Proposal: Towards a Statistical Understanding of Galaxy Evolution
Potential Fellow: David O'Ryan
Potential Mentor(s): Bruno Alteiri & Bruno Merin

Research Rationale

Galaxy Morphology, Photometric Parameterisation and Galaxy Evolution

Galaxy morphology is a key indicator of galactic evolutionary history. Whether a galaxy is elliptical or a disk, is barred or unbarred, is spiral or elliptical, we can immediately form some base assumptions about it. As some examples, spiral galaxies are often gas rich and star forming when compared to often gas poor and quiescent elliptical galaxies (Smethurst et al., 2022), evidence is growing that strongly barred galaxies are more likely to have active nuclear regions than their weakly barred or non-barred counterparts (Garland et al., 2023) and disk-dominated galaxies are likely to have a merger-free history (Simmons et al., 2017). To fully understand the link between morphology and a ~~galaxies~~ evolutionary history we need large catalogues of morphologically classified galaxies. A current example of these catalogues are created by volunteers and citizen scientists visually inspecting galaxy images in the Galaxy Zoo collaboration Lintott et al. (2008). However, the time constraints of visually inspecting sources means the size of such catalogues is often limited. Once we start to control these catalogues for various morphologies, we often find that our samples cannot hold statistical significance.

A recent development in morphology classification has been the accuracy and efficiency of a new model: `Zoobot` Walmsley et al. (2023a). It is a convolutional neural network (CNN) trained to answer all thirty four classification tasks in the Galaxy Zoo workflow. It has had significant success in a number of works, from finding interacting galaxies in the *Hubble Space Telescope* (HST) archives O'Ryan et al. (2023), to successfully matching the volunteer responses to 314,000 galaxies in the Galaxy Zoo DECaLS project Walmsley et al. (2022), to successfully classifying the morphology of 8.7 million galaxies in the Galaxy Zoo: DESI project Walmsley et al. (2023b). Thus, providing large sub-samples of different galaxy morphologies and pioneering insights into the interplay between morphology and the underlying physical processes driving galaxy evolution.

Another set of classifications important to galaxy evolution is photometric parameterisation. These are parameters measured from a galaxy's emission lines using spectral energy distribution (SED) fitting. The resulting catalogues provide the stellar masses, star formation rates, nuclear activity and a host of other parameters. These SED matches are conducted with well used astrophysical softwares such as the Easy and Accurate zPhot from Yale (EAzY; Brammer et al. (2008)) and the Fitting and Assessment of Synthetic Templates (FAST; Kriek et al. (2009)). Many examples of such ancillary catalogues have been produced, a prime example being the Cosmic Evolution Survey (COSMOS; Capak et al., 2007) catalogue. These contain the parameterisations of over 1.8 million galaxies focused on a $3^\circ \times 3^\circ$ area of the sky. This catalogue has seen significant use by the community, with the initial data release having over 650 citations.

However, the size of catalogues which overlap morphological and ancillary data remains limited. Therefore, a key opportunity exists to enhance our understanding of the interplay between observable galactic morphology, their parameterisations and specific evolutionary processes. I will create a combined, statistically significant catalogue completely overlapping ancillary and morphological data which will be the largest to date. I will then conduct initial analysis to demonstrate its applicability. This will be achieved by combining the tried and tested CNN `Zoobot` with ESA Datalabs and utilising the incoming survey data of Euclid.

Impact & Past Achievements

Galaxy Evolution Over the Last 8 Billion Years

Throughout my PhD, my research has been focused on the effects of interaction and merging on galaxy evolution. The specific influence of galaxy interaction, when two galaxies come close enough to have a gravitational effect upon one another, are poorly understood. In general, interaction leads to a sudden increase in star formation (Moreno et al., 2021) often causing quenching (Das et al., 2023), severe morphological disturbance (Toomre & Toomre, 1972) and is often linked to the ignition of nuclear activity (Comerford et al., 2015). To explore the variations in these effects, and quantify their impact on evolution, we require large catalogues of interacting and merging galaxies to fully explore the underlying parameter space.

Creating these catalogues by visual inspection is a notoriously difficult task. Numerous works have shown that classifications from citizen scientists (Darg et al., 2010) or machine learning (Pearson et al., 2022) techniques often lead to significant levels of contamination by close pairs. I.e., contamination from galaxies which appear close together in the two dimensional plane of the sky but are at very different redshifts. I have ~~worked on~~ numerous projects ~~which are attempting~~ to overcome this difficulty (Margalef-Bentabol et al., sub; Martin et al., 2022) but my primary contribution was using `Zoobot` to build a pure catalogue. In O'Ryan et al. (2023), I used the `Zoobot` with ESA Datalabs to explore the HST archives for interacting galaxies. A total of 92 million sources were classified into interacting and non-interacting, creating a catalogue of 21,926 interacting galaxies. The largest to date. This was the first published work where `Zoobot` was used to such a scale on a specific classification problem. This initial use of `Zoobot` was described in the official release paper of the model (Walmsley et al., 2023a).

An unintended result of using `Zoobot` in this way with ESA Datalabs was that a host of other objects of astrophysical interest were discovered in the archives. Because interacting galaxies are often morphologically disturbed and highly irregular, `Zoobot` misclassified many objects of morphological or astrophysical interest as interacting galaxies. As a byproduct of my approach, I therefore created some of the largest catalogues of various astrophysical objects to date. Examples of these extra objects include ring galaxies, overlapping galaxies, quasars and edge-on proto-planetary disks. The latter objects we only have three known examples of. This approach, and internship, gave me a unique insight into how to use ESA Datalabs to efficiently provides me with a large repository of existing pipelines to stream line the ramp up of this project.

Now that a large catalogue of interacting galaxies has been created, it is possible to explore their underlying parameters and attempt to make inferences into their interplay with morphology and parameterisations. One way I am investigating this is by using simulations. I have developed software to directly compare observations and simulations. This code mimics the methodology of Holincheck et al. (2016), a successful project constraining the stellar masses, orientations and sizes of interacting galaxies with citizen scientists.

Another approach I am adopting is to use ancillary data from different catalogues to find the underlying parameters of interaction. I have matched the O'Ryan et al. (2023) catalogue to the COSMOS2020 (Weaver et al., 2022) catalogue and created a sample of 4,000 interacting galaxies (all galaxies which overlap between the catalogues) with parameterisation data readily available. This work remains in progress, but the relation between galaxy interaction, environment, nuclear activity, star formation enhancement and projected separation is being investigated. This work, which will be published before I begin this fellowship, and the pipelines being built will form the basis of the analysis work of this project.

		Year 1				Year 2				Year 3			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
WP1	T1.1: Building Pipeline for Cutout Creation and Running				D1: Repository of 3-Colour Cutouts for Entire JWST and Hubble Public Archives								
	T1.2: Testing and Creating Diagnostics for Zoobot Model												
	T1.3: Running Classification and Analysis of Morphology Classifications					R1: Release of Morphology Catalogue							
WP2	T2.1: Creating Pipeline for Matching Euclid Data												
	T2.2 Creating Pipeline for Parameterising Euclid Data and Running												
	T2.3 Analysis of Euclid Ancillary Data and Diagnostics							R2: Release of Full Ancillary Catalogue					
WP3	T3.1: Matching Morphology and Ancillary Catalogues												
	T3.2: Diagnostics and Initial Analysis of Advanced Catalogue								R3: Release of Combined Catalogue and Initial Analysis				
WP4	T4.1: Selecting Interacting Galaxies and Diagnostics									D2: A Spectroscopically Confirmed Catalogue of Interacting Galaxies			
	T4.2: Identifying AGN in Sample												
	T4.3: Morphological and Parametric Analysis												R4: Analysis of Interacting Galaxy Catalogue and Exploration of Parameter Space

Figure 1: GANTT chart of proposed project timeline. Includes workpackes (WPN), scientific results (RN) and deliverables (DN).

Research Plans for Fellowship

This project has six key goals which will be achieved with four scientific results (RN) and two deliverables (DN). I have defined four **work packages** (WPs) that follow the full process of this project across eleven different tasks. For a full breakdown of the expected timeline, specific tasks and their WPs please see Figure 1. The project goals are:

1. To create a fully open, and accessible image repository of every extended source at rest frame in the *Hubble* and JWST public archives (T1.1; D1).
2. To apply `Zoobot` to the above created repository and gain morphological classifications of an unprecedented number of galaxies (T1.2, T1.3; R1).
3. To utilise Euclid data to create a catalogue of ancillary data to the scale of the morphological catalogue. The size of this catalogue will also be unprecedented (T2.1, T2.2, T2.3; R2).
4. To match these catalogues and create an advanced catalogue, while fully diagnosing their accuracy and completeness (T3.1, T3.2; R3).
5. To create the largest catalogue of spectroscopically confirmed interacting galaxies to date (T4.1; D2).
6. To analyse the advanced catalogues in relation to the interplay between galaxy interaction and the ignition of nuclear activity, demonstrating their use (T4.2, T4.3; R4).

Synergies with ESA Missions

~~As stated throughout this proposal,~~ this project is directly related to multiple ESA missions. The two primary missions it will take advantage of is the development of the archives with ESA Datalabs and the new flagship mission Euclid. The archives, and the new infrastructure built around it, form a key cornerstone of this project. Without ESA Datalabs and the unique kind of access to archival data it provides, this project would not be possible. The second cornerstone of this project is Euclid. To date, no such all-sky spectroscopic survey to the scale of Euclid exists. Combining archival data with Euclids spectroscopy will lead to an advanced catalogue of a size yet seen in the field.

This project will also use the ESA-NASA mission JWST. The volume of available JWST archival data is now impossible to ignore and not utilise. This archival data is specifically available on ESA datalabs, and can be accessed by simple adaption of our pipelines, as they exist. The ancillary data JWST provides, as well as its unique spectroscopy and imaging of high redshift systems will allow us to create a more complete catalogue which accounts for this regime.

My previous experience with ESA Datalabs, `Zoobot` and utilising the ancillary data of the COSMOS catalogue put me in a unique position to create and analyse these new advanced catalogues. While building the pipelines to create cutouts, or conduct classifications, would certainly take time to build the fact that I have much of these pipelines ready for use will significantly streamline the ramp up of the project. From this project, I also hope to fully demonstrate the power of Euclid combined with that of ESA Datalabs and the archives. Giving a new dataset to the field which is unprecedented in scope.

References

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Capak P., et al., 2007, ApJS, 172, 99

Comerford J. M., Pooley D., Barrows R. S., Greene J. E., Zakamska N. L., Madejski G. M., Cooper M. C., 2015, ApJ, 806, 219

Darg D. W., et al., 2010, MNRAS, 401, 1552

Das A., Pandey B., Sarkar S., 2023, Research in Astronomy and Astrophysics, 23, 025016

Garland I., et al., 2023, MNRAS, submitted

Holincheck A. J., et al., 2016, MNRAS, 459, 720

Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, ApJ, 700, 221

Lintott C. J., et al., 2008, MNRAS, 389, 1179

Margalef-Bentabol B., et al., sub, A&A, submitted

Martin G., et al., 2022, MNRAS, 513, 1459

Moreno J., et al., 2021, MNRAS, 503, 3113

O'Ryan D., et al., 2023, ApJ, 948, 40

Pearson W. J., et al., 2022, A&A, 661, A52

Simmons B. D., Smethurst R. J., Lintott C., 2017, MNRAS, 470, 1559

Smethurst R. J., et al., 2022, MNRAS, 510, 4126

Toomre A., Toomre J., 1972, ApJ, 178, 623

Walmsley M., et al., 2022, MNRAS, 509, 3966

Walmsley M., et al., 2023a, The Journal of Open Source Software, 8, 5312

Walmsley M., et al., 2023b, MNRAS, submitted

Weaver J. R., et al., 2022, ApJS, 258, 11