Using different sources of ground truths and transfer learning to improve the generalization of photometric redshift estimation

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

Srinath Saikrishnan[1]
srinathsai22@g.ucla.edu

Vikram Seenivasan[1]
vikrams25@g.ucla.edu

Bernie Boscoe[2]
boscoeb@sou.edu

Jack Singal[3]
jsingal@richmond.edu

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

[1] Physics and Astronomy Department, UCLA, Los Angeles, CA 90024
[2] Computer Science Department, Southern Oregon University, Ashland, OR 97520
[3] Physics Department, University of Richmond, Richmond, VA 23173

Abstract

In this work, we explore methods to improve galaxy redshift predictions by combining different ground truths. Traditional machine learning models rely on training sets with known spectroscopic redshifts, which are precise but only represent a limited sample of galaxies. To make redshift models more generalizable to the broader galaxy population, we investigate transfer learning and directly combining ground truth redshifts derived from photometry and spectroscopy. We use the COSMOS2020 survey to create a dataset, TransferZ, which includes photometric redshift estimates derived from up to 35 imaging filters using template fitting. This dataset spans a wider range of galaxy types and colors compared to spectroscopic samples, though its redshift estimates are less accurate. We first train a base neural network on TransferZ and then refine it using transfer learning on a dataset of galaxies with more precise spectroscopic redshifts (GalaxiesML). In addition, we train a neural network on a combined dataset of TransferZ and GalaxiesML. Both methods reduce bias by $\sim$ 5x, RMS error by $\sim$ 1.5x, and catastrophic outlier rates by 1.3x on GalaxiesML, compared to a baseline trained only on TransferZ. However, we also find a reduction in performance for RMS and bias when evaluated on TransferZ data. Overall, our results demonstrate these approaches can meet cosmological requirements.

1 Introduction

Astronomers are increasingly adopting machine learning methods for redshift estimation, which is crucial for measuring the distances to galaxies in cosmology. Spectroscopic redshifts (spec-z's) are the most accurate, but they are time-consuming and thus impractical for large-scale surveys involving billions of galaxies [e.g., 21, 41, 8, 11]. Photometric redshifts (photo-z's) are derived from measurements of the brightness of galaxies (photometry) from images taken at different wavelengths. They are less precise but enable the analysis of much larger datasets [37]. Photometric redshift methods generally fall into two categories: template-fitting and data-driven approaches. In template-fitting [e.g., 2, 19, 7] a library of broad-band galaxy photometry and redshifts is compared to observed

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

photometry to estimate redshifts. Data-driven methods, often involving machine learning, train models on known redshift samples to predict redshifts for new data [e.g., 5, 12, 9, 37, 22].

A critical factor in the success of machine learning models is the quality and representativeness of the training data. For redshift prediction models, the most accurate training data comes from spectroscopic measurements, which precisely probe emission lines and achieve redshift uncertainties as low as 2×10^{-4} [e.g., 47]. However, these measurements are typically limited to bright galaxies with strong emission lines, representing only a small subset of the galaxies in the Universe. The COSMOS2020 survey [51] offers a broader dataset, covering a wider range of galaxy types. However, the median precision of its redshift measurements is approximately 0.03—about 100 times less precise than spectroscopic redshifts.

The limitation in representativeness highlights the need for methods that can generalize across different types of data. In this paper we explore two approaches of incorporating ground truths from real data for training photometric redshift models: transfer learning and mixing ground truths. Transfer learning [39, 52] offers a promising solution in this regard, allowing models trained on broader, less-precise datasets like COSMOS2020 to be fine tuned on precise but narrower spectroscopic datasets to improve their performance. Mixing ground truths approach is an alternative strategy that combines different sources of redshift measurements at the start of training allowing the model to simultaneously learn from complementary strengths of spectroscopic and photometric redshift datasets. Our approach is novel in its exploration of model generalization by incorporating different sources of ground truth from real data for training photometric redshift models. Understanding these approaches is particularly important as we look forward to large surveys, where astronomers will need to overcome the gaps in spectroscopic dataset coverage and volume in the initial years.

2 Data

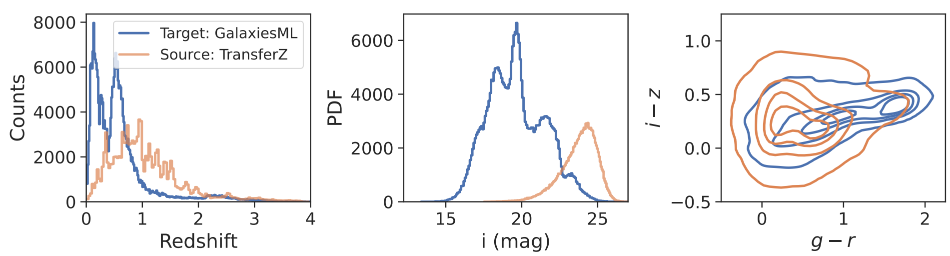

Figure 1: Two datasets: GalaxiesML [15] with spectroscopic redshift ground truth and TransferZ with COSMOS2020 survey [51] multi-band imaging redshift ground truth. The distribution of the dataset in redshift (left), i-band magnitude (center), and color (right) shows how the datasets complement each other to help the model generalize beyond the range of brightness and color sampled by the spectroscopic surveys.

Table 1: Data Summary

Dataset	Number of Sources	Redshift 90th percentile	Median Redshift Uncertainty	i-band mag 90th percentile	No. Filters
TransferZ	116,335	1.9	0.03	25	5
GalaxiesML	286,401	1.2	0.0002	22	5
Combo Data	402,408	1.5	0.0006	24	5

In this work we base our analyses on 5-band photometry to approximate the conditions for the Legacy Survey in Space and Time (LSST), a major upcoming survey [8, 21]. We created the TransferZ dataset by integrating data from two sources: the HSC PDR2 [1] wide field survey, which provides 5-band *grizy* photometry for a query of 3 million sources, and COSMOS2020 [51], which offers up to 35-band photometry for 1.7 million sources along with photometric redshifts. With 35 bands of photometry extending from the ultraviolet to the infrared, the photometric redshifts that can be estimated are much more accurate and precise than those resulting from five-band photometry. This is a reasonable basis for a ground truth redshift value [20, 45]. From COSMOS2020 we choose redshifts computed using LePhare template fitting [2, 19] with at least 30-band photometry from the CLASSIC subset [51]. To create TransferZ, we cross-match sources from COSMOS2020 with HSC PDR2 data, filtering for galaxies, and applying quality cuts to ensure reliable ground truth redshifts, resulting in a refined dataset of 116,335 galaxies with 5-band *grizy* photometry (*g*: 4754 Å, *r*: 6175 Å, *i*: 7711 Å, *z*: 8898 Å, *y*: 9762 Å) from HSC PDR2 and reliable redshifts from COSMOS2020. For more details, see Appendix A.

We use TransferZ as a broader and more general galaxy sample for redshift estimation to train the baseline model and then transfer learn using GalaxiesML [15], which has ground truth for redshifts from spectroscopy. The two datasets complement each other (Fig. 1). GalaxiesML has 286,401 galaxies, with 90% of the galaxies having $i < 22$ mag and most galaxies have redshifts < 1.2 (note that larger magnitude values mean the galaxies are fainter). This dataset is built on HSC PDR2 [1] and its associated spectroscopic database [29, 6, 32, 46, 34, 27, 17, 30, 14, 38, 10, 13]. TransferZ has 90% of its galaxies with $i < 25$ mag and redshifts < 1.9. TransferZ contains a higher number of galaxies in the cosmologically relevant range of $0.3 < z < 1.5$, potentially enabling representational analysis at higher redshifts than GalaxiesML alone. While TransferZ probes much fainter galaxies and more galaxy types, the redshift uncertainties from the 35-band photometry are typically 100 times larger than GalaxiesML with spectroscopy (Table 1). We note that there are 500 galaxies (about 0.1% of the total) in common between both TransferZ and GalaxiesML (Fig. 1). We assume the impact of this overlap is negligible for this experiment, but this can be verified in the future.

We also created a combination dataset called Combo that combines both TransferZ and GalaxiesML to test whether combining two types of ground truth is equivalent to transfer learning from one dataset to another. When there are both spectroscopic and COSMOS2020 photometric redshifts for the same galaxy, we choose to include only the spectroscopic redshift, because it is more accurate (about 500 galaxies are affected). The combo dataset consists of 402,408 galaxies. The datasets are split into 80% training, 10% validation, and 10% testing sets. In the following sections, we refer to TransferZ as the source data, GalaxiesML as the target data, and Combo as combo data. TransferZ is made available on Zenodo with a DOI: 10.5281/zenodo.14218996.

3 Methodology & metrics

We employed a neural network (NN) architecture based on [24, 22] for photometric redshift estimation, consisting of four fully connected layers with ReLU activation and a skip connection. Hyperparameter tuning was performed with the source training and validation data using HyperBand [28]. The Hyperband search space for training the NN includes 1 to 10 hidden layers with 32 to 2048 neurons per layer, whether to include a skip connection, and whether to add additional dense layers. If additional dense layers are included, they range from 1 to 10 hidden layers with 32 to 4096 neurons per layer. The final model has the four initial hidden layers with 200 neurons each followed by a skip connection and two additional hidden layers with 2000 neurons each. All hidden layers use the rectified linear unit (ReLU) activation function.

The base model (NN-Base) was trained on the TransferZ dataset using the Adam optimizer with a learning rate of 5×10^{-4}, a batch size of 512, and for 500 epochs. Transfer learning (NN-TL) was then applied by fine-tuning the NN-Base model on the GalaxiesML dataset, freezing all layers except the input and the first and fifth dense layers, with a reduced learning rate of 5×10^{-10} and trained for 1000 epochs.

The model trained on the combined data (NN-Combo) is hyperparameter optimized similarly to the NN-Base model. The final model has 6 hidden layers and a skip connection between inputs and the hidden layers. NN-Combo was trained on the Combo dataset using the Adam optimizer with a learning rate of 5×10^{-4}, a batch size of 512, and for 2000 epochs. The three model trainings achieve optimal performance before the reported epochs since learning curves plateau earlier.

A custom loss function, $L(\Delta z) = 1 - \frac{1}{1+(\Delta z/0.15)^2}$, is used [47] for training. The photometry data is normalized separately for each training stage. Performance was evaluated using bias, root mean square error (RMS), and the catastrophic outlier rate on the test sets within the redshift range of $0.3 < z < 1.5$. The bias metric is the median of the bias distribution defined as $b = (z_{photo} - z_{truth})/(1 + z_{truth})$ where z_{photo} and z_{truth} is the estimated photometric redshift and the ground truth redshift, respectively. The RMS is defined as the interquartile range of the bias distribution divided by 1.349 weighted by the median redshift in the bin $(1 + \overline{z_{truth}})$ [18]. This definition of RMS is less sensitive to outlier rates than standard definitions. The catastrophic outlier rate is defined as the fraction of objects where the absolute difference between photometric and true redshifts exceeds 1.0, expressed as $|z_{photo} - z_{truth}| > 1.0$. These are among the most important metrics for cosmology [4, 21]. The metrics are evaluated for NN-Base, NN-TL, and NN-Combo test datasets. We compare our metrics to those in [23], where they use a NN trained on GalaxiesML. The comparison highlights the benefits of mixing ground truths against a spectroscopic ground truth.

4 Results & Discussion

In this work, we test different ways of combining different sources of ground truth for photometric redshifts - either through transfer learning or by combining the training datasets. We find that both methods are better than the base model, which is only trained on the TransferZ dataset. This suggests that it is possible to improve photometric redshift estimates by combing multiple sources of ground truth. The choice between the two methods should be based on which metric best serves the scientific objectives. Below we summarize the benefits and limitations of our approaches: transfer learning of the NN model on the source (TransferZ) and target (GalaxiesML) datasets, and a model trained on the combination of GalaxiesML and TransferZ (Combo) dataset.

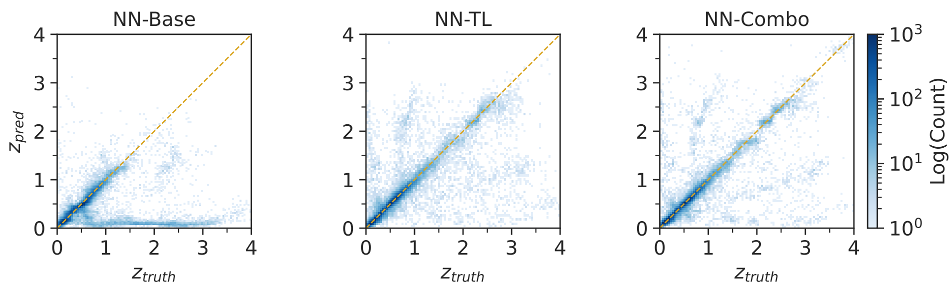

Figure 2: Comparison of redshift predictions from the three neural network models in this work (NN-Base, NN-TL, and NN-Combo) against true redshift values. We show results for the GalaxiesML test dataset using spectroscopic redshift as ground truth. Results shown are from one randomly selected run out of 100 total iterations.

Table 2: Model Metrics Summary

		Bias ($\times 10^{-3}$)	RMS ($\times 10^{-3}$)	Cat. Outlier Rate ($\times 10^{-2}$)
NN-Base	TransferZ	-0.69 $\pm$ 2.37	22.6 $\pm$ 0.35	1.29 $\pm$ 0.25
	GalaxiesML	-10.9 $\pm$ 5.71	22.4 $\pm$ 1.23	2.49 $\pm$ 0.15
NN-TL	TransferZ	7.45 $\pm$ 3.12	28.5 $\pm$ 0.86	0.40 $\pm$ 0.13
	GalaxiesML	-1.51 $\pm$ 1.66	15.4 $\pm$ 0.27	1.68 $\pm$ 0.14
NN-Combo	TransferZ	1.14 $\pm$ 1.74	23.0 $\pm$ 0.33	1.33 $\pm$ 0.29
	GalaxiesML	-1.92 $\pm$ 1.68	15.0 $\pm$ 0.23	1.89 $\pm$ 0.17

Model predictions are evaluated against test datasets comprising of 40,914 galaxies from GalaxiesML and 11,633 from TransferZ. Both NN-TL and NN-Combo perform comparably within the redshift range of $0.3 \leq z \leq 1.5$, exhibiting similar prediction patterns and improving upon the base model (Fig. 2). Both NN-TL and NN-Combo have small scatter in their predictions at these redshift ranges compared to the base model. Quantitatively, we evaluate the model's predictions using three metrics – bias, RMS, and catastrophic outlier rate, these are reported in Table 2. We compare the fractional change of the metrics on both target and source data against NN-Base metrics. Evaluating NN-TL (NN-Combo) on the target data shows 7.19 (5.65) times lower bias, 1.45 (1.49) times lower RMS, and 1.48 (1.32) times lower catastrophic outlier rate. The largest magnitude change can be seen for bias, which is important as confidence in a redshift bin assignment is crucial for cosmological measurements requiring precise redshift distributions. The catastrophic outlier rate remains below

10% across all models, showing that neural network can effectively control a major systematic contaminant in photometric redshift measurements.

For the source data metrics, we find NN-TL reduces the catastrophic outlier rate compared to NN-Base, but performs worse on bias and RMS. The metrics of NN-Combo are comparable to those of NN-Base on the source data. Evaluating NN-TL (NN-Combo) on the source data shows 3.24 times lower (1.23 times higher) catastrophic outlier rate. However, it shows a bias 10.7 (1.64) times higher and RMS 1.26 (1.02) times higher. The increase in bias on the source data after transfer learning on a new ground truth suggests some loss of the originally learned features. The model trained on combined ground truths is similarly affected but to a lesser extent, meaning it better leverages both sources of truth.

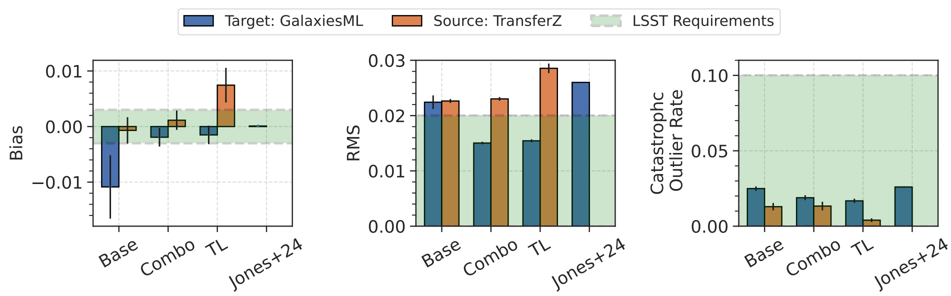

Figure 3: From left to right, comparison of the bias, outlier and RMS metrics between the baseline NN, transfer-learnt NN, combo NN, and [23]. The metrics are evaluated on the target (blue) and source (orange) data within the range of $0.3 \leq z \leq 1.5$. The error bars are generated from 100 random initializations of the model training. We report [23] scatter value for RMS. While [23] use a different RMS definition, our RMS calculation is equivalent to their reported scatter value.

The results also show both transfer learning approaches and models trained on combined ground truth data improve photometric redshift predictions compared to single-dataset training. Our transfer learning model performs better than the one from [23] (henceforth J24), which uses only the target dataset (Fig. 3). In comparing our two approaches for handling mixed ground truth data, the advantage of the models depends on the metric. The Combo model performs better than the transfer learn model on bias (6 times lower) and RMS (1.2 times lower) on target data, and marginally on RMS (1.03 times lower) for source. However, it performs worse on bias (1.27 times higher) on the target dataset. Additionally, the catastrophic outlier rate on target data is 1.12 times higher and on source data 3.24 times higher. Depending on the science needs, the best approach needs to align with the science goal.

There are several limitations to both NN-TL and NN-Combo. The model from J24 achieved a bias of one order magnitude lower than our approach in NN-TL and NN-Combo evaluated on the GalaxiesML. In addition, there are some notable features in the figure showing the prediction versus true redshift, such as the clustered predictions at $0.5 \leq z_{truth} \leq 1$ and $1.5 \leq z_{pred} \leq 3$ in the target dataset (Fig. 2). This suggests there is a source of systematic error (to be analyzed in a forthcoming paper). NN-TL also shows limited accuracy at $z_{truth} > 3$ from sparse training data in these redshift ranges at each training step, while NN-Combo achieves accurate predictions through its training approach of combining both datasets in one training step.

Previous work explored the use of transfer learning in the context of improving redshift estimates using a combination of simulated and real data [16, 36], but to our knowledge, this is the first work to apply transfer learning from only real data to generalize photometric redshift predictions. Both of our models (NN-TL and NN-Combo) are capable of incorporating two sources of ground truth, but there are a number of ways that the models can be improved upon. For example, our models do not produce uncertainties. Extending transfer learning to probabilistic ML models will be important for more scientific applications [e.g, 3, 40, 50, 22]. Additional tests of how well this model generalizes will be important to validate the improvements. One approach could be to use this model to detect galaxy clusters in existing data. Galaxies within the same cluster will be at the same redshift, but will consist of many galaxy types, which allows us to independently quantify the generalizability of the models. Ultimately, this could help solve one of the most challenging problems for surveys like LSST: providing reliable photometric redshift measurements for testing models of cosmology.

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, 2053-051X. DOI: 10.1093/pasj/psz103. arXiv: 1905.12221 [astro-ph].

[2] S. Arnouts et al. "Measuring and Modelling the Redshift Evolution of Clustering: The Hubble Deep Field North". In: *Monthly Notices of the Royal Astronomical Society* 310.2 (Dec. 1999), pp. 540–556. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.1999.02978.x.

[3] Narciso Benítez. "Bayesian Photometric Redshift Estimation". In: *The Astrophysical Journal* 536 (June 2000), pp. 571–583. ISSN: 0004-637X. DOI: 10.1086/308947.

[4] Chris Blake and Sarah Bridle. "Cosmology with Photometric Redshift Surveys". In: *Monthly Notices of the Royal Astronomical Society* 363.4 (Nov. 2005), pp. 1329–1348. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09526.x.

[5] 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.

[6] E. J. Bradshaw et al. "High-Velocity Outflows from Young Star-Forming Galaxies in the UKIDSS Ultra-Deep Survey". In: *Monthly Notices of the Royal Astronomical Society* 433 (July 2013), pp. 194–208. ISSN: 0035-8711. DOI: 10.1093/mnras/stt715.

[7] Gabriel B. Brammer, Pieter G. Van Dokkum, and Paolo Coppi. "EAZY: A Fast, Public Photometric Redshift Code". en. In: *The Astrophysical Journal* 686.2 (Oct. 2008), pp. 1503–1513. ISSN: 0004-637X, 1538-4357. DOI: 10.1086/591786.

[8] Katelyn Breivik et al. *From Data to Software to Science with the Rubin Observatory LSST*. Aug. 2022.

[9] D. Carrasco et al. "Photometric Classification of Quasars from RCS-2 Using Random Forest". en. In: *Astronomy & Astrophysics* 584 (Dec. 2015), A44. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201525752.

[10] Alison L. Coil et al. "The PRIsm MUlti-Object Survey (PRIMUS) I: Survey Overview and Characteristics". In: *The Astrophysical Journal* 741.1 (Nov. 2011), p. 8. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/741/1/8. arXiv: 1011.4307 [astro-ph].

[11] Euclid Collaboration et al. *Euclid. I. Overview of the Euclid Mission*. May 2024. DOI: 10.48550/arXiv.2405.13491. arXiv: 2405.13491 [astro-ph].

[12] Adrian A. Collister and Ofer Lahav. "ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks". en. In: *Publications of the Astronomical Society of the Pacific* 116.818 (Mar. 2004), p. 345. ISSN: 1538-3873. DOI: 10.1086/383254.

[13] Richard J. Cool et al. "The PRIsm MUlti-object Survey (PRIMUS). II. Data Reduction and Redshift Fitting". In: *The Astrophysical Journal* 767.2 (Apr. 2013), p. 118. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/767/2/118. arXiv: 1303.2672 [astro-ph].

[14] Marc Davis et al. "Science Objectives and Early Results of the DEEP2 Redshift Survey". In: 4834 (Feb. 2003), pp. 161–172. DOI: 10.1117/12.457897.

[15] Tuan Do et al. *GalaxiesML: An Imaging and Photometric Dataset of Galaxies for Machine Learning*. May 2024. DOI: 10.5281/zenodo.11117528.

[16] M Eriksen et al. "The PAU Survey: Photometric Redshifts Using Transfer Learning from Simulations". In: *Monthly Notices of the Royal Astronomical Society* 497.4 (Oct. 2020), pp. 4565–4579. ISSN: 0035-8711. DOI: 10.1093/mnras/staa2265.

[17] B. Garilli et al. "The VIMOS Public Extragalactic Survey (VIPERS): First Data Release of 57 204 Spectroscopic Measurements". en. In: *Astronomy & Astrophysics* 562 (Feb. 2014), A23. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201322790.

[18] Melissa L. Graham et al. "Photometric Redshifts with the LSST. II. The Impact of Near-infrared and Near-ultraviolet Photometry". en. In: *The Astronomical Journal* 159.6 (May 2020), p. 258. ISSN: 1538-3881. DOI: 10.3847/1538-3881/ab8a43.

[19] O. Ilbert et al. "Accurate Photometric Redshifts for the CFHT Legacy Survey Calibrated Using the VIMOS VLT Deep Survey". en. In: *Astronomy & Astrophysics* 457.3 (Oct. 2006), pp. 841–856. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361:20065138.

[20] O. Ilbert et al. "COSMOS PHOTOMETRIC REDSHIFTS WITH 30-BANDS FOR 2-Deg2". en. In: *The Astrophysical Journal* 690.2 (Dec. 2008), p. 1236. ISSN: 0004-637X. DOI: 10.1088/0004-637X/690/2/1236.

[21] Željko Ivezić et al. "LSST: From Science Drivers to Reference Design and Anticipated Data Products". en. In: (May 2008). DOI: 10.3847/1538-4357/ab042c.

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

[23] Evan Jones et al. "Improving Photometric Redshift Estimation for Cosmology with LSST Using Bayesian Neural Networks". In: *The Astrophysical Journal* 964 (Apr. 2024), p. 130. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ad2070.

[24] Evan Jones et al. *Photometric Redshifts for Cosmology: Improving Accuracy and Uncertainty Estimates Using Bayesian Neural Networks*. Feb. 2022. DOI: 10.48550/arXiv.2202.07121. arXiv: 2202.07121 [astro-ph].

[25] C. Laigle et al. "THE COSMOS2015 CATALOG: EXPLORING THE 1 < z < 6 UNIVERSE WITH HALF A MILLION GALAXIES". en. In: *The Astrophysical Journal Supplement Series* 224.2 (June 2016), p. 24. ISSN: 0067-0049. DOI: 10.3847/0067-0049/224/2/24.

[26] Dustin Lang, David W. Hogg, and David Mykytyn. *The Tractor: Probabilistic astronomical source detection and measurement*. Astrophysics Source Code Library, record ascl:1604.008. Apr. 2016.

[27] O. Le Fèvre et al. "The VIMOS VLT Deep Survey Final Data Release: A Spectroscopic Sample of 35 016 Galaxies and AGN out to z ~ 6.7 Selected with $17.5 \leq i_{AB} \leq 24.75$". en. In: *Astronomy and Astrophysics* 559 (Nov. 2013), A14. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201322179.

[28] Lisha Li et al. "Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization". In: *Journal of Machine Learning Research* 18.185 (2018), pp. 1–52. ISSN: 1533-7928.

[29] Simon J. Lilly et al. "The zCOSMOS 10k-Bright Spectroscopic Sample". In: *The Astrophysical Journal Supplement Series* 184 (Oct. 2009), pp. 218–229. ISSN: 0067-0049. DOI: 10.1088/0067-0049/184/2/218.

[30] J. Liske et al. "Galaxy And Mass Assembly (GAMA): End of Survey Report and Data Release 2". In: *Monthly Notices of the Royal Astronomical Society* 452 (Sept. 2015), pp. 2087–2126. ISSN: 0035-8711. DOI: 10.1093/mnras/stv1436.

[31] H. J. McCracken et al. "UltraVISTA: A New Ultra-Deep near-Infrared Survey in COSMOS". en. In: *Astronomy & Astrophysics* 544 (Aug. 2012), A156. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201219507.

[32] R. J. McLure et al. "The Sizes, Masses and Specific Star-Formation Rates of Massive Galaxies at 1.3". en. In: (May 2012). DOI: 10.1093/mnras/sts092.

[33] Bo Milvang-Jensen et al. "On-Sky Characterisation of the VISTA NB118 Narrow-Band Filters at 1.19 μm". en. In: *Astronomy & Astrophysics* 560 (Dec. 2013), A94. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201321814.

[34] Ivelina G. Momcheva et al. "The 3D-HST Survey: Hubble Space Telescope WFC3/G141 Grism Spectra, Redshifts, and Emission Line Measurements for $\sim 100,000$ Galaxies". In: *The Astrophysical Journal Supplement Series* 225.2 (Aug. 2016), p. 27. ISSN: 1538-4365. DOI: 10.3847/0067-0049/225/2/27. arXiv: 1510.02106 [astro-ph].

[35] A. Moneti et al. "Euclid Preparation - XVII. Cosmic Dawn Survey: Spitzer Space Telescope Observations of the Euclid Deep Fields and Calibration Fields". en. In: *Astronomy & Astrophysics* 658 (Feb. 2022), A126. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202142361.

[36] Irene Moskowitz et al. "Improving Photometric Redshift Estimates with Training Sample Augmentation". In: *The Astrophysical Journal* 967 (May 2024), p. L6. ISSN: 0004-637X. DOI: 10.3847/2041-8213/ad4039.

[37] Jeffrey A. Newman and Daniel Gruen. "Photometric Redshifts for Next-Generation Surveys". In: *Annual Review of Astronomy and Astrophysics* 60.1 (2022), pp. 363–414. DOI: 10.1146/annurev-astro-032122-014611.

[38] Jeffrey A. Newman et al. "The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts". In: *The Astrophysical Journal Supplement Series* 208.1 (Aug. 2013), p. 5. ISSN: 0067-0049, 1538-4365. DOI: 10.1088/0067-0049/208/1/5. arXiv: 1203.3192 [astro-ph].

[39] Sinno Jialin Pan and Qiang Yang. “A Survey on Transfer Learning”. In: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (Oct. 2010), pp. 1345–1359. ISSN: 1558-2191. DOI: 10.1109/TKDE.2009.191.

[40] 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.

[41] Giuseppe D. Racca et al. “The Euclid Mission Design”. In: July 2016, 99040O. DOI: 10.1117/12.2230762. arXiv: 1610.05508 [astro-ph].

[42] D. B. Sanders et al. “S-COSMOS: The Spitzer Legacy Survey of the Hubble Space Telescope ACS 2 Deg2 COSMOS Field I: Survey Strategy and First Analysis*”. en. In: *The Astrophysical Journal Supplement Series* 172.1 (Sept. 2007), p. 86. ISSN: 0067-0049. DOI: 10.1086/517885.

[43] Marcin Sawicki et al. “The CFHT Large Area U-band Deep Survey (CLAUDS)”. In: *Monthly Notices of the Royal Astronomical Society* 489.4 (Nov. 2019), pp. 5202–5217. ISSN: 0035-8711. DOI: 10.1093/mnras/stz2522.

[44] N. Scoville et al. “The Cosmic Evolution Survey (COSMOS): Overview”. In: *The Astrophysical Journal Supplement Series* 172 (Sept. 2007), pp. 1–8. ISSN: 0067-0049. DOI: 10.1086/516585.

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

[46] Rosalind E. Skelton et al. “3D-HST WFC3-selected Photometric Catalogs in the Five CANDELS/3D-HST Fields: Photometry, Photometric Redshifts, and Stellar Masses”. In: *The Astrophysical Journal Supplement Series* 214 (Oct. 2014), p. 24. ISSN: 0067-0049. DOI: 10.1088/0067-0049/214/2/24.

[47] 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. DOI: 10.1093/pasj/psx077.

[48] Y. Taniguchi et al. “The Cosmic Evolution Survey (COSMOS): Subaru Observations of the HST Cosmos Field*”. en. In: *The Astrophysical Journal Supplement Series* 172.1 (Sept. 2007), p. 9. ISSN: 0067-0049. DOI: 10.1086/516596.

[49] Yoshiaki Taniguchi et al. “The Subaru COSMOS 20: Subaru Optical Imaging of the HST COSMOS Field with 20 Filters*”. In: *Publications of the Astronomical Society of Japan* 67.6 (Dec. 2015), p. 104. ISSN: 0004-6264. DOI: 10.1093/pasj/psv106.

[50] M. Treyer et al. “CNN Photometric Redshifts in the SDSS at r $\leq$ 20”. In: *Monthly Notices of the Royal Astronomical Society* (Oct. 2023). ISSN: 0035-8711. DOI: 10.1093/mnras/stad3171.

[51] J. R. Weaver et al. “COSMOS2020: A Panchromatic View of the Universe to z 10 from Two Complementary Catalogs”. In: *The Astrophysical Journal Supplement Series* 258 (Jan. 2022), p. 11. ISSN: 0067-0049. DOI: 10.3847/1538-4365/ac3078.

[52] Karl Weiss, Taghi M. Khoshgoftaar, and DingDing Wang. “A Survey of Transfer Learning”. In: *Journal of Big Data* 3.1 (May 2016), p. 9. ISSN: 2196-1115. DOI: 10.1186/s40537-016-0043-6.

[53] M. A. Zamojski et al. “Deep GALEX Imaging of the COSMOS HST Field: A First Look at the Morphology of z ~ 0.7 Star-forming Galaxies”. en. In: *The Astrophysical Journal Supplement Series* 172.1 (Sept. 2007), p. 468. ISSN: 0067-0049. DOI: 10.1086/516593.

A Data Creation

A.1 COSMOS2020 Photometric Redshifts

In this work, we make use of photometric redshifts from the latest release of the Cosmic Evolution Survey (COSMOS; [44]) catalog (COSMOS2020; [51]) consisting of over 1 million sources. The COSMOS field covers about 1.7 deg^2 of the sky and has been observed across the electomagnetic spectrum by the Galaxy Evolution Explorer (GALEX; [53]) in the far-UV to near-UV, the Canada-France-Hawaii Telescope Large Area U-band Deep Survey (CLAUDS; [43]), the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; [1]) and Suprime-Cam (SC) data ([48, 49]) on the Subaru telescope in the optical, the UltraVISTA survey (UVISTA; [31, 33]) in the near-infrared, and the Cosmic Dawn Survey [42, 35] using Spitzer in the mid-infrared. The surveys cover the area in the X-ray, optical, and infrared, enhancing our studies in galaxy evolution and nature of dark matter.

The COSMOS2020 catalog is composed of two separate catalogs labeled `CLASSIC` and `THE FARMER`. `CLASSIC` uses aperture photometry methods while `THE FARMER` uses profile fitting methods (The Tractor;[26]) for the photometry measurements of extended sources. In addition, the catalogs provide photometric redshifts from two independent template fitting codes `LePhare` [2] and `EAZY` [7] along with additional outputs.

In this work, we make use of sources that are common to both catalogs. The `CLASSIC` catalog contains $1,720,700$ sources with photometric redshifts computed using up to 35-bands while `THE FARMER` consists of $964,506$ sources with photometric redshifts computed using up to 30-bands. The discrepancy arises from the variable Point Spread Function (PSF) of the Suprime-Cam medium bands, which `THE FARMER` could not overcome. After cross-matching using internal IDs, there is $923,403$ common sources among the two catalogs.

Our decisions in Section A.3 depend on the reliability of the photometric redshifts. Among the different combinations of photometry (`CLASSIC`/`THE FARMER`) and photometric redshift codes (`LePhare`/`EAZY`), the numbers of bands fit by the photo-z codes is not consistent. `CLASSIC`/`LePhare` uses up to 35 bands. `CLASSIC`/`EAZY` uses up to 30 bands, excluding GALEX data and Suprime-Cam broad bands due to their limited depth and PSF issues (Suprime-Cam and UltraVISTA narrow bands are also excluded, though no explicit reason is provided). `THE FARMER`/`CLASSIC` uses up to 30 bands, excluding the Suprime-Cam broad bands. `THE FARMER`/`EAZY` uses up to 27 bands, excluding the Suprime-Cam broad bands, GALEX FUV/NUV bands, and all narrow bands. While photometry cannot reach the precision of spectroscopy, more photometric bands reduces degeneracies in redshift measurements. Since `CLASSIC`/`LePhare`'s photometric redshifts are estimated using the largest number of bands, this set is chosen as the labels used in our neural network training involving TransferZ. Specifically, we use `lp_zPDF` for a galaxy corresponding to the median of the photometric redshift likelihood distribution.

A.2 HSC Photometry

For our analysis we compile five-band (*grizy*) photometry to approximate data produced by large scale surveys in comparable depth [21, 8, 11, 41]. We use data from the HSC Subaru Strategic Program's (HSC-SSP) second data release (HSC PDR2; [1]) in the wide field that reaches similar depths as LSST but over a smaller area coverage. The HSC wide-field camera is mounted on the Subaru Telescope with a FOV of 1.8 deg^2. The survey has observed over 900 deg^2 of the sky across five optical filters (*grizy*) with a median seeing in the i-band of 0.58" and a median 5σ depth of 26.2. The HSC-SSP survey does not have a bluer band unlike LSST which is expected to have six optical filters *ugrizy*.

Our analysis draws from a query of over 3 million sources around the COSMOS field from the HSC-PDR2 Wide catalogs. Our photometry data is queried from HSC-SSP's data release site. No initial constraints are placed on this query of our data. In addition, we query for a sample of galaxies with spectroscopic redshifts for validation. In Section A.3, we use these spectroscopic redshifts to guide the quality of our cuts.

A.3 TransferZ

We created the TransferZ dataset, a dataset consisting of galaxies with *grizy* photometry and reliable photometric redshifts following the workflow shown in Figure 4. The creation of TransferZ is split into two main steps: combining galaxy HSC data with COSMOS2020 data and performing quality cuts to ensure reliable features and labels. In this work, HSC *grizy* photometry serve as our features and `LePhare` photometric redshift `lp_zPDF` from the `CLASSIC` catalog of COSMOS2020 serve as our labels. See Section A.1 for more details about these criteria.

A.3.1 Cross-match Datasets and Galaxy Filter

The combination of HSC and COSMOS2020 data involved two steps. (1) A positional cross-match of HSC PDR2 sources with the COSMOS2020 catalog subset from Section A.1 within 1". (2) We filter

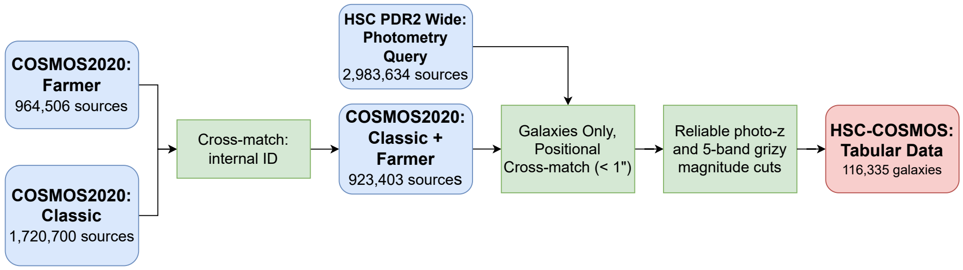

Figure 4: Flow chart showing the steps used in creating the TransferZ dataset. Green rectangles represent processes and blue rectangles represent inputs. The red rectangle is the dataset released with this paper.

for galaxies using `LePhare` classification method `lp_type = 0` when available, otherwise defaulting to HSC source classification `i_extendedness_value = 1`. We are aware that [51] advises against using `LePhare` classifcation; however, we find a strong agreement between HSC and COSMOS2020 classification methods, with 88% accuracy when we consider COSMOS2020 classification as the ground truth. The step reduces our dataset to 670,053 galaxies.

A.3.2 Quality Cuts

The quality cuts are similar to [45] where the cuts are split in two categories: COSMOS2020 photometry cuts and direct cuts to the photometric redshift quality. We note that the analysis in their work makes use of COSMOS2015 [25], a previous iteration of the COSMOS catalog which consists of aperture based photometry and photometric redshifts computed by `LePhare` photo-z code.

First, we implement cuts on COSMOS2020 photometry quality used in their redshift estimation. We require quality measurements in the bands used for their redshift estimation, as poor quality or missing photometry degrades estimated redshifts. We limit the requirement of photometry between 0 and 50 mag for 23 bands from HSC (*grizy*), Suprime-Cam (medium band filters), UVISTA (*YJHKs*), and *Spitzer*/IRAC (channel 1 and channel 2). These filters are consistent across estimated photometric redshifts from the four configurations (aperture based and profile-fitting based photometry processed by `LePhare` and `EAZY`). The u and u* band filters are ignored as galaxies at redshift $z \approx 3.3$ and beyond would not be detected in these bands due to absorption by neutral hydrogen (this absorption creates what is known as the Lyman break at a rest-frame wavelength of 912 Å).

Second, we apply cuts to ensure photometric redshift quality and agreement between all redshift configurations in COSMOS2020. We first exclude galaxies lacking redshift measurement across all configuration and require robust χ^2 fits in both photo-z codes. Following [45], we applied a threshold of $\chi^2 < 1$ for fits from `CLASSIC/LePhare`. For the other configuration, we define the threshold to remove the same fraction of galaxies as the `CLASSIC/LePhare` threshold. Additionally, for `LePhare` redshift estimates, we require close agreement ($|$`lp_zPDF - lp_zMinChi2`$| < 0.1$) between the redshift at the peak of the likelihood distribution and the redshift that minimizes χ^2. We also require `CLASSIC/LePhare` photometric redshifts to agree with the other configurations, as `CLASSIC/LePhare` photometric redshifts serve as the ground truth labels in our training. We restrict our sample to $z < 4$.

Finally, we implement cuts on our training features which we use the HSC-PDR2 *grizy* photometry `cModel` magnitudes. We require these values to be between 0 and 50 mag. After all cuts, TransferZ contains 116,335 galaxies.