

Applying Deep Learning to the Galaxy Photo-z Problem

Samantha E. Thrush^{1*} and Robert J. Brunner^{1,2,3,4}

¹*Department of Astronomy, University of Illinois, Urbana, IL 61801 USA*

²*Department of Physics, University of Illinois, Urbana, IL 61801 USA*

³*Department of Statistics, University of Illinois, Champaign, IL 61820 USA*

⁴*National Center for Supercomputing Applications, Urbana, IL 61801 USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

When large data sets of galaxies are collected, their associated redshifts are usually collected by finding their photometric redshifts, which takes additional time for a survey to complete. However, if deep learning via deep convolutional neural networks (ConvNets) are utilized, photometric redshifts of galaxies can be predicted without the need for photometry to be completed by the associated survey. In this paper, we present an application of ConvNets on reduced and pixelated galaxy images and redshift data from the Sloan Digital Sky Survey (SDSS). We will show that ConvNets are able to accurately predict the redshift of a sample galaxy image in a way that is competitive with previous works on this subject.

Key words: methods: data analysis – techniques: image processing – methods: statistical – surveys – galaxies:statistics.

1 INTRODUCTION

Before the advent of machine learning, finding a galaxy’s redshift required the collection of its spectra so as to ascertain its recession speed and thus its distance and redshift from the observer using Hubble’s Law. Unfortunately, spectroscopy can only be done on a select number of targets at a time due to time constraints of both setting up the spectrograph (be it a slit or a plate and optic fiber setup) and exposure time (York 2000). In contrast, photometric images are much quicker to observe as very little setup is needed in order to obtain the image (apart from calibrations) and each image taken requires a comparatively shorter exposure time (citation needed). Because of this, the idea of only using an object’s image to ascertain its redshift is very attractive.

Using an object’s photometric image to find its redshift (here-in known as photo-z) is not a new concept (citations needed, look at PHAT paper), but with the advent of machine learning

In this paper, we will first review previous works on this subject in Section 2, followed by an explanation of the key ways this body of work differs from previous works. In Section 3, the Deep Learning code and the selection of the data set from the Sloan Digital Sky Survey (SDSS) will be discussed. Section 4 will contain the results of the Deep Learning code, which will be subsequently discussed and analyzed. Finally, Section 5 will contain the conclusions of this paper.

2 LITERATURE REVIEW

Ascertaining the redshifts of galaxies using spectroscopic redshifts has classically been a time-consuming pursuit. Fortunately, with the advent of machine learning, it is possible to ascertain the redshift of a galaxy only using photometric data, thus ascertaining the photometric redshift or "photo-z" as it is commonly called. The project at hand is composed of many parts, so in order to organize the logic behind all of the choices and decisions made, it is important to reflect upon previous research. First, the Sloan Digital Sky survey and its bearing on the project will be discussed. Then, we will consider previous projects that utilize Neural Networks for its applications to images and regressing galaxy images to find their redshifts.

2.1 Sloan Digital Sky Survey

The Sloan Digital Sky Survey Data Release 12 (SDSS-DR12; Alam et al. 2015), which includes data from phases I-III (April 2000 through 14 July 2014), has served the astronomical community by providing photometric data on 14,555 square degrees of the sky (approximately one-third of the sky), resulting in the observation of approximately 470 million individual targets in five different bands (u, g, r, i, and z), with magnitudes brighter, or comparable to 22 (the limiting magnitude). Of these targets, only approximately 200 million are galaxies. In addition to photometric information, SDSS-DR12 also provides spectroscopic data on over 4.35 million targets, 2.4 million of which are galaxies.

* thrush2@illinois.edu

2.2 Neural Networks and Photometric Redshifts

There are many different ways of ascertaining a galaxy's photometric redshift using machine learning, as was shown in the PHAT Photo-z data trial (Hildebrandt et al. 2010).

3 METHODOLOGY

3.1 Data Selection

The various band pass trials described below were completed on SDSS galaxy images. These galaxies were selected with CasJobs (Li and Thakar 2008) in such a fashion so as to maximize fidelity of the data. This was achieved by selecting for galaxies with clean photometry (photo.clean = 1), no known issues with its spectra (spec.zWarning = 0), small redshift errors (spec.zErr < 0.1) and redshifts under 2 (spec.z < 2). In addition, the exponential scale fit radius and the de Vaucouleurs fit scale radius were required to be below 30 arc seconds (so as to exclude bad photometric images).

100,000 galaxies selected from SDSS-DR12 were randomly selected from the subset of galaxies fitting the criteria above using CASJOBS (Li & Thakar 2008). These selection criteria were taken and modified from Edward Kim's work (used in Kim and Brunner 2016).

3.2 Data Processing

The collected

3.3 Neural Network Code

Through the use of a supervised neural network code created by Edward Kim (used in Kim and Brunner 2016), the data collected from SDSS

3.4 Band Pass Trials

In order to ascertain the effectiveness of different band pass image combinations on the Neural Network's ability to assign an accurate redshift to the image,

4 RESULTS AND DISCUSSION

4.1 Five Band Pass Results

4.2 Four Band Pass Results

4.3 Three Band Pass Results

4.4 Two Band Pass Results

4.5 One Band Pass Results

4.6 Discussion

5 CONCLUSIONS

ACKNOWLEDGEMENTS

We acknowledge support from the National Science Foundation Grant No. AST-1313415. RJB acknowledges support as an Associate within the Center for Advanced Study at the University of Illinois.

Funding for SDSS-III has been provided by the Alfred P.

Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is <http://www.sdss3.org/>.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

REFERENCES

- Alam S., et al., 2015, *The Astrophysical Journal Supplement Series*, 219, 12
Hildebrandt H., et al., 2010, *Astronomy and Astrophysics*, 523, A31
Li N., Thakar A. R., 2008, *Computing in Science Engineering*, 10, 18
York D. G., 2000, *The Astronomical Journal*, 120, 1579

This paper has been typeset from a \LaTeX file prepared by the author.