Mem. S.A.It. Vol. 94, 77

Memorie della

CNN Lesson Learned from Two Largest Galaxy Morphological Classification Catalogues

Ting-Yun Cheng[1], H. Domínguez Sánchez[2,3], J. Vega-Ferrero[4], C. J. Conselice[5], M. Siudek[6,3]

[1] Centre for Extragalactic Astronomy, Durham University, South Road, Durham DH1 3LE, UK
[2] Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Plaza de San Juan, 1, 44001 Teruel, Spain
[3] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrerde Can Magrans, s/n, 08193 Barcelona, Spain
[4] Instituto de Astrofíısica de Canarias (IAC) La Laguna, 38205, Spain
[5] Jodrell Bank Centre for Astrophysics, University of Manchester, Oxford Road, Manchester M13 9PL, UK
[6] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, 08193 Bellaterra (Barcelona), Spain
e-mail: `ting-yun.cheng@durham.ac.uk`

Received: 29-11-2022; Accepted: 23-02-2023

Abstract. We compare two large morphological catalogues built by applying two different convolutional neural networks (CNN) methodologies to the Dark Energy Survey (DES) Year 3 dataset. One work trains CNN with bright galaxies ($i < 18$) using i-band images of linear, logarithmic, and gradient scales, while the other work's CNN is trained with fainter emulated galaxies based on the bright samples and *gri*-band images. Despite the different CNN approaches used for the construction of the catalogues, the agreement between the two catalogues is excellent up to $i < 19$, demonstrating that CNN predictions are reliable for samples at least one magnitude fainter than the training sample limit. It also shows that morphological classifications based on monochromatic images are comparable to those based on *gri*-band images, at least in the bright regime. By studying the mismatched cases we are able to identify lenticular galaxies (at least up to $i < 19$), which are difficult to distinguish using standard classification approaches.

Key words. methods: data analysis – methods: statistical – galaxies: structure

1. Introduction

Galaxy morphology describes the visual features of a galaxy and the structure of its light distribution. In addition, galaxy morphologies are related to their stellar populations, likewise for galaxy stellar masses, star formation rates, ages, and metallicities (e.g., Conselice 2006). By investigating the connection of these physical properties with galaxy morphology helps understanding galaxy evolutionary history and stages.

Traditionally, the morphological classification of galaxies has been based on visual inspection. Due to the fast growth in the size of galaxy datasets such as Dark Energy Survey (DES Collaboration 2005; DES Collaboration et al. 2016, hereafter, DES), Euclid Space Telescope, and the Vera Rubin Observatory Legacy Survey of Space and Time (Ivezić et al. 2019), a broad range of of machine learning techniques have been applied to accelerate the classification process. Since machine learning techniques are continuously being introduced and applied to astronomical studies including diverse approaches that can be applied within a single technique, there are not many comparison of a machine learning technique that can be used to assess the potential pros and cons in our current way of using them.

With the approaches using convolutional neural networks (CNN), Vega-Ferrero et al. (2021, hereafter V21) and Cheng et al. (2021, hereafter C21) built two of the largest morphological classification catalogues to date using DES imaging data. These two catalogues include largely overlapping samples (over 17 million galaxies) from DES using the same technique (i.e. CNN) but different approaches which provide one of the first opportunities to statistically assess and discuss the impact of different ways of using CNN. Cheng et al. (2023), the work this manuscript is based on, carried out a massive comparison between the two catalogues. This can not only further validate the classification of the two catalogues, but also provides an inspection of different approaches to use CNN for future users of galaxy morphology studies.

2. Comparison between Catalogues

The two catalogues are built by analyzing DES data: C21 includes ∼21 million galaxies with an i-band magnitude range $16 \leq i < 21$ and at redshift $z < 1$, and V21 contains ∼27 million galaxies with r-band magnitude brighter than 21.5 and without a specific redshift cut. Due to different approaches in CNNs, different criteria in selecting initial samples are applied. This results in an overlap of ∼17 million galaxies, and ∼ 9 million and ∼ 3 million unique galaxies in V21 and C21, respectively. The union of the two catalogues increases the total number of morphological classifications to ∼30 million galaxies.

Apart from the sample selection, the CNN architectures used in the two works are different, such that V21 uses four convolutional layers and one dense layer while C21 uses three convolutional layers and two dense layers. The hyperparameters used in the architectures are also different. This could impact the classifications, but this work cannot provide a fair discussion due to many differences between the training datasets as discussed later. This investigation would be better carried out using the same datasets, and with architectures optimised for the same task, to separate the effect from others (see a paper of computer science on this topic, Alzubaidi et al. 2021). Since each work reaches a high accuracy on its own classification task, we simply focus on the impact of following factors on the classifications:

Training labels: V21 used the T-Type presented in Domínguez Sánchez et al. (2018, hereafter, DS18), which are based on a deep learning model trained on the T-Type provided by the visual classification of Nair & Abraham (2010). Their labels are early-type galaxies (ETG; for galaxies with T-Type < -0.5) or late-type galaxies (LTG; for galaxies with T-Type > 0.5). Galaxies with intermediate T-Types ($-0.5 <$T-Type< 0.5) were excluded from the training sample. C21 used the classifications of spiral and elliptical galaxies from Galaxy Zoo 1 catalogue (GZ1; Lintott et al. 2008, 2011), and therefore C21 separates elliptical (Es) from spiral galaxies (Sp). A correction in the GZ1 visual classifications was applied due to the better resolution and deeper images of the DES data compared to SDSS (Cheng et al. 2020). After this correction, the Sp class in C21 includes galaxies with disk structures such as lenticular galaxies.

Brightness of the training sample: Both catalogues use bright SDSS galaxies ($r < 17.7$ in V21 and $i < 18$ in C21) as the basis of their training sample. While C21 only used the DES i-band images of bright galaxies as training sample, V21 artificially created images at higher redshift and fainter magnitudes

($r < 22.5$) using the bright samples by considering flux and size corrections, k-correction, and evolutionary effects (details on the emulation procedure are described in section 2.3 of V21). They included these faint galaxies in their training sample keeping their original morphological labels.

Input to the CNN: V21 used g, r, i band images (after normalising each band individually for each galaxy) while C21 used only i-band images, but combined linear, logarithmic, and gradient images. This means that the V21 machine focuses on different structures that are shown in different wavelengths, while the C21 machine considers different structures emphasised in different scales, but uses a single band image.

Classifications from two works are defined in a different way. For simplicity, we focus only on the 'robust' classifications for V21, i.e. for ETG and LTG when $max\,(P_i) < 0.3$ and $min\,(P_i) > 0.7$, respectively, where P_i represents the median probability obtained from 5 k-folded models. For C21, we use classifications based on the probability thresholds mentioned above, i.e. $\overline{P} \geq 0.8$. These classifications are referred to as 'certain' type in the discussion. Hereafter, we will refer to Sp/Es for C21 classifications and to ETG/LTG for V21 ones.

3. Morphology Comparison

To be able to compare galaxies one-to-one, we restrict the analysis to the intersection of the two catalogues (17,821,250 galaxies) [see Section 5 in Cheng et al. (2023) for the discussion of all samples]. We define the agreement as the fraction of matched classifications (i.e. Es & ETG or Sp & LTG) from the total number of galaxies with a certain classification. Regardless of the significant differences in both approaches, Fig. 1 shows the overall agreement is very good, larger than 92% in all magnitude ranges. However, at fainter magnitudes, the agreement is largely driven by the matched samples that classifed as Sp by C21 and LTG by V21, and in the last magnitude bin, i=[20,21), only ~0.1% of the galaxies are classified as Es & ETG. To further investigate the agreement of each morphology class, we show

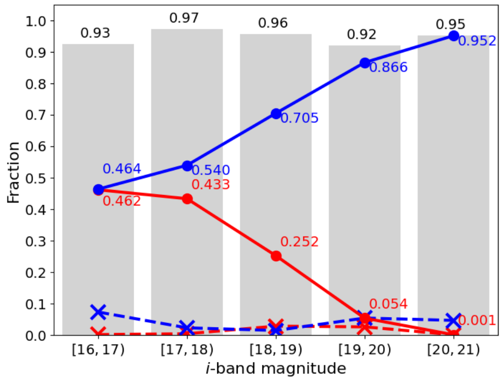

Fig. 1. Agreement of certain types within different magnitude bins in i-band. Grey bars show the percentage agreement of all galaxies with certain classifications. Solid lines and dashed lines represent the fraction of galaxies with matched (i.e., either Sp & LTG or Es & ETG) and mismatched (i.e., either Sp & ETG or Es & LTG) classifications, respectively. Blue dashed lines show the fraction of galaxies that are Sp in C21 but ETG in V21 while red lines present the fraction of ones with a class of Es in C21 and LTG in V21.

in Fig. 2 the confusion matrices in 4 magnitude bins. There is an excellent agreement between both morphology classes up to $i \leq 19$. Assuming that the V21 classification is correct, this indicates that the CNN predictions of C21 are reliable for samples at least one magnitude fainter than the training sample.

Unfortunately, the absence of 'ground truth' for the faint DES galaxies prevents us from claiming which are the 'right' or 'wrong' classifications. We then further investigate the properties of galaxies with mismatched classifications to shed some light on their nature and their true morphological class [see Fig.8-12 in Cheng et al. (2023)].

For the mismatched case of Sp & ETG, at bright magnitudes ($16 \leq i < 18$), 94% of them are labelled as lenticular galaxies according to DS18. This suggests that bright galaxies with a classification of Sp & ETG are likely lenticular galaxies, and the mismatch are simply due to different definitions in labels. Sp & ETG galaxies have intermediate to large Sersic index ($\sim 2-4$), high stellar mass (peaking at $log_{10}\,M^*/\mathrm{M}_\odot \sim 11$), and similar colour distri-

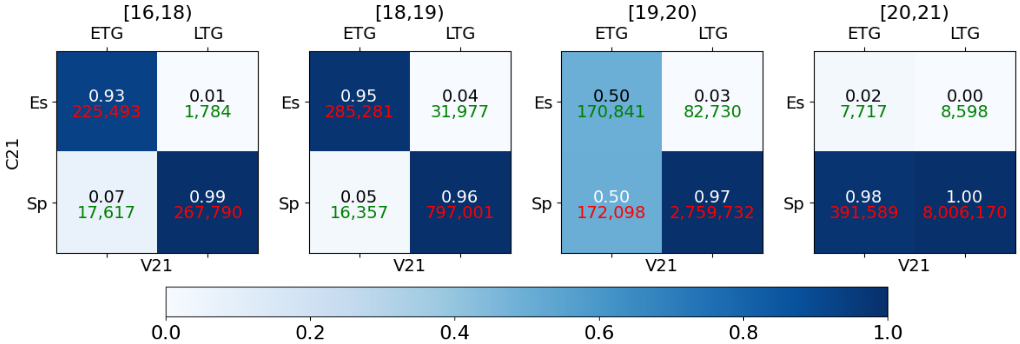

Fig. 2. Confusion matrices of the intersection samples in different magnitude bins. The red or green text in each quadrant represents the number of galaxies with classification in agreement between C21 and V21. The number above is the fraction of these galaxies with respect to the V21 classification.

butions to the Es & ETG ($g - r$ peaking at ∼1). Fainter Sp & ETG galaxies ($i \geq 19$) follow a similar Sérsic index distribution of bright Sp & ETG, peaking at intermediate range ($n \sim 3$). They are also red and massive objects, with very similar colour and mass distributions to the case at bright magnitudes. However, the images at fainter magnitudes are very noisy and it is difficult to confirm their morphology by eye.

For the other case of Es & LTG, these contribute less than 0.4% and ∼ 1% to the galaxies with certain classification in the bright ($16 \leq i < 18$) and faint magnitude ($i \geq 19$) ranges, respectively. They are generally blue ($g - r$ peaking at ∼0.5), round galaxies (95% have ellipticity smaller than 0.5) with intermediate Sèrsic indices. At a bright magnitude, the investigation confirms the improvement by emulating galaxies to a fainter magnitude. However, a bimodal Sérsic index distribution appears at fainter magnitudes [see Fig. 12 in Cheng et al. (2023)]. These galaxies are difficult to classify and could be a mixture of face-on disk galaxies with no signs of spiral structure, lenticular, or elliptical galaxies. This indicates that the emulation process does not increase the diversity of galaxy morphologies, and more complete population of galaxies will be needed in future studies.

4. Summary

In this work we compare the two largest galaxy morphological classification catalogues to date. We examine the agreement (defined in Section 3) between the two catalogues using the intersection sample, and the agreement is as high as ∼95% (Fig. 1). However, the large agreement is mostly driven by the Sp & LTG population. If separately analysing different morphology types, there is an excellent agreement between the two catalogues up to $i < 19$ (Fig. 2).

One of the main results from this comparison is that the C21 machine can push its predictions accurately one magnitude fainter than its training samples. This result also indicates that the use of multi-band images does not provide a significant improvement or it poses similar effect in the morphological classifications of galaxies when compared to the use of monochromatic images with different scales (i.e. linear, logarithmic, and gradient).

By investigating the mismatched cases, we summarise that the galaxies classified as Sp by C21 and ETG by V21 are likely lenticular galaxies, in particular at bright magnitudes, while the properties of the ones classified as Es by C21 and LTG by V21 are closer to the case of LTGs. The former indicates that the use of different labels can be used to identify difficult types such as lenticulars. The latter shows that by emulating bright galaxies to fainter magnitude helps to improve the classification of Es

& LTG. This however does not increase the diversity of galaxy morphologies, and results in confused classifications at fainter magnitudes. One may consider hydrodynamic simulation to build samples with complete population of galaxies.

Acknowledgements. TYC acknowledges the support of STFC grant ST/T000244/1 and Royal Society grant RF/ERE/210326. HDS acknowledges the support by the PID2020-115098RJ-I00 grant from MCIN/AEI/10.13039/501100011033.

References

Alzubaidi, L., Zhang, J., Humaidi, A. J., et al. 2021, Journal of Big Data, 8

Cheng, T.-Y., Conselice, C. J., Aragón-Salamanca, A., et al. 2021, MNRAS, 507, 4425

Cheng, T.-Y., Conselice, C. J., Aragón-Salamanca, A., et al. 2020, MNRAS, 493, 4209

Cheng, T. Y., Domínguez Sánchez, H., Vega-Ferrero, J., et al. 2023, MNRAS, 518, 2794

Conselice, C. J. 2006, MNRAS, 373, 1389

DES Collaboration. 2005, arXiv e-prints, astro

DES Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, MNRAS, 476, 3661

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111

Lintott, C., Schawinski, K., Bamford, S., et al. 2011, MNRAS, 410, 166

Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179

Nair, P. B. & Abraham, R. G. 2010, ApJS, 186, 427

Vega-Ferrero, J., Domínguez Sánchez, H., Bernardi, M., et al. 2021, MNRAS, 506, 1927