HubbleCLIP: Associating Astronomical Observations and Natural Language with Multi-Modal Models

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

Yiding Song ydsong@mit.edu
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Jesse Thaler jthaler@mit.edu
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract

We present a multi-modal model which associates astronomical observations imaged by the *Hubble* Space Telescope (HST) with natural language. The model is fine-tuned from a base CLIP model using summarized proposal abstracts corresponding to HST observations. We show that the model embodies a meaningful joint representation between observations and text through experiments targeting observation retrieval (i.e., retrieving most relevant set of observations using natural language queries) and description retrieval (i.e., querying the astrophysical object classes and science use cases most relevant to a given observation). The model demonstrates the potential for using generalist rather than task-specific models for astrophysics research, in particular by leveraging text as an interface.

1 Introduction

Machine learning (ML) is starting to have a significant impact on the sciences, with astrophysics being no exception. Machine learning methods have demonstrated promise when applied to every part of the astrophysics pipeline, from instrument design, to data acquisition, to its analysis. Until recently, most applications of ML within astrophysics have focused on replacing or augmenting traditional techniques with ML counterparts in order to improve performance on specific tasks.

The *Foundation Model* paradigm, in contrast, seeks to develop generalist models which can be deployed on a wide range of tasks. The paradigm has been highly successful in domains like computer vision and natural language processing, as demonstrated by the widespread adoption of tools like CLIP, ChatGPT, Dall-E, and Stable Diffusion. These models are typically pre-trained on massive amounts of unlabeled data using self-supervised learning techniques, enabling them to learn powerful representations which can be optionally fine-tuned to address domain-specific tasks. At the heart of the paradigm lies the triumph of scale – scaling up model size, dataset size, and compute. However, foundation models can often be fine-tuned using only small amounts of domain-specific data, increasing their usefulness when applied to those specific domains.

There is considerable interest in developing custom foundation models for the sciences, with astrophysics being ripe for such an effort due to several reasons. The first is the availability of large amounts of publicly-available data as a contingency of publicy-funded data-taking efforts. The second is the multi-modality inherent astrophysical observations, with different types of data (e.g., images, spectra, light curves, textual descriptions) often available for each observation. This multi-modality was recently exploited to train AstroCLIP – a joint representation between multi-band images and optical spectra from the Dark Energy Spectroscopic Instrument (DESI). AstroLLaMA is another recent effort to fine-tune a publicly-available model (Llama-2) on astrophysics-specific textual data from the arXiv.

The success of the foundation model paradigm partly relies on the ability to leverage text as a *universal interface*. In this work, we take this outlook and train a joint representation between observations taken by the *Hubble* Space Telescope (HST) and natural language. We do so by using the associations between observation proposals and corresponding downstream observations. We show that fine-tuning a CLIP (Contrastive Language-Image Pre-training) model on this data enables learning meaningful joint representations.

The paper is organized as follows. In Sec. 2, we describe the dataset used in this work, including its curation and processing. In Sec. 3, we describe the methodology used to train and evaluate the model. In Sec. 4, we present the results of our experiments on image and text retrieval tasks. We discuss future prospects and conclude in Sec. 5.

2 Dataset and Processing

2.1 Summarization via guided generation

3 Methodology

3.1 Language-Image Pre-training

3.2 Pre-trained CLIP Model

3.3 Fine-tuning Objectives

4 Results and Discussion

4.1 Fine-tuned Retrieval Accuracy

4.2 'Zero-shot' Hypothesis and Object Retrieval

4.3 Text-to-Image Retrieval

5 Conclusion

Hinton et al. (2006)

Broader Impact Statement

In this optional section, TMLR encourages authors to discuss possible repercussions of their work, notably any potential negative impact that a user of this research should be aware of. Authors should consult the TMLR Ethics Guidelines available on the TMLR website for guidance on how to approach this subject.

Author Contributions

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors. Only add this information once your submission is accepted and deanonymized.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper. Only add this information once your submission is accepted and deanonymized.

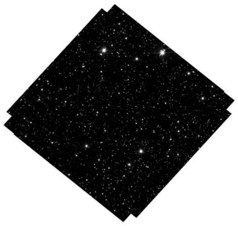
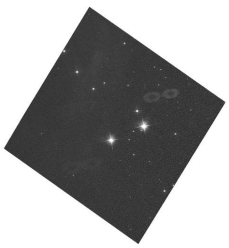
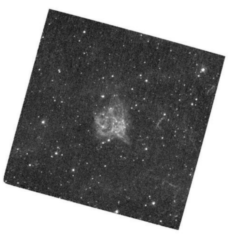
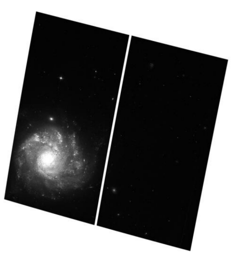
Image	Obs. cycle	Prop. ID	Summarized abstract
	26	15513	isolated black holes, old stellar population, background stars, Galactic bulge, microlensing events, light deflection, astrometric microlensing, 100-day events, low-mass stars, relative proper motions; long-duration events, T>300-day events, observed durations, BHs with mass > 10 Msun, T>300 days, very-long duration events, OGLE detects six T>300 days events each year, monitoring a few T>300-day events
	19	12577	interstellar dust, Cas A outburst, historical supernova SN, supernova remnant (SNR), astrophysical objects, light echoes, dust in the Milky Way, cooling envelope, shock breakout, progenitor star; observing its light echoes, studying large numbers of objects, connecting how a single object should change with time, time series of spectra, spatially resolved light curve, measuring the properties of the cooling envelope, estimating the radius of the progenitor star, connecting the progenitor star to the explosion to the SN to the SNR
	7	7340	oxygen-rich supernova remnants, LMC, SMC, O IIILambda5007 emission-line, S IILambdaLambda6724 emission-line, O IILambdaLambda3727 emission-line, ionization structure, chemically peculiar debris, metal-rich plasmas, supernova explosions, active pulsar, synchrotron nebula, young supernova remnant, extended complex of young stars ; models for nucleosynthesis in massive stars, excitation mechanisms in extremely metal-rich plasmas, the dynamics of supernova explosions, validating observed phenomena in astrophysical contexts
	22	13757	type Iax supernovae, white dwarfs, luminous blue systems, S1, companion stars, accretion onto white dwarfs, massive stars, unrelated stars, HST pre-explosion data, NGC 1309; progenitor systems, thermonuclear white dwarf supernovae, white dwarf supernovae, confirmation, characterization

Table 1: Data overview

References

Geoffrey E. Hinton, Simon Osindero, and Yee Whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.

A Appendix

You may include other additional sections here.

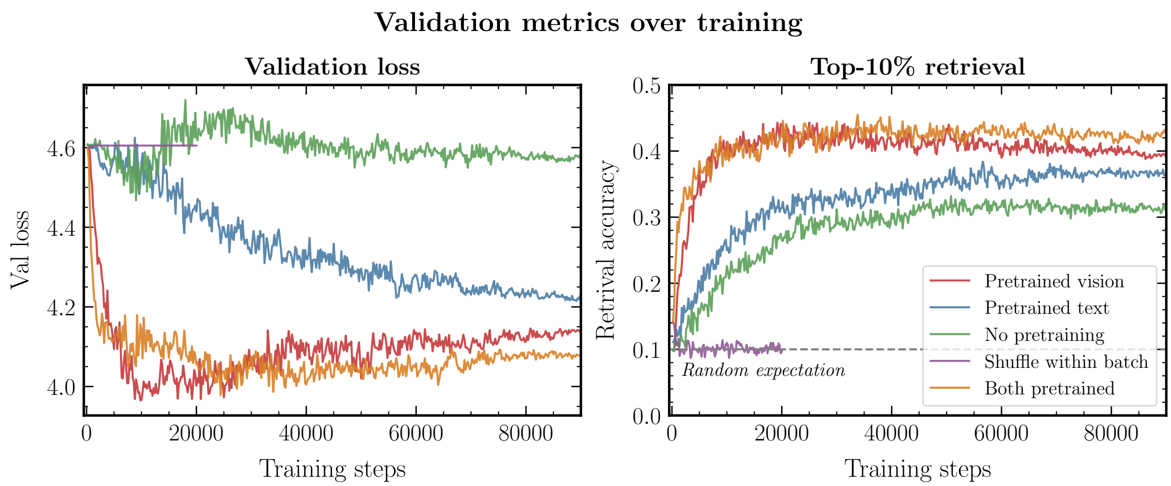


Figure 1: Retrieval accuracy

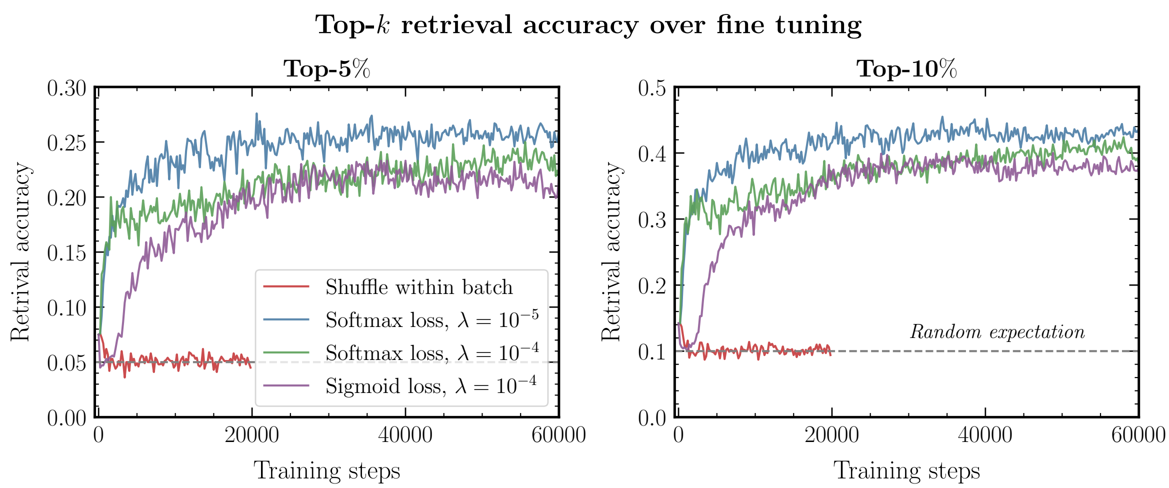


Figure 2: Retrieval accuracy

Image-to-Text Retrieval

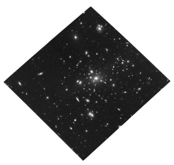
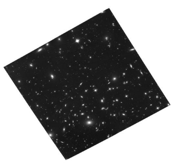
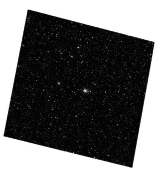
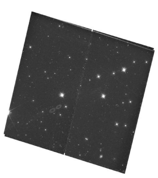
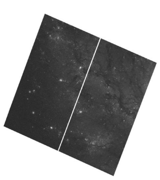
Image	Top classes (fine-tuned)	Top classes (base)	Abstract
	1. dark matter 2. Einstein rings 3. galaxy mergers 4. gravitational lensing 5. galaxy clusters	1. galaxy clusters 2. ultra diffuse galaxies 3. dwarf galaxies 4. gravitational lensing 5. crowded stellar field	Category: COSMOLOGY. We propose to study the physical nature of dark matter by using massive, merging clusters of galaxies. As shown with the Bullet Cluster {1E0657-56}, such massive well-measured systems are critical for our understanding of dark matter. By more than doubling the number of clusters in the sample and obtaining systems at different observation angles, impact parameters, geometrical arrangements, and merger velocities, the systematic uncertainties in the dark matter cross section calculations can be improved substantially, allowing us to move from rough order of magnitude estimates to measurements with quantifiable uncertainties that can be compared usefully with the predictions from numerical simulations, and the constraints on alternate gravity models become unambiguous. Our proposed targets are three extraordinary, merging galaxy clusters with X-ray and optical offsets that are placed at ideal redshifts for such a study; A520, A1758N, and A2163. To pin down the position of the dark matter component we require high resolution, absolutely calibrated mass maps. High resolution gravitational lensing data is needed to attain this goal, which can only be achieved with the excellent resolving power of the HST.
	1. dark matter 2. galaxy mergers 3. Einstein rings 4. gravitational lensing 5. dark energy	1. ultra diffuse galaxies 2. galaxy clusters 3. gravitational lensing 4. high-redshift quasars 5. dwarf galaxies	Category: COSMOLOGY. We propose to study the physical nature of dark matter by using massive, merging clusters of galaxies. As shown with the Bullet Cluster {1E0657-56}, such massive well-measured systems are critical for our understanding of dark matter. By more than doubling the number of clusters in the sample and obtaining systems at different observation angles, impact parameters, geometrical arrangements, and merger velocities, the systematic uncertainties in the dark matter cross section calculations can be improved substantially, allowing us to move from rough order of magnitude estimates to measurements with quantifiable uncertainties that can be compared usefully with the predictions from numerical simulations, and the constraints on alternate gravity models become unambiguous. Our proposed targets are three extraordinary, merging galaxy clusters with X-ray and optical offsets that are placed at ideal redshifts for such a study; A520, A1758N, and A2163. To pin down the position of the dark matter component we require high resolution, absolutely calibrated mass maps. High resolution gravitational lensing data is needed to attain this goal, which can only be achieved with the excellent resolving power of the HST.
	1. crowded stellar field 2. supernova remnants 3. compact stellar remnants 4. primordial black holes 5. pre-main sequence stars	1. stellar abundances 2. stellar populations 3. interstellar medium 4. pre-main sequence stars 5. Cepheid variables	Category: RESOLVED STELLAR POPULATIONS. Exploiting the full power of the Wide Field Camera 3 {WFC3}, we propose deep panchromatic imaging of four fields in the Galactic bulge. These data will enable a sensitive dissection of its stellar populations, using a new set of reddening-free photometric indices we have constructed from broad-band filters across UV, optical, and near-IR wavelengths. These indices will provide accurate temperatures and metallicities for hundreds of thousands of individual bulge stars. Proper motions of these stars derived from multi-epoch observations will allow separation of pure bulge samples from foreground disk contamination. Our catalogs of proper motions and panchromatic photometry will support a wide range of bulge studies. Using these photometric and astrometric tools, we will reconstruct the detailed star-formation history as a function of position within the bulge, and thus differentiate between rapid- and extended-formation scenarios. We will also measure the dependence of the stellar mass function on metallicity, revealing how the characteristic mass of star formation varies with chemistry. Our sample of bulge stars with accurate metallicities will include 12 candidate hosts of extrasolar planets. Planet frequency is correlated with metallicity in the solar neighborhood; our measurements will extend this knowledge to a remote environment with a very distinct chemistry. Our proposal also includes observations of six well-studied globular and open star clusters; these observations will serve to calibrate our photometric indices, provide empirical population templates, and transform the theoretical isochrone libraries into the WFC3 filter system. Besides enabling our own program, these products will provide powerful new tools for a host of other stellar-population investigations with HST/WFC3. We will deliver all of the products from this Treasury Program to the community in a timely fashion.
	1. low surface brightness galaxies 2. star formation histories 3. galaxy formation 4. ultra diffuse galaxies 5. circumgalactic medium	1. gravitational lensing 2. high-redshift quasars 3. brown dwarfs 4. trans-Neptunian objects 5. Kuiper Belt objects	Category: Stellar Populations and the Interstellar Medium. Observations of the ultra-faint dwarfs (UFDs), as relics of the epoch of reionization, allow us to probe the earliest epochs of star formation (SF). In particular, the UFDs in low density environments for most of their lifetimes provide unique tools to probe the effects of early environmental conditions on the SF histories (SFHs) of the UFDs and their sub-solar initial mass function (IMF) because they likely maintain the best \`fossil' record of early local environments. We propose to obtain deep ACS and UVIS imaging in F606W and F814W for 2 LMC satellite UFDs that are on their first approach to our Galaxy, and thus resided in the outskirts of the Local Group at high redshift. This program is designed to identify systematic differences in the stellar populations of recently captured UFDs vs. long-term MW satellites (data available from previous programs) by using high-fidelity color-magnitude diagrams constructed from deep HST imaging as well as spectroscopically measured metallicity distribution functions. We will: (1) Establish whether SF is quenched at different times with different rate in UFDs in low density environment at early times, probing the patchiness of reionization by directly comparing with theoretical predictions; (2) Identify variations in the sub-Solar IMF across UFDs born in different environments; (3) Pave the way for a more accurate constraint on the MW halo mass.
	1. supernovae 2. Cepheid variables 3. star clusters 4. star forming galaxies 5. dust	1. high-redshift quasars 2. gravitational lensing 3. Kuiper Belt objects 4. compact stellar remnants 5. ultra diffuse galaxies	Category: RESOLVED STELLAR POPULATIONS. We propose to test two of the clearest predictions of the theory of evolution of massive-star evolution: 1} The formation of Wolf-Rayet stars depends strongly on these stars' metallicity {Z}, with relatively fewer WR stars forming at lower Z, and 2} Wolf-Rayet stars die as Type Ib or Ic supernovae. To carry out these tests we propose a deep, narrowband imaging survey of the massive star populations in the ScI spiral galaxy M101. Just as important, we will test the hypothesis that Superclusters like 30 Doradus are always richly populated with WR stars, and by implication that these complexes are responsible for the spectral signatures of starburst galaxies. Our previous HST survey of the HII regions in the ScIII galaxy NGC 2403 suggested that the distribution of WR stars and RSG is a sensitive diagnostic of the recent star-forming history of these large complexes: young cores of O and WR stars are surrounded by older halos containing RSG. Theory predicts that this must change with metallicity; relatively fewer WR stars form at lower Z. A key goal of our proposal is to directly test this paradigm in a single galaxy, M101 being the ideal target. The abundance gradient across M101 {a factor of 20} suggests that relatively many more WR will be found in the inner parts of this galaxy than in the outer "suburbs". Second, we note that WR stars are predicted to end their lives as core-collapse or pair-instability supernovae. The WR population in M101 may be abundant enough for one to erupt as a Type Ib or Ic supernova within a generation. The clear a priori identification of a WR progenitor would be a major legacy of HST. Third, we will also determine if "superclusters", heavily populated by WR stars, are common in M101. It is widely claimed that such Superclusters produce the integrated spectral signatures of Starburst galaxies. We will be able to directly measure the numbers and emission-line luminosities of thousands of Wolf Rayet stars located in hundreds of M101 Superclusters, and correlate those numbers against the Supercluster sizes and luminosities. It is likely {but far from certain} that Supercluster sizes and emission-line luminosities are driven by their Wolf-Rayet star content. Our sample will be the largest and best-ever Supercluster/Wolf Rayet sample, an excellent local proxy for characterizing starburst galaxies' Superclusters.

Figure 3: Retrieval accuracy

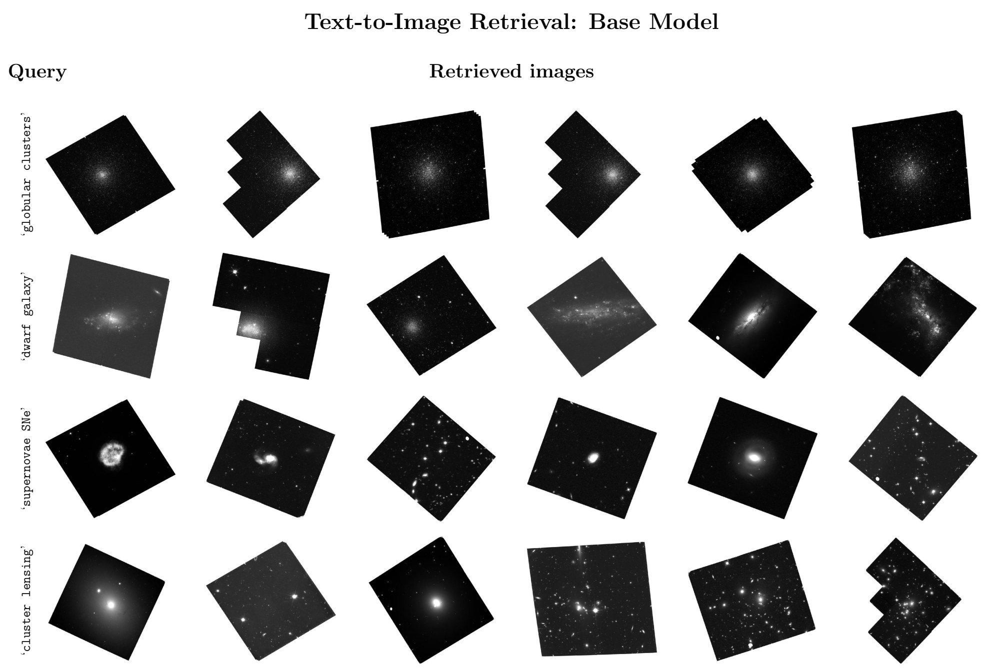


Figure 4: Retrieval accuracy

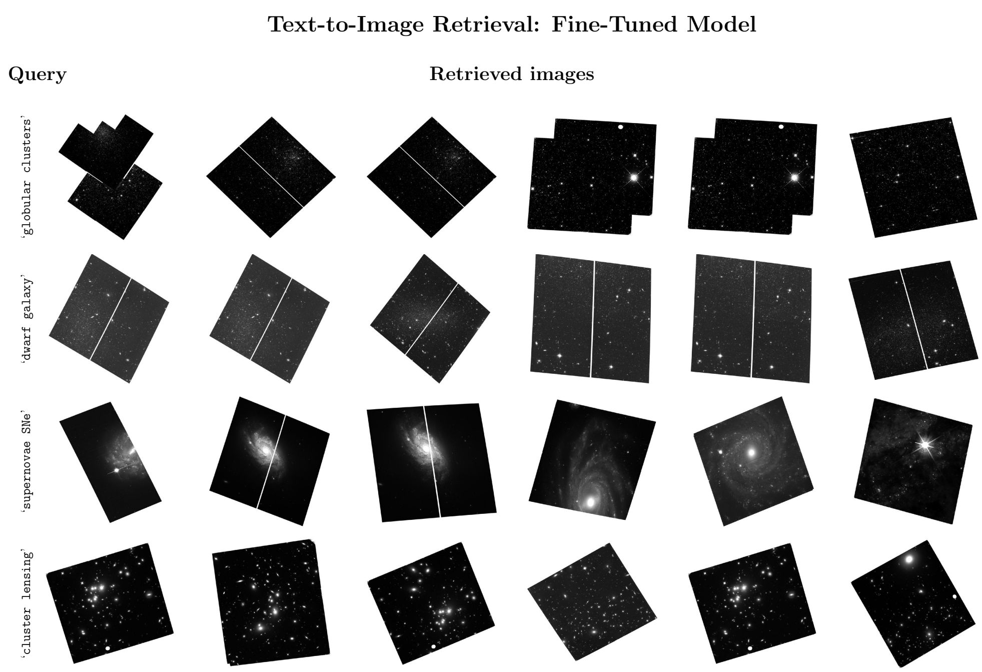


Figure 5: Retrieval accuracy