Investigating Galaxy Photometry and Redshift by Deep Learning

Siou–Yu Chang[1], Yu–Yen Chang[1,2]

1. Department of Physics, National Tsing Hua University, Taiwan
2. Academia Sinica Institute of Astronomy and Astrophysics, Taiwan

Abstract We use deep learning techniques to analyze galaxy parameters with Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP) Public Data Release 2 (PDR2) data. With Keras sequential model from Python package, we predict y-band data from g-,r-,i-, and z- band photometry, as well as estimate redshifts of galaxies from g-, r-, i-, z-, and y- band photometry. We optimize our inputs by normalizing fluxes, magnitudes, and colors. As a result, flux with individual galaxy normalization can obtain the best performance of r^2 values (>0.95 for photometry estimation and >0.70 for redshift estimation). We separate the sample to different redshift bins as various classifications, but do not find better results than our regression models. In the future, we will also include five-band HSC imaging as input data, and apply our methods to global galaxy properties.

SED fitting problem

- The traditional way to know the redshift of galaxies is to use spectrometer (spec-z), but in the absence of sufficient telescope time to determine a spectroscopic redshift for each object, the technique of photometric redshifts (photo-z) provides a method to determine at least the qualitative characterization of a redshift.

- However, Spectral Energy Distributions (SED) fitting also needs plenty of multiwavelength information to fit. We would like to find another method which can achieve the same result of SED fitting with less band data. Our first idea is to use deep learning techniques in machine learning to help us to achieve this.

Data

- We use Hyper Suprime-Cam Subaru Strategic Program(HSC-SSP) Public Data Release 2 (PDR2) Wide data.
- We select the data with effective spec-z, and select the range of redshift between 0.0 to 1.2.
- We use bootstrapping method to calculate the standard deviation.
- We compare our result with "photo-z best", which is the photo-z calculated from the HSC-SSP team (Mizuki; see Tanaka et al, 2018 for more details).

Method

- In the test, the goal is to fit five bands data to know the redshift of galaxies. We use flux, magnitude, flux plus its uncertainty (fluxsigma), and color to test whether different types of data of galaxies has an impact on the results.

- We mainly use Keras sequential model in Python packages to do deep learning. We also predict redshift by using various algorithms in Machine Learning (Table 2).

	Train r^2	Test r^2
Linear Regression	0.14	0.13
Medium Tree	0.65	0.67
Bagged Trees	0.75	0.76
Trilayered Neural Network	0.70	0.69

Table 1: *Accuracy of redshift prediction by using various algorithms in Machine Learning.*

Further work (Quasar)

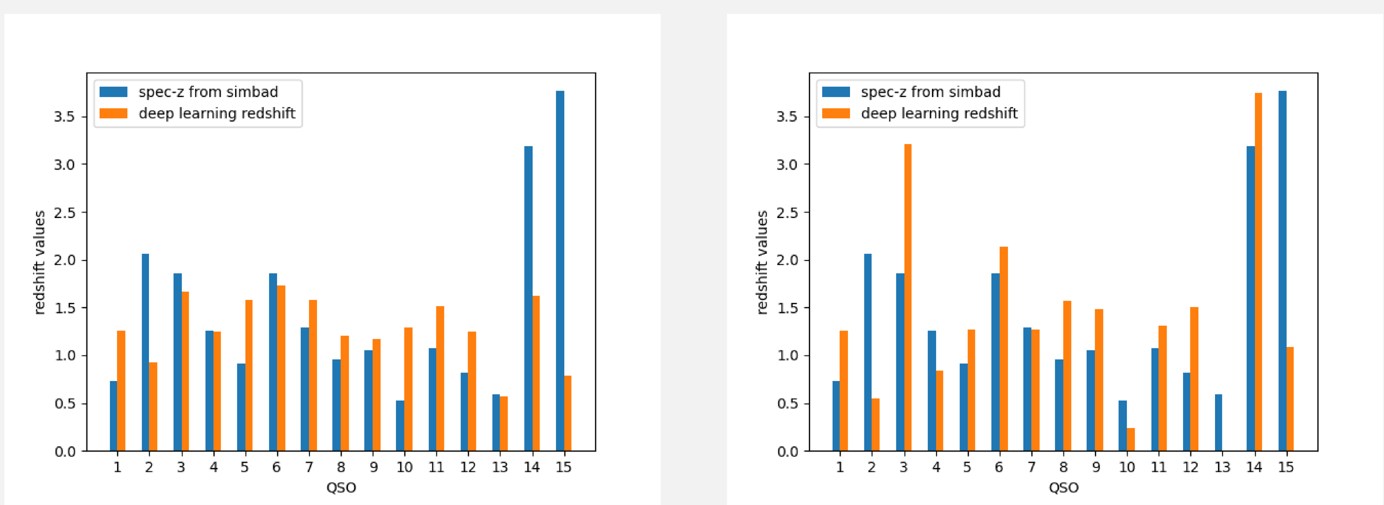

Fig. 4: *We also try to train the QSO data based on the built model. The left figure is the prediction result using QSO colors with different redshift; the right figure is using flux plus sigma as input.*

References

[1] Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114A
[2] Gupta, A., Zorrilla Matilla, J. M., Hsu, D., et al. 2018, PhRvD, 97, 10
[3] Hung-yi, L. 2016, online open course: ML Lecture
[4] Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
[5] Keras.io. 2015, https://keras.io/
[6] Tanaka, M., Coupon, J., Hsieh, B. C., et al. 2018, PASJ, 70, 9T

Results

- We have tested many kinds of data, including magnitude, color, flux and flux plus fluxsigma, our result suggests that sequential model does not work well if the data range is out of 0 to 1. This may be one of the reasons why color fitting does not require normalization (Fig. 1).
- The best result is using flux and fluxsigma with normalizing the same galaxy (r^2=0.710). It is reasonable, the more completed band information we consider, the better fitting results can be obtained (Fig. 2, Fig. 3).
- We also compare our result with photo-z, we find that photo-z's predictions are evenly distributed over the entire redshift range, while our model prefers to predict data at low redshifts (z <~ 1.2) (Table3, Table4).

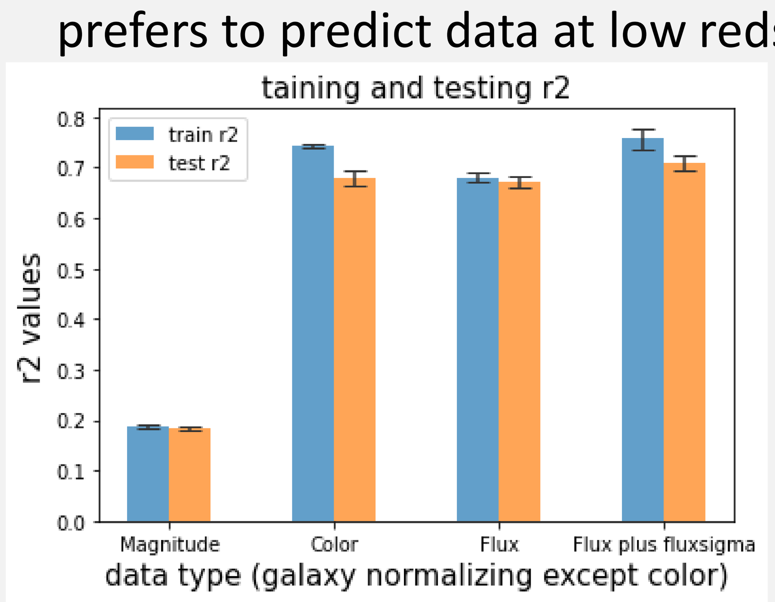

Fig. 1: *Estimate redshifts of galaxies from different type g-, r-, i-,z-, and y- band data. We find that the best result is by using flux and fluxsigma with normalizing the same galaxy. It can get a result of r^2=0.710.*

	r^2	MAE	η	σ
Deep learning (this work)	0.713	0.100	9.7%	0.052
Photo-z	0.209	0.116	12.3%	0.046

Table 2: *Comparison of our model prediction and photo-z in the redshift selection of 0.0 to 1.2.*

	r^2	MAE	η	σ
Deep learning (this work)	0.461	0.140	15.2%	0.066
Photo-z	0.378	0.138	14.1%	0.049

Table 3: *Comparison of our model prediction and photo-z in full-range redshift selection.*

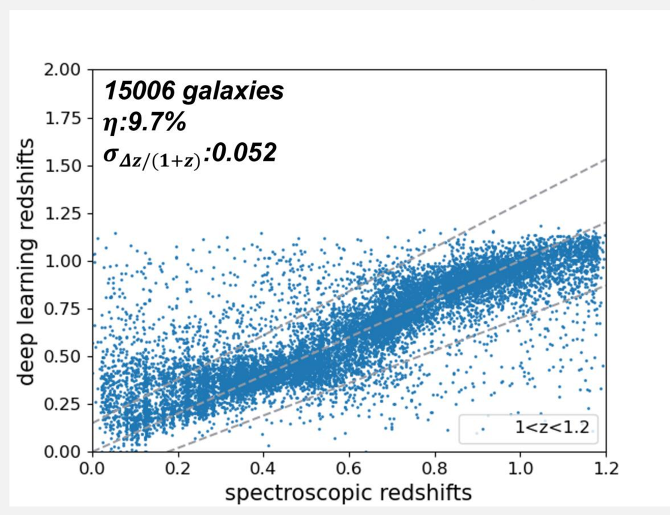

Fig. 2: *Our model gives good result when predicting low redshift data.*

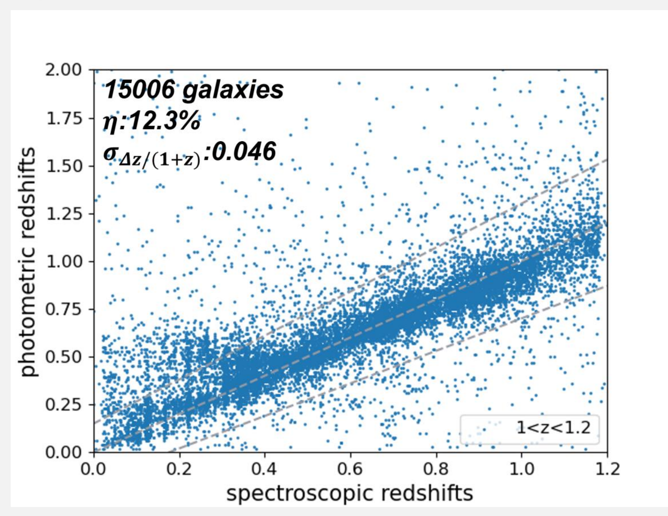

Fig. 3: *In the case of high redshift prediction, the result given by photo-z is definitely more reasonable.*

Summary

1. For redshift estimations of galaxies from g-, r-, i-, z-, and y- band, we put many different galaxy information and do some adjustment. The best result for spectroscopic redshift estimation is by using flux and its uncertainty (fluxsigma) with normalizing the same galaxy. It can get a result of r^2=0.710 and the redshift accuracy of $\sigma_{\Delta z/(1+z)}$ can be 0.052.

2. The results obtained by our Keras model are not inferior to photo-z provided by other teams, and it seems that our results have some advantages in the number of predictions and the predictions of the low redshift selected.

3. Although photo-z performs slightly worse overall, it far outperforms our model at predicting high redshifts, and it can also predict in the presence of missing values.