Galaxy Morphology Classifier and Photometric Redshift Estimation using Deep Learning

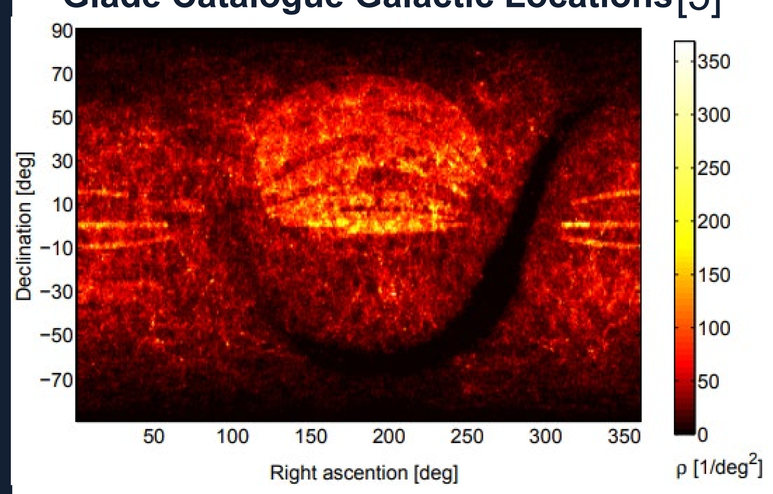
Andrew Engel and Gautham Narayan
Department of Astronomy, College of Liberal Arts and Sciences, University of Illinois at Urbana-Champaign

Abstract

We are developing a new model for morphological classification and photometric redshift estimation of galaxies. Predicting morphology and redshift photometrically is critical to rapid follow-up of gravitational wave alerts and to time-domain studies without the benefit of detailed spectroscopic information. Our model makes use of advances in deep learning to sort galactic survey images into spiral and elliptical classes with the Galaxy Zoo 1 catalogue providing training targets. We present methods whereby adding more layers, mixing, and providing an additional channel of photometric magnitudes increased classification accuracy for greater redshifted spiral galaxies. Utilizing full image inputs, photometric magnitudes, and angular size, our model can classify Galaxy Zoo 1 high debiased confidence galaxies ($Z_{avg} \approx 0.05$) to 0.97 accuracy, and low debiased confidence galaxies ($Z_{avg} \approx 0.15$) to 0.80 accuracy. We will use our model to build a photometric redshift catalog for the northern sky using Pan-STARRS1, which we will incorporate into the ANTARES alert broker that is currently processing the ZTF alert stream.

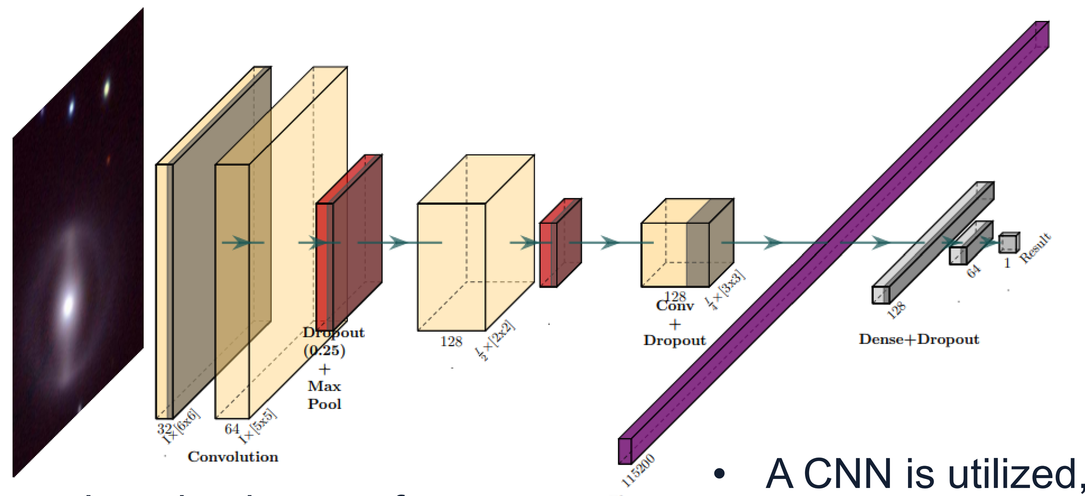
PROBLEM

Glade Catalogue Galactic Locations[5]



- Multi-messenger and transient-event astrophysics rely on accurate catalogues of galaxy depths to identify hosts
- Spectroscopy is expensive and therefore galaxy catalogues are incomplete, (see left, [5])
- Therefore, photometric redshifts are used to match the time scale and field of view of modern surveys
- However, photometric redshifts are biased by morphological effects like edge-on-disk reddening
- We start with a simpler problem of galaxy classification

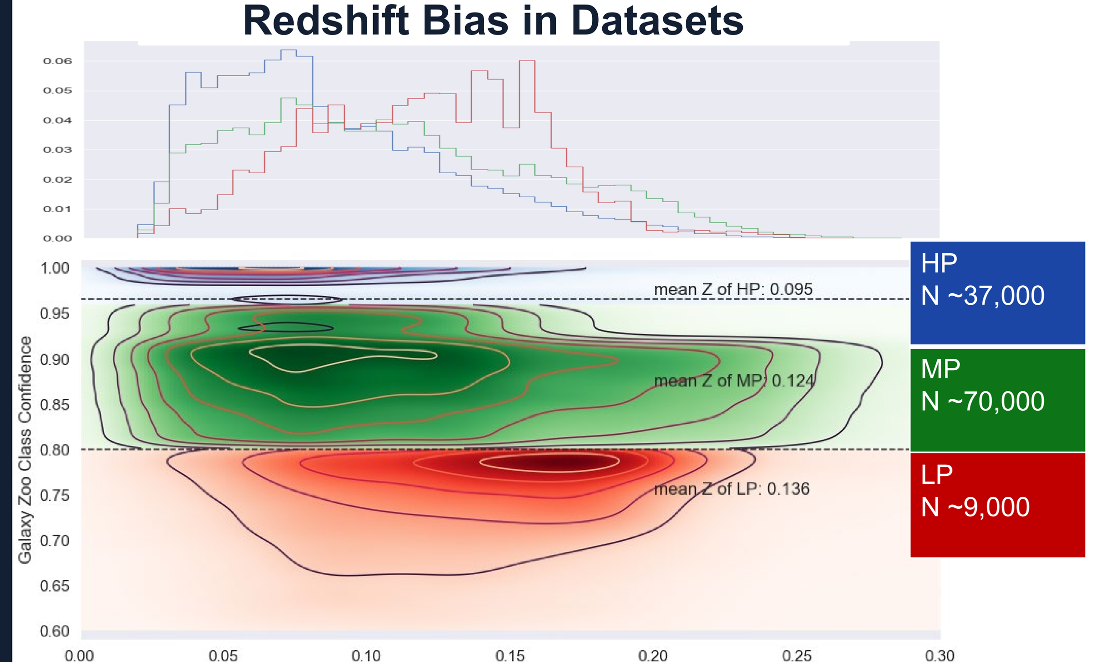
OUR APPROACH



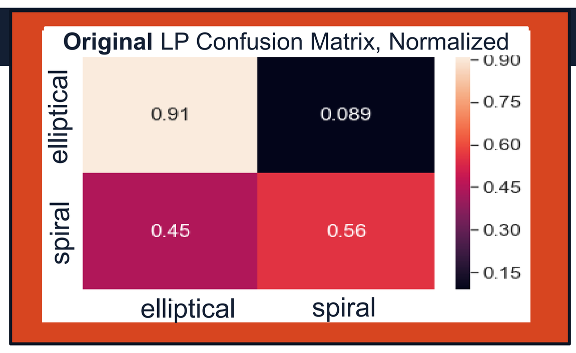
- Deep learning is one of many empirical methods for estimating redshift
- NN are scalable, versatile, and benefit from advancements in GPUs
- Convolutional Neural Networks (CNN) are known for success in machine vision problems
- NN also have been used for similar galaxy morphology classification tasks [1][2]

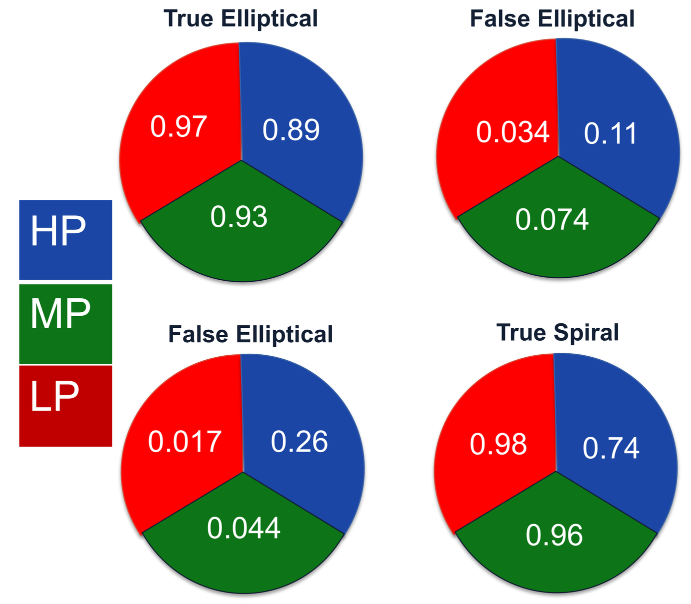
- A CNN is utilized, broadly inspired by [1]
- Augmentation is done on the fly to reduce overfitting by adding random rotations, flips, and translations
- Trained for 25 epochs on a Quadro K620 GPU, ~18 hours.
- Based on results of [2] only trained on HP dataset

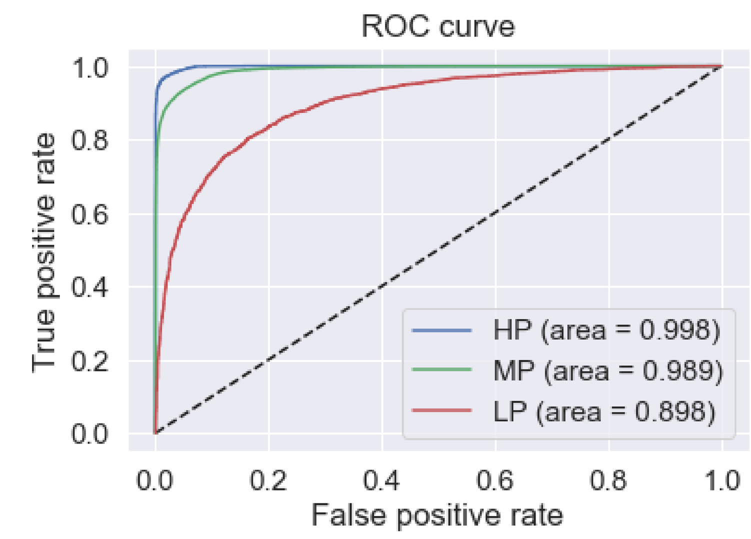
Redshift Bias in Datasets

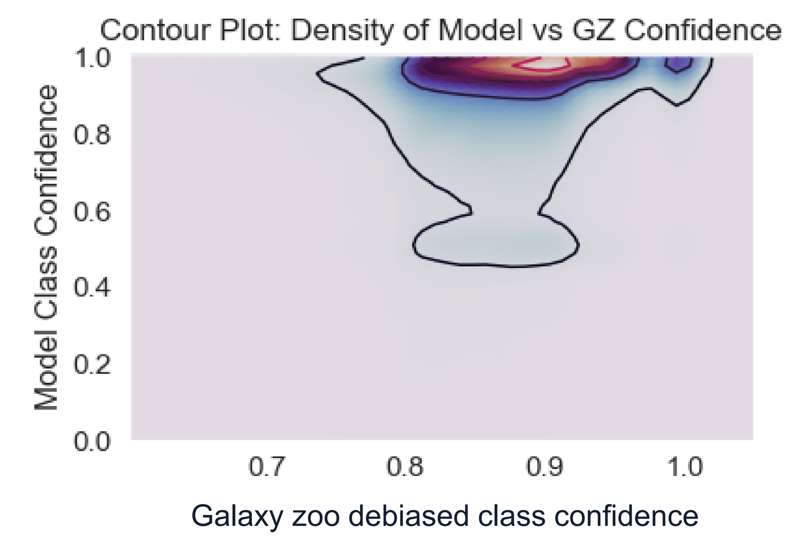


- Training data taken from the Sloan Digital Sky Survey [3] matched with the Galaxy Zoo 1 Catalogue (GZ) [4]
- GZ was a citizen-science project that used volunteers to sort galaxies into spiral or elliptical classes
- Split data into three sets: High Probability, Mid Probability and Low Probability (HP,MP,LP), along GZ debiased class confidence scores (C) for ease of exploration
- Later, added sixth channel of magnitudes, petrosian magnitudes, and angular size, boosting performance







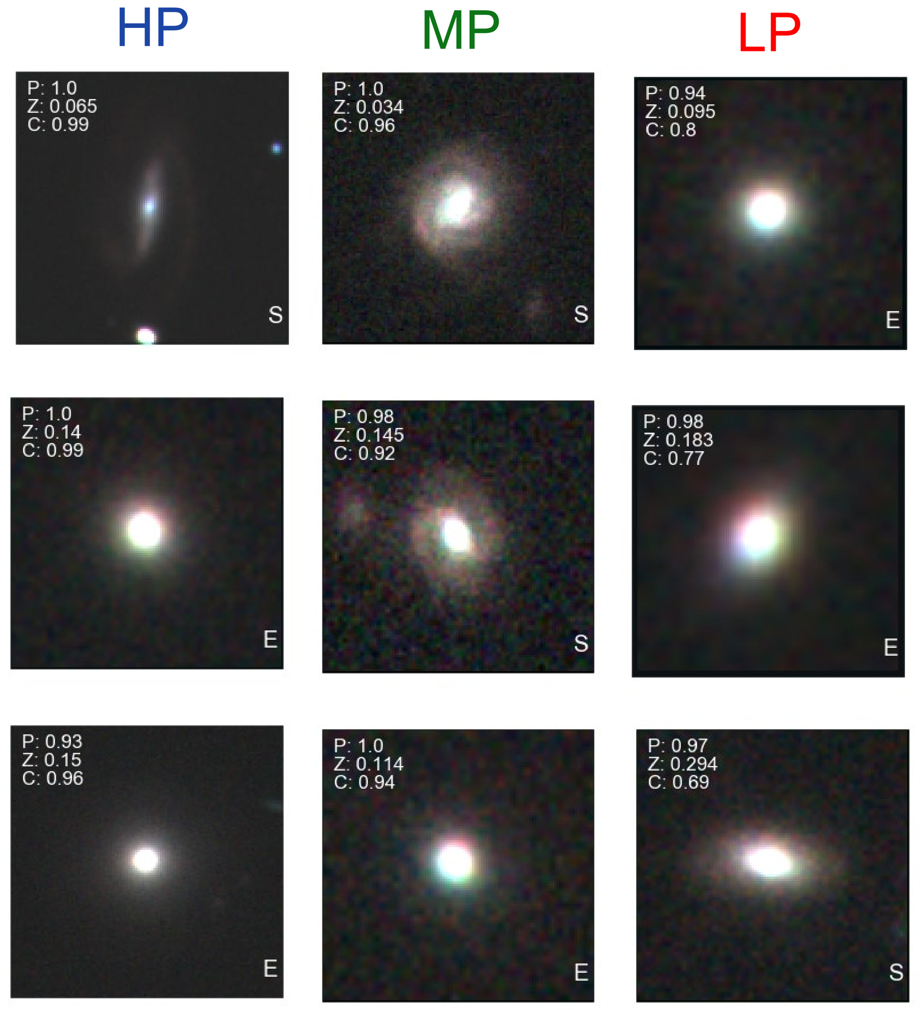


Left: Confusion matrices of each dataset: top left is an old model's confusion matrix depicting the performance on the LP dataset before 'mixing,' increasing model depth, and adding sixth channel. Performance in general degrades as human accuracy degrades. Improvements increased LP spiral performance by ~20%

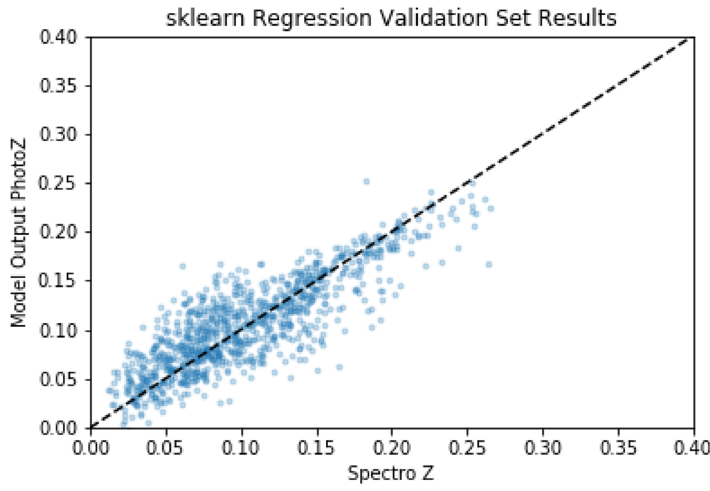
Left: Receiver Operating Characteristic curve for the HP validation dataset after training, A perfect performing model would have an area under the curve = 1.0

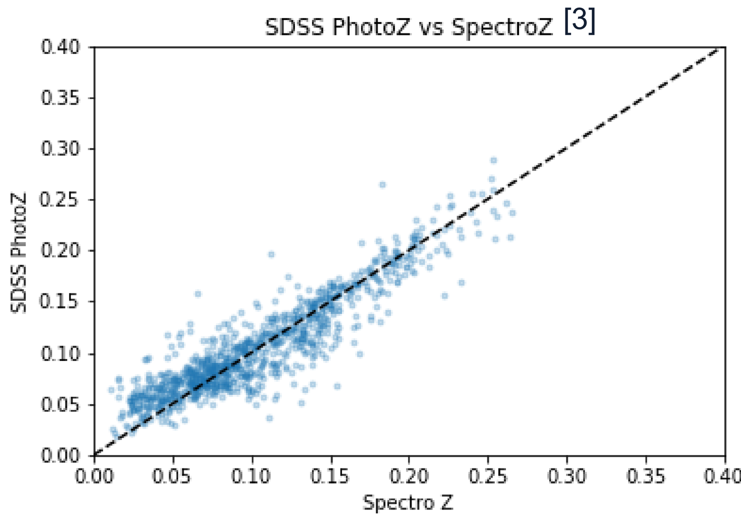
Left: Model's Class Confidence against GZ's Class Confidence. Notice: a band around 0.5 indicates there are some layers which remain un-trained yet, though our model did converge.

Below: Representative outputs of our model, with cases of both spiral ('s') and elliptical ('e'). Our model continues to have trouble classifying spiral galaxies with low disk visibility, despite significant improvements.



Preliminary Results for Photometric Redshift Estimation





- With a morphological classifier, we can utilize our output to inform a photometric Z estimator
- New sixth channel will be implemented with angular size of galaxies, extinction coefficients, and morphological class from this model
- Shown left, top is the output of a model trained on dereddened magnitudes, colors, angular size. Shown left, bottom, is the performance of SDSS DR7 Photo-Z algorithm on the same dataset.
- With a proof of concept complete, in the future we will see how performance changes with a morphological input, and upgrade to a full CNN with galactic images as inputs.

References:

[2] K. Asad, E. Huerta, *et al,* Deep learning at scale for the construction of galaxy catalogs in the Dark Energy Survey, Phy. Let. B., 795, (2019), p. 248-258

[1] H. Domínguez Sánchez, Huertas-Company, *et al.* arXiv e-prints arXiv:1807.00807 [astro-ph.GA] (2018)

[3] **SDSS**, SDSS skyserver, http://skyserver.sdss.org/dr7/en/tools/search/sql.asp (2018)

[4] C. Lintott, K. Schawinski, *et al.* Galaxy Zoo 1 : Data Release of Morphological Classifications for nearly 900,000 galaxies. Mon. Not. R. Astron. Soc., 410 (2011), p. 166

[5] G. Dálya, G. Galgóczi, *et al*, GLADE: A galaxy catalogue for multimessenger searches in the advanced gravitational-wave detector era, Monthly Notices of the Royal Astronomical Society, Volume 479, Issue 2, September 2018, Pages 2374–2381

ANTARES

ILLINOIS