuki Kodama, Michitaro Koike, Mitsuru Kokubo, Yutaka Komiyama, Xiangchong Li, Yongming Liang, Yen-Ting Lin, Robert H. Lupton, Nate B. Lust, Lauren A. MacArthur, Ken Mawatari, Sogo Mineo, Hironao Miyatake, Satoshi Miyazaki, Surhud More, Takahiro Morishima, Hitoshi Murayama, Kimihiko Nakajima, Fumiaki Nakata, Atsushi J. Nishizawa, Masamune Oguri, Nobuhiro Okabe, Yuki Okura, Yoshiaki Ono, Ken Osato, Masami Ouchi, Yen-Chen Pan, Andrés A. Plazas Malagón, Paul A. Price, Sophie L. Reed, Eli S. Rykoff, Takatoshi Shibuya, Mirko Simunovic, Michael A. Strauss, Kanako Sugimori, Yasushi

Suto, Nao Suzuki, Masahiro Takada, Yuhei Takagi, Tadafumi Takata, Satoshi Takita, Masayuki Tanaka, Shenli Tang, Dan S. Taranu, Tsuyoshi Terai, Yoshiki Toba, Edwin L. Turner, Hisakazu Uchiyama, Bovornpratch Vijarnwannaluk, Christopher Z. Waters, Yoshihiko Yamada, Naoaki Yamamoto, and Takuji Yamashita. Third data release of the Hyper Suprime-Cam Subaru Strategic Program. PASJ, 74(2):247–272, April 2022.

[5] Hiroaki Aihara, Nobuo Arimoto, Robert Armstrong, Stéphane Arnouts, Neta A. Bahcall, Steven Bickerton, James Bosch, Kevin Bundy, Peter L. Capak, James H. H. Chan, Masashi Chiba, Jean Coupon, Eiichi Egami, Motohiro Enoki, Francois Finet, Hiroki Fujimori, Seiji Fujimoto, Hisanori Furusawa, Junko Furusawa, Tomotsugu Goto, Andy Goulding, Johnny P. Greco, Jenny E. Greene, James E. Gunn, Takashi Hamana, Yuichi Harikane, Yasuhiro Hashimoto, Takashi Hattori, Masao Hayashi, Yusuke Hayashi, Krzysztof G. Hełminiak, Ryo Higuchi, Chiaki Hikage, Paul T. P. Ho, Bau-Ching Hsieh, Kuiyun Huang, Song Huang, Hiroyuki Ikeda, Masatoshi Imanishi, Akio K. Inoue, Kazushi Iwasawa, Ikuru Iwata, Anton T. Jaelani, Hung-Yu Jian, Yukiko Kamata, Hiroshi Karoji, Nobunari Kashikawa, Nobuhiko Katayama, Satoshi Kawanomoto, Issha Kayo, Jin Koda, Michitaro Koike, Takashi Kojima, Yutaka Komiyama, Akira Konno, Shintaro Koshida, Yusei Koyama, Haruka Kusakabe, Alexie Leauthaud, Chien-Hsiu Lee, Lihwai Lin, Yen-Ting Lin, Robert H. Lupton, Rachel Mandelbaum, Yoshiki Matsuoka, Elinor Medezinski, Sogo Mineo, Shoken Miyama, Hironao Miyatake, Satoshi Miyazaki, Rieko Momose, Anupreeta More, Surhud More, Yuki Moritani, Takashi J. Moriya, Tomoki Morokuma, Shiro Mukae, Ryoma Murata, Hitoshi Murayama, Tohru Nagao, Fumiaki Nakata, Mana Niida, Hiroko Niikura, Atsushi J. Nishizawa, Yoshiyuki Obuchi, Masamune Oguri, Yukie Oishi, Nobuhiro Okabe, Sakurako Okamoto, Yuki Okura, Yoshiaki Ono, Masato Onodera, Masafusa Onoue, Ken Osato, Masami Ouchi, Paul A. Price, Tae-Soo Pyo, Masao Sako, Marcin Sawicki, Takatoshi Shibuya, Kazuhiro Shimasaku, Atsushi Shimono, Masato Shirasaki, John D. Silverman, Melanie Simet, Joshua Speagle, David N. Spergel, Michael A. Strauss, Yuma Sugahara, Naoshi Sugiyama, Yasushi Suto, Sherry H. Suyu, Nao Suzuki, Philip J. Tait, Masahiro Takada, Tadafumi Takata, Naoyuki Tamura, Manobu M. Tanaka, Masaomi Tanaka, Masayuki Tanaka, Yoko Tanaka, Tsuyoshi Terai, Yuichi Terashima, Yoshiki Toba, Nozomu Tominaga, Jun Toshikawa, Edwin L. Turner, Tomohisa Uchida, Hisakazu Uchiyama, Keiichi Umetsu, Fumihiro Uraguchi, Yuji Urata, Tomonori Usuda, Yousuke Utsumi, Shiang-Yu Wang, Wei-Hao Wang, Kenneth C. Wong, Kiyoto Yabe, Yoshihiko Yamada, Hitomi Yamanoi, Naoki Yasuda, Sherry Yeh, Atsunori Yonehara, and Suraphong Yuma. The Hyper Suprime-Cam SSP Survey: Overview and survey design. PASJ, 70:S4, January 2018.

[6] Shadab Alam, Franco D. Albareti, Carlos Allende Prieto, F. Anders, Scott F. Anderson, Timothy Anderton, Brett H. Andrews, Eric Armengaud, Éric Aubourg, Stephen Bailey, Sarbani Basu, Julian E. Bautista, Rachael L. Beaton, Timothy C. Beers, Chad F. Bender, Andreas A. Berlind, Florian Beutler, Vaishali Bhardwaj, Jonathan C. Bird, Dmitry Bizyaev, Cullen H. Blake, Michael R. Blanton, Michael Blomqvist, John J. Bochanski, Adam S. Bolton, Jo Bovy, A. Shelden Bradley, W. N. Brandt, D. E. Brauer, J. Brinkmann, Peter J. Brown, Joel R. Brownstein, Angela Burden, Etienne Burtin, Nicolás G. Busca, Zheng Cai, Diego Capozzi, Aurelio Carnero Rosell, Michael A. Carr, Ricardo Carrera, K. C. Chambers, William James Chaplin, Yen-Chi Chen, Cristina Chiappini, S. Drew Chojnowski, Chia-Hsun Chuang, Nicolas Clerc, Johan Comparat, Kevin Covey, Rupert A. C. Croft, Antonio J. Cuesta, Katia Cunha, Luiz N. da Costa, Nicola Da Rio, James R. A. Davenport, Kyle S. Dawson, Nathan De Lee, Timothée Delubac, Rohit Deshpande, Saurav Dhital, Letícia Dutra-Ferreira, Tom Dwelly, Anne Ealet, Garrett L. Ebelke, Edward M. Edmondson, Daniel J. Eisenstein, Tristan Ellsworth, Yvonne Elsworth, Courtney R. Epstein, Michael Eracleous, Stephanie Escoffier, Massimiliano Esposito, Michael L. Evans, Xiaohui Fan, Emma Fernández-Alvar, Diane Feuillet, Nurten Filiz Ak, Hayley Finley, Alexis Finoguenov, Kevin Flaherty, Scott W. Fleming, Andreu Font-Ribera, Jonathan Foster, Peter M. Frinchaboy, J. G. Galbraith-Frew, Rafael A. García, D. A. García-Hernández, Ana E. García Pérez, Patrick Gaulme, Jian Ge, R. Génova-Santos, A. Georgakakis, Luan Ghezzi, Bruce A. Gillespie, Léo Girardi, Daniel Goddard, Satya Gontcho A. Gontcho, Jonay I. González Hernández, Eva K. Grebel, Paul J. Green, Jan Niklas Grieb, Nolan Grieves, James E. Gunn, Hong Guo, Paul Harding, Sten Hasselquist, Suzanne L. Hawley, Michael Hayden,

Fred R. Hearty, Saskia Hekker, Shirley Ho, David W. Hogg, Kelly Holley-Bockelmann, Jon A. Holtzman, Klaus Honscheid, Daniel Huber, Joseph Huehnerhoff, Inese I. Ivans, Linhua Jiang, Jennifer A. Johnson, Karen Kinemuchi, David Kirkby, Francisco Kitaura, Mark A. Klaene, Gillian R. Knapp, Jean-Paul Kneib, Xavier P. Koenig, Charles R. Lam, Ting-Wen Lan, Dustin Lang, Pierre Laurent, Jean-Marc Le Goff, Alexie Leauthaud, Khee-Gan Lee, Young Sun Lee, Timothy C. Licquia, Jian Liu, Daniel C. Long, Martín López-Corredoira, Diego Lorenzo-Oliveira, Sara Lucatello, Britt Lundgren, Robert H. Lupton, III Mack, Claude E., Suvrath Mahadevan, Marcio A. G. Maia, Steven R. Majewski, Elena Malanushenko, Viktor Malanushenko, A. Manchado, Marc Manera, Qingqing Mao, Claudia Maraston, Robert C. Marchwinski, Daniel Margala, Sarah L. Martell, Marie Martig, Karen L. Masters, Savita Mathur, Cameron K. McBride, Peregrine M. McGehee, Ian D. McGreer, Richard G. McMahon, Brice Ménard, Marie-Luise Menzel, Andrea Merloni, Szabolcs Mészáros, Adam A. Miller, Jordi Miralda-Escudé, Hironao Miyatake, Antonio D. Montero-Dorta, Surhud More, Eric Morganson, Xan Morice-Atkinson, Heather L. Morrison, Benôit Mosser, Demitri Muna, Adam D. Myers, Kirpal Nandra, Jeffrey A. Newman, Mark Neyrinck, Duy Cuong Nguyen, Robert C. Nichol, David L. Nidever, Pasquier Noterdaeme, Sebastián E. Nuza, Julia E. O'Connell, Robert W. O'Connell, Ross O'Connell, Ricardo L. C. Ogando, Matthew D. Olmstead, Audrey E. Oravetz, Daniel J. Oravetz, Keisuke Osumi, Russell Owen, Deborah L. Padgett, Nikhil Padmanabhan, Martin Paegert, Nathalie Palanque-Delabrouille, Kaike Pan, John K. Parejko, Isabelle Pâris, Changbom Park, Petchara Pattarakijwanich, M. Pellejero-Ibanez, Joshua Pepper, Will J. Percival, Ismael Pérez-Fournon, Ignasi Pérez-Ràfols, Patrick Petitjean, Matthew M. Pieri, Marc H. Pinsonneault, Gustavo F. Porto de Mello, Francisco Prada, Abhishek Prakash, Adrian M. Price-Whelan, Pavlos Protopapas, M. Jordan Raddick, Mubdi Rahman, Beth A. Reid, James Rich, Hans-Walter Rix, Annie C. Robin, Constance M. Rockosi, Thaíse S. Rodrigues, Sergio Rodríguez-Torres, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, John J. Ruan, J. A. Rubiño-Martín, Eli S. Rykoff, Salvador Salazar-Albornoz, Mara Salvato, Lado Samushia, Ariel G. Sánchez, Basílio Santiago, Conor Sayres, Ricardo P. Schiavon, David J. Schlegel, Sarah J. Schmidt, Donald P. Schneider, Mathias Schultheis, Axel D. Schwope, C. G. Scóccola, Caroline Scott, Kris Sellgren, Hee-Jong Seo, Aldo Serenelli, Neville Shane, Yue Shen, Matthew Shetrone, Yiping Shu, V. Silva Aguirre, Thirupathi Sivarani, M. F. Skrutskie, Anže Slosar, Verne V. Smith, Flávia Sobreira, Diogo Souto, Keivan G. Stassun, Matthias Steinmetz, Dennis Stello, Michael A. Strauss, Alina Streblyanska, Nao Suzuki, Molly E. C. Swanson, Jonathan C. Tan, Jamie Tayar, Ryan C. Terrien, Aniruddha R. Thakar, Daniel Thomas, Neil Thomas, Benjamin A. Thompson, Jeremy L. Tinker, Rita Tojeiro, Nicholas W. Troup, Mariana Vargas-Magaña, Jose A. Vazquez, Licia Verde, Matteo Viel, Nicole P. Vogt, David A. Wake, Ji Wang, Benjamin A. Weaver, David H. Weinberg, Benjamin J. Weiner, Martin White, John C. Wilson, John P. Wisniewski, W. M. Wood-Vasey, Christophe Ye'che, Donald G. York, Nadia L. Zakamska, O. Zamora, Gail Zasowski, Idit Zehavi, Gong-Bo Zhao, Zheng Zheng, Xu Zhou, Zhimin Zhou, Hu Zou, and Guangtun Zhu. The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. ApJS, 219(1):12, July 2015.

[7] P. D. Aleo, K. Malanchev, S. Sharief, D. O. Jones, G. Narayan, R. J. Foley, V. A. Villar, C. R. Angus, V. F. Baldassare, M. J. Bustamante-Rosell, D. Chatterjee, C. Cold, D. A. Coulter, K. W. Davis, S. Dhawan, M. R. Drout, A. Engel, K. D. French, A. Gagliano, C. Gall, J. Hjorth, M. E. Huber, W. V. Jacobson-Galán, C. D. Kilpatrick, D. Langeroodi, P. Macias, K. S. Mandel, R. Margutti, F. Matašić, P. McGill, J. D. R. Pierel, E. Ramirez-Ruiz, C. L. Ransome, C. Rojas-Bravo, M. R. Siebert, K. W. Smith, K. M. de Soto, M. C. Stroh, S. Tinyanont, K. Taggart, S. M. Ward, R. Wojtak, K. Auchettl, P. K. Blanchard, T. J. L. de Boer, B. M. Boyd, C. M. Carroll, K. C. Chambers, L. DeMarchi, G. Dimitriadis, S. A. Dodd, N. Earl, D. Farias, H. Gao, S. Gomez, M. Grayling, C. Grillo, E. E. Hayes, T. Hung, L. Izzo, N. Khetan, A. N. Kolborg, J. A. P. Law-Smith, N. LeBaron, C. C. Lin, Y. Luo, E. A. Magnier, D. Matthews, B. Mockler, A. J. G. O'Grady, Y. C. Pan, C. A. Politsch, S. I. Raimundo, A. Rest, R. Ridden-Harper, A. Sarangi, S. L. Schrøder, S. J. Smartt, G. Terreran, S. Thorp, J. Vazquez, R. J. Wainscoat, Q. Wang, A. R. Wasserman, S. K. Yadavalli, R. Yarza, Y. Zenati,

and Young Supernova Experiment. The Young Supernova Experiment Data Release 1 (YSE DR1): Light Curves and Photometric Classification of 1975 Supernovae. ApJS, 266(1):9, May 2023.

[8] Catarina S. Alves, Hiranya V. Peiris, Michelle Lochner, Jason D. McEwen, Tarek Allam, Rahul Biswas, and LSST Dark Energy Science Collaboration. Considerations for Optimizing the Photometric Classification of Supernovae from the Rubin Observatory. ApJS, 258(2):23, February 2022.

[9] R. Amanullah, C. Lidman, D. Rubin, G. Aldering, P. Astier, K. Barbary, M. S. Burns, A. Conley, K. S. Dawson, S. E. Deustua, M. Doi, S. Fabbro, L. Faccioli, H. K. Fakhouri, G. Folatelli, A. S. Fruchter, H. Furusawa, G. Garavini, G. Goldhaber, A. Goobar, D. E. Groom, I. Hook, D. A. Howell, N. Kashikawa, A. G. Kim, R. A. Knop, M. Kowalski, E. Linder, J. Meyers, T. Morokuma, S. Nobili, J. Nordin, P. E. Nugent, L. Östman, R. Pain, N. Panagia, S. Perlmutter, J. Raux, P. Ruiz-Lapuente, A. L. Spadafora, M. Strovink, N. Suzuki, L. Wang, W. M. Wood-Vasey, N. Yasuda, and The Supernova Cosmology Project. Spectra and Hubble Space Telescope Light Curves of Six Type Ia Supernovae at $0.511 < z < 1.12$ and the Union2 Compilation. ApJ, 716(1):712–738, June 2010.

[10] AstroDave, Christopher Read @ Winton AstroTom, joycenv, and Kyle Willett. Galaxy zoo - the galaxy challenge, 2013.

[11] J. Audenaert, J. S. Kuszlewicz, R. Handberg, A. Tkachenko, D. J. Armstrong, M. Hon, R. Kgoadi, M. N. Lund, K. J. Bell, L. Bugnet, D. M. Bowman, C. Johnston, R. A. García, D. Stello, L. Molnár, E. Plachy, D. Buzasi, C. Aerts, and T'DA collaboration. TESS Data for Asteroseismology (T'DA) Stellar Variability Classification Pipeline: Setup and Application to the Kepler Q9 Data. AJ, 162(5):209, November 2021.

[12] J. Audenaert and A. Tkachenko. Multiscale entropy analysis of astronomical time series. Discovering subclusters of hybrid pulsators. A&A, 666:A76, October 2022.

[13] Micaela B. Bagley, Steven L. Finkelstein, Anton M. Koekemoer, Henry C. Ferguson, Pablo Arrabal Haro, Mark Dickinson, Jeyhan S. Kartaltepe, Casey Papovich, Pablo G. Pérez-González, Nor Pirzkal, Rachel S. Somerville, Christopher N. A. Willmer, Guang Yang, L. Y. Aaron Yung, Adriano Fontana, Andrea Grazian, Norman A. Grogin, Michaela Hirschmann, Lisa J. Kewley, Allison Kirkpatrick, Dale D. Kocevski, Jennifer M. Lotz, Aubrey Medrano, Alexa M. Morales, Laura Pentericci, Swara Ravindranath, Jonathan R. Trump, Stephen M. Wilkins, Antonello Calabrò, M. C. Cooper, Luca Costantin, Alexander de la Vega, Bryan Hilbert, Taylor A. Hutchison, Rebecca L. Larson, Ray A. Lucas, Elizabeth J. McGrath, Russell Ryan, Xin Wang, and Stijn Wuyts. CEERS Epoch 1 NIRCam Imaging: Reduction Methods and Simulations Enabling Early JWST Science Results. ApJ, 946(1):L12, March 2023.

[14] Micaela B. Bagley, Nor Pirzkal, Steven L. Finkelstein, Casey Papovich, Danielle A. Berg, Jennifer M. Lotz, Gene C. K. Leung, Henry C. Ferguson, Anton M. Koekemoer, Mark Dickinson, Jeyhan S. Kartaltepe, Dale D. Kocevski, Rachel S. Somerville, L. Y. Aaron Yung, Bren E. Backhaus, Caitlin M. Casey, Marco Castellano, Óscar A. Chávez Ortiz, Katherine Chworowsky, Isabella G. Cox, Romeel Davé, Kelcey Davis, Vicente Estrada-Carpenter, Adriano Fontana, Seiji Fujimoto, Jonathan P. Gardner, Mauro Giavalisco, Andrea Grazian, Norman A. Grogin, Nimish P. Hathi, Taylor A. Hutchison, Anne E. Jaskot, Intae Jung, Lisa J. Kewley, Allison Kirkpatrick, Rebecca L. Larson, Jasleen Matharu, Priyamvada Natarajan, Laura Pentericci, Pablo G. Pérez-González, Swara Ravindranath, Barry Rothberg, Russell Ryan, Lu Shen, Raymond C. Simons, Gregory F. Snyder, Jonathan R. Trump, and Stephen M. Wilkins. The Next Generation Deep Extragalactic Exploratory Public (NGDEEP) Survey. ApJ, 965(1):L6, April 2024.

[15] Eric C. Bellm, Shrinivas R. Kulkarni, Tom Barlow, Ulrich Feindt, Matthew J. Graham, Ariel Goobar, Thomas Kupfer, Chow-Choong Ngeow, Peter Nugent, Eran Ofek, Thomas A. Prince, Reed Riddle, Richard Walters, and Quan-Zhi Ye. The Zwicky Transient Facility: Surveys and Scheduler. PASP, 131(1000):068003, June 2019.

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

[17] M. Betoule, R. Kessler, J. Guy, J. Mosher, D. Hardin, R. Biswas, P. Astier, P. El-Hage, M. Konig, S. Kuhlmann, J. Marriner, R. Pain, N. Regnault, C. Balland, B. A. Bassett, P. J. Brown, H. Campbell, R. G. Carlberg, F. Cellier-Holzem, D. Cinabro, A. Conley, C. B. D'Andrea, D. L. DePoy, M. Doi, R. S. Ellis, S. Fabbro, A. V. Filippenko, R. J. Foley, J. A. Frieman, D. Fouchez, L. Galbany, A. Goobar, R. R. Gupta, G. J. Hill, R. Hlozek, C. J. Hogan, I. M. Hook, D. A. Howell, S. W. Jha, L. Le Guillou, G. Leloudas, C. Lidman, J. L. Marshall, A. Möller, A. M. Mourão, J. Neveu, R. Nichol, M. D. Olmstead, N. Palanque-Delabrouille, S. Perlmutter, J. L. Prieto, C. J. Pritchet, M. Richmond, A. G. Riess, V. Ruhlmann-Kleider, M. Sako, K. Schahmaneche, D. P. Schneider, M. Smith, J. Sollerman, M. Sullivan, N. A. Walton, and C. J. Wheeler. Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. A&A, 568:A22, August 2014.

[18] F. B. Bianco, M. Modjaz, M. Hicken, A. Friedman, R. P. Kirshner, J. S. Bloom, P. Challis, G. H. Marion, W. M. Wood-Vasey, and A. Rest. Multi-color Optical and Near-infrared Light Curves of 64 Stripped-envelope Core-Collapse Supernovae. ApJS, 213(2):19, August 2014.

[19] Michael R. Blanton, Matthew A. Bershady, Bela Abolfathi, Franco D. Albareti, Carlos Allende Prieto, Andres Almeida, Javier Alonso-García, Friedrich Anders, Scott F. Anderson, Brett Andrews, Erik Aquino-Ortíz, Alfonso Aragón-Salamanca, Maria Argudo-Fernández, Eric Armengaud, Eric Aubourg, Vladimir Avila-Reese, Carles Badenes, Stephen Bailey, Kathleen A. Barger, Jorge Barrera-Ballesteros, Curtis Bartosz, Dominic Bates, Falk Baumgarten, Julian Bautista, Rachael Beaton, Timothy C. Beers, Francesco Belfiore, Chad F. Bender, Andreas A. Berlind, Mariangela Bernardi, Florian Beutler, Jonathan C. Bird, Dmitry Bizyaev, Guillermo A. Blanc, Michael Blomqvist, Adam S. Bolton, Médéric Boquien, Jura Borissova, Remco van den Bosch, Jo Bovy, William N. Brandt, Jonathan Brinkmann, Joel R. Brownstein, Kevin Bundy, Adam J. Burgasser, Etienne Burtin, Nicolás G. Busca, Michele Cappellari, Maria Leticia Delgado Carigi, Joleen K. Carlberg, Aurelio Carnero Rosell, Ricardo Carrera, Nancy J. Chanover, Brian Cherinka, Edmond Cheung, Yilen Gómez Maqueo Chew, Cristina Chiappini, Peter Doohyun Choi, Drew Chojnowski, Chia-Hsun Chuang, Haeun Chung, Rafael Fernando Cirolini, Nicolas Clerc, Roger E. Cohen, Johan Comparat, Luiz da Costa, Marie-Claude Cousinou, Kevin Covey, Jeffrey D. Crane, Rupert A. C. Croft, Irene Cruz-Gonzalez, Daniel Garrido Cuadra, Katia Cunha, Guillermo J.

Damke, Jeremy Darling, Roger Davies, Kyle Dawson, Axel de la Macorra, Flavia Dell'Agli, Nathan De Lee, Timothée Delubac, Francesco Di Mille, Aleks Diamond-Stanic, Mariana Cano-Díaz, John Donor, Juan José Downes, Niv Drory, Hélion du Mas des Bourboux, Christopher J. Duckworth, Tom Dwelly, Jamie Dyer, Garrett Ebelke, Arthur D. Eigenbrot, Daniel J. Eisenstein, Eric Emsellem, Mike Eracleous, Stephanie Escoffier, Michael L. Evans, Xiaohui Fan, Emma Fernández-Alvar, J. G. Fernandez-Trincado, Diane K. Feuillet, Alexis Finoguenov, Scott W. Fleming, Andreu Font-Ribera, Alexander Fredrickson, Gordon Freischlad, Peter M. Frinchaboy, Carla E. Fuentes, Lluís Galbany, R. Garcia-Dias, D. A. García-Hernández, Patrick Gaulme, Doug Geisler, Joseph D. Gelfand, Héctor Gil-Marín, Bruce A. Gillespie, Daniel Goddard, Violeta Gonzalez-Perez, Kathleen Grabowski, Paul J. Green, Catherine J. Grier, James E. Gunn, Hong Guo, Julien Guy, Alex Hagen, ChangHoon Hahn, Matthew Hall, Paul Harding, Sten Hasselquist, Suzanne L. Hawley, Fred Hearty, Jonay I. Gonzalez Hernández, Shirley Ho, David W. Hogg, Kelly Holley-Bockelmann, Jon A. Holtzman, Parker H. Holzer, Joseph Huehnerhoff, Timothy A. Hutchinson, Ho Seong Hwang, Héctor J. Ibarra-Medel, Gabriele da Silva Ilha, Inese I. Ivans, KeShawn Ivory, Kelly Jackson, Trey W. Jensen, Jennifer A. Johnson, Amy Jones, Henrik Jönsson, Eric Jullo, Vikrant Kamble, Karen Kinemuchi, David Kirkby, Francisco-Shu Kitaura, Mark Klaene, Gillian R. Knapp, Jean-Paul Kneib, Juna A. Kollmeier, Ivan Lacerna, Richard R. Lane, Dustin Lang, David R. Law, Daniel Lazarz, Youngbae Lee, Jean-Marc Le Goff, Fu-Heng Liang, Cheng Li, Hongyu Li, Jianhui Lian, Marcos Lima, Lihwai Lin, Yen-Ting Lin, Sara Bertran de Lis, Chao Liu, Miguel Angel C. de Icaza Lizaola, Dan Long, Sara Lucatello, Britt Lundgren, Nicholas K. MacDonald, Alice Deconto Machado, Chelsea L. MacLeod, Suvrath Mahadevan, Marcio Antonio Geimba Maia, Roberto Maiolino, Steven R. Majewski, Elena Malanushenko, Viktor Malanushenko, Arturo Manchado, Shude Mao, Claudia Maraston, Rui Marques-Chaves, Thomas Masseron, Karen L. Masters, Cameron K. McBride, Richard M. McDermid, Brianne McGrath, Ian D. McGreer, Nicolás Medina Peña, Matthew Melendez, Andrea Merloni, Michael R. Merrifield, Szabolcs Meszaros, Andres Meza, Ivan Minchev, Dante Minniti, Takamitsu Miyaji, Surhud More, John Mulchaey, Francisco Müller-Sánchez, Demitri Muna, Ricardo R. Munoz, Adam D. Myers, Preethi Nair, Kirpal Nandra, Janaina Correa do Nascimento, Alenka Negrete, Melissa Ness, Jeffrey A. Newman, Robert C. Nichol, David L. Nidever, Christian Nitschelm, Pierros Ntelis, Julia E. O'Connell, Ryan J. Oelkers, Audrey Oravetz, Daniel Oravetz, Zach Pace, Nelson Padilla, Nathalie Palanque-Delabrouille, Pedro Alonso Palicio, Kaike Pan, John K. Parejko, Taniya Parikh, Isabelle Pâris, Changbom Park, Alim Y. Patten, Sebastien Peirani, Marcos Pellejero-Ibanez, Samantha Penny, Will J. Percival, Ismael Perez-Fournon, Patrick Petitjean, Matthew M. Pieri, Marc Pinsonneault, Alice Pisani, Radosław Poleski, Francisco Prada, Abhishek Prakash, Anna Bárbara de Andrade Queiroz, M. Jordan Raddick, Anand Raichoor, Sandro Barboza Rembold, Hannah Richstein, Rogemar A. Riffel, Rogério Riffel, Hans-Walter Rix, Annie C. Robin, Constance M. Rockosi, Sergio Rodríguez-Torres, A. Roman-Lopes, Carlos Román-Zúñiga, Margarita Rosado, Ashley J. Ross, Graziano Rossi, John Ruan, Rossana Ruggeri, Eli S. Rykoff, Salvador Salazar-Albornoz, Mara Salvato, Ariel G. Sánchez, D. S. Aguado, José R. Sánchez-Gallego, Felipe A. Santana, Basílio Xavier Santiago, Conor Sayres, Ricardo P. Schiavon, Jaderson da Silva Schimoia, Edward F. Schlafly, David J. Schlegel, Donald P. Schneider, Mathias Schultheis, William J. Schuster, Axel Schwope, Hee-Jong Seo, Zhengyi Shao, Shiyin Shen, Matthew Shetrone, Michael Shull, Joshua D. Simon, Danielle Skinner, M. F. Skrutskie, Anže Slosar, Verne V. Smith, Jennifer S. Sobeck, Flavia Sobreira, Garrett Somers, Diogo Souto, David V. Stark, Keivan Stassun, Fritz Stauffer, Matthias Steinmetz, Thaisa Storchi-Bergmann, Alina Streblyanska, Guy S. Stringfellow, Genaro Suárez, Jing Sun, Nao Suzuki, Laszlo Szigeti, Manuchehr Taghizadeh-Popp, Baitian Tang, Charling Tao, Jamie Tayar, Mita Tembe, Johanna Teske, Aniruddha R. Thakar, Daniel Thomas, Benjamin A. Thompson, Jeremy L. Tinker, Patricia Tissera, Rita Tojeiro, Hector Hernandez Toledo, Sylvain de la Torre, Christy Tremonti, Nicholas W. Troup, Octavio Valenzuela, Inma Martinez Valpuesta, Jaime Vargas-González, Mariana Vargas-Magaña, Jose Alberto Vazquez, Sandro Villanova, M. Vivek, Nicole Vogt, David Wake, Rene Walterbos, Yuting Wang, Benjamin Alan Weaver, Anne-Marie Weijmans, David H. Weinberg, Kyle B. Westfall, David G. Whelan, Vivienne Wild, John Wilson,

W. M. Wood-Vasey, Dominika Wylezalek, Ting Xiao, Renbin Yan, Meng Yang, Jason E. Ybarra, Christophe Yèche, Nadia Zakamska, Olga Zamora, Pauline Zarrouk, Gail Zasowski, Kai Zhang, Gong-Bo Zhao, Zheng Zheng, Zheng Zheng, Xu Zhou, Zhi-Min Zhou, Guangtun B. Zhu, Manuela Zoccali, and Hu Zou. Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. AJ, 154(1):28, July 2017.

[20] Kyle Boone. Avocado: Photometric classification of astronomical transients with gaussian process augmentation. *The Astronomical Journal*, 158(6):257, December 2019.

[21] William J. Borucki, David Koch, Gibor Basri, Natalie Batalha, Timothy Brown, Douglas Caldwell, John Caldwell, Jørgen Christensen-Dalsgaard, William D. Cochran, Edna DeVore, Edward W. Dunham, Andrea K. Dupree, Thomas N. Gautier, John C. Geary, Ronald Gilliland, Alan Gould, Steve B. Howell, Jon M. Jenkins, Yoji Kondo, David W. Latham, Geoffrey W. Marcy, Søren Meibom, Hans Kjeldsen, Jack J. Lissauer, David G. Monet, David Morrison, Dimitar Sasselov, Jill Tarter, Alan Boss, Don Brownlee, Toby Owen, Derek Buzasi, David Charbonneau, Laurance Doyle, Jonathan Fortney, Eric B. Ford, Matthew J. Holman, Sara Seager, Jason H. Steffen, William F. Welsh, Jason Rowe, Howard Anderson, Lars Buchhave, David Ciardi, Lucianne Walkowicz, William Sherry, Elliott Horch, Howard Isaacson, Mark E. Everett, Debra Fischer, Guillermo Torres, John Asher Johnson, Michael Endl, Phillip MacQueen, Stephen T. Bryson, Jessie Dotson, Michael Haas, Jeffrey Kolodziejczak, Jeffrey Van Cleve, Hema Chandrasekaran, Joseph D. Twicken, Elisa V. Quintana, Bruce D. Clarke, Christopher Allen, Jie Li, Haley Wu, Peter Tenenbaum, Ekaterina Verner, Frederick Bruhweiler, Jason Barnes, and Andrej Prsa. Kepler Planet-Detection Mission: Introduction and First Results. *Science*, 327(5968):977, February 2010.

[22] Micah Bowles, Hongming Tang, Eleni Vardoulaki, Emma L Alexander, Yan Luo, Lawrence Rudnick, Mike Walmsley, Fiona Porter, Anna M M Scaife, Inigo Val Slijepcevic, Elizabeth A K Adams, Alexander Drabent, Thomas Dugdale, Gülay Gürkan, Andrew M Hopkins, Eric F Jimenez-Andrade, Denis A Leahy, Ray P Norris, Syed Faisal ur Rahman, Xichang Ouyang, Gary Segal, Stanislav S Shabala, and O Ivy Wong. Radio galaxy zoo EMU: towards a semantic radio galaxy morphology taxonomy. *Monthly Notices of the Royal Astronomical Society*, 522(2):2584–2600, 04 2023.

[23] Micah Bowles, Hongming Tang, Eleni Vardoulaki, Emma L. Alexander, Yan Luo, Lawrence Rudnick, Mike Walmsley, Fiona Porter, Anna M. M. Scaife, Inigo Val Slijepcevic, and Gary Segal. A New Task: Deriving Semantic Class Targets for the Physical Sciences. *NeurIPS 2022: Machine Learning and the Physical Sciences Workshop*, page arXiv:2210.14760, October 2022.

[24] D. Brout, M. Sako, D. Scolnic, R. Kessler, C. B. D'Andrea, T. M. Davis, S. R. Hinton, A. G. Kim, J. Lasker, E. Macaulay, A. Möller, R. C. Nichol, M. Smith, M. Sullivan, R. C. Wolf, S. Allam, B. A. Bassett, P. Brown, F. J. Castander, M. Childress, R. J. Foley, L. Galbany, K. Herner, E. Kasai, M. March, E. Morganson, P. Nugent, Y. C. Pan, R. C. Thomas, B. E. Tucker, W. Wester, T. M. C. Abbott, J. Annis, S. Avila, E. Bertin, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, M. Crocce, C. E. Cunha, L. N. da Costa, C. Davis, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, T. F. Eifler, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, D. A. Goldstein, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, D. L. Hollowood, K. Honscheid, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, T. S. Li, M. Lima, J. L. Marshall, P. Martini, R. Miquel, B. Nord, A. A. Plazas, A. Roodman, E. S. Rykoff, E. Sanchez, V. Scarpine, R. Schindler, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, D. L. Tucker, A. R. Walker, B. Yanny, Y. Zhang, and DES COLLABORATION. First Cosmology Results Using Type Ia Supernovae from the Dark Energy Survey: Photometric Pipeline and Light-curve Data Release. ApJ, 874(1):106, March 2019.

[25] D. Brout, M. Sako, D. Scolnic, R. Kessler, C. B. D'Andrea, T. M. Davis, S. R. Hinton, A. G. Kim, J. Lasker, E. Macaulay, A. Möller, R. C. Nichol, M. Smith, M. Sullivan,

R. C. Wolf, S. Allam, B. A. Bassett, P. Brown, F. J. Castander, M. Childress, R. J. Foley, L. Galbany, K. Herner, E. Kasai, M. March, E. Morganson, P. Nugent, Y. C. Pan, R. C. Thomas, B. E. Tucker, W. Wester, T. M. C. Abbott, J. Annis, S. Avila, E. Bertin, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, M. Crocce, C. E. Cunha, L. N. da Costa, C. Davis, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, T. F. Eifler, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, D. A. Goldstein, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, D. L. Hollowood, K. Honscheid, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, T. S. Li, M. Lima, J. L. Marshall, P. Martini, R. Miquel, B. Nord, A. A. Plazas, A. Roodman, E. S. Rykoff, E. Sanchez, V. Scarpine, R. Schindler, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, D. L. Tucker, A. R. Walker, B. Yanny, Y. Zhang, and DES COLLABORATION. First Cosmology Results Using Type Ia Supernovae from the Dark Energy Survey: Photometric Pipeline and Light-curve Data Release. ApJ, 874(1):106, March 2019.

[26] Peter J. Brown, Alice A. Breeveld, Stephen Holland, Paul Kuin, and Tyler Pritchard. SOUSA: the Swift Optical/Ultraviolet Supernova Archive. Ap&SS, 354(1):89–96, November 2014.

[27] Peter J. Brown, Alice A. Breeveld, Stephen Holland, Paul Kuin, and Tyler Pritchard. SOUSA: the Swift Optical/Ultraviolet Supernova Archive. Ap&SS, 354(1):89–96, November 2014.

[28] Sven Buder, Sanjib Sharma, Janez Kos, Anish M. Amarsi, Thomas Nordlander, Karin Lind, Sarah L. Martell, Martin Asplund, Joss Bland-Hawthorn, Andrew R. Casey, Gayandhi M. de Silva, Valentina D'Orazi, Ken C. Freeman, Michael R. Hayden, Geraint F. Lewis, Jane Lin, Katharine J. Schlesinger, Jeffrey D. Simpson, Dennis Stello, Daniel B. Zucker, Tomaž Zwitter, Kevin L. Beeson, Tobias Buck, Luca Casagrande, Jake T. Clark, Klemen Čotar, Gary S. da Costa, Richard de Grijs, Diane Feuillet, Jonathan Horner, Prajwal R. Kafle, Shourya Khanna, Chiaki Kobayashi, Fan Liu, Benjamin T. Montet, Govind Nandakumar, David M. Nataf, Melissa K. Ness, Lorenzo Spina, Thor Tepper-García, Yuan-Sen Ting, Gregor Traven, Rok Vogrinčič, Robert A. Wittenmyer, Rosemary F. G. Wyse, Maruša Žerjal, and Galah Collaboration. The GALAH+ survey: Third data release. MNRAS, 506(1):150–201, September 2021.

[29] Kevin Bundy, Matthew A. Bershady, David R. Law, Renbin Yan, Niv Drory, Nicholas MacDonald, David A. Wake, Brian Cherinka, José R. Sánchez-Gallego, Anne-Marie Weijmans, Daniel Thomas, Christy Tremonti, Karen Masters, Lodovico Coccato, Aleksandar M. Diamond-Stanic, Alfonso Aragón-Salamanca, Vladimir Avila-Reese, Carles Badenes, Jésus Falcón-Barroso, Francesco Belfiore, Dmitry Bizyaev, Guillermo A. Blanc, Joss Bland-Hawthorn, Michael R. Blanton, Joel R. Brownstein, Nell Byler, Michele Cappellari, Charlie Conroy, Aaron A. Dutton, Eric Emsellem, James Etherington, Peter M. Frinchaboy, Hai Fu, James E. Gunn, Paul Harding, Evelyn J. Johnston, Guinevere Kauffmann, Karen Kinemuchi, Mark A. Klaene, Johan H. Knapen, Alexie Leauthaud, Cheng Li, Lihwai Lin, Roberto Maiolino, Viktor Malanushenko, Elena Malanushenko, Shude Mao, Claudia Maraston, Richard M. McDermid, Michael R. Merrifield, Robert C. Nichol, Daniel Oravetz, Kaike Pan, John K. Parejko, Sebastian F. Sanchez, David Schlegel, Audrey Simmons, Oliver Steele, Matthias Steinmetz, Karun Thanjavur, Benjamin A. Thompson, Jeremy L. Tinker, Remco C. E. van den Bosch, Kyle B. Westfall, David Wilkinson, Shelley Wright, Ting Xiao, and Kai Zhang. Overview of the SDSS-IV MaNGA Survey: Mapping nearby Galaxies at Apache Point Observatory. ApJ, 798(1):7, January 2015.

[30] U. F. Burhanudin, J. R. Maund, T. Killestein, K. Ackley, M. J. Dyer, J. Lyman, K. Ulaczyk, R. Cutter, Y. L. Mong, D. Steeghs, D. K. Galloway, V. Dhillon, P. O'Brien, G. Ramsay, K. Noysena, R. Kotak, R. P. Breton, L. Nuttall, E. Pallé, D. Pollacco, E. Thrane, S. Awiphan, P. Chote, A. Chrimes, E. Daw, C. Duffy, R. Eyles-Ferris, B. Gompertz, T. Heikkilä, P. Irawati, M. R. Kennedy, A. Levan, S. Littlefair, L. Makrygianni, D. Mata-Sánchez, S. Mattila, J. McCormac, D. Mkrtichian, J. Mullaney,

U. Sawangwit, E. Stanway, R. Starling, P. Strøm, S. Tooke, and K. Wiersema. Light-curve classification with recurrent neural networks for GOTO: dealing with imbalanced data. MNRAS, 505(3):4345–4361, August 2021.

[31] Umar F. Burhanudin and Justyn R. Maund. Pan-chromatic photometric classification of supernovae from multiple surveys and transfer learning for future surveys. MNRAS, 521(2):1601–1619, May 2023.

[32] Christopher R. Burns, Emilie Parent, M. M. Phillips, Maximilian Stritzinger, Kevin Krisciunas, Nicholas B. Suntzeff, Eric Y. Hsiao, Carlos Contreras, Jorge Anais, Luis Boldt, Luis Busta, Abdo Campillay, Sergio Castellón, Gastón Folatelli, Wendy L. Freedman, Consuelo González, Mario Hamuy, Peter Heoflich, Wojtek Krzeminski, Barry F. Madore, Nidia Morrell, S. E. Persson, Miguel Roth, Francisco Salgado, Jacqueline Serón, and Simón Torres. The Carnegie Supernova Project: Absolute Calibration and the Hubble Constant. ApJ, 869(1):56, December 2018.

[33] Douglas A. Caldwell, Peter Tenenbaum, Joseph D. Twicken, Jon M. Jenkins, Eric Ting, Jeffrey C. Smith, Christina Hedges, Michael M. Fausnaugh, Mark Rose, and Christopher Burke. TESS Science Processing Operations Center FFI Target List Products. *Research Notes of the American Astronomical Society*, 4(11):201, November 2020.

[34] Anthony Carr, Tamara M. Davis, Dan Scolnic, Khaled Said, Dillon Brout, Erik R. Peterson, and Richard Kessler. The pantheon+ analysis: Improving the redshifts and peculiar velocities of type ia supernovae used in cosmological analyses. *Publications of the Astronomical Society of Australia*, 39:e046, 2022.

[35] Brian Cherinka, Brett H. Andrews, José Sánchez-Gallego, Joel Brownstein, María Argudo-Fernández, Michael Blanton, Kevin Bundy, Amy Jones, Karen Masters, David R. Law, Kate Rowlands, Anne-Marie Weijmans, Kyle Westfall, and Renbin Yan. Marvin: A Tool Kit for Streamlined Access and Visualization of the SDSS-IV MaNGA Data Set. AJ, 158(2):74, August 2019.

[36] Carlos Contreras, Mario Hamuy, M. M. Phillips, Gastón Folatelli, Nicholas B. Suntzeff, S. E. Persson, Maximilian Stritzinger, Luis Boldt, Sergio González, Wojtek Krzeminski, Nidia Morrell, Miguel Roth, Francisco Salgado, María José Maureira, Christopher R. Burns, W. L. Freedman, Barry F. Madore, David Murphy, Pamela Wyatt, Weidong Li, and Alexei V. Filippenko. The Carnegie Supernova Project: First Photometry Data Release of Low-Redshift Type Ia Supernovae. AJ, 139(2):519–539, February 2010.

[37] Kaiming Cui, D. J. Armstrong, and Fabo Feng. Identifying Light-curve Signals with a Deep-learning-based Object Detection Algorithm. II. A General Light-curve Classification Framework. ApJS, 274(2):29, October 2024.

[38] F. De Angeli, M. Weiler, P. Montegriffo, D. W. Evans, M. Riello, R. Andrae, J. M. Carrasco, G. Busso, P. W. Burgess, C. Cacciari, M. Davidson, D. L. Harrison, S. T. Hodgkin, C. Jordi, P. J. Osborne, E. Pancino, G. Altavilla, M. A. Barstow, C. A. L. Bailer-Jones, M. Bellazzini, A. G. A. Brown, M. Castellani, S. Cowell, L. Delchambre, F. De Luise, C. Diener, C. Fabricius, M. Fouesneau, Y. Frémat, G. Gilmore, G. Giuffrida, N. C. Hambly, S. Hidalgo, G. Holland, Z. Kostrzewa-Rutkowska, F. van Leeuwen, A. Lobel, S. Marinoni, N. Miller, C. Pagani, L. Palaversa, A. M. Piersimoni, L. Pulone, S. Ragaini, M. Rainer, P. J. Richards, G. T. Rixon, D. Ruz-Mieres, N. Sanna, L. M. Sarro, N. Rowell, R. Sordo, N. A. Walton, and A. Yoldas. Gaiadata release 3: Processing and validation of bp/rp low-resolution spectral data. *Astronomy & Astrophysics*, 674:A2, June 2023.

[39] G. M. De Silva, K. C. Freeman, J. Bland-Hawthorn, S. Martell, E. Wylie de Boer, M. Asplund, S. Keller, S. Sharma, D. B. Zucker, T. Zwitter, B. Anguiano, C. Bacigalupo, D. Bayliss, M. A. Beavis, M. Bergemann, S. Campbell, R. Cannon, D. Carollo, L. Casagrande, A. R. Casey, G. Da Costa, V. D'Orazi, A. Dotter, L. Duong, A. Heger, M. J. Ireland, P. R. Kafle, J. Kos, J. Lattanzio, G. F. Lewis, J. Lin, K. Lind, U. Munari, D. M. Nataf, S. O'Toole, Q. Parker, W. Reid, K. J. Schlesinger, A. Sheinis, J. D.

Simpson, D. Stello, Y. S. Ting, G. Traven, F. Watson, R. Wittenmyer, D. Yong, and M. Žerjal. The GALAH survey: scientific motivation. MNRAS, 449(3):2604–2617, May 2015.

[40] DESI Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, G. Aldering, D. M. Alexander, R. Alfarsy, C. Allende Prieto, M. Alvarez, O. Alves, A. Anand, F. Andrade-Oliveira, E. Armengaud, J. Asorey, S. Avila, A. Aviles, S. Bailey, A. Balaguera-Antolínez, O. Ballester, C. Baltay, A. Bault, J. Bautista, J. Behera, S. F. Beltran, S. BenZvi, L. Beraldo e Silva, J. R. Bermejo-Climent, A. Berti, R. Besuner, F. Beutler, D. Bianchi, C. Blake, R. Blum, A. S. Bolton, S. Brieden, A. Brodzeller, D. Brooks, Z. Brown, E. Buckley-Geer, E. Burtin, L. Cabayol-Garcia, Z. Cai, R. Canning, L. Cardiel-Sas, A. Carnero Rosell, F. J. Castander, J. L. Cervantes-Cota, S. Chabanier, E. Chaussidon, J. Chaves-Montero, S. Chen, C. Chuang, T. Claybaugh, S. Cole, A. P. Cooper, A. Cuceu, T. M. Davis, K. Dawson, R. de Belsunce, R. de la Cruz, A. de la Macorra, A. de Mattia, R. Demina, U. Demirbozan, J. DeRose, A. Dey, B. Dey, G. Dhungana, J. Ding, Z. Ding, P. Doel, R. Doshi, K. Douglass, A. Edge, S. Eftekharzadeh, D. J. Eisenstein, A. Elliott, S. Escoffier, P. Fagrelius, X. Fan, K. Fanning, V. A. Fawcett, S. Ferraro, J. Ereza, B. Flaugher, A. Font-Ribera, D. Forero-Sánchez, J. E. Forero-Romero, C. S. Frenk, B. T. Gänsicke, L. Á. García, J. García-Bellido, C. Garcia-Quintero, L. H. Garrison, H. Gil-Marín, J. Golden-Marx, S. Gontcho A Gontcho, A. X. Gonzalez-Morales, V. Gonzalez-Perez, C. Gordon, O. Graur, D. Green, D. Gruen, J. Guy, B. Hadzhiyska, C. Hahn, J. J. Han, M. M. S Hanif, H. K. Herrera-Alcantar, K. Honscheid, J. Hou, C. Howlett, D. Huterer, V. Iršič, M. Ishak, A. Jacques, A. Jana, L. Jiang, J. Jimenez, Y. P. Jing, S. Joudaki, E. Jullo, S. Juneau, N. Kizhuprakkat, N. G. Karaçaylı, T. Karim, R. Kehoe, S. Kent, A. Khederlarian, S. Kim, D. Kirkby, T. Kisner, F. Kitaura, J. Kneib, S. E. Koposov, A. Kovács, A. Kremin, A. Krolewski, B. L'Huillier, A. Lambert, C. Lamman, T. W. Lan, M. Landriau, D. Lang, J. U. Lange, J. Lasker, L. Le Guillou, A. Leauthaud, M. E. Levi, T. S. Li, E. Linder, A. Lyons, C. Magneville, M. Manera, C. J. Manser, D. Margala, P. Martini, P. McDonald, G. E. Medina, L. Medina-Varela, A. Meisner, J. Mena-Fernández, J. Meneses-Rizo, M. Mezcua, R. Miquel, P. Montero-Camacho, J. Moon, S. Moore, J. Moustakas, E. Mueller, J. Mundet, A. Muñoz-Gutiérrez, A. D. Myers, S. Nadathur, L. Napolitano, R. Neveux, J. A. Newman, J. Nie, R. Nikutta, G. Niz, P. Norberg, H. E. Noriega, E. Paillas, N. Palanque-Delabrouille, A. Palmese, P. Zhiwei, D. Parkinson, S. Penmetsa, W. J. Percival, A. Pérez-Fernández, I. Pérez-Ràfols, M. Pieri, C. Poppett, A. Porredon, S. Pothier, F. Prada, R. Pucha, A. Raichoor, C. Ramírez-Pérez, S. Ramirez-Solano, M. Rashkovetskyi, C. Ravoux, A. Rocher, C. Rockosi, A. J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, C. G. Sabiu, K. Said, A. Saintonge, L. Samushia, E. Sanchez, C. Saulder, E. Schaan, E. F. Schlafly, D. Schlegel, D. Scholte, M. Schubnell, H. Seo, A. Shafieloo, R. Sharples, W. Sheu, J. Silber, F. Sinigaglia, M. Siudek, Z. Slepian, A. Smith, D. Sprayberry, L. Stephey, J. Suárez-Pérez, Z. Sun, T. Tan, G. Tarlé, R. Tojeiro, L. A. Ureña-López, R. Vaisakh, D. Valcin, F. Valdes, M. Valluri, M. Vargas-Magaña, A. Variu, L. Verde, M. Walther, B. Wang, M. S. Wang, B. A. Weaver, N. Weaverdyck, R. H. Wechsler, M. White, Y. Xie, J. Yang, C. Yèche, J. Yu, S. Yuan, H. Zhang, Z. Zhang, C. Zhao, Z. Zheng, R. Zhou, Z. Zhou, H. Zou, S. Zou, and Y. Zu. The Early Data Release of the Dark Energy Spectroscopic Instrument. *arXiv e-prints*, page arXiv:2306.06308, June 2023.

[41] DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, Charles Baltay, Lucas Beaufore, Chris Bebek, Timothy C. Beers, Eric F. Bell, José Luis Bernal, Robert Besuner, Florian Beutler, Chris Blake, Hannes Bleuler, Michael Blomqvist, Robert Blum, Adam S. Bolton, Cesar Briceno, David Brooks, Joel R. Brownstein, Elizabeth Buckley-Geer, Angela Burden, Etienne Burtin, Nicolas G. Busca, Robert N. Cahn, Yan-Chuan Cai, Laia Cardiel-Sas, Raymond G. Carlberg, Pierre-Henri Carton, Ricard Casas, Francisco J. Castander, Jorge L. Cervantes-Cota, Todd M. Claybaugh, Madeline Close, Carl T. Coker, Shaun Cole, Johan Comparat, Andrew P. Cooper, M. C. Cousinou, Martin Crocce, Jean-Gabriel Cuby, Daniel P. Cunningham, Tamara M. Davis, Kyle S. Dawson, Axel de la Macorra, Juan De Vicente, Timothée Delubac, Mark Derwent, Arjun Dey, Govinda Dhungana, Zhejie Ding, Peter Doel, Yutong T. Duan, Anne Ealet, Jerry Edelstein, Sarah Eftekharzadeh,

Daniel J. Eisenstein, Ann Elliott, Stéphanie Escoffier, Matthew Evatt, Parker Fagrelius, Xiaohui Fan, Kevin Fanning, Arya Farahi, Jay Farihi, Ginevra Favole, Yu Feng, Enrique Fernandez, Joseph R. Findlay, Douglas P. Finkbeiner, Michael J. Fitzpatrick, Brenna Flaugher, Samuel Flender, Andreu Font-Ribera, Jaime E. Forero-Romero, Pablo Fosalba, Carlos S. Frenk, Michele Fumagalli, Boris T. Gaensicke, Giuseppe Gallo, Juan Garcia-Bellido, Enrique Gaztanaga, Nicola Pietro Gentile Fusillo, Terry Gerard, Irena Gershkovich, Tommaso Giannantonio, Denis Gillet, Guillermo Gonzalez-de-Rivera, Violeta Gonzalez-Perez, Shelby Gott, Or Graur, Gaston Gutierrez, Julien Guy, Salman Habib, Henry Heetderks, Ian Heetderks, Katrin Heitmann, Wojciech A. Hellwing, David A. Herrera, Shirley Ho, Stephen Holland, Klaus Honscheid, Eric Huff, Timothy A. Hutchinson, Dragan Huterer, Ho Seong Hwang, Joseph Maria Illa Laguna, Yuzo Ishikawa, Dianna Jacobs, Niall Jeffrey, Patrick Jelinsky, Elise Jennings, Linhua Jiang, Jorge Jimenez, Jennifer Johnson, Richard Joyce, Eric Jullo, Stéphanie Juneau, Sami Kama, Armin Karcher, Sonia Karkar, Robert Kehoe, Noble Kennamer, Stephen Kent, Martin Kilbinger, Alex G. Kim, David Kirkby, Theodore Kisner, Ellie Kitanidis, Jean-Paul Kneib, Sergey Koposov, Eve Kovacs, Kazuya Koyama, Anthony Kremin, Richard Kron, Luzius Kronig, Andrea Kueter-Young, Cedric G. Lacey, Robin Lafever, Ofer Lahav, Andrew Lambert, Michael Lampton, Martin Landriau, Dustin Lang, Tod R. Lauer, Jean-Marc Le Goff, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Su-Jeong Lee, Daniela Leitner, Michael Lesser, Michael E. Levi, Benjamin L'Huillier, Baojiu Li, Ming Liang, Huan Lin, Eric Linder, Sarah R. Loebman, Zarija Lukić, Jun Ma, Niall MacCrann, Christophe Magneville, Laleh Makarem, Marc Manera, Christopher J. Manser, Robert Marshall, Paul Martini, Richard Massey, Thomas Matheson, Jeremy McCauley, Patrick McDonald, Ian D. McGreer, Aaron Meisner, Nigel Metcalfe, Timothy N. Miller, Ramon Miquel, John Moustakas, Adam Myers, Milind Naik, Jeffrey A. Newman, Robert C. Nichol, Andrina Nicola, Luiz Nicolati da Costa, Jundan Nie, Gustavo Niz, Peder Norberg, Brian Nord, Dara Norman, Peter Nugent, Thomas O'Brien, Minji Oh, Knut A. G. Olsen, Cristobal Padilla, Hamsa Padmanabhan, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Antonella Palmese, Daniel Pappalardo, Isabelle Pâris, Changbom Park, Anna Patej, John A. Peacock, Hiranya V. Peiris, Xiyan Peng, Will J. Percival, Sandrine Perruchot, Matthew M. Pieri, Richard Pogge, Jennifer E. Pollack, Claire Poppett, Francisco Prada, Abhishek Prakash, Ronald G. Probst, David Rabinowitz, Anand Raichoor, Chang Hee Ree, Alexandre Refregier, Xavier Regal, Beth Reid, Kevin Reil, Mehdi Rezaie, Constance M. Rockosi, Natalie Roe, Samuel Ronayette, Aaron Roodman, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Eduardo Rozo, Vanina Ruhlmann-Kleider, Eli S. Rykoff, Cristiano Sabiu, Lado Samushia, Eusebio Sanchez, Javier Sanchez, David J. Schlegel, Michael Schneider, Michael Schubnell, Aurélia Secroun, Uros Seljak, Hee-Jong Seo, Santiago Serrano, Arman Shafieloo, Huanyuan Shan, Ray Sharples, Michael J. Sholl, William V. Shourt, Joseph H. Silber, David R. Silva, Martin M. Sirk, Anze Slosar, Alex Smith, George F. Smoot, Debopam Som, Yong-Seon Song, David Sprayberry, Ryan Staten, Andy Stefanik, Gregory Tarle, Suk Sien Tie, Jeremy L. Tinker, Rita Tojeiro, Francisco Valdes, Octavio Valenzuela, Monica Valluri, Mariana Vargas-Magana, Licia Verde, Alistair R. Walker, Jiali Wang, Yuting Wang, Benjamin A. Weaver, Curtis Weaverdyck, Risa H. Wechsler, David H. Weinberg, Martin White, Qian Yang, Christophe Yeche, Tianmeng Zhang, Gong-Bo Zhao, Yi Zheng, Xu Zhou, Zhimin Zhou, Yaling Zhu, Hu Zou, and Ying Zu. The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, page arXiv:1611.00036, October 2016.

[42] Arjun Dey, David J Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, et al. Overview of the desi legacy imaging surveys. *The Astronomical Journal*, 157(5):168, 2019.

[43] Arjun Dey, David J. Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R. Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, Martin Landriau, Michael Levi, Ian McGreer, Aaron Meisner, Adam D. Myers, John Moustakas, Peter Nugent, Anna Patej, Edward F. Schlafly, Alistair R. Walker, Francisco Valdes, Benjamin A. Weaver, Christophe Yèche, Hu Zou, Xu Zhou, Behzad Abareshi, T. M. C.

Abbott, Bela Abolfathi, C. Aguilera, Shadab Alam, Lori Allen, A. Alvarez, James Annis, Behzad Ansarinejad, Marie Aubert, Jacqueline Beechert, Eric F. Bell, Segev Y. BenZvi, Florian Beutler, Richard M. Bielby, Adam S. Bolton, César Briceño, Elizabeth J. Buckley-Geer, Karen Butler, Annalisa Calamida, Raymond G. Carlberg, Paul Carter, Ricard Casas, Francisco J. Castander, Yumi Choi, Johan Comparat, Elena Cukanovaite, Timothée Delubac, Kaitlin DeVries, Sharmila Dey, Govinda Dhungana, Mark Dickinson, Zhejie Ding, John B. Donaldson, Yutong Duan, Christopher J. Duckworth, Sarah Eftekharzadeh, Daniel J. Eisenstein, Thomas Etourneau, Parker A. Fagrelius, Jay Farihi, Mike Fitzpatrick, Andreu Font-Ribera, Leah Fulmer, Boris T. Gänsicke, Enrique Gaztanaga, Koshy George, David W. Gerdes, Satya Gontcho A. Gontcho, Claudio Gorgoni, Gregory Green, Julien Guy, Diane Harmer, M. Hernandez, Klaus Honscheid, Lijuan Wendy Huang, David J. James, Buell T. Jannuzi, Linhua Jiang, Richard Joyce, Armin Karcher, Sonia Karkar, Robert Kehoe, Jean-Paul Kneib, Andrea Kueter-Young, Ting-Wen Lan, Tod R. Lauer, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Michael Lesser, Laurence Perreault Levasseur, Ting S. Li, Justin L. Mann, Robert Marshall, C. E. Martínez-Vázquez, Paul Martini, Hélion du Mas des Bourboux, Sean McManus, Tobias Gabriel Meier, Brice Ménard, Nigel Metcalfe, Andrea Muñoz-Gutiérrez, Joan Najita, Kevin Napier, Gautham Narayan, Jeffrey A. Newman, Jundan Nie, Brian Nord, Dara J. Norman, Knut A. G. Olsen, Anthony Paat, Nathalie Palanque-Delabrouille, Xiyan Peng, Claire L. Poppett, Megan R. Poremba, Abhishek Prakash, David Rabinowitz, Anand Raichoor, Mehdi Rezaie, A. N. Robertson, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Gregory Rudnick, Sasha Safonova, Abhijit Saha, F. Javier Sánchez, Elodie Savary, Heidi Schweiker, Adam Scott, Hee-Jong Seo, Huanyuan Shan, David R. Silva, Zachary Slepian, Christian Soto, David Sprayberry, Ryan Staten, Coley M. Stillman, Robert J. Stupak, David L. Summers, Suk Sien Tie, H. Tirado, Mariana Vargas-Magaña, A. Katherina Vivas, Risa H. Wechsler, Doug Williams, Jinyi Yang, Qian Yang, Tolga Yapici, Dennis Zaritsky, A. Zenteno, Kai Zhang, Tianmeng Zhang, Rongpu Zhou, and Zhimin Zhou. Overview of the DESI Legacy Imaging Surveys. AJ, 157(5):168, May 2019.

[44] Arjun Dey, David J. Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R. Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, Martin Landriau, Michael Levi, Ian McGreer, Aaron Meisner, Adam D. Myers, John Moustakas, Peter Nugent, Anna Patej, Edward F. Schlafly, Alistair R. Walker, Francisco Valdes, Benjamin A. Weaver, Christophe Yèche, Hu Zou, Xu Zhou, Behzad Abareshi, T. M. C. Abbott, Bela Abolfathi, C. Aguilera, Shadab Alam, Lori Allen, A. Alvarez, James Annis, Behzad Ansarinejad, Marie Aubert, Jacqueline Beechert, Eric F. Bell, Segev Y. BenZvi, Florian Beutler, Richard M. Bielby, Adam S. Bolton, César Briceño, Elizabeth J. Buckley-Geer, Karen Butler, Annalisa Calamida, Raymond G. Carlberg, Paul Carter, Ricard Casas, Francisco J. Castander, Yumi Choi, Johan Comparat, Elena Cukanovaite, Timothée Delubac, Kaitlin DeVries, Sharmila Dey, Govinda Dhungana, Mark Dickinson, Zhejie Ding, John B. Donaldson, Yutong Duan, Christopher J. Duckworth, Sarah Eftekharzadeh, Daniel J. Eisenstein, Thomas Etourneau, Parker A. Fagrelius, Jay Farihi, Mike Fitzpatrick, Andreu Font-Ribera, Leah Fulmer, Boris T. Gänsicke, Enrique Gaztanaga, Koshy George, David W. Gerdes, Satya Gontcho A. Gontcho, Claudio Gorgoni, Gregory Green, Julien Guy, Diane Harmer, M. Hernandez, Klaus Honscheid, Lijuan Wendy Huang, David J. James, Buell T. Jannuzi, Linhua Jiang, Richard Joyce, Armin Karcher, Sonia Karkar, Robert Kehoe, Jean-Paul Kneib, Andrea Kueter-Young, Ting-Wen Lan, Tod R. Lauer, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Michael Lesser, Laurence Perreault Levasseur, Ting S. Li, Justin L. Mann, Robert Marshall, C. E. Martínez-Vázquez, Paul Martini, Hélion du Mas des Bourboux, Sean McManus, Tobias Gabriel Meier, Brice Ménard, Nigel Metcalfe, Andrea Muñoz-Gutiérrez, Joan Najita, Kevin Napier, Gautham Narayan, Jeffrey A. Newman, Jundan Nie, Brian Nord, Dara J. Norman, Knut A. G. Olsen, Anthony Paat, Nathalie Palanque-Delabrouille, Xiyan Peng, Claire L. Poppett, Megan R. Poremba, Abhishek Prakash, David Rabinowitz, Anand Raichoor, Mehdi Rezaie, A. N. Robertson, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Gregory Rudnick, Sasha Safonova, Abhijit Saha, F. Javier Sánchez, Elodie Savary, Heidi Schweiker, Adam Scott, Hee-Jong Seo, Huanyuan Shan, David R. Silva, Zachary Slepian, Christian Soto, David

Sprayberry, Ryan Staten, Coley M. Stillman, Robert J. Stupak, David L. Summers, Suk Sien Tie, H. Tirado, Mariana Vargas-Magaña, A. Katherina Vivas, Risa H. Wechsler, Doug Williams, Jinyi Yang, Qian Yang, Tolga Yapici, Dennis Zaritsky, A. Zenteno, Kai Zhang, Tianmeng Zhang, Rongpu Zhou, and Zhimin Zhou. Overview of the DESI Legacy Imaging Surveys. AJ, 157(5):168, May 2019.

[45] Sander Dieleman, Kyle W. Willett, and Joni Dambre. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly Notices of the Royal Astronomical Society*, 450(2):1441–1459, 04 2015.

[46] S. Dobryakov, K. Malanchev, D. Derkach, and M. Hushchyn. Photometric data-driven classification of Type Ia supernovae in the open Supernova Catalog. *Astronomy and Computing*, 35:100451, April 2021.

[47] H. Domínguez Sánchez, M. Huertas-Company, M. Bernardi, S. Kaviraj, J. L. Fischer, T. M. C. Abbott, F. B. Abdalla, J. Annis, S. Avila, D. Brooks, E. Buckley-Geer, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, C. Davis, J. De Vicente, P. Doel, A. E. Evrard, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, D. L. Hollowood, K. Honscheid, B. Hoyle, D. J. James, K. Kuehn, N. Kuropatkin, O. Lahav, M. A. G. Maia, M. March, P. Melchior, F. Menanteau, R. Miquel, B. Nord, A. A. Plazas, E. Sanchez, V. Scarpine, R. Schindler, M. Schubnell, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, A. R. Walker, and J. Zuntz. Transfer learning for galaxy morphology from one survey to another. MNRAS, 484(1):93–100, March 2019.

[48] James S. Dunlop, Roberto G. Abraham, Matthew L. N. Ashby, Micaela Bagley, Philip N. Best, Angela Bongiorno, Rychard Bouwens, Rebecca A. A. Bowler, Gabriel Brammer, Malcolm Bremer, Antonello Calabro', Adam Carnall, Marco Castellano, Michele Cirasuolo, Christopher Conselice, Fergus Cullen, Romeel Dave, Pratika Dayal, Avishai Dekel, Mark Dickinson, Kenneth James Duncan, David Elbaz, Richard S. Ellis, Harry C. Ferguson, Andrea Ferrara, Steven L. Finkelstein, Adriano Fontana, Steven Furlanetto, Johan P. U. Fynbo, Simona Gallerani, Jonathan P. Gardner, Mauro Giavalisco, Andrea Grazian, Norman Grogin, Yuichi Harikane, Philip F. Hopkins, Olivier Ilbert, Garth D. Illingworth, Stephanie Juneau, Intae Jung, Jeyhan Kartaltepe, Susan Kassin, Olivier Benjamin Kauffmann, Sadegh Khochfar, Allison Kirkpatrick, Dale D. Kocevski, Anton M. Koekemoer, Ivo Labbe, Nicolas Laporte, Rebecca L. Larson, Ray A. Lucas, Daniel K. Magee, Charlotte Mason, Henry Joy McCracken, Derek McLeod, Ross McLure, Emiliano Merlin, Andrei Mesinger, Bo Milvang-Jensen, Jeffrey Allen Newman, Pascal Oesch, Masami Ouchi, Camilla Pacifici, Casey Papovich, John Peacock, Molly Peeples, Laura Pentericci, Pablo G. Perez-Gonzalez, Norbert Pirzkal, Alexandra Pope, John P. Pye, Naveen A. Reddy, Brant Robertson, Mara Salvato, Paola Santini, Daniel Schaerer, Alice E. Shapley, Raymond Simons, Renske Smit, Britton D. Smith, Greg Snyder, Rachel S. Somerville, Elizabeth R. Stanway, Mauro Stefanon, Lidia Tasca, Tuomo Tikkanen, Laurence Tresse, Jonathan R. Trump, Katherine E. Whitaker, Stephen Matthew Wilkins, Gillian Wright, J. Stuart B. Wyithe, Pieter van Dokkum, and Paul van der Werf. PRIMER: Public Release IMaging for Extragalactic Research. JWST Proposal. Cycle 1, ID. #1837, March 2021.

[49] Daniel J. Eisenstein, David H. Weinberg, Eric Agol, Hiroaki Aihara, Carlos Allende Prieto, Scott F. Anderson, James A. Arns, Éric Aubourg, Stephen Bailey, Eduardo Balbinot, Robert Barkhouser, Timothy C. Beers, Andreas A. Berlind, Steven J. Bickerton, Dmitry Bizyaev, Michael R. Blanton, John J. Bochanski, Adam S. Bolton, Casey T. Bosman, Jo Bovy, W. N. Brandt, Ben Breslauer, Howard J. Brewington, J. Brinkmann, Peter J. Brown, Joel R. Brownstein, Dan Burger, Nicolas G. Busca, Heather Campbell, Phillip A. Cargile, William C. Carithers, Joleen K. Carlberg, Michael A. Carr, Liang Chang, Yanmei Chen, Cristina Chiappini, Johan Comparat, Natalia Connolly, Marina Cortes, Rupert A. C. Croft, Katia Cunha, Luiz N. da Costa, James R. A. Davenport, Kyle Dawson, Nathan De Lee, Gustavo F. Porto de Mello, Fernando de Simoni, Janice Dean, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Edward M.

Edmondson, Jacob M. Eiting, Stephanie Escoffier, Massimiliano Esposito, Michael L. Evans, Xiaohui Fan, Bruno Femenía Castellá, Leticia Dutra Ferreira, Greg Fitzgerald, Scott W. Fleming, Andreu Font-Ribera, Eric B. Ford, Peter M. Frinchaboy, Ana Elia García Pérez, B. Scott Gaudi, Jian Ge, Luan Ghezzi, Bruce A. Gillespie, G. Gilmore, Léo Girardi, J. Richard Gott, Andrew Gould, Eva K. Grebel, James E. Gunn, Jean-Christophe Hamilton, Paul Harding, David W. Harris, Suzanne L. Hawley, Frederick R. Hearty, Joseph F. Hennawi, Jonay I. González Hernández, Shirley Ho, David W. Hogg, Jon A. Holtzman, Klaus Honscheid, Naohisa Inada, Inese I. Ivans, Linhua Jiang, Peng Jiang, Jennifer A. Johnson, Cathy Jordan, Wendell P. Jordan, Guinevere Kauffmann, Eyal Kazin, David Kirkby, Mark A. Klaene, G. R. Knapp, Jean-Paul Kneib, C. S. Kochanek, Lars Koesterke, Juna A. Kollmeier, Richard G. Kron, Hubert Lampeitl, Dustin Lang, James E. Lawler, Jean-Marc Le Goff, Brian L. Lee, Young Sun Lee, Jarron M. Leisenring, Yen-Ting Lin, Jian Liu, Daniel C. Long, Craig P. Loomis, Sara Lucatello, Britt Lundgren, Robert H. Lupton, Bo Ma, Zhibo Ma, Nicholas MacDonald, Claude Mack, Suvrath Mahadevan, Marcio A. G. Maia, Steven R. Majewski, Martin Makler, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Claudia Maraston, Daniel Margala, Paul Maseman, Karen L. Masters, Cameron K. McBride, Patrick McDonald, Ian D. McGreer, Richard G. McMahon, Olga Mena Requejo, Brice Ménard, Jordi Miralda-Escudé, Heather L. Morrison, Fergal Mullally, Demitri Muna, Hitoshi Murayama, Adam D. Myers, Tracy Naugle, Angelo Fausti Neto, Duy Cuong Nguyen, Robert C. Nichol, David L. Nidever, Robert W. O'Connell, Ricardo L. C. Ogando, Matthew D. Olmstead, Daniel J. Oravetz, Nikhil Padmanabhan, Martin Paegert, Nathalie Palanque-Delabrouille, Kaike Pan, Parul Pandey, John K. Parejko, Isabelle Pâris, Paulo Pellegrini, Joshua Pepper, Will J. Percival, Patrick Petitjean, Robert Pfaffenberger, Janine Pforr, Stefanie Phleps, Christophe Pichon, Matthew M. Pieri, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Beatriz H. F. Ramos, I. Neill Reid, Celine Reyle, James Rich, Gordon T. Richards, George H. Rieke, Marcia J. Rieke, Hans-Walter Rix, Annie C. Robin, Helio J. Rocha-Pinto, Constance M. Rockosi, Natalie A. Roe, Emmanuel Rollinde, Ashley J. Ross, Nicholas P. Ross, Bruno Rossetto, Ariel G. Sánchez, Basilio Santiago, Conor Sayres, Ricardo Schiavon, David J. Schlegel, Katharine J. Schlesinger, Sarah J. Schmidt, Donald P. Schneider, Kris Sellgren, Alaina Shelden, Erin Sheldon, Matthew Shetrone, Yiping Shu, John D. Silverman, Jennifer Simmerer, Audrey E. Simmons, Thirupathi Sivarani, M. F. Skrutskie, Anže Slosar, Stephen Smee, Verne V. Smith, Stephanie A. Snedden, Keivan G. Stassun, Oliver Steele, Matthias Steinmetz, Mark H. Stockett, Todd Stollberg, Michael A. Strauss, Alexander S. Szalay, Masayuki Tanaka, Aniruddha R. Thakar, Daniel Thomas, Jeremy L. Tinker, Benjamin M. Tofflemire, Rita Tojeiro, Christy A. Tremonti, Mariana Vargas Magaña, Licia Verde, Nicole P. Vogt, David A. Wake, Xiaoke Wan, Ji Wang, Benjamin A. Weaver, Martin White, Simon D. M. White, John C. Wilson, John P. Wisniewski, W. Michael Wood-Vasey, Brian Yanny, Naoki Yasuda, Christophe Yèche, Donald G. York, Erick Young, Gail Zasowski, Idit Zehavi, and Bo Zhao. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. AJ, 142(3):72, September 2011.

[50] Daniel J. Eisenstein, Chris Willott, Stacey Alberts, Santiago Arribas, Nina Bonaventura, Andrew J. Bunker, Alex J. Cameron, Stefano Carniani, Stephane Charlot, Emma Curtis-Lake, Francesco D'Eugenio, Ryan Endsley, Pierre Ferruit, Giovanna Giardino, Kevin Hainline, Ryan Hausen, Peter Jakobsen, Benjamin D. Johnson, Roberto Maiolino, Marcia Rieke, George Rieke, Hans-Walter Rix, Brant Robertson, Daniel P. Stark, Sandro Tacchella, Christina C. Williams, Christopher N. A. Willmer, William M. Baker, Stefi Baum, Rachana Bhatawdekar, Kristan Boyett, Zuyi Chen, Jacopo Chevallard, Chiara Circosta, Mirko Curti, A. Lola Danhaive, Christa DeCoursey, Anna de Graaff, Alan Dressler, Eiichi Egami, Jakob M. Helton, Raphael E. Hviding, Zhiyuan Ji, Gareth C. Jones, Nimisha Kumari, Nora Lützgendorf, Isaac Laseter, Tobias J. Looser, Jianwei Lyu, Michael V. Maseda, Erica Nelson, Eleonora Parlanti, Michele Perna, Dávid Puskás, Tim Rawle, Bruno Rodríguez Del Pino, Lester Sandles, Aayush Saxena, Jan Scholtz, Katherine Sharpe, Irene Shivaei, Maddie S. Silcock, Charlotte Simmonds, Maya Skarbinski, Renske Smit, Meredith Stone, Katherine A. Suess, Fengwu Sun, Mengtao Tang, Michael W. Topping, Hannah Übler, Natalia C. Villanueva, Imaan E. B.

Wallace, Lily Whitler, Joris Witstok, and Charity Woodrum. Overview of the JWST Advanced Deep Extragalactic Survey (JADES). *arXiv e-prints*, page arXiv:2306.02465, June 2023.

[51] Ian N. Evans, Francis A. Primini, Kenny J. Glotfelty, Craig S. Anderson, Nina R. Bonaventura, Judy C. Chen, John E. Davis, Stephen M. Doe, Janet D. Evans, Giuseppina Fabbiano, Elizabeth C. Galle, II Gibbs, Danny G., John D. Grier, Roger M. Hain, Diane M. Hall, Peter N. Harbo, Xiangqun Helen He, John C. Houck, Margarita Karovska, Vinay L. Kashyap, Jennifer Lauer, Michael L. McCollough, Jonathan C. McDowell, Joseph B. Miller, Arik W. Mitschang, Douglas L. Morgan, Amy E. Mossman, Joy S. Nichols, Michael A. Nowak, David A. Plummer, Brian L. Refsdal, Arnold H. Rots, Aneta Siemiginowska, Beth A. Sundheim, Michael S. Tibbetts, David W. Van Stone, Sherry L. Winkelman, and Panagoula Zografou. The Chandra Source Catalog. ApJS, 189(1):37–82, July 2010.

[52] Steven L. Finkelstein, Micaela B. Bagley, Henry C. Ferguson, Stephen M. Wilkins, Jeyhan S. Kartaltepe, Casey Papovich, L. Y. Aaron Yung, Pablo Arrabal Haro, Peter Behroozi, Mark Dickinson, Dale D. Kocevski, Anton M. Koekemoer, Rebecca L. Larson, Aurélien Le Bail, Alexa M. Morales, Pablo G. Pérez-González, Denis Burgarella, Romeel Davé, Michaela Hirschmann, Rachel S. Somerville, Stijn Wuyts, Volker Bromm, Caitlin M. Casey, Adriano Fontana, Seiji Fujimoto, Jonathan P. Gardner, Mauro Giavalisco, Andrea Grazian, Norman A. Grogin, Nimish P. Hathi, Taylor A. Hutchison, Saurabh W. Jha, Shardha Jogee, Lisa J. Kewley, Allison Kirkpatrick, Arianna S. Long, Jennifer M. Lotz, Laura Pentericci, Justin D. R. Pierel, Nor Pirzkal, Swara Ravindranath, Russell E. Ryan, Jonathan R. Trump, Guang Yang, Rachana Bhatawdekar, Laura Bisigello, Véronique Buat, Antonello Calabrò, Marco Castellano, Nikko J. Cleri, M. C. Cooper, Darren Croton, Emanuele Daddi, Avishai Dekel, David Elbaz, Maximilien Franco, Eric Gawiser, Benne W. Holwerda, Marc Huertas-Company, Anne E. Jaskot, Gene C. K. Leung, Ray A. Lucas, Bahram Mobasher, Viraj Pandya, Sandro Tacchella, Benjamin J. Weiner, and Jorge A. Zavala. CEERS Key Paper. I. An Early Look into the First 500 Myr of Galaxy Formation with JWST. ApJ, 946(1):L13, March 2023.

[53] Ryan J. Foley, Daniel Scolnic, Armin Rest, S. W. Jha, Y. C. Pan, A. G. Riess, P. Challis, K. C. Chambers, D. A. Coulter, K. G. Dettman, M. M. Foley, O. D. Fox, M. E. Huber, D. O. Jones, C. D. Kilpatrick, R. P. Kirshner, A. S. B. Schultz, M. R. Siebert, H. A. Flewelling, B. Gibson, E. A. Magnier, J. A. Miller, N. Primak, S. J. Smartt, K. W. Smith, R. J. Wainscoat, C. Waters, and M. Willman. The Foundation Supernova Survey: motivation, design, implementation, and first data release. MNRAS, 475(1):193–219, March 2018.

[54] A. Francis and M. Czerkawski. Major TOM: Expandable Datasets for Earth Observation. *ArXiv e-prints*, page arXiv:2402.12095, 2024.

[55] Wendy L. Freedman, Barry F. Madore, Dylan Hatt, Taylor J. Hoyt, In Sung Jang, Rachael L. Beaton, Christopher R. Burns, Myung Gyoon Lee, Andrew J. Monson, Jillian R. Neeley, M. M. Phillips, Jeffrey A. Rich, and Mark Seibert. The Carnegie-Chicago Hubble Program. VIII. An Independent Determination of the Hubble Constant Based on the Tip of the Red Giant Branch. ApJ, 882(1):34, September 2019.

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

[57] Gaia Collaboration, R. Drimmel, M. Romero-Gómez, L. Chemin, P. Ramos, E. Poggio, V. Ripepi, R. Andrae, R. Blomme, T. Cantat-Gaudin, A. Castro-Ginard, G. Clementini,

F. Figueras, M. Fouesneau, Y. Frémat, K. Jardine, S. Khanna, A. Lobel, D. J. Marshall, T. Muraveva, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, F. Arenou, C. Babusiaux, M. Biermann, O. L. Creevey, C. Ducourant, D. W. Evans, L. Eyer, R. Guerra, A. Hutton, C. Jordi, S. A. Klioner, U. L. Lammers, L. Lindegren, X. Luri, F. Mignard, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, C. Soubiran, P. Tanga, N. A. Walton, C. A. L. Bailer-Jones, U. Bastian, F. Jansen, D. Katz, M. G. Lattanzi, F. van Leeuwen, J. Bakker, C. Cacciari, J. Castañeda, F. De Angeli, C. Fabricius, L. Galluccio, A. Guerrier, U. Heiter, E. Masana, R. Messineo, N. Mowlavi, C. Nicolas, K. Nienartowicz, F. Pailler, P. Panuzzo, F. Riclet, W. Roux, G. M. Seabroke, R. Sordo, F. Thévenin, G. Gracia-Abril, J. Portell, D. Teyssier, M. Altmann, M. Audard, I. Bellas-Velidis, K. Benson, J. Berthier, P. W. Burgess, D. Busonero, G. Busso, H. Cánovas, B. Carry, A. Cellino, N. Cheek, Y. Damerdji, M. Davidson, P. de Teodoro, M. Nuñez Campos, L. Delchambre, A. Dell'Oro, P. Esquej, J. Fernández-Hernández, E. Fraile, D. Garabato, P. García-Lario, E. Gosset, R. Haigron, J. L. Halbwachs, N. C. Hambly, D. L. Harrison, J. Hernández, D. Hestroffer, S. T. Hodgkin, B. Holl, K. Janßen, G. Jevardat de Fombelle, S. Jordan, A. Krone-Martins, A. C. Lanzafame, W. Löffler, O. Marchal, P. M. Marrese, A. Moitinho, K. Muinonen, P. Osborne, E. Pancino, T. Pauwels, A. Recio-Blanco, C. Reylé, M. Riello, L. Rimoldini, T. Roegiers, J. Rybizki, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, E. Utrilla, M. van Leeuwen, U. Abbas, P. Ábrahám, A. Abreu Aramburu, C. Aerts, J. J. Aguado, M. Ajaj, F. Aldea-Montero, G. Altavilla, M. A. Álvarez, J. Alves, F. Anders, R. I. Anderson, E. Anglada Varela, T. Antoja, D. Baines, S. G. Baker, L. Balaguer-Núñez, E. Balbinot, Z. Balog, C. Barache, D. Barbato, M. Barros, M. A. Barstow, S. Bartolomé, J. L. Bassilana, N. Bauchet, U. Becciani, M. Bellazzini, A. Berihuete, M. Bernet, S. Bertone, L. Bianchi, A. Binnenfeld, S. Blanco-Cuaresma, T. Boch, A. Bombrun, D. Bossini, S. Bouquillon, A. Bragaglia, L. Bramante, E. Breedt, A. Bressan, N. Brouillet, E. Brugaletta, B. Bucciarelli, A. Burlacu, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, R. Carballo, T. Carlucci, M. I. Carnerero, J. M. Carrasco, L. Casamiquela, M. Castellani, L. Chaoul, P. Charlot, V. Chiaramida, A. Chiavassa, N. Chornay, G. Comoretto, G. Contursi, W. J. Cooper, T. Cornez, S. Cowell, F. Crifo, M. Cropper, M. Crosta, C. Crowley, C. Dafonte, A. Dapergolas, P. David, P. de Laverny, F. De Luise, R. De March, J. De Ridder, R. de Souza, A. de Torres, E. F. del Peloso, E. del Pozo, M. Delbo, A. Delgado, J. B. Delisle, C. Demouchy, T. E. Dharmawardena, P. Di Matteo, S. Diakite, C. Diener, E. Distefano, C. Dolding, H. Enke, C. Fabre, M. Fabrizio, S. Faigler, G. Fedorets, P. Fernique, Y. Fournier, C. Fouron, F. Fragkoudi, M. Gai, A. Garcia-Gutierrez, M. Garcia-Reinaldos, M. García-Torres, A. Garofalo, A. Gavel, P. Gavras, E. Gerlach, R. Geyer, P. Giacobbe, G. Gilmore, S. Girona, G. Giuffrida, R. Gomel, A. Gomez, J. González-Núñez, I. González-Santamaría, J. J. González-Vidal, M. Granvik, P. Guillout, J. Guiraud, R. Gutiérrez-Sánchez, L. P. Guy, D. Hatzidimitriou, M. Hauser, M. Haywood, A. Helmer, A. Helmi, M. H. Sarmiento, S. L. Hidalgo, N. Hładczuk, D. Hobbs, G. Holland, H. E. Huckle, G. Jasniewicz, A. Jean-Antoine Piccolo, Ó. Jiménez-Arranz, J. Juaristi Campillo, F. Julbe, L. Karbevska, P. Kervella, G. Kordopatis, A. J. Korn, Á. Kóspál, Z. Kostrzewa-Rutkowska, K. Kruszyńska, M. Kun, P. Laizeau, S. Lambert, A. F. Lanza, Y. Lasne, J. F. Le Campion, Y. Lebreton, T. Lebzelter, S. Leccia, N. Leclerc, I. Lecoeur-Taibi, S. Liao, E. L. Licata, H. E. P. Lindstrøm, T. A. Lister, E. Livanou, A. Lorca, C. Loup, P. Madrero Pardo, A. Magdaleno Romeo, S. Managau, R. G. Mann, M. Manteiga, J. M. Marchant, M. Marconi, J. Marcos, M. M. S. Marcos Santos, D. Marín Pina, S. Marinoni, F. Marocco, L. Martin Polo, J. M. Martín-Fleitas, G. Marton, N. Mary, A. Masip, D. Massari, A. Mastrobuono-Battisti, T. Mazeh, P. J. McMillan, S. Messina, D. Michalik, N. R. Millar, A. Mints, D. Molina, R. Molinaro, L. Molnár, G. Monari, M. Monguió, P. Montegriffo, A. Montero, R. Mor, A. Mora, R. Morbidelli, T. Morel, D. Morris, C. P. Murphy, I. Musella, Z. Nagy, L. Noval, F. Ocaña, A. Ogden, C. Ordenovic, J. O. Osinde, C. Pagani, I. Pagano, L. Palaversa, P. A. Palicio, L. Pallas-Quintela, A. Panahi, S. Payne-Wardenaar, X. Peñalosa Esteller, A. Penttilä, B. Pichon, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, A. Prša, L. Pulone, E. Racero, S. Ragaini, M. Rainer, C. M. Raiteri, M. Ramos-Lerate, P. Re Fiorentin, S. Regibo, P. J. Richards, C. Rios Diaz, A. Riva, H. W. Rix, G. Rixon, N. Robichon, A. C. Robin, C. Robin, M. Roelens, H. R. O. Rogues, L. Rohrbasser, N. Rowell, F. Royer, D. Ruz

Mieres, K. A. Rybicki, G. Sadowski, A. Sáez Núñez, A. Sagristà Sellés, J. Sahlmann, E. Salguero, N. Samaras, V. Sanchez Gimenez, N. Sanna, R. Santoveña, M. Sarasso, M. S. Schultheis, E. Sciacca, M. Segol, J. C. Segovia, D. Ségransan, D. Semeux, S. Shahaf, H. I. Siddiqui, A. Siebert, L. Siltala, A. Silvelo, E. Slezak, I. Slezak, R. L. Smart, O. N. Snaith, E. Solano, F. Solitro, D. Souami, J. Souchay, A. Spagna, L. Spina, F. Spoto, I. A. Steele, H. Steidelmüller, C. A. Stephenson, M. Süveges, J. Surdej, L. Szabados, E. Szegedi-Elek, F. Taris, M. B. Taylor, R. Teixeira, L. Tolomei, N. Tonello, F. Torra, J. Torra, G. Torralba Elipe, M. Trabucchi, A. T. Tsounis, C. Turon, A. Ulla, N. Unger, M. V. Vaillant, E. van Dillen, W. van Reeven, O. Vanel, A. Vecchiato, Y. Viala, D. Vicente, S. Voutsinas, M. Weiler, T. Wevers, Ł. Wyrzykowski, A. Yoldas, P. Yvard, H. Zhao, J. Zorec, S. Zucker, and T. Zwitter. Gaia Data Release 3. Mapping the asymmetric disc of the Milky Way. A&A, 674:A37, June 2023.

[58] Gaia Collaboration, A. Recio-Blanco, G. Kordopatis, P. de Laverny, P. A. Palicio, A. Spagna, L. Spina, D. Katz, P. Re Fiorentin, E. Poggio, P. J. McMillan, A. Vallenari, M. G. Lattanzi, G. M. Seabroke, L. Casamiquela, A. Bragaglia, T. Antoja, C. A. L. Bailer-Jones, M. Schultheis, R. Andrae, M. Fouesneau, M. Cropper, T. Cantat-Gaudin, A. Bijaoui, U. Heiter, A. G. A. Brown, T. Prusti, J. H. J. de Bruijne, F. Arenou, C. Babusiaux, M. Biermann, O. L. Creevey, C. Ducourant, D. W. Evans, L. Eyer, R. Guerra, A. Hutton, C. Jordi, S. A. Klioner, U. L. Lammers, L. Lindegren, X. Luri, F. Mignard, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, C. Soubiran, P. Tanga, N. A. Walton, U. Bastian, R. Drimmel, F. Jansen, F. van Leeuwen, J. Bakker, C. Cacciari, J. Castañeda, F. De Angeli, C. Fabricius, Y. Frémat, L. Galluccio, A. Guerrier, E. Masana, R. Messineo, N. Mowlavi, C. Nicolas, K. Nienartowicz, F. Pailler, P. Panuzzo, F. Riclet, W. Roux, R. Sordo, F. Thévenin, G. Gracia-Abril, J. Portell, D. Teyssier, M. Altmann, M. Audard, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, P. W. Burgess, D. Busonero, G. Busso, H. Cánovas, B. Carry, A. Cellino, N. Cheek, G. Clementini, Y. Damerdji, M. Davidson, P. de Teodoro, M. Nuñez Campos, L. Delchambre, A. Dell'Oro, P. Esquej, J. Fernández-Hernández, E. Fraile, D. Garabato, P. García-Lario, E. Gosset, R. Haigron, J. L. Halbwachs, N. C. Hambly, D. L. Harrison, J. Hernández, D. Hestroffer, S. T. Hodgkin, B. Holl, K. Janßen, G. Jevardat de Fombelle, S. Jordan, A. Krone-Martins, A. C. Lanzafame, W. Löffler, O. Marchal, P. M. Marrese, A. Moitinho, K. Muinonen, P. Osborne, E. Pancino, T. Pauwels, C. Reylé, M. Riello, L. Rimoldini, T. Roegiers, J. Rybizki, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, E. Utrilla, M. van Leeuwen, U. Abbas, P. Ábrahám, A. Abreu Aramburu, C. Aerts, J. J. Aguado, M. Ajaj, F. Aldea-Montero, G. Altavilla, M. A. Álvarez, J. Alves, F. Anders, R. I. Anderson, E. Anglada Varela, D. Baines, S. G. Baker, L. Balaguer-Núñez, E. Balbinot, Z. Balog, C. Barache, D. Barbato, M. Barros, M. A. Barstow, S. Bartolomé, J. L. Bassilana, N. Bauchet, U. Becciani, M. Bellazzini, A. Berihuete, M. Bernet, S. Bertone, L. Bianchi, A. Binnenfeld, S. Blanco-Cuaresma, T. Boch, A. Bombrun, D. Bossini, S. Bouquillon, L. Bramante, E. Breedt, A. Bressan, N. Brouillet, E. Brugaletta, B. Bucciarelli, A. Burlacu, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, R. Carballo, T. Carlucci, M. I. Carnerero, J. M. Carrasco, M. Castellani, A. Castro-Ginard, L. Chaoul, P. Charlot, L. Chemin, V. Chiaramida, A. Chiavassa, N. Chornay, G. Comoretto, G. Contursi, W. J. Cooper, T. Cornez, S. Cowell, F. Crifo, M. Crosta, C. Crowley, C. Dafonte, A. Dapergolas, P. David, F. De Luise, R. De March, J. De Ridder, R. de Souza, A. de Torres, E. F. del Peloso, E. del Pozo, M. Delbo, A. Delgado, J. B. Delisle, C. Demouchy, T. E. Dharmawardena, P. Di Matteo, S. Diakite, C. Diener, E. Distefano, C. Dolding, B. Edvardsson, H. Enke, C. Fabre, M. Fabrizio, S. Faigler, G. Fedorets, P. Fernique, F. Figueras, Y. Fournier, C. Fouron, F. Fragkoudi, M. Gai, A. Garcia-Gutierrez, M. Garcia-Reinaldos, M. García-Torres, A. Garofalo, A. Gavel, P. Gavras, E. Gerlach, R. Geyer, P. Giacobbe, G. Gilmore, S. Girona, G. Giuffrida, R. Gomel, A. Gomez, J. González-Núñez, I. González-Santamaría, J. J. González-Vidal, M. Granvik, P. Guillout, J. Guiraud, R. Gutiérrez-Sánchez, L. P. Guy, D. Hatzidimitriou, M. Hauser, M. Haywood, A. Helmer, A. Helmi, M. H. Sarmiento, S. L. Hidalgo, N. Hładczuk, D. Hobbs, G. Holland, H. E. Huckle, K. Jardine, G. Jasniewicz, A. Jean-Antoine Piccolo, Ó. Jiménez-Arranz, J. Juaristi Campillo, F. Julbe, L. Karbevska, P. Kervella, S. Khanna, A. J. Korn, Á. Kóspál, Z. Kostrzewa-Rutkowska, K. Kruszyńska, M. Kun, P. Laizeau, S. Lambert, A. F. Lanza, Y. Lasne, J. F. Le

Campion, Y. Lebreton, T. Lebzelter, S. Leccia, N. Leclerc, I. Lecoeur-Taibi, S. Liao, E. L. Licata, H. E. P. Lindstrøm, T. A. Lister, E. Livanou, A. Lobel, A. Lorca, C. Loup, P. Madrero Pardo, A. Magdaleno Romeo, S. Managau, R. G. Mann, M. Manteiga, J. M. Marchant, M. Marconi, J. Marcos, M. M. S. Marcos Santos, D. Marín Pina, S. Marinoni, F. Marocco, D. J. Marshall, L. Martin Polo, J. M. Martín-Fleitas, G. Marton, N. Mary, A. Masip, D. Massari, A. Mastrobuono-Battisti, T. Mazeh, S. Messina, D. Michalik, N. R. Millar, A. Mints, D. Molina, R. Molinaro, L. Molnár, G. Monari, M. Monguió, P. Montegriffo, A. Montero, R. Mor, A. Mora, R. Morbidelli, T. Morel, D. Morris, T. Muraveva, C. P. Murphy, I. Musella, Z. Nagy, L. Noval, F. Ocaña, A. Ogden, C. Ordenovic, J. O. Osinde, C. Pagani, I. Pagano, L. Palaversa, L. Pallas-Quintela, A. Panahi, S. Payne-Wardenaar, X. Peñalosa Esteller, A. Penttilä, B. Pichon, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, A. Prša, L. Pulone, E. Racero, S. Ragaini, M. Rainer, C. M. Raiteri, P. Ramos, M. Ramos-Lerate, S. Regibo, P. J. Richards, C. Rios Diaz, V. Ripepi, A. Riva, H. W. Rix, G. Rixon, N. Robichon, A. C. Robin, C. Robin, M. Roelens, H. R. O. Rogues, L. Rohrbasser, M. Romero-Gómez, N. Rowell, F. Royer, D. Ruz Mieres, K. A. Rybicki, G. Sadowski, A. Sáez Núñez, A. Sagristà Sellés, J. Sahlmann, E. Salguero, N. Samaras, V. Sanchez Gimenez, N. Sanna, R. Santoveña, M. Sarasso, E. Sciacca, M. Segol, J. C. Segovia, D. Ségransan, D. Semeux, S. Shahaf, H. I. Siddiqui, A. Siebert, L. Siltala, A. Silvelo, E. Slezak, I. Slezak, R. L. Smart, O. N. Snaith, E. Solano, F. Solitro, D. Souami, J. Souchay, F. Spoto, I. A. Steele, H. Steidelmüller, C. A. Stephenson, M. Süveges, J. Surdej, L. Szabados, E. Szegedi-Elek, F. Taris, M. B. Taylor, R. Teixeira, L. Tolomei, N. Tonello, F. Torra, J. Torra, G. Torralba Elipe, M. Trabucchi, A. T. Tsounis, C. Turon, A. Ulla, N. Unger, M. V. Vaillant, E. van Dillen, W. van Reeven, O. Vanel, A. Vecchiato, Y. Viala, D. Vicente, S. Voutsinas, M. Weiler, T. Wevers, Ł. Wyrzykowski, A. Yoldas, P. Yvard, H. Zhao, J. Zorec, S. Zucker, and T. Zwitter. Gaia Data Release 3. Chemical cartography of the Milky Way. A&A, 674:A38, June 2023.

[59] Gaia Collaboration, A. Vallenari, A. G. A. Brown, T. Prusti, J. H. J. de Bruijne, F. Arenou, C. Babusiaux, M. Biermann, O. L. Creevey, C. Ducourant, D. W. Evans, L. Eyer, R. Guerra, A. Hutton, C. Jordi, S. A. Klioner, U. L. Lammers, L. Lindegren, X. Luri, F. Mignard, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, C. Soubiran, P. Tanga, N. A. Walton, C. A. L. Bailer-Jones, U. Bastian, R. Drimmel, F. Jansen, D. Katz, M. G. Lattanzi, F. van Leeuwen, J. Bakker, C. Cacciari, J. Castañeda, F. De Angeli, C. Fabricius, M. Fouesneau, Y. Frémat, L. Galluccio, A. Guerrier, U. Heiter, E. Masana, R. Messineo, N. Mowlavi, C. Nicolas, K. Nienartowicz, F. Pailler, P. Panuzzo, F. Riclet, W. Roux, G. M. Seabroke, R. Sordo, F. Thévenin, G. Gracia-Abril, J. Portell, D. Teyssier, M. Altmann, R. Andrae, M. Audard, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, P. W. Burgess, D. Busonero, G. Busso, H. Cánovas, B. Carry, A. Cellino, N. Cheek, G. Clementini, Y. Damerdji, M. Davidson, P. de Teodoro, M. Nuñez Campos, L. Delchambre, A. Dell'Oro, P. Esquej, J. Fernández-Hernández, E. Fraile, D. Garabato, P. García-Lario, E. Gosset, R. Haigron, J. L. Halbwachs, N. C. Hambly, D. L. Harrison, J. Hernández, D. Hestroffer, S. T. Hodgkin, B. Holl, K. Janßen, G. Jevardat de Fombelle, S. Jordan, A. Krone-Martins, A. C. Lanzafame, W. Löffler, O. Marchal, P. M. Marrese, A. Moitinho, K. Muinonen, P. Osborne, E. Pancino, T. Pauwels, A. Recio-Blanco, C. Reylé, M. Riello, L. Rimoldini, T. Roegiers, J. Rybizki, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, E. Utrilla, M. van Leeuwen, U. Abbas, P. Ábrahám, A. Abreu Aramburu, C. Aerts, J. J. Aguado, M. Ajaj, F. Aldea-Montero, G. Altavilla, M. A. Álvarez, J. Alves, F. Anders, R. I. Anderson, E. Anglada Varela, T. Antoja, D. Baines, S. G. Baker, L. Balaguer-Núñez, E. Balbinot, Z. Balog, C. Barache, D. Barbato, M. Barros, M. A. Barstow, S. Bartolomé, J. L. Bassilana, N. Bauchet, U. Becciani, M. Bellazzini, A. Berihuete, M. Bernet, S. Bertone, L. Bianchi, A. Binnenfeld, S. Blanco-Cuaresma, A. Blazere, T. Boch, A. Bombrun, D. Bossini, S. Bouquillon, A. Bragaglia, L. Bramante, E. Breedt, A. Bressan, N. Brouillet, E. Brugaletta, B. Bucciarelli, A. Burlacu, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, T. Cantat-Gaudin, R. Carballo, T. Carlucci, M. I. Carnerero, J. M. Carrasco, L. Casamiquela, M. Castellani, A. Castro-Ginard, L. Chaoul, P. Charlot, L. Chemin, V. Chiaramida, A. Chiavassa, N. Chornay, G. Comoretto, G. Contursi, W. J. Cooper, T. Cornez, S. Cowell, F. Crifo, M. Cropper, M. Crosta, C. Crowley, C. Dafonte,

A. Dapergolas, M. David, P. David, P. de Laverny, F. De Luise, R. De March, J. De Ridder, R. de Souza, A. de Torres, E. F. del Peloso, E. del Pozo, M. Delbo, A. Delgado, J. B. Delisle, C. Demouchy, T. E. Dharmawardena, P. Di Matteo, S. Diakite, C. Diener, E. Distefano, C. Dolding, B. Edvardsson, H. Enke, C. Fabre, M. Fabrizio, S. Faigler, G. Fedorets, P. Fernique, A. Fienga, F. Figueras, Y. Fournier, C. Fouron, F. Fragkoudi, M. Gai, A. Garcia-Gutierrez, M. Garcia-Reinaldos, M. García-Torres, A. Garofalo, A. Gavel, P. Gavras, E. Gerlach, R. Geyer, P. Giacobbe, G. Gilmore, S. Girona, G. Giuffrida, R. Gomel, A. Gomez, J. González-Núñez, I. González-Santamaría, J. J. González-Vidal, M. Granvik, P. Guillout, J. Guiraud, R. Gutiérrez-Sánchez, L. P. Guy, D. Hatzidimitriou, M. Hauser, M. Haywood, A. Helmer, A. Helmi, M. H. Sarmiento, S. L. Hidalgo, T. Hilger, N. Hładczuk, D. Hobbs, G. Holland, H. E. Huckle, K. Jardine, G. Jasniewicz, A. Jean-Antoine Piccolo, Ó. Jiménez-Arranz, A. Jorissen, J. Juaristi Campillo, F. Julbe, L. Karbevska, P. Kervella, S. Khanna, M. Kontizas, G. Kordopatis, A. J. Korn, Á. Kóspál, Z. Kostrzewa-Rutkowska, K. Kruszyńska, M. Kun, P. Laizeau, S. Lambert, A. F. Lanza, Y. Lasne, J. F. Le Campion, Y. Lebreton, T. Lebzelter, S. Leccia, N. Leclerc, I. Lecoeur-Taibi, S. Liao, E. L. Licata, H. E. P. Lindstrøm, T. A. Lister, E. Livanou, A. Lobel, A. Lorca, C. Loup, P. Madrero Pardo, A. Magdaleno Romeo, S. Managau, R. G. Mann, M. Manteiga, J. M. Marchant, M. Marconi, J. Marcos, M. M. S. Marcos Santos, D. Marín Pina, S. Marinoni, F. Marocco, D. J. Marshall, L. Martin Polo, J. M. Martín-Fleitas, G. Marton, N. Mary, A. Masip, D. Massari, A. Mastrobuono-Battisti, T. Mazeh, P. J. McMillan, S. Messina, D. Michalik, N. R. Millar, A. Mints, D. Molina, R. Molinaro, L. Molnár, G. Monari, M. Monguió, P. Montegriffo, A. Montero, R. Mor, A. Mora, R. Morbidelli, T. Morel, D. Morris, T. Muraveva, C. P. Murphy, I. Musella, Z. Nagy, L. Noval, F. Ocaña, A. Ogden, C. Ordenovic, J. O. Osinde, C. Pagani, I. Pagano, L. Palaversa, P. A. Palicio, L. Pallas-Quintela, A. Panahi, S. Payne-Wardenaar, X. Peñalosa Esteller, A. Penttilä, B. Pichon, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, E. Poggio, A. Prša, L. Pulone, E. Racero, S. Ragaini, M. Rainer, C. M. Raiteri, N. Rambaux, P. Ramos, M. Ramos-Lerate, P. Re Fiorentin, S. Regibo, P. J. Richards, C. Rios Diaz, V. Ripepi, A. Riva, H. W. Rix, G. Rixon, N. Robichon, A. C. Robin, C. Robin, M. Roelens, H. R. O. Rogues, L. Rohrbasser, M. Romero-Gómez, N. Rowell, F. Royer, D. Ruz Mieres, K. A. Rybicki, G. Sadowski, A. Sáez Núñez, A. Sagristà Sellés, J. Sahlmann, E. Salguero, N. Samaras, V. Sanchez Gimenez, N. Sanna, R. Santoveña, M. Sarasso, M. Schultheis, E. Sciacca, M. Segol, J. C. Segovia, D. Ségransan, D. Semeux, S. Shahaf, H. I. Siddiqui, A. Siebert, L. Siltala, A. Silvelo, E. Slezak, I. Slezak, R. L. Smart, O. N. Snaith, E. Solano, F. Solitro, D. Souami, J. Souchay, A. Spagna, L. Spina, F. Spoto, I. A. Steele, H. Steidelmüller, C. A. Stephenson, M. Süveges, J. Surdej, L. Szabados, E. Szegedi-Elek, F. Taris, M. B. Taylor, R. Teixeira, L. Tolomei, N. Tonello, F. Torra, J. Torra, G. Torralba Elipe, M. Trabucchi, A. T. Tsounis, C. Turon, A. Ulla, N. Unger, M. V. Vaillant, E. van Dillen, W. van Reeven, O. Vanel, A. Vecchiato, Y. Viala, D. Vicente, S. Voutsinas, M. Weiler, T. Wevers, Ł. Wyrzykowski, A. Yoldas, P. Yvard, H. Zhao, J. Zorec, S. Zucker, and T. Zwitter. Gaia Data Release 3. Summary of the content and survey properties. A&A, 674:A1, June 2023.

[60] Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al. The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*, 2020.

[61] Ana E. García Pérez, Carlos Allende Prieto, Jon A. Holtzman, Matthew Shetrone, Szabolcs Mészáros, Dmitry Bizyaev, Ricardo Carrera, Katia Cunha, D. A. García-Hernández, Jennifer A. Johnson, Steven R. Majewski, David L. Nidever, Ricardo P. Schiavon, Neville Shane, Verne V. Smith, Jennifer Sobeck, Nicholas Troup, Olga Zamora, David H. Weinberg, Jo Bovy, Daniel J. Eisenstein, Diane Feuillet, Peter M. Frinchaboy, Michael R. Hayden, Fred R. Hearty, Duy C. Nguyen, Robert W. O'Connell, Marc H. Pinsonneault, John C. Wilson, and Gail Zasowski. ASPCAP: The APOGEE Stellar Parameter and Chemical Abundances Pipeline. AJ, 151(6):144, June 2016.

[62] M. Goswami, K. Szafer, A. Choudhry, Y. Cai, S. Li, and A. Dubrawski. MOMENT: A Family of Open Time-series Foundation Models. *ArXiv e-prints*, page arXiv:2402.03885, 2024.

[63] J. Guy, M. Sullivan, A. Conley, N. Regnault, P. Astier, C. Balland, S. Basa, R. G. Carlberg, D. Fouchez, D. Hardin, I. M. Hook, D. A. Howell, R. Pain, N. Palanque-Delabrouille, K. M. Perrett, C. J. Pritchet, J. Rich, V. Ruhlmann-Kleider, D. Balam, S. Baumont, R. S. Ellis, S. Fabbro, H. K. Fakhouri, N. Fourmanoit, S. González-Gaitán, M. L. Graham, E. Hsiao, T. Kronborg, C. Lidman, A. M. Mourao, S. Perlmutter, P. Ripoche, N. Suzuki, and E. S. Walker. The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints. A&A, 523:A7, November 2010.

[64] J. Guy, M. Sullivan, A. Conley, N. Regnault, P. Astier, C. Balland, S. Basa, R. G. Carlberg, D. Fouchez, D. Hardin, I. M. Hook, D. A. Howell, R. Pain, N. Palanque-Delabrouille, K. M. Perrett, C. J. Pritchet, J. Rich, V. Ruhlmann-Kleider, D. Balam, S. Baumont, R. S. Ellis, S. Fabbro, H. K. Fakhouri, N. Fourmanoit, S. González-Gaitán, M. L. Graham, E. Hsiao, T. Kronborg, C. Lidman, A. M. Mourao, S. Perlmutter, P. Ripoche, N. Suzuki, and E. S. Walker. The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints. A&A, 523:A7, November 2010.

[65] ChangHoon Hahn, K. J. Kwon, Rita Tojeiro, Malgorzata Siudek, Rebecca E. A. Canning, Mar Mezcua, Jeremy L. Tinker, David Brooks, Peter Doel, Kevin Fanning, Enrique Gaztañaga, Robert Kehoe, Martin Landriau, Aaron Meisner, John Moustakas, Claire Poppett, Gregory Tarle, Benjamin Weiner, and Hu Zou. The desi probabilistic value-added bright galaxy survey (provabgs) mock challenge. *The Astrophysical Journal*, 945(1):16, March 2023.

[66] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[67] Pavan R. Hebbar and Craig O. Heinke. Machine Learning Applied to X-Ray Spectra: Separating Stars in Orion Nebula Cluster from Active Galactic Nuclei in CDFS. ApJ, 949(1):12, May 2023.

[68] Malcolm Hicken, Peter Challis, Saurabh Jha, Robert P. Kirshner, Tom Matheson, Maryam Modjaz, Armin Rest, W. Michael Wood-Vasey, Gaspar Bakos, Elizabeth J. Barton, Perry Berlind, Ann Bragg, Cesar Briceño, Warren R. Brown, Nelson Caldwell, Mike Calkins, Richard Cho, Larry Ciupik, Maria Contreras, Kristi-Concannon Dendy, Anil Dosaj, Nick Durham, Kris Eriksen, Gil Esquerdo, Mark Everett, Emilio Falco, Jose Fernandez, Alejandro Gaba, Peter Garnavich, Genevieve Graves, Paul Green, Ted Groner, Carl Hergenrother, Matthew J. Holman, Vit Hradecky, John Huchra, Bob Hutchison, Diab Jerius, Andres Jordan, Roy Kilgard, Miriam Krauss, Kevin Luhman, Lucas Macri, Daniel Marrone, Jonathan McDowell, Daniel McIntosh, Brian McNamara, Tom Megeath, Barbara Mochejska, Diego Munoz, James Muzerolle, Orlando Naranjo, Gautham Narayan, Michael Pahre, Wayne Peters, Dawn Peterson, Ken Rines, Ben Ripman, Anna Roussanova, Rudolph Schild, Aurora Sicilia-Aguilar, Jennifer Sokoloski, Kyle Smalley, Andy Smith, Tim Spahr, K. Z. Stanek, Pauline Barmby, Stéphane Blondin, Christopher W. Stubbs, Andrew Szentgyorgyi, Manuel A. P. Torres, Amili Vaz, Alexey Vikhlinin, Zhong Wang, Mike Westover, Deborah Woods, and Ping Zhao. CfA3: 185 Type Ia Supernova Light Curves from the CfA. ApJ, 700(1):331–357, July 2009.

[69] Malcolm Hicken, Peter Challis, Robert P. Kirshner, Armin Rest, Claire E. Cramer, W. Michael Wood-Vasey, Gaspar Bakos, Perry Berlind, Warren R. Brown, Nelson Caldwell, Mike Calkins, Thayne Currie, Kathy de Kleer, Gil Esquerdo, Mark Everett, Emilio Falco, Jose Fernandez, Andrew S. Friedman, Ted Groner, Joel Hartman, Matthew J. Holman, Robert Hutchins, Sonia Keys, David Kipping, Dave Latham, George H. Marion, Gautham Narayan, Michael Pahre, Andras Pal, Wayne Peters, Gopakumar Perumpilly, Ben Ripman, Brigitta Sipocz, Andrew Szentgyorgyi, Sumin Tang, Manuel A. P. Torres, Amali Vaz, Scott Wolk, and Andreas Zezas. CfA4: Light Curves for 94 Type Ia Supernovae. ApJS, 200(2):12, June 2012.

[70] Malcolm Hicken, Andrew S. Friedman, Stephane Blondin, Peter Challis, Perry Berlind, Mike Calkins, Gil Esquerdo, Thomas Matheson, Maryam Modjaz, Armin Rest, and Robert P. Kirshner. Type II Supernova Light Curves and Spectra from the CfA. ApJS, 233(1):6, November 2017.

[71] R. Hložek, A. I. Malz, K. A. Ponder, M. Dai, G. Narayan, E. E. O. Ishida, Jr. Allam, T., A. Bahmanyar, X. Bi, R. Biswas, K. Boone, S. Chen, N. Du, A. Erdem, L. Galbany, A. Garreta, S. W. Jha, D. O. Jones, R. Kessler, M. Lin, J. Liu, M. Lochner, A. A. Mahabal, K. S. Mandel, P. Margolis, J. R. Martínez-Galarza, J. D. McEwen, D. Muthukrishna, Y. Nakatsuka, T. Noumi, T. Oya, H. V. Peiris, C. M. Peters, J. F. Puget, C. N. Setzer, Siddhartha, S. Stefanov, T. Xie, L. Yan, K. H. Yeh, and W. Zuo. Results of the Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC). ApJS, 267(2):25, August 2023.

[72] Marc Hon, Daniel Huber, James S. Kuszlewicz, Dennis Stello, Sanjib Sharma, Jamie Tayar, Joel C. Zinn, Mathieu Vrard, and Marc H. Pinsonneault. A "Quick Look" at All-sky Galactic Archeology with TESS: 158,000 Oscillating Red Giants from the MIT Quick-look Pipeline. ApJ, 919(2):131, October 2021.

[73] Chelsea X. Huang, Andrew Vanderburg, Andras Pál, Lizhou Sha, Liang Yu, Willie Fong, Michael Fausnaugh, Avi Shporer, Natalia Guerrero, Roland Vanderspek, and George Ricker. Photometry of 10 Million Stars from the First Two Years of TESS Full Frame Images: Part I. *Research Notes of the American Astronomical Society*, 4(11):204, November 2020.

[74] Chelsea X. Huang, Andrew Vanderburg, Andras Pál, Lizhou Sha, Liang Yu, Willie Fong, Michael Fausnaugh, Avi Shporer, Natalia Guerrero, Roland Vanderspek, and George Ricker. Photometry of 10 Million Stars from the First Two Years of TESS Full Frame Images: Part II. *Research Notes of the American Astronomical Society*, 4(11):206, November 2020.

[75] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[76] X. Huang, C. Storfer, V. Ravi, A. Pilon, M. Domingo, D. J. Schlegel, S. Bailey, A. Dey, R. R. Gupta, D. Herrera, S. Juneau, M. Landriau, D. Lang, A. Meisner, J. Moustakas, A. D. Myers, E. F. Schlafly, F. Valdes, B. A. Weaver, J. Yang, and C. Yèche. Finding Strong Gravitational Lenses in the DESI DECam Legacy Survey. ApJ, 894(1):78, May 2020.

[77] M. Huertas-Company, K. G. Iyer, E. Angeloudi, M. B. Bagley, S. L. Finkelstein, J. Kartaltepe, E. J. McGrath, R. Sarmiento, J. Vega-Ferrero, P. Arrabal Haro, P. Behroozi, F. Buitrago, Y. Cheng, L. Costantin, A. Dekel, M. Dickinson, D. Elbaz, N. A. Grogin, N. P. Hathi, B. W. Holwerda, A. M. Koekemoer, R. A. Lucas, C. Papovich, P. G. Pérez-González, N. Pirzkal, L. M. Seillé, A. de la Vega, S. Wuyts, G. Yang, and L. Y. A. Yung. Galaxy morphology from z ∼ 6 through the lens of JWST. A&A, 685:A48, May 2024.

[78] M. Huertas-Company and F. Lanusse. The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys. PASA, 40:e001, January 2023.

[79] Željko Ivezić, Steven M. Kahn, J. Anthony Tyson, Bob Abel, Emily Acosta, Robyn Allsman, David Alonso, Yusra AlSayyad, Scott F. Anderson, John Andrew, James Roger P. Angel, George Z. Angeli, Reza Ansari, Pierre Antilogus, Constanza Araujo, Robert Armstrong, Kirk T. Arndt, Pierre Astier, Éric Aubourg, Nicole Auza, Tim S. Axelrod, Deborah J. Bard, Jeff D. Barr, Aurelian Barrau, James G. Bartlett, Amanda E. Bauer, Brian J. Bauman, Sylvain Baumont, Ellen Bechtol, Keith Bechtol, Andrew C. Becker, Jacek Becla, Cristina Beldica, Steve Bellavia, Federica B. Bianco, Rahul Biswas, Guillaume Blanc, Jonathan Blazek, Roger D. Blandford, Josh S. Bloom, Joanne Bogart, Tim W. Bond, Michael T. Booth, Anders W. Borgland, Kirk Borne, James F. Bosch, Dominique Boutigny, Craig A. Brackett, Andrew Bradshaw, William Nielsen

Brandt, Michael E. Brown, James S. Bullock, Patricia Burchat, David L. Burke, Gianpietro Cagnoli, Daniel Calabrese, Shawn Callahan, Alice L. Callen, Jeffrey L. Carlin, Erin L. Carlson, Srinivasan Chandrasekharan, Glenaver Charles-Emerson, Steve Chesley, Elliott C. Cheu, Hsin-Fang Chiang, James Chiang, Carol Chirino, Derek Chow, David R. Ciardi, Charles F. Claver, Johann Cohen-Tanugi, Joseph J. Cockrum, Rebecca Coles, Andrew J. Connolly, Kem H. Cook, Asantha Cooray, Kevin R. Covey, Chris Cribbs, Wei Cui, Roc Cutri, Philip N. Daly, Scott F. Daniel, Felipe Daruich, Guillaume Daubard, Greg Daues, William Dawson, Francisco Delgado, Alfred Dellapenna, Robert de Peyster, Miguel de Val-Borro, Seth W. Digel, Peter Doherty, Richard Dubois, Gregory P. Dubois-Felsmann, Josef Durech, Frossie Economou, Tim Eifler, Michael Eracleous, Benjamin L. Emmons, Angelo Fausti Neto, Henry Ferguson, Enrique Figueroa, Merlin Fisher-Levine, Warren Focke, Michael D. Foss, James Frank, Michael D. Freemon, Emmanuel Gangler, Eric Gawiser, John C. Geary, Perry Gee, Marla Geha, Charles J. B. Gessner, Robert R. Gibson, D. Kirk Gilmore, Thomas Glanzman, William Glick, Tatiana Goldina, Daniel A. Goldstein, Iain Goodenow, Melissa L. Graham, William J. Gressler, Philippe Gris, Leanne P. Guy, Augustin Guyonnet, Gunther Haller, Ron Harris, Patrick A. Hascall, Justine Haupt, Fabio Hernandez, Sven Herrmann, Edward Hileman, Joshua Hoblitt, John A. Hodgson, Craig Hogan, James D. Howard, Dajun Huang, Michael E. Huffer, Patrick Ingraham, Walter R. Innes, Suzanne H. Jacoby, Bhuvnesh Jain, Fabrice Jammes, M. James Jee, Tim Jenness, Garrett Jernigan, Darko Jevremović, Kenneth Johns, Anthony S. Johnson, Margaret W. G. Johnson, R. Lynne Jones, Claire Juramy-Gilles, Mario Jurić, Jason S. Kalirai, Nitya J. Kallivayalil, Bryce Kalmbach, Jeffrey P. Kantor, Pierre Karst, Mansi M. Kasliwal, Heather Kelly, Richard Kessler, Veronica Kinnison, David Kirkby, Lloyd Knox, Ivan V. Kotov, Victor L. Krabbendam, K. Simon Krughoff, Petr Kubánek, John Kuczewski, Shri Kulkarni, John Ku, Nadine R. Kurita, Craig S. Lage, Ron Lambert, Travis Lange, J. Brian Langton, Laurent Le Guillou, Deborah Levine, Ming Liang, Kian-Tat Lim, Chris J. Lintott, Kevin E. Long, Margaux Lopez, Paul J. Lotz, Robert H. Lupton, Nate B. Lust, Lauren A. MacArthur, Ashish Mahabal, Rachel Mandelbaum, Thomas W. Markiewicz, Darren S. Marsh, Philip J. Marshall, Stuart Marshall, Morgan May, Robert McKercher, Michelle McQueen, Joshua Meyers, Myriam Migliore, Michelle Miller, David J. Mills, Connor Miraval, Joachim Moeyens, Fred E. Moolekamp, David G. Monet, Marc Moniez, Serge Monkewitz, Christopher Montgomery, Christopher B. Morrison, Fritz Mueller, Gary P. Muller, Freddy Muñoz Arancibia, Douglas R. Neill, Scott P. Newbry, Jean-Yves Nief, Andrei Nomerotski, Martin Nordby, Paul O'Connor, John Oliver, Scot S. Olivier, Knut Olsen, William O'Mullane, Sandra Ortiz, Shawn Osier, Russell E. Owen, Reynald Pain, Paul E. Palecek, John K. Parejko, James B. Parsons, Nathan M. Pease, J. Matt Peterson, John R. Peterson, Donald L. Petravick, M. E. Libby Petrick, Cathy E. Petry, Francesco Pierfederici, Stephen Pietrowicz, Rob Pike, Philip A. Pinto, Raymond Plante, Stephen Plate, Joel P. Plutchak, Paul A. Price, Michael Prouza, Veljko Radeka, Jayadev Rajagopal, Andrew P. Rasmussen, Nicolas Regnault, Kevin A. Reil, David J. Reiss, Michael A. Reuter, Stephen T. Ridgway, Vincent J. Riot, Steve Ritz, Sean Robinson, William Roby, Aaron Roodman, Wayne Rosing, Cecille Roucelle, Matthew R. Rumore, Stefano Russo, Abhijit Saha, Benoit Sassolas, Terry L. Schalk, Pim Schellart, Rafe H. Schindler, Samuel Schmidt, Donald P. Schneider, Michael D. Schneider, William Schoening, German Schumacher, Megan E. Schwamb, Jacques Sebag, Brian Selvy, Glenn H. Sembroski, Lynn G. Seppala, Andrew Serio, Eduardo Serrano, Richard A. Shaw, Ian Shipsey, Jonathan Sick, Nicole Silvestri, Colin T. Slater, J. Allyn Smith, R. Chris Smith, Shahram Sobhani, Christine Soldahl, Lisa Storrie-Lombardi, Edward Stover, Michael A. Strauss, Rachel A. Street, Christopher W. Stubbs, Ian S. Sullivan, Donald Sweeney, John D. Swinbank, Alexander Szalay, Peter Takacs, Stephen A. Tether, Jon J. Thaler, John Gregg Thayer, Sandrine Thomas, Adam J. Thornton, Vaikunth Thukral, Jeffrey Tice, David E. Trilling, Max Turri, Richard Van Berg, Daniel Vanden Berk, Kurt Vetter, Francoise Virieux, Tomislav Vucina, William Wahl, Lucianne Walkowicz, Brian Walsh, Christopher W. Walter, Daniel L. Wang, Shin-Yawn Wang, Michael Warner, Oliver Wiecha, Beth Willman, Scott E. Winters, David Wittman, Sidney C. Wolff, W. Michael Wood-Vasey, Xiuqin Wu, Bo Xin, Peter Yoachim, and Hu Zhan. LSST: From Science Drivers to Reference Design and Anticipated Data

Products. ApJ, 873(2):111, March 2019.

[80] Jon M. Jenkins, Joseph D. Twicken, Sean McCauliff, Jennifer Campbell, Dwight Sanderfer, David Lung, Masoud Mansouri-Samani, Forrest Girouard, Peter Tenenbaum, Todd Klaus, Jeffrey C. Smith, Douglas A. Caldwell, A. D. Chacon, Christopher Henze, Cory Heiges, David W. Latham, Edward Morgan, Daryl Swade, Stephen Rinehart, and Roland Vanderspek. The TESS science processing operations center. In Gianluca Chiozzi and Juan C. Guzman, editors, *Software and Cyberinfrastructure for Astronomy IV*, volume 9913 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 99133E, August 2016.

[81] D. O. Jones, D. M. Scolnic, R. J. Foley, A. Rest, R. Kessler, P. M. Challis, K. C. Chambers, D. A. Coulter, K. G. Dettman, M. M. Foley, M. E. Huber, S. W. Jha, E. Johnson, C. D. Kilpatrick, R. P. Kirshner, J. Manuel, G. Narayan, Y. C. Pan, A. G. Riess, A. S. B. Schultz, M. R. Siebert, E. Berger, R. Chornock, H. Flewelling, E. A. Magnier, S. J. Smartt, K. W. Smith, R. J. Wainscoat, C. Waters, and M. Willman. The Foundation Supernova Survey: Measuring Cosmological Parameters with Supernovae from a Single Telescope. ApJ, 881(1):19, August 2019.

[82] N. V. Karpenka, F. Feroz, and M. P. Hobson. A simple and robust method for automated photometric classification of supernovae using neural networks. MNRAS, 429(2):1278–1285, February 2013.

[83] Richard Kessler, Bruce Bassett, Pavel Belov, Vasudha Bhatnagar, Heather Campbell, Alex Conley, Joshua A. Frieman, Alexandre Glazov, Santiago González-Gaitán, Renée Hlozek, Saurabh Jha, Stephen Kuhlmann, Martin Kunz, Hubert Lampeitl, Ashish Mahabal, James Newling, Robert C. Nichol, David Parkinson, Ninan Sajeeth Philip, Dovi Poznanski, Joseph W. Richards, Steven A. Rodney, Masao Sako, Donald P. Schneider, Mathew Smith, Maximilian Stritzinger, and Melvin Varughese. Results from the Supernova Photometric Classification Challenge. PASP, 122(898):1415, December 2010.

[84] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[85] David G. Koch, William J. Borucki, Gibor Basri, Natalie M. Batalha, Timothy M. Brown, Douglas Caldwell, Jørgen Christensen-Dalsgaard, William D. Cochran, Edna DeVore, Edward W. Dunham, III Gautier, Thomas N., John C. Geary, Ronald L. Gilliland, Alan Gould, Jon Jenkins, Yoji Kondo, David W. Latham, Jack J. Lissauer, Geoffrey Marcy, David Monet, Dimitar Sasselov, Alan Boss, Donald Brownlee, John Caldwell, Andrea K. Dupree, Steve B. Howell, Hans Kjeldsen, Søren Meibom, David Morrison, Tobias Owen, Harold Reitsema, Jill Tarter, Stephen T. Bryson, Jessie L. Dotson, Paul Gazis, Michael R. Haas, Jeffrey Kolodziejczak, Jason F. Rowe, Jeffrey E. Van Cleve, Christopher Allen, Hema Chandrasekaran, Bruce D. Clarke, Jie Li, Elisa V. Quintana, Peter Tenenbaum, Joseph D. Twicken, and Hayley Wu. Kepler Mission Design, Realized Photometric Performance, and Early Science. ApJ, 713(2):L79–L86, April 2010.

[86] Kevin Krisciunas, Carlos Contreras, Christopher R. Burns, M. M. Phillips, Maximilian D. Stritzinger, Nidia Morrell, Mario Hamuy, Jorge Anais, Luis Boldt, Luis Busta, Abdo Campillay, Sergio Castellón, Gastón Folatelli, Wendy L. Freedman, Consuelo González, Eric Y. Hsiao, Wojtek Krzeminski, Sven Eric Persson, Miguel Roth, Francisco Salgado, Jacqueline Serón, Nicholas B. Suntzeff, Simón Torres, Alexei V. Filippenko, Weidong Li, Barry F. Madore, D. L. DePoy, Jennifer L. Marshall, Jean-Philippe Rheault, and Steven Villanueva. The Carnegie Supernova Project. I. Third Photometry Data Release of Low-redshift Type Ia Supernovae and Other White Dwarf Explosions. AJ, 154(5):211, November 2017.

[87] Michelle Kunimoto, Chelsea Huang, Evan Tey, Willie Fong, Katharine Hesse, Avi Shporer, Natalia Guerrero, Michael Fausnaugh, Roland Vanderspek, and George Ricker. Quick-look Pipeline Lightcurves for 9.1 Million Stars Observed over the First Year of

the TESS Extended Mission. *Research Notes of the American Astronomical Society*, 5(10):234, October 2021.

[88] Michelle Kunimoto, Evan Tey, Willie Fong, Katharine Hesse, Avi Shporer, Michael Fausnaugh, Roland Vanderspek, and George Ricker. Quick-look Pipeline Light Curves for 5.7 Million Stars Observed Over the Second Year of TESS' First Extended Mission. *Research Notes of the American Astronomical Society*, 6(11):236, November 2022.

[89] David R. Law, Brian Cherinka, Renbin Yan, Brett H. Andrews, Matthew A. Bershady, Dmitry Bizyaev, Guillermo A. Blanc, Michael R. Blanton, Adam S. Bolton, Joel R. Brownstein, Kevin Bundy, Yanmei Chen, Niv Drory, Richard D'Souza, Hai Fu, Amy Jones, Guinevere Kauffmann, Nicholas MacDonald, Karen L. Masters, Jeffrey A. Newman, John K. Parejko, José R. Sánchez-Gallego, Sebastian F. Sánchez, David J. Schlegel, Daniel Thomas, David A. Wake, Anne-Marie Weijmans, Kyle B. Westfall, and Kai Zhang. The Data Reduction Pipeline for the SDSS-IV MaNGA IFU Galaxy Survey. AJ, 152(4):83, October 2016.

[90] Gregory R. Lee, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt, and Aaron O'Leary. Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019.

[91] M. Leoni, E. E. O. Ishida, J. Peloton, and A. Möller. Fink: Early supernovae Ia classification using active learning. A&A, 663:A13, July 2022.

[92] Henry Leung and Jo Bovy. Galaxy10 decals dataset. https://astronn.readthedocs.io/en/latest/galaxy10.html, 2023. Accessed: 2024-06-06.

[93] Henry W Leung and Jo Bovy. Deep learning of multi-element abundances from high-resolution spectroscopic data. *Monthly Notices of the Royal Astronomical Society*, 483(3):3255–3277, 2019.

[94] Henry W. Leung and Jo Bovy. Towards an astronomical foundation model for stars with a transformer-based model. MNRAS, 527(1):1494–1520, January 2024.

[95] Yang Li, Si Si, Gang Li, Cho-Jui Hsieh, and Samy Bengio. Learnable fourier features for multi-dimensional spatial positional encoding. *Advances in Neural Information Processing Systems*, 34:15816–15829, 2021.

[96] Yan Liang, Peter Melchior, ChangHoon Hahn, Jeff Shen, Andy Goulding, and Charlotte Ward. Outlier Detection in the DESI Bright Galaxy Survey. ApJ, 956(1):L6, October 2023.

[97] Chris Lintott, Kevin Schawinski, Steven Bamford, Anå¾e Slosar, Kate Land, Daniel Thomas, Edd Edmondson, Karen Masters, Robert C. Nichol, M. Jordan Raddick, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg. Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies. MNRAS, 410(1):166–178, January 2011.

[98] Chris Lintott, Kevin Schawinski, Steven Bamford, Anže Slosar, Kate Land, Daniel Thomas, Edd Edmondson, Karen Masters, Robert C Nichol, M Jordan Raddick, et al. Galaxy zoo 1: data release of morphological classifications for nearly 900 000 galaxies. *Monthly Notices of the Royal Astronomical Society*, 410(1):166–178, 2011.

[99] Chris J. Lintott, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick, Robert C. Nichol, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg. Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey . *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189, September 2008.

[100] Michelle Lochner, Catarina Alves, Hiranya Peiris, Jason McEwen, Jr. Allam, Tarek, Rahul Biswas, Johnny Holland, Ofer Lahav, Robert Schuhmann, Christian Setzer, and Max Winter. snmachine: Photometric supernova classification. Astrophysics Source Code Library, record ascl:2107.006, July 2021.

[101] Michelle Lochner, Jason D. McEwen, Hiranya V. Peiris, Ofer Lahav, and Max K. Winter. Photometric Supernova Classification with Machine Learning. ApJS, 225(2):31, August 2016.

[102] Steven R. Majewski, Ricardo P. Schiavon, Peter M. Frinchaboy, Carlos Allende Prieto, Robert Barkhouser, Dmitry Bizyaev, Basil Blank, Sophia Brunner, Adam Burton, Ricardo Carrera, S. Drew Chojnowski, Kátia Cunha, Courtney Epstein, Greg Fitzgerald, Ana E. García Pérez, Fred R. Hearty, Chuck Henderson, Jon A. Holtzman, Jennifer A. Johnson, Charles R. Lam, James E. Lawler, Paul Maseman, Szabolcs Mészáros, Matthew Nelson, Duy Coung Nguyen, David L. Nidever, Marc Pinsonneault, Matthew Shetrone, Stephen Smee, Verne V. Smith, Todd Stolberg, Michael F. Skrutskie, Eric Walker, John C. Wilson, Gail Zasowski, Friedrich Anders, Sarbani Basu, Stephane Beland, Michael R. Blanton, Jo Bovy, Joel R. Brownstein, Joleen Carlberg, William Chaplin, Cristina Chiappini, Daniel J. Eisenstein, Yvonne Elsworth, Diane Feuillet, Scott W. Fleming, Jessica Galbraith-Frew, Rafael A. García, D. Aníbal García-Hernández, Bruce A. Gillespie, Léo Girardi, James E. Gunn, Sten Hasselquist, Michael R. Hayden, Saskia Hekker, Inese Ivans, Karen Kinemuchi, Mark Klaene, Suvrath Mahadevan, Savita Mathur, Benoît Mosser, Demitri Muna, Jeffrey A. Munn, Robert C. Nichol, Robert W. O'Connell, John K. Parejko, A. C. Robin, Helio Rocha-Pinto, Matthias Schultheis, Aldo M. Serenelli, Neville Shane, Victor Silva Aguirre, Jennifer S. Sobeck, Benjamin Thompson, Nicholas W. Troup, David H. Weinberg, and Olga Zamora. The Apache Point Observatory Galactic Evolution Experiment (APOGEE). AJ, 154(3):94, September 2017.

[103] K Małek, A Solarz, Agnieszka Pollo, A Fritz, B Garilli, M Scodeggio, A Iovino, BR Granett, U Abbas, C Adami, et al. The vimos public extragalactic redshift survey (vipers)-a support vector machine classification of galaxies, stars, and agns. *Astronomy & Astrophysics*, 557:A16, 2013.

[104] Peter Melchior, Yan Liang, ChangHoon Hahn, and Andy Goulding. Autoencoding galaxy spectra. i. architecture. *The Astronomical Journal*, 166(2):74, 2023.

[105] Siddharth Mishra-Sharma, Yiding Song, and Jesse Thaler. Paperclip: Associating astronomical observations and natural language with multi-modal models, 2024.

[106] Devina Mohan, Anna M M Scaife, Fiona Porter, Mike Walmsley, and Micah Bowles. Quantifying uncertainty in deep learning approaches to radio galaxy classification. *Monthly Notices of the Royal Astronomical Society*, 511(3):3722–3740, 01 2022.

[107] A. Möller, V. Ruhlmann-Kleider, C. Leloup, J. Neveu, N. Palanque-Delabrouille, J. Rich, R. Carlberg, C. Lidman, and C. Pritchet. Photometric classification of type Ia supernovae in the SuperNova Legacy Survey with supervised learning. J. Cosmology Astropart. Phys., 2016(12):008, December 2016.

[108] J. Newling, B. Bassett, R. Hlozek, M. Kunz, M. Smith, and M. Varughese. Parameter estimation with Bayesian estimation applied to multiple species in the presence of biases and correlations. MNRAS, 421(2):913–925, April 2012.

[109] Tuan Dung Nguyen, Yuan-Sen Ting, I. Ciucă, Charlie O'Neill, Ze-Chang Sun, Maja Jablo'nska, S. Kruk, Ernest Perkowski, Jack W. Miller, Jason Li, Josh Peek, K. Iyer, Tomasz R'o.za'nski, P. Khetarpal, Sharaf Zaman, D. Brodrick, Sergio J. Rodr'iguez M'endez, Thang Bui, Alyssa Goodman, A. Accomazzi, Jill P. Naiman, Jesse Cranney, K. Schawinski, and UniverseTBD. Astrollama: Towards specialized foundation models in astronomy. *ArXiv*, abs/2309.06126, 2023.

[110] Greg Olmschenk, Richard K. Barry, Stela Ishitani Silva, Brian P. Powell, Ethan Kruse, Jeremy D. Schnittman, Agnieszka M. Cieplak, Thomas Barclay, Siddhant Solanki, Bianca Ortega, John Baker, and Yesenia Helem Salinas Mamani. Short-Period Variables in TESS Full-Frame Image Light Curves Identified via Convolutional Neural Networks. *arXiv e-prints*, page arXiv:2402.12369, February 2024.

[111] Jia-Shu Pan, Yuan-Sen Ting, and Jie Yu. Astroconformer: The prospects of analysing stellar light curves with transformer-based deep learning models. MNRAS, 528(4):5890–5903, March 2024.

[112] Liam Parker, Francois Lanusse, Siavash Golkar, Leopoldo Sarra, Miles Cranmer, Alberto Bietti, Michael Eickenberg, Geraud Krawezik, Michael McCabe, Rudy Morel, et al. Astroclip: A cross-modal foundation model for galaxies. *Monthly Notices of the Royal Astronomical Society*, page stae1450, 2024.

[113] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Andreas Müller, Joel Nothman, Gilles Louppe, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, October 2011.

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

[115] M. M. Phillips, Carlos Contreras, E. Y. Hsiao, Nidia Morrell, Christopher R. Burns, Maximilian Stritzinger, C. Ashall, Wendy L. Freedman, P. Hoeflich, S. E. Persson, Anthony L. Piro, Nicholas B. Suntzeff, Syed A. Uddin, Jorge Anais, E. Baron, Luis Busta, Abdo Campillay, Sergio Castellón, Carlos Corco, T. Diamond, Christa Gall, Consuelo Gonzalez, Simon Holmbo, Kevin Krisciunas, Miguel Roth, Jacqueline Serón, F. Taddia, Simón Torres, J. P. Anderson, C. Baltay, Gastón Folatelli, L. Galbany, A. Goobar, Ellie Hadjiyska, Mario Hamuy, Mansi Kasliwal, C. Lidman, Peter E. Nugent, S. Perlmutter, David Rabinowitz, Stuart D. Ryder, Brian P. Schmidt, B. J. Shappee, and Emma S. Walker. Carnegie Supernova Project-II: Extending the Near-infrared Hubble Diagram for Type Ia Supernovae to z ∼ 0.1. PASP, 131(995):014001, January 2019.

[116] Helen Qu, Masao Sako, Anais Möller, and Cyrille Doux. SCONE: Supernova Classification with a Convolutional Neural Network. AJ, 162(2):67, August 2021.

[117] Helen Qu and Sang Michael Xie. Connect later: Improving fine-tuning for robustness with targeted augmentations. 2024.

[118] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. *arXiv e-prints*, page arXiv:2103.00020, February 2021.

[119] Nabeel Rehemtulla, Adam A. Miller, Michael W. Coughlin, and Theophile Jegou du Laz. BTSbot: A Multi-input Convolutional Neural Network to Automate and Expedite Bright Transient Identification for the Zwicky Transient Facility. *arXiv e-prints*, page arXiv:2307.07618, July 2023.

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

[121] George R. Ricker, Joshua N. Winn, Roland Vanderspek, David W. Latham, Gáspár Á. Bakos, Jacob L. Bean, Zachory K. Berta-Thompson, Timothy M. Brown, Lars Buchhave, Nathaniel R. Butler, R. Paul Butler, William J. Chaplin, David Charbonneau, Jørgen Christensen-Dalsgaard, Mark Clampin, Drake Deming, John Doty, Nathan De Lee, Courtney Dressing, Edward W. Dunham, Michael Endl, Francois Fressin, Jian Ge, Thomas Henning, Matthew J. Holman, Andrew W. Howard, Shigeru Ida, Jon M. Jenkins, Garrett Jernigan, John Asher Johnson, Lisa Kaltenegger, Nobuyuki Kawai, Hans Kjeldsen, Gregory Laughlin, Alan M. Levine, Douglas Lin, Jack J. Lissauer, Phillip MacQueen, Geoffrey Marcy, Peter R. McCullough, Timothy D. Morton, Norio Narita, Martin Paegert, Enric Palle, Francesco Pepe, Joshua Pepper, Andreas Quirrenbach, Stephen A. Rinehart, Dimitar Sasselov, Bun'ei Sato, Sara Seager, Alessandro Sozzetti, Keivan G. Stassun, Peter Sullivan, Andrew Szentgyorgyi, Guillermo Torres, Stephane Udry, and Joel Villasenor. Transiting Exoplanet Survey Satellite (TESS). *Journal of Astronomical Telescopes, Instruments, and Systems*, 1:014003, January 2015.

[122] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv e-prints*, page arXiv:1505.04597, May 2015.

[123] David Rubin, Greg Aldering, Marc Betoule, Andy Fruchter, Xiaosheng Huang, Alex G. Kim, Chris Lidman, Eric Linder, Saul Perlmutter, Pilar Ruiz-Lapuente, and Nao Suzuki. Union Through UNITY: Cosmology with 2,000 SNe Using a Unified Bayesian Framework. *arXiv e-prints*, page arXiv:2311.12098, November 2023.

[124] Kevin Schawinski, C. Megan Urry, Brooke D. Simmons, Lucy Fortson, Sugata Kaviraj, William C. Keel, Chris J. Lintott, Karen L. Masters, Robert C. Nichol, Marc Sarzi, Ramin Skibba, Ezequiel Treister, Kyle W. Willett, O. Ivy Wong, and Sukyoung K. Yi. The green valley is a red herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies . *Monthly Notices of the Royal Astronomical Society*, 440(1):889–907, 03 2014.

[125] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. LAION-5B: An open large-scale dataset for training next generation image-text models. *arXiv e-prints*, page arXiv:2210.08402, October 2022.

[126] MARCO Scodeggio, L Guzzo, BIANCA Garilli, BR Granett, M Bolzonella, S De La Torre, U Abbas, C Adami, S Arnouts, D Bottini, et al. The vimos public extragalactic redshift survey (vipers)-full spectroscopic data and auxiliary information release (pdr-2). *Astronomy & Astrophysics*, 609:A84, 2018.

[127] D. M. Scolnic, D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, R. J. Foley, M. E. Huber, R. Kessler, G. Narayan, A. G. Riess, S. Rodney, E. Berger, D. J. Brout, P. J. Challis, M. Drout, D. Finkbeiner, R. Lunnan, R. P. Kirshner, N. E. Sanders, E. Schlafly, S. Smartt, C. W. Stubbs, J. Tonry, W. M. Wood-Vasey, M. Foley, J. Hand, E. Johnson, W. S. Burgett, K. C. Chambers, P. W. Draper, K. W. Hodapp, N. Kaiser, R. P. Kudritzki, E. A. Magnier, N. Metcalfe, F. Bresolin, E. Gall, R. Kotak, M. McCrum, and K. W. Smith. The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. ApJ, 859(2):101, June 2018.

[128] Dan Scolnic, Dillon Brout, Anthony Carr, Adam G. Riess, Tamara M. Davis, Arianna Dwomoh, David O. Jones, Noor Ali, Pranav Charvu, Rebecca Chen, Erik R. Peterson, Brodie Popovic, Benjamin M. Rose, Charlotte M. Wood, Peter J. Brown, Ken Chambers, David A. Coulter, Kyle G. Dettman, Georgios Dimitriadis, Alexei V. Filippenko, Ryan J. Foley, Saurabh W. Jha, Charles D. Kilpatrick, Robert P. Kirshner, Yen-Chen Pan, Armin Rest, Cesar Rojas-Bravo, Matthew R. Siebert, Benjamin E. Stahl, and WeiKang Zheng. The Pantheon+ Analysis: The Full Data Set and Light-curve Release. ApJ, 938(2):113, October 2022.

[129] M Siudek, K Małek, Agnieszka Pollo, T Krakowski, Angela Iovino, MARCO Scodeggio, T Moutard, G Zamorani, L Guzzo, BIANCA Garilli, et al. The vimos public extragalactic redshift survey (vipers)-the complexity of galaxy populations at 0.4< z< 1.3 revealed with unsupervised machine-learning algorithms. *Astronomy & Astrophysics*, 617:A70, 2018.

[130] M. J. Smith and J. E. Geach. Astronomia ex machina: a history, primer and outlook on neural networks in astronomy. *Royal Society Open Science*, 10(5), 2023.

[131] M. J. Smith, R. J. Roberts, E. Angeloudi, and M. Huertas-Company. AstroPT: Scaling Large Observation Models for Astronomy. *ArXiv e-prints*, page arXiv:2405.14930, 2024.

[132] George Stein, Jacqueline Blaum, Peter Harrington, Tomislav Medan, and Zarija Lukić. Mining for Strong Gravitational Lenses with Self-supervised Learning. ApJ, 932(2):107, June 2022.

[133] George Stein, Jacqueline Blaum, Peter Harrington, Tomislav Medan, and Zarija Lukić. Mining for Strong Gravitational Lenses with Self-supervised Learning. ApJ, 932(2):107, June 2022.

[134] George Stein, Peter Harrington, Jacqueline Blaum, Tomislav Medan, and Zarija Lukic. Self-supervised similarity search for large scientific datasets. *arXiv e-prints*, page arXiv:2110.13151, October 2021.

[135] Maximilian D. Stritzinger, M. M. Phillips, Luis N. Boldt, Chris Burns, Abdo Campillay, Carlos Contreras, Sergio Gonzalez, Gastón Folatelli, Nidia Morrell, Wojtek Krzeminski, Miguel Roth, Francisco Salgado, D. L. DePoy, Mario Hamuy, Wendy L. Freedman, Barry F. Madore, J. L. Marshall, Sven E. Persson, Jean-Philippe Rheault, Nicholas B. Suntzeff, Steven Villanueva, Weidong Li, and Alexei V. Filippenko. The Carnegie Supernova Project: Second Photometry Data Release of Low-redshift Type Ia Supernovae. AJ, 142(5):156, November 2011.

[136] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019.

[137] Evan Tey, Dan Moldovan, Michelle Kunimoto, Chelsea X. Huang, Avi Shporer, Tansu Daylan, Daniel Muthukrishna, Andrew Vanderburg, Anne Dattilo, George R. Ricker, and S. Seager. Identifying Exoplanets with Deep Learning. V. Improved Light-curve Classification for TESS Full-frame Image Observations. AJ, 165(3):95, March 2023.

[138] The PLAsTiCC team, Jr. Allam, Tarek, Anita Bahmanyar, Rahul Biswas, Mi Dai, Lluís Galbany, Renée Hložek, Emille E. O. Ishida, Saurabh W. Jha, David O. Jones, Richard Kessler, Michelle Lochner, Ashish A. Mahabal, Alex I. Malz, Kaisey S. Mandel, Juan Rafael Martínez-Galarza, Jason D. McEwen, Daniel Muthukrishna, Gautham Narayan, Hiranya Peiris, Christina M. Peters, Kara Ponder, Christian N. Setzer, The LSST Dark Energy Science Collaboration, The LSST Transients, and Variable Stars Science Collaboration. The Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC): Data set. *arXiv e-prints*, page arXiv:1810.00001, September 2018.

[139] Tijmen2. Cosmosage_v2. https://huggingface.co/Tijmen2/cosmosage_v2, 2024. Accessed: 2024-06-06.

[140] Syed A. Uddin, Christopher R. Burns, Mark M. Phillips, Nicholas B. Suntzeff, Wendy L. Freedman, Peter J. Brown, Nidia Morrell, Mario Hamuy, Kevin Krisciunas, Lifan Wang, Eric Y. Hsiao, Ariel Goobar, Saul Perlmutter, Jing Lu, Maximilian Stritzinger, Joseph P. Anderson, Chris Ashall, Peter Hoeflich, Benjamin J. Shappee, S. E. Persson, Anthony L. Piro, Eddie Baron, Carlos Contreras, Lluís Galbany, Sahana Kumar, Melissa Shahbandeh, Scott Davis, Jorge Anais, Luis Busta, Abdo Campillay, Sergio

Castellón, Carlos Corco, Tiara Diamond, Christa Gall, Consuelo Gonzalez, Simon Holmbo, Miguel Roth, Jacqueline Serón, Francesco Taddia, Simón Torres, Charles Baltay, Gastón Folatelli, Ellie Hadjiyska, Mansi Kasliwal, Peter E. Nugent, David Rabinowitz, and Stuart D. Ryder. Carnegie Supernova Project-I and -II: Measurements of H_0 using Cepheid, TRGB, and SBF Distance Calibration to Type Ia Supernovae. *arXiv e-prints*, page arXiv:2308.01875, August 2023.

[141] Francesco Valentino, Gabriel Brammer, Katriona M. L. Gould, Vasily Kokorev, Seiji Fujimoto, Christian Kragh Jespersen, Aswin P. Vijayan, John R. Weaver, Kei Ito, Masayuki Tanaka, Olivier Ilbert, Georgios E. Magdis, Katherine E. Whitaker, Andreas L. Faisst, Anna Gallazzi, Steven Gillman, Clara Giménez-Arteaga, Carlos Gómez-Guijarro, Mariko Kubo, Kasper E. Heintz, Michaela Hirschmann, Pascal Oesch, Masato Onodera, Francesca Rizzo, Minju Lee, Victoria Strait, and Sune Toft. An Atlas of Color-selected Quiescent Galaxies at z > 3 in Public JWST Fields. ApJ, 947(1):20, April 2023.

[142] Jesús Vega-Ferrero, Marc Huertas-Company, Luca Costantin, Pablo G. Pérez-González, Regina Sarmiento, Jeyhan S. Kartaltepe, Annalisa Pillepich, Micaela B. Bagley, Steven L. Finkelstein, Elizabeth J. McGrath, Johan H. Knapen, Pablo Arrabal Haro, Eric F. Bell, Fernando Buitrago, Antonello Calabrò, Avishai Dekel, Mark Dickinson, Helena Domínguez Sánchez, David Elbaz, Henry C. Ferguson, Mauro Giavalisco, Benne W. Holwerda, Dale D. Kocesvski, Anton M. Koekemoer, Viraj Pandya, Casey Papovich, Nor Pirzkal, Joel Primack, and L. Y. Aaron Yung. On the Nature of Disks at High Redshift Seen by JWST/CEERS with Contrastive Learning and Cosmological Simulations. ApJ, 961(1):51, January 2024.

[143] V. A. Villar, E. Berger, G. Miller, R. Chornock, A. Rest, D. O. Jones, M. R. Drout, R. J. Foley, R. Kirshner, R. Lunnan, E. Magnier, D. Milisavljevic, N. Sanders, and D. Scolnic. Supernova Photometric Classification Pipelines Trained on Spectroscopically Classified Supernovae from the Pan-STARRS1 Medium-deep Survey. ApJ, 884(1):83, October 2019.

[144] Mike Walmsley, Micah Bowles, Anna M. M. Scaife, Jason Shingirai Makechemu, Alexander J. Gordon, Annette M. N. Ferguson, Robert G. Mann, James Pearson, Jürgen J. Popp, Jo Bovy, Josh Speagle, Hugh Dickinson, Lucy Fortson, Tobias Géron, Sandor Kruk, Chris J. Lintott, Kameswara Mantha, Devina Mohan, David O'Ryan, and Inigo V. Slijepevic. Scaling laws for galaxy images, 2024.

[145] Mike Walmsley, Tobias Géron, Sandor Kruk, Anna M. M. Scaife, Chris Lintott, Karen L. Masters, James M. Dawson, Hugh Dickinson, Lucy Fortson, Izzy L. Garland, Kameswara Mantha, David O'Ryan, Jürgen Popp, Brooke Simmons, Elisabeth M. Baeten, and Christine Macmillan. Galaxy Zoo DESI: Detailed morphology measurements for 8.7M galaxies in the DESI Legacy Imaging Surveys. MNRAS, 526(3):4768–4786, December 2023.

[146] Mike Walmsley, Tobias Géron, Sandor Kruk, Anna M. M. Scaife, Chris Lintott, Karen L. Masters, James M. Dawson, Hugh Dickinson, Lucy Fortson, Izzy L. Garland, Kameswara Mantha, David O'Ryan, Jürgen Popp, Brooke Simmons, Elisabeth M. Baeten, and Christine Macmillan. Galaxy Zoo DESI: Detailed morphology measurements for 8.7M galaxies in the DESI Legacy Imaging Surveys. MNRAS, 526(3):4768–4786, December 2023.

[147] Mike Walmsley, Chris Lintott, Tobias Géron, Sandor Kruk, Coleman Krawczyk, Kyle W. Willett, Steven Bamford, Lee S. Kelvin, Lucy Fortson, Yarin Gal, William Keel, Karen L. Masters, Vihang Mehta, Brooke D. Simmons, Rebecca Smethurst, Lewis Smith, Elisabeth M. Baeten, and Christine Macmillan. Galaxy Zoo DECaLS: Detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies. MNRAS, 509(3):3966–3988, January 2022.

[148] Mike Walmsley, Lewis Smith, Chris Lintott, Yarin Gal, Steven Bamford, Hugh Dickinson, Lucy Fortson, Sandor Kruk, Karen Masters, Claudia Scarlata, Brooke Simmons, Rebecca Smethurst, and Darryl Wright. Galaxy Zoo: probabilistic morphology through

Bayesian CNNs and active learning. *Monthly Notices of the Royal Astronomical Society*, 491(2):1554–1574, 10 2019.

[149] Kyle B. Westfall, Michele Cappellari, Matthew A. Bershady, Kevin Bundy, Francesco Belfiore, Xihan Ji, David R. Law, Adam Schaefer, Shravan Shetty, Christy A. Tremonti, Renbin Yan, Brett H. Andrews, Joel R. Brownstein, Brian Cherinka, Lodovico Coccato, Niv Drory, Claudia Maraston, Taniya Parikh, José R. Sánchez-Gallego, Daniel Thomas, Anne-Marie Weijmans, Jorge Barrera-Ballesteros, Cheng Du, Daniel Goddard, Niu Li, Karen Masters, Héctor Javier Ibarra Medel, Sebastián F. Sánchez, Meng Yang, Zheng Zheng, and Shuang Zhou. The Data Analysis Pipeline for the SDSS-IV MaNGA IFU Galaxy Survey: Overview. AJ, 158(6):231, December 2019.

[150] Belinda Wilkes and Wallace Tucker. *The Chandra X-ray Observatory; Exploring the high energy universe.* 2019.

[151] J. C. Wilson, F. R. Hearty, M. F. Skrutskie, S. R. Majewski, J. A. Holtzman, D. Eisenstein, J. Gunn, B. Blank, C. Henderson, S. Smee, M. Nelson, D. Nidever, J. Arns, R. Barkhouser, J. Barr, S. Beland, M. A. Bershady, M. R. Blanton, S. Brunner, A. Burton, L. Carey, M. Carr, J. P. Colque, J. Crane, G. J. Damke, Jr. Davidson, J. W., J. Dean, F. Di Mille, K. W. Don, G. Ebelke, M. Evans, G. Fitzgerald, B. Gillespie, M. Hall, A. Harding, P. Harding, R. Hammond, D. Hancock, C. Harrison, S. Hope, T. Horne, J. Karakla, C. Lam, F. Leger, N. MacDonald, P. Maseman, J. Matsunari, S. Melton, T. Mitcheltree, T. O'Brien, R. W. O'Connell, A. Patten, W. Richardson, G. Rieke, M. Rieke, A. Roman-Lopes, R. P. Schiavon, J. S. Sobeck, T. Stolberg, R. Stoll, M. Tembe, J. D. Trujillo, A. Uomoto, M. Vernieri, E. Walker, D. H. Weinberg, E. Young, B. Anthony-Brumfield, D. Bizyaev, B. Breslauer, N. De Lee, J. Downey, S. Halverson, J. Huehnerhoff, M. Klaene, E. Leon, D. Long, S. Mahadevan, E. Malanushenko, D. C. Nguyen, R. Owen, J. R. Sánchez-Gallego, C. Sayres, N. Shane, S. A. Shectman, M. Shetrone, D. Skinner, F. Stauffer, and B. Zhao. The Apache Point Observatory Galactic Evolution Experiment (APOGEE) Spectrographs. PASP, 131(999):055001, May 2019.

[152] John F. Wu, Alina Hyk, Kiera McCormick, Christine Ye, Simone Astarita, Elina Baral, Jo Ciuca, Jesse Cranney, Anjalie Field, Kartheik Iyer, Philipp Koehn, Jenn Kotler, Sandor Kruk, Michelle Ntampaka, Charles O'Neill, Joshua E. G. Peek, Sanjib Sharma, and Mikaeel Yunus. Designing an evaluation framework for large language models in astronomy research, 2024.

[153] Donald G. York, J. Adelman, Jr. Anderson, John E., Scott F. Anderson, James Annis, Neta A. Bahcall, J. A. Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, William N. Boroski, Steve Bracker, Charlie Briegel, John W. Briggs, J. Brinkmann, Robert Brunner, Scott Burles, Larry Carey, Michael A. Carr, Francisco J. Castander, Bing Chen, Patrick L. Colestock, A. J. Connolly, J. H. Crocker, István Csabai, Paul C. Czarapata, John Eric Davis, Mamoru Doi, Tom Dombeck, Daniel Eisenstein, Nancy Ellman, Brian R. Elms, Michael L. Evans, Xiaohui Fan, Glenn R. Federwitz, Larry Fiscelli, Scott Friedman, Joshua A. Frieman, Masataka Fukugita, Bruce Gillespie, James E. Gunn, Vijay K. Gurbani, Ernst de Haas, Merle Haldeman, Frederick H. Harris, J. Hayes, Timothy M. Heckman, G. S. Hennessy, Robert B. Hindsley, Scott Holm, Donald J. Holmgren, Chi-hao Huang, Charles Hull, Don Husby, Shin-Ichi Ichikawa, Takashi Ichikawa, Željko Ivezić, Stephen Kent, Rita S. J. Kim, E. Kinney, Mark Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, John Korienek, Richard G. Kron, Peter Z. Kunszt, D. Q. Lamb, B. Lee, R. French Leger, Siriluk Limmongkol, Carl Lindenmeyer, Daniel C. Long, Craig Loomis, Jon Loveday, Rich Lucinio, Robert H. Lupton, Bryan MacKinnon, Edward J. Mannery, P. M. Mantsch, Bruce Margon, Peregrine McGehee, Timothy A. McKay, Avery Meiksin, Aronne Merelli, David G. Monet, Jeffrey A. Munn, Vijay K. Narayanan, Thomas Nash, Eric Neilsen, Rich Neswold, Heidi Jo Newberg, R. C. Nichol, Tom Nicinski, Mario Nonino, Norio Okada, Sadanori Okamura, Jeremiah P. Ostriker, Russell Owen, A. George Pauls, John Peoples, R. L. Peterson, Donald Petravick, Jeffrey R. Pier, Adrian Pope, Ruth Pordes, Angela Prosapio, Ron Rechenmacher, Thomas R. Quinn, Gordon T. Richards, Michael W. Richmond, Claudio H. Rivetta, Constance M. Rockosi, Kurt Ruthmansdorfer, Dale

Sandford, David J. Schlegel, Donald P. Schneider, Maki Sekiguchi, Gary Sergey, Kazuhiro Shimasaku, Walter A. Siegmund, Stephen Smee, J. Allyn Smith, S. Snedden, R. Stone, Chris Stoughton, Michael A. Strauss, Christopher Stubbs, Mark SubbaRao, Alexander S. Szalay, Istvan Szapudi, Gyula P. Szokoly, Anirudda R. Thakar, Christy Tremonti, Douglas L. Tucker, Alan Uomoto, Dan Vanden Berk, Michael S. Vogeley, Patrick Waddell, Shu-i. Wang, Masaru Watanabe, David H. Weinberg, Brian Yanny, Naoki Yasuda, and SDSS Collaboration. The Sloan Digital Sky Survey: Technical Summary. AJ, 120(3):1579–1587, September 2000.

[154] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

[155] A Ćiprijanović, A Lewis, K Pedro, S Madireddy, B Nord, G N Perdue, and S M Wild. Deepastrouda: semi-supervised universal domain adaptation for cross-survey galaxy morphology classification and anomaly detection. *Machine Learning: Science and Technology*, 4(2):025013, apr 2023.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See section 5
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] We include an impact statement in appendix.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]

3. If you ran experiments (e.g. for benchmarks)...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] All codes are included on our projects GitHub repository `https://github.com/MultimodalUniverse/MultimodalUniverse`, and links are included in the appendix.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] Training details can be found as configuration files in the source code, and documented in the appendix.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] For reasons of computational budget, and quoted results from the astrophysics literature often do not originally contain error estimates.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Information provided in appendix

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [Yes] Documented in supplementary material for each dataset.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] The compiled dataset, available from the project GitHub page.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] All collected datasets are publicly available scientific datasets.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No] All collected data is exclusively tied to astrophysics.

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

Appendix Table of Contents

A Multimodal Universe Datasheet

A.1 Motivation

The MULTIMODAL UNIVERSE dataset was compiled to enable the development of large-scale and multimodal scientific machine learning models and brings together hundreds of millions of astronomical observations, constituting over 70TB of data that includes multi-channel and hyperspectral images, spectra, multivariate time series, and scientific measurements from the field of astronomy. It further addresses a critical gap in the availability of standardized and machine learning-ready datasets in astronomy. Despite the data-rich nature of the field, existing datasets are often fragmented and tailored for traditional analysis rather than machine learning workflows. In addition, MULTIMODAL UNIVERSE puts significant emphasis on the inclusion and standardization of the metadata necessary to interpret these measurements, thus supporting the development of metadata-aware machine learning models which are necessary for the proper interpretation of scientific data. The data collected in MULTIMODAL UNIVERSE is gathered from a multitude of astronomical surveys and instruments, and is organized into subsets corresponding to one particular parent survey. The data is further organized as to allow the cross-matching of its different components in order to build pairs of different modalities for the same objects.

A.2 Distribution

We openly distribute the MULTIMODAL UNIVERSE dataset under the Creative Commons Attribution (CC BY) 4.0 license, noting however that when using specific subsets, the license and conditions of utilisation should be respected. Any work using the MULTIMODAL UNIVERSE should therefore cite the main paper and appropriately cite or acknowledge the individual dataset sources, as described in this appendix. The Multimodal Universe package[4] offers citation tools to automatically retrieve the citations and acknowledgements of the various datasets.

The full dataset content in native HDF5 format is hosted at the Flatiron Institute. The data described in this paper is `v1` (version 1), with future versions expected to be hosted through the same system, and are available either through HTTPS or through GLOBUS[5] at the following links:

- https://users.flatironinstitute.org/~polymathic/data/
  MultimodalUniverse

---

[4]https://github.com/MultimodalUniverse/MultimodalUniverse/
[5]https://www.globus.org/

- https://app.globus.org/file-manager?origin_id=
  57136152-fc1d-418e-b74e-75ca52bddd21&origin_path=/

GLOBUS is much preferable when downloading large amounts of data, or a large number of files. Local download of the full data in its native HDF5 format is necessary for using the provided cross-matching utilities.

In addition, for convenience, fast streaming access, and ease of prototyping, a copy of the dataset in parquet format is maintained on the Hugging Face Hub, as individual datasets under the MULTIMODAL UNIVERSE organization:

- https://huggingface.co/MultimodalUniverse

Finally, the source code used to compile all datasets, along with documentation and examples is hosted on the project's GitHub repository: https://github.com/MultimodalUniverse/MultimodalUniverse

A.3   Image Datasets

A.3.1   Legacy Surveys DR10: `legacysurvey` ☁ 🔗 🤗 🥐

Figure 3: Illustration of grz images of galaxies from the Legacy Survey DR10 sample.

Description   The DESI Legacy Imaging Surveys (hereafter Legacy Surveys) [44] is originally a combination of dedicated large imaging surveys conducted with three different telescopes (the Blanco telescope at the Cerro Tololo Inter-American Observatory; the Mayall Telescope at the Kitt Peak National Observatory; and the University of Arizona Steward Observatory 2.3m (90inch) Bart Bok Telescope at Kitt Peak National Observatory), and covering over 14,000 square degrees in three optical bands (g,r,z). With its most recent data release, DR10, the Legacy Surveys were further extended in the southern hemisphere with additional observations in four bands (g,r,i,z) from the Dark Energy Camera (DECam), pushing the total survey area to almost 20,000 square degrees (i.e. half of the sky). Beyond the large joint footprint that the Legacy Surveys represent, their main characteristic is that they are

all analysed self-consistently with the same processing and inference pipeline, providing a large and uniform dataset over a large fraction of the sky. We specifically base this dataset on the 4-band imaging of the southern sky part of Data Release 10, and representing 15,000 square degrees.

Composition   The dataset we compile consists of 124M postage stamps of size 160x160 pixels with pixel size of 0.262 arcsec in the (g,r,i,z) bands, combined with a set of measurements from the Legacy Survey analysis pipeline. The sample was selected based on the following criteria:

- Extended objects: TYPE != 'PSF'

- Magnitude cut: mag_z < 21

- Full Color: observations in all 4 bands

- Quality cuts: all objects with the following flags are excluded[6]: BRIGHT, SATUR_G, SATUR_R, SATUR_Z ALLMASK_G, ALLMASK_R, ALLMASK_Z, MEDIUM, SATUR_I, ALLMASK_I

Collection Process   We first download from the Legacy Survey data archive[7] all coadded bricks from the DR10 southern region which possess observations in all 4 bands. Each of these bricks is an image of size 3600x3600 pixels, from which we extract 160x160 cutouts centered on the coordinates of detected objects that pass the selection cuts mentioned above. We further associate each of these cutouts with the corresponding entry in the Legacy Survey sweep catalogs which contain the most commonly used fields provided by the analysis pipeline.

Preprocessing / cleaning   No further processing is applied beyond the selection cuts reported above. This means in particular that cutouts may contain problematic regions (e.g. saturated stars) around the central object. These regions can be identified by the user through the `mask` field included with each cutout and that mark unreliable pixels in the image. The motivation for not including more aggressive cleaning is that defects are ubiquitous in astronomical images, and part of the motivation for this dataset is to build machine learning models that are robust to such defects.

Uses   It is the first time that this specific collection of data is assembled, however similar datasets based on Legacy Survey images have been used for a range of machine learning applications, such as identifying strong gravitational lenses [133, 76], or classifying galaxy morphologies [146].

Acknowledgements and Terms of Use   When used in scientific publications, the formal acknowledgement described at this URL should be used: https://www.legacysurvey.org/acknowledgment/#toc-entry-2

A.3.2   Legacy Surveys North: `ssl_legacysurvey` ☁ 🎮 🤗 🥐

Description   This dataset is a collection of galaxy postage stamps extracted from the northern sky of the Legacy Surveys Data Release 9 in [134] in the context of building one of the first examples of self-supervised learning in astrophysics (hence the name `ssl_legacysurvey`). The northern component of DR9 contains (g,r,z) images from the Beijing-Arizona Sky Survey (BASS), and the Mayall z-band Legacy Survey (MzLS) which self consistently analysed through the Legacy Surveys pipeline. This dataset is complementary to subsubsection A.3.1 by providing a large set of imaging data in the northern part of the sky.

Composition   The dataset consists of 15M galaxies, each imaged in three optical bands (g, r, z) at a pixel scale of 0.262 arcsec. The images are provided in 152×152 cut-outs. Each galaxy image is also associated with a set of measurement from the Legacy Survey catalog.

---

[6]Description available at: https://www.legacysurvey.org/dr10/bitmasks/

[7]https://www.legacysurvey.org/dr10/files/

The sample was selected based on the following criteria:

- Extended objects: TYPE != 'PSF'
- Magnitude cut: $20 <$ mag_z <21
- Quality cuts[8]: all objects with the following flags are excluded: BRIGHT, ALL-MASK_G, ALLMASK_R, ALLMASK_Z, MEDIUM

Collection Process The original collection process used in [134] for this dataset is based on extracted cutouts from coadded images for each galaxy in the selected sample using the `imagine`[9] library. This processing was performed at the National Energy Research Scientific Computing Center (NERSC) using the local copy of the Legacy Survey data hosted at the facility. The collected image cutouts along with relevant catalog information was then packaged as a set of HDF5 files and made publicly available through GLOBUS [10]. The content of our dataset is directly and exactly extracted from the northern component of this dataset.

Preprocessing/Cleaning No further processing is applied to the dataset compiled by [134].

Uses This specific dataset was used to build the self-supervised representation learning model introduced in [134] and used to identify strong gravitational lensing in [132]. More recently, this dataset was used to build the image-spectrum contrastive learning model presented in [112].

Acknowledgements and Terms of Use When using this dataset in scientific publications, the data compilation paper should be cited [132] and the formal acknowledgement described at this URL should be used: https://www.legacysurvey.org/acknowledgment/#toc-entry-2

A.3.3 HSC: `hsc` ☁️ 🐙 🤗 🥐

Description The Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP) [5] provides a wide-field imaging survey using the Hyper Suprime-Cam on the Subaru Telescope. Covering approximately 1,200 square degrees of the sky in five broad-band filters (g, r, i, z, y), the survey boasts high-quality images with a median seeing of about 0.6 arcseconds in the i-band and depth reaching 26.4 magnitudes in the r-band, which leads to significantly better image quality than the Legacy Surveys. In this dataset, we specifically use data from the Deep/Ultra Deep fields of the third Public Data Release[PDR3, 4]. These fields represent a much smaller fraction of the sky (27 square degrees), but are much deeper (i.e. have less noise) than the wide survey.

Composition The dataset is composed of 400k 160x160 cutouts in five bands (g,r,i,z,y) at a pixel scale of 0.168 arcsec. Each image is also associated with a set of measurements from the analysis pipeline. The sample was selected from the Deep/UltraDeep catalog based on the following criteria:

- Magnitude cut: mag_i < 22.5
- Full depth Full Color: At least 3 observations in all bands
- Quality cuts: Remove all objects affected by bright stars, objects intersecting edges, objects with saturated or interpolated pixels, and objects with unreliable cmodel magnitude.

These selection cuts lead to a sample of 400k objects (which may include stars in addition to galaxies).

---

[8]the meaning of the flags is described here: https://www.legacysurvey.org/dr9/bitmasks/

[9]https://github.com/legacysurvey/imagine

[10]https://github.com/georgestein/ssl-legacysurvey/tree/main?tab=readme-ov-file#data-access

Collection Process We first download all coadded images (specifically `calexp` files from the `deepCoadd_result` folder) for the Deep/UltraDeep fields from the HSC PDR3 data archive[11]. Each of these images is of size 4200x4200, from which we extract cutouts centered on the coordinates of the objects in the sample selected above. We associate each of these cutouts with their corresponding entry in the object catalog.

Preprocessing/cleaning No further processing is applied to the images beyond the selection cuts mentioned above. Similarly to previous imaging datasets, images can contain bad pixel or artifacts surrounding the central object, those can be identified through the `mask` field that identifies good pixels.

Uses It is the first time that this specific dataset is compiled, it will have similar use cases as the other imaging datasets, i.e. machine learning methods for detecting rare objects, or classifying or characterizing galaxy properties.

Acknowledgements and Terms of Use Data from the HSC SSP can be used without prior permission within the following scopes:

- Extent of free use stipulated by Japanese copyright law (private use, educational use, news reporting, etc.)
- Usage in academic research, education, and learning activities
- Usage by news organizations
- Usage in printed media, or in websites and social networks

The full description of restricted use-cases is provided here: https://www.nao.ac.jp/en/terms/copyright.html. In all cases, please explicitly include the credit "NAOJ / HSC Collaboration".

A.3.4 BTSbot: `btsbot` ☁️ 🎧 🤗 🥐

Description The Zwicky Transient Facility (ZTF) [15, 16] Bright Transient Survey (BTS) [56, 114] is the largest spectroscopic supernova survey ever constructed, surveying the entire nighttime Northern sky every 2-3 nights in two filters aiming to acquire spectra for a complete sample of bright, local transients. While identification of new transient sources has traditionally been led by human scanners assessing images, BTS have developed the BTSbot model and accompanying dataset [120] with the goal of automating transient identification. Each example in the BTSbot dataset corresponds to an 'alert', which will be generated automatically after observations and passed to the community to identify interesting objects and plan potential followup observations. The BTSbot model was trained specifically to identify sources expected to fit BTS's selection criteria.

Composition The BTSbot dataset contains 409,107 transient alerts. Note that while each alert is treated separately, the same transient objects will feature multiple times in the data set. For example, real transients will be harder to identify in earlier alerts when they are fainter and data is noisier but will become more apparent in later alerts as they get brighter. Each transient alert comprises a set of 3 single-band images referred to as the 'science', 'reference', and 'difference' images. The science image shows the latest image of a potential transient event and its surrounding environment, for example the galaxy it is located in. The reference image is an archival image of the same area before the transient was present, while the difference image is the residual between the two which isolates the transient. We collect these views as different channels in a $3{\times}63{\times}63$ image in the same way that other multi-band imaging datasets organise images seen through different filters. The dataset is also multi-modal as each image triplet is accompanied by metadata containing extracted features of the transient source which are informative for classification.

---

[11]https://hsc-release.mtk.nao.ac.jp/archive/filetree/pdr3_dud/ (requires authentication)

Collection Process A description of how the BTSbot dataset was compiled is given in the original paper [120]. We accessed the dataset at the authors' zenodo page[12].

Preprocessing/Cleaning The BTSbot dataset is included within the MULTIMODAL UNIVERSE without any modifications (aside from reformatting).

Uses The BTSbot dataset was used to develop the BTSbot model described in the original paper [120].

Acknowledgements and Terms of Use The BTSbot dataset is distributed under the CC BY 4.0 license. Attribution is given to the authors of the BTSbot paper [120].

A.3.5 JWST: `jwst` ☁️ 🐙 🤗 🥐

Description The JWST dataset contains NIRCam images from several of the first James Webb Space Telescope (JWST) deep field observations: CEERS [52], PRIMER [48], JADES [50] and NGDEEP [14].

Composition Fixed size postage stamps (96×96 pixels) centered on the positions of sources in the photometric catalogs. Each deep survey is an independent instance of the JWST imaging survey and has a different wavelength coverage with 6 or 7 filters in the $\sim 0.9\mu m$ to $\sim 4.4\mu m$ range. The combined dataset contains of the order of 300K images. The list of available filters is: F090W ($0.9\mu m$), F115W ($1.15\mu m$), F150W ($1.5\mu m$), F200W ($2\mu m$), F277W ($2.7\mu m$), F356W ($3.56\mu m$), F444W ($4.4\mu m$).

Collection Process For consistency, we use the public reductions and associated photometric catalogs from the Dawn JWST Archive[13]. Basic details of the NIRCam data reduction are presented in [141]. Images have been reduced with the grizli pipeline[14].

Preprocessing/cleaning Where bands are missing, we store an array of zeros instead to simplify data loading.

Uses Images from these deep JWST surveys have been used in a large number of scientific publications. However, the specific dataset compiled in this work has not been used in any prior publication. Some examples of machine learning works using JWST images are: [77, 142]

Acknowledgements and Terms of Use All the data included in this dataset is of the public domain. The data products presented herein were retrieved from the Dawn JWST Archive (DJA). DJA is an initiative of the Cosmic Dawn Center (DAWN), which is funded by the Danish National Research Foundation under grant DNRF140.

A.4 Time Series

A.4.1 PLAsTiCC: `plasticc` ☁️ 🐙 🤗 🥐

Description The PLAsTiCC dataset is a collection of simulated light curves from the Photometric LSST Astronomical Time-Series Classification Challenge. This diverse dataset contains 14 types of astronomical time-varying objects, simulated using the expected instrument characteristics and survey strategy of the upcoming Legacy Survey of Space and Time [LSST 79] conducted at the Vera C. Rubin Observatory (Figure 4). The dataset includes two overall categories of time-series objects: *transients*, which are short-lived events such as supernovae, and *variable* sources, which are objects with fluctuating brightness such as pulsating stars. Specifically, the dataset includes the following transients: type Ia supernovae (SNIa), SNIax, SNIa-91bg, SNIbc, SNII, superluminous supernovae (SLSN), tidal disruption

---

[12]https://zenodo.org/records/10839691

[13]https://dawn-cph.github.io/dja/index.html

[14]https://github.com/gbrammer/grizli, https://zenodo.org/records/8370018
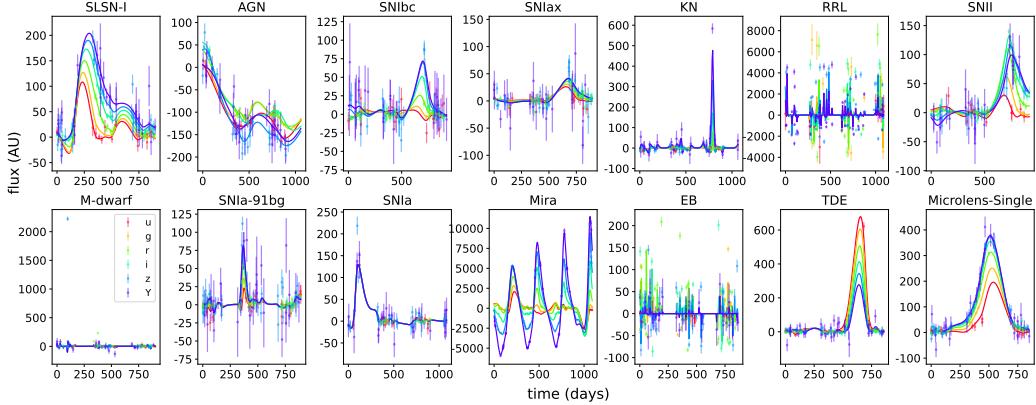
Figure 4: Examples of the 14 object types included in the PLAsTiCC dataset. Observations taken in different wavelength ranges, or *filters*, are shown in different colors. We overplot lines from Gaussian process interpolations of the data for visibility.

events (TDE), and single lens microlensing events (μLens-Single); and the following variable objects: active galactic nuclei (AGN), Mira variables, eclipsing binary systems (EB), and RR Lyrae (RRL); details in Table 7.

Table 7: Breakdown of the PLAsTiCC train and test datasets by object type. The small training set size coupled with the class imbalances make this a challenging classification dataset.

Object Type	Train Samples	Test Samples
SNIa	2,313	1,659,831
SNIa-91bg	208	40,193
SNIax	183	63,664
SNII	1193	1,000,150
SNIbc	484	175,094
SLSN-I	175	35,782
TDE	495	13,555
KN	102	133
AGN	370	101,424
RRL	239	197,155
M-dwarf	981	93,494
EB	924	96,572
Mira	30	1,453
μ-Lens-Single	151	1,303
Total	**7,848**	**3,492,890**

Composition The dataset contains a collection of 3,492,890 time-series, providing in each case several fields: time of observation, flux, flux error, and filter in which the flux is measured. Observations in 6 filters are included (u,g,r,i,z,Y). In addition, each time series is associated with a class representing the type of object, and a redshift for that object.

Collection Process The PLAsTiCC dataset is available on Zenodo at https://zenodo.org/records/2539456. We simply download and reformat the data to match the MULTIMODAL UNIVERSE standard.

Preprocessing/cleaning No further pre-processing is applied to the data.

Uses The PLAsTiCC dataset has been widely used to develop and test supernova classification model [e.g. 116, 20].

Acknowledgements and Terms of Use The dataset is released under the CC BY 4.0 License, it can be cited using the following DOI: 10.5281/zenodo.2539456.

A.4.2 TESS: tess ☁ 🐙 🤗 🥐

Description The NASA Transiting Exoplanet Survey Satellite (TESS)[15] [121] is an all-sky photometric survey observing millions of sources to discover exoplanets and study variable stars. Stars are observed for a period of about 27 days at a time, before the satellite shifts to its next observing field (sector). Stars can be observed multiple times in different sectors over the course of the mission. The observations have a a cadence (sampling time) between 20 sec and 30 min, depending on the mission cycle and data product.

Composition The TESS data set in the MULTIMODAL UNIVERSE currently contains the light curves (brightness over time) for around 1,120,000 stars observed in sectors 56, 58, 60, 63, 64, 67 and 69, but the data can be downloaded for all available sectors by running the respective download scripts. The light curves consist of six fields: `object_id`, `time` (BTJD; Barycenter corrected TESS Julian Date), `flux` (electrons per second) and `flux_error` (electrons per second). In the future, we will also include the much larger set of light curves delivered by the MIT Quick-look Pipeline (QLP)[16] [73, 74, 87, 88], which contains up to a few million light curves per sector.

Collection Process The TESS Science Processing Operations Center (SPOC) pipeline generates light curves [80] for up to 160,000 targets per sector [33] from the shorter cadence Target Pixel Files (TPFs) and the longer cadence Full-Frame Images (FFIs). The data sets were accessed through the Mikulski Archive for Space Telescopes (MAST) servers.

Preprocessing/cleaning We use the Pre-search Data Conditioning (`PDC_SAP`) flux measurements, in which long term systematic trends have been removed using Co-trending Basis Vectors. We have further cleaned the light curves by applying a hard quality flag bitmask that removes data points with possible anomalies as described in the TESS Science Data Products Description Document and the TESS Data Release Notes.

Uses Light curves delivered by TESS, or light curves with similar characteristics, such as those from the NASA Kepler mission [21, 85], have been used in a number of machine learning publications. Light curve classification (e.g., transiting exoplanet, pulsating star or eclipsing binary star) examples are [11, 72, 12, 137, 110, 37], while a stellar parameter estimation example is [111].

Acknowledgements and Terms of Use The data collected by the TESS mission are publicly available from the Mikulski Archive for Space Telescopes (MAST). The light curves in this data set have been released under the CC BY 4.0 License. More detailed information on the data products can be found on MAST: https://archive.stsci.edu/hlsp/tess-spoc. Funding for the TESS mission is provided by NASA's Science Mission directorate.

A.4.3 CfA Datasets: cfa ☁ 🐙 🤗 🥐

Description The Harvard-Smithsonian Center for Astrophysics (CfA) Supernova Group has published light curves of various classes of supernovae observed between 2000-2011.

Composition The four datasets in the MULTIMODAL UNIVERSE are as follows: CfA3 presents light curves of 185 Type Ia supernovae in the UBVRIr'i' bands in the standard photometric system, CfA4 provides light curves of and additional 94 Type Ia supernovae in the u'UBVr'i' bands in the standard photometric system, CfA-SECCSN presents light curves of 64 stripped-envelope core-collapse supernovae in the u'UBVRIr'i'JHK$_s$ bands. Lastly CfA-SNII has light curves of 60 Type II supernovae in the UBVRIu'r'i'JHKs bands.

We present an equivalent breakdown in Table 8.

Each dataset has a row for each supernova with two float fields for sky coordinates (ra, dec), two string fields (object_id and obj_type) describing each supernova's name and classification,

---

[15]https://tess.mit.edu/
[16]https://archive.stsci.edu/hlsp/qlp

Table 8: Breakdown of the CfA Samples.

Source Survey	Shape	Number of samples	Main science
CfA 4 SNe Ia	Variable	341	Type I Supernovae
CfA 3 SNe Ia	Variable	185	Type I Supernovae
CfA Stripped-Core SNe	Variable	249	Stripped-envelope core-collapse Supernovae
CfA SNe II	Variable	400	Type II Supernovae

and a lightcurve field with rows for each observation of that particular supernova. Each row has four subfields. The time subfield contains float values for the modified Julian date of observation (in units of days). The band subfield contains strings describing the band of the observation. The flux and flux_err fields are replaced by mag and mag_err fields, which present the original photometric measurements in the standard photometric system.

Collection Process The compilation of each dataset is described in the corresponding papers [68, 69, 18, 70]. The datasets were accessed from the CfA Supernova Group's website https://lweb.cfa.harvard.edu/supernova/.

Preprocessing/cleaning We have included supplemental data for each object, including RA, Dec, and spectroscopic classifications from the transient name server (TNS)[17], the International Astronomical Union's official repository for reporting astronomical transients.

Uses These datasets were introduced in their respective papers, and have been used in many studies since. In particular, the Type Ia supernovae from CfA3 and CfA4 are included in the Pantheon [127] and Pantheon+ samples [128], the Joint Light-curve Analysis sample [17], the DESSN-3YR sample [25], the Union2 [9] and Union3 samples [123], and many more.

Acknowledgements and Terms of Use This research has made use of the CfA Supernova Archive, which is funded in part by the National Science Foundation through grant AST 0907903. If you make use of CfA data, please include the same acknowledgement as was made here and cite the appropriate references for the data (see the 'Collection Process' paragraph above) in your publication. The CfA3 data is provided under the CC BY 4.0 license and the CfA-SNII data under the CC BY 3.0 license. Refer to the licensing policies of IOPscience[18] and AAS[19] for the CfA4 and CfA-SECCSN datasets.

A.4.4 YSE: yse ☁ 🐙 🤗 🥐

Description A collection of 1975 supernova light curves from the Young Supernova Experiment Data Release 1 (YSE DR1) [7]. An example use of this dataset would be for photometric classification.

Composition We present the data as light curve objects containing time of observation, flux, flux error and band pass filter. Measurements are in $griz$ band pass filters. The supernova classification is included as metadata. Additional metadata includes information on coordinates, redshifts and host galaxy masses.

Collection Process The data is presented in full as shown in the original paper [7] and is accessed through Zenodo[20].

Acknowledgements and Terms of Use 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 [7].

---

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

[18]https://publishingsupport.iopscience.iop.org/creative-commons-licences/

[19]https://journals.aas.org/article-charges-and-copyright/

[20]https://zenodo.org/records/7317476

A.4.5 PS1 SNe Ia: `ps1_sne_ia` ☁️ 🎧 🤗 🥐

Description   A collection of 369 spectroscopically confirmed type Ia supernova light curves from Pan-STARRS1 (PS1 SNe Ia) [127]. This dataset can be used for photometric redshift prediction or light curve inpainting.

Composition   We present the data as light curve objects containing time of observation, flux, flux error and band pass filter. Measurements are in $griz$ band pass filters. The supernova classification is included as metadata. Additional metadata includes information on coordinates, redshifts and host galaxy masses.

Collection Process   The data is collected from the Pantheon+ compilation [128] that also performs a series of selection cuts. The data is accessed through the compilation's GitHub: https://github.com/PantheonPlusSH0ES/DataRelease/tree/main/Pantheon%2B_Data/1_DATA/photometry/Pantheon_PS1MD.

Acknowledgements and Terms of Use   Original content from the work [127] may be used under the terms of the Creative Commons Attribution 4.0 license. If you use this data, please cite the appropriate papers and acknowledge the dataset as stated on the Pan-STARRS webpage[21]. The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

A.4.6 DES Y3 SNe Ia: `des_y3_sne_ia` ☁️ 🎧 🤗 🥐

Description   A collection of 248 spectroscopically confirmed type Ia supernova light curves from the Dark Energy Survey Year 3 (DES Y3 SNe Ia) dataset [24]. This dataset can be used for photometric redshift prediction or light curve inpainting.

Composition   We present the data as light curve objects containing time of observation, flux, flux error and band pass filter. Measurements are in $griz$ band pass filters. The supernova classification is included as metadata. Additional metadata includes information on coordinates, redshifts and host galaxy masses.

Collection Process   The data is collected from the Pantheon+ compilation [128] that also performs a series of selection cuts. The data is accessed through the compilation's GitHub: https://github.com/PantheonPlusSH0ES/DataRelease/tree/main/Pantheon%2B_Data/1_DATA/photometry/DES3YR_DES_COMBINED_TEXT.

Acknowledgements and Terms of Use   Original content from the work [24] may be used under the terms of the Creative Commons Attribution 4.0 license. If you use this data please cite [34, 24] and add the following acknowledgement to your work: This project used public archival data from the Dark Energy Survey (DES). 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 FacilitiesCouncil 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

---

[21]https://outerspace.stsci.edu/display/PANSTARRS/

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, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the OzDES Membership Consortium, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University. Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

A.4.7 SNLS: `snls` ☁️ 🎮 🤗 🥐

Description   A collection of 239 spectroscopically confirmed type Ia supernova light curves from the Supernova Legacy Survey (SNLS) [63]. This dataset can be used for photometric redshift prediction or light curve inpainting.

Composition   We present the data as light curve objects containing time of observation, flux, flux error and band pass filter. Measurements are in $griz$ band pass filters. The supernova classification is included as metadata. Additional metadata includes information on coordinates, redshifts and host galaxy masses.

Collection Process   The data is collected from the Pantheon+ compilation [128] that also performs a series of selection cuts. The data is accessed through the compilation's GitHub: https://github.com/PantheonPlusSH0ES/DataRelease/tree/main/Pantheon% 2B_Data/1_DATA/photometry/JLA2014_SNLS_DS17.

Acknowledgements and Terms of Use   Original content from the work [63] may be used under the terms of the GNU General Public License v3.0. If you use this data, please attribute the original data source [63] appropriately.

A.4.8 Foundation: `foundation` ☁️ 🎮 🤗 🥐

Description   A collection of 180 spectroscopically confirmed type Ia supernova light curves from Foundation Data Release 3 [Foundation DR3 53, 81]. This dataset can be used for photometric redshift prediction or light curve inpainting.

Composition   We present the data as light curve objects containing time of observation, flux, flux error and band pass filter. Measurements are in $griz$ band pass filters. The supernova classification is included as metadata. Additional metadata includes information on coordinates, redshifts and host galaxy masses.

Collection Process   The data is collected from the Pantheon+ compilation [128] that also performs a series of selection cuts. The data is accessed through the compilation's GitHub: https://github.com/PantheonPlusSH0ES/DataRelease/tree/main/Pantheon% 2B_Data/1_DATA/photometry/Foundation_DJ17.

Acknowledgements and Terms of Use Original content from the work [53, 81] may be used under the terms of the Creative Commons Attribution 3.0 license. Any further distribution of this work must include attribution to the author(s), the title of the respective work, journal citation and DOI.

A.4.9 CSP: `csp` ☁ ◯ 🤗 🥐

Description The third data release of the first stage of the Carnegie Supernova Project (CSP) contains light curves of 134 spectroscopically confirmed Type Ia supernovae observed between 2004-2009 [86].

Composition Aside from the use of different bands (ugriBVYJH), the data fields are identical to those of the CfA datasets.

Collection Process The compilation of the CSP data is described in the data release paper [86]. The dataset was accessed from the CSP website's data products page `https://csp.obs.carnegiescience.edu/data`.

Preprocessing/cleaning No pre-processing has been applied.

Uses The CSP dataset has been used in many studies including those mentioned in the CfA section, as well as several CSP studies [e.g. 32, 115, 55, 140].

Acknowledgements and Terms of Use The CSP dataset is provided under the CC BY 4.0 license. If you use CSP, please cite the respective data release paper [86] in your paper.

A.4.10 Swift SNe Ia: `swift_sne_ia` ☁ ◯ 🤗 🥐

Description A collection of 117 spectroscopically confirmed type Ia supernova light curves from the Swift Type Ia Supernova (Swift SN Ia) dataset [27]. This dataset can be used for photometric redshift prediction or light curve inpainting.

Composition We present the data as light curve objects containing time of observation, flux, flux error and band pass filter. Measurements are in UV_{w1}, u, b, UV_{w2}, v and UV_{m2} band pass filters. The supernova classification is included as metadata. Additional metadata includes information on coordinates, redshifts and host galaxy masses.

Collection Process The data is collected from the Pantheon+ compilation [128] that also performs a series of selection cuts. The data is accessed through the compilation's GitHub: `https://github.com/PantheonPlusSH0ES/DataRelease/tree/main/Pantheon%2B_Data/1_DATA/photometry/SWIFT`.

Acknowledgements and Terms of Use Original content from the work [27] may be used under the terms of the GNU Lesser General Public License.

A.5 Spectra Datasets

A.5.1 *Gaia* BP/RP: `gaia` ☁ ◯ 🤗 🥐

Refer to Subsection A.7.1.

A.5.2 SDSS: `sdss` ☁ ◯ 🤗 🥐

Description The Sloan Digital Sky Survey (SDSS; [153]) is a comprehensive astronomical survey that has mapped the sky using a dedicated 2.5-meter wide-angle optical telescope located at the Apache Point Observatory in New Mexico, USA. SDSS has collected detailed images and spectra of millions of celestial objects, covering a significant portion of the sky in multiple bands. In this dataset we collect data from all SDSS optical spectra including SDSS Legacy survey, SEGUE-1, SEGUE-2, BOSS, and eBOSS surveys [153, 49, 19].

Composition   The SDSS dataset includes spectra spanning a wavelength range from $3,800$ to $9,200$ for spectra acquired with the SDSS spectrograph and $3,650$-$10,400$ for data from the BOSS spectrograph. In addition, for various processing reasons, not all spectra have the same lengths. Each spectrum comes with associated metadata, including flux measurements, wavelength values, and inverse variance (`ivar`) for the flux. Additionally, pixel-level masks indicate potential quality issues. The dataset contains a total of roughly 4 million spectra.

We apply a minimal selection on this data by only imposing the following cuts:

- Avoid duplicated spectra for the same object: SPECPRIMARY = 1
- Only good plates: PLATEQUALITY = 'good'
- Only science targets (no sky, calibration, etc.): TARGETTYPE = 'science'

Collection Process   We use the publicly accessible data from the SDSS 18th data release (SDSS DR18) [22], and systematically download all processed spectra files for spectra that pass the selection cut described above.

Preprocessing/Cleaning   No further cuts of processing steps are performed.

Uses   SDSS spectra have been used in numerous scientific publications. Notable examples involving machine learning include building data-driven representations of galaxy spectra and looking for outliers [104].

Acknowledgements and Terms of Use   All SDSS data released in public data releases is considered to be in the public domain. When using the data included in this compilation, the following acknowledgements should be included:

- **SDSS and SEGUE-1 samples**: Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.
  The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

- **BOSS and SEGUE-2 samples**: Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/.
  SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New

---

[22]https://www.sdss.org/dr18/data_access/bulk/

York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

A.5.3 DESI: `desi` ☁ 🐙 🤗 🥐

Description   The Dark Energy Spectroscopic Instrument (DESI; [41]) is designed to measure the effect of dark energy on the expansion of the universe by mapping the distributions of galaxies in three dimensions. To do so, DESI uses a 4-meter telescope at Kitt Peak National Observatory to collect spectra from millions of galaxies and quasars over a five-year period. This experiment is still on-going and most of the data is not yet public. In this dataset, we specifically make use of the DESI Early Data Release (EDR) [40] which contains 1% survey of the final survey.

Composition   The DESI EDR sample includes spectra with a fixed wavelength range from $3,600$ to $9,800$ and 7081 pixels. Each spectrum is accompanied by flux measurements, wavelength values, and inverse variance (`ivar`) for the flux. To select the sample including in this dataset we only apply the following minimal set of selection cuts:

- Only use data from the one percent survey: SURVEY = 'sv3'

- Only use the primary spectrum for each object: SV_PRIMARY is true

- Only use targets (ignore sky and others): OBJTYPE = 'TGT'

- Only use fibers with good status: COADD_FIBERSTATUS = 0

This results in a set of approximately 1M spectra of galaxies, quasars, and stars.

Collection Process   We systematically download all spectra files for the SV3 sample through GLOBUS from the DESI EDR archive https://data.desi.lbl.gov/doc/releases/.

Preprocessing/Cleaning   We apply no further pre-processing beyond the set of cuts described above.

Uses   The DESI EDR being still very recent, few machine learning works have been developed on this dataset. Still, a few examples are the `AstroClip` multimodal representation learning model [112], and the detection of outliers with the `spender` spectrum auto-encoding model [96].

A.5.4   APOGEE: apogee ☁ 🐙 🤗 🥐

Description   Apache Point Observatory Galactic Evolution Experiment (APOGEE; [19, 2]) within the Sloan Digital Sky Survey (SDSS) is a high-resolution ($R \sim 22,000$), high signal-to-noise (100 per pixel typically) stellar spectroscopic survey with 2.5-m telescopes in both northern and southern hemisphere in the near infrared H-band wavelength region [102, 151].

Composition   Fixed size APOGEE stellar spectra (1×7514) spanning a wavelength range of 1.5 to $1.7\mu m$. Each stellar spectrum comes with both flux and continuum normalized flux (i.e. flux with continuum removed), its associated λ vector representing the corresponding wavelength of each pixel, as well the inverse variance (`ivar`) for both flux and continuum normalized flux and a pixel-level mask indicating potential quality issues. The dataset contains a total of roughly 700,000 stellar spectra. Surface temperature T_{eff}, surface gravity $\log g$, overall metallicity $[M/H]$, overall alpha abundance $[\alpha/H]$ and radial velocity are derived using the APOGEE Stellar Parameter and Chemical Abundances Pipeline [61] from the APOGEE spectra.

Collection Process   We use the publicly accessible data from the APOGEE 17th data release (APOGEE DR17) catalog [23].

Preprocessing/cleaning   The dataset only include APOGEE spectra that has signal-to-noise (SNR) greater than 30, where duplicated stars in the public data released are removed due to some stars being observed in multiple programs multiple times. Preprocessing of the raw data are done internally hence minimal processing are needed for the public data release which we are using.

---

[23]https://www.sdss4.org/dr17/irspec/

Uses   The public APOGEE DR17 dataset has been used in a number of scientific publications in both stellar and Galactic astronomy, as well as in machine learning specific applications.

A.5.5   GALAH: `galah` ☁ 🖸 🤗 🥐

Description   Galactic Archaeology with HERMES (GALAH; [39]) is an Australian-led survey aimed at obtaining high-quality one-dimensional stellar spectra for deriving stellar atmospheric parameters and elemental abundances.

Composition   The dataset includes spectra and derived stellar parameters such as effective temperature and several elemental abundances. The publicly available Data Release 3 (DR3; [28]) contains spectra for 325,518 objects after applying recommended quality cuts.

Collection Process   The GALAH survey utilizes the HERMES spectrograph on the Anglo-Australian Telescope to collect high-resolution spectra. The observations are processed and made publicly available through periodic data releases, with DR3 being the latest release used for this dataset. We download the data through https://www.galah-survey.org/dr3/the_spectra/

Preprocessing/Cleaning   We apply recommended cuts on spectra quality[24] to ensure the reliability of the data, resulting in a dataset with high-quality spectra and accurate stellar parameters.

Uses   The GALAH dataset is used in studies of stellar populations, Galactic archaeology, and the chemical evolution of the Milky Way.

A.5.6   VIPERS: `vipers` ☁ 🖸 🤗 🥐

Description   The "VIMOS Public Extragalactic Redshift Survey" (VIPERS) [126] dataset contains measured optical spectra of galaxies in the $0.5 < z < 1.0$ range.

Composition   Fixed size spectra (1×557) spanning a wavelength range of 5514 Å to 9484 Å, taken at the center of each galaxy. Each spectrum comes with its associated λ vector,

---

[24]https://www.galah-survey.org/dr3/using_the_data/

as well the inverse variance (`ivar`) and a mask of each flux measurement in the spectrum. The dataset contins a total of roughly 90,000 galaxies with red magnitude I(AB) brighter than 22.5 over an overall area of nearly 24 square degrees.

Collection Process For consistency, we use the publicly accessible data from the second VIPERS public data release (PDR-2) catalog[25].

Preprocessing/cleaning The VIPERS spectra flux is natively stored in units of $\mathrm{erg\ cm^{-2}s^{-1}\mathring{A}^{-1}}$, and correspondingly the noise is stored in $\mathrm{erg\ cm^{-2}s^{-1}\mathring{A}^{-1}}$ and the wavelength in angstroms (Å). To be consistent with the other spectra datasets, we normalize the spectra fluxes to $10^{-17}\mathrm{erg\ cm^{-2}s^{-1}\mathring{A}^{-1}}$. Correspondingly, we convert the spectra noise to inverse variance, also in units $10^{-17}\mathrm{erg\ cm^{-2}s^{-1}\mathring{A}^{-1}}$.

Uses This dataset has been used in a number of scientific publications, as well as in machine learning specific applications, including source identification with SVMs [103] and galaxy classification [129] with unsupervised methods.

Acknowledgements and Terms of Use This paper uses data from the VIMOS Public Extragalactic Redshift Survey (VIPERS). VIPERS has been performed using the ESO Very Large Telescope, under the "Large Programme" 182.A-0886. The participating institutions and funding agencies are listed at http://vipers.inaf.it.

A.5.7 Chandra: `chandra` 🟣 🔗 🤗 🥐
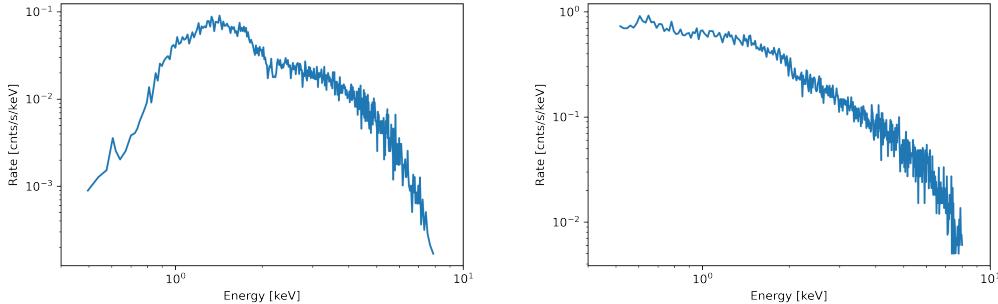
Figure 5: Examples of Chandra X-ray spectra included in MULTIMODAL UNIVERSE. Spectra of sources 2CXO J002029.1+591651 (left) and 2CXO J023437.8-084716 (right) are shown.

Description The Chandra X-ray Observatory is one of NASA's flagship observatories, studying the X-ray emission from the most extreme environments in the Universe, including the surroundings of stellar-mass black holes and super massive black holes (SMBHs), supernova remnants, and violent coronal mass ejections from young stars [150]. The observatory records these observations using two instruments, the Advanced CCD Imaging Spectrometer (ACIS), and the High Resolution Camera (HRC), with CCD chips arranged to cover the telescope's field of view. The Chandra processing pipeline produces the individual X-ray photon recordings for each observation in the form of event files, which can be thought of as multivariate time series of the photon's energies and spatial location on the detector, with the energies covering the range between approximately 0.5 keV and 8 keV. MULTIMODAL UNIVERSE contains ACIS X-ray spectra of astrophysical sources, derived from the event files through the proper application of the instrument's spectral response.

X-ray spectra (photon energy versus count rate) from astrophysical sources detected by the Chandra X-ray Observatory in the energy range between 0.5keV and 8keV, binned at 10 counts per energy bin. The spectral files are downloaded from the Chandra Source Catalog

---

(CSC) version 2.1 server. (https://cxc.cfa.harvard.edu/csc/). We provide all CSC spectra that have at least 40 total counts and a detection signal to noise ratio of 4 or larger.

Composition Each entry consist of the source identifiers and coordinates, as well as arrays that contain the low and high limits of the energy bins, the center of the bin, the measured count rate per bin, and the error in that count rate. In addition, we provide features that characterize the observed source: the aperture photometry X-ray flux in physical units (ergs s^{-1} cm^{-2}), the flux estimated using a blackbody fit to the spectrum, the flux significance, or signal-to-noise ratio, three hardness ratios which characterize the spectral shape, and measures of the flux variability as a function of time, both as a probability and as an index that quantifies the statistical significance of the variability.

Collection Process The data was originally collected as part of the Chandra Source Catalog project, which detects and characterizes all sources serendipitously detected by Chandra, and produces both tables and data products for those sources, including spectral files, storing them as a database in a dedicated server hosted at the Smithsonian Astrophysical Observatory. For MULTIMODAL UNIVERSE, we have accessed these files using International Virtual Observatory (IVOA) interfaces and HTTP requests via command line interface.

Preprocessing/cleaning The observatory records these observations using two instruments, the Advanced CCD Imaging Spectrometer (ACIS), and the High Resolution Camera (HRC), with CCD chips arranged to cover the telescope's field of view. The Chandra processing pipeline produces the individual X-ray photon recordings for each observation in the form of event files, which can be thought of as multivariate time series of the photon's energies and spatial location on the detector, with the energies covering the range between approximately 0.5 keV and 8 keV. MULTIMODAL UNIVERSE contains ACIS X-ray spectra of astrophysical sources, derived from the event files through the proper application of the instrument's spectral response, and processed using the Chandra Interactive Analysis of Observations (CIAO) software (https://cxc.cfa.harvard.edu/ciao/), in particular the Sherpa package.

Uses Chandra studies the X-ray emission from the most extreme environments in the Universe, including the surroundings of stellar-mass black holes and super massive black holes (SMBHs), supernova remnants, and violent coronal mass ejections from young stars [150]. Many of the serendipitous sources compiled in the CSC lack and astrophysical classification. A similar dataset to the one presented here has been used to perform supervised machine learning classification of Chandra sources [67].

Acknowledgements and Terms of Use This research has made use of data obtained from the Chandra Source Catalog, provided by the Chandra X-ray Center (CXC). The CSC (doi:10.25574/csc2) is publicly available. The current version is CSC 2.1. CXC provides up to date guidance[26] on acknowledging and citing their work, which we ask you do if you use Chandra data.

A.6 HyperSpectral Imaging Datasets

A.6.1 MaNGA `manga` ☁ ⦿ 🤗 🐚

Description The SDSS-IV MaNGA survey[29] is a wide-field, optical, IFU survey of 10,000 nearby galaxies. MaNGA's goal is to understand the "life history" of present day galaxies from imprinted clues of their birth and assembly, through their ongoing growth via star formation and merging, to their death from quenching at late times.

The primary MaNGA data products are composed of 3-D calibrated data cubes produced by the DRP[89] and 2-D maps of derived quantities, such as emission line fluxes, gas and stellar kinematics, and stellar population properties, produced by the DAP[149] from those cubes. The 3-D data cubes are constructed from a few tens to a few thousands of individual spectra that have been combined onto a regular grid. The 2-D maps of derived quantities

---

[26]https://cxc.cfa.harvard.edu/csc2.1/cite.html
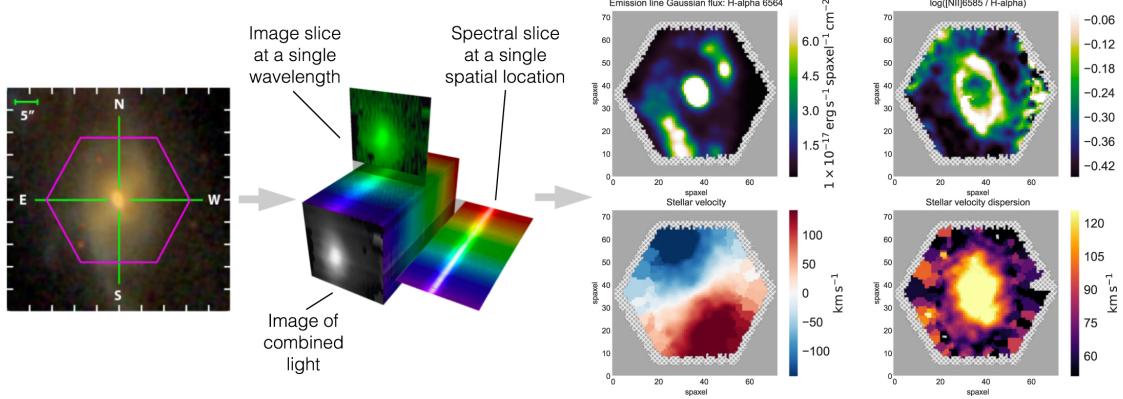
Figure 6: Left: gri image of MaNGA 1-596678 with the IFU field of view shown in purple. Middle: IFU observations produce three-dimensional data cubes, with one spectral and two spatial dimensions (credit: Stephen Todd and Douglas Pierce-Price; IFS Wiki http://ifs.wikidot.com). Right: spectral analysis of individual spaxels produces hundreds of two-dimensional maps for each galaxy spanning a wide range of physical properties. The four example maps shown for 1-596678 are flux (top left), log([N ii] 6585/) flux (top right), stellar velocity (bottom left), and stellar velocity dispersion corrected for instrumental broadening (bottom right). Original figure from the SDSS MaNGA Marvin paper,[35].

are constructed by analyzing individual or binned groups of spaxels and constructing maps of the quantities at the relevant on-sky location.

This dataset contains information pulled from the MaNGA DRP LOGCUBE[27] files, as well as the DAP "HYB10-MILESHC-MASTARSSP" MAPS[28] files.

Composition   Each entry in the dataset is a single MaNGA galaxy observation, identified by its unique id designation, the "plate-IFU", e.g. "8485-1901", with information aggregated from both the IFU data cubes and the derived analsyis maps, for that target. For each galaxy, the following features are provided:

- **object_id**: the MaNGA plate-IFU
- **ra**: the Right Ascension of the target
- **dec**: the Declination of the target
- **healpix**: the healpix id
- **z**: the target redshift
- **spaxel_size**: the size of the spaxel (spatial pixel)
- **spaxel_size_units**: the units of the spaxel size
- **spaxels** a list of spaxels, with properties
- **images**: a list of reconstructed images, with properties
- **maps**: a list of derived maps, with properties

Each **spaxel** item has the following features:

- **flux**: the flux spectrum array in the spaxel
- **ivar**: the inverse variance array in the spaxel
- **mask**: the pixel quality mask array in the spaxel
- **lsf**: the line-spread-function array in the spaxel

---

[27]https://www.sdss4.org/dr17/manga/manga-data/data-model/#CUBEFiles
[28]https://www.sdss4.org/dr17/manga/manga-data/data-model/#DataAnalysisPipelineOutput

- **lambda**: the wavelength array, same for all spaxels
- **x**: the x-array index spaxel coordinate
- **y**: the y-array index spaxel coordinate
- **spaxel_idx**: the 1d array index of the numpy.raveled spaxel
- **flux_units**: the physical units of the flux array
- **lambda_units**: the physical units of the wavelength array
- **skycoo_x**: the sky x offset in RA, in arcsec from galaxy center
- **skycoo_y**: the sky y offset in DEC, in arcsec from galaxy center
- **ellcoo_r**: the elliptical polar coordinate radial offset, from galaxy center, R in arcsec
- **ellcoo_rre**: the elliptical polar coordinate radial offset, from galaxy center, as R/Re
- **ellcoo_rkpc**: the elliptical polar coordinate radial offset, from galaxy center, R in h-1 kpc
- **ellcoo_theta**: the elliptical polar coordinate angle offset, from galaxy center, theta in degrees
- **skycoo_units**: the physical units of skycoo_x/y
- **ellcoo_r_units**: the physical units of ellcoo_r
- **ellcoo_rre_units**: the physical units of ellcoo_rre
- **ellcoo_rkpc_units**: the physical units of ellcoo_rkpc
- **ellcoo_theta_units**: the physical units of ellcoo_theta

Each **images** item has the following features:

- **filter**: the image band or filter, e.g. griz
- **array**: the 2d array pixel values
- **array_units**: the physical units of the data
- **psf**: the reconstructed PSF pixel values
- **psf_units**: the physical units of the data
- **scale**: the image pixel scale
- **scale_units**: the units of the image scale

Each **maps** item has the following features:

- **group**: the original MaNGA DAP extension, e.g. 'emline_gflux'
- **label**: the full map label, made from the HDU extension name + channel, e.g. 'emline_gflux_ha_6564'.
- **array**: the 2d array of pixel values
- **ivar**: the 2d array of inverse variance values
- **mask**: the 2d array of pixel quality mask values
- **array_units**: the physical units of the map data

Collection Process We use the publicly accessible data for MaNGA from the SDSS-IV 17th data release (DR17)[29].

Preprocessing/cleaning During processing all IFU cubes, and maps have been resized to the same spatial grid of 96 x 96, with zero-padded elements added around the edges of the data. Thus all cubes have shape 96 x 96 x 4563 (42,882,048 pixels), and all maps have shape 96 x 96 (9,216 pixels).

---

[29]https://www.sdss4.org/dr17/manga/

Uses   The public MaNGA DR17 dataset has been used in a number of scientific publications in extra-galactic astronomy, as well as in machine learning specific applications.

A.7   Tabular Datasets

A.7.1   *Gaia* DR3: `gaia` ☁️ 🐙 🤗 🥐

Description   The *Gaia* mission is primarily intended for very precisely measuring astrometry of stars in the Milky Way (MW) from space at an unprecedented scale. Data Release 3 (DR3), in addition to astrometric measurements (position on the sky, parallaxes, proper motions, as well as associated uncertainties), also provides, for example: photometry; physical parameters (stellar surface properties like surface gravity and effective temperature); epoch photometry for variables; orbital elements for binary systems; high-resolution spectra from the radial velocity (RV) spectrometer and low-resolution spectra from the blue and red photometers (BP and RP).

Composition   The dataset currently contains the following:

- Low-resolution optical spectra [38]: the spectra are stored as 110 coefficients of Gauss-Hermite polynomials (55 for BP, and 55 for RP). These can be sampled to an arbitrary wavelength grid using the GaiaXPy tool[30].

- Astrometry: on-sky position, parallax, and proper motions, along with the full covariance matrix.

- Photometry: mean G, G_{BP}, and G_{RP} magnitudes and fluxes, along with errors.

- Photometrically-estimated parameters: stellar parameters (e.g., distance, surface gravity, metallicity, surface temperature) and lower/upper bounds derived from BP/RP spectra, apparent G magnitudes, and parallax.

- Radial velocities: RVs, errors, and stellar parameters of the template used to derive the RVs.

- Corrections and flags: commonly used variables used to make, for example, radial velocity and parallax zero-point corrections.

---

[30]https://www.cosmos.esa.int/web/gaia/gaiaxpy

Collection Process We use the publicly accessible DR3 data through the mirror hosted by the Flatiron Institute[31].

Preprocessing/cleaning We select the subset of 220M stars in DR3 for which *Gaia* mean BP/RP spectra are provided. For these stars, we provide the contents described above. In the future we will include the entire dataset of the nearly 2B sources measured in DR3.

Uses DR3 data has already been used extensively in MW science and in machine learning applications. For example, using DR3, [57] explored and identified non-axisymmetric features in the disc, [58] have created a detailed chemodynamical map of the MW, and [94] have trained a transformer-based model that can generate stellar spectra, estimate physical parameters, and inpaint unobserved spectral regions.

Acknowledgements and Terms of Use This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. The Gaia data are open and free to use, provided credit is given to 'ESA/Gaia/DPAC'. If you use Gaia DR3 data in your research, please acknowledge it as above.

A.7.2 Galaxy10 DECaLS `gz10` ☁️ 🎧 🤗 🥐

Description Galaxy Zoo (https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/) is a long running citizen science (crowd-sourcing) effort recruiting volunteers to annotate images of galaxies online according to the galaxies structure such as whether or not they have spiral arms or bulges [99]. These structures are useful, and traditionally expensive, tracers of underlying physics such as galaxy evolution [124]. Galaxy10 DECaLS [93] represents a subset of the Galaxy Zoo data, where the original decision tree for Galaxy Zoo is split into ten classes described below.
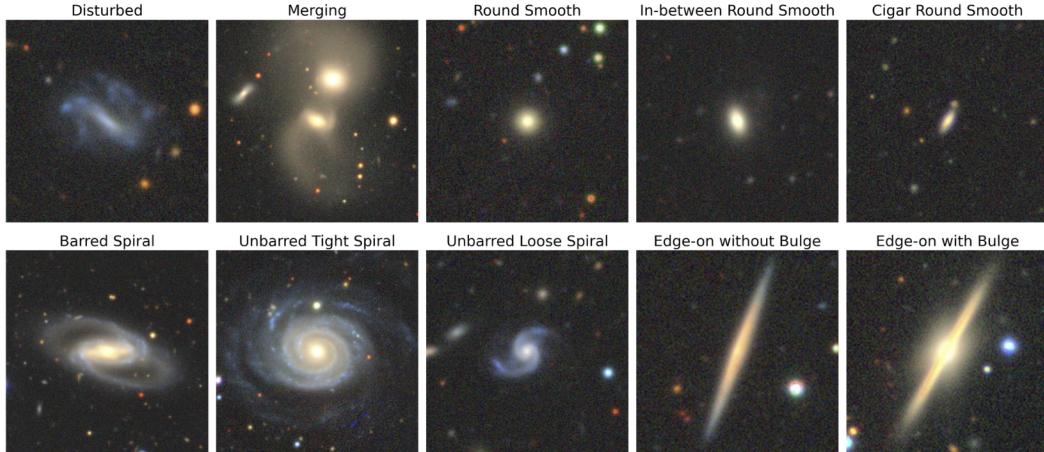
Figure 7: Examples of each class of the GalaxyZoo10 DECaLS data set with their RGB composite image. Adapted from original figure from [93].

Composition [92] The data set is comprised of the images which volunteers saw and an aggregation of their resulting classifications into 10 subsets class labels for the objects spanning 17,736 samples split among the following classes:

1. Disturbed Galaxies (1,081)

---

2. Merging Galaxies (1,853)

3. Round Smooth Galaxies (2,645)

4. In-between Round Smooth Galaxies (2,027)

5. Cigar Shaped Smooth Galaxies (334)

6. Barred Spiral Galaxies (2,043)

7. Unbarred Tight Spiral Galaxies (1,829)

8. Unbarred Loose Spiral Galaxies (2,628)

9. Edge-on Galaxies without Bulge (1,423)

10. Edge-on Galaxies with Bulge (1,873)

The data can be loaded in two broad configurations: Firstly, as just the labels with auxiliary data, or the labels with the respective `uint8` PNGs (3x256x256). The auxiliary data includes the right ascension `ra`, declination `dec`, `redshift`, and Multimodal Universe specific object id `object_id`. Additionally, if the images are loaded, the pixel scale of the images is included as `rgb_pixel_scale`.

Collection Process The images presented in this dataset are RGB composites [145] of data collected by the DESI legacy imaging survey [44] and are analogous to what the volunteers were shown while assigning the labels to the galaxies. The collection process for the labels of the underlying data are detailed in [145, 97].

Preprocessing/cleaning There are a number of selection effects made on what data received labels, see [145]. The labels, constructed from the original decision tree (https://data.galaxyzoo.org/gz_trees/gz_trees.html) are collated to select confident and clearly distinguishable classes [92] for simplicity. Only confident labels are maintained as this dataset was intended as a benchmark dataset with minimal or no label noise. Several baseline models for this benchmark are provided, see subsection B.1 Additionally, the images are not the full scientific measurement of the respective object, as done by DESI [44], but rather a RGB composite designed to highlight the respective structure. These images are what the volunteers would have seen whilst labelling the respective objects.

Uses A number of publications have made use of this data, presented on the homepage (https://astronn.readthedocs.io/en/latest/galaxy10.html#some-papers-that-used-galaxy-10). These efforts are largely bench marking the efficacy of various approaches for galaxy morphology classification, as the labels in this dataset are clean and relatively simple in comparison to the raw labels of the decision trees, such as those presented in [145].

Acknowledgements and Terms of Use Galaxy Zoo is described in [99], the GalaxyZoo Data Release 2 is described in [98], Galaxy Zoo DECaLS Campaign is described in Galaxy Zoo is described in [99], the GalaxyZoo Data Release 2 is described in [98], Galaxy Zoo DECaLS Campaign is described in [145], DESI Legacy Imaging Surveys is described in [42].

The Legacy Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS; Proposal ID #2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing-Arizona Sky Survey (BASS; NOAO Prop. ID #2015A-0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall z-band Legacy Survey (MzLS; Prop. ID #2016A-0453; PI: Arjun Dey). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF's NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. The Legacy Surveys project is honored to be permitted to conduct astronomical research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

A.7.3 PROVABGS `desi_provabgs` ☁️ 🐙 🤗 🐋

Description The PRObabilistic Value-Added Bright Galaxy Survey (PROVABGS) [65] dataset provides measurements of galaxy properties, such as log stellar mass ($\log M_*$), star

formation rate (SFR), mass-weighted stellar metallicity (Z_{MW}), and mass-weighted stellar age ($t_{age,MW}$) for galaxy spectra in the DESI Bright Galaxy Survey (BGS). These are inferred using a full Bayesian inference framework and state-of-the-art Spectral Energy Distribution (SED) modeling of the DESI spectra. At present, as the DESI ruvey is still on-going, the PROVABGS catalog reports these properties for the early data release (EDR) of the DESI sample.

Composition The best-fit parameters of the PROVABGS SED modeling for $\log M_*$, SFR, Z_{MW}, $t_{age,MW}$. Additionally, it also includes the 100×13 array of 100 samples drawn from the 13-dimensional PROVABGS posterior. The 13 parameters correspond to $\log(M_{form})$[32], β_1, β_2, β_3, β_4, f_{burst}, t_{burst}, γ_1, γ_2, τ_{BC}, τ_{ISM}, n_{dust}, f_{fiber}, from which the above properties are derived. We refer the reader to the original PROVABGS documentation[33] for the full details. Overall, the PROVABGS modeling of the DESI EDR contains physical properties for roughly $221,000$ galaxy spectra.

Collection Process We use the reported PROVABGS posteriors from the publicly available DESI EDR BGS PROVABGS catalog[34].

Preprocessing/cleaning We extract the best fit parameters of the PROVABGS SED modeling for $\log M_*$, SFR, Z_{MW}, $t_{age,MW}$ using the PROVABGS SPS model with non-parametric star formation and metallicity histories and flexible dust attenuation model.

Uses The PROVABGS catalog has been used for physical property estimation from both images and spectra in both self-supervised learning and supervised learning contexts [112].

Acknowledgements and Terms of Use All of the data included in this dataset is of the public domain. We are grateful to the PROVABGS team for making the data available to the community. This research used data obtained with the Dark Energy Spectroscopic Instrument (DESI). DESI construction and operations is managed by the Lawrence Berkeley National Laboratory. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High-Energy Physics, under Contract No. DE–AC02–05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico (CONACYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: www.desi.lbl.gov/collaborating-institutions. The DESI collaboration is honored to be permitted to conduct scientific research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation. 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 U.S. National Science Foundation, the U.S. Department of Energy, or any of the listed funding agencies.

A.8 Maintenance

To ensure the continued evolution and relevance of MULTIMODAL UNIVERSE, the project is organized as an open and collaborative effort. Contributions from researchers are actively encouraged, with a well-defined process for submitting new data, benchmarks, and improvements via GitHub, as described in the project's CONTRIBUTING document. A dedicated team of maintainers oversees the integration of these contributions, ensuring that the dataset remains up-to-date and continues to meet the needs of the scientific and machine learning communities. This framework will enable regular updates to incorporate new data from

---

[32]Note that $\log(M_{form})$ is different from log(stellar mass).

[33]https://data.desi.lbl.gov/doc/releases/edr/vac/provabgs/

[34]https://data.desi.lbl.gov/public/edr/vac/edr/provabgs/v1.0/README.md

ongoing and future astronomical surveys, as well as periodic evaluations and improvements based on feedback from the research community.

A.9   Author statement

The authors of this scientific paper bear full responsibility for any violation of rights that may arise from the collection of the data included in this research.

B Details on Benchmark tasks

B.1 Galaxy10 DECaLS Morphology Classification

For the Galaxy10 DECaLS morphology classification baselines, we train all models using an 80/20 train-test split on the GalaxyZoo10 DECaLS classification labels and the corresponding Legacy Survey RGB-converted images. We train three canonical computer vision baseline networks - ResNet-18 [66], EfficientNetB0 [136], and DenseNet121 [75] - to classify these images. As they are already range compressed in their conversion to RGB images, we do not apply any range compression. To prevent overfitting, we use data augmentation: random horizontal and vertical flips, random gaussian blurring, random affine translation, and color jittering[35]. We train all models using a cross-entropy loss on the training dataset with the Adam optimizer [84]. Training is performed for 10 epochs on a single A100 GPU for each model with a batch size of 128 and a learning rate of $\lambda = 10^{-4}$. We report the Top-1 accuracy on the held-out validation test dataset.

B.2 Physical Property Estimation

For the physical property inference baselines, we train the image and spectrum models using an 80/20 train-test split on the cross-match between the PROVABGS dataset and the Legacy Survey imaging datasets and the DESI spectrum datasets respectively.

For the image models, we use canonical computer vision baseline networks - ResNet-18 [66], EfficientNetB0 [136], and DenseNet121 [75] - to regress the PROVABGS galaxy properties from the Legacy Survey images. Due to the high dynamic range of the images, we use a range compression scheme to reduce the dynamic range for better optimization. Specifically, we apply *arcsinh* range compression and some additional clamping to the images. We also use data augmentation - random horizontal and vertical rotations and random gaussian blurring - to reduce overfitting. For the spectrum model, we use a convolutional+attention network with the same architecture as the encoder from the state-of-the-art spectrum encoder from [104]. For the photometry model, we use a simple 5-layer MLP with $d = 512$ hidden dimension. Additionally, we Z-score the photometry to normalize the data.

We train all models using a mean-squared error (MSE) loss on the training dataset with the Adam optimizer [84]. We Z-score the PROVABGS galaxy properties. Additionally, we convert metallicity (Z_{MW}) to log metallicity and star-formation rate (SFR) to $sSFR = \log SFR - \log Z_{MW}$. We train all models for 25 epochs on a single A100 GPU with a batch size of 256 and a learning rate of $\lambda = 10^{-4}$. We report the R^2 performance of all models on the held-out test datasets.

B.3 PLAsTiCC Photometric Classification and Redshift Estimation

For the PLAsTiCC baseline model, we use an encoder-only Informer model [154] with 8 encoder layers of 12 attention heads each. The model hidden dimension was chosen to be 768 and the layer MLPs have hidden dimension 256. Due to the 2-dimensional position data (each element of the time-series has an associated time and photometric band/wavelength) and irregular sampling of our dataset, we train a positional encoding based on learnable Fourier features following [95]. We also select a random window of length 300 from each example (and zero-pad examples with fewer than 300 observations) to produce inputs of uniform shape. We first pretrain with a masked autoencoding objective:

$$\mathcal{L}_{\mathrm{MAE}}(\phi) = \mathbb{E}_{x \sim P_U, x' \sim \mathcal{A}_{\mathrm{pre}}(\cdot|x)}[(\phi(x') - x)^2] \tag{1}$$

Here, P_U is the distribution of unlabeled data that we pretrain with, $\mathcal{A}_{\mathrm{pre}}$ is the distribution of augmented (masked) inputs, and ϕ is the model. We perform pretraining by randomly masking 60% of observations of each time-series, a batch size of 256, and learning rate 1e-4

---

[35]In other astrophysical contexts color jittering tends to be avoided due to its potential to corrupt valuable astrophysical information present in the galaxy colors; however, as we are only concerned with galaxy morphology in this application, we apply it here.

(selected from 1e-3 ∼ 1e-6) for 75,000 steps. We finetune the pretrained model with linear probing for 20,000 steps and learning rate 1e-4, then fine-tuning for 10,000 steps at learning rate of 4e-5. The redshift estimation task uses LP learning rate of 5e-4 and FT learning rate of 1e-4. We decrease the learning rate per step with a linear scheduler.

B.4 YSE Photometric classification

We demonstrate an implementation of the `snmachine` package [100, 8] on the YSE data set with classes mapped to SNe Ia, SNe II, SNe Ib/c, or other. `snmachine` offers feature extraction methods based on SN Ia templates [64], parametric models [108, 82], and wavelet analysis [90], along with wrappers for several classifiers from `scikit-learn` [113]. We find that wavelet-based feature extraction combined with the Random Forest classifier, with or without the ensemble AdaBoost classifier, achieves a AUC of 0.90in SN Ia one-vs-rest classification.

B.5 BTSbot Candidate Identification

We report on the performance of BTSbot - a multimodal convolutional neural network which combines images of potential new transient detections with associated metadata - on the BTSbot training sample [119, 120]. Applied to a validation data set comprising 10 per cent of the entire data set, BTSbot achieves an AUC of 0.985 [120]. Given that the BTSbot training set includes multiple alerts for the same transient objects as they evolve, it is important to ensure that train/test splits are based on unique transient objects rather than just unique transient alerts, as in the public data set. While the overall AUC is an important measure, it is also valuable to consider the performance on an object-by-object basis rather than just considering all the individual alerts to determine how many unique transient objects are being identified. Early classification is also of value for this data set to allow transients to be targeted for further observations, therefore a model which is able to accurately identify real transient candidates as early as possible in their light curves will be of great benefit.

B.6 Contrastive Image-Spectrum Pretraining

We replicate the results of `AstroCLIP` method [112] using the Multimodal Universe cross-matched DESI and Legacy Surveys. We split the dataset using an 80/20 train-test split. We use the same hyperparameter and training suite as in [112] for training the model on the training suite. We then evaluate model performance by evaluating the zero-shot performance of a k-NN algorithm with $k = 16$ nearest neighbors on the cross-matched DESI and Legacy Survey and PROVABGS dataset. We report the R^2 performance of the model on the held-out test set. Note that the model has never seen the test set during either pretraining or zero-shot "training".

B.7 From MaNGA Hyperspectral Datacubes to Physical Parameter Images

We demonstrate how 3D hyperspectral datacubes of galaxies can be used to infer 2D image "maps" corresponding to physical parameters, a process typically performed through complete spectrum analysis at each spaxel. In this use case, we train the canonical computer vision baseline network, U-Net [122], on a subset of ∼ 1100 galaxy datacubes from the MaNGA dataset (with a shape of 96 x 96 x 4563) to infer the stellar velocity dispersion map of each galaxy (96 x 96). We split the dataset using an 90/10 train-test split.

The model is trained using a mean absolute error (MAE) loss on the training dataset with the Adam optimizer [84] for 10 epochs. The R^2 performance of the model on the held-out test dataset is 0.78, which drops to 0.35 when zero pixels are not considered in the calculation. The reconstruction of the 2D stellar velocity dispersion map for four random galaxies from the test set is shown in Figure 8.
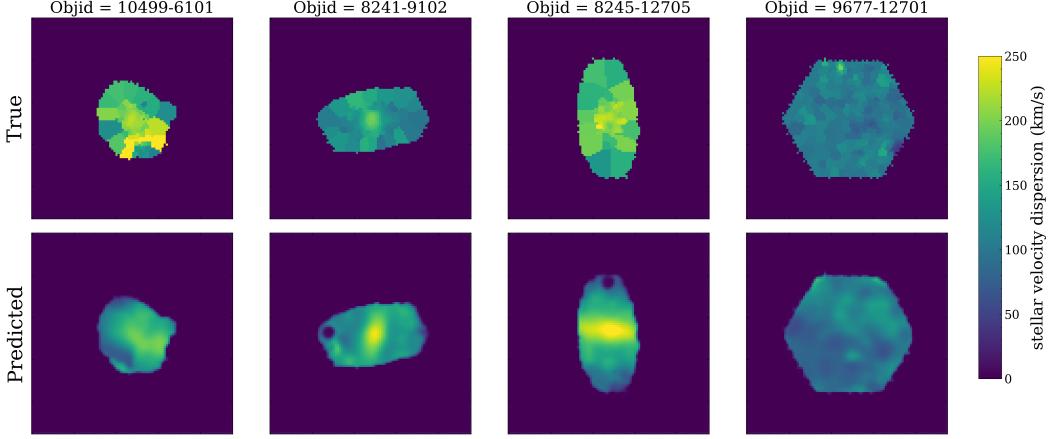
Figure 8: Reconstruction of the stellar velocity dispersion maps by the U-net using the hyperspectral datacubes as input for four galaxies from the MaNGA test set.

C   Infrastructure

To handle data at the 100TB scale, key decisions around data storage were made to accommodate both I/O and flexibility requirements. The data is stored in HDF5 files for consistency and performance. To allow efficient data access across various file systems, the samples are partitioned into HEALPix[36] regions. This enables simple spatial indexing for objects within similar regions (constrained by the grid size). This partitioning also allows downstream users to download only samples pertaining to specific regions or to match samples across different MULTIMODAL UNIVERSE modalities or datasets. Finally, the data is chunked into HDF5 files with a variable number of samples to prevent individual files from becoming too large.

C.1   Cross Matching Utility

The MULTIMODAL UNIVERSE framework provides utilities for cross-matching between different astronomical datasets. Below we demonstrate the simplicity of performing cross-matching operations using our provided tools:

```
1 from datasets import load_dataset_builder
2 from mmu.utils import cross_match_datasets
3
4 # Load the dataset descriptions from local copy of the data
5 sdss = load_dataset_builder("MultimodalUniverse/sdss",
6                             trust_remote_code=True)
7 hsc = load_dataset_builder("MultimodalUniverse/hsc",
8                            trust_remote_code=True)
9
10 # Use the cross matching utility to return a new HF dataset
11 dset = cross_match_datasets(
12     sdss,  # Left dataset, contains spectroscopy
13     hsc,   # Right dataset, contains imaging data
14     matching_radius=1.0,  # Distance in arcsec
15 )
```

C.2   Contributing New Datasets

Detailed guidelines for contributing new datasets to the `MultimodalUniverse`[37] framework are provided in the project's `DESIGN.md` and `CONTRIBUTING.md` documentation. These

---

[36]https://healpix.sourceforge.io/
[37]https://github.com/MultimodalUniverse/MultimodalUniverse/

documents are regularly updated with new tutorials and examples to facilitate community contributions.

For the most up-to-date examples and tutorials, users can refer to our getting started notebook[38].

---
[38]https://colab.research.google.com/github/MultimodalUniverse/MultimodalUniverse/blob/main/notebooks/getting_started.ipynb