Identification of interacting galaxies in the S-PLUS database

G.B. Oliveira Schwarz[1,2], N.M. Cardoso[3], A. Cortesi[4,5], A.V. Smith Castelli[6,7], G. Lucatelli[2,8], J. Saponara[9], Y. Jaffé[10], L.A. Gutiérrez-Soto[2], C.R. Bom[4], A. Krabbe[11], J.A. Hernandez-Fernandez[11], F. Ferrari[12], A.L. Chies-Santos[13], H.M. Pereira[13], L. Sodré[2], C. Lima-Dias[14], E.V. Lima[2], F. Caro[15] & C. Mendes de Oliveira[2]

[1] *Universidade Presbiteriana Mackenzie, Brasil*
[2] *Instituto de Astronomia, Geofísica e Ciências Atmosféricas, USP, Brasil*
[3] *Escola Politécnica, USP, Brasil*
[4] *Centro Brasileiro de Pesquisas Físicas, Brasil*
[5] *Observatório do Valongo, UFRJ, Brasil*
[6] *Instituto de Astrofísica de La Plata, CONICET–UNLP, Argentina*
[7] *Facultad de Ciencias Astronómicas y Geofísicas, UNLP, Argentina*
[8] *Jodrell Bank Centre for Astrophysics, UM, Reino Unido*
[9] *Instituto Argentino de Radioastronomía, CONICET–CICPBA–UNLP, Argentina*
[10] *Instituto de Física y Astronomía, Universidad de Valparaíso, Chile*
[11] *UNIVAP, Brasil*
[12] *Instituto de Matemática Estatística e Física, FURG, Brasil*
[13] *Departamento de Astronomia, UFRGS, Brasil*
[14] *Departamento de Astronomía, ULS, Chile*
[15] *LERMA, Observatoire de Paris, Francia*

Contact / gustavo.b.schwarz@gmail.com

Resumen / Mostramos resultados preliminares relacionados a la identificación de galaxias interactuantes en un área de cielo de 1000 grados cuadrados utilizando la segunda liberación de datos de S-PLUS. Nuestro primer objetivo es encontrar galaxias interactuantes muy cercanas ($z < 0.019$) en el cúmulo de galaxias de Fornax y otros sistemas próximos tomando ventaja de la gran cobertura areal de S-PLUS mediante el uso de imágenes en la banda de Hα. Nuestro segundo objetivo es obtener una muestra lo suficientemente representativa de galaxias interactuantes con $r < 17.5$ mag, realizando diferentes tipos de búsquedas, para ser utilizada luego como muestra de entrenamiento supervisado para un modelo de *deep learning* que es capaz de identificar en forma automática nuevas galaxias de ese tipo en S-PLUS. Aquí presentamos, como resultado parcial del proyecto, una muestra de galaxias interactuantes encontradas por dicho modelo en la DR2 de S-PLUS. La precisión del modelo en la muestra de testeo es de $\approx 91\,\%$.

Abstract / We show preliminary results of searches for interacting galaxies over an area of about 1000 square degrees in the southern hemisphere, using the second data release of the Southern Photometric Local Universe Survey. Our first goal is to find very nearby interacting galaxies ($z < 0.019$) with a high fraction of gas in the Fornax cluster and other nearby systems, taking advantage of the large area covered by S-PLUS with images in the Hα band. Our second goal is to obtain a sufficiently representative dataset of interacting galaxies with $r < 17.5$ mag, using different methods, to feed a supervised training of a deep learning model that is capable of finding new interacting galaxies in S-PLUS automatically. Here we show a sample of interacting galaxies found by the deep learning model under development as a partial result of the project. The accuracy of the model in the test set is $\approx 91\%$.

Keywords / surveys — methods: observational — galaxies: interactions

1. Introduction

Interacting galaxies have been used to study galaxy evolution (e.g. merger rate as a function of redshift) and formation (e.g. birth of new objects in galaxy-galaxy collisions). Largely due to the Sloan Digital Sky Survey, SDSS (Blanton et al., 2017), there are a number of interacting galaxy catalogs - and studies that explore those - in the northern sky, but the southern sky lacks similar catalogs and studies. The Southern Photometric Local Universe Survey (S-PLUS) may change this situation, since it is mapping 9000 square degrees of the southern sky with an 80-cm robotic telescope located at Tololo, Chile, in 7-narrow and 5-broad bands, to a depth of 21.5 AB in the SDSS-like filters (Mendes de Oliveira et al., 2019). When complete, S-PLUS will be an excellent database for searches of rare objects in the nearby universe, in particular of interacting galaxies.

This contribution describes searches of interacting galaxies using S-PLUS catalogs and images. Collisions of two similar-mass galaxies, called mergers, that produce beautiful astronomical images, are known to be rare in the local universe but they are widespread at high redshifts. More frequent are the so-called accre-

Poster contribution

tion events, with small objects falling onto larger ones, within a short merging timescale. As for other collisional interactions, perhaps the most amazing objects besides the mergers themselves are the so called tidal-dwarf galaxies, created in the merging process. Those are observed, for example, at the tips of the tails of the Antennae galaxies. As for hydrodynamical interactions, some of the most spectacular cases of the so-called jellyfish galaxies, which are extreme cases of ram-pressure stripped galaxies entering a cluster environment, are also fairly rare and one can certainly benefit from the large coverage of S-PLUS and its unique Hα filter to enhance the numbers of known highly disrupted jellyfish galaxies in nearby clusters. In this work, we are after these and other kinds of interacting objects, using a number of more traditional techniques (e.g. color-color diagrams and non-parametric measurements), but we also use deep learning techniques to automate the search for interacting galaxies. The novelty is that we take the output sample of traditional methods as input for the training and validation of a deep learning model with the goal of searching for a more complete sample of interacting galaxies.

2. Sample and Data Access

S-PLUS is mapping nearly half of the southern celestial sphere in 7-narrow and 5-broad bands. When finished, it will have accumulated a total of 500 Terabytes of images and catalogs of sources. A public database was created for handling S-PLUS data[*]. It uses tools for the distribution of catalogs that follows the IVOA standards. It also contains astronomical images in FITS format and it allows the creation of color images in real time. Data can also be accessed through a Python package called SPLUSDATA[**].

3. Analysis and Results

The techniques to find interacting galaxies used in this work are the following: (1) search in the database for the strongest emission-line galaxies using the Hα-line contrast and/or with color-color diagrams. (2) search for objects with high asymmetry and low concentration, and (3) search using Deep Learning (DL) methods. These three methods are described in the following sections.

3.1. Search for nearby z < 0.019 interacting galaxies

Using the method of the Hα contrast over the r and i bands, we first performed a search for interacting galaxies in the Fornax cluster and other nearby systems where we expect to have the lines fall within the corresponding filters. We searched for fairly large galaxies (Class_star_r[***] < 0.1 and FWHM > 10 pix), with magnitudes less than r=17.5 and $\mathrm{F660}_{aper} - (r+i)/2 > 0.4$ mag, where r, i and F660 are the galaxy total magnitudes in the r, i and F660 (Hα) bands.

We also used colour-colour diagrams in order to detect Hα-emitting objects, see Wevers et al. (2017) and Gutierrez-Soto (in prep.) for details. These yielded a sample of nearby strongly-emitting galaxies (with $z < 0.019$), many of which were interacting and merging. Some examples of the objects found are shown in Fig. 1.

The result of this search was used to feed a Zooniverse experiment[****] that will also involve the public, called "The Green Dots project", given the green color of the typical Hα-emitting blobs, soon to be released.

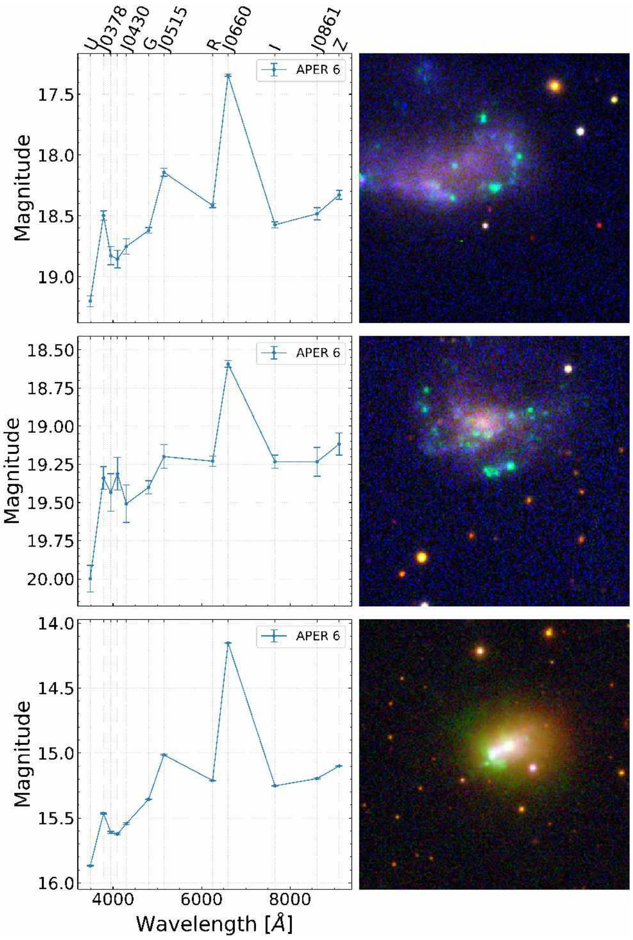

Figure 1: Interacting objects with strong Hα in emission. *Left*: Spectral energy distribution of each source. *Right*: The S-PLUS RGB images using the 12 filters. The green colour of parts of the image indicates Hα in emission. For legibility, the labels of the filters J0395 and J0410, between J0378 and J0430, were not indicated in the figure.

3.2. Search for interacting pairs at all redshifts

A non-parametric fitting of the galaxies with $r < 17.5$ was performed by implementing the program Morfom-

[*]https://splus.cloud

[**]https://pypi.org/project/splusdata

[***]Class_star is an output of SExtractor, software for source extraction (Bertin & Arnouts, 1996) - the low value used here indicates a high chance of the object being a galaxy.

[****]https://www.zooniverse.org/

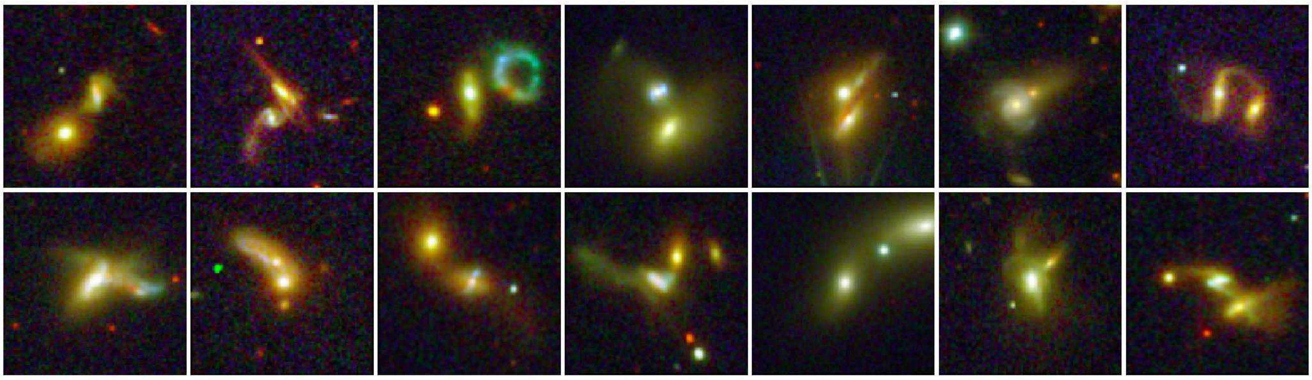

Figure 2: Sample of interacting galaxies in S-PLUS found in DR2 with DL techniques.

etryka (Ferrari et al., 2015). We obtained asymmetry and concentration indices and we restricted the sample to objects with asymmetries larger than 0.2 and concentrations less than 0.7 to select objects that had a high chance to be interacting or merging. This criterion was obtained from studying where a known sample of interacting and merging galaxies lies in the asymmetry-concentration diagram. This sample was then supplemented by a sample of merging galaxies from the Galaxy Zoo sample (Lintott et al., 2008), using only objects with very high probability of being a merger ($P_{merger} > 0.95$).The result was a final sample of about 500 bonafide interacting and merging pairs. We used this list as a training set for DL-based techniques to search for new objects in the whole DR2.

3.3. Deep Learning methods used to find new interacting pairs

The 500 objects that passed our visual inspection (mentioned above) and a sample of about 1500 spiral and elliptical galaxies from the work of Bom et al. (2021) were used for validation and test of the DL method. The sample of Bom et al. (2021) encompasses data that have already been extensively visually inspected, to verify that it contains bonafide elliptical and spiral galaxies. The network input data are RGB images of size 128×128 pixels (70.4×70.4 arcsec) from the SPLUS DR2. The dataset split was 80% for training and 20% for the validation of the network. The data preprocessing was done by mapping images from the 12 bands into a 3-band RGB image, using the trilogy package[†].

As a very first experiment, we used the VGG-16 convolutional neural network - a high performing neural network for computer vision validated in several computer visions challenge datasets, less prone to overfit since it uses fewer parameters than other similar models - with weights from a previously trained model from ImageNet as an image feature extractor and using dense layers, all available in the TensorFlow python package (Abadi et al., 2016). This simple experiment, with a network that the group had already used previously, will be supplemented in the near future: we are working on new convolutional neural network architectures, as well as joining the best architectures into a single classifier. With the configuration described above, the VGG-16 model reached 91% of prediction accuracy in the test sample. Figure 2 shows a number of objects found when passing the already trained neural network in the whole SPLUS DR2, which contains 111625 objects with high probability of being galaxies. We found that about 1.1% (1287 objects) may be in interacting systems, according to the deep learning method used here. We are in the process of improving the purity of the output sample of interacting galaxies, given that our visual inspection of the new objects revealed several cases of star-galaxy superposition (not bona-fide interacting galaxies).

[†]`https://www.stsci.edu/~dcoe/trilogy/Intro.html`

4. Conclusions

We used traditional and DL-based techniques to search for interacting galaxies in the S-PLUS DR2 dataset. The traditional search generated a list of very nearby interacting systems, with some interesting cases of jellyfish galaxies. On the other hand, the DL-based code generated a list of mostly interacting pairs, in the premerger and interacting phases, over the whole DR2. These are interesting systems for future further studies.

Acknowledgements: We thank the referee for her/his report that helped to improve the content of this contribution. This work was funded with grants from FAPESP, CNPq, Fondecyt, ANID, CONICET, Agencia I+D+i and Universidad Nacional de La Plata.

References

Abadi M., et al., 2016, *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, 265–283, OSDI'16, USENIX Association, USA
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Blanton M.R., et al., 2017, ApJ, 154, 28
Bom C.R., et al., 2021, MNRAS, 507, 1937
Ferrari F., et al., 2015, ApJ, 814, 55
Lintott C.J., et al., 2008, MNRAS, 389, 1179
Mendes de Oliveira C., et al., 2019, MNRAS, 489, 241
Wevers T., et al., 2017, MNRAS, 466, 163