Applying deep learning to galaxy morphology

Paulo H. Barchi[a1], Reinaldo R. de Carvalho[a], R. R. Rosa[a], Rubens Sautter[a], Marcelle Soares-Santos[b], Bruno A. D. Marques[c] and Esteban Clua[c]

[a]National Institute for Space Research (INPE), São José dos Campos, SP, Brazil
[b]Brandeis University, Waltham, MA, USA
[c]Fluminense Federal University (UFF), Niterói, RJ, Brazil

Received on January 20, 2019 / accepted on *****, 2019

Abstract

In observational cosmology, one of the most fundamental and basic procedures is the morphological classification of galaxies into a taxonomy system. The challenge is to build up a robust methodology to perform a reliable classification. The main objective of this work is to investigate how to substantially improve the classification of galaxies within large datasets by mimicking human classification. We combine accurate visual classifications (from Galaxy Zoo project) with Deep Learning methodology.

The first classification system, by [1, 2], distinguishes galaxies with dominant bulge component – also known as Early-Type Galaxies (ETGs) – from galaxies with a prominent disk component – named Late-Type Galaxies (LTGs). LTGs are commonly referred to as spiral galaxies because of their prominent spiral arms, while ETGs are commonly referred to as elliptical (E) galaxies as they have a simpler ellipsoidal structure, with less structural differentiation (less information).

Morphology reveals structural, intrinsic and environmental properties of galaxies. In the local universe, ETGs are mostly situated in the center of galaxy clusters, have a larger mass, less gas, higher velocity dispersions, and older stellar populations than LTGs, which are rich star-forming systems [3, 4, 5]. By mapping where the ETGs are, it is possible to map the large-scale structure of the universe. Therefore, galaxy morphology is of paramount importance for extragalactic research as it relates to stellar properties and key aspects of the evolution and structure of the universe.

Galaxy Zoo 1 (GZ1) [6, 7] is the first phase of a citizen science project which provides the distinction between ETGs and LTGs. More refined classifications fork spirals into two groups: barred (SB) and unbarred (S) galaxies. These two groups can also be refined even further by their spiral arms strength. Galaxy Zoo also provides a more detailed catalog in its second phase (GZ2) [8].

[1]E-mail Corresponding Author: paulobarchi@gmail.com

2 classes	3 classes	7 classes	9 classes	11 classes
99.5	82.7	70.0	67.4	65.2

Table 1: Overall Accuracy (OA in percentage) achieved with Deep Learning techniques for all problems considering different numbers of classes.

We use two samples in this work: (1) 58,030 galaxies classified by GZ1; and, (2) 67,637 galaxies from GZ2 for more detailed classification. Considering 11 classes, we use classifications with the following prefixes Er, Ei, Ec, Sa, Sb, Sc, Sd, SBa, SBc, SBd as supervision guide to the learning process. Other three different scenarios are explored with GZ2 supervision. Classification considering 9 classes (same as 11 classes except that we have one class for all elliptical galaxies united), 7 classes (same as previous but disconsidering the faintest galaxy types: Sd and SBd) and 3 classes: E (elliptical), S (non-barred spiral) and SB (barred spiral galaxies). The datasets used in this work are composed by galaxies from Sloan Digital Sky Survey Data Release 7 (SDSS DR7) [9].

Deep Convolutional Neural Network is a well-established methodology to classify images [10]. Without the need of a feature extractor, the network itself adjusts its parameters in the learning process to extract the features. We use the GoogLeNet Inception architecture, a Deep Convolutional Neural Network with 22 layers directly applied to the images [11].

We explore the class imbalance problem considering 3 classes, with undersampling, oversampling and Synthetic Minority Over-sampling Technique (SMOTE) experiments experiments [12]. Performance is measured by Overall Accuracy (OA), also Precision (P) and Recall (R) for specific cases. We achieve 99% OA for separating galaxies into 2 classes and $\approx$ 83% OA for 3 classes (E, S and SB). The Overall Acurracy of all models are presented in Table 1.

The main product of these years of research is a complete catalog with morphology for $\approx$ 670.000 galaxies – which will be published soon with the paper that is currently under revision [13]. The result of this ongoing PhD research has already classified hundreds of thousands of galaxies from SDSS and has potential to provide morphological classification for millions of galaxies from different surveys.

This work is intended to be presented as a talk at the special session SS01 coordinated by R. R. Rosa and C. A. Careta.

Keywords: Galaxy morphology, Deep learning.

Acknowledgments. This work was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. R.R.dC. and R.R.R. acknowledge financial support from FAPESP through grant # 2014/11156-4. RRR and PHB thank Santos Dumont Supercomputer-LNCC for providing 500KUAs for the partial development of this research. The authors thank to MCTIC/FINEP (CT-INFRA grant 0112052700), the Embrace Space Weather Program for the computing facilities at INPE and NVIDIA for providing GPUs. We thank Dr. Bjoern Penning for thoughtful comments about the paper.

References

[1] E. P. Hubble, 1926, *The Astrophysical Journal*, 64, 321.

[2] E. P. Hubble, 1936, Realm of the Nebulae. *New Haven: Yale University Press.*

[3] M. S. Roberts and M. P. Haynes. Physical Parameters Along the Hubble Sequence. *Annual Review of Astronomy and Astrophysics*, 32(1):115–152, 1994.

[4] M. R. Blanton and J. Moustakas. Physical Properties and Environments of Nearby Galaxies. *Annual Review of Astronomy and Astrophysics*, 47(1):159–210, 2009.

[5] L. Pozzetti. et al. zCOSMOS - 10k-bright Spectroscopic Sample - The Bimodality in the Galaxy Stellar Mass Function: Exploring its Evolution with Redshift. *Astronomy and Astrophysics*, 523:A13, 2010.

[6] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, et al. Galaxy Zoo: Morphologies Derived from Visual Inspection of Galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189, 2008.

[7] C. Lintott, K. Schawinski, S. Bamford, A. Slosar, K. Land, D. Thomas, E. Edmondson, K. Masters, R. C. Nichol, M. Jordan Raddick, A. Szalay, D. Andreescu, P. Murray, and J. Vandenberg. Galaxy Zoo 1: Data Release of Morphological Classifications for Nearly 900,000 Galaxies*. *Monthly notices of the royal astronomical society*, 410(1):166–178, 2011.

[8] K. W. Willett, C. J. Lintott, S. P. Bamford, K. L. Masters, B. D. Simmons, K. R. V. Casteels, E. M. Edmondson, L. F. Fortson, S. Kaviraj,

W. C. Keel, T. Melvin, R. C. Nichol, M. J. Raddick, K. Schawinski, R. J. Simpson, R. A. Skibba, A. M. Smith, and D. Thomas. Galaxy Zoo 2: Detailed Morphological Classifications for 304,122 Galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 435(4):2835–2860, 2013.

[9] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. A. Prieto, S. F. Anderson, J. A. Arns, E. Aubourg, S. Bailey, and E. et al. Balbinot. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *The Astronomical Journal*, 142(3):72, 2011.

[10] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[11] C. Szegedy et al. Going Deeper with Convolutions. Computer Vision and Pattern Recognition, 2015. `http://arxiv.org/abs/1409.4842`.

[12] F. Pedregosa et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, abs/1201.0490, 2012. `http://arxiv.org/abs/1201.0490`.

[13] P. H. Barchi, R. R. de Carvalho, R. R. Rosa, R. Sautter, M. Soares-Santos, B. A. D. Marques, E. Clua. Machine and Deep Learning Applied to Galaxy Morphology - A Complete Classification Catalog. *Monthly Notices of the Royal Astronomical Society*, Accepted/Under Review, 2019.