Photometric Redshifts for Cosmology: Improving Accuracy and Uncertainty Estimates Using Bayesian Neural Networks

Evan Jones [1], Tuan Do [1], Bernadette Boscoe [1,3], Yujie Wan[1], Zoey Nguyen[1], Jack Singal [2]
[1] Physics and Astronomy Department, University of California, Los Angeles
[2] Physics Department, University of Richmond
[3] Computer Science Department, Occidental College

Abstract

We present results exploring the role that probabilistic deep learning models can play in cosmology from large scale astronomical surveys through estimating the distances to galaxies (redshifts) from photometry. Due to the massive scale of data coming from these new and upcoming sky surveys, machine learning techniques using galaxy photometry are increasingly adopted to predict galactic redshifts which are important for inferring cosmological parameters such as the nature of Dark Energy. Associated uncertainty estimates are also critical measurements, however, common machine learning methods typically provide only point estimates and lack uncertainty information as outputs. We turn to Bayesian Neural Networks (BNNs) as a promising way to provide accurate predictions of redshift values. We have compiled a new galaxy training dataset from the Hyper Suprime-Cam Survey, designed to mimic large surveys, but over a smaller portion of the sky. We evaluate the performance and accuracy of photometric redshift (photo-z) predictions from photometry using machine learning, astronomical and probabilistic metrics. We find that while the Bayesian Neural Network did not perform as well as non-Bayesian Neural Networks if evaluated solely by point estimate photo-z values, BNNs can provide uncertainty estimates that are necessary for cosmology.

Data: Galaxy Observations

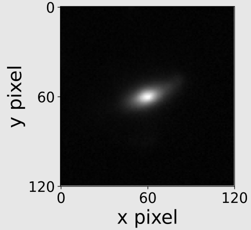
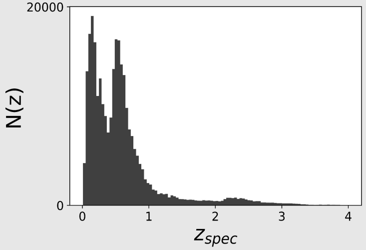
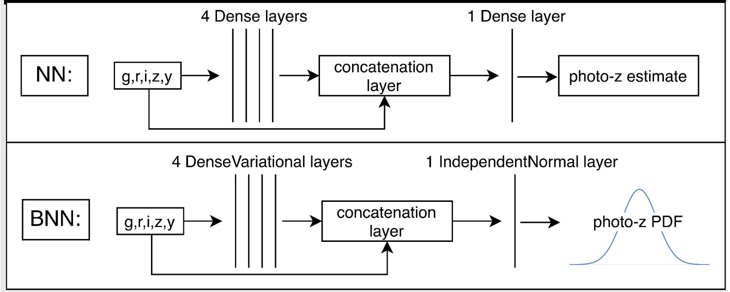


Fig 1. (left): Typical galaxy image in the i-band. We use g,r,i,z,y photometry from the Hyper-Suprime Cam (HSC) Public Data Release 2 (PDR2)[1], which is designed to reach similar depths as LSST[2] but over a smaller portion of the sky.

Fig. 2 (right): Number of galaxies in our sample as a function of redshift. In total, our data consists of 286,401 galaxies with broad-band g,r,i,z,y photometry and known spectroscopic redshifts[3]. We use 80% for training and 20% for testing.



Network architectures



Comparison of a Neural Network to Bayesian Neural Network

We find that the non-Bayesian NN produces better point estimate photo-z predictions compared to the BNN, however the BNN has the advantage of producing uncertainties for photo-z predictions. We note that the photo-z uncertainties are overestimated at low redshifts ($0 < z < 2.5$) and underestimated at higher redshifts. We believe the BNN may require more information than is present in the g,r,i,z,y inputs in order to improve performance; in a future work we will investigate galaxy images as input in a bayesian convolutional neural network.

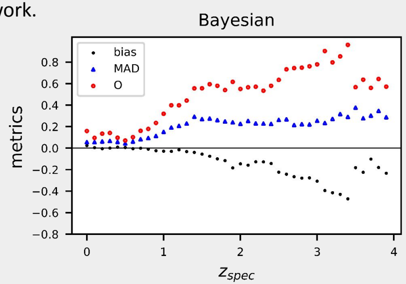
$$\mathrm{bias} = \frac{\mathrm{z_{phot}} - \mathrm{z_{spec}}}{1 + \mathrm{z_{spec}}}$$

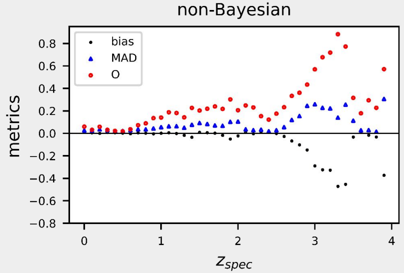
$$\mathrm{O}: \frac{|\mathrm{z_{phot}} - \mathrm{z_{spec}}|}{1 + \mathrm{z_{spec}}} > .15$$

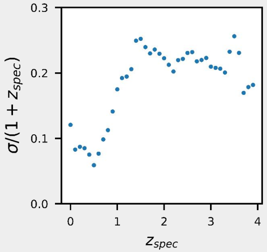
$$\mathrm{O_b}: \frac{|\mathrm{z_{phot}} - \mathrm{z_{spec}}| - \sigma}{1 + \mathrm{z_{spec}}} > .15$$

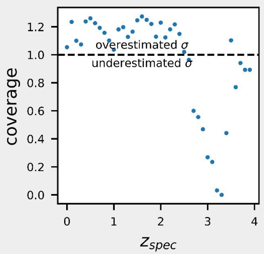
$$\mathrm{MAD} = \mathrm{Median}(|\Delta z - \mathrm{Median}(\Delta z_i)|)$$

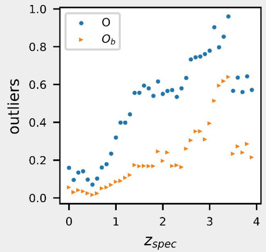
$$\mathrm{coverage} = \sum_{\mathrm{i}}^{\mathrm{n_{gals}}} \frac{(\bar{\mathrm{z}}_{\mathrm{pdf,i}} - \mathrm{z_{spec,i}}) < \mathrm{z}_{\sigma,\mathrm{i}}}{\mathrm{n_{gals}}}$$











References

[1] Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, Pasj

[2] LSST Collaboration 2021, arXiv:1809.01669 [astro-ph].

[3] https://hsc-release.mtk.nao.ac.jp/doc/index.php/dr1_specz/