The Platonic Universe: Do Foundation Models See the Same Sky?

UniverseTBD

Kshitij Duraphe*[1]
kshitijduraphe5@gmail.com

Michael J. Smith*[2,3,4]
mike@mjjsmith.com

Shashwat Sourav*[5]
s.shashwat@wustl.edu

John F. Wu*[6,7]
jowu@stsci.edu

[1]Independent Researcher [2]AstroAI [3]Harvard-Smithsonian CfA
[4]University of Hertfordshire [5]Washington University St. Louis
[6]Space Telescope Science Institute [7]Johns Hopkins University

Abstract

We test the Platonic Representation Hypothesis (PRH) in astronomy by measuring representational convergence across a range of foundation models trained on different data types. Using spectroscopic and imaging observations from JWST, HSC, Legacy Survey, and DESI, we compare representations from vision transformers, self-supervised models, and astronomy-specific architectures via mutual k-nearest neighbour analysis. We observe consistent scaling: representational alignment generally increases with model capacity across our tested architectures, supporting convergence toward a shared representation of galaxy astrophysics. Our results suggest that astronomical foundation models can use pre-trained general-purpose architectures, allowing us to capitalise on the broader machine learning community's already-spent computational investment.

1 Astronomy and the Platonic Representation Hypothesis

Three historical waves of increasingly automated connectionism have lapped the shores of astronomy. The late 1980s brought with them MLPs tuned for astronomical applications on manually selected inputs (e.g. Adorf and Johnston, 1988; Angel et al., 1990; Odewahn et al., 1992). With the advent of CNNs, RNNs, and deep learning, these MLP models gave way to raw data ingestion (e.g. Dieleman et al., 2015; Charnock et al., 2018; Wu and Peek, 2020). And the third wave of unsupervised and self-supervised learning entirely removed the need for human supervision, with connectionist methods inferring astronomical knowledge directly from the raw data (e.g. Sarmiento et al., 2021; Smith et al., 2022). The swell of a fourth wave has now broken upon astronomy's shores—the application of foundation models to astronomical observations, publications, and survey data (Smith and Geach, 2023). The fourth wave has brought with it diverse approaches in the search for a viable path towards a single, canonical, astro-foundation model. As a rough overview, research groups have explored contrastive methods (e.g. Mishra-Sharma et al., 2024; Parker et al., 2024; Slijepcevic et al., 2024; Zhao et al., 2025), generative architectures (e.g. Leung and Bovy, 2023; Koblischke and Bovy, 2024; Ore et al., 2024), autoregressive modelling (e.g. J.-S. Pan et al., 2024; Smith et al., 2024; Euclid Collaboration et al., 2025; Heneka et al., 2025; Moriwaki et al., 2025; Zuo et al., 2025),

*Equal contribution.

Machine Learning and the Physical Sciences Workshop at NeurIPS 2025.

and approaches that directly finetune large language models on astronomical text (e.g. Nguyen et al., 2023; de Haan et al., 2024; Perkowski et al., 2024; Zaman et al., 2025).

In this paper we explore the hypothesis that the neural architecture and training regime of our eventual canonical astronomical foundation model *does not matter*: that any sufficient neural network will converge to an equivalent embedding space when pre-trained on enough data with enough compute. This conjecture has already gained traction in the deep learning community as the 'Platonic Representation Hypothesis' (PRH), with perhaps the best known example being put forward as a position paper at ICML 2024[2] (Huh et al., 2024).

The PRH as defined by Huh et al. (2024) proposes that neural networks trained with different objectives on different data modalities are converging toward a shared statistical model of reality in their representation spaces. The authors draw inspiration from Plato's 'Allegory of the Cave', where the cave-dwellers mistake shadows on a wall for reality itself (Plato, c. 375 BCE). In this analogy, our training data are the shadowy projections of an underlying reality, and our models are learning to recover representations (or 'Forms') of the reality that generates these data. As our models become larger and are trained on more diverse tasks, they converge toward a Platonic ideal representation: a perfect lossless Form of our underlying reality. This convergence is driven by three key mechanisms: *task generality* (models trained on more diverse tasks require representations that capture more information about underlying reality), *model capacity* (larger models are more likely to find optimal representations), and *simplicity bias* (neural networks naturally favour simpler solutions that generalise better). Under the PRH, we expect progressively larger models to exhibit more similar representations, even if models are trained across different data modalities. We now quantitatively test whether this is true.

2 Astronomical Data as Imperfect Phenomena of Forms

Astronomical observations provide an important testbed for the PRH due to their fundamental nature as different projections of the same underlying cosmic reality. These observations are inherently linked through shared physical processes; a galaxy's morphology (captured in images), chemical composition (revealed through spectroscopy), and integrated properties (measured via photometry) all emerge from the same stellar populations, gas and dust dynamics, and underlying matter distributions. This shared physical origin suggests that foundation models viewing different astronomical modalities should converge toward representations that capture the underlying fundamental physics governing these phenomena. All the pieces are in place to test the PRH in astronomy: the scale and diversity of modern surveys provide the data volume necessary to test convergence across multiple model architectures and training objectives, and recent work has eased the crossmodal[3] use of such data (The Multimodal Universe Collaboration, 2024).

We therefore test the PRH on a selection of vision and spectra foundation models using datasets compiled by the Multimodal Universe. Below, we briefly describe and motivate our chosen data and model architectures below. Further information about the crossmatched datasets and model specifications can be found in Appendix C.

Data. We test across four crossmatched astronomical datasets that capture fundamentally different projections of galaxy Forms: Hyper Suprime-Cam (HSC; Miyazaki et al., 2018), DESI Legacy Imaging Survey (Dey et al., 2019), and James Webb Space Telescope (JWST; Gardner et al., 2023) images from public surveys (Valentino et al., 2023); and DESI spectra (DESI Collaboration et al., 2024).[4] We use ground-based HSC imaging as our reference baseline. We include DESI spectroscopy to enable cross-modal representation alignment testing between images and spectra. The DESI Legacy Survey's inclusion allows us to test representational alignment across different ground-based imaging survey strategies. JWST NIRCam imaging represents the most extreme imaging test: the telescope produces space-based infrared observations that reveal dust emission and dust-obscured

[2]With—as always—many related rumblings preceding this work (e.g. Lenc and Vedaldi, 2014; Bansal et al., 2021; Liu et al., 2023).

[3]We define an astronomical 'mode' or 'modality' as the information captured by a specific instrument. Under this definition, for example, JWST and HSC imaging are separate modes.

[4]We also initially tested Sloan Digital Sky Survey spectra (SDSS I & II; York et al., 2000). However, our embedding model exhibits out-of-domain behaviour for SDSS spectra (even when we reprocess the SDSS spectra to match DESI data). See Appendix B for more details.

stellar populations invisible to our HSC and Legacy optical surveys. We use the Multimodal Universe (MMU) to crossmatch between data modalities (The Multimodal Universe Collaboration, 2024).

We rescale our images by setting the min and max channel-wise pixel values to their respective 5th and 99th percentile values, computed using batches of up to 10000 images per dataset. For HSC and Legacy Survey we take the z, r, and g bands as our RGB image channels, and for JWST we take the F444W, F277W, and F090W bands, ensuring maximum wavelength coverage while remaining suitable for our foundation models trained on RGB natural images. We perform any further data pre-processing steps described in the original model authors.

Model architectures. We test across six fundamentally different neural architectures and training paradigms: ViT, ConvNeXtv2, DINOv2, IJEPA, AstroPT, and Specformer. ConvNeXTv2 and ViT represent two approaches (convolutional and self-attentional) that are trained under a 'traditional' supervised paradigm, having been pre-trained or fine-tuned under classification objectives (Dosovitskiy et al., 2020; Woo et al., 2023). DINOv2 employs self-supervised learning through knowledge distillation (Caron et al., 2021; Oquab et al., 2023), and IJEPA employs a non-generative self-supervised approach that predicts abstract representations of image regions rather than reconstructing pixel-level details (LeCun, 2022; Assran et al., 2023). DINO and IJEPA's inclusion allows us to test representation convergence across a range of self-supervised approaches. AstroPTv2 is an autoregressive decoder transformer designed for astronomical applications (Smith et al., 2024). As a model pre-trained exclusively on astronomical observations, AstroPT's inclusion tests whether models pre-trained on specialised datasets converge toward the same representations as models pre-trained on more general data. Our other astronomy-specific model, Specformer, processes one-dimensional astronomical spectra via a transformer, and represents an entirely distinct data modality compared to image-based models (Parker et al., 2024). Specformer's inclusion tests the most extreme case: whether models pre-trained on fundamentally different input types adhere to the PRH.

Measuring representational alignment. Following the methodology established in the original PRH work, we measure representational alignment using the mutual k-nearest neighbour (MKNN) metric (Chechik et al., 2010). Given two embeddings $(\mathbf{z}_1, \mathbf{z}_2)$ corresponding to the same object as viewed by two different instruments or models, the MKNN score is computed as the cardinality of intersections for each object's k-nearest neighbours in the embedding space: $\text{MKNN}(\mathbf{z}_1, \mathbf{z}_2) = k^{-1}|N_k(\mathbf{z}_1) \cap N_k(\mathbf{z}_2)|$ where N_k is the k-nearest neighbours operation, and $|\cdot|$ denotes set cardinality.

We test representational alignment via the MKNN score *across* astronomical modes (*crossmodal*), and *within* a mode (*intramodal*). For the intramodal case, we calculate the MKNN on embeddings produced by two different sizes of the same model architecture, given the same modality. For the crossmodal case we take a model of a specified architecture type and compute the embeddings for two crossmatched astronomical modalities for a range of model sizes. If the PRH holds, then the intramodal and crossmodal MKNN scores should consistently increase with increasing model size.

3 Convergence Toward Shared Representations

Table 1: Intramodal embedding alignment within a model family. The PRH predicts that the MKNN score will increase as we compare the embeddings of larger model pairs within a model family, since larger models will generate embeddings closer to the Platonic Ideal.

Model Pairs	JWST	Legacy	HSC
AstroPTv2 Small vs Base	49.7%	8.1%	10.3%
AstroPTv2 Base vs Large	56.2%	10.0%	13.5%
ConvNeXtv2 Nano vs Tiny	33.3%	4.5%	5.3%
ConvNeXtv2 Tiny vs Base	29.5%	3.8%	4.4%
ConvNeXtv2 Base vs Large	35.8%	6.3%	7.8%
DINOv2 Small vs Base	32.8%	4.2%	4.6%
DINOv2 Base vs Large	32.1%	5.6%	5.7%
DINOv2 Large vs Giant	40.2%	10.2%	10.9%
ViT Base vs Large	28.7%	3.1%	4.3%
ViT Large vs Huge	32.6%	4.4%	5.0%

Results. We show results from the *intramodal* trials in Tab. 1 and the results from the *crossmodal* alignment trials in Fig. 1. Both sets of results show significant correlations between MKNN score and model size.

We find statistically significant evidence that larger models, even when trained across different data modalities, converge towards more similar representations. Table 1 shows that intramodal MKNN scores increase for 14 of the 18 pairwise comparisons. For example, we see an increase for JWST between AstroPTv2 Small vs Base (49.7%) and Base vs Large (56.2%). Under a random binomial test, $p = 1.54 \times 10^{-2}$. Meanwhile, crossmodal MKNN scores increase for 28 out of 33 trials, with a binomial test $p = 3.31 \times 10^{-5}$.

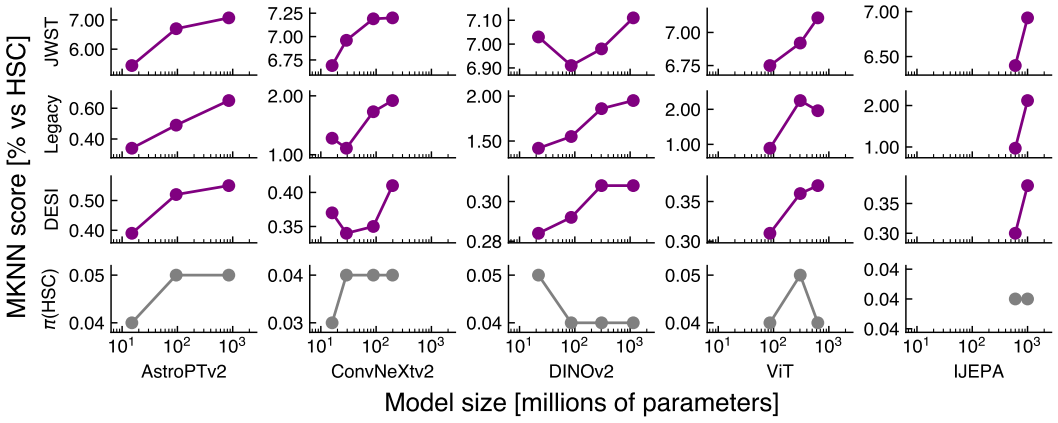

Figure 1: Model size vs crossmodal embedding alignment for our tested models. Each of our modality embeddings are compared to crossmatched embeddings from a paired HSC image dataset. π denotes a random permutation along the zeroth axis, such that π(HSC) denotes a comparison to a randomised HSC embedding dataset. A table of the MKNN scores can be found in Tab. 2 of Appendix A.

Discussion. Aside from AstroPT and Specformer, our tested models are not significantly pre-trained on astronomical data. That these models identify any correspondence between fundamentally different astronomical observations supports the PRH: suggesting that sufficiently scaled up neural networks learn universal structural patterns transcending their training domains. We can also see that our natural image-trained models achieve embedding alignment that increases with model size with Specformer's DESI spectral embeddings, therefore showing correspondence between fundamentally different modalities and data types they have never encountered. Interestingly, AstroPTv2 (pre-trained on imaging from the DESI Legacy Survey) yields comparable rather than dominant performance, also suggesting convergence toward shared representations rather than domain-specific features.

While our results provide compelling evidence for the PRH, we note several limitations that suggest avenues for future exploration. Primarily, some modality comparisons rely on relatively small datasets (e.g. 1.67k objects in the case of JWST vs HSC), which may not capture the full diversity of astronomical phenomena. Future work will prioritise increasing the size of crossmatched catalogues and adding additional testing modalities to the MMU. Our choice of the MKNN metric provides only one perspective on representation similarity; future work will explore further similarity metrics like centred kernel alignment (Kornblith et al., 2019) and mutual information (Y. Li et al., 2015) measurements. We also plan on comparing more diverse modalities and architectures, as our tested astronomical modes and architectures represent only a small slice of all possible varieties. Notable absences include but are not limited to LLMs, diffusion models, and multimodal architectures on the modelling side, and time series and tabular data, and other wavelength regimes as new astronomical modalities.

Summary. We observe general improvement in representational alignment at larger model scales, suggesting that each architecture is converging towards a shared representation. Taken to its conclusion, this convergence implies that future efforts in astronomical foundation modelling should focus less on astronomy-specific architectures and more on scale and data diversity. It also follows that the astronomy community should embrace pre-trained foundation models rather than training from scratch: if all architectures converge toward the same representations, then starting from models pre-trained on natural images or text—with their billions of parameters and massive computational investment already spent—offers both superior performance and dramatic reductions in environmental impact. The broader open source machine learning community has already invested the

GPU-centuries needed to learn general-purpose representations, we need now only gently guide these models toward astronomical use-cases.

Acknowledgements

This work has made use of the University of Hertfordshire's high-performance computing facility (`http://uhhpc.herts.ac.uk/`).

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.

This work is based in part on observations made with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST.

The DESI Legacy Imaging Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS), the Beijing-Arizona Sky Survey (BASS), and the Mayall z-band Legacy Survey (MzLS). 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. NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. Pipeline processing and analyses of the data were supported by NOIRLab and the Lawrence Berkeley National Laboratory (LBNL). Legacy Surveys also uses data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), a project of the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. Legacy Surveys was supported by: the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility; the U.S. National Science Foundation, Division of Astronomical Sciences; the National Astronomical Observatories of China, the Chinese Academy of Sciences and the Chinese National Natural Science Foundation. LBNL is managed by the Regents of the University of California under contract to the U.S. Department of Energy. The complete acknowledgments can be found at `https://www.legacysurvey.org/acknowledgment/`.

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.

References

Adorf, H. M. and M. D. Johnston (1988). "Artificial neural nets in astronomy". In: *Arbeitspapier der Gesellschaft für Mathematik and Datenverarbeitung*. Vol. 329. Arbeitspapier der Gesellschaft für Mathematik and Datenverarbeitung.

Angel, J. R.P. et al. (1990). "Adaptive optics for array telescopes using neural-network techniques". English (US). In: *Nature* 348.6298, pp. 221–224. ISSN: 0028-0836.

Assran, M. et al. (2023). "Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture". In: *ArXiv e-prints*. eprint: 2301.08243.

Astropy Collaboration et al. (Oct. 2013). "Astropy: A community Python package for astronomy". In: *Astronomy and Astrophysics* 558, A33, A33. arXiv: 1307.6212 [astro-ph.IM].

Astropy Collaboration et al. (Sept. 2018). "The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package". In: *The Astronomical Journal* 156.3, 123, p. 123. arXiv: 1801.02634 [astro-ph.IM].

Astropy Collaboration et al. (Aug. 2022). "The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package". In: *The Astrophysical Journal* 935.2, 167, p. 167. arXiv: 2206.14220 [astro-ph.IM].

Bansal, Y., P. Nakkiran, and B. Barak (2021). "Revisiting Model Stitching to Compare Neural Representations". In: *ArXiv e-prints*. eprint: 2106.07682.

Caron, M. et al. (2021). "Emerging Properties in Self-Supervised Vision Transformers". In: *ArXiv e-prints*. eprint: 2104.14294.

Charnock, T., G. Lavaux, and B. D. Wandelt (2018). "Automatic physical inference with information maximizing neural networks". In: *Physical Review D* 97.8, p. 083004. ISSN: 2470-0029.

Chechik, G. et al. (2010). "Large Scale Online Learning of Image Similarity Through Ranking". In: *Journal of Machine Learning Research* 11.36, pp. 1109–1135.

de Haan, T. et al. (2024). "AstroMLab 3: Achieving GPT-4o Level Performance in Astronomy with a Specialized 8B-Parameter Large Language Model". In: *ArXiv e-prints*. eprint: 2411.09012.

DESI Collaboration et al. (2024). "The Early Data Release of the Dark Energy Spectroscopic Instrument". In: *Astronomical Journal* 168.2, p. 58. ISSN: 1538-3881.

Dey, A. et al. (2019). "Overview of the DESI Legacy Imaging Surveys". In: *Astronomical Journal* 157.5, p. 168. ISSN: 1538-3881.

Dieleman, S., K. W. Willett, and J. Dambre (2015). "Rotation-invariant convolutional neural networks for galaxy morphology prediction". In: *Monthly Notices of the Royal Astronomical Society* 450.2, pp. 1441–1459. arXiv: 1503.07077 [astro-ph.IM].

Dosovitskiy, A. et al. (2020). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *ArXiv e-prints*. eprint: 2010.11929.

Earl, N. et al. (2025). *astropy/specutils: v2.1.0*. Version v2.1.0.

Euclid Collaboration et al. (2025). "Euclid Quick Data Release (Q1) Exploring galaxy properties with a multi-modal foundation model". In: *ArXiv e-prints*. eprint: 2503.15312.

Gardner, J. P. et al. (2023). "The James Webb Space Telescope Mission". In: *Publications of the Astronomical Society of the Pacific* 135.1048, 068001, p. 068001. arXiv: 2304.04869 [astro-ph.IM].

Heneka, C. et al. (2025). "Large Language Models – the Future of Fundamental Physics?" In: *ArXiv e-prints*. eprint: 2506.14757.

Huh, M. et al. (2024). "Position: The Platonic Representation Hypothesis". In: *International Conference on Machine Learning*. PMLR, pp. 20617–20642.

Koblischke, N. and J. Bovy (2024). "SpectraFM: Tuning into Stellar Foundation Models". In: *ArXiv e-prints*. eprint: 2411.04750.

Kornblith, S. et al. (2019). "Similarity of neural network representations revisited". In: *International conference on machine learning*. PMlR, pp. 3519–3529.

LeCun, Y. (2022). *A Path Towards Autonomous Machine Intelligence*. [Online; accessed 27. Jul. 2025].

Lenc, K. and A. Vedaldi (2014). "Understanding image representations by measuring their equivariance and equivalence". In: *ArXiv e-prints*. eprint: 1411.5908.

Leung, H. W. and J. Bovy (2023). "Towards an astronomical foundation model for stars with a transformer-based model". In: *Monthly Notices of the Royal Astronomical Society* 527.1, pp. 1494–1520. ISSN: 0035-8711.

Li, Y. et al. (2015). "Convergent Learning: Do different neural networks learn the same representations?" In: *Proceedings of the 1st International Workshop on Feature Extraction: Modern Questions and Challenges at NIPS 2015*. Ed. by Dmitry Storcheus, Afshin Rostamizadeh, and Sanjiv Kumar. Vol. 44. Proceedings of Machine Learning Research. Montreal, Canada: PMLR, pp. 196–212.

Liu, H. et al. (2023). "Visual Instruction Tuning". In: *ArXiv e-prints*. eprint: 2304.08485.

Mishra-Sharma, S., Y. Song, and J. Thaler (2024). "PAPERCLIP: Associating Astronomical Observations and Natural Language with Multi-Modal Models". In: *ArXiv e-prints*. eprint: 2403.08851.

Miyazaki, S. et al. (2018). "Hyper Suprime-Cam: System design and verification of image quality". In: *Publications of the Astronomical Society of Japan* 70.SP1, S1. ISSN: 0004-6264.

Moriwaki, K. et al. (2025). "CosmoGLINT: Cosmological Generative Model for Line Intensity Mapping with Transformer". In: *ArXiv e-prints*. eprint: 2506.16843.

Nguyen, T. D. et al. (2023). "AstroLLaMA: Towards Specialized Foundation Models in Astronomy". In: *ArXiv e-prints*. eprint: 2309.06126.

Odewahn, S. C. et al. (1992). "Automated Star/Galaxy Discrimination With Neural Networks". In: *The Astronomical Journal* 103, p. 318.

Oquab, M. et al. (2023). "DINOv2: Learning Robust Visual Features without Supervision". In: *ArXiv e-prints*. eprint: 2304.07193.

Ore, A., C. Heneka, and T. Plehn (2024). "SKATR: A Self-Supervised Summary Transformer for SKA". In: *ArXiv e-prints*. eprint: 2410.18899.

Pan, J.-S. et al. (2024). "The Scaling Law in Stellar Light Curves". In: *ArXiv e-prints*. eprint: 2405.17156.

Parker, L. et al. (2024). "AstroCLIP: a cross-modal foundation model for galaxies". In: *Monthly Notices of the Royal Astronomical Society* 531.4, pp. 4990–5011. ISSN: 0035-8711.

Perkowski, E. et al. (2024). "AstroLLaMA-Chat: Scaling AstroLLaMA with Conversational and Diverse Datasets". In: *Research Notes of the AAS* 8.1, p. 7. ISSN: 2515-5172.

Plato (c. 375 BCE). *The Republic*, 514a–520a.

Sarmiento, R. et al. (2021). "Capturing the Physics of MaNGA Galaxies with Self-supervised Machine Learning". In: *The Astrophysical Journal* 921.2, p. 177. ISSN: 0004-637X.

Slijepcevic, I. V. et al. (2024). "Radio galaxy zoo: towards building the first multipurpose foundation model for radio astronomy with self-supervised learning". In: *RAS Techniques and Instruments* 3.1, pp. 19–32. ISSN: 2752-8200.

Smith, M. J. and J. E. Geach (2023). "Astronomia ex machina: a history, primer and outlook on neural networks in astronomy". In: *R. Soc. Open Sci.* 10.5, p. 221454. ISSN: 2054-5703.

Smith, M. J. et al. (2022). "Realistic galaxy image simulation via score-based generative models". In: *Monthly Notices of the Royal Astronomical Society* 511.2, pp. 1808–1818. arXiv: 2111.01713 [astro-ph.IM].

Smith, M. J. et al. (2024). "AstroPT: Scaling Large Observation Models for Astronomy". In: *ArXiv e-prints*. eprint: 2405.14930.

The Multimodal Universe Collaboration (2024). "The Multimodal Universe: Enabling Large-Scale Machine Learning with 100 TB of Astronomical Scientific Data". In: *Advances in Neural Information Processing Systems* 37, pp. 57841–57913.

Valentino, F. et al. (2023). "An Atlas of Color-selected Quiescent Galaxies at z > 3 in Public JWST Fields". In: *The Astrophysical Journal* 947.1, 20, p. 20. arXiv: 2302.10936 [astro-ph.GA].

Woo, S. et al. (2023). "ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders". In: *ArXiv e-prints*. eprint: 2301.00808.

Wu, J. F. and J. E. G. Peek (2020). "Predicting galaxy spectra from images with hybrid convolutional neural networks". In: *ArXiv e-prints*. eprint: 2009.12318.

York, D. G. et al. (2000). "The Sloan Digital Sky Survey: Technical Summary". In: *Astronomical Journal* 120.3, pp. 1579–1587. arXiv: astro-ph/0006396 [astro-ph].

Zaman, S. et al. (2025). "AstroLLaVA: towards the unification of astronomical data and natural language". In: *ArXiv e-prints*. eprint: 2504.08583.

Zhao, X. et al. (2025). "SpecCLIP: Aligning and Translating Spectroscopic Measurements for Stars". In: *ArXiv e-prints*. eprint: 2507.01939.

Zuo, X. et al. (2025). "FALCO: a Foundation model of Astronomical Light Curves for time dOmain astronomy". In: *ArXiv e-prints*. eprint: 2504.20290.

A Full Crossmodal Results Table

We list the full results across pairs of data modalities for each model variant in Table 2. Each entry is the MKNN score as a percentage. Our main experimental results are shown in the first three columns (JWST imaging vs HSC imaging, Legacy imaging vs HSC imagina, and DESI spectra vs HSC imaging). The fourth column shows results from the SDSS spectra vs HSC imaging; however, SDSS is out-of-domain for the Specformer embedding model, and we only include these results for transparency (see Appendix B for more details).

As a null test, we include a final column that shows MKNN scores for randomly shuffled HSC embeddings, π(HSC), against unshuffled HSC embeddings. Here we use the same dataset as in the 'DESI vs HSC' experiment (with 18.6k galaxies).

Table 2: Crossmodal MKNN scores. The SDSS and DESI spectra are encoded by Specformer, with the spectra embedding compared to an embedding generated by the vision model on the left.

Model	JWST vs HSC	Legacy vs HSC	DESI vs HSC	SDSS vs HSC	π(HSC) vs HSC
AstroPTv2 Small	5.44%	0.34%	0.39%	0.44%	0.04%
AstroPTv2 Base	6.70%	0.49%	0.52%	0.50%	0.05%
AstroPTv2 Large	7.07%	0.65%	0.55%	0.47%	0.05%
ConvNeXtv2 Nano	6.69%	1.28%	0.37%	0.36%	0.03%
ConvNeXtv2 Tiny	6.96%	1.11%	0.34%	0.43%	0.04%
ConvNeXtv2 Base	7.19%	1.73%	0.35%	0.46%	0.04%
ConvNeXtv2 Large	7.20%	1.92%	0.41%	0.41%	0.04%
DINOv2 Small	7.03%	1.42%	0.28%	0.34%	0.05%
DINOv2 Base	6.91%	1.55%	0.29%	0.43%	0.04%
DINOv2 Large	6.98%	1.86%	0.31%	0.37%	0.04%
DINOv2 Giant	7.11%	1.95%	0.31%	0.45%	0.04%
IJEPA Huge	6.40%	0.97%	0.30%	0.44%	0.04%
IJEPA Giant	6.93%	2.12%	0.38%	0.39%	0.04%
ViT Base	6.75%	0.89%	0.31%	0.43%	0.04%
ViT Large	6.92%	2.25%	0.36%	0.41%	0.05%
ViT Huge	7.11%	1.96%	0.37%	0.43%	0.04%

B SDSS Spectra and Domain Shift Challenges

In addition to the datasets listed in Section 2, we also initially experimented with SDSS galaxy spectra (York et al., 2000). However, the Specformer embedding model is trained on DESI spectra (Parker et al., 2024), which requires a different wavelength grid (i.e., with different wavelength ranges and with a different spectral resolution). We attempted to circumvent this by interpolating SDSS spectra via the (Astropy-affiliated; Astropy Collaboration et al. (2013, 2018, 2022)) Specutils package (Earl et al., 2025) to resample SDSS spectra onto DESI's wavelength grid (7781 pixels, 3600–9800 Å), transforming them from their native format (4000 pixels, 3800–9200 Å), albeit with loss of spectral resolution.

However, DESI and SDSS spectra differ in terms of other observing systematics, e.g., observing conditions and sites, integration times and observing depth, detector systematics, and calibration pipelines, as well as galaxy population-level trends owing to survey design and selection effects. All of these subtle biases are implicitly encoded in the Specformer model, leaving the SDSS spectra significantly out of domain. This is true even when we attempt to correct the wavelength-dependent effects.

Thus, we do not expect the SDSS Specformer embeddings to be meaningful. Indeed, the crossmodal MKNN scores for SDSS vs HSC are no better than random. Table 2 shows that 6 of the 11 scores are increasing, with a binomial test $p = 0.5$. While the magnitude of the MKNN scores for SDSS vs HSC are comparable to those of DESI vs HSC, this only suggests that spectral-to-imaging alignment typically falls in the MKNN $\sim 0.3 - 0.5\%$ range.

C Further Information On Used Datasets and Models

Here we list the datasets used, as well as the models used for this study. For each dataset and model we provide a link to the publicly available data or weights. Cross-matching between surveys is performed using the Multimodal Universe framework on MMU v1 (The Multimodal Universe Collaboration, 2024) with a 1 arcsecond matching radius.

We release all of our code on Github at `github.com/UniverseTBD/platonic-universe`.

Table 3: Foundation models and astronomical datasets used in this study.

Category	Name	Size	Hugging Face source
Models	AstroPTv2	15M (Small)	`Smith42/astroPT_v2.0`
		95M (Base)	`Smith42/astroPT_v2.0`
		850M (Large)	`Smith42/astroPT_v2.0`
	ConvNeXtv2	15M (Nano)	`facebook/convnextv2-nano-22k-224`
		28M (Tiny)	`facebook/convnextv2-tiny-22k-224`
		89M (Base)	`facebook/convnextv2-base-22k-224`
		198M (Large)	`facebook/convnextv2-large-22k-224`
	DINOv2	22M (Small)	`facebook/dinov2-with-registers-small`
		86M (Base)	`facebook/dinov2-with-registers-base`
		304M (Large)	`facebook/dinov2-with-registers-large`
		1.1B (Giant)	`facebook/dinov2-with-registers-giant`
	IJEPA	630M (Huge)	`facebook/ijepa_vith16_22k`
		1.0B (Giant)	`facebook/ijepa_vitg14_22k`
	Specformer	43M (Base)	`polymathic-ai/specformer`
	ViT	86M (Base)	`google/vit-base-patch16-224-in21k`
		304M (Large)	`google/vit-large-patch16-224-in21k`
		632M (Huge)	`google/vit-huge-patch14-224-in21k`
Crossmatches	JWST vs HSC	1.67k	`Smith42/jwst_hsc_crossmatched`
	Legacy vs HSC	102k	`Smith42/legacysurvey_hsc_crossmatched`
	DESI vs HSC	18.6k	`Smith42/desi_hsc_crossmatched`
	SDSS vs HSC	2.32k	`Smith42/sdss_hsc_crossmatched`
Embeddings	JWST vs HSC	1.67k	`UniverseTBD/jwst_hsc_embeddings`
	Legacy vs HSC	102k	`UniverseTBD/legacysurvey_hsc_embeddings`
	DESI vs HSC	18.6k	`UniverseTBD/desi_hsc_embeddings`
	SDSS vs HSC	2.32k	`UniverseTBD/sdss_hsc_embeddings`