Integral Field Unit Spectroscopy with One Fiber

Zehao Peng (彭则皓)[1,2], **Biprateep Dey**[1,2]*, **Chris J. Maddison**[1,2], **Joshua S. Speagle** (沈佳士)[1]

[1]University of Toronto [2]Vector Institute

Abstract

Integral field unit (IFU) spectroscopy provides spatially resolved spectra across galaxies, offering crucial insights into their evolution. However, its high observational cost limits current IFU datasets to $\sim 10^4$ objects. We present a multi-modal, probabilistic foundation model that predicts high-resolution spectra with calibrated uncertainties at arbitrary spatial locations within a galaxy directly from broadband images. Built on a masked autoencoder framework, our architecture injects fiber positional encodings and redshift aware wavelength encodings, enabling spatially conditioned predictions. Trained on 4.7 million images and single fiber spectroscopic observations from the Dark Energy Spectroscopic Instrument (DESI) survey, our model exploits the natural variance of fiber placements and the morphological self-similarity of galaxies to achieve IFU-like capabilities without any IFU training data. Predicted emission line flux maps match independent IFU observations from the Mapping Nearby Galaxies at APO (MaNGA) survey, with performance comparable to a supervised baseline trained directly on IFU data.

1 Introduction

Understanding galaxy evolution requires characterizing galaxy properties such as chemical composition, star formation rates, and stellar ages, both at the population level and also within individual galaxies. Spatially resolved mappings of these features can reveal critical variations in stellar populations, gas kinematics, and chemical abundances across galactic discs and bulges. This allows us to investigate the relationship between star formation and galactic structure, as well as the impact of the surrounding environment on local regions within a galaxy.

Large astronomical surveys, such as the Sloan Digital Sky Survey (SDSS; York et al., 2000) and the Dark Energy Spectroscopic Instrument survey (DESI; Dey et al., 2019; DESI Collaboration et al., 2022), collect broadband images that reveal morphology and single-aperture spectra for tens of millions of objects. These spectra provide detailed population properties, but lack spatial resolution within individual galaxies. Integral field unit (IFU) spectroscopy captures these physical details as spatially resolved spectral datacubes; however, the high cost of allocating numerous optical fibers per target limits contemporary IFU datasets, such as Mapping Nearby Galaxies at APO (MaNGA; Bundy et al., 2015), to roughly 10^4 galaxies. Therefore, bridging the gap between the statistical volume of these broad surveys and the complete spatio-spectral resolution provided by IFUs is essential for unlocking the full scientific potential of modern astronomical observations.

Previous works have applied advances in self-supervised foundation models to bridge this observational gap by predicting spectra directly from broadband images (Wu and Peek, 2020; Parker et al., 2025). However, all existing frameworks to our knowledge produce a single spectrum per galaxy, typically an integrated or object-center prediction, without exploiting the natural variance of fiber placement locations present in large spectroscopic data sets, or capitalizing on the morphological self-similarity of galaxies to construct fully resolved spatio-spectral representations. Furthermore,

*Corresponding author: Biprateep Dey (biprateep@pitt.edu)

Conference on Physics and AI at Stanford University (PAI 2026).

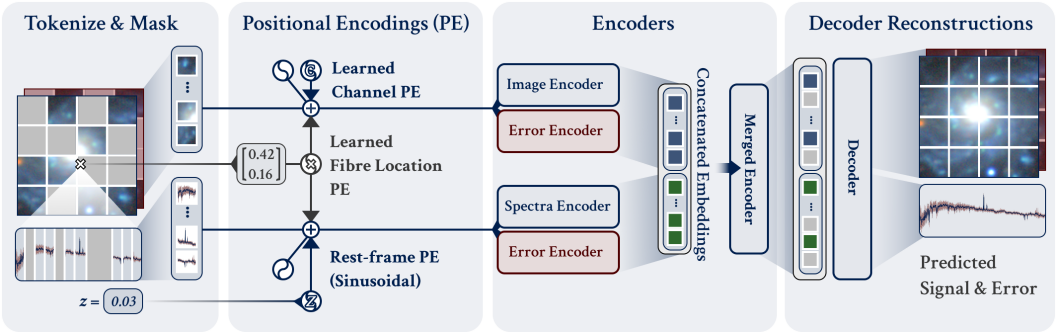

Figure 1: **Schematic representation of our multi-modal Masked Autoencoder architecture.** The model processes six inputs: a galaxy's image, spectrum, their respective measurement errors, the on-sky fiber location and redshift. Following independent tokenization and masking, the image and spectral sequences receive specialized positional encodings (PE). The fiber location is injected into both modalities for spatial conditioning. Spectral tokens additionally incorporate a redshift-informed sinusoidal PE to establish a rest-frame latent space. Finally, these encoded sequences are concatenated and passed through a merged cross-modal encoder, followed by a joint decoder that reconstructs the full images and spectra alongside their predicted uncertainties.

current architectures neither account for heteroscedastic observational uncertainties nor provide calibrated predictive uncertainties, which significantly limits their scientific utility.

In this work, we present a novel uncertainty-aware, multi-modal foundation model based on the self-supervised Masked Autoencoder (MAE; He et al., 2021) framework. To effectively bridge the imaging-spectroscopy gap, our architecture enforces cross-modal fusion through masked reconstructions and uniquely incorporates both fiber location and cosmological redshift as positional encodings. This spatial and spectral conditioning establishes a coherent rest-frame latent space and enables spatially resolved predictions of spectra. By actively modeling input measurement and output prediction uncertainties to produce well-calibrated probabilistic reconstructions, our framework synthesizes robust representations that demonstrate powerful generalization capabilities.

2 Data and Methods

Data Sets Used We cross-match photometry from the DESI Legacy Imaging Survey Data Release 9 (DR9) with spectra from DESI DR1 (DESI Collaboration et al., 2025), selecting spectroscopically confirmed galaxies and quasars with extinction-corrected r-band magnitudes less than 19.5 and reliable redshift measurements ($\Delta\chi^2 > 40$). For each object, we extract 128×128 pixel image cutouts across 5 or 6 channels (g, r, z, $W1$, $W2$, plus DR10 z-band when available) alongside corresponding spectra spanning 3600–9824 Å (7781 pixels at 0.8 Å/pixel). To enable spatial conditioning and uncertainty modeling, we retain precise on-sky coordinates for both images and spectroscopic fibers, and convert pixel-wise inverse variances into standard measurement errors. We mask missing channels and non-finite measurements. The final curated dataset comprises 4.7 million objects, with 98% training, 1% held-out validation for hyperparameter selection, and 1% held-out testing.

Tokenization and Positional Encoding Following vision transformers (ViT; Dosovitskiy et al., 2021), images are divided into 16×16 pixel patches per channel and spectra into 31-pixel patches, with fluxes and errors tokenized separately and projected to a 256 dimensional embedding. Channels are tokenized independently to enable direct cross-channel mixing and explicit association between spectral features and broadband morphology. To encourage robust cross-modal representation learning, we independently mask the image and spectral sequences. Image tokens receive a fixed 2D sinusoidal spatial encoding and a learnable channel encoding (Gehring et al., 2017). We explicitly inject the spectroscopic fiber's on-sky location as a fiber positional encoding by projecting its relative spatial offset from the image center via a learned linear transformation, adding this to each image token. For the spectrum, we directly add the 2D sinusoidal spatial encoding of the fiber's center to all spectral tokens, establishing a spatially conditioned mapping. Because galaxies at different redshifts have their spectra shifted to different observed wavelengths, we use similar sinusoidal encodings to inject the rest-frame wavelengths, enforcing a latent representation that is invariant to redshifts.

Model Architecture Our Transformer-based (Vaswani et al., 2017) masked autoencoder randomly removes tokens from the encoder sequence and is trained to reconstruct the complete data. Masking is applied independently to each modality. To avoid overfitting to a fixed masking regime, we dynamically sample the mask ratio $r \in \{0, 1/14, \ldots, 13/14, 1\}$, with $P(r = 1) = 0.3$ and uniform probability 0.05 otherwise. This also disproportionately exposes the model to cases where an entire modality is missing, encouraging cross-modal reconstruction. In addition, each ratio is paired with a contiguous chunk size $c \in \{1, 2, 4, 8, 16, 32, 64, 128\}$, chosen to avoid pathological masking patterns. Modality-specific encoders (4 layers, 8 heads) process masked sequences separately; their outputs are concatenated and passed through a merged cross-modal encoder (4 layers, 8 heads). For reconstruction, sequences are projected to 512 dimensions, padded with learned mask tokens, re-injected with positional encodings, and processed by a joint decoder (8 layers, 16 heads) that routes to four feed-forward heads reconstructing image fluxes, image errors, spectral fluxes, and spectral errors. Figure 1 depicts our architecture and full model details are available in our code repository[2].

Loss Function Following Laroche and Speagle (2023), we formulate a probabilistic reconstruction objective for heteroscedastic noise. Let $\mathbf{x}_{\text{obs}}$ denote the observed fluxes with pixel-wise known variances $\boldsymbol{\sigma}^2_{\text{obs}}$. We model the true underlying flux as a normal distribution centered on our prediction $\mathbf{x}_{\text{pred}}$ with learned predictive variance $\boldsymbol{\sigma}^2_{\text{pred}}$. This yields a χ^2-type loss function:

$$\sum \left[\frac{(\mathbf{x}_{\text{obs}} - \mathbf{x}_{\text{pred}})^2}{\boldsymbol{\sigma}^2_{\text{obs}} + \boldsymbol{\sigma}^2_{\text{pred}}} + \log(\boldsymbol{\sigma}^2_{\text{obs}} + \boldsymbol{\sigma}^2_{\text{pred}}) \right]$$

summed over all image and spectral pixels. The log normalization term penalizes variance inflation where observed data exists. To discourage the model from over-inflating $\boldsymbol{\sigma}^2_{\text{pred}}$ on masked patches where $\mathbf{x}_{\text{obs}}$ is unavailable, we apply a complementary L_2 regularization penalty directly to $\boldsymbol{\sigma}^2_{\text{pred}}$ in those regions. During training, we apply random spatial translations to simulate varied off-center fiber placements. The model is optimized using Adam (Kingma and Ba, 2015) with a cosine warmup learning rate schedule.

3 Results and Discussion

Our model achieves strong reconstruction performance across image and spectral generation tasks over a wide range of masking ratios. Figure 2 (right) demonstrates that the model accurately reconstructs DESI spectra at precise fiber locations, recovering global continuum shapes and local structures, such as oxygen doublets and Ca H and K lines, directly from the broadband image. Because the model explicitly accounts for observational uncertainties, the predicted spectrum represents a denoised inference of the true underlying spectrum.

By systematically scanning the input fiber location across the image at the spatial resolution of the MaNGA IFU, we generate a full synthetic datacube directly from a single broadband image. The predicted spatially resolved emission-line maps and individual spaxel spectra adhere closely to independent MaNGA Data Release 17 ground-truth observations, with slight under-prediction of Hα fluxes in select star-forming regions. To quantify performance, we compare our zero-shot predictions against a supervised ridge regression baseline trained on image patches paired with MaNGA Hα flux measurements across 470 galaxies. On a held-out test set of 48 galaxies, our foundation model achieves comparable or superior performance to this baseline (Pearson $r \sim 0.71$ for both, and normalized RMSE 1.51 vs. 1.63 for the baseline). This proves that unsupervised cross-modal pretraining effectively captures the necessary spatio-spectral structures, successfully generalizing to completely unseen data in a zero-shot capacity.

We further evaluate the calibration of the model's uncertainty estimates on the held-out test set. The distribution of prediction errors weighted by total variance is expected to follow $\mathcal{N}(0, 1)$ if the uncertainty estimates are calibrated. Across varying masking ratios, we find that the distribution resembles a Gaussian with a standard deviation between 1.01 and 1.02 for spectra and 0.74 to 0.99 for images, and the median ranges from 0.01 to 0.04 across both modalities. Although the spectral predictions are slightly over-confident and image predictions moderately under-confident, the uncertainties remain well-calibrated overall.

[2] `https://github.com/harrypenguin/MultiModal`

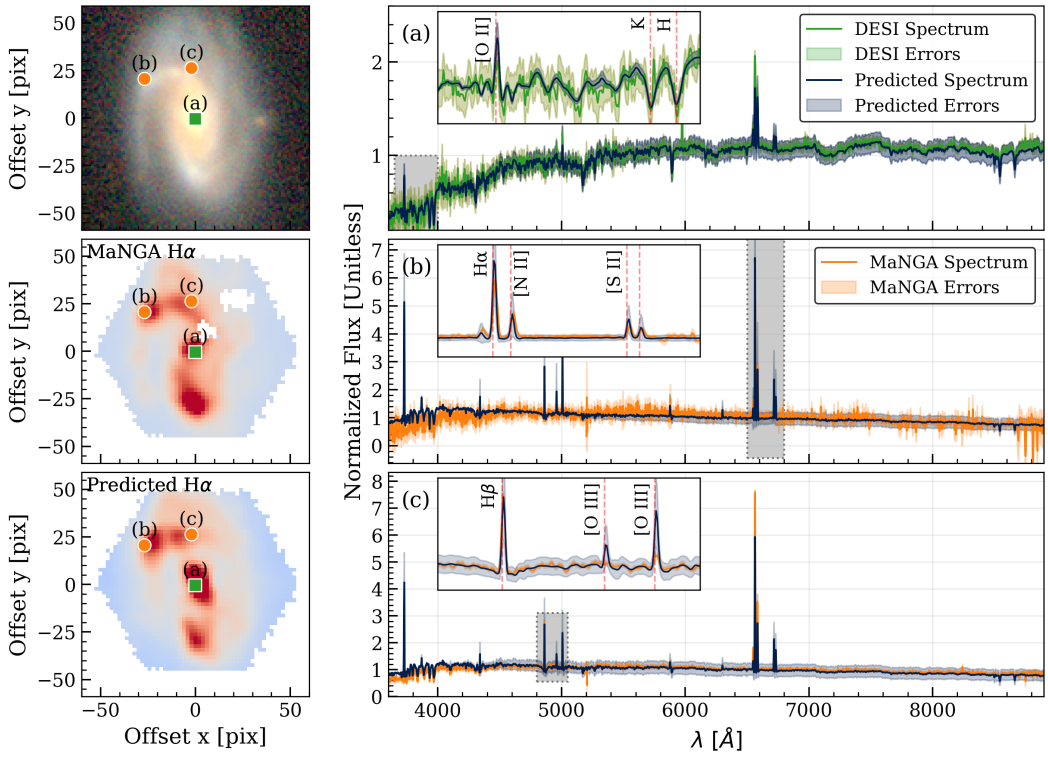

Figure 2: **Predicted spectra at various spatial locations for an example galaxy. Left Column**: (Top) DESI Legacy Survey image of DESI TARGETID 39628444518058848 (MaNGA-ID 9042-9101) at redshift ~ 0.07. The model is trained to infer spectra based solely on the broadband image, the redshift of the object, and a specified fiber location. (Middle) True Hα flux map of the object, derived from the MaNGA IFU datacube. (Bottom) Reconstructed Hα flux map derived from spectra predicted by our model. These spectra are generated using only the image and an input fiber location that is systematically scanned across the image at the spatial resolution of the MaNGA IFU. Three spaxels are labelled (a), (b), and (c). Our model qualitatively recovers the Hα map and successfully resolves star-forming regions along the galactic arms, denoted by localized hotspots. **Right Column**: Each panel corresponds to a chosen spaxel, displaying the predicted spectrum and uncertainties in dark blue. Panel (a) compares the model prediction to the available DESI spectrum observed at the central location (green), whereas panels (b) and (c) compare off-center predictions to the corresponding spectra from the MaNGA IFU (orange). The insets zoom into areas of interest to highlight the accurate reconstruction of prominent emission and absorption features. Notably, even though our model was never trained on MaNGA data, it successfully reconstructs off-center spectra and captures the spatially varying physical properties of the galaxy. These predictions trained on DESI data remarkably match a completely independent dataset (up to a normalization constant from flux calibration differences between the two instruments).

This near-IFU capability reflects a powerful inductive bias learned from the training data distribution. Across 4.7 million DESI observations with naturally varying off-center fiber placements, the model learns that morphologically similar sub-regions (e.g., spiral arms, bulges, and star-forming regions) produce similar spectra regardless of their host galaxy. While this self-similarity assumption is well-grounded for the bulk galaxy population, performance may degrade for morphologically peculiar systems such as ongoing mergers or AGN-dominated galaxies.

Currently, our model ingests redshift as a fixed, uniform input across the spatial dimensions of a given galaxy. This flattens the distribution along the velocity axis and precludes the generation of kinematic maps. This limitation could be addressed by masking the redshift during training and introducing a corresponding reconstruction objective. Additionally, our assumption of Gaussian uncertainty is rigid; future work should explore more flexible probabilistic inference frameworks, such as flow matching or diffusion models.

In conclusion, we demonstrate for the first time the prediction of high-resolution, spatially resolved spectra with calibrated uncertainties at arbitrary locations within a galaxy directly from broadband imaging. By exploiting the natural variance of fiber placements and galactic self-similarity, our model develops near-IFU capabilities using only single-aperture training data. Combined with reliable

predictive uncertainties, this approach can significantly expand the scientific yield of current and future photometric surveys to study localized galactic properties. The predictions from our method could enable us to examine specific regions within a galaxy and assess how the environment influences local star formation. Such data sets could also uncover radial gradients in age and metallicity and allow us to study the link between star formation and galactic structure for galaxies without existing IFU data. Furthermore, our method holds substantial potential for broader multi-modal applications in fields such as remote sensing and biology.

Acknowledgments and Disclosure of Funding

B. Dey is a postdoctoral fellow at the University of Toronto in the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship Program, a program of Schmidt Sciences.

The authors advocate for the judicious and ethical use of artificial intelligence in scientific discourse. In accordance with the American Physical Society's guidelines on appropriate use of AI-based writing tools, we confirm that AI assistance was utilized in the preparation of this statement solely to polish, condense, and edit our original ideas. The authors retain full responsibility for the accuracy, integrity, and content of this document.

This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02- 05CH11231 using NERSC award HEP-ERCAP0033572.

Resources used in preparing this research were provided, in part, by the Province of Ontario, the Government of Canada through CIFAR, and companies sponsoring the Vector Institute (`www.vectorinstitute.ai/partnerships/`).

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 Humanities, Science and Technology of Mexico (CONAHCYT); 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 I'oligam 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.

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. Pipeline processing and analyses of the data were supported by NOIRLab and the Lawrence Berkeley National Laboratory (LBNL). 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.

NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. LBNL is managed by the Regents of the University of California under contract to the U.S. Department of Energy.

This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, 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, Fundacao Carlos Chagas Filho de Amparo, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Cientifico e Tecnologico and the Ministerio da Ciencia, Tecnologia e Inovacao, 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 Energeticas, Medioambientales y Tecnologicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenossische Technische Hochschule (ETH) Zurich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciencies de l'Espai (IEEC/CSIC), the Institut de Fisica d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig Maximilians Universitat Munchen and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

BASS is a key project of the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories of China, the Chinese Academy of Sciences (the Strategic Priority Research Program "The Emergence of Cosmological Structures" Grant # XDB09000000), and the Special Fund for Astronomy from the Ministry of Finance. The BASS is also supported by the External Cooperation Program of Chinese Academy of Sciences (Grant # 114A11KYSB20160057), and Chinese National Natural Science Foundation (Grant # 12120101003, # 11433005).

The Legacy Survey team makes use of data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), which is a project of the Jet Propulsion Laboratory/California Institute of Technology. NEOWISE is funded by the National Aeronautics and Space Administration.

The Legacy Surveys imaging of the DESI footprint is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH1123, by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; and by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to NOAO.

References

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. doi: 10.1088/0004-637X/798/1/7.

DESI Collaboration, B. Abareshi, J. Aguilar, S. Ahlen, Shadab Alam, David M. Alexander, R. Alfarsy, L. Allen, C. Allende Prieto, O. Alves, J. Ameel, E. Armengaud, J. Asorey, Alejandro Aviles, S. Bailey, A. Balaguera-Antolínez, O. Ballester, C. Baltay, A. Bault, S. F. Beltran, B. Benavides,

S. BenZvi, A. Berti, R. Besuner, Florian Beutler, D. Bianchi, C. Blake, P. Blanc, R. Blum, A. Bolton, S. Bose, D. Bramall, S. Brieden, A. Brodzeller, D. Brooks, C. Brownewell, E. Buckley-Geer, R. N. Cahn, Z. Cai, R. Canning, R. Capasso, A. Carnero Rosell, P. Carton, R. Casas, F. J. Castander, J. L. Cervantes-Cota, S. Chabanier, E. Chaussidon, C. Chuang, C. Circosta, S. Cole, A. P. Cooper, L. da Costa, M.-C. Cousinou, A. Cuceu, T. M. Davis, K. Dawson, R. de la Cruz-Noriega, A. de la Macorra, A. de Mattia, J. Della Costa, P. Demmer, M. Derwent, A. Dey, B. Dey, G. Dhungana, Z. Ding, C. Dobson, P. Doel, J. Donald-McCann, J. Donaldson, K. Douglass, Y. Duan, P. Dunlop, J. Edelstein, S. Eftekharzadeh, D. J. Eisenstein, M. Enriquez-Vargas, S. Escoffier, M. Evatt, P. Fagrelius, X. Fan, K. Fanning, V. A. Fawcett, S. Ferraro, J. Ereza, B. Flaugher, A. Font-Ribera, J. E. Forero-Romero, C. S. Frenk, S. Fromenteau, B. T. Gänsicke, C. Garcia-Quintero, L. Garrison, E. Gaztañaga, F. Gerardi, H. Gil-Marín, S. Gontcho A Gontcho, Alma X. Gonzalez-Morales, G. Gonzalez-de-Rivera, V. Gonzalez-Perez, C. Gordon, O. Graur, D. Green, C. Grove, D. Gruen, G. Gutierrez, J. Guy, C. Hahn, S. Harris, D. Herrera, Hiram K. Herrera-Alcantar, K. Honscheid, C. Howlett, D. Huterer, V. Iršič, M. Ishak, P. Jelinsky, L. Jiang, J. Jimenez, Y. P. Jing, R. Joyce, E. Jullo, S. Juneau, N. G. Karaçaylı, M. Karamanis, A. Karcher, T. Karim, R. Kehoe, S. Kent, D. Kirkby, T. Kisner, F. Kitaura, S. E. Koposov, A. Kovács, A. Kremin, Alex Krolewski, B. L'Huillier, O. Lahav, A. Lambert, C. Lamman, Ting-Wen Lan, M. Landriau, S. Lane, D. Lang, J. U. Lange, J. Lasker, L. Le Guillou, A. Leauthaud, A. Le Van Suu, Michael E. Levi, T. S. Li, C. Magneville, M. Manera, Christopher J. Manser, B. Marshall, Paul Martini, W. McCollam, P. McDonald, Aaron M. Meisner, J. Mena-Fernández, J. Meneses-Rizo, M. Mezcua, T. Miller, R. Miquel, P. Montero-Camacho, J. Moon, J. Moustakas, E. Mueller, Andrea Muñoz-Gutiérrez, Adam D. Myers, S. Nadathur, J. Najita, L. Napolitano, E. Neilsen, Jeffrey A. Newman, J. D. Nie, Y. Ning, G. Niz, P. Norberg, Hernán E. Noriega, T. O'Brien, A. Obuljen, N. Palanque-Delabrouille, A. Palmese, P. Zhiwei, D. Pappalardo, X. PENG, W. J. Percival, S. Perruchot, R. Pogge, C. Poppett, A. Porredon, F. Prada, J. Prochaska, R. Pucha, A. Pérez-Fernández, I. Pérez-Ràfols, D. Rabinowitz, and A. Raichoor. Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument. AJ, 164(5):207, November 2022. doi: 10.3847/1538-3881/ac882b.

DESI Collaboration, M. Abdul Karim, A. G. Adame, D. Aguado, J. Aguilar, S. Ahlen, S. Alam, G. Aldering, D. M. Alexander, R. Alfarsy, L. Allen, C. Allende Prieto, O. Alves, A. Anand, U. Andrade, E. Armengaud, S. Avila, A. Aviles, H. Awan, S. Bailey, A. Baleato Lizancos, O. Ballester, A. Bault, J. Bautista, R. Bean, J. Behera, S. BenZvi, L. Beraldo e Silva, J. R. Bermejo-Climent, F. Beutler, D. Bianchi, C. Blake, R. Blum, A. S. Bolton, M. Bonici, S. Brieden, A. Brodzeller, D. Brooks, E. Buckley-Geer, E. Burtin, A. Byström, R. Canning, A. Carnero Rosell, A. Carr, P. Carrilho, L. Casas, F. J. Castander, R. Cereskaite, J. L. Cervantes-Cota, E. Chaussidon, J. Chaves-Montero, S. Chen, X. Chen, C. Circosta, T. Claybaugh, S. Cole, A. P. Cooper, M.-C. Cousinou, A. Cuceu, T. M. Davis, K. S. Dawson, R. de Belsunce, R. de la Cruz, A. de la Macorra, A. de Mattia, N. Deiosso, J. Della Costa, R. Demina, U. Demirbozan, J. DeRose, A. Dey, B. Dey, J. Ding, Z. Ding, P. Doel, K. Douglass, M. Dowicz, H. Ebina, J. Edelstein, D. J. Eisenstein, W. Elbers, N. Emas, S. Escoffier, P. Fagrelius, X. Fan, K. Fanning, G. Favole, V. A. Fawcett, E. Fernández-García, S. Ferraro, N. Findlay, A. Font-Ribera, J. E. Forero-Romero, D. Forero-Sánchez, C. S. Frenk, B. T. Gänsicke, L. Galbany, J. García-Bellido, C. Garcia-Quintero, L. H. Garrison, E. Gaztañaga, H. Gil-Marín, A. Gloudemans, O. Y. Gnedin, S. Gontcho A Gontcho, D. Gonzalez, A. X. Gonzalez-Morales, V. Gonzalez-Perez, C. Gordon, O. Graur, D. Green, D. Gruen, R. Gsponer, C. Guandalin, G. Gutierrez, J. Guy, C. Hahn, J. J. Han, J. Han, S. He, H. K. Herrera-Alcantar, S. Heydenreich, K. Honscheid, J. Hou, C. Howlett, D. Huterer, V. Iršič, M. Ishak, A. Jacques, L. Jiang, J. Jimenez, Y. P. Jing, B. Joachimi, S. Joudaki, R. Joyce, E. Jullo, S. Juneau, N. G. Karaçaylı, T. Karim, R. Kehoe, S. Kent, A. Khederlarian, D. Kirkby, T. Kisner, F.-S. Kitaura, N. Kizhuprakkat, H. Kong, S. E. Koposov, A. Kremin, A. Krolewski, O. Lahav, Y. Lai, C. Lamman, T.-W. Lan, M. Landriau, D. Lang, J. U. Lange, J. Lasker, J. M. Le Goff, L. Le Guillou, A. Leauthaud, M. E. Levi, S. Li, T. S. Li, W. Liu, K. Lodha, M. Lokken, Y. Luo, Y. Luo, C. Magneville, M. Manera, C. J. Manser, D. Margala, P. Martini, M. Maus, J. McCullough, P. McDonald, G. E. Medina, L. Medina-Varela, A. Meisner, J. Mena-Fernández, A. Menegas, J. Meneses-Rizo, M. Mezcua, R. Miquel, P. Montero-Camacho, J. Moon, J. Moustakas, A. Muñoz-Gutiérrez, D. Muñoz-Santos, A. D. Myers, J. Myles, S. Nadathur, J. Najita, L. Napolitano, J. A. Newman, F. Nikakhtar, R. Nikutta, G. Niz, and H. E. Noriega. Data Release 1 of the Dark Energy Spectroscopic Instrument. *arXiv e-prints*, art. arXiv:2503.14745, March 2025. doi: 10.48550/arXiv.2503.14745.

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. doi: 10.3847/1538-3881/ab089d.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021. URL `https://arxiv.org/abs/2010.11929`.

Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. Convolutional sequence to sequence learning. In *International conference on machine learning*, pages 1243–1252. PMLR, 2017.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked Autoencoders Are Scalable Vision Learners. *arXiv e-prints*, art. arXiv:2111.06377, November 2021. doi: 10.48550/arXiv.2111.06377.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. URL `https://arxiv.org/abs/1412.6980`.

Alexander Laroche and Joshua S. Speagle. Closing the stellar labels gap: An unsupervised, generative model for *Gaia* BP/RP spectra. *arXiv e-prints*, art. arXiv:2307.06378, July 2023. doi: 10.48550/arXiv.2307.06378.

Liam Parker, Francois Lanusse, Jeff Shen, Ollie Liu, Tom Hehir, Leopoldo Sarra, Lucas Meyer, Micah Bowles, Sebastian Wagner-Carena, Helen Qu, Siavash Golkar, Alberto Bietti, Hatim Bourfoune, Nathan Casserau, Pierre Cornette, Keiya Hirashima, Geraud Krawezik, Ruben Ohana, Nicholas Lourie, Michael McCabe, Rudy Morel, Payel Mukhopadhyay, Mariel Pettee, Bruno Regaldo-Saint Blancard, Kyunghyun Cho, Miles Cranmer, and Shirley Ho. AION-1: Omnimodal Foundation Model for Astronomical Sciences. *arXiv e-prints*, art. arXiv:2510.17960, October 2025. doi: 10.48550/arXiv.2510.17960.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

John F. Wu and J. E. G. Peek. Predicting galaxy spectra from images with hybrid convolutional neural networks. *arXiv e-prints*, art. arXiv:2009.12318, September 2020. doi: 10.48550/arXiv.2009.12318.

Donald G. York, J. Adelman, John E. Anderson, Jr., 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. doi: 10.1086/301513.