Combining datasets with different ground truths using Low-Rank Adaptation to generalize image-based CNN models for photometric redshift prediction

Vikram Seenivasan[1]
vikrams25@ucla.edu

Srinath Saikrishnan[2]
srinathsai22@ucla.edu

Andrew Lizarraga[3]
andrewlizarraga@ucla.edu

Jonathan Soriano[1]
jsoriano@astro.ucla.edu

Bernie Boscoe[4]
boscoeb@sou.edu

Tuan Do[1]
tdo@astro.ucla.edu

[1] Department of Physics and Astronomy, UCLA, Los Angeles, CA 90095
[2] Department of Computer Science, UCLA, Los Angeles, CA 90095
[3] Department of Statistics and Data Science, UCLA, Los Angeles, CA 90095
[4] Department of Computer Science, Southern Oregon University, Ashland, OR, 97520

Abstract

In this work, we demonstrate how Low-Rank Adaptation (LoRA) can be used to combine different galaxy imaging datasets to improve redshift estimation for cosmology with CNN models. LoRA is an established technique for large language models that adds adapter networks to adjust model weights and biases to efficiently fine-tune large base models without retraining. We train a base model using a photometric redshift ground truth dataset, which contains broad galaxy types but is less accurate. We then fine-tune using LoRA on a spectroscopic redshift ground truth dataset. These are more accurate but limited to bright galaxies and take orders of magnitude more time to obtain, so are less available for large surveys. Ideally, the combination of the two datasets would yield more accurate models that generalize well. The LoRA model performs better than a traditional transfer learning method, with $\sim 2.5\times$ less bias and $\sim$2.2$\times$ less scatter. Retraining the model on a combined dataset yields a model that generalizes better than LoRA but at a cost of greater computation time. Our work shows that LoRA is useful for fine-tuning regression models in astrophysics by providing a middle ground between full retraining and zero-shot generalization. LoRA shows potential in allowing us to leverage existing pretrained astrophysical models, especially for data sparse tasks.

1 Introduction

Astronomy is in the era of Big Data, with several current and upcoming surveys such as LSST [13] and Euclid [17] imaging billions of galaxies to constrain cosmological models governing the origin and evolution of the universe. Galaxy redshifts are used to determine distances to galaxies, allowing us to map the structure of the universe through cosmic time, to determine properties of dark energy and dark matter that govern the accelerated expansion of the universe. Astrophysicists are increasingly adopting machine learning techniques to efficiently make measurements, such as galaxy redshifts, from large datasets [26, 20]. Such measurements often require leveraging different sources of ground truth. In this paper, we explore using Low-Rank adaptation (LoRA)[12] and traditional transfer learning to fine-tune redshift prediction models for galaxies by integrating datasets with different sources of ground truth.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2025.

1.1 Related Works

Redshift estimation: Traditionally, redshifts are obtained using spectroscopy. These are highly accurate but limited to bright galaxies that have strong emission lines, and are not available at the scale of large upcoming surveys as they are expensive [20]. Photometric redshifts are derived from galaxy *photometry* (brightness) in different wavelength bands to sample spectral properties. These are less accurate [2] but enable the analysis of much larger datasets. Recently, many machine learning models have been proposed to estimate photometric redshifts [8, 6, 4, 2, 25, 14, 15, 26, 27]. For example, convolution neural networks (CNN) produce some of the most precise redshift estimates [21, 27, 9, 16] by using images directly. These methods usually rely on spectroscopic redshift as ground truth for training. Recent work by [5, 29] explored methods of incorporating photometric redshifts from multi-band photometry as a different source of ground truth. However, these methods are trained on photometry alone, losing out on the wealth of information in images [18].

LoRA and fine-tuning: LoRA is an established technique for fine-tuning large language models that adds adapter networks with lower-rank weight matrices to pre-trained models. This allows efficient fine-tuning of large base models, with only a tiny fraction of the parameters being updated [12]. Low-Rank adaptation has been used in astrophysics to fine-tune pre-trained models from other domains for astrophysics tasks. For example, LLMs for token-based redshift estimation [23], foundation models for radio astronomy [24], and diffusion models for galaxy image generation [3]. Its utility for regression-based tasks is not fully explored in astrophysics.

1.2 Contribution

We explore using LoRA and transfer learning to integrate different sources of ground truth for a redshift regression CNN that takes 5-band galaxy images as input. We adapt the recently developed LoRA implementations [19] for CNNs for astrophysical data. To our knowledge, this work is the first application of LoRA for regression-based tasks for astrophysics. Previous work in astrophysics used pre-trained models from other domains– this is the first work that uses baseline models trained on astronomical data. These are key considerations vital for large-scale surveys that will leverage previous astrophysics knowledge.

2 Data

In this work we use three datasets, one for the base model, one for transfer learning, and a combination dataset of the two. The dataset used to train the base model is derived from TransferZ [29] with the addition of 5-band images from the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP) Second Public Data Release [1] and additional quality cuts as specified in [10]. This results in a dataset we call TransferZ-Images [1] consisting of 100,442 galaxies with 5-band *g, r, i, z, y* images. We use the GalaxiesML[10] [2]dataset for fine-tuning. These datasets have been used for redshift estimation in the context of cosmology, and are ML-ready and complement each other by covering different galaxy color types and redshift ranges.

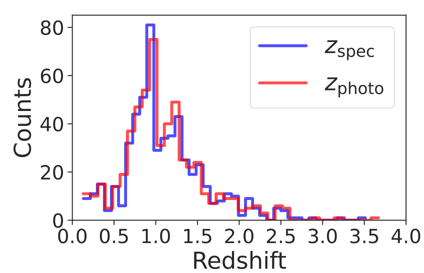

Figure 1: The redshift number distributions of the 557 common galaxies using spectroscopic ground truth from GalaxiesML (blue) and photometric ground truth (red).

The TransferZ-Images dataset has redshift ground truths from photometric redshift estimates while GalaxiesML has ground truth from spectroscopy. The TransferZ-Images photometric redshifts are less accurate than GalaxiesML but contain a fainter and larger variation in galaxy types. Ideally, it is a good dataset to train the base model. We aim to fine-tune on GalaxiesML to improve the base model with more accurate redshifts. In principle, the combination should produce a model that is more accurate and generalizable. We additionally create the Combo dataset by combining TransferZ-Images and GalaxiesML, resulting

[1]TransferZ-Images is available on Zenodo here.
[2]GalaxiesML is available on Zenodo here.

in 386,286 galaxies. For the 557 common galaxies in both datasets we use the GalaxiesML spectroscopic redshift as ground truth. Additionally, comparing the spectroscopic and photometric ground truth redshifts for the overlapping galaxies, we find no systematic bias in both redshift distributions (see Fig. 1). The data used in this study is summarized in Table 1. The three datasets are split into 80% training, 10% validation and 10% testing subsets.

Table 1: Data Summary.

Dataset	Ground Truth z	No. of Sources	z 90th %ile	Median z Uncertainty	i-mag 90th %ile	No. Filters
TransferZ-Images	z_{phot}	100,442	1.8	0.03	25	5
GalaxiesML	z_{spec}	286,401	1.2	0.0002	22	5
Combo	z_{phot} and z_{spec}	386,286	1.4	0.0006	24	5

3 Methodology and Metrics

We train four models for this study to compare the effectiveness of LoRA to generalize redshift estimation models. These are (1) the base model (CNN-Base) trained on TransferZ; (2) the traditionally transfer learned model, built from fine-tuning CNN-base on GalaxiesML (CNN-TL); (3) the LoRA model, built from fine-tuning CNN-Base on GalaxiesML using LoRA (CNN-LoRA); and (4) the full retraining model (CNN-Combo) trained on Combo. These models share a ResNet18 architecture [11], modified to accept five images as input, with a regressor network attached to produce a single redshift value as output. We perform hyperparameter tuning on learning rate, batch size, and regressor architecture. The learning rate was the most sensitive parameter. The regressor consists of two linear layers with 512 and 256 neurons, with dropout layers (using standard dropout of 0.5) and SiLU activation between them to promote stable model training.

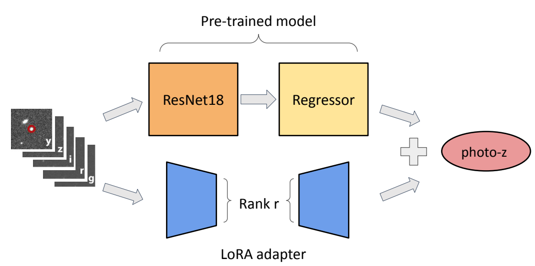

Figure 2: Visualization of LoRA implementation on a ResNet + regressor Model, based on Fig 1. in [12].

To train these models, we use a `ReduceLROnPlateau` scheduler with the same parameters as in [11], the Adam optimizer, and a custom loss function, $L(\Delta z) = 1 - \frac{1}{1+(\Delta z/0.15)^2}$, from [30]. This loss function encapsulates photometric redshift accuracy metrics, such as outlier rate, bias and scatter which are the most important metrics for cosmology [16](see below). We use an exponential moving average (EMA) of the validation loss to determine when to stop training. The models are set to train for 500 epochs with early stopping which ends training if the model's EMA validation loss does not improve after 20 epochs. We use an AMD Ryzen Threadripper PRO 3955WX with 16-Cores and NVIDIA RTX A6000 to train the models. For CNN-Base, we use a learning rate of 1e-3, and it stops training after 56 epochs taking 26.7 minutes with a final learning rate of 1e-8. This is used as a base to train CNN-TL and CNN-LoRA. To create CNN-TL, all the model weights in CNN-Base are frozen except the input and regressor layers, and it is trained on GalaxiesML with the final learning rate of 1e-8. It trains for the full 500 epochs, taking 4.4 hours. To create CNN-LoRA, we use the `peft` library (Parameter Efficient Fine-Tuning) [31]. We create a LoRA adapter that spans all the linear and convolutional layers in the model (Fig. 2). The adapter has a rank of 4, alpha parameter of 16 and dropout of 0.1 chosen from hyperparameter tuning. The LoRA adapter is trained with GalaxiesML, stopping in 121 epochs, taking 1.4 hours. Finally, CNN-Combo is trained with the Combo dataset and a learning rate of 1e-3 for 121 epochs without early stopping, taking 2.4 hours. We choose the same number of epochs for CNN-Combo as CNN-LoRA to quantify computational efficiency gains. All models are implemented in Pytorch [22]. The code is made available through Github.

To evaluate these models, we use the bias, scatter and catastrophic outlier rate metrics [30]. The bias metric is the median of the bias distribution defined as $b = (z_{pred} - z_{truth})/(1 + z_{truth})$ where z_{pred}

and z_{truth} are the predicted redshift and the ground truth redshift, respectively. The scatter is defined as the median absolute deviation of the bias distribution multiplied by 1.4826. The catastrophic outlier rate is the fraction of objects satisfying $|z_{photo} - z_{truth}| > 1.0$ [28]. We calculate the metrics in the redshift range $0.3 < z < 1.5$. These metrics are important for most cosmology science applications, in particular LSST requires bias $< |0.003|$ and scatter < 0.02 for photometric redshift estimates [7]. We qualitatively evaluate model "generalizability" based on how well it performs across all testing datasets.

4 Results and Discussion

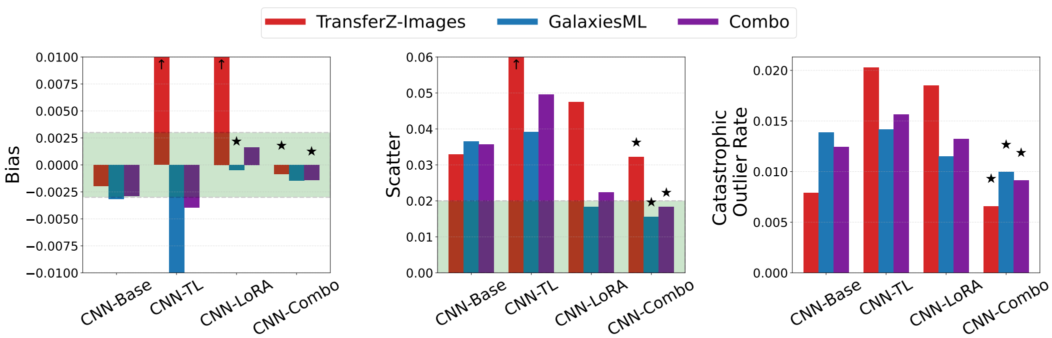

Figure 3: Bias (left), scatter (middle), and catastrophic outlier rate performance metrics (right) for all four models. The metrics are evaluated on all three datasets (TransferZ-Images (red), GalaxiesML (blue), Combo (purple)). The green band for Bias and Scatter are the LSST Science Requirements. The black stars represent the best (lowest) metric for each dataset. Upward black arrows indicates the bar overflows the plot. CNN-LoRA performs better than traditional transfer learning (CNN-TL), but not as well as retraining the entire dataset (CNN-Combo).

We find that Low-Rank adaptation (CNN-LoRA) results in more generalizable models than traditional transfer learning (CNN-TL), but does not perform as well as retraining on combined ground truths (CNN-Combo). We evaluate the performance of the four models by testing them on the test sets of 10,027 galaxies from TransferZ-Images, 40,914 galaxies from GalaxiesML and 50,871 galaxies from Combo. The model performance metrics are displayed in Fig. 3 (see Table 2 for values). CNN-LoRA has $\sim$2.5$\times$ less bias and $\sim$2.2$\times$ less scatter on Combo compared to CNN-TL. Compared to CNN-Combo, CNN-LoRA has similar bias ($\sim$1.1$\times$ higher) and scatter ($\sim$1.2$\times$ higher) on the Combo testing set, though slightly higher. When looking at the GalaxiesML fine-tuning dataset in particular, CNN-LoRA has $\sim$3$\times$ lower bias than CNN-Combo, and $\sim$1.2$\times$ higher scatter. However, CNN-LoRA's performance on the baseline data, TransferZ-Images, is worse, with $\sim$20$\times$ higher bias and $\sim$1.5$\times$ higher scatter. This indicates that the CNN-LoRA model forgot information it learned in the baseline training on TransferZ-Images. The choice of method of mixing ground truths has a trade-off between performance requirements and resources available. Thus, LoRA may be a good choice if retraining the model with combined data is too computationally intensive and diminished performance on the base dataset does not detract from science goals.

We perform a further test of CNN-LoRA by training the model with different fractions of GalaxiesML. Training with just 10% of GalaxiesML, the model's scatter on GalaxiesML improves significantly and continues improving until about a 40-50 % data fraction (Fig. 4). Beyond this, there is no significant improvement in model performance. Further, even with a 100% data fraction, CNN-LoRA trains in $\sim$1.5$\times$ less time than CNN-Combo when both models are trained to the same number of epochs. Training time is even lower with smaller data fractions. Thus, LoRA offers a way to fine-tune models with smaller fractions of data and still achieve maximum model performance on the fine-tuning dataset all while saving computational resources, although it still performs worse than retraining on combined ground truths.

When looking at transfer learning in particular, previous work suggests that neurons are co-adapted between model layers and later models layers can be highly specialized to baseline data [32]. CNN-LoRA is able to adjust weights and biases across the entire model with its learned adapters, resulting

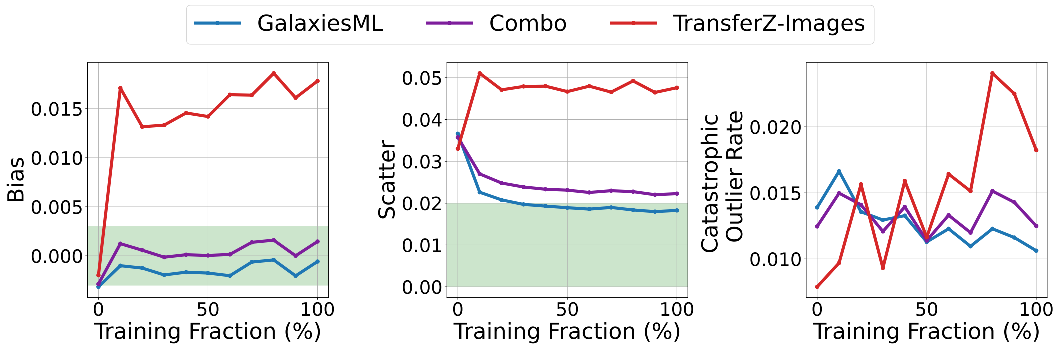

Figure 4: Bias (left), scatter (middle) and catastrophic outlier rate (right) as a function of fraction of GalaxiesML used to fine-tune using CNN-LoRA. The green band for Bias and Scatter are the LSST Science Requirements. Using just 10% of the data, the model's scatter on GalaxiesML improves and on TransferZ reduces drastically. The performance of CNN-LoRA plateaus when 40-50% of GalaxiesML is used for training; only a fraction of the dataset is needed for the best fine-tuning.

in its better performance compared to CNN-TL which only trains on the input and regressor layers. Thus, LoRA offers a standardized way to do transfer learning on models without the need for extensive optimization of layers to unfreeze in traditional transfer learning.

Table 2: Model Metrics Summary

		Bias	Scatter	Catastrophic Outlier Rate
CNN-Base	TransferZ-Images	-0.00198	0.0330	0.00789
	GalaxiesML	-0.00316	0.0366	0.0139
	Combo	-0.00287	0.0358	0.0125
CNN-TL	TransferZ-Images	0.0502	0.121	0.0203
	GalaxiesML	-0.00995	0.0391	0.0141
	Combo	-0.00396	0.0496	0.0157
CNN-LoRA	TransferZ-Images	0.0176	0.0476	0.0185
	GalaxiesML	-0.000481	0.0184	0.0115
	Combo	0.0016	0.0223	0.0132
CNN-Combo	TransferZ-Images	-0.000868	0.0323	0.00660
	GalaxiesML	-0.00146	0.0156	0.00996
	Combo	-0.00141	0.0183	0.00915
LSST Req.		$< \|0.03\|$	< 0.02	

Future work tuning LoRA parameters can further optimize its performance, and potentially rival CNN-Combo in performance with lower computational costs. An ablation study freezing and unfreezing different layers to create different CNN-TL models can also further characterize traditional transfer learning performance and compare against LoRA fine-tuning. Further work can also explore training CNN-Combo using early stopping and with different data fractions to fully quantify its performance and more robustly compare to LoRA fine-tuning. Ultimately, LoRA and other fine-tuning methods will be crucial as we ramp up to current and upcoming cosmological surveys. LoRA demonstrates potential in leveraging the wide array of existing pre-trained astrophysical models for data sparse tasks, especially when retraining models is too computationally expensive, and can adapt existing models to multiple new datasets, representing a developing avenue to do new precision cosmology.

Acknowledgments and Disclosure of Funding

Partial support for this work was provided by the Alfred P. Sloan Foundation and the NSF DGE2034835. This work used Jetstream2 at Indiana University through allocation # PHY230092 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by U.S. National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at http://dm.lsst.org

This paper is based [in part] on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by the Subaru Telescope and Astronomy Data Center (ADC) at National Astronomical Observatory of Japan. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), National Astronomical Observatory of Japan. The Subaru Telescope is honored and grateful for the opportunity of observing the Universe from Maunakea, which has the cultural, historical and natural significance in Hawaii.

This work is based on observations collected at the European Southern Observatory under ESO programme ID 179.A-2005 and on data products produced by CALET and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium.

References

[1] Hiroaki Aihara et al. "Second Data Release of the Hyper Suprime-Cam Subaru Strategic Program". In: *Publications of the Astronomical Society of Japan* 71.6 (Dec. 2019), p. 114. ISSN: 0004-6264. DOI: 10.1093/pasj/psz103. (Visited on 08/28/2025).

[2] Róbert Beck et al. "Photometric Redshifts for the SDSS Data Release 12". In: *Monthly Notices of the Royal Astronomical Society* 460.2 (Aug. 2016), pp. 1371–1381. ISSN: 0035-8711. DOI: 10.1093/mnras/stw1009. (Visited on 08/28/2025).

[3] Hardik Bishnoi and Tala Talaei Khoei. "Fast Galaxy Image Simulation Using Low-Rank Adapters". In: *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)*. Jan. 2025, pp. 00171–00177. DOI: 10.1109/CCWC62904.2025.10903751. (Visited on 08/27/2025).

[4] Christopher Bonnett. "Using Neural Networks to Estimate Redshift Distributions. An Application to CFHTLenS". In: *Monthly Notices of the Royal Astronomical Society* 449.1 (May 2015), pp. 1043–1056. ISSN: 0035-8711. DOI: 10.1093/mnras/stv230. (Visited on 08/22/2024).

[5] L. Cabayol et al. "The PAU Survey and *Euclid:* Improving Broadband Photometric Redshifts with Multi-Task Learning". In: *Astronomy & Astrophysics* 671 (Mar. 2023), A153. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202245027. (Visited on 04/13/2025).

[6] Matias Carrasco Kind and Robert J. Brunner. "TPZ: Photometric Redshift PDFs and Ancillary Information by Using Prediction Trees and Random Forests". In: *Monthly Notices of the Royal Astronomical Society* 432.2 (June 2013), pp. 1483–1501. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stt574. (Visited on 04/13/2025).

[7] LSST Science Collaboration et al. *LSST Science Book, Version 2.0*. Dec. 2009. DOI: 10.48550/arXiv.0912.0201. arXiv: 0912.0201 [astro-ph]. (Visited on 04/13/2025).

[8] Adrian A. Collister and Ofer Lahav. "ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks". In: *Publications of the Astronomical Society of the Pacific* 116 (Apr. 2004), pp. 345–351. ISSN: 0004-6280. DOI: 10.1086/383254. (Visited on 04/12/2025).

[9] Biprateep Dey et al. "Photometric Redshifts from SDSS Images with an Interpretable Deep Capsule Network". In: *Monthly Notices of the Royal Astronomical Society* 515.4 (Aug. 2022), pp. 5285–5305. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stac2105. arXiv: 2112.03939 [astro-ph]. (Visited on 04/12/2025).

[10] Tuan Do et al. *GalaxiesML: A Dataset of Galaxy Images, Photometry, Redshifts, and Structural Parameters for Machine Learning*. Sept. 2024. DOI: 10.48550/arXiv.2410.00271. arXiv: 2410.00271 [astro-ph]. (Visited on 08/28/2025).

[11] Kaiming He et al. *Deep Residual Learning for Image Recognition*. Dec. 2015. DOI: 10.48550/arXiv.1512.03385. arXiv: 1512.03385 [cs]. (Visited on 08/27/2025).

[12] Edward J. Hu et al. *LoRA: Low-Rank Adaptation of Large Language Models*. Oct. 2021. DOI: 10.48550/arXiv.2106.09685. arXiv: 2106.09685 [cs]. (Visited on 08/27/2025).

[13] Željko Ivezić et al. "LSST: From Science Drivers to Reference Design and Anticipated Data Products". In: *The Astrophysical Journal* 873.2 (Mar. 2019), p. 111. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ab042c. (Visited on 08/28/2025).

[14] E. Jones and J. Singal. "Analysis of a Custom Support Vector Machine for Photometric Redshift Estimation and the Inclusion of Galaxy Shape Information". In: *Astronomy & Astrophysics* 600 (Apr. 2017), A113. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201629558. (Visited on 04/13/2025).

[15] Evan Jones et al. "Improving Photometric Redshift Estimation for Cosmology with LSST Using Bayesian Neural Networks". In: *The Astrophysical Journal* 964.2 (Apr. 2024), p. 130. ISSN: 0004-637X, 1538-4357. DOI: 10.3847/1538-4357/ad2070. (Visited on 04/12/2025).

[16] Evan Jones et al. "Redshift Prediction with Images for Cosmology Using a Bayesian Convolutional Neural Network with Conformal Predictions". In: *The Astrophysical Journal* 974.2 (Oct. 2024), p. 159. ISSN: 0004-637X, 1538-4357. DOI: 10.3847/1538-4357/ad6d5a. (Visited on 04/12/2025).

[17] R. Laureijs et al. *Euclid Definition Study Report*. Oct. 2011. DOI: 10.48550/arXiv.1110.3193. (Visited on 08/28/2025).

[18] Andrew Lizarraga et al. *Understanding Galaxy Morphology Evolution Through Cosmic Time via Redshift Conditioned Diffusion Models*. Nov. 2024. DOI: 10.48550/arXiv.2411.18440. (Visited on 08/28/2025).

[19] Sourab Mangrulkar et al. *PEFT: State-of-the-art Parameter-Efficient Fine-Tuning methods*. https://github.com/huggingface/peft. 2022.

[20] Jeffrey A. Newman and Daniel Gruen. "Photometric Redshifts for Next-Generation Surveys". In: *Annual Review of Astronomy and Astrophysics* 60.Volume 60, 2022 (Sept. 2022), pp. 363–414. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev-astro-032122-014611. (Visited on 08/28/2025).

[21] Johanna Pasquet et al. "Photometric Redshifts from SDSS Images Using a Convolutional Neural Network". In: *Astronomy & Astrophysics* 621 (Jan. 2019), A26. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201833617. arXiv: 1806.06607 [astro-ph]. (Visited on 04/12/2025).

[22] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Dec. 2019. DOI: 10.48550/arXiv.1912.01703. arXiv: 1912.01703 [cs]. (Visited on 08/27/2025).

[23] Nesar Ramachandra et al. *Teaching LLMs to Speak Spectroscopy*. Aug. 2025. DOI: 10.48550/arXiv.2508.10075. arXiv: 2508.10075 [astro-ph]. (Visited on 08/27/2025).

[24] S. Riggi et al. *Radio-Llava: Advancing Vision-Language Models for Radio Astronomical Source Analysis*. Aug. 2025. DOI: 10.48550/arXiv.2503.23859. arXiv: 2503.23859 [astro-ph]. (Visited on 08/28/2025).

[25] Iftach Sadeh, Filipe B. Abdalla, and Ofer Lahav. "ANNz2 - Photometric Redshift and Probability Distribution Function Estimation Using Machine Learning". In: *Publications of the Astronomical Society of the Pacific* 128.968 (Oct. 2016), p. 104502. ISSN: 0004-6280, 1538-3873. DOI: 10.1088/1538-3873/128/968/104502. arXiv: 1507.00490 [astro-ph]. (Visited on 04/13/2025).

[26] Mara Salvato, Olivier Ilbert, and Ben Hoyle. "The Many Flavours of Photometric Redshifts". In: *Nature Astronomy* 3 (June 2019), pp. 212–222. ISSN: 2397-3366. DOI: 10.1038/s41550-018-0478-0. (Visited on 04/13/2025).

[27] S. Schuldt et al. “Photometric Redshift Estimation with a Convolutional Neural Network: NetZ”. In: *Astronomy & Astrophysics* 651 (July 2021), A55. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202039945. (Visited on 07/25/2025).

[28] J. Singal et al. “Machine Learning Classification to Identify Catastrophic Outlier Photometric Redshift Estimates”. In: *The Astrophysical Journal* 928.1 (Mar. 2022), p. 6. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ac53b5. (Visited on 04/13/2025).

[29] Jonathan Soriano et al. *Using Different Sources of Ground Truths and Transfer Learning to Improve the Generalization of Photometric Redshift Estimation*. Nov. 2024. DOI: 10.48550/arXiv.2411.18054. (Visited on 04/12/2025).

[30] Masayuki Tanaka et al. “Photometric Redshifts for Hyper Suprime-Cam Subaru Strategic Program Data Release 1”. In: *Publications of the Astronomical Society of Japan* 70.SP1 (Jan. 2018), S9. ISSN: 0004-6264, 2053-051X. DOI: 10.1093/pasj/psx077. (Visited on 04/13/2025).

[31] Lingling Xu et al. *Parameter-Efficient Fine-Tuning Methods for Pretrained Language Models: A Critical Review and Assessment*. Dec. 2023. DOI: 10.48550/arXiv.2312.12148. arXiv: 2312.12148 [cs]. (Visited on 08/29/2025).

[32] Jason Yosinski et al. *How Transferable Are Features in Deep Neural Networks?* Nov. 2014. DOI: 10.48550/arXiv.1411.1792. arXiv: 1411.1792 [cs]. (Visited on 08/29/2025).

A Fine-tuning in the 'reverse' direction

To fully explore the model fine-tuning methodology, we also explored training on GalaxiesML first, followed by fine-tuning with TransferZ-Images. The GalaxiesML baseline, hereafter CNN-Base-Rev, was trained in the same manner as CNN-Base. It stopped training after 147 epochs taking 1.3 hours. We then fine-tuned using the same LoRA configuration as CNN-LoRA, creating CNN-LoRA-Rev, which stopped training after 56 epochs and took 25.3 minutes. Finally, we created CNN-TL-Rev by using the same frozen layers as for CNN-TL, which stops training after 27 epochs taking 14.7 minutes.

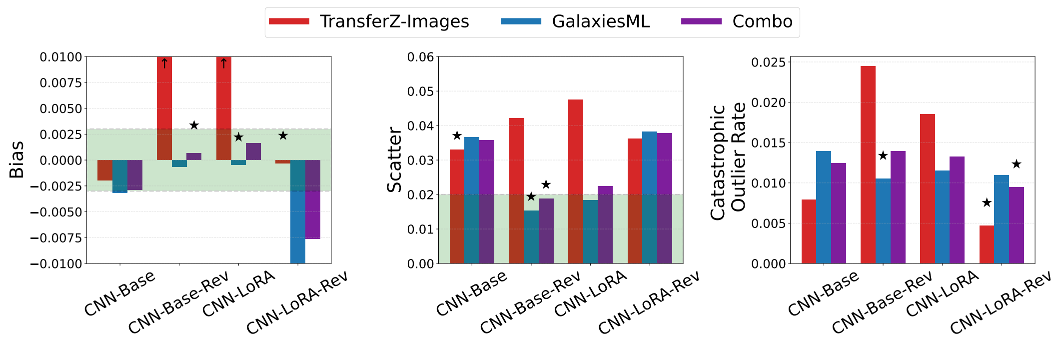

Figure 5: Bias (left), scatter (middle), and catastrophic outlier rate performance metrics (right) CNN-Base, CNN-Base-Rev, CNN-LoRA and CNN-LoRA-Rev. The metrics are evaluated on all three datasets (TransferZ-Images (red), GalaxiesML (blue), Combo (purple)). The green band for Bias and Scatter are the LSST Science Requirements. The black stars represent the best (lowest) metric for each dataset. Upward black arrows indicates the bar overflows the plot. CNN-LoRA-Rev performs worse than CNN-LoRA across most datasets. CNN-Base-Rev performs better than CNN-LoRA on scatter across all datasets, as well as bias on GalaxiesML and Combo.

The performance of four models, CNN-Base, CNN-Base-Rev, CNN-LoRA and CNN-LoRA-Rev are shown in Fig. 5 (see Table 3 for values, as well as CNN-TL and CNN-TL-Rev performance). CNN-TL and CNN-TL-Rev both perform poorly compared to all other models, though CNN-TL is slightly better than CNN-TL-Rev. The limitations of traditional transfer learning is thus further verified. CNN-LoRA-Rev performs worse than CNN-LoRA on TransferZ-Images and Combo. In particular on Combo, it has $\sim$4.7 $\times$ higher bias and $\sim$1.7 $\times$ higher scatter. CNN-Base-Rev performs better than CNN-LoRA on scatter across all datasets, as well as on bias on Combo. This indicates that spectroscopic redshift ground truth contains information that aids the generalizability of models to broader galaxy types than for which spectra are typically obtained. We can additionally speculate that the higher uncertainty in photometric redshift ground truths from TransferZ-Images hinders generalizability even though the types of galaxies in the dataset are more diverse. The photometric ground truth redshifts in TransferZ-Images are derived from 30+ imaging bands, yet these appear to be insufficient to provide precise enough redshifts to train accurate and generalizable machine learning models. However, the quality of datasets impacts model performance as well, and thus this claim may be verified by further study using the TransferZ-Images dataset.

Table 3: Model Metrics Summary Comparing Both Directions of Fine-tuning

		Bias	Scatter	Catastrophic Outlier Rate
CNN-Base	TransferZ-Images	-0.00198	0.0330	0.00789
	GalaxiesML	-0.00316	0.0366	0.0139
	Combo	-0.00287	0.0358	0.0125
CNN-Base-Rev	TransferZ-Images	0.0114	0.0421	0.0244
	GalaxiesML	-0.000647	0.0153	0.0105
	Combo	0.000672	0.0188	0.0139
CNN-LoRA	TransferZ-Images	0.0176	0.0476	0.0185
	GalaxiesML	-0.000481	0.0184	0.0115
	Combo	0.0016	0.0223	0.0132
CNN-LoRA-Rev	TransferZ-Images	-0.000322	0.0362	0.00466
	GalaxiesML	-0.0101	0.0381	0.0110
	Combo	-0.00759	0.0378	0.00944
CNN-TL	TransferZ-Images	0.0502	0.121	0.0203
	GalaxiesML	-0.00995	0.0391	0.0141
	Combo	-0.00396	0.0496	0.0157
CNN-TL-Rev	TransferZ-Images	-0.0638	0.101	0.0259
	GalaxiesML	-0.147	0.0844	0.0149
	Combo	-0.134	0.0973	0.0175
LSST Req.		$< \|0.03\|$	< 0.02	