The Cosmic Evolution of Galaxy Structure and Morphology at $0.5 < z < 8$

Leonardo de Albernaz Ferreira

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September 2022

I dropped down again
From a star
On a desert island
Full of skies
And I saw a boy
Looking up
Dreaming of his future
From my past

Lost-Riverside

* This piece was generated from an artificial neural network especially for this thesis

Supervisors: Prof. Dr Christopher Conselice
Dr. Simon Dye

Examiners: Prof. Chris Lintott (University of Oxford)
Prof. Frazer Pearce (University of Nottingham)

Examination chair:

Submitted: 30th September 2022
Examined: 7th December 2022
Final version: 15th December 2023

Contents

List of Tables

List of Figures

Abstract

This thesis is prepared in two parts. In the first half (Chapter 2 and Chapter 3) we discuss the evolution of galaxy mergers at $0.5 < z < 3.0$ in all the CANDELS fields based on a supervised deep learning model trained on the IllustrisTNG cosmological simulations. The second half is dedicated to the rest-frame optical morphological evolution of galaxies from $z = 1.5$ to 8 as observed by JWST in the SMACS 0723 field, and in the early observations of the CEERS program.

In Chapter 2 we describe a supervised deep learning framework designed for the classification of high redshift galaxy mergers based on data from the IllustrisTNG cosmological simulations. We generate a large dataset of before mergers/post-mergers/non-mergers galaxy mocks labeled with information from IllustrisTNG 300-1 merger trees. These imaging data are prepared to be CANDELS-like and are then used to train deep learning models capable of achieving 90% of accuracy within the simulations. Using these them we describe the evolution of the galaxy merger fractions and rates in the CANDELS fields and we discuss how these deep learning classifications are related to visual classifications. We report the first agreement of galaxy merger rates between galaxy pair statistics methods and morphologically selected mergers, with $\mathcal{R}(z) = 0.02 \pm 0.004 \times (1 + z)^{2.76\pm0.21}$, showing that the highest merger rates are found at the highest redshifts.

We tackle the challenging problem of separating recently coalesced galaxy mergers from non-interacting highly star forming galaxies in Chapter 3. These two populations present ambiguous morphologies due to asymmetric features. We refine our methods reported in Chapter 2 for this particular question, generating a dataset of TNG100-1 post-mergers and star forming galaxies at $0.5 < z < 3.0$, including a full radiative transfer treatment with the SKIRT code, producing $\sim$ 160,000 images with realistic morphologies. We explore the relative populations of post-mergers and non interacting star forming galaxies in this redshift range. We show that the population of high redshift asymmetric galaxies are more likely to be of post-merger origin than their low redshift counterparts. The interpretability of our models is discussed by exploring the feature space extracted from the mock imaging and the real CANDELS galaxies. We show that for this particular problem, deep learning models provide an 30% improvement over quantitative morphology methods.

We focus on the early release JWST observations of the SMACS 0723 cluster in Chapter 4. We report the first ever morphological study of rest-frame optical structure in $1.5 < z < 6$ with NIRCam, within the wavelength range $\lambda = 0.9\mu m$ - $4.4\mu m$. We conduct visual classifications and quantitative morphology measurements on a sample of $\sim$ 200 galaxies previously detected with HST. We report a surprising mismatch between the number of disk galaxies detected with HST and JWST. Around ten times more disks are found. We briefly discuss the implications of this result and how it fits in the galaxy formation and evolution evolution

picture.

Over Chapter 5 we expand the framework of Chapter 4 to the early CEERS JWST observations that have overlap with the EGS observations from the CANDELS fields with HST. We release to the community the biggest sample of visually classified galaxies observed with JWST to-date, with 4265 galaxies that are both observed by HST and JWST. With this dataset, we carefully discuss the evolution of the Hubble sequence up to $z \sim 8$, finding that it is already present at the earliest of times for low to intermediate mass galaxies, while evolution driven by mergers is observed for massive galaxies. We detail the quantitative morphology characteristics of this sample, and how it correlates with visual optical morphology.

We finish with a brief discussion on the results presented in this thesis, how the merger evolution at $0.5 < z < 3.0$ and the general morphological evolution at $z > 3.0$ are linked, and what are the next steps to explore this connection further.

Acknowledgements

Many are the people that I feel indebted to, scientifically or that cared to spare some time to help me throughout my journey. Mentors, family, and friends, this thesis would not be possible without you. The minimum I can offer in return is these words.

My mentor Chris Conselice, this thesis and this work would not be possible without him. I cannot stress enough his importance and the central role he played for the opportunities that were presented to me since the start of my PhD. He was always dedicated and patient, even when at the start I was barely able to formulate a full sentence in English. He trusted me enough to give me independence to let me pursue my own projects and subjects, and was always there when I most needed guidance. He opened many doors to me and I will be forever grateful, as my life will never be the same from now on. During the pandemic, he made sure I never felt lost, and that I could always count on him.

I also am immensely beholden by Ulli Kuchner and her kindness, she was there throughout all my journey, since my first research meeting. She always went out of her way to help me when possible, offering unique points of view, stimulating different perspectives. She was a catalyst for my integration within the department, and every time I felt a bit isolated she was there to motivate me and to help me break some social barriers, both due to language and due the cultural shift.

Fabricio Ferrari who put me in this neverending scientific voyage, former supervisor and scientific father, now friend. Even with the distance, he was always keen to help. A quick chat with him will always warrant you some valuable life lessons. A significant part of my work is building up on what he started with Morfometryka.

The larger Conselice group both in Nottingham and in Manchester, all amazing people I was able to collaborate and become friends with over the years: Alex, Amy, CB, Sunny, Florian, Laura, Nathan, James, Duncan, Rachana, Sakina.

The FLAGS team and the wide PEARLS collaboration, the last couple of months working alongside you and JWST data are among the best of my PhD.

All CAPT, professors, post-docs, staff. You made me feel home since day one, with a unique environment like no other to study galaxies. I felt privileged to be part of this community for these four years. Special thanks to Amelia Fraser-McKelvie, she was always encouraging and helped a lot when I was having panic attacks before hosting seminars or giving talks during my first year. Phil Parry for being an IT wizard and making everything run smoothly, even when I had more than 40 TB of data scattered across all CAPT servers.

In 4 years I made so many good friendships in Nottingham, there is too many memorable moments to list, from weird Halloween parties to being yeeted from the Peak district by park rangers for wild camping. These are among the best memories of my life. The COTD crew, CB, Mick, Liza, Jacob, Roan, and Sukhi, I could not ask for better comrades. CAPT put a lot of friends in my life, stay assured that I have a special memory with each one of you: Swagat, Tomáš, Lizzie T, Lizzie E, Matt, Simon, Kellie, Karel, Steph, Dan, Tom C, Tom P, Finlay, Charutha, mi hermano Agustín, Brad.

I enjoyed my journey very much, but the past four years were not a particularly good time to do a PhD. I had many moments of despair and sorrow, especially during the pandemic. Being an international student is already hard enough, but during these times, there were special challenges. And hereby I must thank my family and friends for enduring this harsh period with me, even if we were 10,000 km apart. For the past 48 months I was only able to visit my family once, and had to cut my visit short due to airport closures. I said goodbye to my grandma at the bus stop back in 2018 to never see her again. I never said farewell to my granddad. The feeling of never knowing when you will be able to see your family again is hard, and I'm sure it hit my mom and my father the hardest. They endured it, I endured it. And we keep on walking hopeful that we will be reunited soon. Mom and Dad, everything I do is for you, I miss you so much.

I have many people to thank for keeping my mental health stable during these years, and you all know who you are, but special thanks goes to my constant, Jessica, for being there for me during all the happy times and rock bottoms. You are my point of equilibrium. All our adventures around Nottingham, enduring lockdown together, or simply binge watching weird TV shows and keeping each others company. Thanks for sharing part of your life with me. You are the love of my life.

Published works

Almost all of the content in this thesis has already been published or submitted for publication in the following four works:

I **Ferreira, L**, Conselice C. J., Duncan K., Cheng T., Griffiths, A., Whitney, A, 2020. *Galaxy Merger Rates up to $z \sim 3$ Using a Bayesian Deep Learning Model: A Major-merger Classifier Using IllustrisTNG Simulation Data.* The Astrophysical Journal, 895, 2, 115

II **Ferreira, L.**, Conselice, C. J., Kuchner U., Tohill C. B., 2022 A Simulation-driven Deep Learning Approach for Separating Mergers and Star-forming Galaxies: The Formation Histories of Clumpy Galaxies in All of the CANDELS Fields. The Astrophysical Journal, 931, 1, 34

III **Ferreira L.**, Adams N., Conselice C. J., Sazonova E., Austin D., Caruana J., Ferrari F., Verma A., Trussler J., Broadhurst T, Diego J, Frye B. L., Pascale M., Wilkins S. M., Windhorst R. A., Zitrin A., 2022. *Panic! At the Disks: First Rest-frame Optical Observations of Galaxy Structure at $z > 3$ with JWST in the SMACS 0723 Field.* The Astrophysical Journal Letters. *Aceppted.*

IV **Ferreira L.**, Conselice C. J., Ferrari F., Sazonova E., Tohill CB., Lucatelli G., Caruana J., Adams N., Irodotou D., Marshall A. M., Trussler J., Duncan A., Verma A., Wilkins S. M., Roper W., Lovell, C., 2022. *The JWST Hubble Sequence: The Rest-Frame Optical Evolution of Galaxy Structure in CEERSx at $1.5 < z < 8$.* The Astrophysical Journal. *Submitted.*

Chapter 2 is based on material published in Paper I, Chapter 3 is based on Paper II, and Chapter 4 is based on Paper III, and Chapter 5 is based on Paper II.

The vast majority of the work presented in this thesis was carried out by the author, with supplementary advice from various co-authors listed above. In any instances where the work includes the product of larger collaborations, this is explicitly mentioned in the relevant Chapter.

Chapter 1

Introduction

Galaxies and their contents – stars, gas, dust, black holes, and dark matter – are the fundamental building blocks of our Universe. Understanding the interplay between galaxies and their environment, how their contents change over time, how their mass is assembled, and the key physical processes behind these changes is fundamental for a correct description of galaxy formation and evolution, and for a description of cosmology. No two galaxies are equal, and they present themselves with rich forms and properties. Naturally, the first step towards this understanding is establishing a taxonomy system for their morphological appearance, much like what is common practice for biological systems and other systems in nature. In fact, one of the backbones of our modern view on galaxies is their morphological classification, visual or quantitative, as to group them by similar features.

In this thesis, I will explore new ways to quantitatively classify galaxy mergers at high redshifts, a type of system that plays a central role in the morphological transformation of galaxies, as well as discuss for the first time the optical rest-frame morphology of $z > 3$ galaxies as observed by the brand new James Webb Space Telescope (JWST).

1.1 Galaxy Morphology

Over the last century, astronomical observations went from uncovering the extragalactic nature of galaxies (Opik, 1922; Hubble, 1929b), known before as 'nebulae', to dozens of facilities scattered across the globe and in space capable of observing millions of galaxies, and soon billions (e.g., SDSS, DES, Rubin, EUCLID, York et al., 2000; The Dark Energy Survey Collaboration, 2005; LSST Science Collaboration et al., 2009; Laureijs et al., 2011; Euclid Collaboration et al., 2022a,b) covering all the electromagnetic spectrum, up to potentially the very first galaxies ever formed (Adams et al., 2022; Atek et al., 2022; Donnan et al., 2022; Castellano et al., 2022; Yan et al., 2022; Naidu et al., 2022). This abundant landscape of astronomical data is very unlike the one available to Hubble when he designed his morphological classification scheme (Hubble, 1926, 1929a, 1936, Figure 1.1). However, his framework continues to be widely adopted today, as it relies on the apparent dominance of some morphological structures which galaxies can display: featureless oblate spheroids (ellipticals, E), feature-rich disks (spiral, S), and irregulars (Irr). Each of these categories can be subdivided into smaller bins, ellipticals based on their roundness and spirals based on emergent features from the disk. These are broad categories that encompass the majority of the galaxies observed in the nearby universe, in fact $\sim 97\%$ of the Shapley-Ames catalog (Sandage & Tammann, 1981) of galaxies are fully accounted for in E or S categories (Sandage, 2005). More recent classification schemes built up on the Hubble Sequence to refine these bins or to fit new discovered types (van den Bergh, 1960, 1976), as it is the case of lenticulars galaxies (S0s, Kormendy & Bender, 2012).

These morphological models encode the dynamics of the stellar orbits, but are not directly related to their stellar masses and other physical properties regard-

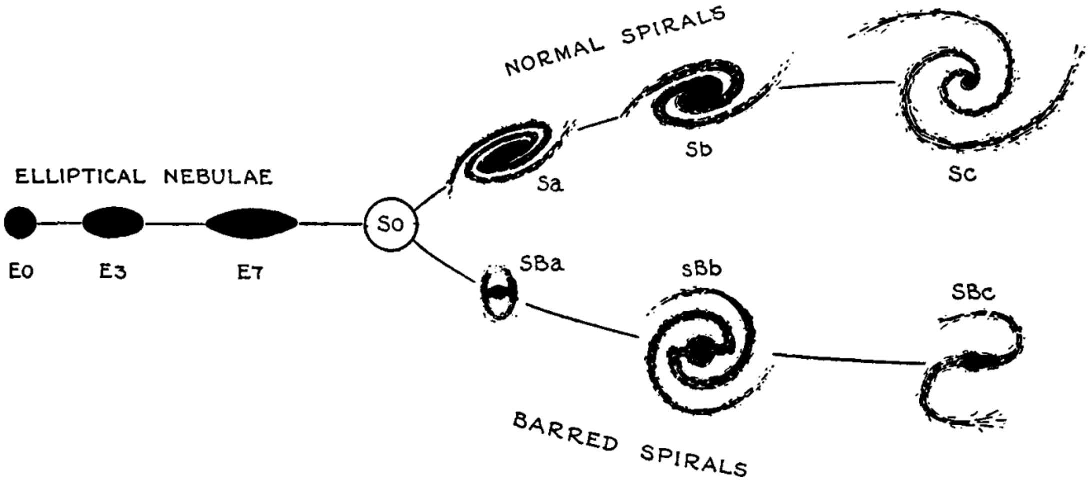


Figure 1.1: The Hubble Sequence as described in Hubble (1936) showing the main morphological groups, as well as the hypothetical (at the time) S0 morphology class between the elliptical and spiral galaxies. The spirals are subdivided in unbarred and barred spirals.

ing the formation and evolution of galaxies, they are at at top level descriptive. Some argue that this is the correct way to establish a morphological classification system (Sandage, 2005), as including physical information could create circular arguments when interrogating the morphology against physical properties. The triumph of the Hubble sequence lies in its simplicity, and how general it is. At higher redshifts, as we will see throughout this thesis, fewer finer details are available compared to nearby galaxies, due to the lower angular resolution. Thus, even with spiral arms, bars, and dust lanes being more difficult to resolve, the simple separation of E/S/Irr is still powerful, or even simpler, the more agnostic terms of disc-dominated and bulge-dominated (Kartaltepe et al., 2015). Tracking how the demographic of these populations changes over cosmic time give us clues on how galaxies evolve and assemble (Conselice, 2014).

Since the establishment of the Hubble sequence, it was noted that morphology correlate with several physical properties of galaxies, and the dichotomy between ellipticals and spirals seems to be fundamental. Galaxies with different morphologies seem to be segregated in colours based on the color-magnitude diagram (de Vaucouleurs, 1961; Strateva et al., 2001; Hogg et al., 2003; Bell et al., 2004; Schawinski et al., 2014), where in general Es are red and not star-forming while Ss are blue and star-forming. Still, types that challenge these categories exist, but are usually rare (e.g., red spirals and blue ellipticals, Melvin et al., 2014; George, 2017). This is consequently associated with age, star formation rates, metallicity and dust contents. Spatially resolved spectroscopy studies of galaxies, such as ATLAS3D, CALIFA and MaNGA (Cappellari et al., 2011a; Sánchez et al., 2012; Bundy et al., 2015), show that morphology correlates with kinematics, as disk galaxies are supported by rotation while ellipticals are in dominated by tri-axial motion with high velocity dispersions (Cappellari et al., 2011b). Fundamental relations between the stellar mass content (or luminosity as a proxy), the overall morphologies and stellar kinematics also exists, such as the Tully-Fisher relation for disks (Tully et al., 1975), and the Faber-Jackson relation for ellipticals (Minkowski, 1962; Faber & Jackson, 1976).

Key to morphology is also the environmental dynamics on which galaxies inhabit. The demographics of the Hubble classes highly also vary depending on the local galaxy density, as in dense clusters and galaxy groups, elliptical morphologies and lenticulars are more common, while star forming spirals are rare. This is known as the Morphology-Density relation (Dressler, 1980; Houghton, 2015; Sazonova et al., 2020). Another piece of the puzzle, as the environment dynamics that galaxies experience have a strong effect on its morphology.

1.1.1 Galaxy Mergers

In early works on galaxy morphology, little attention was given to interacting galaxies. They compromised a tiny fraction ($\sim 1\%$) of the galaxy population in the local Universe, and hence were seem as unimportant. The Atlas of Peculiar Galaxies (Arp, 1966) was one of the first comprehensive catalog of Peculiar galaxies that included a large sample (at the time) of merging galaxies. Figure 1.2 shows an example of a merger as imaged by HST, the IC 1623 galaxy pair, and its fabulous chaotic features, including tidal tails, clumpy star forming regions and strong reddening from dust (Cortijo-Ferrero et al., 2017). Features like these motivated the seminal work of the Toomre brothers on the first computational simulations of pairs of interacting galaxies (Toomre & Toomre, 1972). These simulations showed how such features could be created by the gravitational interaction of galaxies, and their disruptive influence on galactic components. Also noted was that the merging of two disc galaxies could produce a spheroidal remnant. This scenario is now widely accepted as one of the ways galaxies can be transformed morphologically along the Hubble sequence, where similar mass discs merge together to form larger elliptical galaxies, the first physically motivated evolutionary link between classes in the literature. Accordingly, the study of this rare population was critically central to the development of the modern theory of how galaxies form and grow.

More recent studies of morphological features of galaxy mergers show that they are transient on timescales of several million years to a few billion years (Lotz et al., 2008; Conselice, 2009; Pawlik et al., 2016; Snyder et al., 2017; Whitney et al., 2021; McElroy et al., 2022). This poses a challenge for the identification of such systems, as some soon to be merging galaxies may not display any feature as it approaches its companion, or that recently merged system would be dynamically relaxed soon thereafter as to not contain any obvious merging signatures (Bottrell

Figure 1.2: IC 1623 interacting galaxy pair as observed by the Hubble Space Telescope. The chaotic display of forms is a clear deviation from Hubble's sequence Ellipticals and Spirals. The morphology is rich with star-forming regions, tidal features and dust. Image credit: ESA/Hubble & NASA, R. Chandar.

et al., 2019). Additionally, some of these features are faint and difficult to detect. Hence, in the age of the wide field deep sky surveys such as SDSS, DES and CANDELS, techniques to find mergers accurately are sought after, and an active area of research in astronomy (Conselice et al., 2003; Lotz et al., 2004; Lotz et al., 2006; Lotz et al., 2008; Pawlik et al., 2016; Ackermann et al., 2018; Pearson et al., 2019; Bickley et al., 2021).

The importance of mergers is not limited to the morphological transformation, however. Mergers are also believed to be one of the main sources by which galaxies grow their mass (Duncan et al., 2019), trigger AGN (Ellison et al., 2011) by fueling central super massive black holes, enhance star formation (Genzel et al., 2001; Lotz et al., 2004; Patton et al., 2020) by the addition of gas to a galaxy and by mixing its material, and trigger quenching, and thus lead to quenching, as the mechanisms above expel the star-forming gas or prevent it from forming stars (Hopkins et al., 2008; Ellison et al., 2022).

1.1.2 Quantitative Morphology

As a natural development from the visual characterization of morphology, more quantitative methods of morphological study were developed. These can be divided in two types: parametric and non-parametric. Parametric methods estimate the light distribution of sources adjusting it to an analytical form, one common choice is the Sérsic profile. Through aperture photometry, iso-photal ellipses are adjusted over the galaxy light profile and then fitted to a Sérsic law (Sérsic, 1963), of the form

$$I(R) = I_0 \exp\left[-b(n)\frac{R^{\frac{1}{n}}}{R_e} - 1\right], \tag{1.1}$$

where n is known as the Sérsic index, which defines the slope of the light distribution, while $b(n)$ is defined such that R_e is the galaxy effective radius (e.g. half-light radius). This law is a generalization of the de Vaucouleus law (de Vaucouleurs, 1948), where the index was fixed as $n = 4$, and was used to describe the light distribution of elliptical galaxies. The Sérsic law is then more general, and can be used to describe point sources, disks, ellipticals. Galaxies with multiple components can be modeled by co adding Sérsic profiles together. The steepness of the profile is a good indicator if the galaxy is dominated by a disk or a centrally concentrated bulge component (Kelvin et al., 2012), and thus correlates with morphology. This approach is robust when the source is well sampled (Häussler et al., 2007). However, parametric methods assume that the light of the source takes a particular shape, and can not account for fine structures such as bars, rings, and spiral arms without introducing elements of high complexity to the fitting process (e.g., Jiménez-Teja & Benítez, 2012), which in turn can hinder the real world application of such models.

On the other hand, non-parametric methods have no assumption of the underlying shape of the galaxy and are generally measured directly to the pixel values of the image. The Concentration, Asymmetry and Smoothness (CAS, Abraham et al., 1994; Bershady et al., 2000; Conselice et al., 2003) is one of these systems, where it captures how concentrated is the source, how asymmetric, or how clumpy it is through simple transformations of the original image. Another system is the Gini-M20 plane (G-M20, Abraham et al., 1994; Lotz et al., 2004), designed to measure how the light is distributed among the pixels of the source. These measurements are then used together to characterize morphology and as a classification tool. For example, the CAS system and G-M20 are extensively used to find galaxy mergers, as its measurements are sensitive to disturbances are the light profile of sources, or when there is two galaxies in the same image.

Other methods exist, and are designed for specific tasks on characterizing galaxy morphology, such as identifying how many components there is in a galaxy (**?**), how strong non-radial features are (Ferrari et al., 2015) or even if a galaxy is a merger remnant (Pawlik et al., 2016). Most of these (including CAS and G-M20) are widely available in astronomical software such as MORFOMETRYKA (Ferrari et al., 2015), STATMORPH (Rodriguez-Gomez et al., 2019), GALFIT (Peng et al., 2010).

These methods, however, can be strongly affected by contamination, as galaxies with different physical properties can present similar morphologies, which can cluster them together in the parameter space of these measurements, producing regions that are highly ambiguous among classes. Additionally, observational effects, such as seeing and depth, can shift and bias these indices, making it hard to produce selections that are general for all instruments (Ferreira & Ferrari, 2018; Snyder et al., 2015).

1.2 Morphological Classifications

The previous section discussed the morphology of galaxies in general terms, how it is related to galaxies' physical evolution and ways to quantify morphology. However, little was said about *how* galaxies are actually classified. I dedicate this section for the actual act of classification, the main theme running behind this thesis.

When only photographic plates were available, or astronomers could only observe galaxies over the eyepiece of their telescopes, visual classification was all that existed. This is the main reason the first morphological studies were so focused in what the eye could easily distinguish, like the spiral arms emerging from the disk of a galaxy, or other particular features such as rings, bars, dust lanes. Albeit this being the classic way a galaxy image is classified, it is still one of the most robust and reliable ways to identify patterns on astronomical images. Astronomers continue to classify data-sets by eye and to produce highly detailed morphological samples, such as FREI (Frei et al., 1996), EFIGI (Baillard et al., 2011; de Lapparent et al., 2011), the Nair & Abraham sample (Nair & Abraham, 2010), and the CANDELS visual classifications (Kartaltepe et al., 2015), to name a few. These however, are only sub-samples of the overall available data. With the advent of the wide-field surveys, such as SDSS and DES, or the deep NIR surveys, such as CANDELS (Grogin et al., 2011; Koekemoer et al., 2011), the number count of galaxies detected and observed far surpasses the capabilities of any individual astronomer, or groups of astronomers. To classify thousands to millions of galaxies by eye is a huge undertaking, as it is generally necessary for sources to be identified by more than one classifier, as to mitigate subjective effects on classifications.

One very successful take on visual classifications enabled within the World Wide Web is the GalaxyZoo citizen science project (Lintott et al., 2011), where a large number of non-expert volunteers have access to a tool that instructs them how to classify galaxies visually, through a guided approach based on simple questions. The continuous engagement of citizens on this process trains them to become efficient classifiers rivalling to any expert on galaxy morphology (Marshall et al., 2015). Subjectivity is still at play, but the statistical classification of sources (i.e. many volunteers classify the same galaxy) and careful treatment later by the leading scientists in the project generate a powerful framework that produces large visually classified samples of galaxies, reaching the scale of hundreds of thousands of galaxies far surpassing what is possible by experts alone. Not only that, but the engagement of the volunteers with the platform is proven to lead to new discoveries based on particular observations that were not possible before (Cardamone et al., 2009; Masters et al., 2010), and provides good opportunities for science outreach and public engagement.

Astronomers are very good at breaking the order of magnitude scale of their datasets with new instruments and surveys, such that even citizen science projects are not sufficient to classify **all**[1] galaxies. Hence, one of the active and vibrant areas of research in astronomy is designing fast and robust fully automated methods for morphological classifications. Some of these methods use the quantitative morphology measurements described earlier (§ 1.1.2) and other physical processes to define regions on the parameter space that are dominated by one morphological class (or type of merger, for instance). These include, but are not limited to, linear discriminant analysis (LDA, Ferrari et al., 2015), decision trees (Snyder et al., 2019), artificial networks (Storrie-Lombardi et al., 1992) and support vector machines (Cheng et al., 2020). They all rely on previously extracted features from the observations, and only act as a way to organize these on a particular model for the final classification. Some of these methods are very successful, such as the merger classification based on CAS parameters (Bershady et al., 2000; Conselice et al., 2008) or G-M20 (Lotz et al., 2004, 2008; Snyder et al., 2019). However, it is difficult to benchmark their results on new data without prior visual classification, as the decision boundaries for classes move around based on several factors, such as the noise levels/redshift (Thorp et al., 2021). Thereby, visual classifications will never be out of fashion, as they are benchmarking backbones of any automated classification task. More critical even, they have a central role

---

[1] One might ask themselves if this is really necessary, but there is no such thing as too much data.

in end-to-end learning algorithms, where visual classifications are used as training data for deep learning models. Models such as these, including the Convolutional Neural Network, using supervised learning, are the central framework used in Chapters 3 and Chapter 4. Details are given in the appropriate Chapter, however in §1.5 we give a basic outline on how Deep Learning and Convolutional Neural Networks work.

1.3 Hierarchical Structure Assembly and Λ-CDM

Within a dynamic universe, the evolution of galaxies has to be put into context with the evolution of the Universe itself. As contemporary extragalactic astronomy evolved alongside observational cosmology, the foundations for the concordance cosmological model widely adopted today, the Λ-CDM model, are also the foundations for galaxy formation and evolution. The Λ-CDM cosmological model describes a Universe expanding following a Big Bang, around ≈ 14 billion years old, dominated at present day by a matter component and a pressure energy of unknown origin. The total matter density makes up to $\approx 30\%$ of the total energy density of the universe, only 4% of this matter being in the form of baryons (e.g stars, gas, dust), while the rest is in the form of Dark Matter. This high level description seems very elusive, but this is a powerful and simple cosmological framework that explains key observational evidence, such as the Cosmic Microwave Background (CMB, Alpher & Herman, 1948; Gamow, 1946; Penzias & Wilson, 1965; Planck Collaboration et al., 2018), the recession velocity of galaxies (Hubble, 1929a), the chemical abundances of galaxies (Font et al., 2006), and the distribution of matter and the large scale structure of the universe (Colless et al., 2001; Eisenstein et al., 2005).

Dark matter haloes formed from small density anisotropies in the early universe. After recombination, when photons could freely travel, the universe became transparent. As the Universe cooled down and expanded, the radiation pressure started to give way for gravity, allowing baryonic matter to gravitationally collapse in these previously formed dark matter haloes. Therefore, the distribution of matter in the Universe can be traced back to these density fluctuations (Blumenthal et al., 1984). The formation of structure follows suit, with galaxies forming through the hierarchical clustering of these dark matter haloes (White & Rees, 1978; Cole et al., 2000), with smaller galaxies merging together to form larger ones, and also bigger structures, such as galaxy groups and clusters (White & Frenk, 1991). Figure 1.3 illustrates what we call a merger tree of dark matter haloes or galaxies where their formation and evolution is depicted by this hierarchy with smaller components merging together to form larger ones.

Unfortunately, this information is inaccessible from real observations (i.e., it is impossible to recover the full merger tree of a individual galaxy). However, by searching for galaxies undergoing merging, and by a statistical survey across cosmic time, one can track the rate of this clustering over the age of the Universe (Mantha et al., 2018; Duncan et al., 2019) and compare to theoretical models and simulations (Vogelsberger et al., 2014; Schaye et al., 2015). Therefore, one of the best ways of understanding the evolution of galaxies in the Universe is by tracing parallels between numerical simulations of these systems with observations. In Chapter 2 and Chapter 3 we leverage deep learning models as a way to construct this bridge.

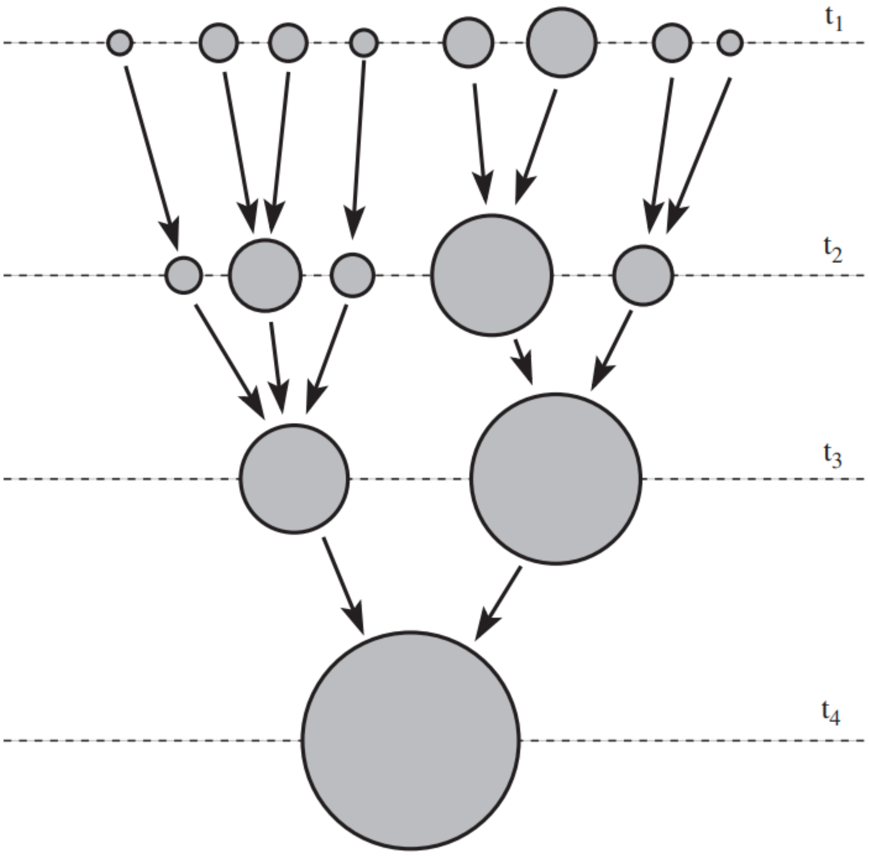


Figure 1.3: Schematic of a galaxy merger tree. Smaller galaxies merge together with other galaxies to form larger, more massive systems, as time progresses and these systems interact gravitationally. This tree shows a simplified case with only a few galaxies; in practice merger trees are very complex, the massive galaxies we see in the nearby universe are the result of hundreds and thousands of these small systems coalescing. Source: Mo et al. (2010)

1.3.1 Cosmological Simulations

Numerical simulations of isolated and interacting galaxies evolved from few particle N-body systems (Toomre & Toomre, 1972) to complex gravito-hydrodynamical simulations such as EAGLE and ILLUSTRIS (Vogelsberger et al., 2014; Schaye et al., 2015; Nelson et al., 2019) capable of simulating entire universes in periodic boxes, including gas physics, stellar particles, black holes, dark matter, feedback mechanisms, and chemical enrichment. These highly sophisticated simulations are capable of generating massive galaxies at $z = 0$ that reproduce similar morphologies to what is observed in the local universe (Huertas-Company et al., 2019), by evolving the initial conditions of the CMB within a Λ-CDM cosmology.

Essentially, data from these cosmological simulations can be forward modeled to the observational domain by simulating how the light of sources within these simulated galaxies interacts with the ISM and travels until it reaches a hypothetical detector (Bottrell et al., 2019). Then, observational effects can be applied to these synthetic images so to make them as similar to observations as possible (Bottrell & Hani, 2022; Marshall et al., 2022b). Then, one can explore the merger trees from the simulations, select galaxies of interest, undergoing particular physical processes, and produce images to interrogate how that type of object would look like in the observations. We discuss this in detail in Chapter 2 and Chapter 3.

1.3.2 Distance Measurements and Redshift Effects

The cosmological principle, which states that the universe in large scales is homogeneous and isotropic – a fair assumption based on contemporary evidence – produces solutions to Einstein's field equations that describe how the Universe as a whole evolves, and how it depends on its matter content, Ω_m, the curvature

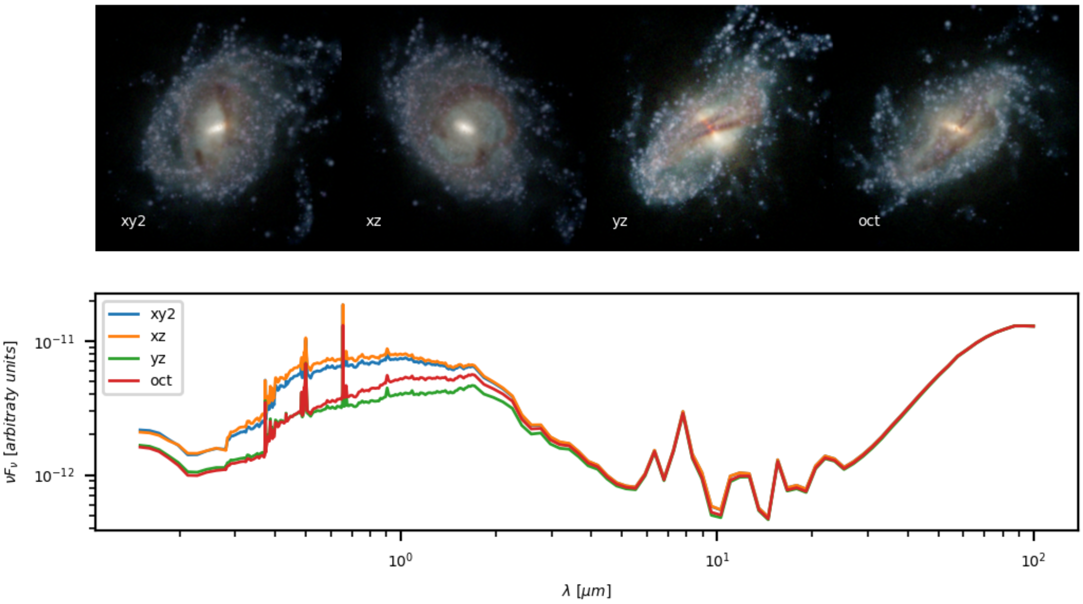


Figure 1.4: Simulated mock imaging of a galaxy after a merger event in IllustrisTNG 50-1 simulation (z=0.79, ID=381587) produced with the radiative transfer pipeline discussed in Chapter 3 in four different orientations (top), and corresponding SEDs to each vieweing angle shown on the bottom. Simulated galaxies show rich structural components, star forming regions, dust effects.

κ and the cosmological constant Λ. These solutions are known as the Friedman equations which ultimately can take the useful form of

$$H(t) = H_0 \left[\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}\right]^{1/2}, \quad H(t) = \left(\frac{\dot{a}}{a}\right) \tag{1.2}$$

where $H(t)$ is the Hubble factor, and a the scale factor encoding the expansion, describes how the scale factor changes with time. H_0 is the Hubble constant, $\Omega_{m,0}$ is the matter density term, and $\Omega_{\Lambda,0}$ the density contribution of dark energy. The dark energy contributions are indirectly inferred by

$$\Omega_{m,0} + \Omega_{\Lambda,0} = 1. \tag{1.3}$$

The form of Eq. 1.2 is useful as it is defined by two parameters that are observable, H_0 and $\Omega_{m,0}$, and from this equation relations can be traced between the observed redshift of galaxies to distance measurements, particularly the co-moving distance is defined as

$$D_c(z) = \frac{c}{H_0} \int_0^z \frac{dz}{\left[\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}\right]^{1/2}}, \tag{1.4}$$

and all other distance relations follow $D_c(z)$ (Hogg, 1999). Many of the redshift dependent effects discussed in the following chapters of this thesis can be derived from Eq 1.2, with the cosmological parameters as measured by Planck Collaboration et al. (2018).

1.4 High Redshift Morphology

Hubble types are still found in the early Universe, but in a different proportion as to what is found in the local Universe (Driver et al., 1995; Schade et al., 1995;

Delgado-Serrano et al., 2010; Mortlock et al., 2013). Likewise, the visual morphology of distant galaxies deviates for nearby galaxies for a number of reasons. First, distant objects are imaged at lower angular resolution, and many of the distinct signatures featured in the original Hubble sequence are unresolved. This is aggravated by observations made from the ground, as the atmospheric seeing imposes a high constraint on angular separations in astronomical images, thus galaxies outside the local universe $z > 0.1$, are mostly poorly resolved for morphology studies. Second, distant galaxies are moving away from us, and their Spectral Energy Distribution (SED) is shifted towards redder wavelengths as $\lambda \times (1 + z)$. This bandpass shifting, consequentially moves away the rest-frame optical emitted light from the galaxy to redder filters out of an optical telescope range (Blanton et al., 2003). Our atmosphere absorbs some of the near infrared to infrared light, limiting what can be detected from the ground. Moreover, due to cosmological effects, the emitted light from larger distances is dimmed, and distant galaxies are fainter and more difficult to detect (Lubin & Sandage, 2001). Figure 1.5 shows simulations of how a nearby galaxy (NGC 497) would be observed at higher redshifts by 4 different instruments, the SDSS survey camera, DES camera, LSST camera and HST ACS (**?**) through the FERENGI code (Barden et al., 2008). From simulations such as the one presented, it's clear that morphological studies from ground-based facilities requires rare observing conditions and use of complex techniques such as adaptive optics. On the other hand, HST is able to resolve morphological features out to high redshifts, being only limited by its wavelength coverage.

These effects are only observational, and do not include any evolutionary trends in galaxies. However, high redshift galaxies are found to experience tremendous evolution beyond the local Universe, and this adds to the complexity of observing

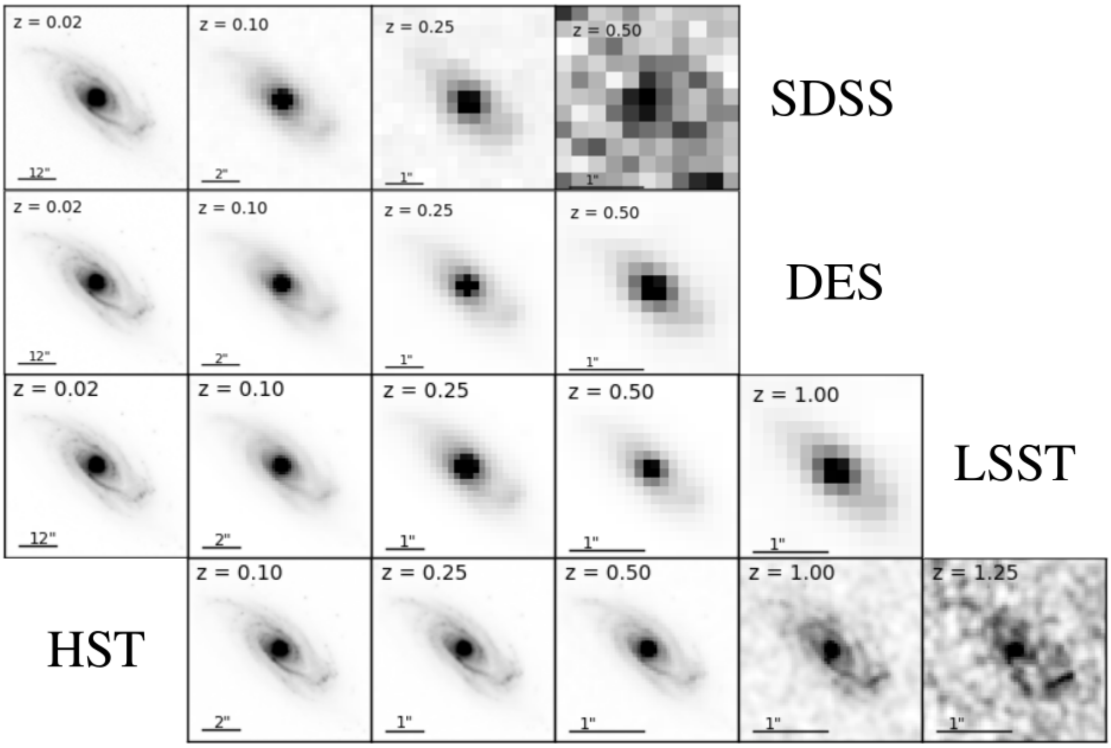


Figure 1.5: Example of redshift simulations of the NGC 497 galaxy using FERENGI (Barden et al., 2008) as it was observed by SDSS, DES, LSST and HST. The resolution difference between ground and space instruments is critical in moderated to high redshifts. Important structures for the classification process, such as spiral arms, become too small or end up vanishing in the image **?**.

distant systems. For example, galaxies were smaller in the past (Whitney et al., 2019) and peculiar galaxies are more common in the early universe (Mortlock et al., 2013) as a consequence of more frequent mergers (Duncan et al., 2019). For these reasons, we focus our efforts on space-based observations with HST and JWST.

The HST wavelength coverage goes up to $\lambda \sim 1.6$ μm, and for that reason is only able to probe the rest-frame UV at $z > 2.5$. Fortunately, the successful launch of the JWST extend this window to $z \sim 10$, as its high resolution NIRCam coverage goes up as $\lambda \sim 4.4$ μm, and highly supersedes the resolution IR capabilities of Spitzer. Thus, for the first time, we are able to investigate the optical morphology of $z > 3.0$ galaxies. Chapter 4 and Chapter 5 are focused on this question.

1.5 Deep Learning

Deep Learning methods gained notoriety for being very efficient and robust on image classification tasks, achieving superhuman performance (Krizhevsky et al., 2017). But even before modern deep learning models were employed, artificial neural networks were already used for classification tasks (Storrie-Lombardi et al., 1992), and deep learning follows the same underlying principles anchored in a robust optimization algorithm known as the gradient-descent (Ruder, 2016), together with the backpropagation (Rumelhart et al., 1986).

Here we limit our discussion to the case of supervised learning, where a dataset of annotated examples (labels) is used to train a model to predict these annotations, and then used on non-annotated data. An approach that can produce models that are heavily tailored at a specific task. This is in contrast to unsupervised learning (clustering) where labels are not a requirement, and representations that cluster different examples together is learned as part of the training,

The overall basic idea behind supervised machine learning is that one wants to map a set of inputs ($\mathbf{x}$) to a set of ground truth values (y) through an unknown function f that is based on a set of learned parameters called weights ($\mathbf{w}$), that is

$$\mathbf{y} = f(\mathbf{x}, \mathbf{w}), \tag{1.5}$$

where $\mathbf{w}$ is learned from data points $\{\mathbf{x}, \mathbf{y}\}$, through the optimization (minimization) of a cost function $\mathcal{L}$, usually taking the form of

$$\mathcal{L} = \frac{1}{2} \sum_{n} \left[f(\mathbf{x}, \mathbf{w}) - \mathbf{y} \right]^2 . \tag{1.6}$$

The weights $\mathbf{w}$ are them updated iteratively by the process called back propagation using the gradient of the cost function with respect to the weights, hence gradient descent,

$$w^i_{t+1} = w^i_t - \alpha \frac{\partial \mathcal{L}}{\partial w^i} \tag{1.7}$$

until convergence or when a set criteria is met, and that this function f is capable of generalizing generalize beyond the already seen data. This framework is general and applies to a wide range of machine learning methods. Deep learning, in this case, defines the functions f through an hierarchical combination of several sets of artificial neurons, that take an input, apply a non-linear function based on its weight w^i, usually known as the activation function, and passes this result to the next layer of neurons, until it reaches the output. The optimization algorithm in this case, adjusts the parameters of each of these neurons such that network is trained for the particular task in hand. The backpropagation algorithm defines the way and order in which these weights need to be updated, usually starting from the last layer (the output) all the way back to the inputs. This creates a highly non-linear model capable of learning patterns from data, encoding them in its weights.

Convolutional Neural Networks (CNNs, for uses in astronomy, see Huertas-Company et al., 2015; Bottrell et al., 2019; Bickley et al., 2021; Walmsley et al., 2022) are a type of artificial neural network that takes as input 2D representations, such as images, and replace some of the neuron operations of some layers by convolutions, so that as the convolutional kernels are learned instead of scalar weights. This makes the CNNs end-to-end models, where it is not necessary to produce scalar features that are going to be used as inputs, the network itself extracts the best features and patterns during optimization as well. In Chapter 2 and Chapter 3 we will discuss the use of CNNs that are similar to the AlexNet type of networks (Krizhevsky et al., 2017). Neural networks are computationally expensive to train, but once their parameters are optmized, the inference time is very fast. Millions of images can be processed in seconds (Tohill et al., 2021), enabling the community to deal with new scales of data. Moreover, with the popularization of Graphics Processing Units (GPUs), together with software optimized for it, such as TENSORFLOW (Abadi et al., 2016) and PYTORCH (Paszke et al., 2019), training deep neural networks is more manageable than ever before.

We are particularly interested in the capabilities of CNNs to encode and learn from information only available in cosmological simulations, such as what is done for low redshift galaxies in Bottrell et al. (2019) and Pearson et al. (2019), and then be used to do classifications of high redshift galaxies in real observations. This framework and the resulting models act as a bridge between simulations and observations.

1.6 Thesis Format and Outline

In this thesis, I will explore the evolution of galaxy morphology and structure throughout the Universe's first 10 billion years, within $0.5 < z < 8$, with a focus on galaxy mergers for $z < 3$, and more general morphological evolution with the newly available JWST data for $z > 3$. The main goal of this thesis is to explore new ways to characterize the morphology and structure of high redshift galaxies, and thus measure how galaxies evolved across cosmic time. Quantitative morphology from ground-based data is limited due to the atmospheric seeing (§ 1.4) having too large an angular size in comparison to the angular size of $z > 0.5$ sources (**?**), and thus limited to $z < 0.5$ studies. Therefore, this thesis is entirely focused on space-based HST and JWST data, exploring morphology on the CANDELS survey fields.

Given the time-frame, and the uncertainty regarding the JWST launch, I did not expect to work on JWST data as part of my PhD. Yet, as soon as this possibility became real, I made all the efforts I could to include it in this thesis. In light of this, this thesis consists of two parts. The first half, Chapter 2 and Chapter 3, is focused on a new way of forward modeling cosmological simulations to the observational domain using convolutional neural networks (CNNs). While on the second half, Chapter 4 and Chapter 5, I discuss for the first time the rest-frame optical morphologies of galaxies at $z > 3$ observed by the JWST.

In Chapter 2 I show that CNNs are a powerful tool to encode the information from the cosmological simulation merger trees, capable of performing merger classification on real high redshift HST CANDELS data. With our trained models on IllustrisTNG TNG300-1 mock observations, we measure galaxy-galaxy merger fractions and galaxy merger rates. This is the first agreement between morphology-based merger rates and pair statistics merger rates.

In Chapter 3 I explore the morphological distinction between already coalesced merging galaxies and non-interacting highly star forming galaxies. These two morphologies pose a challenging problem for automated methods, as visual properties are similar. We show that our CNN framework developed in Chapter 2, together with a refined mock observations pipeline is capable of producing models that reach 80% accuracy on this particular separation. Previous methods based on non-parametric morphologies are highly insensitive to this problem, and are very incomplete. We use these methods to discuss the evolution of post-mergers in the CANDELS fields and provide a detailed discussion on interpretability of our deep learning methods.

In Chapter 4 we report the very first morphological study with JWST, observing galaxies in the SMACS 0723 cluster field. We use a visual classification effort together with quantitative morphology measurements to report a surprising mismatch between the number of disk galaxies detected with HST and JWST. Around ten times more disks are found than expected from previous HST studies. We briefly discuss the implications of this result and how it fits in the galaxy formation and evolution big picture.

Chapter 5 we expand the framework of Chaper 4 to the CEERS observation program that has an overlap with the EGS observations in the CANDELS fields. We release to the community the biggest sample of visually classified galaxies observed with JWST to-date, with 4265 galaxies that are both observed by HST and JWST. With this dataset, we carefully discuss the evolution of the Hubble sequence up to $z \sim 8$, how it seems to be always present for low to intermediate mass galaxies, while an evolution driven by mergers is observed for the massive galaxies. We detail the quantitative morphology characteristics of this sample, and how it correlates with visual optical morphology.

We finish with a brief discussion in Chapter 6 of the results presented in this thesis, how the merger evolution at $0.5 < z < 3.0$ and the general morphological evolution at $z > 3.0$ are linked, and what are the next steps to explore this connection further.

Chapter 2

Galaxy Merger Rates up to z∼3 Using a Bayesian Deep Learning Model: A Major-merger Classifier Using IllustrisTNG Simulation Data

The content of this chapter has been published (Ferreira et al., 2020) in The Astrophysical Journal.

2.1 Abstract

Merging is potentially the dominant process in galaxy formation, yet there is still debate about its history over cosmic time. To address this, we classify major mergers and measure galaxy merger rates up to $z \sim 3$ in all five CANDELS fields (UDS, EGS, GOODS-S, GOODS-N, COSMOS) using deep learning convolutional neural networks trained with simulated galaxies from the IllustrisTNG cosmological sim-

ulation. The deep learning architecture used is objectively selected by a Bayesian optimization process over the range of possible hyperparameters. We show that our model can achieve 90% accuracy when classifying mergers from the simulation and has the additional feature of separating mergers before the infall of stellar masses from post-mergers. We compare our machine-learning classifications on CANDELS galaxies and compare with visual merger classifications from Kartaltepe et al., and show that they are broadly consistent. We finish by demonstrating that our model is capable of measuring galaxy merger rates, $\mathcal{R}$, that are consistent with results found for CANDELS galaxies using close pairs statistics, with $\mathcal{R}(z) = 0.02 \pm 0.004 \times (1+z)^{2.76\pm0.21}$. This is the first general agreement between major mergers measured using pairs and structure at z $<$ 3.

2.2 Introduction

Galaxy mergers are an explicit display of the hierarchical assembly of the universe, where galaxies and their dark matter halos merge together to form more massive systems (e.g. Mo et al., 2010). Indeed, the rate by which galaxies merge is a consequence of how the universe evolved, and can be used as an observable for the history of mass assembly of galaxies (Conselice et al., 2014). The understanding of how mass is assembled by galaxies is a very important piece of the galaxy formation and evolution landscape. It is known to happen in two ways: merging (Duncan et al., 2019) and through the accretion of gas from the environment, resulting in star formation (Almeida et al., 2014). The contribution of star formation to the mass assembly of galaxies is well measured even to high redshifts, where a peak in star formation rates are observed around $z \sim 2$ (Madau & Dickinson, 2014). The contribution from mergers, however, is less straightforward to measure and has some difficulties linked to how we identify merging systems (Conselice, 2006; Lotz et al., 2008; Conselice, 2014; Man et al., 2016).

Overall, two distinct methods are currently used to find galaxy mergers. One consists of finding close pairs of galaxies that fulfill a maximum separation criteria (both in redshift and angular separation) such that their orbits will dynamically decay with time resulting in a merger event. This is a quite successful approach and enabled merger fractions and rates to be estimated up to $z \sim 6$ (e.g. Mundy et al., 2017; Duncan et al., 2019). The second method relies on non-parametric morphological measurements that are robust for finding galaxies with disturbed morphologies, which is a strong suggestion (but not solely) for galaxy merging and interactions. In this case, a suite of measurements, generally the CAS (Concentration, Asymmetry, Smoothness) and the G-M_{20} systems, are used together to generate a parameter space which serves as a diagnostic tool for galaxy morphological classification (Conselice, 2003; Lotz et al., 2004). Some regions of this parameter space are in general dominated by merging galaxies, which then can be used to determine if a galaxy is likely a merger or not (Conselice, 2003; Lotz et al., 2004, 2008).

Both methods have had success (Conselice et al., 2003; Lotz et al., 2004; Conselice, 2009; Mundy et al., 2017; Duncan et al., 2019), but they probe galaxy mergers in different ways and rely on different assumptions. For example, in the case of galaxy pairs, merger fractions and rates are measured taking into consideration that the merger event did not happen yet, and may not happen, while the traditional non-parametric approach is only able to probe around one third of the period of the merger event, when morphologies are disturbed enough to distinguish from normal galaxies (Hubble type galaxies; Conselice, 2006). On top of that, not all regions of the parameter space generated by non-parametric measurements are purely compromised of galaxy mergers, some can be highly contaminated by non-merging galaxies. Other types of galaxies can have signatures that produce similar values, and not all mergers occupy that defined parameter space for the entirety of the merging event. This results in some contamination, generally from star forming galaxies, where star formation regions show themselves as clumpy light in the morphology of the galaxy which can, by eye mimic the appearance of an ongoing merger.

Another problem inherent in measuring merger rates is the knowledge of the time-scales involved in the merger event. It is very difficult to infer time-scales from observations, as we are limited to a single snapshot for each observed galaxy, and the merging timescale depends on several dynamical properties of the system (Lotz et al., 2008; Conselice, 2009). Fortunately, galaxy simulations can be used to estimate such timescales. Not only that, it is also possible to infer timescales attached to each method, for they probe different stages of the merger event (Lotz et al., 2008). Thus, large scale cosmological simulations can be used to estimate the dependence on redshift of merger timescales and visibilities (Snyder et al., 2017).

This scenario motivates us to develop new methods of finding mergers, and to improve upon current methods. One potential way to make progress in this direction is by using Deep Learning techniques where groups and layers of functions are laid out in a structure inspired by how the neurons in our brain works. In fact, some of these techniques, such as Convolutional Neural Networks, are dedicated to solve computer vision problems (CNNs; Goodfellow et al., 2016). For instance, CNNs are widely used in astronomy to tackle several problems, like galaxy morphological classification, segmentation and deblending (e.g. Huertas-Company et al., 2018; Reiman & Göhre, 2019; Huertas-Company et al., 2019; Cheng et al., 2020;

Martin et al., 2020).

One of the attempts to detect galaxy mergers with CNNs was done by Ackermann et al. (2018), where their network was trained with SDSS data labeled with classifications from Darg et al. (2010). They were able to detect new mergers in the SDSS data that were not originally found by Darg et al. (2010). This shows that indeed, CNNs are able to learn imaging aspects of merging galaxies. However, any bias in the classifications from Darg et al. (2010) are also incorporated in the model, since galaxies used for training were classified by eye.

Another experiment was conducted by Pearson et al. (2019), where galaxy mergers from the EAGLE cosmological simulation (Schaye et al., 2015) were used to train a CNN. In cosmological simulations such as this the merger history of all simulation galaxies is available through merger trees generated by Friend-of-Friends methods. This is a potential solution for labelling training data since this represents a ground truth relative to when two galaxies (or more) are merging, in contrast to eyeball classifications that can be uncertain. These authors also conduct cross training experiments, where simulated galaxies are classified with models trained with real galaxies, and the other way around. However, the results from the application of this trained model fails to classify galaxy mergers, even within the simulation. They attribute the performance of the network to the difference between EAGLE galaxies and real galaxies. Their conclusions is that mergers in the simulation have different morphologies from real galaxy mergers. Although these differences exist, their small sample size and low resolution mocks can also play a role on the poor performance achieved.

A different approach was recently employed by Snyder et al. (2019), where the authors used a combination of non-parametric morphological parameters, random forests, and ensemble learning to create a model which is capable of classifying galaxy mergers using the Illustris simulation (Vogelsberger et al., 2014) galaxies as the training sample. This approach however does not use the embedded powerful feature extraction layers present in CNNs and resembles more the classic classification methods in combination with some of the aspects of basic machine learning.

With this background in mind, we further explore how deep learning methods can help us extract more information regarding mergers from imaging data. We do this by training a model with only simulated data labeled with information available from merger trees in cosmological simulations. This has the potential to avoid biases that emerge from visual classifications, and by leveraging all the potential information deep learning methods provides, we can construct a full probabilistic approach to conduct predictions in real galaxies.

To do this, we construct a sample of galaxies from the IllustrisTNG suite of cosmological simulations (Nelson et al., 2019) with their complete merger histories available as a training sample, and then train a CNN to distinguish major mergers from non-merging galaxies with the goal of applying this to The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) fields (Grogin et al., 2011; Koekemoer et al., 2011). We check if our results are consistent with visual classifications from Kartaltepe et al. (2015) and galaxy merger rates from Duncan et al. (2019).

This Chapter is organized as follows: in §2.3 we describe how the data from IllustrisTNG was prepared while we elaborate our Deep Learning architecture in §2.4. We dedicate §2.5 to discuss our results both with the simulation data and real data and we summarize the Chapter in §2.6. All transformations and measurements here assume the same cosmological model used by IllustrisTNG, which are consistent with Planck Collaboration et al. (2018) results that show

$\Omega_{\Lambda,0} = 0.6911$, $\Omega_{m,0} = 0.3089$ and $h = 0.6774$. Magnitudes are quoted in the AB system (Oke & Gunn, 1983) unless otherwise specified.

2.3 Data

Our goal is to develop a major-merger classifier model trained with galaxies from cosmological simulations and explore whether it is capable of carrying out predictions on real galaxies. In these simulations, a galaxy's complete merger history is generally available through merger trees (Rodriguez-Gomez et al., 2015). This approach enables us to use a completely objective way of labelling our training data, bypassing any visual bias that might affect visual classifications, especially in this merger/non-merger classification task that deals with morphological features that can be the result of several processes, not only merging. However, this comes with drawbacks. The resolution of the simulation must be good enough to generate similar morphologies to the ones present in real galaxies. Not only that, but post-processing steps are necessary to mimic the same observational effects and characteristic noise of the data where predictions will be conducted. It is of utmost importance that the simulation is able to provide enough galaxy numbers for the classification task (i.e tens of thousands), as we expect it to be able to generalize to a different dataset. We also want to probe galaxies to moderate redshifts ($0 < z \leq 3$) so we can estimate galaxy merger rates using our predictions.

2.3.1 IllustrisTNG

All these requirements lead us to the IllustrisTNG project (Nelson et al., 2019), a suite of cosmological, gravo-magnetohydrodynamical simulation runs, ranging within a diverse set of particle resolutions for three comoving simulation boxes of length size, $50, 100, 300$ Mpc h^{-1}, named TNG50, TNG100 and TNG300, respectively. Each of these simulations probe a different resolution regime, in a trade-off between galaxy numbers and simulation resolution. As we are interested in building a large training sample, we recur to the largest simulation available, TNG300. Within each simulation box there are also different setups, with variations in the number of gas and dark matter particles. We limit ourselves to the highest resolution available in the largest simulation box, namely TNG300-1[1]. We briefly discuss how subfind defines a galaxy (subhalo) in the following section.

It is important to note, however, that the physical resolution of TNG300-1 does not perfectly match the CANDELS resolution, especially at higher redshifts. TNG100-1 and TNG50 would provide better resolution matched candidates if the dominant concern was physical resolution. Instead, our choice here was driven by the simulation volume, and the need to have the largest number of galaxies available to train our machine learning. As a way to mitigate potential issues that could come with this resolution mismatch we only use in our analysis massive galaxies with $M_* > 10^{10} M_\odot$ and major mergers in the case of mergers.

From TNG300-1 we draw two samples: a major-mergers (hereafter **MM**) only sample and a sample of non-interacting galaxies (hereafter **NM**). Details on how both samples are selected are described in §2.3.1 and §2.3.1, respectively. After selecting and creating a sample of clean galaxy images from IllustrisTNG, we need to apply effects to the imaging data to generate realistic galaxy mocks, this process is described in §2.3.3. For our sample of real galaxies, we choose to use galaxies in all of the CANDELS fields (COSMOS, UDS, GOODS-S, GOODS-N and EGS).

---

[1] As a comparison, the TNG100-1 simulation has approximately 4.3 million subfind groups at $z = 0$ while TNG300-1 has 14.4 million. These groups are sets of simulation particles that are bound together by the Sublink algorithm, which in a general sense can represent galaxies.

How we select galaxies from CANDELS is described in §2.3.2.

Subfind Friends-of-Friends algorithm

`Subfind` is a software that detects substructures by identifying clusters of particles with higher density that are bounded gravitationally (Springel et al., 2001). The process begins by using the Friends-of-Friends algorithm to identify a main cluster. Then, for each particle, a local density is estimated using a kernel estimation method with a fixed number of smoothing neighbors. From these isolated density peaks, particles with decreasing density are sequentially added. If a saddle point is reached in the global density field that connects two separate high-density regions, the smaller cluster is considered a substructure candidate, and the two regions are merged. The substructure candidates undergo an iterative unbinding process using a tree-based calculation of the potential.

Major-Merger (MM) Sample

All our samples are selected through available merger trees. First, we limit our exploration to $z \leq 3$ (snapshots 99 to 25). As we will later use near-infrared imaging, this redshift limit is applied to ensure that we are not probing rest-frame UV observations. We limit this work to the near-infrared to mitigate the effects of dust attenuation, as the IllustrisTNG imaging data used here is not produced by a proper radiative transfer process. As such, it is essential to avoid probing the rest-frame UV of the simulated galaxies where the effects of dust would be extreme. Thus, within our redshift range we expect the impact of dust to increase as our rest-frame wavelength is closer to the UV rest-frame. A full radiative transfer treatment of the images would be necessary to completely avoid this problem.

An alternative would be to use longer wavelengths, which will be possible with JWST imaging in the future. However, both solutions are beyond the scope of this Chapter.

Then, for each galaxy at $z = 0$ (snapshot 99), we climb the merger tree by checking for cases where there is more than one progenitor in a previous snapshot that fulfill the major-merger mass ratio, μ, criteria,

$$\mu \geq \frac{1}{4}, \tag{2.1}$$

and at least one of the progenitors has $M_* \geq 10^{10}\ M_\odot$. This mass criteria is used to ensure that our galaxies are well sampled within the mass resolution of TNG300-1. If that is the case, we select the snapshot where these criteria are met as the central snapshot of the merger event. This means that this is the snapshot where the subfind algorithm decided that particles from its progenitors became one descendant. However, it is still possible that in the central snapshot such galaxies are still separated by some distance in the sky, but will appear as only one galaxy in snapshots moving forward. With the central snapshot defined, we select all progenitors and descendants within ± 0.3 Gyr of the central snapshot as mergers as well. By doing so, we are selecting galaxy mergers in different stages of the merger event around a well defined time-scale. Galaxies in this selection window can appear as pairs, disturbed morphologies that indicate recent infall, and also cases where two or more galaxies already merged and little to no disturbance is visible.

For all selections before the central snapshot, we measure the distance between each progenitor, D_n. Here we apply an additional cut by limiting the distance between each pair of galaxies by $D_n < 20$ kpc h^{-1}. We are only interested in galaxies that are close enough to appear as if they are going to merge in the future. Such distance separation is within the range generally used for close-pair studies (e.g., Duncan et al., 2019), but we use it in the lower limit so that all pairs of galaxies involved in a merger event can be sampled in the image's field of view used in this work.

This selection procedure yields $\sim 30,000$ distinct major-merger candidates. The information in each selected object with respect to its central snapshot enables us to also categorize this sample further in different cases of mergers. All selected objects that have redshifts higher or equal to the redshift of the central snapshot are marked as merger candidates before the merger event (hereafter **BM**) and the cases with redshifts lower than the central snapshot's redshift are considered post-mergers (hereafter **PM**).

This will not limit our approach towards classifying galaxy mergers only in these two classes, as in §2.4.1 we will show that we can still use the prior probability to do a **MM**/**NM** classification instead of a **BM**/**PM**/**NM** classification. The only difference when moving from specialized classes to general mergers is using appropriate corresponding observing timescales. It is necessary to use $\tau_{\text{obs}} = 0.3$ Gyr when working with **BM** and **PM** classes, and $\tau_{\text{obs}} = 0.6$ Gyr when working with **MM** in general, to appropriately reflect our sampling windows. To help with the visualization of our method, we show in Fig. (2.1) a simplified sketch of our selection criteria for two galaxies undergoing a merger.

The division between major mergers and minor mergers adopted ($\mu > 1/4$) is arbitrary and follows the conventional use among merger studies (Conselice et al., 2003; Lotz et al., 2004, 2008; Snyder et al., 2017; Duncan et al., 2019).

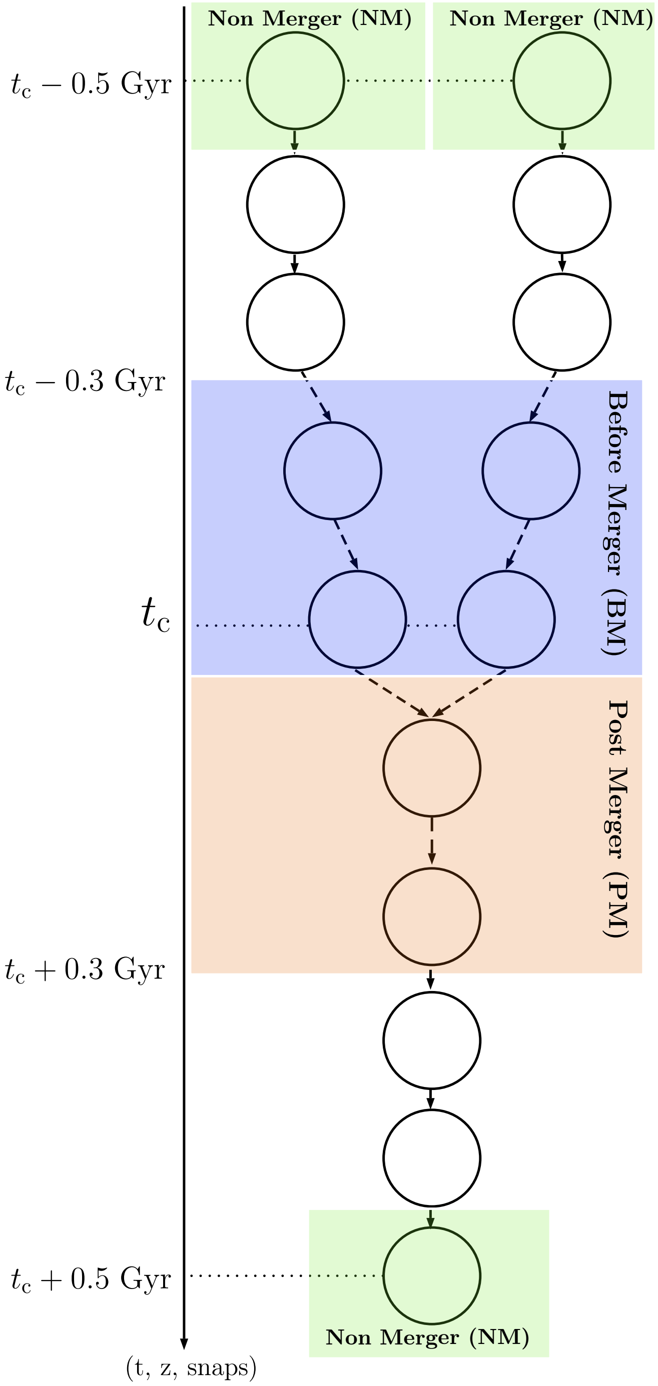


Figure 2.1: Diagram with a simplified example of two galaxies merging and the resulting label selection for each object and snapshot. Area in blue shows galaxies selected with **BM** labels, orange represent galaxies with **PM** labels and in green **NM**. Both **BMs** and **PMs** are selected with our selection timescale, $\tau_{\text{obs}} = 0.3$ Gyr , whilst **NMs** are defined with a longer interval from the central snapshot. Selection windows are drawn based on the central snapshot, $t_c \pm \tau_{\text{obs}}$. The **BM** window include the central snapshot. Note that the average time between snapshots in IllustrisTNG is of about $\tau \sim 0.15$ Gyr, which makes out selection window around 2 snapshots wide.

Non-Merger (NM) sample

A sample of non-mergers is a requirement for our classification task, and necessary for our model to learn how to distinguish major-mergers from other types of galaxies. As there are many more galaxies in the simulation than just major-mergers, we use the number of major-mergers found in the **MM** sample selection as a guideline to define a control sample of non-interacting galaxies.

First we apply redshift and stellar mass cuts to select galaxies in the same range as the **MM** sample, with $z < 3$ and $M_* \geq 10^{10}$ $\mathrm{M}_\odot$. Next, we clean this pre-selection from interacting galaxies as best as possible. This can not be done by just simply removing the galaxies found in the **MM** sample from this new selection as there are other mergers occurring, with lower mass ratios, and cases where a merger event can have longer timescales than $\tau_{\mathrm{obs}} \pm 0.3$ Gyr, for selecting the **MM** sample. This means that it is possible to have merging morphologies with broader timescales in the simulation. Thus, to solve this we do a broader search of merging galaxies, looking at all mass ratios and mergers occurring in ± 0.5 Gyr. Then, we proceed to remove all galaxies found in this way from the initial redshift and stellar mass cut. The resulting sample is then separated in the same bins of redshift as the major merger sample, enabling us to draw randomly the same number of galaxies for each redshift bin in order to construct a sample that has a similar redshift distribution, as shown in Fig (2.2) (in the outer plot by the blue solid line and green dashed line, for mergers and non-mergers, respectively).

Nevertheless, these selections are made only within the simulation merger trees. We still need to produce the imaging data that will be used to train our model. However, it is important first to define the data in which we are going to apply our model to make predictions, as we have to apply similar instrumental and

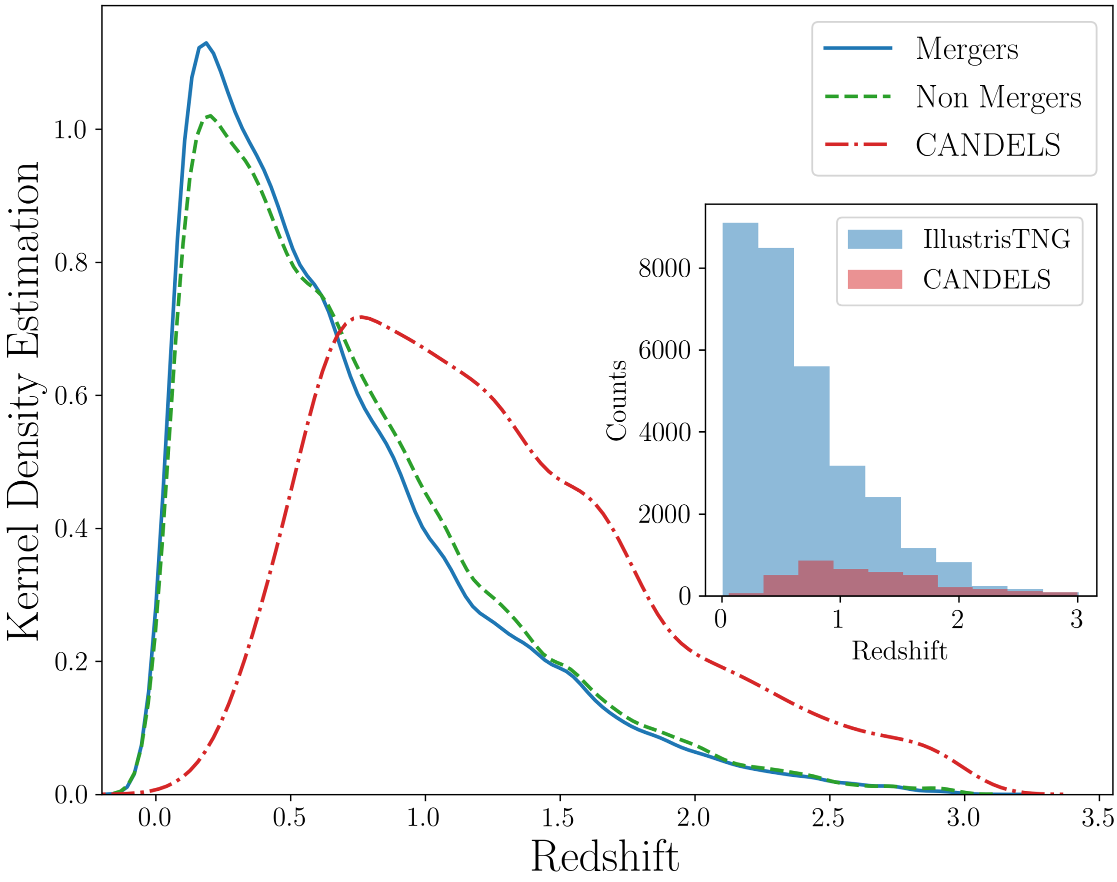


Figure 2.2: Redshift distribution for the simulated Major Merger sample (blue solid line), simulated non-interacting sample (green dashed line) and the CANDELS sample (red dot dashed line). The redshift distribution for our IllustrisTNG mergers and non-merger samples are by construction very similar. We also display the CANDELS redshift distribution to show that it does not match the redshift distribution of the samples used for training, but its numbers are within the range of the simulation distribution, as demonstrated by the unnormalized redshift histogram in the inner plot, showing all the IllustrisTNG galaxies in blue and CANDELS galaxies in red.

observational effects in order to mimic the data the best way possible. In our case, we want to apply our model to galaxies in the CANDELS fields.

2.3.2 CANDELS Fields

One goal of this work is to do predictions on CANDELS WFC3/IR imaging data (Grogin et al., 2011; Koekemoer et al., 2011). This consists of wide field data with enough depth to detect galaxies in the limit of our selection on the simulation data. This data was already used extensively within galaxy merger studies, with merger rates estimated up to $z \sim 6$ (Duncan et al., 2019). There are also visual morphology classification catalogues (Kartaltepe et al., 2015), photometric redshifts and stellar mass estimates (Duncan et al., 2019), which are essential if we want to make the same selection cuts as the ones done in IllustrisTNG simulation data, as we are only interested in predictions on a similar parameter space.

Here our selection is similar to the one applied to the IllustrisTNG merger trees, with the exception that we do not use any merger classifications available to select it. The first step consists in removing all objects that have problems with quality flags in the original photometry catalogue and the Kartaltepe et al. (2015) catalogues, as we want to avoid edges, artifacts and stars. Then, we apply a magnitude cut in the H band of $H < 24.5$ mag following the same cut used in Huertas-Company et al. (2016) and Kartaltepe et al. (2015). A signal-to-noise (SNR) cut of SNR > 20 is also applied, as the magnitude cut would bias the SNR of our sample against extended sources. Then we proceed with the same cuts we made to the IllustrisTNG selection, using $z < 3$ and $M_* > 10^{10} M_\odot$. This results in a sample of 3759 galaxies wish high enough SNR.

Fig. (2.2) shows the redshift distribution of this subsample of CANDELS galaxies (red dot dashed line). It can be seen that this redshift distribution does not match the redshift distribution for IllustrisTNG galaxies. However, the inner plot shows an unnormalized redshift histogram of IllustrisTNG (blue) and CANDELS galaxies (red), which demonstrates that our training sample of IllustrisTNG galaxies is large enough to have at least similar galaxy counts to the CANDELS sample at higher redshifts. One might argue that it would be ideal to construct the training sample with the same redshift distribution as the data we are planning to do predictions with, but in this case, we are limited by resolution, which requires us to limit the scope to massive galaxies ($M_* \geq 10^{10} M_\odot$) only. At the same time, we are not introducing redshift information during training, apart from embedded instrumental and cosmological effects, so the variability on merger morphologies available in the regime where both redshift distributions disagree ($z < 0.5$) is essential to the learning model.

In the training step we tested matching the redshift distribution of the training sample with the CANDELS redshift distribution by removing low redshift galaxies from the training sample. However, our findings suggest that the performance of the model suffers from the smaller training sample by over predicting mergers at low redshifts. This is due the lack of generalization by the model when limited to smaller training samples. In this way, additional tests with different training samples are left for future work. Even though galaxies at different redshifts are intrinsically different, their morphologies can be degenerate.

Finally, we produce cutouts from the imaging data that represents a field of view of 50 kpc $\times$ 50 kpc using available redshift. In this way, we choose to rely on the redshift information available instead of using any assumption about the sizes of galaxies in our samples, as it is difficult to define it when two or more galaxies are interacting in the field of view. By using this approach, we are also preserving relative sizes between galaxies within our samples, which might provide important information for the network to use during the classification. As we are using CANDELS Near IR data, we proceed to produce galaxy images from IllustrisTNG and apply instrumental and cosmological effects to the images so that they are a realistic representation of CANDELS galaxies.

2.3.3 IllustrisTNG Imaging Data

We take advantage of the tools available in the IllustrisTNG API and website to select stellar maps for a given object in the simulation. The 'Galaxy and Halos Vizualization'[2] (Nelson et al., 2019) tool enables us to select a galaxy by combining the simulation run, snapshot and subfind identification to visualize a given object in several filters. It uses a pipeline coupled with CLOUDY (Ferland et al., 2017) photoionization code and Flexible Stellar Population Synthesis (FSPS)[3] through `python-fsps` (Conroy et al., 2009; Conroy & Gunn, 2010), a stellar population synthesis code, generating stellar density maps for the appropriate ages and metalicities (in rest or observational frames), as selected by the chosen filter, refer to Nelson et al. (2019) for details. However, this procedure has its limitations, as described earlier, as it does not include a full radiative transfer treatment, and does not account for dust.

This could impact some of the morphologies presented, especially for the star forming galaxies. Although studies using IllustrisTNG mocks generally use a complete radiative transfer approach for galaxies with high star formation rates (**?**Rodriguez-Gomez et al., 2019; Huertas-Company et al., 2019), we limit our sample only to near-infrared filters as a way to mitigate potential biases due the

---

[2] `http://www.tng-project.org/data/vis/`

[3] FSPS uses Kroupa IMF whilst stellar masses in our CANDELS catalogs are measured with Chabrier IMF, a $\sim 5\%$ offset is expected.

absence of dust in our treatment. Bottrell et al. (2019) shows that realistic instrumental effects, such as noise and an appropriate PSF, are more important than radiative transfer effects when training deep learning models, where the slight improvement in performance comes with a huge computational cost of producing galaxy mocks with full radiative transfer, especially for large samples of galaxies. Moreover, we do not explicitly use any color information in our model. In this way, one might use our galaxy mocks as stellar density maps, which will be closely related to the true morphology of the galaxy.

The following is a brief overview of our complete mock pipeline. The first step consists of the selection pipelines described in §2.3.1 and §2.3.1. The result of the selection is a list with each galaxy snapshot, subfindID and redshift. This is then fed to the Illustris API, requesting the mock produced by the Galaxy and Halos Vizualization pipeline. These images have field of views of 120 kpc × 120 kpc and are imaged in the observed frame for the HST F125W and F160W filters, which are available for the CANDELS fields. For each subsample, we randomly request 80% of the galaxies as face-on and 20% as edge-on, as we do not have the freedom to choose arbitrary orientations using this tool[4]. This proportion of face-on and edge-on galaxies is draw from axis ratio statistics from real galaxies in the CANDELS fields (e.g., Ravindranath et al., 2004; Mowla et al., 2019). This produces a set of clean images from the IllustrisTNG in the appropriate band, with cosmological dimming and k-correction applied. However, it is necessary to apply transformations in order to make mocks of these images as if they were observed by HST.

We apply cosmological geometric effects based on 'redshifting' (e.g., Conselice

---

[4] As this Chapter goes to press a new feature in IllustrisTNG API enable the user to use different projections and orientations instead of only face-on and edge-on orientation. This was not available when we generated our sample and we advise anyone doing a similar approach to use this new feature instead of only edge-on and face-on cases.

et al., 2003; Barden et al., 2008) approaches and add features of image realism (Bottrell et al., 2019) by appropriately simulating characteristics of CANDELS images, such as noise, PSF and adding the resulting image to a patch of the sky from the CANDELS fields. First, for each galaxy we apply a random rotation to the image following a crop to 50 kpc × 50 kpc field of view for both filters. The reason why images have such large fields of view is to have an adequate window for image transformations. If one would crop a galaxy image after a random rotation, artifacts would be noticeable around the edges, especially for cases with intermediate rotation angles. Then, as we know the exact pixel scale of the clean image, we can transform it to 60 mas/pixel HST WFC3/IR pixel scale and apply PSF effects by convolving it with a simulated PSF produced with `TinyTim` (Krist et al., 2004).

Noise is then added by converting the image to e/s^{-1}, multiplying it by an appropriate exposure time, and drawing a sample of it from a Poisson distribution. This is done to ensure that our mock images have similar shot noise to the real data. Then the resulting distribution is added to a empty sky region of the CANDELS fields. This region is selected randomly from a pool of pre-prepared regions. This is necessary, as the CANDELS fields are produced by a stack of multi-epoch sky subtracted images, which creates correlated noise (Koekemoer et al., 2011). These regions are empty since we expect the impact from crowding to be small in the redshift range probed here. Bottrell et al. (2019) shows that the presence of neighbour sources during training is important for the success of the deep learning model, but their simulations are limited to low redshifts. However, we show in §2.5.1 that the presence of crowded sky regions impacts the model negatively.

After all of these effects are introduced to the image, we prepare it for the CNN by re-sampling it to 128x128 pixels. This is the same as changing the pixel scale once more, but in most cases we are oversampling the image, as by this stage all images should be smaller than 128x128 pixels, thus we are not losing information by doing this. This particular resolution is selected so as to provide the CNN with the possibility of having more convolutional layers. Then, we package the whole sample in a HDF5 file with its train, test and validation split, including normalization. This is the package that is then used by the CNN.

The result of the selection and imaging data pipeline is summarized in Table (2.1).

2.4 Methods

We employ a Deep Learning approach with Convolutional Neural Networks (CNNs) to our images, a state of the art tool to solve computer vision problems (Goodfellow et al., 2016) that is gaining popularity among galaxy merger studies (Ackermann et al., 2018; Pearson et al., 2019; Bottrell et al., 2019). In a CNN, convolutional layers use convolution operations on multidimensional data, such as images, to extract features that can then be used for classification tasks in regular fully connected layers at the top of the CNN architecture. The convolutional part of the network can be divided into convolutional blocks, which can then nest more types of layers than just convolutional layers. However, each block is generally limited to probe a specific resolution range of the input data. Pooling operations are usually located between convolutional blocks with the goal of changing the input image to a lower (or higher) resolution. How these blocks and layers are organized and how wide the network is, including the number of filters, size of the kernels, and other properties, are defined by hyperparameters.

We briefly describe our method for finding a good model with an optimization

Redshift	**Snaps**	**Number of Galaxies**			**Before Merger**			**After Merger**			**Non Interacting**		
		Train	Test	Val	Train	Test	Val	Train	Test	Val	Train	Test	Val
$0.0 \leq z < 0.5$	99-66	19633	4257	4214	5331	1076	1171	4966	1117	1035	9336	2064	2008
$0.5 \leq z < 1.0$	67-51	13410	2837	2931	3240	669	726	3434	697	755	6736	1471	1450
$1.0 \leq z < 1.5$	50-41	6127	1342	1299	1377	292	295	1599	348	320	3151	702	684
$1.5 \leq z < 2.0$	40-33	2821	563	551	715	148	122	681	141	137	1425	274	292
$2.0 \leq z < 2.58$	33-27	993	213	216	257	62	60	240	44	44	496	107	112
$2.58 \leq z < 3.0$	28-25	210	44	45	57	12	14	51	7	9	102	25	22
Totals		43194	9256	9256	10977	2259	2388	10971	2354	2300	21246	4643	4568
		61706			15624			15625			30457		

Table 2.1: Summary of the IllustrisTNG samples of major-mergers and non interacting galaxies separated in redshift bins, label and the Training, Testing and Validation subsamples.

approach in §2.4.1, together with a short description of each hyperparameter; We describe the metrics used to evaluate the performance of our models and the architecture found by our optimization approach in §2.4.2.

2.4.1 Bayesian Optmization of Hyperparameters

Generally, CNNs and other Deep Learning methods are regarded as black boxes since their parameters are adjusted by an automated training process in order to maximize its performance, with little control over it apart from the architecture of the network. Its architecture is defined by a set of parameters that control how big a network is, how many layers there are, the learning rate and batch size, among other configurations. The results produced by a network model are highly dependent on its hyperparameters, so it is of utmost importance to fine-tune them as best as possible (Hacohen & Weinshall, 2019). Unfortunately, there is no method that is capable of finding the best set of hyperparameters without training the network and assessing its performance. Often, this is done by brute force methods such as grid searches, where a large domain of possible values for each hyperparameter is defined and portions of the domain are evaluated by training the corresponding network. If a high number of hyperparameters are present, the result is a very expensive task and might not lead to the best model.

To avoid this treatment, we use a Bayesian Optimization approach to find a good set of hyperparameters by modeling our architecture as a surrogate gaussian function $g(\mathbf{x_1}, ..., \mathbf{x_n})$, where $\mathbf{x_1}, ..., \mathbf{x_n}$ are the hyperparameters. Each possible combination of hyperpameters is a different model. This function is very expensive to evaluate, but with few samples it is possible to reach a set of hyperparameters that best optimizes the performance of the model by updating the posterior at each sample, using it to make informed guesses for the next observation. This technique is faster and can yield a set of hyperparameters that results in models with better performances than ones optimized manually, reducing the number of configurations necessary to reach a good model (Snoek & Larochelle, 2017).

Hyperparameters

We first define what will be considered a hyperparameter in our architecture by defining what aspects of it can be changed, setting a domain for each case. Here we briefly describe each of the hyperparameters of the architecture while a summary is displayed in Table (2.2).

We define a convolutional block as a group of convolutional layers that probe similar input resolutions. Each block is separated by pooling layers that change the size of the input for the next block by a factor of 2. The number of convolutional blocks, `number_conv_blocks`, is one of the main hyperparameters to define how long the convolutional portion of the network will be. Thus, the number of layers in each block, `number_conv_per_block` is also a hyperparameter. Every convolutional layer in a given block has the same number of filters and kernel size. The possible number of blocks varies between 1 and 5 while each block can have from 1 to 3 convolutional layers. Convolutional blocks not only group convolutional layers, but their activation and other auxiliary counterparts as well. Additionally, we set the number of filters in the first convolutional block, `initial_number_filters`, and the kernel size of the first convolutional block, `initial_kernel_size`, as hyperparameters. In a analogous way to the number convolutional layers, we consider the number of fully connected layers, `number_fullyconnected_layers`, and their size, `size_fullyconeccted_layers`, as hyperparameters as well. The loss function adopted for our classification problem is the binary cross-entropy log loss

(Good, 1952), the loss generally used for training with one hot encoded labels.

In neural networks, an optimizing function is used to maximize the performance of the network (minimize an error function). There are several distinct methods to accomplish this and different methods work better for different problems, as they represent strategies to find minima in the topology generated by parameters in parameter space. Here we choose from a pool of all optimizers available in Keras (Chollet & others, 2015) and let it also act as a hyperparameter of the architecture, even though it is not usually considered a hyperparameter.

We dedicate two hyperparameters to control the regularization of the architecture, namely the L2 regularization λ term, `l2_regularization`, and the dropout rate, `dropout`. The former act as a way to regularize the weights of the convolutional portion of the network by adding a penalty to the loss function in order to prevent spiked weights in favor of more diffuse configurations, while the later applies regularization to the fully connected layers by deactivating a percentage of the neurons for each layer equal to the dropout rate (`dropout`). By using dropout we will also be able to assess uncertainties in the network predictions. This is done by measuring probability distributions for each prediction by running the model for the same input with the dropout layers several times, as each time only portions of the fully connected layers are going to be used by the model. This approach is known as a Monte Carlo dropout (Cook et al., 2000; Huertas-Company et al., 2019).

Finally, we set a range of possible batch sizes, `batch_size`, and possible initial learning rates, `initial_learning_rate`, as hyperparameters.

Hyperparameter	Best Model
batch_size	256
number_conv_blocks	2
number_conv_per_block	2
initial_number_filters	32
initial_kernel_size	11
number_fullyconnected_layers	2
size_fullyconnected_layers	1024
optimizer	Adadelta
initial_learning_rate	0.1
l2_regularization	0.62
dropout	0.38

Table 2.2: Set of hyperparameters of our architecture and the best parameters found by doing Bayesian Optimization.

2.4.2 Performance Metrics and Best Model

In order to evaluate each of the possible models within our domain of hyperparameters, we first define how our models are going to be evaluated, since the Bayesian Optimization employed here runs as an automated process which tries to find the set of hyperparameters resulting in the best performance. This is assessed by training the network as a binary classifier of **MM**/**NM** (see §2.3.1 for definitions) with the training sample and performance evaluated in the testing sample. As we are not concerned with class imbalance problems at the moment, we simply try to minimize the loss function within our architecture. Models with low loss will represent models with high performance metrics. We also track the accuracy, precision and recall of each model, which inversely follow the loss very closely.

We perform the Bayesian optmization in the domain described with the GPyOpt python package (The GPyOpt, 2016). The model with the lowest validation loss is shown in Table 2.2.

2.4.3 Bayesian Neural Networks

Even though we carry out the hyperparameter optimization with the binary **MM/NM** classification, it is also important for us to probe if our CNN is capable of separating merger classes into further sub-classes, where galaxies are undergoing mergers at different stages. An easy distinction that we use from our selection procedure (Section 2.3.1) is to have a **BM/PM/NM** classifier. We follow a similar approach as is done by Huertas-Company et al. (2019), where a hierarchy of binary classifiers are used to develop classifiers that are specialized in a specific separation task. In our case, this means that we will have a **MM/NM** classifier trained with all our sample and another one trained only with mergers to separate them into **BM/PM**. Then, the output for this set of binary classifiers can be combined with Bayes Theorem to yield the probability in each merger class by:

$$P(\mathbf{BM}) = \mathrm{P}(\mathbf{MM}) \times \mathrm{P}\left(\frac{\mathbf{BM}}{\mathbf{MM}}\right), \tag{2.2}$$

$$P(\mathbf{PM}) = \mathrm{P}(\mathbf{MM}) \times \mathrm{P}\left(\frac{\mathbf{PM}}{\mathbf{MM}}\right), \tag{2.3}$$

where the probability of being a **NM** is simply the output for the **NM** class in the **MM/NM** classifier. In this sense, the **MM** acts as a prior probability.

By combining multiple binary classifiers together to do multi-class classification we are combining models refined to perform very specific tasks instead of using only one classifier that has to share all its weights and parameters among all classes. However, even though in some cases the output probabilities will not have any meaning, they can still be used to investigate the classification process. For example, a relatively high $P(\mathbf{PM})$ value for **NM** galaxies might indicate that their morphology has aspects resembling a disturbed galaxy. A high value of $P(\mathbf{BM})$ in a **NM** galaxy might indicate that the galaxy has companions. Nevertheless, this should not be common within the simulation data but might be useful when performing predictions in real data where no labels are available.

2.5 Results

With the architecture and the sample from the simulation described in Section (2.3.1), we train our model and explore how it performs in the validation sample. In this way it is possible to analyze how the model generalizes to simulation data it has not seen. This is necessary before we apply it to real data. After checking if the results are what we would expect within the simulation, we apply our model to the sub-sample of galaxies from all the CANDELS fields as described in §2.3.2.

2.5.1 Predictions using IllustrisTNG

By exploring how our models perform in the validation data, it is possible to identify its performance in a sample of galaxies from the simulation that the model has not seen during training or testing. Even though it should follow the performance of the testing set, this procedure enables us to verify if there are any biases in our set of classifiers. These, if present, can then be used to adjust predictions on real data later. We apply our model to the validation data to classify all galaxies in the sample in three classes: **BM**, **PM** and **NM**, as defined in §2.3.1. In Fig. (2.3) we show the distribution of probabilities assigned to each class using predictions within our hierarchy of models, as described in §2.4.3. We can see that

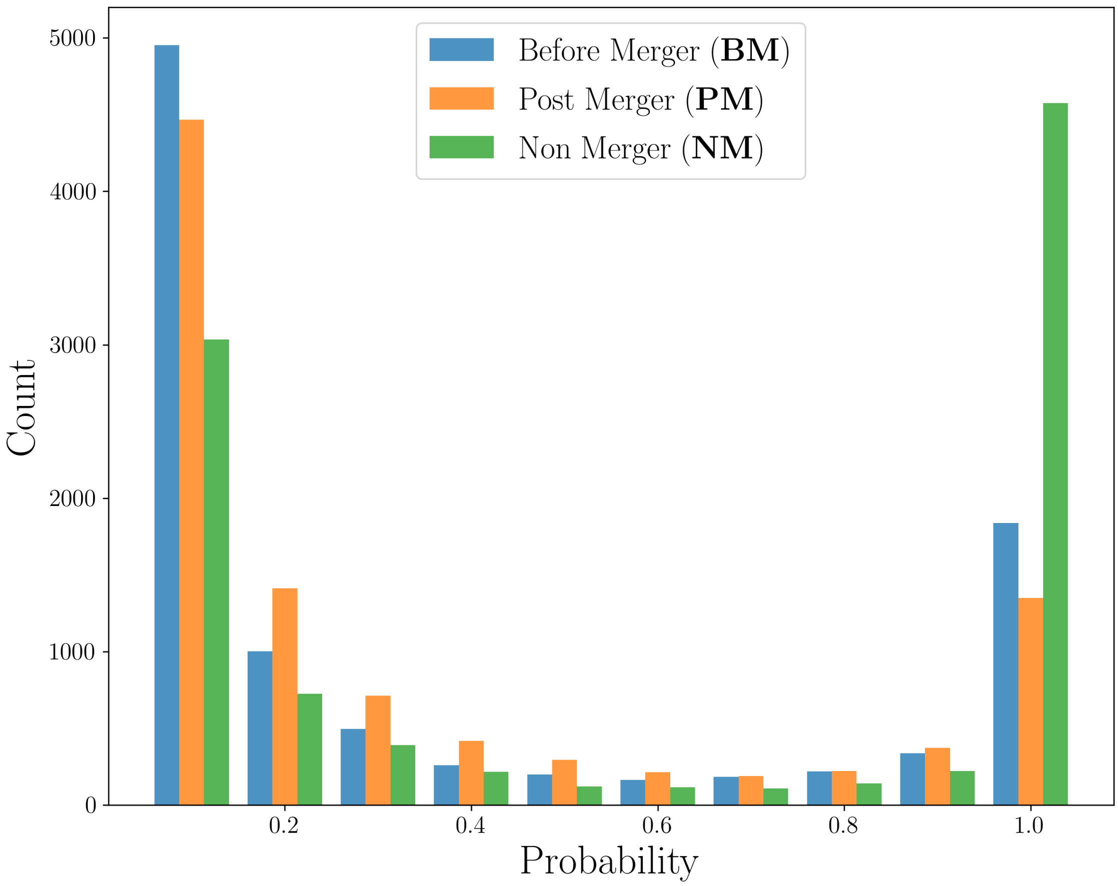
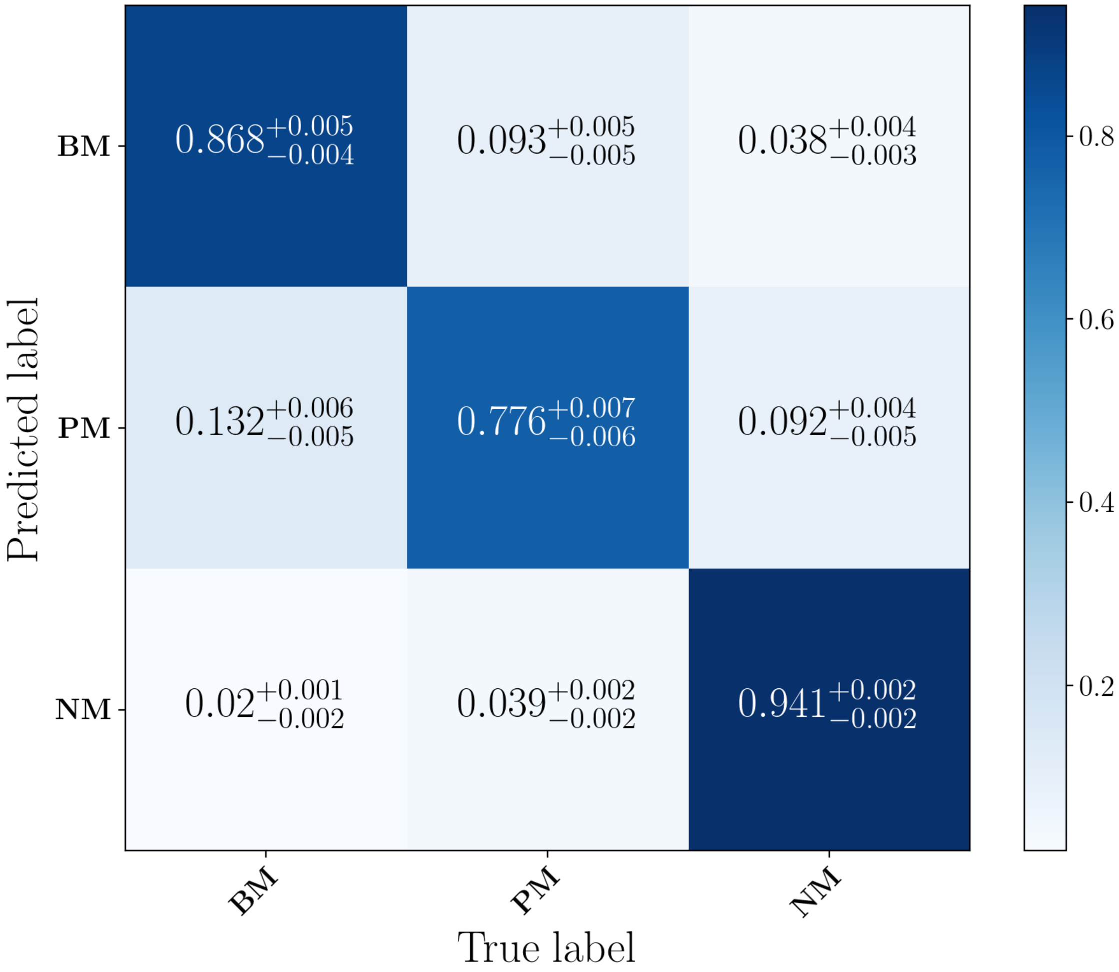


Figure 2.3: Class probability distribution of IllustrisTNG galaxies in the validation sample for each class in bins of 0.1 probability. This shows that our network has high confidence in the **NM** classifications whilst the probability distribution for the merger classes are more spread out. There is also a discrepancy between **BM** and **PM** in P $>$ 0.9, a sign that the **PM** class is the case that the network is less sure about, which has more ambiguity among the other types.

the classifier is fairly balanced between **MM** and **NM**, which is expected since the distribution of our simulation data is balanced. However, when comparing merger sub-classes, the distribution is skewed towards **BM**, as the network is less sure about **PM** classifications.

The class probability distributions shown in Fig. (2.3) are not enough to draw conclusions about our CNN's performance, we further explore performance metrics with our validation sample. We evaluate our hierarchy of models by looking at its normalized confusion matrix, which is shown in Fig. (2.4). The confusion matrix gives us an overview of the performance of the model by comparing the predicted labels with the true labels for each class. It shows this by listing the precision of each class in the diagonal, the fraction of correct classifications among all examples for the given class, while also showing the relative miss-classifications between each pair of classes. Our model is capable of identifying **BM** and **NM** types with 87% and 94% accuracy, respectively, with a contamination between both classes of less than 5%. However, in the **PM** case, the model has a lower performance, with 78% correct classifications with 13% contamination with **BM** and 9% contamination with **NM**. Even though it has almost a 10% performance difference with the other classes, almost two thirds of its miss-classifications are still merger classifications. Also, as in some cases the morphology of **PM** systems have no clear distortions, we therefore expected it to have some degeneracy with **NM** galaxies, while this is not true for the **BM** and **NM** classes.

It is also useful to verify the model with other metrics, especially the Receiver Operating Characteristic curves (ROC curves) and Precision-Recall diagrams (Powers, 2011). These are important because they also take classification threshold into account, while the confusion matrix only uses one threshold specified before-hand (i.e predictions should be in binary form). In Fig. (2.5) we show ROC curves for each class in the left panel and the Precision-Recall curves in the right panel. Precision-Recall curves can also be thought of as Purity-Completeness diagrams, which are a more common convention in astronomy. As we are using Monte Carlo dropout, we have ways of estimating the uncertainty of our classifications. Due to this feature of our model, we can plot the mean curves for each diagram with confidence intervals. This can be seen in each of the plots in Fig. (2.5) by the shaded area, which represents $\pm 4\ \sigma$ from the mean of the model, shown as a solid line. For the ROC curves, this uncertainty is very small and all



Predicted label \ True label	BM	PM	NM
BM	$0.868^{+0.005}_{-0.004}$	$0.093^{+0.005}_{-0.005}$	$0.038^{+0.004}_{-0.003}$
PM	$0.132^{+0.006}_{-0.005}$	$0.776^{+0.007}_{-0.006}$	$0.092^{+0.004}_{-0.005}$
NM	$0.02^{+0.001}_{-0.002}$	$0.039^{+0.002}_{-0.002}$	$0.941^{+0.002}_{-0.002}$

Figure 2.4: The normalized confusion matrix for our classifier hierarchy. Each column represents the true labels for each class while rows represent the predicted class. The diagonal of a multi-class classifier present the precision for each class, while other cells show the contamination between each possible pair of classes. It is important to note that almost two thirds of the contamination of **PM** happens with **PM** being classified as **BM**, which is still a merger classification. Errors shown are measured with the Monte Carlo dropout. This confusion matrix is measured within our balanced validation sample and do not represent the performance of the method with real galaxies.

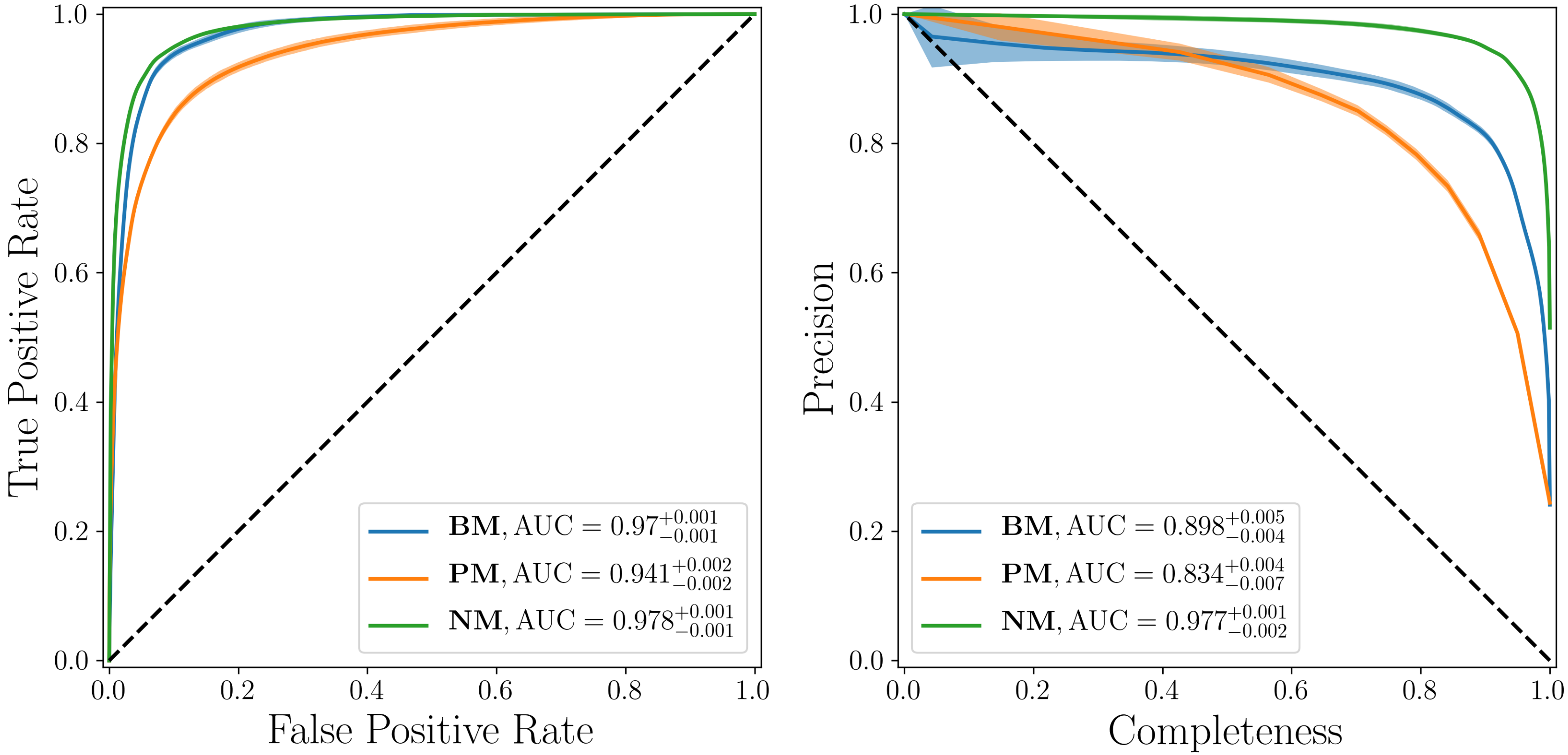


Figure 2.5: Performance metrics for classifications using the validation data. ROC curves for each class are shown in the left plot with **BMs**, **PMs**, **NMs** in blue, orange and green, respectively. The compromise between completeness and precision is shown in the right with the same color code. The performance shown here is based on the balanced validation sample, real galaxy samples will have very unbalanced configurations and hence this metric does not translate directly to applications on real galaxies.

classes follow a similar trend to what we might expect for a model with a confusion matrix equal to the one presented in Fig. (2.4). The area under the curve is also shown in the legend.

For the Precision-Recall diagram in the right panel of Fig. (2.5), it is possible to check that the uncertainties in our model are more apparent in the region of high precision. This is due to the fact that in this regime the threshold is very high, limiting the model to only very precise classifications. This results in smaller sets of classified galaxies, with very poor completeness, that are more prone to variability.

For visualization purposes, we plot a mosaic of images with galaxies randomly drawn for each class in Fig (2.6). Every galaxy plot shows the probabilities for the three classes, P(**BM**), P(**PM**), P(**NM**). Thus, as these galaxies are randomly selected, we also have cases that are miss-classifications. It is important to note that the threshold used here is the binary threshold, for probabilities $P > 0.5$, so this show the standard performance of the model, based on the confusion matrix of Fig. (2.4).

It is also useful to characterize each type of miss-classification produced by the network. In our case, this represents 6 different kinds of miss-classifications, one for each possible pair of classes in our three class hierarchy. We plot in Fig. (2.7) a panel of 15 miss-classified galaxies for each possible pair. The title of each panel refers to the true class, and what was the classification based on the probability from the model. Here, we see that the classifier uses very clear characteristics of merging for classifying galaxies as **BM**, as all galaxies misclassified as **BM** look as though they have two nuclei, or featuring two or more galaxies very close together. This even appear for **NM** systems classified as **BM**, a clue that our selection process for **NM** has some, even though small, contamination from galaxies with

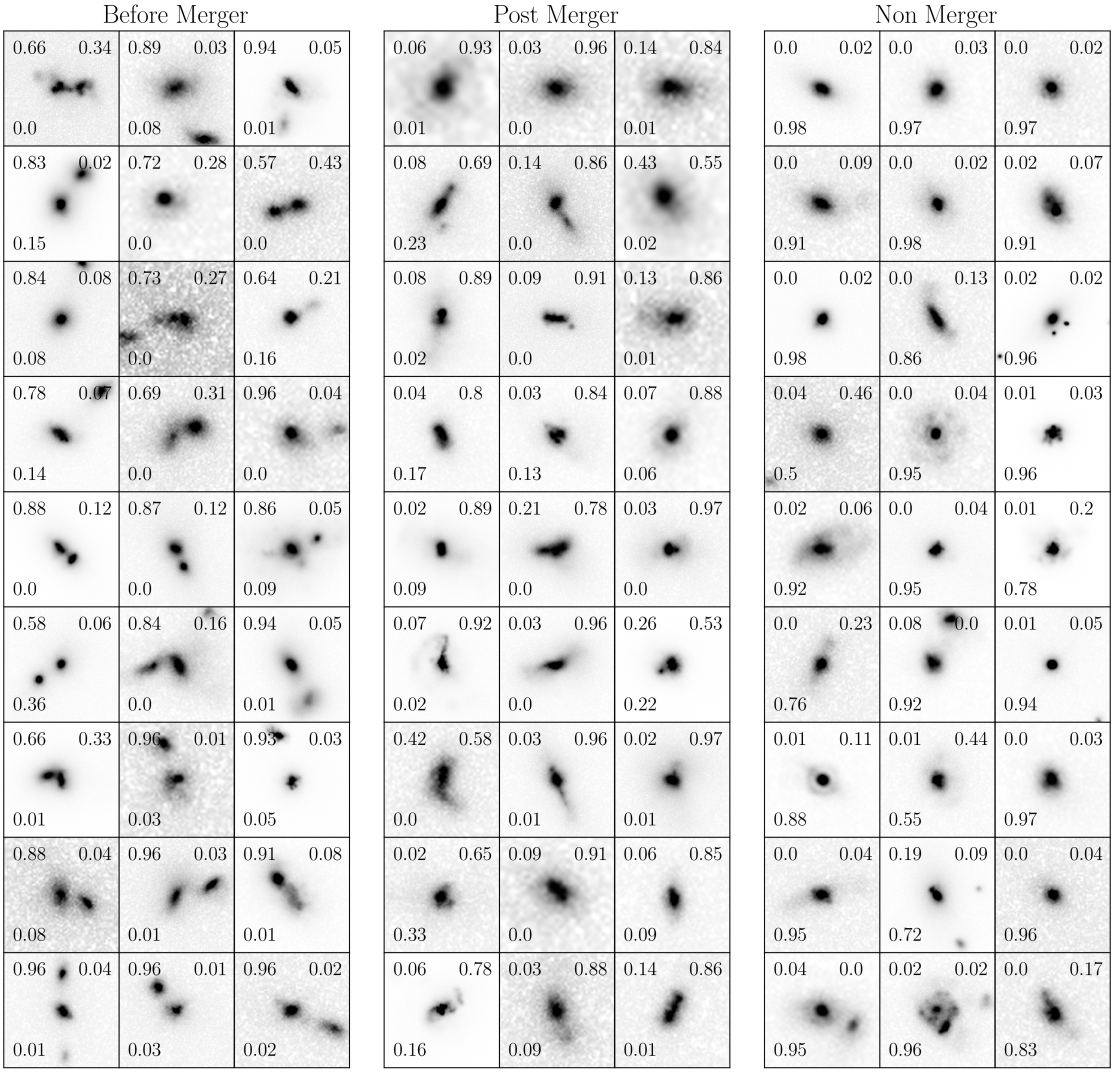
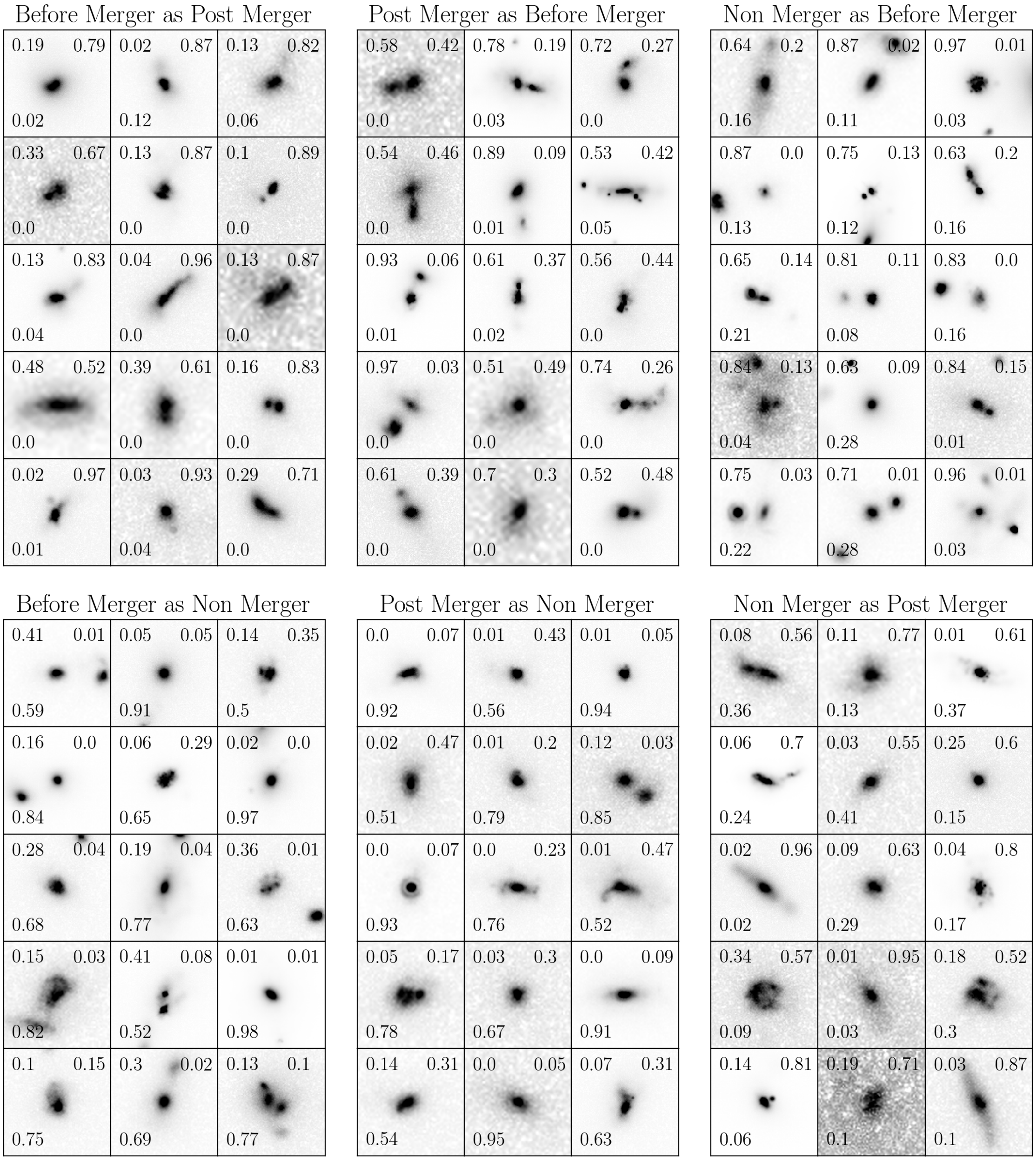


Figure 2.6: Mosaics for each class as classified by our model using simulated IllustrisTNG data. All galaxies were randomly drawn from the validation sample. In each galaxy image, all three probabilities are shown on each image. P(Before Merger), P(Post Merger), P(Non Merger), top-left, top-right and bottom, respectively. Varying signal-to-noise in the images are due to the varying intrinsic luminosity of the simulated galaxies or due to cosmological dimming.

close companions. It is possible that the selection is not accounting for some types of mergers. Likewise, galaxies misclassified as **NM** are in general more symmetric than their true counterparts. For instance, **BM**s classified as **NM** still show companions and some sort of interaction, but are more symmetric than most **BM** in Fig. (2.6).

We also see that **BM** systems classified as **PMs** show clearly signs of two nuclei, but for those which are closer together than regular **BM** systems. This is a sign of some degeneracy on the Sublink algorithm. Even if two galaxies are roughly in the same space, such that can still be regarded as two distinct galaxies. A similar pattern is seen in the case of **NMs** classified as **PMs**, as these non-interacting galaxies are more disturbed than their true counterparts. This shows us, overall, that the miss-classifications say a lot about how our model classifies a galaxy, as it follows properties that would also be used in visual classifications. Often, miss-classifications happen for cases where the morphology is really degenerate between classes, which would be expected. These are generally regarded as hard cases to learn, a natural limitation to the method based on visual structure, as they represent less than 3% of the training data which is not enough to represent significant shift in the weights of the model.

Yet another meaningful test is to generate images of pure random noise to check how our methods deal with images that are not representative of the parameter space we are interested in. As the model has to assign probabilities that sum to 1 to any image given to it, it will by design likely classify a random noise image as one of the possible classes. By generating a relatively large sample of random noise images we can inspect the output probabilities to check the behavior of the network in this case. To do so we generate 1000 random images within two filters



Before Merger as Post Merger

Post Merger as Before Merger

Non Merger as Before Merger

Before Merger as Non Merger

Post Merger as Non Merger

Non Merger as Post Merger

Figure 2.7: Mosaics for each possible case of miss-classification in the simulated IllustrisTNG data. Each title describes what is the truth class being miss-classified as a different class (truth class as wrong class) on given panel. All galaxies were randomly drawn from the validation sample for each specific case. In each galaxy image, all three probabilities are shown in each plot. P(BM), P(PM), P(NM), top-left, top-right and bottom, respectively. Varying signal-to-noise in the images are due to the varying intrinsic luminosity of the simulated galaxies or due to cosmological dimming.

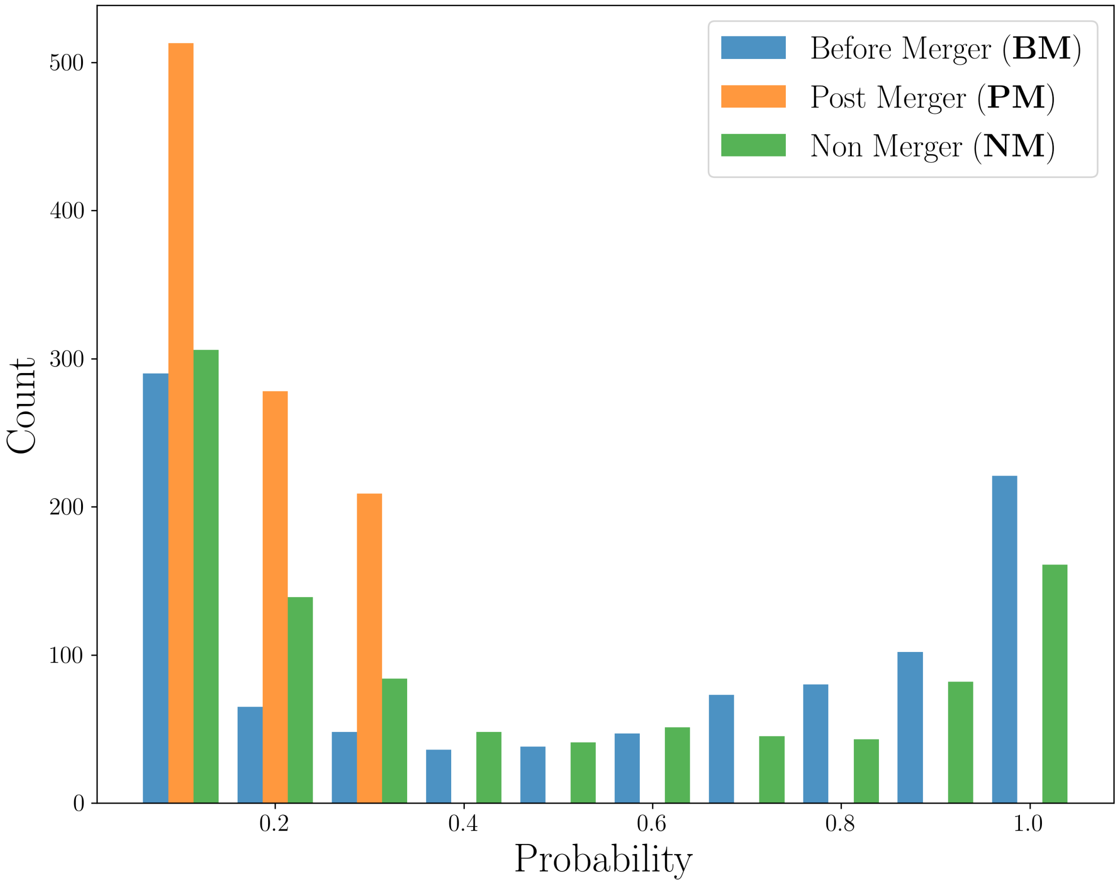


Figure 2.8: Mean Posterior Probabilities for all images in the random noise sample. Our hierarchy of model tends to classify most of the random noise images as **BM** and **NM** while none of the high probability noise images are classified as **PM**.

each[5], representative of the filters of our regular input data, and feed it to the network. We explore the probability distribution of each class in Fig. (2.8).

These probabilities show that our model tends to classify $\sim 60\%$ of the noisy images as **BM** and $\sim 40\%$ as **NM**. This is a good sign, as we have two opposite classes that show a similar behavior towards noise. The network did not classify any of the input random images as **PM**, where the maximum probability among all classifications was $P(PM) = 0.48$. This means that we can be fairly secure that miss-classification of **PM**s due to image quality effects, like noise, will be rare.

---

[5] We also investigated completely random noise and different images for each filter and the same random noise for both filters, with similar results.

Finally, we assess how the presence of crowded sky regions impacts our model classification. Bottrell et al. (2019) shows that the presence of contamination from neighboring sources is important during training when using simulated galaxies at low redshift. To show if this statement is true for the data used here, we retrain our model with a new dataset of simulated galaxies prepared with random patches of the sky from the CANDELS fields. These random regions are selected by searching for places that are centrally empty but have neighbor sources around the center.

The confusion matrix displayed in Fig. (2.9) shows that in this situation the classification precision of **BMs** slightly improves from 87% to 91%, whilst **PMs** and **NMs** decrease, from 78% to 67% and 94% to 92%, respectively. Even though our results for the presence of crowded backgrounds diverge from what is shown in Bottrell et al. (2019), we attribute it to the difference in scope of our data. We probe higher redshifts ($0 < z \leq 3$) and different wavelengths with simulated galaxies from cosmological simulations, which have lower resolution than galaxy-galaxy simulations. This experiment, however, shows that in crowded regions we should expect our model to display worse performances for **PMs**. In the case of galaxies in the CANDELS fields, we are selecting small fields of view and expect low contamination from crowded regions. As the overall results are worse with crowded regions of the sky, we conduct the rest of the Chapter with the class hierarchy trained with the original dataset.

It is important to note, however, that all performance metrics shown in this section are valid within the scope of our simulation validation sample. This needs to be taken into account when applying our classifier hierarchy to real data, as we expect to have an unbalanced sample of BMs, PMs and NMs. As we do not have ways to directly assess the performance of this classifier in the real data, we have to make comparisons with visual classifications and galaxy merger rates to test it.

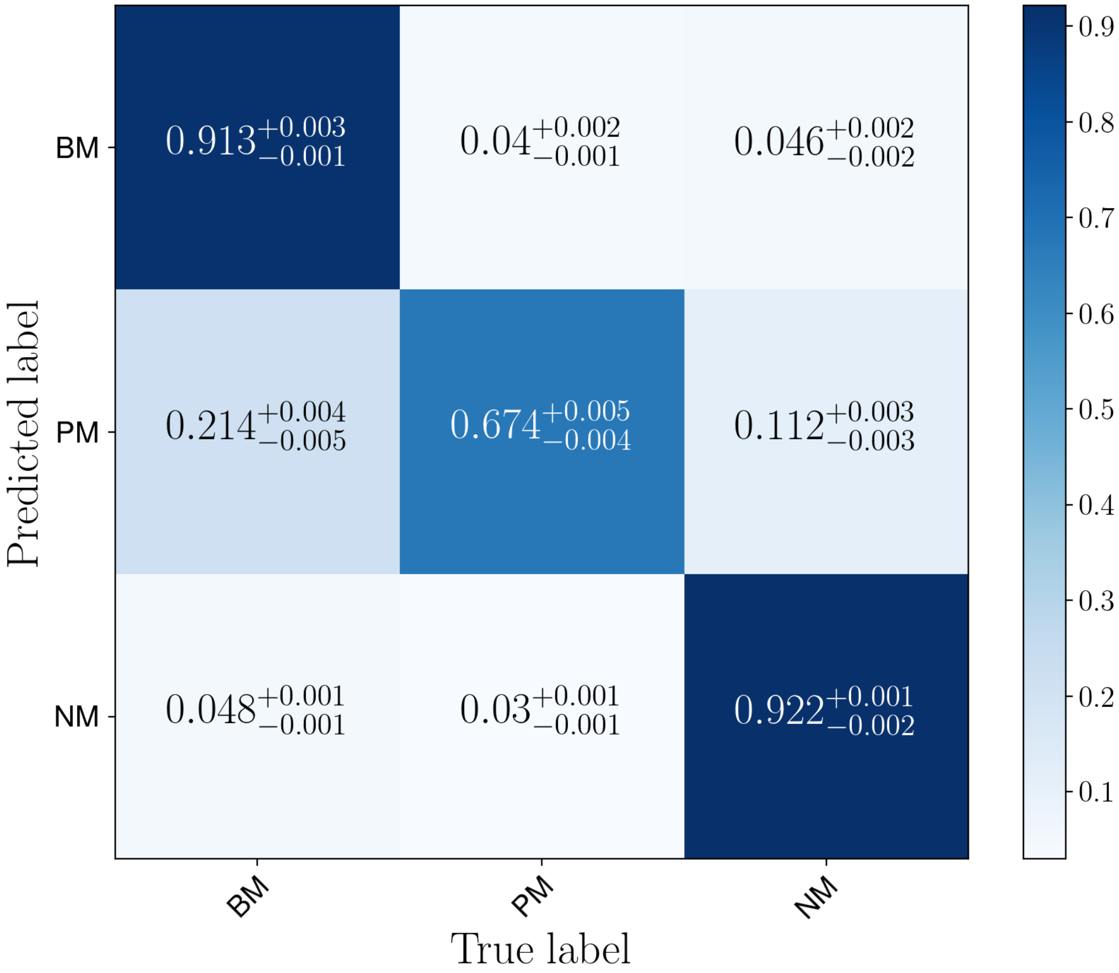


Figure 2.9: The normalized confusion matrix for our classifier hierarchy trained with simulated galaxies included in crowded patches of the sky from the CANDELS fields. Each column represents the true labels for each class while rows represent the predicted class. The diagonal of a multi-class classifier present the precision for each class, while other cells show the contamination between each possible pair of classes. It is important to note that almost two thirds of the contamination of **PM** happens with **PM** being classified as **BM**, which is still a merger classification. Errors shown are measured with Monte Carlo dropout.

2.5.2 Predictions on CANDELS

We test our methodology on CANDELS imaging data described in §2.3.2. For predicting classes on real data, we use an independent indicator to check if the observed galaxies are mergers or not. We rely on the visual classification of the CANDELS fields conducted in Kartaltepe et al. (2015), where detailed information about the morphology is available. Using this, we have a set of indicators that can help us decide if the galaxy looks like a merger or not. With this subsample of CANDELS galaxies that have similar properties to our simulation galaxies, we carry out predictions in the same way as we do for the validation data, as shown in Fig. (2.10). However, it is important to keep in mind that these visual indicators are not ground truths and are prone to the subjectivity of the classifiers. The apparent morphology of a galaxy merger can be produced by other physical processes.

Visual Classification

The Kartaltepe et al. (2015) classification effort on CANDELS galaxies includes a set of indicators dedicated to describe galaxy mergers, with the goal to develop a group of characteristics only related to merging aspects of the morphology of the galaxy. Here, in order to assess how our model performs using real CANDELS galaxies, we compare how its classification relates to these indicators.

Namely, we use the classification fractions `f_any`, `f_int1`, `f_int2`, `f_none`, `f_merger`, `f_comp`, plus two indicators that are not in the set of merger indicators but might relate to mergers, `f_tadpole` and `f_irr`. These fractions represent the overall fraction of total classifiers that marked the galaxy with given property. We briefly discuss each of these indicators here, for a full discussion please refer to

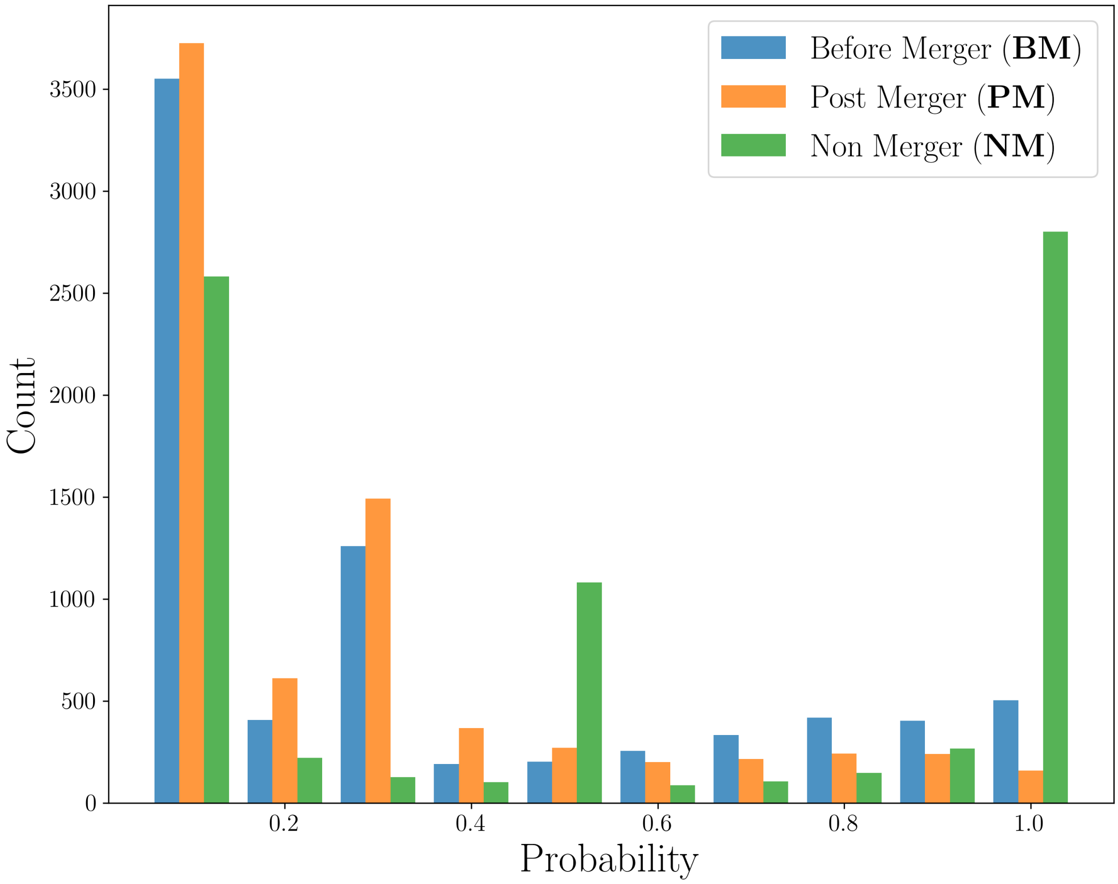


Figure 2.10: Probability distribution for the three classes that are classified by our hierarchy of models in the CANDELS selected sample. Overall these distributions are very distinct from the validation data. Here they are more irregular, especially those with intermediate confidence probabilities. This shows signs that the network is less certain about the classes in general than with was in the validation sample. This is expected since the validation sample is prepared to look very similar to but it is not equal to the CANDELS data.

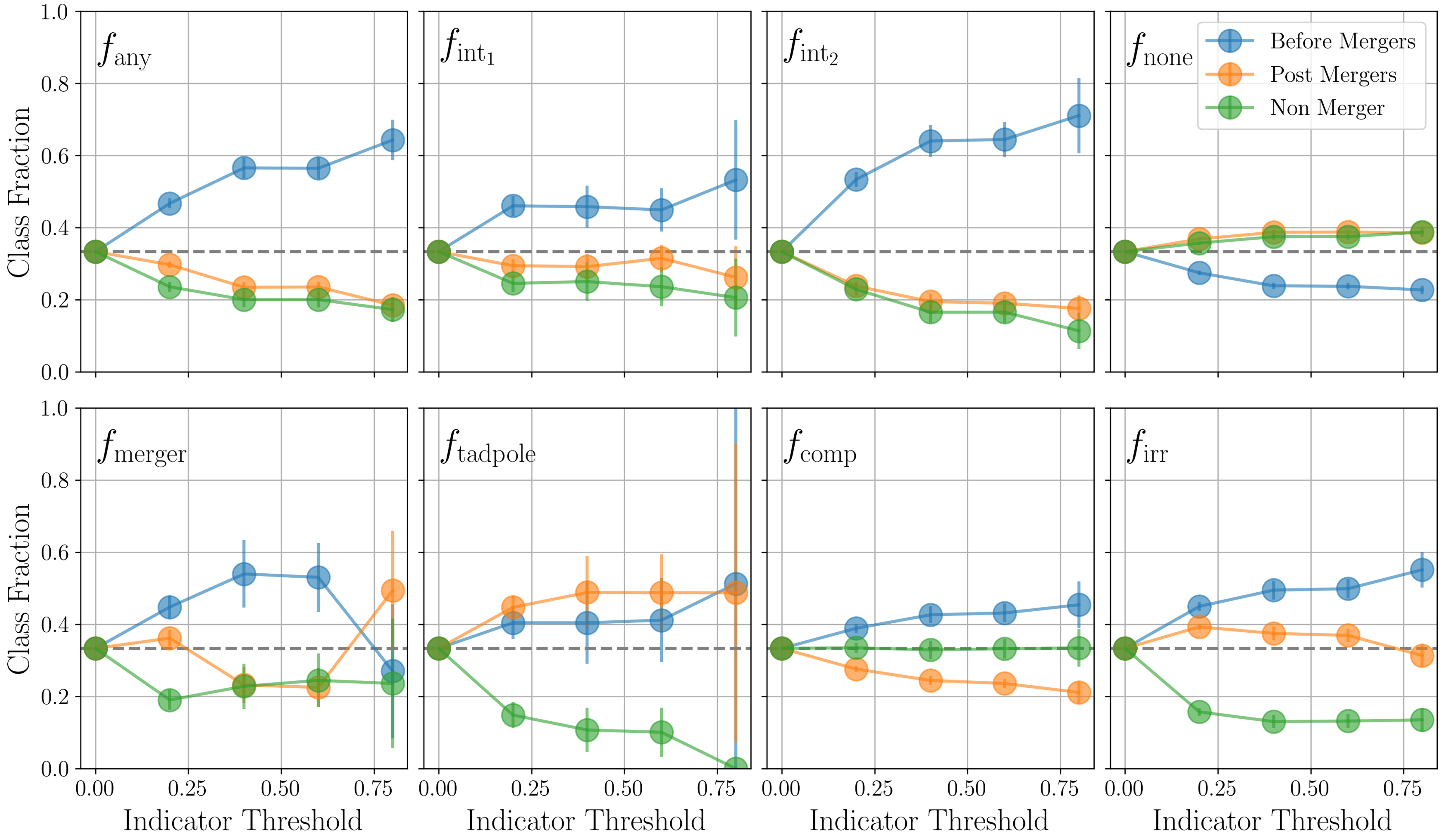


Figure 2.11: Mean class fractions from 100 samplings of a class balanced sub-sample (700 galaxies of each class) of CANDELS galaxies with the given indicator from visual classifications above the shown threshold. The first point represent the mean of the complete sub-sample of evenly distributed classes, while following points show only the fraction of those galaxies above the threshold. Error bars show $1 \pm \sigma$ for class fractions among all samples. **BM**, **PM** and **NM** are displayed in blue, orange and green, respectively.

Kartaltepe et al. (2015).

`f_any` is used when the galaxy has any type of interaction. Usually, if a classifier marked a galaxy in any of the other indicators, it will also be marked with `f_any`; `f_int1` represent galaxies with interactions within their segmap, while `f_int2` is for galaxies with interactions beyond their segmap; `f_none` is used when the galaxy has no signs of interaction and `f_merger` when the galaxy look like it underwent a recent merger event; `f_comp` indicates if the galaxy has a non-interacting companion, with no signs of interaction and tidal features; The other two non-merger indicators, `f_tadpole` and `f_irr`, represents whether the galaxy look like a tadpole galaxy with strong tidal features, or if the galaxy has an irregular morphology, which in general might be a sign of merging, but not uniquely. So each indicator represents the fraction of classifiers that mark the galaxy as having the assigned characteristics. Thus, this fraction is related to how obvious and how unified the classification was among all expert classifiers. A fraction of 0 represents a galaxy that no classifier marked as having those characteristics, while a fraction of 1 represents the cases where all classifiers marked the galaxy with the given indicator. Intermediate fractions might result from morphologies that are ambiguous, thus objects with higher fractions represent less ambiguous morphologies. However, it is important to none that for some indicators very few objects were unanimously classified. Thus these indicators are subject to the subjectivity of the classifiers, while a higher fraction means that the classification is less prone to biases.

To explore how our model's classification of CANDELS galaxies correlates with the visual classification available from Kartaltepe et al. (2015), we randomly generate 100 balanced sub-samples based on the model classification with 700 galaxies in each class. We do this as our resulting sample of CANDELS classified galaxies is very imbalanced towards non-mergers as shown in Fig. (2.10). If we use the entire sample, trends in our class fraction would be more difficult to visualize, especially for the case of **PMs**, which consists of the class with the fewer number of classified objects. We then compare each sub-sample against increasing thresholds within the given indicator. Fig. (2.11) show the class fraction mean ±1 σ for each class among all sub-sample for an increasing threshold. The **BMs** are shown in blue, **PMs** in orange and **NMs** in green.

The overall trend with all merger indicators (`f_any`, `f_int1`, `f_int2`, `f_merger`) is dominated by an increase in the fraction of **BM** classifications, as one would expect. Plus, the fraction of **PMs** do not follow this trend with **BMs**, a sign that both classes represent different objects. Indeed, by solely following these merger indicators, one might assume that **PM** and **NM** represent the same type of objects since `f_none` shows the fraction of **NM** and **PM** to be similar. However, `f_tadpole` and `f_irr` show similar trends for **BM** and **PM**. In this case, **PMs** classified by our model might represent galaxies without companions and clear signs of recent merger interactions by disturbed morphologies. Meanwhile, `f_comp` show different behaviors for each class with a very small scatter, which suggest that **PM**s as classified by our network are isolated galaxies, with no clear signs of companions, while **NM** can have companions but no signs of interactions. This might represent a bias from the network towards objects without any companion in the field, which indicates that **BM** might have a significant impact from sky projections. On the other hand, this is expected since we do not factor in any redshift information in the central and neighbor galaxies in our classification method. The introduction of this information in the classification pipeline might further improve the quality of the model, but this is left for a future work.

In Fig. (2.12) we show CANDELS galaxies as classified by our method with corresponding probabilities for each class, similarly to Fig. (2.6).

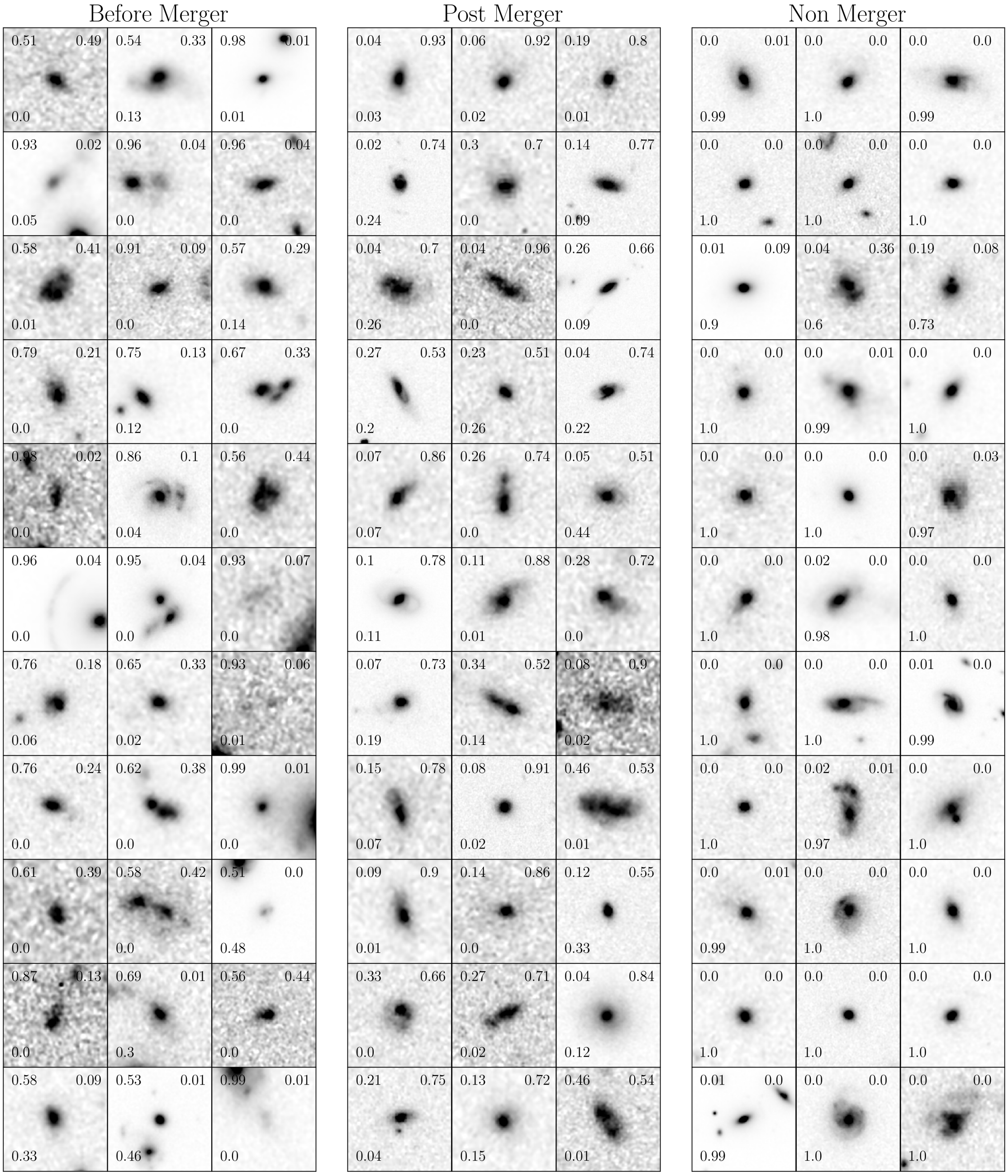


Figure 2.12: Mosaic with classifications done on CANDELS data for each class, **BM**, **PM** and **NM**, respectively. Mean probabilities for each class are shown in each image, top values represent merger classes (**BM** and **PM**) while bottom value represents the **NM** probability. The low probabilities represent cases where the network is more unsure and appears ambiguous. Increasing the probability threshold would produce more precise classifications with more clearly distinct morphologies, but we display here classifications above 50% probability as this represents the peak completeness of our classifications and the threshold used throughout this Chapter.

Merger Fractions and Merger Rates

One of our main goals in this Chapter is to estimate galaxy merger fractions, f_{m} and galaxy merger rates, $\mathcal{R}$, with our CNN method. We proceed to estimate f_{m} by counting merger classifications with probabilities $P(\text{class}) > 0.5$ in $\Delta z = 0.5$ bins of redshift in the range $0.5 < z < 3$. We do this for both merger sub-classes, **BM**, **PM** and also for **MM**. Even though we train our model with low redshift galaxies, our CANDELS samples have only a few galaxies with redshifts $z < 0.5$, which results in poor statistics for merger fractions in that regime. The measured merger fractions we derive are shown in Table (2.3).

We estimate galaxy merger rates by using merger fractions and appropriate timescales for each class, with $\tau_{\mathrm{obs}} = 0.3$ Gyr for **BM** and **PM**, and $\tau_{\mathrm{obs}} = 0.6$ Gyr for **MM**. Our timescales are defined by our sample selection steps, as described in §2.3.1. Although a consistent merger rate measurement does not validate individual classifications, it would represent that the overall statistics of the sample of classifications would follow one expected from other classification methods. By comparing merger rates estimated by our method with previous results we demonstrate a real application of our approach.

Redshift	BM	PM	MM
$0.5 \leq z < 1.0$	0.041 ± 0.008	0.014 ± 0.004	0.055 ± 0.009
$1.0 \leq z < 1.5$	0.048 ± 0.009	0.059 ± 0.010	0.107 ± 0.013
$1.5 \leq z < 2.0$	0.110 ± 0.016	0.084 ± 0.014	0.196 ± 0.021
$2.0 \leq z < 2.5$	0.180 ± 0.032	0.112 ± 0.026	0.292 ± 0.037
$2.5 \leq z < 3.0$	0.181 ± 0.043	0.206 ± 0.044	0.383 ± 0.052

Table 2.3: **BM**, **PM** and **MM** fractions in bins of redshift based on the classification from our models.

We estimate merger rates using our model by simply taking our merger fractions averaged over our timescale, that is

$$\mathcal{R} = \frac{f_m}{\tau_{\rm obs}}. \tag{2.4}$$

We plot our estimated merger fractions and rates in Fig. (2.13), in the left panel and right panel respectively, comparing with the results of merger fractions and rates as estimated with CANDELS galaxies from Mundy et al. (2017) and Duncan et al. (2019).

One important point is that our model was not prepared to measure merger fractions by construction, as it was trained with a balanced sample of mergers and non-mergers. Additionally, no redshift bias for mergers was used. In fact, the redshift distribution of our training sample is also balanced between mergers and non-mergers (Fig. 2.2).

It is possible to check in Fig. (2.13) that our results are in general consistent with merger rates found by Mundy et al. (2017) and Duncan et al. (2019). Here, even though we are making comparisons to close pairs statistics results, we do not make any assumptions on the fraction of pairs that will actually merge, C_{pair}, in $\mathcal{R}$ as all galaxies considered as mergers in our training sample are actually mergers, as we use information from IllustrisTNG's merger trees. Moreover, based on our selection approach, we are also not introducing information about the simulation's intrinsic merger rates into our model.

We fit power laws to our merger fractions and rates of the form

$$f_m(z) = f_0 \times (1+z)^m \tag{2.5}$$

$$\mathcal{R}(z) = \mathcal{R}_0 \times (1+z)^m, \tag{2.6}$$

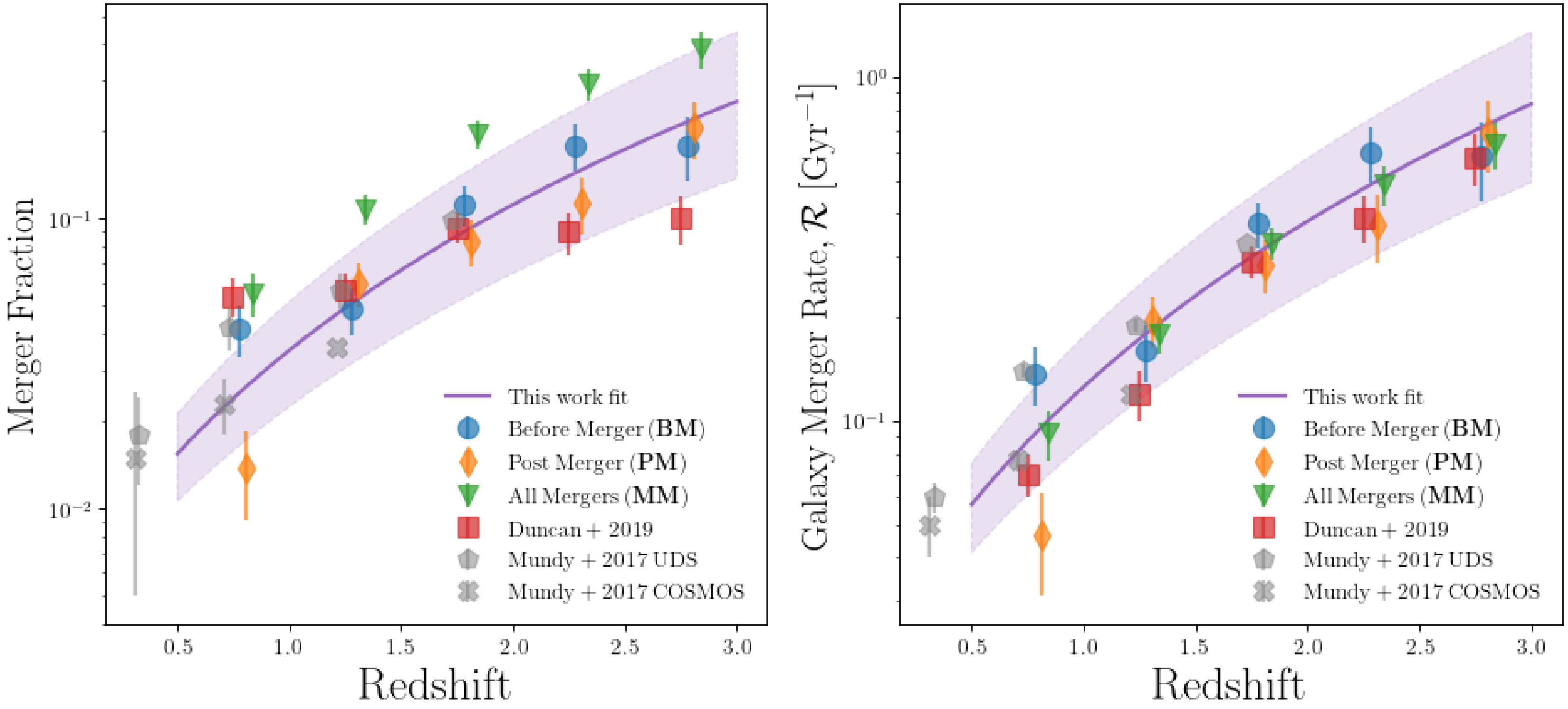


Figure 2.13: Merger fractions f_m (left) and galaxy merger rates $\mathcal{R}$ (right) in bins of redshift for our **BM** (blue circles), **PM** (orange diamonds) and **MM** (green triangles) classifications. Error bars represent ±1 σ uncertainties and account for the accuracies displayed in the confusion matrix in Fig. (2.4). We fit a power law for fractions and rates and show the best fit in purple together with ±1 σ uncertainties of the fit in the shaded area. We show results from Duncan et al. (2019) (red squares) and from Mundy et al. (2017) (gray X's and hexagons) for comparison. Overall, the trend estimated by our model agrees very well with previous results.

to our merger fractions and rates respectively. We do this fit by a simple least squares fit to all our data points, including **BM**, **PM** and **MM**, and show the uncertainty based on ±1 σ (shaded region in Fig. 2.13). We find

$$f_m(z) = 0.005 \pm 0.001 \times (1+z)^{2.85\pm0.24}, \tag{2.7}$$

and

$$\mathcal{R}(z) = 0.02 \pm 0.004 \times (1+z)^{2.76\pm0.21}, \tag{2.8}$$

which is expected since our observing timescale, $\tau_{\rm obs}$, is flat and defined by our selection (§2.3.1). These fits were done to all our data points (**PM**, **BM**) as well its associated error bars. We extract the uncertainties from the covariance matrix of the fit, and use it to show confidence levels of 1 σ in the purple shaded area, based on the lower and upper bounds of the power law two parameters. The lower limits are drawn from

$$f_m(z) = 0.004 \times (1+z)^{2.61}, \tag{2.9}$$

and

$$\mathcal{R}(z) = 0.0196 \times (1+z)^{2.55}, \tag{2.10}$$

while the upper limits are from

$$f_m(z) = 0.006 \times (1+z)^{3.09}, \tag{2.11}$$

and

$$\mathcal{R}(z) = 0.0204 \times (1+z)^{2.97}. \tag{2.12}$$

Overall this shows that the trend represented by our findings using major merger classifications by a deep learning model agrees with the trend found by Duncan et al. (2019) using close pair statistics for all the CANDELS fields, where within the redshift range probed here $0.5 < z < 3$, the highest merger rates, $\mathcal{R}$, are found in the highest redshift probed. Different assumptions regarding timescales and a different method of identifying mergers yield similar results, and even though our uncertainty is larger at all redshifts, the mean of our classifications match pairs well.

We cannot probe higher redshifts with our current model as it is limited by our training data, which was prepared to probe redshifts up to $z = 3$ with observed near-infrared data. One could expand the model to probe higher redshifts by training it with rest-frame UV data, but in this case the effects of dust and the lack of a radiative transfer treatment would become more important and the training sample should be prepared in a different manner, however this will be examined in a future study.

2.6 Summary

In this work we show that it is possible to train deep learning models to find galaxy mergers using only simulated galaxies and then to carry out predictions on real data by training a deep learning Convolutional Neural Network (CNN) model. We do this by classifying galaxy mergers with IllustrisTNG data and then carrying out predictions on real CANDELS galaxies. We show that

- Using automated methods for optimizing deep learning hyperparameters is a good way of achieving high performance architectures for solving astronomy classification tasks. This not only speeds up the training step of working with deep learning networks, but removes some of the subjectivity present when fine tuning such hyperparameters by hand.

- It is possible to train a model capable of achieving ∼90% accuracy in classifying galaxy mergers within the simulated balanced validation sample. Not only that, but our model can classify mergers in two stages: mergers before the merger event (**BM**) and post mergers **PM**, with 87% and 78% accuracy, respectively. The performance of the model using simulated galaxies from IllustrisTNG does not directly translate to the same performance that would be achieved using real galaxies, as the validation sample is balanced in the simulation, which is not true in our CANDELS sample. The quality of the model with real galaxies must be assessed by the visual classification comparison and the estimated galaxy merger rates.

- We show that predictions using real galaxy images are possible, and galaxies classified in the validation and CANDELS samples share similarities. We show that our model classifications follows visual classification indicators for mergers from Kartaltepe et al. (2015). Even though merger classifications can be ambiguous between visual classifiers, our blind classifications based on the information from mergers trees from the IllustrisTNG show that galaxy mergers classified by our network have similar visual cues to those classified by visual experts. This is shown by the different trends for mergers before the merger event, post mergers and non-mergers when compared to merger indicators from visual classifications. Galaxies before the merger event (**BM**) dominate samples selected with higher thresholds of the merger indicators from the visual classification.

- By using our model to classify CANDELS galaxies we measure galaxy merger fractions and rates between $0.5 \leq z \leq 3$ that are consistent with previous results for CANDELS galaxies estimated with close pair statistics from Duncan et al. (2019). This was done without any prior merger fraction or rate information embedded in our training step. Our model, by construction, was not prepared to do such measurements and this is an independent method of estimating merger fractions and rates, even though the uncertainties are higher than when using other methods.

Our results are based on a sample of simulated galaxies with several constraints: our mocks do not account for the effects of dust, we do not explore arbitrary orientations besides face-on and edge-on orientations, and our results are only limited to massive galaxies with $M_* > 10^{10} M_\odot$. Addressing these points will further improve results when carrying out predictions on real galaxies, as it would serve to lessen the gap between simulated and real galaxies. This approach is limited by the quality of the training data, and improvements in the post-processing of the simulation data should further improve the results displayed here. It is of utmost importance to always use large training samples, as the parameter space in the training step is crucial for the learning of the model.

This work shows the potential of using a combination of galaxy simulations and machine learning techniques as an avenue for solving problems where observables are impossible or expensive to estimate from real observations of galaxy mergers. Approaches like the one presented here will naturally improve alongside cosmological simulations.

Chapter 3

A Simulation-driven Deep Learning Approach for Separating Mergers and Star-forming Galaxies: The Formation Histories of Clumpy Galaxies in All of the CANDELS Fields

The content of this chapter has been published (Ferreira et al., 2022a) in The Astrophysical Journal.

3.1 Abstract

Being able to distinguish between galaxies that have recently undergone major-merger events, or are experiencing intense star formation, is crucial for making progress in our understanding of the formation and evolution of galaxies. As such, we have developed a machine-learning framework based on a convolutional neural network to separate star-forming galaxies from post-mergers using a data set of 160,000 simulated images from IllustrisTNG100 that resemble observed deep imaging of galaxies with Hubble. We improve upon previous methods of machine learning with imaging by developing a new approach to deal with the complexities of contamination from neighboring sources in crowded fields and define a quality control limit based on overlapping sources and background flux. Our pipeline successfully separates post-mergers from star-forming galaxies in IllustrisTNG 80% of the time, which is an improvement by at least 25% in comparison to a classification using the asymmetry (A) of the galaxy. Compared with measured Sérsic profiles, we show that star-forming galaxies in the CANDELS fields are predominantly disk-dominated systems while post-mergers show distributions of transitioning disks to bulge-dominated galaxies. With these new measurements, we trace the rate of post-mergers among asymmetric galaxies in the universe, finding an increase from 20% at $z = 0.5$ to 50% at $z = 2$. Additionally, we do not find strong evidence that the scattering above the star-forming main sequence can be attributed to major post-mergers. Finally, we use our new approach to update our previous measurements of galaxy merger rates $R = 0.022 \pm 0.006 \times (1 + z)^{2.71 \pm 0.31}$.

3.2 Introduction

The first deep Hubble Space Telescope (HST) images of the distant universe revealed that many distant and faint galaxies are in fact irregular/peculiar in appearance (e.g., Williams et al., 1996). Because the first cameras on HST, WFPC1/WFPC2 were sensitive in optical wavelengths only, probing distant galaxies was limited to their rest-frame ultra-violet light, due to the effects of redshift. It was unclear whether the peculiar appearances were the result of observational limitations or real. The question thus remained whether the observed irregularities were in fact just the star forming areas of these galaxies, while the older stars remained below detection. When the NICMOS camera was launched in 1998 on HST it became clear that the morphologies of distant galaxies were peculiar in their rest-frame optical wavelengths as well, implying that the bulk stellar mass in these galaxies was indeed out of equilibrium (e.g., Dickinson et al., 2000; Papovich et al., 2005; Conselice et al., 2005; Mortlock et al., 2013; Whitney et al., 2021). The common consensus was that distant galaxies are indeed intrinsically peculiar. However, it remained unclear why and how this finding relates to the various possible modes that could be responsible for producing these irregularities in galaxy structures at high redshifts. The peculiar appearance is likely linked to the formation process of the galaxies, but details of the origin of the observed irregular structures have proven difficult to fully understand.

Since then, it has become clear that, overall, galaxies gradually transition from peculiar galaxies at higher redshifts to ellipticals and disc systems at lower redshifts (e.g., Conselice et al., 2003; Lotz et al., 2004; Mortlock et al., 2013; Huertas-Company et al., 2015). This conclusion was made possible with the advent of the WFC3 camera on HST which allowed astronomers to trace the morphological evolution of galaxies over large areas of the sky. Galaxies are therefore undergoing a transformation, and their irregular origins reveal clues about the processes which drive galaxy formation. One popular and well explored hypothesis is that these systems are in fact undergoing hierarchical mergers to form larger systems. The basic idea is that two galaxies in the early universe smash together to form a larger galaxy, a process which is predicted to be a critical element in the cosmological context of galaxy formation within a Cold Dark Matter (CDM) universe, with well defined predictions of this process (e.g., Jogee et al., 2009; Bertone & Conselice, 2009; Mundy et al., 2017).

Today, the merger rate can be accurately measured to high redshifts ($z \sim 3$) using galaxy structure (e.g., Conselice et al., 2003; Conselice et al., 2008; Man et al., 2016; Mantha et al., 2018; Ferreira et al., 2020; Whitney et al., 2021). Using e.g., CAS parameters, the measurements show that the merger rate increases at higher redshifts up to $z \sim 3$, such that $\mathrm{f}_{\mathrm{merger}} \sim (1+z)^{2-3}$ (e.g., Conselice, 2014), an evolution which scales similarly to the density of the universe, which evolves as $\sim (1+z)^3$. This implies that with identifications of mergers at both high and low redshifts, we are able to trace the galaxy merger history and investigate the role of mergers within the formation of galaxies over time (e.g. Conselice, 2006; Mundy et al., 2017).

In addition to high merger rates, distant galaxies have much higher star formation rates than today, peaking at $z \sim 2$ (e.g., Madau & Dickinson, 2014). We further know that galaxy structure is highly dependent on the star formation rate in the sense that intensely star forming galaxies generally appear more clumpy and irregular than quiescent galaxies at all redshifts (e.g., Windhorst et al., 2002; Guo et al., 2015; Mager et al., 2018; Guo et al., 2018; Sazonova et al., 2021). In fact, these two different types of galaxies – mergers and non interacting intensely star forming galaxies – can look very similar by eye, which complicates visual classifications. Even kinematically it can be challenging to distinguish mergers from rotating galaxies with high dispersions (e.g., Simons et al., 2019; Bottrell et al., 2021). In addition, classifications and selections of galaxies after a merger event (post-mergers) are highly contaminated by misclassified isolated galaxies with high specific star formation rates (sSFR). This is because their star forming regions and dusty inter stellar medium (ISM) can generate asymmetric features reminiscent of (post-) merger features. It is therefore currently unknown if and how we can correctly distinguish whether a galaxy is undergoing intense star formation, or some type of merger using galaxy structures and morphologies.

One way to approach this question is through novel techniques using machine learning. Recently, tremendous progress has been made in applying supervised deep learning methods to investigate galaxy morphology (e.g. Huertas-Company et al., 2018; Reiman & Göhre, 2019; Huertas-Company et al., 2019; Cheng et al., 2020; Martin et al., 2020; Huertas-Company et al., 2020; Walmsley et al., 2020, 2022). These end-to-end techniques are also very promising for investigating galaxy mergers specifically (Ackermann et al., 2018; Pearson et al., 2019; Pearson et al., 2019; Bottrell et al., 2019; Wang et al., 2020; Ferreira et al., 2020; Bickley et al., 2021), as they are able to extract features from the images in an automated fashion, without the need for feature processing prior to the classification task. Additionally, one can also leverage information not only from visual classifications and observations, but also by forward modeling cosmological simulations to the observational domain (Ćiprijanović et al., 2020, 2021).

We have recently started a machine learning exercise to determine the merger history of galaxies using cosmological simulation runs from IllustrisTNG (Vogelsberger et al., 2014; Pillepich et al., 2018b; Nelson et al., 2019). In Ferreira et al.

(2020) we were able to separate mergers from other types of galaxies in IllustrisTNG to a success rate of 90% up to $z \sim 3$. The present Chapter is a followup to Chapter 2, in which we now investigate whether it is possible to distinguish merging galaxies from intensely star forming galaxies (above the SFR-stellar mass main sequence). These galaxies have the lowest success rates in classifications from Ferreira et al. (2020). Our task in this Chapter is to correctly distinguish mergers from star forming galaxies by only using their morphology and structure.

This Chaper is organized as follows: in §3.3 we describe the data sets we constructed for this task, from IllustrisTNG (simulations) and CANDELS (observations). A description of the methods we used to train a Deep Learning model and how we measure the structure of the galaxies in our samples is given in §3.4. We present our results in §3.5 while a discussion on the implications is laid out in §3.6. Finally, we summarize and conclude our findings in §3.7.

3.3 Data

To test our new Deep Learning approach, we use simulated galaxies from cosmological simulations post-processed with the SKIRT (Camps & Baes, 2015, 2020) dusty radiative transfer code. The simulations are based on IllustrisTNG (Sec. §3.3.1) which are used for the construction of the training sample for a Convolutional Neural Network (CNN) that is subsequently applied to observed galaxies from the CANDELS fields (Sec. §3.3.3). Our sample definitions for post-mergers and star forming galaxies are given in Sec. §3.3.2. We discuss the pipeline used to generate CANDELIZED mock images from IllustrisTNG in Sec. §3.3.4.

3.3.1 IllustrisTNG

As already described in Chapter 2, IllustrisTNG is a suite of cosmological, gravo-magneto-hydrodynamical simulation runs with a diverse set of particle resolutions (Pillepich et al., 2018a; Naiman et al., 2018; Nelson et al., 2019; Springel et al., 2018; Marinacci et al., 2018; Nelson et al., 2019; Pillepich et al., 2019).

However, as opposed to the previous chapter, we use the higher resolution TNG100-1 simulation (intead of TNG300-1), which has proven to be a good compromise between resolution and volume[1]. TNG100 has already been used extensively in studies that analyze galaxy morphologies and structures, including the comparison between simulations and observations (Huertas-Company et al., 2019; Blumenthal et al., 2020), including as training datasets for deep learning (Wang et al., 2020; Bickley et al., 2021; Bottrell & Hani, 2022). Specifically, Zanisi et al. (2021) show that TNG100 galaxies reproduce observed objects well, especially disc-dominated sources. While there are some deviations in the small-scale structure of highly concentrated spheroidal systems, this is a minor issue in our analysis since they only make up a small fraction of our sample. In addition, our galaxies are resolution limited at the current redshift of interest, meaning that tiny details of structure are not relevant in this analysis.

To counterbalance any limitations from resolution we limit our analysis to galaxies with $M_* > 10^{9.5} M_\odot$. Above this limit, and at our explored redshift range $z > 0.5$, galaxies are represented by thousands of stellar particles. This enabled sampling of the simulated galaxies into resolutions comparable to that of the observed CANDELS data. Specifically, the gravitational softening length of the simulation, ϵ, is not a limitation when compared to the HST ACS and WFC3

---

[1] We tested our selections on TNG50-1 but resulting samples are too small for Deep Learning training.

cameras resolution. This is a lower stellar mass limit than the one used in Chapter 3 as we are levering the higher resolution TNG100-1 simulation run.

This approach is a noticeable refinement to our previous treatment in Ferreira et al. (2020), where the research question did not demand the resolution of fine morphological features like clumpy regions and tidal features which are required for the present analysis.

To select appropriate galaxies from TNG100-1, we isolate galaxies with $M_* > 10^{9.5} M_\odot$, in the redshift range $0.5 < z < 3$. To limit contamination in our sample, we use a minimum dark matter to total mass ratio of

$$\frac{M_{\rm DM}}{M_{\rm total}} > 0.1, \tag{3.1}$$

as a way to avoid subhalos created as a result of disk fragmentation. This means that at least 10% of the subhalo's mass needs to be in the form of dark matter. We acknowledge that this could also inadvertently remove galaxies that had their dark matter stripped, however this number is small and does not impact the final sample. This criteria removes $\approx 2\%$ of galaxies from the overall pool of available sources (subhalos) in TNG100.

We also remove objects that are smaller than the ACS PSF size from the selection. To identify these objects, we first convert the half mass radius $R_{1/2}$ provided in the simulation group catalogs in kpc, to a pixel scale based on the cosmological model adopted by IllustrisTNG and the ACS pixel scale

$$R_{1/2Mass,pix} = \frac{a(z)}{h} R_{1/2Mass,kpc}, \tag{3.2}$$

where $a(z)$ is the angular size at z and h the Hubble constant / 100. Any galaxy with $R_{1/2Mass,pix} < 3$ pix was then filtered out from our selections. This step removes $\approx 3\%$ of galaxies from the total pool of available sources.

3.3.2 Sample Definitions

Our goal is to separate star forming galaxies from post-mergers at intermediate to high redshifts based on their morphology. We define post-mergers as galaxies with at least one major merger event with a mass ratio

$$\mu = \frac{M_2}{M_1}, \quad \mu \geq 0.25, \tag{3.3}$$

where M_1, M_2 are the stellar masses of the galaxy pair involved in the merging event, ranked by their stellar mass respectively, with $M_1 > M_2$. Galaxies are considered post-mergers if they have coalesced into a single galaxy in the past 500 Myrs, where a single galaxy is represented by a subhalo in the simulation as identified by friends-of-friends algorithms (Rodriguez-Gomez et al., 2015). This selection window timescale is motivated by the observability timescales of disrupted structures caused by mergers identified by structure measurements in IllustrisTNG (Whitney et al., 2021), and are higher than what was previously used in Ferreira et al. (2020). We allow post-mergers to have low sSFRs. Their asymmetric features likely arise from the merging process rather than from star forming clumps. In contrast, non-interacting star forming galaxies are defined here as galaxies that have sSFRs above the star-forming main sequence (SFMS) at redshift $z \sim 0$ with the following threshold,

$$sSFR > 10^{-9.5}\ \mathrm{yr}^{-1} \tag{3.4}$$

and are not interacting with other galaxies. To isolate non-interacting cases, we exclude any galaxy from the simulation that had major or minor merger events ($\mu > 0.1$) around ±1 Gyr of its current redshift. Minor mergers are excluded completely from both definitions, and any conclusions presented in this Chapter should be considered with this in mind. Importantly, this selection is not intended to limit the non-interacting cases to extreme starbursting episodes alone, but to select non-interacting galaxies with sufficiently high sSFR to produce clumpy and asymmetric features that could be mistaken for merging signatures.

In summary, this selection results in a sample of $\sim 6,000$ post-mergers and $\sim 110,000$ non-interacting star forming galaxies. While this may be a realistic representation of actual fractions (only $\sim 5\%$ of the sample are post-mergers), training the network requires a balanced dataset. We thus use the post-merger sample as the baseline and separate it in bins of redshift, stellar mass, and size, randomly sampling the same number of non-interacting galaxies within each bin. We remove bins without adequate matched numbers of star forming galaxies. This becomes noticeable in the higher mass bins where post-mergers dominate and very few star forming galaxies are present.

After matching the samples, we count $\sim 4,000$ galaxies in each class as our final sample. A summary of this sample separated by class and redshifts is available in Table 3.1. The distribution of redshifts, star forming rates, stellar masses, and stellar half-mass radius are shown in Figure 3.1 for post-mergers in red and star forming galaxies in blue. Both classes have very similar physical properties, with a small excess of large, passive and massive post-mergers in comparison to the star forming galaxies. Additionally, the top right and bottom right panels of Figure 3.1 show the time since the last major merger event, τ, and the mass ratio, respectively, for post-mergers. The nature of the distribution for τ arises from the

Redshift	**Post-Mergers**	**Star Forming**	**Total**
$0.5 \leq z < 1.0$	1214	1167	2381
$1.0 \leq z < 1.5$	1082	1140	2222
$1.5 \leq z < 2.0$	847	881	1728
$2.0 \leq z < 2.5$	589	556	1145
$2.5 \leq z < 3.0$	333	321	645

Table 3.1: Summary of the initial IllustrisTNG sample. The numbers in this table present the sample before each galaxy was post-processed with SKIRT and CANDELIZER (see text for detail), during which each image was augmented by 20 for 4 orientations and 5 different fields.

average time between snapshots in the simulation of around ~ 0.15 Gyr. This timescale represents 1 to 3 snapshots after the coalescence of stellar masses.

3.3.3 CANDELS Fields

One of the main goals of this work is to predict star forming and post-mergers galaxies in the observed CANDELS imaging data (Grogin et al., 2011; Koekemoer et al., 2011), which comprises high-quality HST observations from COSMOS, UDS, EGS, GOODS-South, GOODS-North (Grogin et al., 2011; Koekemoer et al., 2011). CANDELS data has been already discussed in Chapter 2, and here we focus on how we select our new sample.

To select CANDELS galaxies we first remove all problematic objects according to their quality flags as recorded in the photometric catalogue and in the Kartaltepe et al. (2015) catalogues to avoid edges, artifacts and stars. Following Huertas-Company et al. (2016) and Kartaltepe et al. (2015), we then select galaxies with H-band magnitudes $H < 24.5$ mag. Because this cut can bias our sample against extended sources, we also include a signal-to-noise (SNR) lower limit of SNR > 50 to exclude any compact source with only a few bright pixels.

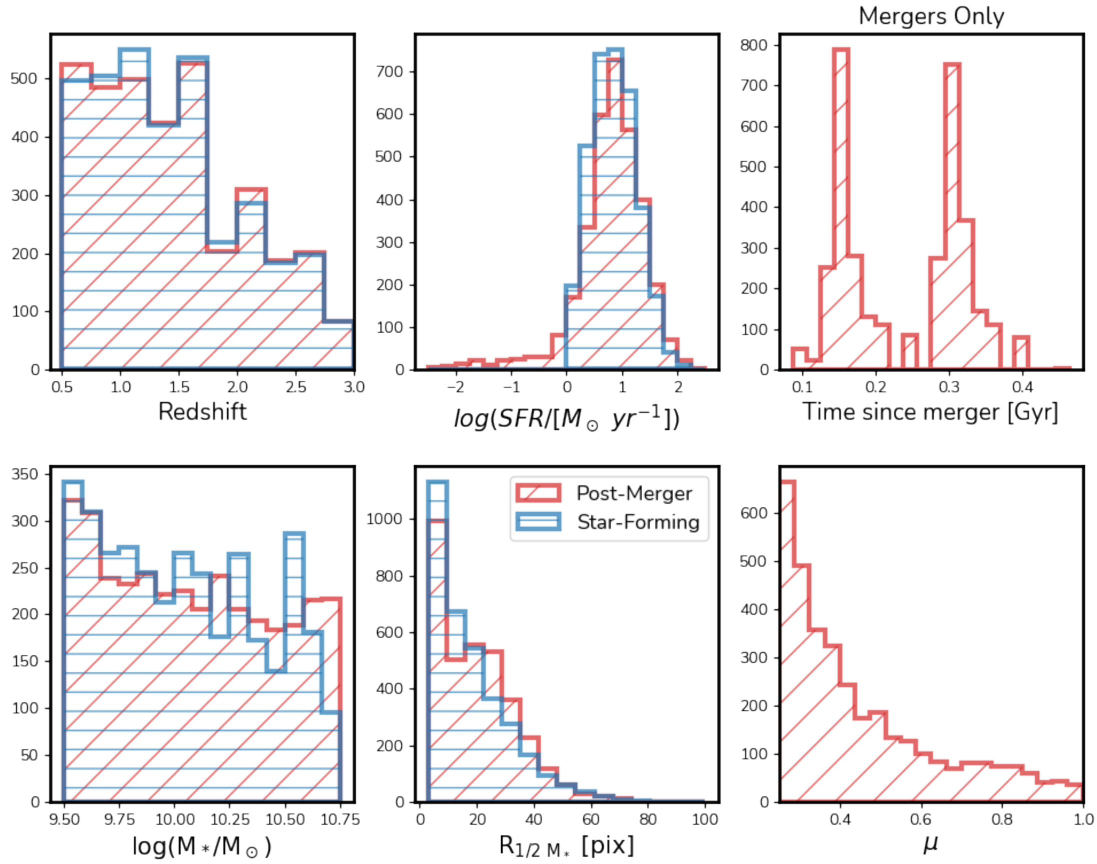


Figure 3.1: Physical properties of the 8,000 IllustrisTNG TNG100-1 selected simulated galaxies. For both types of galaxies we show distributions for redshifts (top left), star formation rates (top middle), stellar masses (bottom left) and stellar half-mass radius (bottom middle) in red for post-mergers, and blue for star forming galaxies. Distributions agree in general, with a small excess of stellar mass and size for the post-mergers. The time since the last major merging event and the mass ratio, μ, – properties unique to the post-mergers – are shown in top right and bottom right, respectively.

This magnitude cut removes 1074 sources, while the SNR cut further removes 430 sources. Then we proceed with the same cuts we used to select IllustrisTNG galaxies, using $0.5 < z < 3$ and $M_* > 10^{9.5} M_\odot$. We apply a final cut using the asymmetry ($A > 0.1$, Sec. §3.4.3) to remove regular unambiguous galaxies with no apparent disturbed or asymmetric features. This ultimately results in a sample of 23,494 galaxies from all the CANDELS fields combined.

Finally, we produce cutouts for I_{814}, J_{125} and H_{160W} bands centering on each selected CANDELS galaxy, each with a field of view of 50 kpc × 50 kpc, using photometric redshifts from Duncan et al. (2019), preserving relative sizes between galaxies. Importantly, this selection does not rely on size measurements that could easily be spurious in interacting or merging galaxies. We do not find any bias in our classifications that could be attributed to small changes of the field of view caused by the photometric redshift uncertainties.

3.3.4 Pipeline to produce CANDELIZED Mocks

In order to guarantee realistic representations of CANDELS galaxies in the simulated sample, we must include instrumental and cosmological effects to the images of the IllustrisTNG galaxies. An overview of the steps are shown in Fig. 3.2 and are detailed in this section. IllustrisTNG data holds information on the stellar, gas and dark matter particles for each source. Each particle represents a large physical region that can be described by rich stellar populations, that vary depending on age, mass and metallicity. The resampling of the star forming regions is particularly important to avoid problems with the coarse representations (Camps et al., 2016; Trayford et al., 2017).

To create mock broadband images, we thus process each stellar particle with

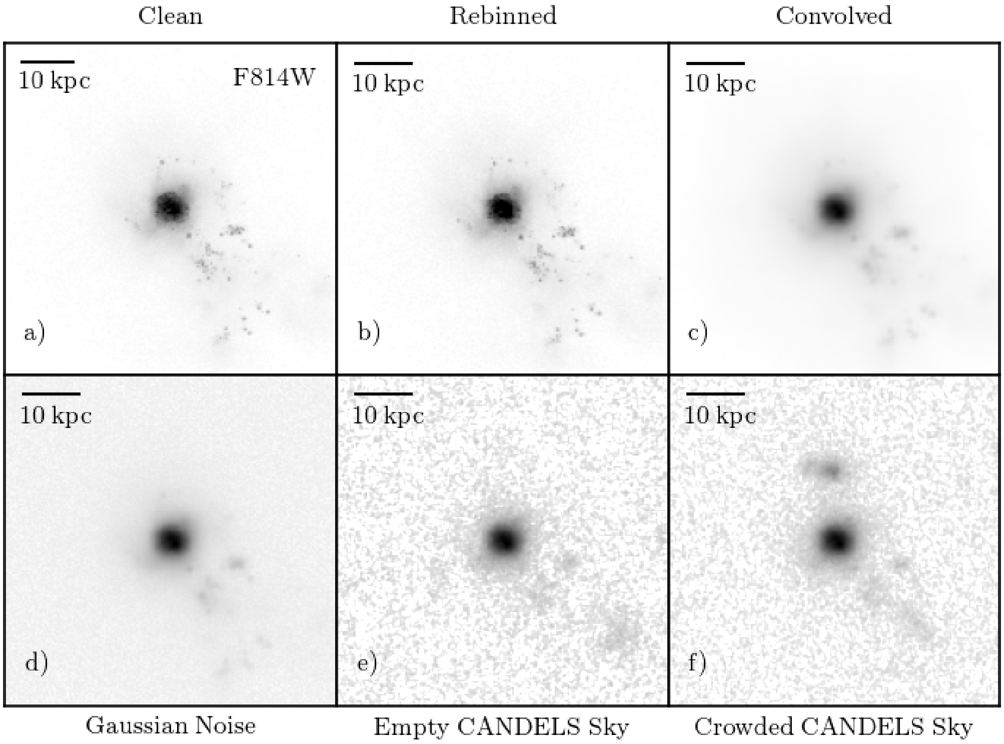


Figure 3.2: Example of the processing steps of our mock pipeline. **a)** Noiseless F814W broadband image generated from the simulated galaxy datacube with 0.03 $''/pix$ pixel scale. **b)** The same image after rebinning from z=0.5 to z=0.6. **c)** Image convolved by the HST F814W PSF. **d)** Image with Gaussian noise added. **e)** Image added on top of a random patch of the sky within a CANDELS field with no neighbouring sources. **f)** Image added randomly to a patch of sky with other sources in the field of view. As this patch of the sky is randomly selected, all final images have varying levels of contamination from nearby sources. We quantify this by the total flux in the sky patch before adding the simulated source to it.

a population synthesis model following the recipes from Trayford et al. (2017) & Vogelsberger et al. (2020). This entails post-processing the simulation data with the Monte Carlo dusty radiative transfer code SKIRT (Camps & Baes, 2015, 2020).

Each stellar particle in the simulation is considered as a Single Stellar Population (SSP) with GALAXEV (Bruzual & Charlot, 2003) or MAPPINGSIII (Allen et al., 2008) SEDs based on its stellar mass, absolute metallicity, and age. We choose to adopt these particular templates because, firstly, they are implemented in SKIRT and, secondly, they had been tested previously in similar pipelines to generate mock observations from cosmological simulations (Trayford et al., 2017; Rodriguez-Gomez et al., 2019). Finally, Bruzual & Charlot (2003) are also the templates used to derive stellar masses and star formation rates for all the CANDELS fields in Duncan et al. (2019) which are used in this study.

To account for the fact that each stellar particle represents an extended area (rather than treating them as a point source), we model the particles with a smoothing length of a truncated Gaussian emissivity profile equal to the distance to its 64th neighbor particle (Trayford et al., 2017). We then define a grid of wavelengths covering all spectral features we want to probe within the HST filter response functions, similar to the grid used in Trayford et al. (2017). For each wavelength bin of this grid, we launch 10^6 photon packets, assuming isotropic emission until they reach the virtual detector.

This process produces IFU datacubes over the SKIRT wavelength grid which we then reduce to broadband images with the same properties as the CANDELS HST images. SKIRT's reference frame used to generate the datacubes is located at a distance of 10 Mpc (initial redshift z_0) of the sources. We must therefore shift

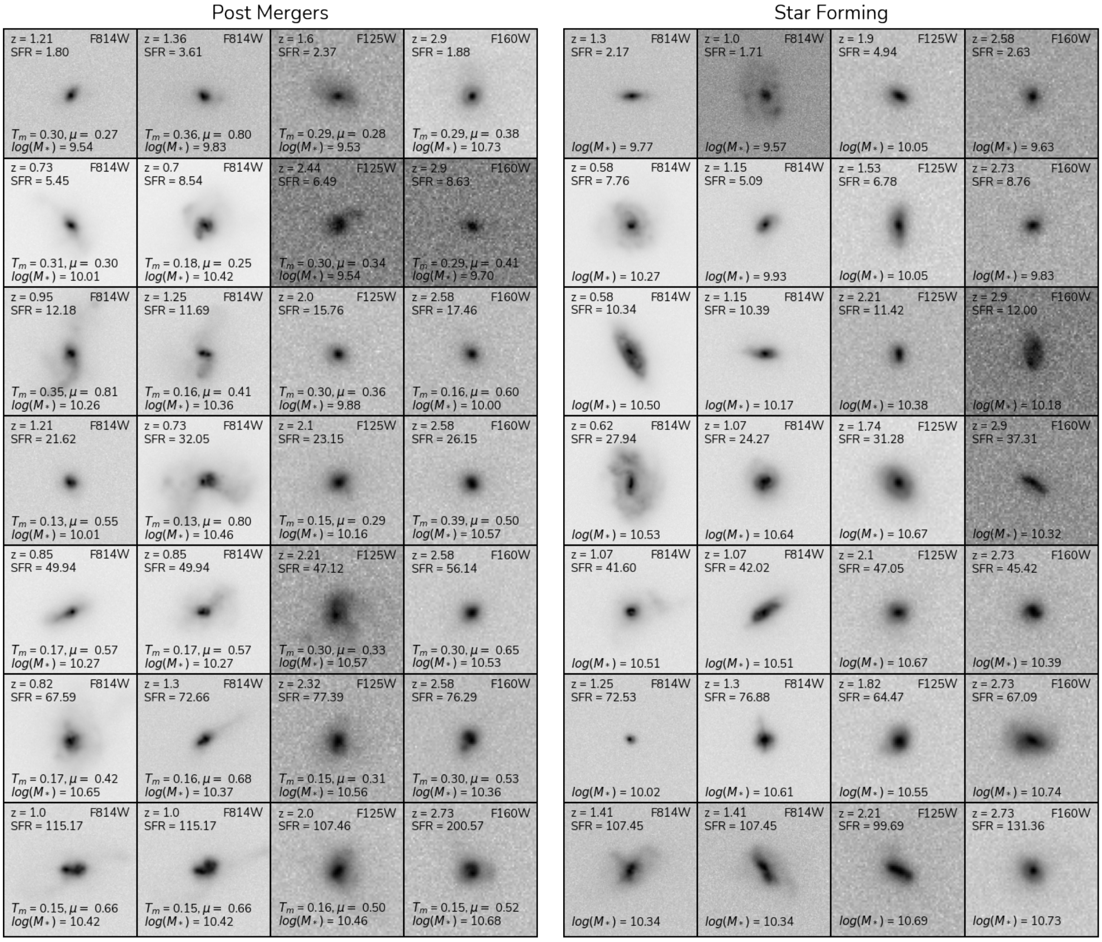


Figure 3.3: A random selection of IllustrisTNG simulated galaxies in our test sample; post-mergers (left) and star forming galaxies (right), with their redshifts, SFRs, and stellar masses printed in each stamp. Images are ordered from left to right in redshift, and top to bottom in SFR. For post-mergers we also display the time since merger, T_m, and the mass ratio μ. All stamps use a square-root normalization.

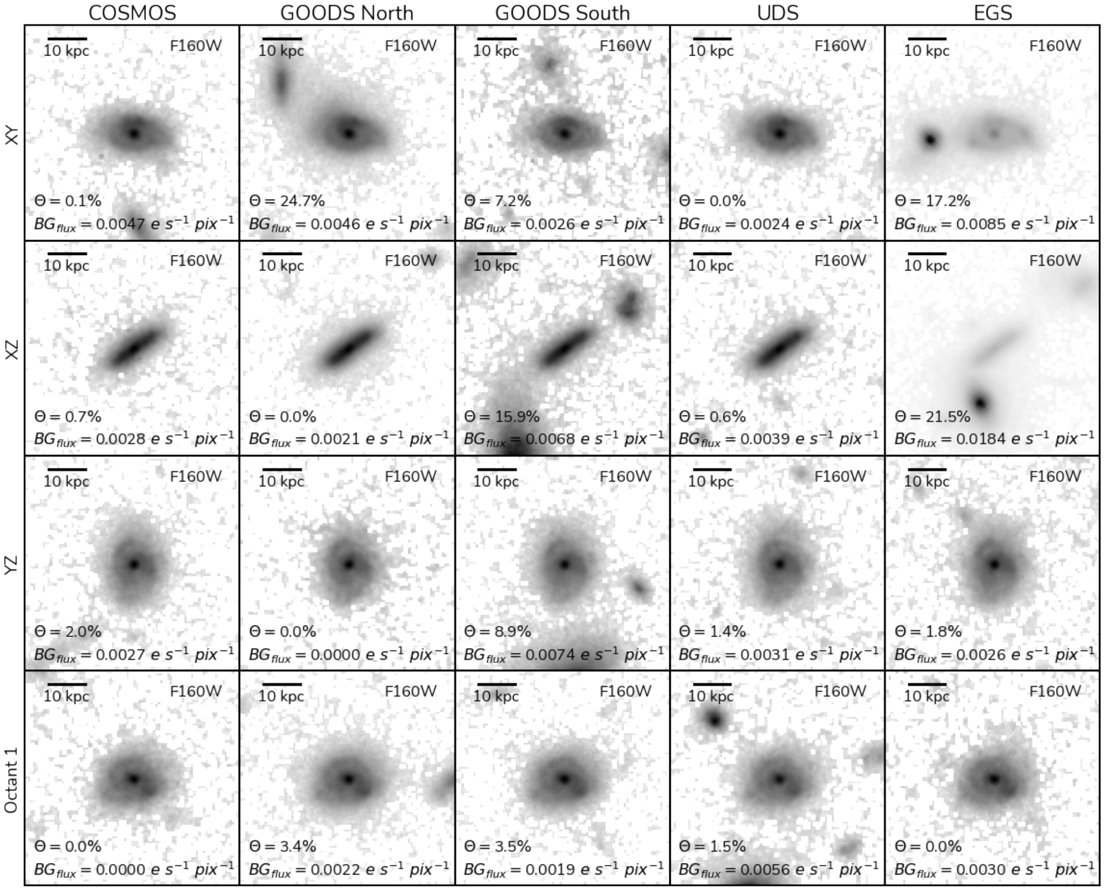


Figure 3.4: Demonstration of the augmentation pipeline for one random galaxy from TNG100-1 (ID=192802, z=0.55, at different orientations). We increase our sample by augmenting the dataset, reproducing it in four orientations (rows) in each of the CANDELS fields (columns). The simulated galaxy is placed in a random patch of the sky in the CANDELS fields and thus can have other sources in the final cutout. The amount of contamination from neighbouring sources varies widely due to the random sampling of the background described in §3.3.4. This contamination is quantified by the overlapping percentage, Θ, and the average flux of the background patch, BG_{flux}.

the IFU data to each target's redshift, z_t, by $(1 + z_t)$ while dimming its flux by

$$\frac{f_t}{f_o} = (1 + z_t)^{-1} \left(\frac{D_{Lo}}{D_{Lt}}\right)^2, \tag{3.5}$$

due to cosmological dimming (Hogg, 1999, eq. 15). Next, we convolve the IFU data with the broadband filters response functions for I_{814}, J_{125} and H_{160W}. The results are clean, noiseless images from the simulation galaxies at 30 mas/pix (matching the ACS pixel scale) before adding any PSF effects (Fig. 3.2, images a) to c). We rebin J and H bands from 30 mas/pix to the WFC3 images pixel scales of 60 mas/pix. Examples for stamps where the background was added can be seen in Fig. 3.2, images e) and f). Figure 3.3 shows randomly selected examples of galaxies in our sample before any contamination from the CANDELS sky is included, separated by their class.

The data-driven paradigm of Deep Learning methods imposes high requirements on the amount of data necessary to train a model that is capable of generalizing the training data well. In practice this means that for the majority of models, a successful approach requires tens, hundreds or even millions of examples. We are far away from these numbers in cosmological simulations. Our initial selection results in a balanced set of ~ 4000 examples of each class (Sec. 3.3.1). Fortunately, in the case of galaxy images there are ways to increase the initial dataset by exploiting aspects of the final image that do not depend directly on the simulated galaxy. In our case, we apply data augmentation to our dataset in three ways outlined below. An example of this approach is shown in Fig. 3.4, following the same galaxy in each possible combination of orientation/field.

First, since IllustrisTNG provides the 3D distribution of all particles associated with a galaxy, we generate each galaxy with different line of sight projections, treating each new representation as a new galaxy. We select four different projections, three aligned with the axis of the simulation, XY, XZ and YZ, respectively, and a fourth line of sight aligned with one octant of the simulation cube.

Secondly, each CANDELS field has unique observational properties (e.g. different noise levels, depth). We exploit this aspect and reproduce each of the different orientations from the previous step on top of a random patch of sky of each CANDELS field, taking care to use appropriate noise levels for the simulated galaxy. To find empty patches of sky, we randomly sample the RA and DEC within each field, and make a large cutout of the area that is 4 times larger than the final size of the cutout. Using positions given in the CANDELS catalogs, we then identify all sources within this cutout and reselect a new RA and DEC within the cutout that does not centrally overlap with another source. We allow some degree of overlapping source, but require a unique central position. We do this interactively until a patch of sky that matches all above criteria is found. This, combined with all the orientations, augments our data set 20 times. In addition, this also helps the network to generalize the impact of contamination from neighbouring sources, as the same galaxy in one field might be isolated in its cutout, but in a denser environment in another.

Finally, we apply random flips, rotations and small zoom-in/zoom-outs around the central source on the fly during training as a regularization technique. This does not increase the overall size of the sample, but at each training epoch the network sees different realizations of the same sample.

Overall, our sample increases from $\sim 8,000$ examples to $\sim 160,000$. However, having multiples of similar galaxies in our dataset can result in overfitting. To reduce this risk, we do not allow different realizations of the same galaxy to fall in both the training sample and the test sample. This ensures that testing and validating are performed on unique datasets.

3.3.5 Contamination Quantification

Providing realistic levels of contamination and the inclusion of neighbouring sources are some of the most important requirements for a good generalization between samples of simulated galaxies and real observations (Bottrell et al., 2019). In an update to what was done in Ferreira et al. (2020), we included realistic contamination in our IllustrisTNG sample, as described in §3.3.4. By comparing clean galaxy realizations to their respective background-added images, we can thus test how our methods behave when faced with a variety of contamination levels, drawing direct conclusions for real world applications. We quantify the degree of contamination in each image using two measurements, which are also listed in Fig. 3.4.

First, we define how much of the background sources is covered by the central source. We call this the overlapping percentage, Θ. For this we measure segmentation maps both for the central source and all background sources of each image stamp. Θ is the percentage of the segmentation map of the central galaxy that is covered by segmentation map(s) of background sources and ranges from 0%, for no overlap, to 100%, where the central galaxy is completely covered by another galaxy in the field.

Second, we estimate the average flux (per pixel) of all background sources, $\mathrm{BG}_{\mathrm{flux}}$, by averaging the flux of the sources within the segmentation map over its area. $\mathrm{BG}_{\mathrm{flux}}$ values are given in units of $e\ s^{-1}\ pix^{-1}$. This ranges from $\mathrm{BG}_{\mathrm{flux}} \sim 0$, where there is no apparent or very faint source in the background, to values that are comparable to or even higher than the flux of the central source. Very high

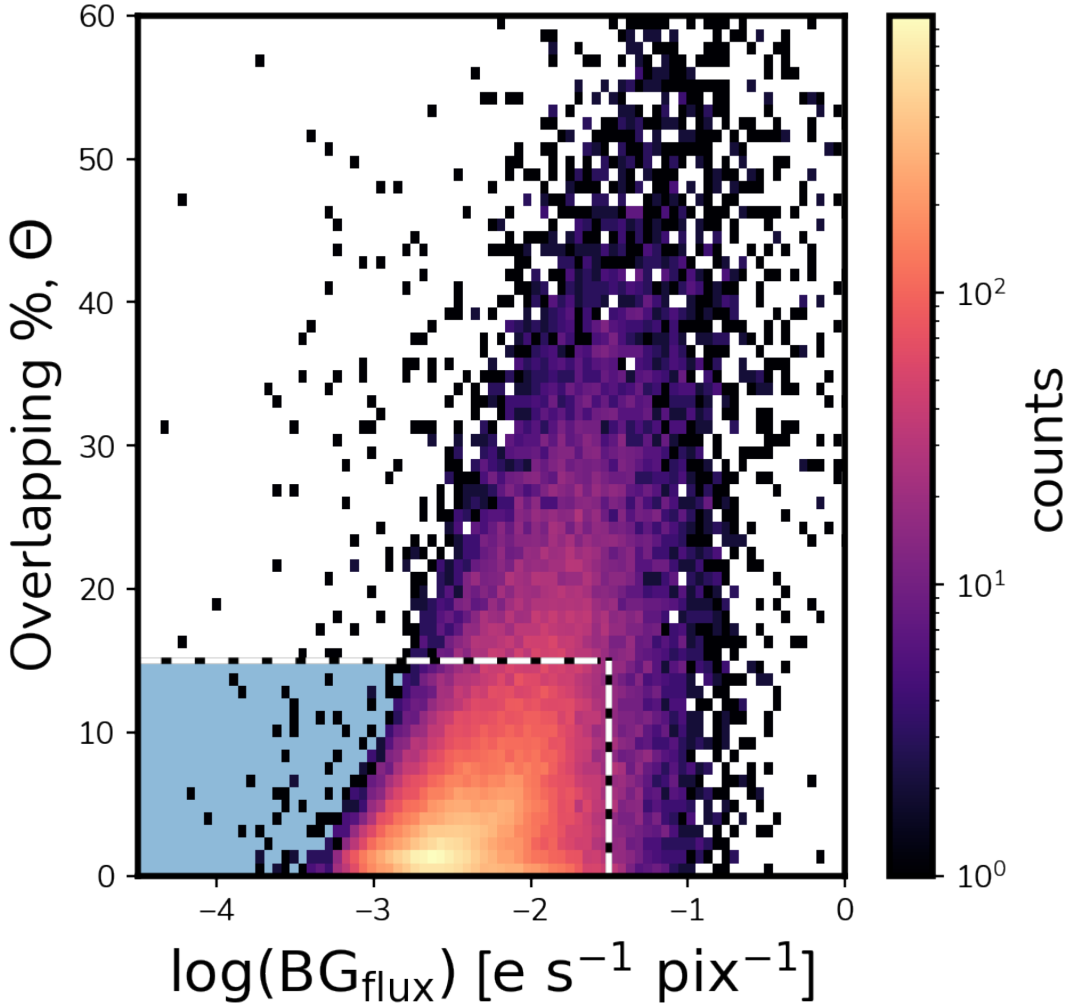


Figure 3.5: Contamination characterization for 162, 000 IllustrisTNG simulated images in our sample. We show the logarithm of the average flux per pixel of the background measured in each cutout, log(BG_{flux}) vs. the overlapping percentage, Θ, which indicates how much the central galaxy segmentation map is covered by the segmentation map of the sources in the background. We define a conservative region of low contamination shown by the dashed line and blue area, which contains 90% of the whole sample. Every point represents at least one image.

values may be due to bright neighbouring sources that outshine the central galaxy. Stars can also be identified by this method.

We use the overlapping percentage, Θ, and the flux of background sources, $\mathrm{BG}_{\mathrm{flux}}$, to define galaxy images with low contamination. Figure 3.5 shows the parameter space formed by these two measurements for the entire sample of $\sim 160,000$ simulated and *candelized* images. The blue box framed by the dashed line defines a region of galaxies with low contamination,

$$\Theta < 15\%, \quad \mathrm{BG}_{\mathrm{flux}} < \times 10^{-1.5}\ \mathrm{e\ s^{-1}\ pix^{-1}}, \tag{3.6}$$

which can be considered as a conservative choice. We find that $\sim 90\%$ of our sample is located in this region. We do not remove the remaining 10% of the galaxies from our sample, because such highly contaminated cases will also be present in observations. We use these contamination estimates to understand how our methods are impacted by it.

These two properties form a simple and powerful way to characterize the contamination of our sample, as they control different contributions to contamination. Because these are challenging to measure directly in real CANDELS observations, we trained a deep learning model to predict the same values in real images. We describe this exercise in § 3.3.5. By inference, any discussion based on contamination measurements in our simulation sample is also valid for the CANDELS observations.

Contamination Network

Based on the contamination measurements described in §3.3.5, we are in a privileged position to provide the community with a model capable of estimating the overlapping percentage (Θ) and the background flux (BG_{flux}) measurements from real observations, as this is also a dataset that can be used for training. In this way, contamination thresholds can be applied to real observational samples in a similar way to what is done in the simulations. However, the sample de-contamination provided by such models are not a requirement for the results discussed in this Chapter, as the imaging mock dataset contains all the contamination expected from real surveys.

The contamination quantification depends on our ability to separate the background patch of the sky from the central source, a feature that is only available when we are post-processing simulated galaxies. In the case of real CANDELS observations, directly measuring these properties is difficult, because it is not straightforward to de-blend background/foreground sources if they are projected on top of one another or are close enough to be a potential interaction.

We use all the contamination information from our data pipeline (§3.3.4) to train a neural network to predict these values from the final image, without separating source and background. We use the same network architecture described in this work, but replacing the final sigmoid layer with a linear activation function, changing the loss function as well. The result is a model that can be directly applied to real observations, where the image is the input and the outputs are values for Θ and BG_{flux}.

Figure 3.6 displays the performance of these predictions based on the original measurements, together with Pearson and Spearman correlation indices. In general, the performance of the model is in good agreement with the original measurements, with root mean square errors in the order of $\sim 10^{-3}$ for BG_{flux} and $\sim 5\%$ for Θ. These limits are well within the region of the parameter space formed by these indices that we defined as a low contamination region. Apart from the small bias making the predictions undervalue the truth values, the performance is good enough to separate high contamination cases from the rest of the sample, which is ultimately our goal. In Figure 3.7 we show examples of different combinations of Θ and BG_{flux}.

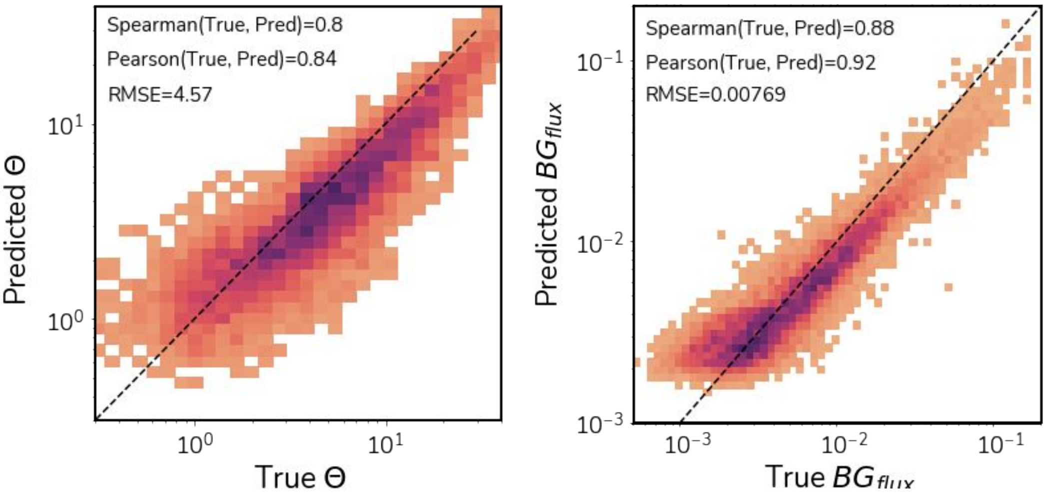


Figure 3.6: Performance of the contamination quantification network. (top) The relationship between true and predicted values for BG_{flux} and (bottom) relationship between truth and predicted values for Θ. Pearson and Spearman correlation indices are displayed for each case, as well as the root mean square error.

Even though this network is designed to be used within the context of this work as a way to reproduce contamination quantification in the same manner as what was done with the simulated images, we recognize that this can be useful for a wider application. For example, this can be used as a fast selection tool that can remove catastrophically bad cases from big samples in just a couple of seconds, thus it can be a powerful tool for quick exploration. In this regard, we release this model independent of the classification models presented in 3.4.1.

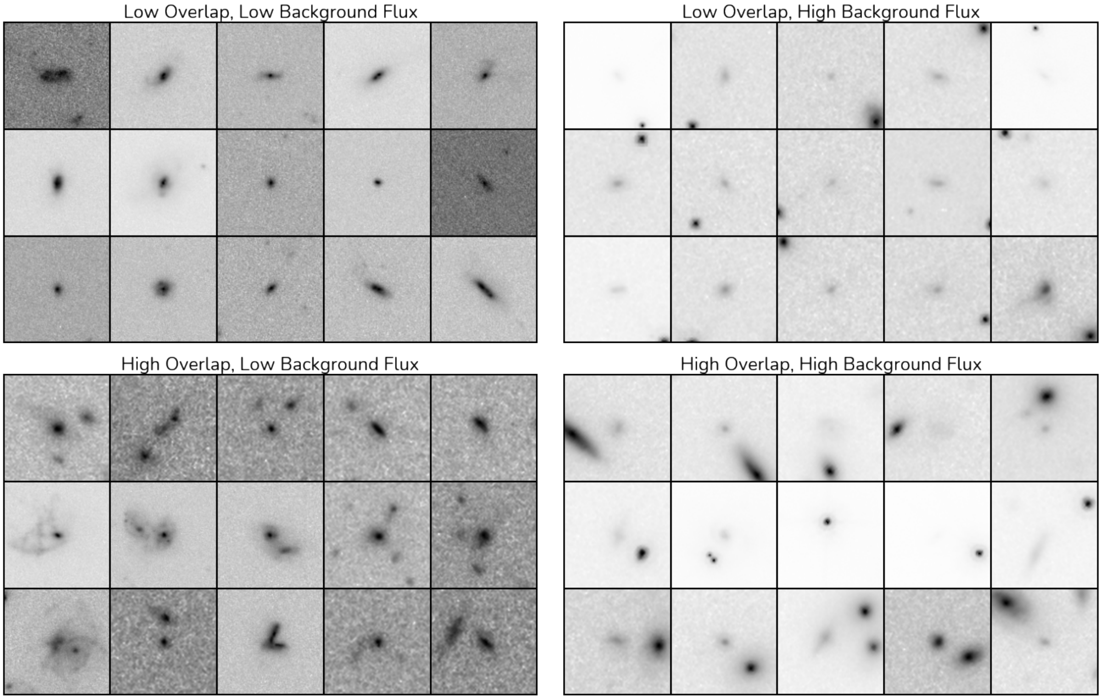


Figure 3.7: Panel with cutouts of IllustrisTNG galaxies demonstrating four different selections on the O_f and $< f_{BG} >$ parameter space. Isolated galaxies with almost no noticeable contamination have low overlap and low background flux (top left). Low overlap and high background flux show cases where the central galaxy is overshadowed by a bright companion, but with no overlapping (top right). High overlap and low background flux show galaxies overlapping with similar brightness, cases where projection effects can be misinterpreted as a major merger (bottom left). High overlap and high background flux show central galaxies with very large and bright companions that extend over its segmentation map (bottom right). This illustrates how useful these two measurements can be for proper selections.

3.4 Methods

We use a deep learning framework with a Convolutional Neural Network (CNN) based on Chapter 2 but with significant updates related to the improved and more robust data pipeline that was discussed in §3.3.4. In this section we describe our Deep Learning analysis (§3.4.1), where we also highlight the improvements to Chapter 2. In Section §3.4.2, we discuss how to avoid overfitting due to the augmentation of the TNG sample, which was part of our sample pipeline. We further wish to compare the resulting classifications to “traditional” classifications. We thus measure non-parametric morphology indices, structural parameters and Sérsic profiles for both the TNG sample and the CANDELS sample with MORFOMETRYKA (Ferrari et al., 2015; Albernaz Ferreira & Ferrari, 2018; Lucatelli & Ferrari, 2019), for which we provide a brief overview in Section §3.4.3.

3.4.1 Deep Learning Classifications

We employ neural networks to forward model the simulations into the observational domain. The neural network takes galaxy images as input and outputs a probability associated with its classification, in this case whether it is a post-merger or a star forming galaxy.

Neural networks are known for being able to approximate complex functions where no analytical approach is feasible, based on the universal approximation theorem (Lu et al., 2017). Deep neural nets combine several layers of nodes (neurons) in a feed forward fashion, mapping inputs to outputs using non-linear activation functions. As a data-driven method, the underlying rules are not explicitly programmed into the network but learned from pattern recognition on the relationship between inputs and outputs of data. These rules are found by minimizing a loss function between the true outputs and the predicted outputs. It is optimized by adjusting the weights and biases of the network so that the loss function reaches a minimum.

A Convolutional Neural Network (CNN) is an end-to-end method, where the most meaningful spatial features are also learned from the data itself through convolution operations. These features are then combined for a classification task, producing the desired outcome based on the input.

In this work, we use an improved version of the CNN architecture described in Ferreira et al. (2020). This consists of a feed-forward network with an input image size of 128x128 pixels, where the number of convolutional blocks, convolutional layers, fully connected layers, number of filters, and kernel sizes, are all defined by the following hyper-parameters:

- `number_conv_blocks` define the number of convolutional blocks, each will probe features of different scales;
- `number_conv_per_block` describe how many convolutions each block will have;
- `initial_number_filters` define the starting number of filters, that are then doubled after each convolutional block;
- `initial_kernel_size` is the initial size of the convolutional kernel, that is then reduced by 2 after each block, down to a minimum of 3;
- `n_fc_layers` and `size_fc_layers` define the number of hidden layers and their respective size, respectively;
- `l2_regularization` and `dropout` are the degrees for each regularization

technique used, respectively. l2 regularization is applied to all convolutional layers, where dropout is applied only after the hidden layers;

The approach of variable depth and width for neural networks is similar to the family of networks described in Tan & Le (2019). However, in our case the networks are smaller due to the smaller image size used.

We modify the methods from Ferreira et al. (2020) to improve generalization of our models. First, instead of using two binary classification networks and combining their predictions to construct a multi-class classification, we now only use one network for the binary classification of post-mergers and star forming galaxies.

Second, we treat the learning rate differently. In Ferreira et al. (2020), we monitored the learning rate decays during training as a hyper-parameter. Here, we use Cosine Annealing, a type of learning rate scheduling (e.g., Loshchilov & Hutter, 2016, for an explanation) combined with a regular Stochastic Gradient Descent (SGD) optimizer (Zhou et al., 2020). This approach probes several different learning rate regimes during training and uses cyclic resets that serve as a way to avoid unstable local minima, improving generalization of the solutions.

All hyper-parameters are determined by a Bayesian Optimization process (The GPyOpt, 2016), and the values for the best model used here are summarized in Table 3.2. These values can be directly used in conjunction with our public KERAS implementation.

3.4.2 Augmentations and Overfitting Avoidance

To avoid overfitting pitfalls from using our CANDELS background augmentation pipeline (§3.3.4), we train a suite of models, one for each CANDELS field. Because we have included areas of all the CANDELS fields as background in our training

Hyperparameter	Best Model
batch_size	128
number_conv_blocks	3
number_conv_per_block	2
initial_number_filters	32
initial_kernel_size	11
number_fc_layers	2
size_fc_layers	128
l2_regularization	0.1
dropout	0.5

Table 3.2: The best hyper-parameters of our architecture found through Bayesian Optimization (The GPyOpt, 2016). These define the depth, width, and number of trainable parameters of our architecture. This process is done using our set-aside validation samples. The same model is used for all the CANDELS datasets.

set, the network could potentially memorize these and use them for predictions, impairing the results. To ensure this is not the case, each CANDELS field has two models – one at low redshift, $0.5 < z < 1.5$, and one at high redshift, $1.5 < z < 3.0$ – trained only with images augmented with regions of the other four fields. All datasets (training, validation and test) are restricted in this way, guaranteeing that any overfitting of the CANDELS background will have no impact in the final application of our models.

An example of this process is outlined in Figure 3.8, for the models that will be used for predictions in the GOODS North (GDN) field. The training set contains galaxies augmented with the COSMOS (COS), GOODS South (GDS), Extended Groth Strip (EGS) and The UltraDeep Survey (UDS) fields while the validation and test sets only contain galaxies from GDN.

This ensures that each model is tailored to one CANDELS field and that no source from that particular field is used during training, i.e., the network never sees any of its data. We further apply a regularization method that makes use of random rotations and image flips on the fly during the training time.

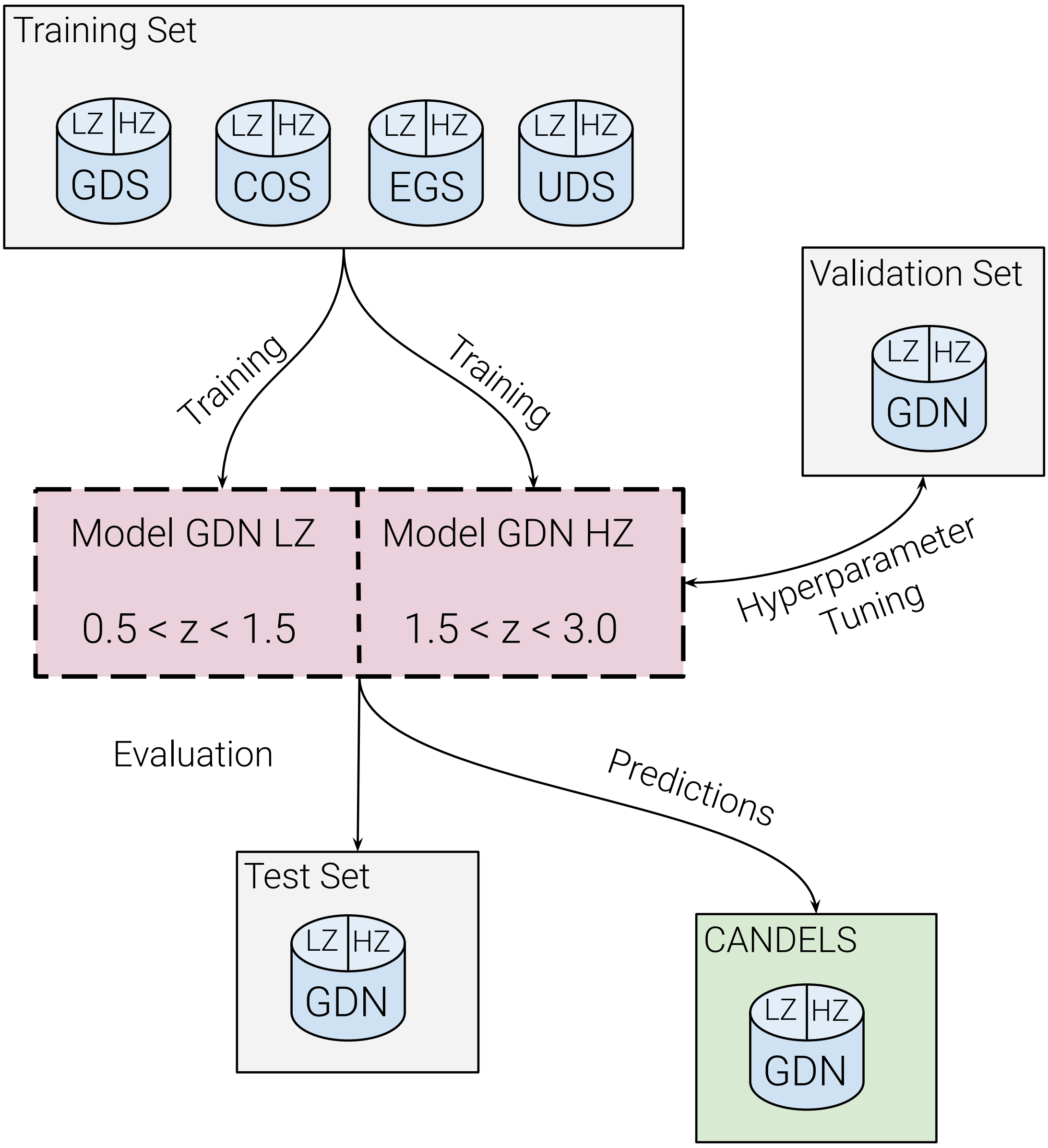


Figure 3.8: Schematics of the training pipeline leveraging multiple fields for augmentation. Each pair of models, at low redshift (LZ), and high redshift (HZ), is trained only with data that is augmented with the CANDELS fields that are not the target for the model. In this example we show a model designed for predictions on GOODS North (GDN), trained on data augmented with characteristics of all the remaining four fields (GDS, COS, EGS, UDS). This model is also tuned and evaluated in validation and test sets that have only of target CANDELS field augmentations, ensuring that no overfitting of neighbouring sources is part of the predictive process.

3.4.3 Galaxy Structure and Morphology

Non-parameteric structure measurements of galaxies are a traditional way to select galaxy mergers (Conselice et al., 2003; Lotz et al., 2004, 2008; Snyder et al., 2017). To measure structures for our sample, we fit Sérsic profiles to all galaxies, using the software Morfometryka (Ferrari et al., 2015; Albernaz Ferreira & Ferrari, 2018; Lucatelli & Ferrari, 2019). Morfometryka measures asymmetry (A) concentrations (C), the Gini coefficient (G), moment of light of the brightest pixels (M_{20}), normalized information entropy (H) and others. It also measures several structural parameters and fits 1D and 2D Sérsic profiles. For our purpose, we are particularly interested in the asymmetry of the galaxies (A), as well as their smoothness (S) since, together, they define a common criterion for finding galaxy mergers:

$$(A > 0.35), \quad (A > S).$$

The asymmetry is defined as the pixelwise normalized difference between the original image and the same image rotated by 180 degrees,

$$A = \frac{\Sigma|I - I_{180}|}{\Sigma|I|} - A_{bg},$$

where I is the image, I_{180} is the rotated image and A_{bg} is an asymmetry term associated with the background (e.g., Conselice, 2014). We measure A_{bg} in each cell of a meshgrid overlayed onto the image, omitting the area occupied by the segmentation map of the central galaxy. We then use the median of these values as A_{bg}. This ensures a robust modeling of the impact of the background in the resulting asymmetry of the image (e.g., Tohill et al., 2021).

Finally, as we are especially interested in investigating the nature of the peculiar/irregular cases, we follow the hybrid method proposed by Bickley et al. (2021). We first filter out regular symmetric galaxies from the sample using the asymmetry (A). Instead of using the widely used cut for selecting mergers ($A > 0.35$), we choose a conservative selection of galaxies with,

$$(A > 0.1).$$

This will remove cases that are irrelevant for our research question. These are galaxies without any disturbances that would classify them as peculiar or irregular.

In Fig. 3.9 we show the distribution of asymmetries A measured with Morfometryka for star forming galaxies (in blue) and post-mergers (in red) for the simulated galaxies. The distributions largely overlap, though asymmetries for post-mergers are generally slightly higher. The difference between both distributions is small enough that using solely the asymmetry ($A > 0.35$) will produce samples with low completeness and purity, and given that the fraction of merging galaxies is lower than regular star forming galaxies, it is likely that this approach produces very contaminated samples.

3.5 Results

Here we discuss what our trained models reveal, first from the test dataset of IllustrisTNG selected galaxies (§3.5.1), and then applied to the CANDELS fields (§3.5.2).

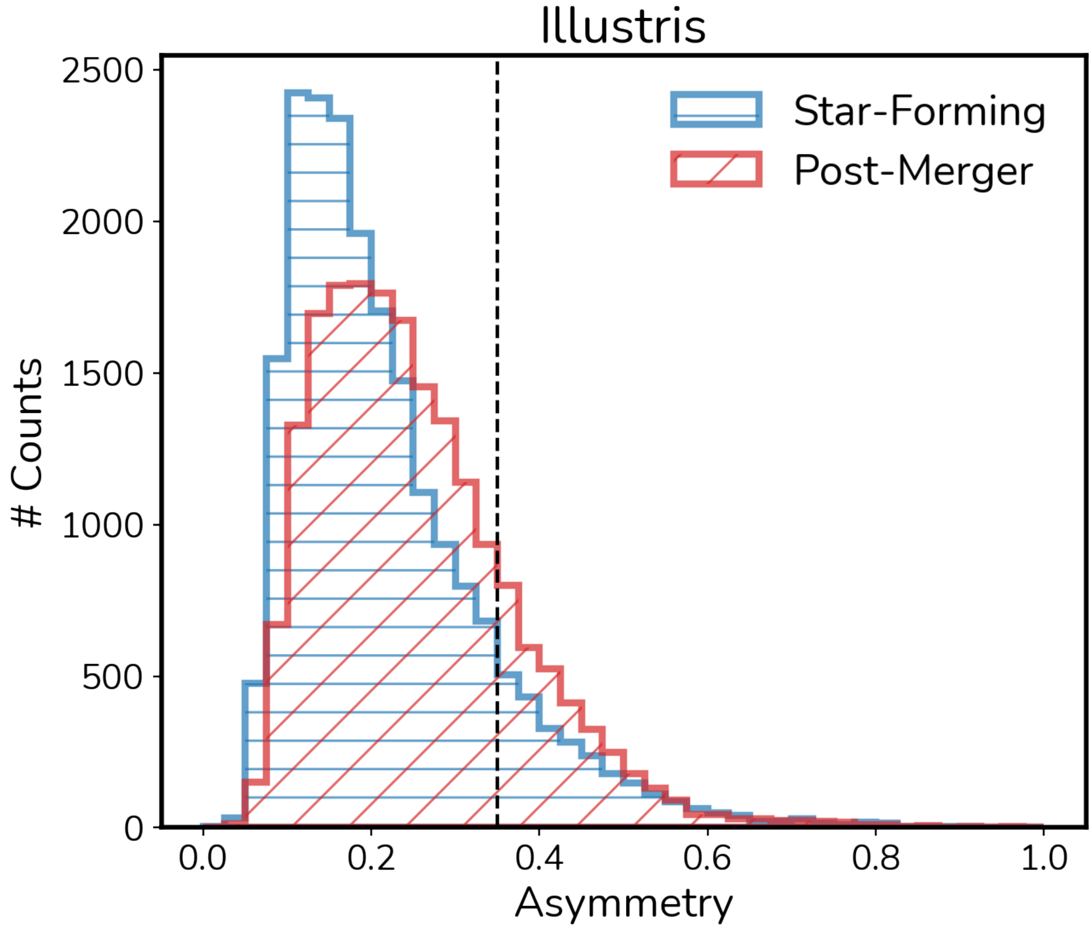


Figure 3.9: Distribution of asymmetries A measured with Morfometryka for our TNG100-1 sample of galaxies. Star forming non-mergers and post-mergers are shown in blue and red, respectively. The dashed vertical line illustrate the typical threshold ($A > 0.35$) used to classify galaxies as mergers.

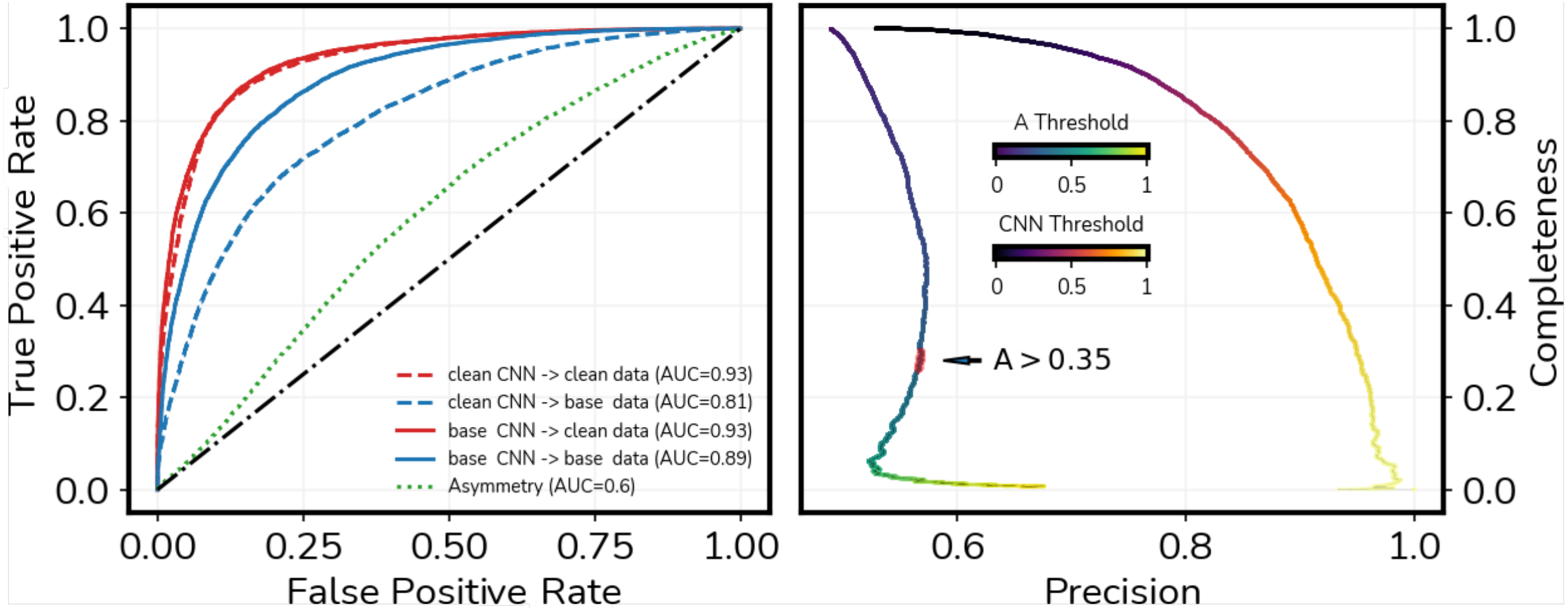


Figure 3.10: Performance metrics for our four trained models and comparison with the classical Asymmetry index A for the simulated images. **Left:** ROC curves for both the network trained with the `pristine mocks` dataset (dashed lines) and with the `realistic mocks` dataset (solid lines) applied to both datasets, color coded in red (`pristine`) and blue (`realistic`). The green dotted line indicates the ROC curve for a classifier using only the asymmetry A. The area under each curve can be read in the label. **Right:** Precision-Completeness diagrams for the baseline network trained with the CANDELS matched mocks with asymmetry A, color-coded by classification threshold levels for CNN (`inferno`) and asymmetry (`viridis`). A small region in red is printed over the asymmetry curve to point out the region where the classification threshold is $A > 0.35$.

3.5.1 Predictions within IllustrisTNG

We measure the performance of our trained models in our prepared test sets. This is done by training the network with two realizations of the test datasets, one with full HST-matched properties including a CANDELS background patch of the sky (§3.3.4, §3.4.2) (which we call `realistic mocks`) and one with clean mocks with no sky noise and contamination included (which we call `pristine mocks`). For simplicity, in cases where we only mention the `realistic mocks` without specifying which CANDELS fields it was augmented with, we consider the average of all 20 models described in §3.4.2.

To compare between models and realizations of these datasets, we use traditional performance metrics common for evaluating Machine Learning model performance. These consist of Receiver Operating Characteristic (ROC) curves and Precision-Completeness diagrams (Powers, 2011), as well as confusion matrices and their individual indices. Here, we are dealing with a single binary classification task, such that the probabilities of both classes respect the condition $\mathrm{P(NMSF)} + \mathrm{P(PM)} = 1$. Fig. 3.10 displays the overall performance for each network.

The left panel shows four different realizations of the network for comparison purposes. The network is trained twice to generate two different types of models: one labeled `base` that consists of a network trained with the `realistic mocks`, and a second labeled `clean`, which is trained with the `pristine mocks`. Then, each model is applied to both datasets. We do this to measure the best case scenario within the simulations, in the absence of any contamination or impact from observational effects. Models trained with the `realistic mocks` dataset are plotted as solid lines, while models trained with the `pristine mocks` dataset are shown by dashed lines. Furthermore, the color conveys the dataset in which the model was applied to, red and blue for `pristine mocks` and `realistic mocks`, respectively. In addition to these, a single parameter classifier based on the asymmetry (A) is also evaluated and displayed as the green dotted line. The area under the curve (AUC) for each case can be found in the legend of the left panel.

The different realizations of our network (`base` and `clean`) cross-correlated with the `realistic mocks` and `pristine mocks` datasets confirms the importance of realistic observational modeling of the mocks (discussed in detail by Bottrell et al. (2019)). This is especially important when crossing domains from cosmological simulations to real observations. Figure 3.10 shows that the `base` network performs just as well as the `clean` network when applied to the `pristine mocks`, resulting in similar performance metrics, as can be seen by the overlapping red curves. However, the `base` network outperforms by $\sim 10\%$ the `clean` network when applied to the `realistic mocks` dataset, as displayed by the difference between the blue curves in Figure 3.10. This demonstrates that correctly modeled observational features increase the generalization capabilities of the resulting models. A network that is only trained on pristine images will perform poorly in the real observations domain.

Importantly, all cases outperform the asymmetry by $20-30\%$. To some extent, this is expected because asymmetries of post-mergers are lower than asymmetries of galaxies that are just in the beginning of their merging event, including cases of closely interacting galaxies. Evidently, the asymmetry function is a much more general morphological descriptor while the network is very specialized for the particular task of dividing post-mergers from star forming galaxies. Here, even the `clean` CNN applied to the `realistic mocks` overperforms the asymmetry by 20%, while in the best case scenario the `base` network applied to the `pristine` dataset overperforms the asymmetry by 30%.

We compare the performance of asymmetry (A) and CNN predictions further and show completeness-purity diagrams in the right panel of Figure 3.10. It displays outcomes for our ensemble of CNN models in `inferno` colormap, and for the classic asymmetry parameter in `viridis` corlormap. The commonly used asymmetry value to classify galaxy mergers is generally higher than ($A > 0.35$), which is shown in the figure by the red patch over the curve. However, here we compare an asymmetry classifier with our neural network to exemplify how one can use the classification threshold of the network as a way to control the trade-off between precision and completeness. This is a useful feature when dealing with unbalanced datasets, like the case for galaxy mergers.

The precision and completeness of the asymmetry behaves in unpredictable ways. First, the precision of the selection increases slowly, then it decreases again around ($A \sim 0.2$), and spikes above 0.6 precision for ($A > 0.8$), but with very low completeness. We do not seek to redefine its use, but merely contrast it with our deep learning approach, and show in broad terms when it might fail when dealing with ambiguous morphologies.

Our network is able to correctly identify post-mergers and star forming galaxies from the IllustrisTNG simulation in $\sim 80\%$ of the cases. Figure 3.11 shows the confusion matrix for the `realistic mocks` dataset identified within each individual CANDELS field, as well as for the `pristine mocks` sample, where accuracy reaches $\sim 90\%$. All classifications are done with the model trained with the `realistic mocks`. We show True Positives (TP) and True Negatives (TN) in blue, and False Positives (FP) and False Negatives in pink. The `CLEAN` case represents the best case scenario, where our current method and dataset achieves an even higher performance of $\sim 91\%$ True Positives. A histogram of the redshift distribution for each cell helps to visualize any possible biases in redshift for the miss-classification cases. This demonstrates that the models are more likely to correctly classify low redshift galaxies, as they represent the majority of the samples.

Impact of Redshift

With the goal to apply our models to a wide range of redshifts, we explore how our performance metrics are impacted by increasing redshifts. Following the angular size – distance-relation, galaxies at increasingly larger distances from low to intermediate redshifts will be greatly impacted by decreasing resolution, which means that morphological features are less well sampled. The right panel of Figure 3.12 shows this effect on the performance of our models, where the scores of the metrics gradually decrease with increasing redshift, going from 85% accuracy at $z = 0.5$ to around 80% at $z = 2$. The errorbars – sampled from boostraping our testing samples – follow accordingly.

Contamination impact on classification

We use the contamination estimates measured in §3.3.5 to find the contamination failure threshold of our classifier, comparing performance metrics for subsets of the test set selected in bins of both the overlapping percentage, Θ, and the average background flux per pixel, BG_{flux}, as shown in Fig.3.12. The horizontal black dashed line at 0.8 shows the accuracy of the model when evaluated in the complete test set (80%). The metrics outperform this baseline in sub-samples of images with low contamination, decreasing as we increase each of the contamination factors.

As described in Section §3.3.5, we select the point where the average mean values for each metric falls below the dashed line, which is our contamination

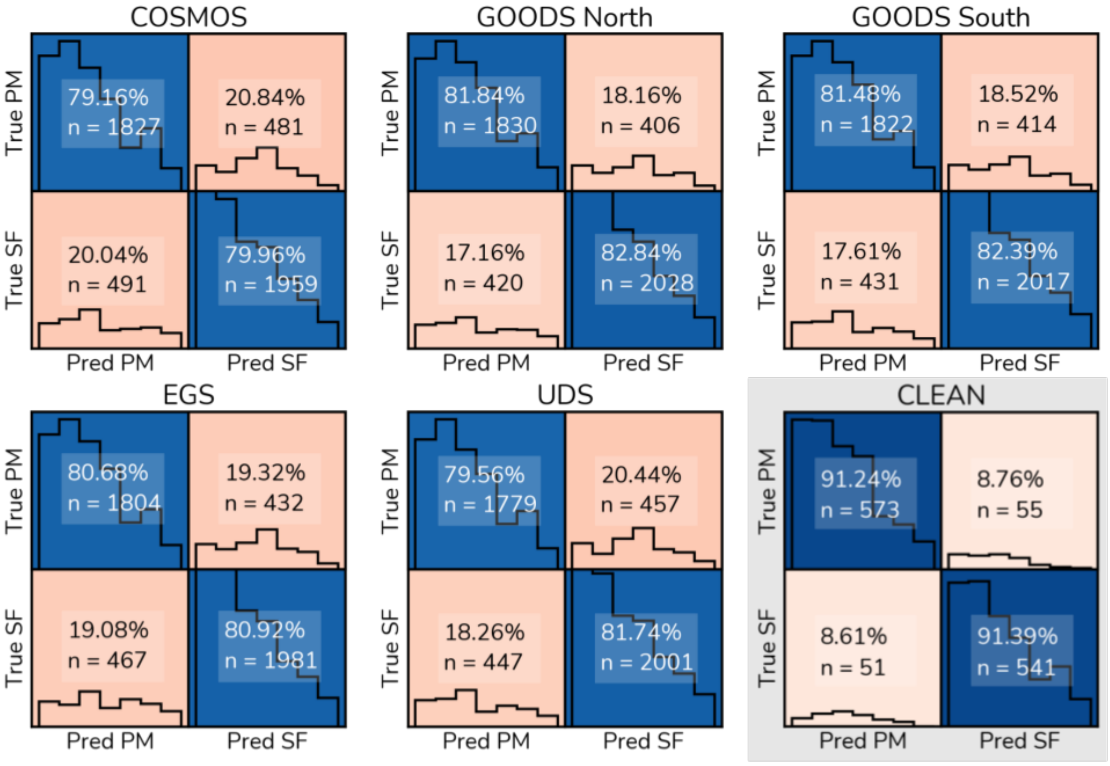


Figure 3.11: Confusion matrix for all the samples matched to CANDELS fields as well as the the `pristine` sample (highlighted by gray shading in the bottom right). These confusion matrices were evaluated with the ensemble of models trained with the CANDELS matched mocks. We show True Negative (TN) and True Positives (TP) highlighted in blue while the False Negatives (FN) and False Positives (FP) are shown in pink. The colors are based on the rate percentage, which is also printed in each cell. All the CANDELS fields have TP and FN rates of around $\sim 80\%$. For the `pristine` case performance can reach as high as $\sim 90\%$, marking the intrinsic limit of our method based on the data available. The histograms show the redshift distribution for the galaxies in each category, which demonstrate that it is easier to recover correct classifications at lower redshifts.

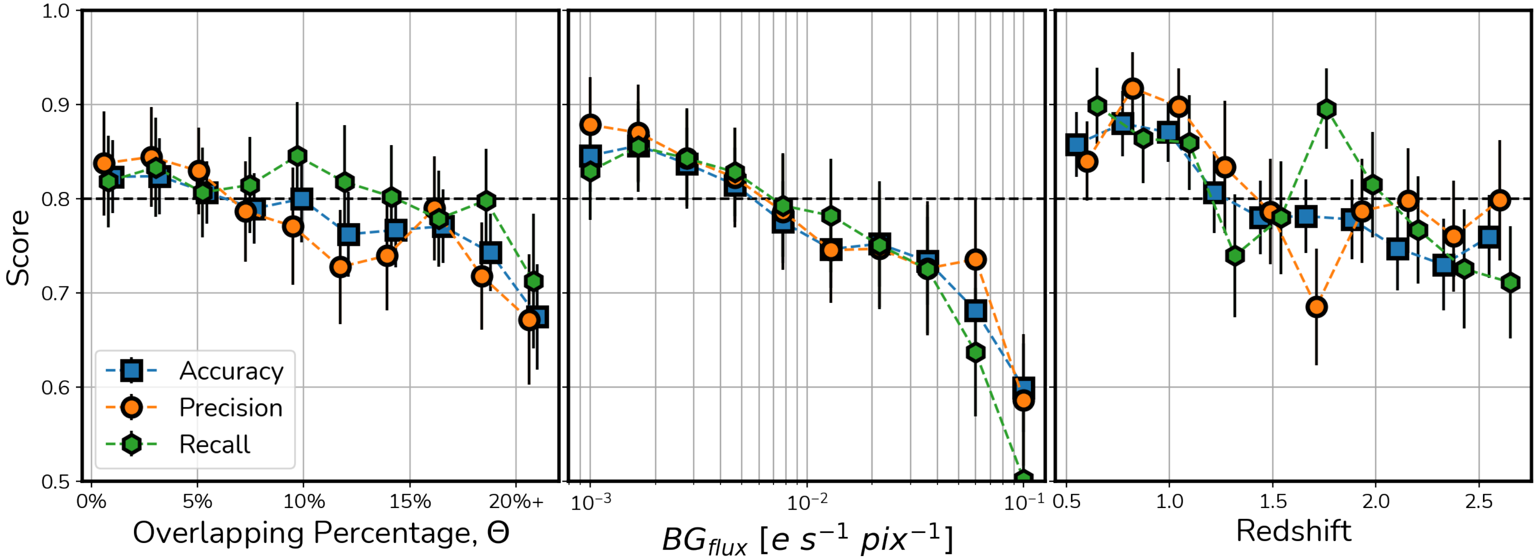


Figure 3.12: Impact of contamination and redshifts on the performance of our models. The left and central panels show how accuracy, precision and recall (blue squares, orange circles and green hexagons, respectively) behave for increasing percentages of overlap (Θ) and for increased background flux (BG_{flux}). In the right panel we show how the accuracy, precision and recall of our methods change in bins of $\Delta z = 0.25$ redshift. Errorbars are sampled from bootstraping the test sample. The performance gradually decreases with z, decreasing below 80% beyond $z = 2$. There is a slight uptick at $z = 2.5$, but with large errorbars. The cutoff at $z > 2.5$ is the result of a combination of small sample size and redshift effects. The black dashed line at $score = 0.8$ indicates the overall accuracy of the model in the complete test set.

cutoff, i.e.,

$$\Theta \sim 15\%, \quad BG_{flux} \sim 10^{-1.5}\ e\ s^{-1}\ \text{pix}^{-1}.$$

Since it is not possible to directly measure the contamination parameters in the real observations, we refer the reader to our deep learning model trained to measure the contamination in Appendix §3.3.5.

3.5.2 Classifications on CANDELS

We use our network to carry out predictions in all real CANDELS galaxies at $0.5 < z < 3$, $M_* \geq 10^{9.5} M_\odot$, $S/N > 50$ and $H_{\text{MAG}} < 24.5$. We filter out regular galaxies using a conservative asymmetry cut of $A > 0.1$ as we are interested only in asymmetric, irregular/peculiar systems. This selection results in a sample of $23,494$ galaxies, for which $14,410$ have visual classifications from Kartaltepe et al. (2015). Based on the classifications from our networks, we separate these galaxies in post-mergers and non-interacting star forming galaxies using a threshold probability of 60%. Galaxies with probabilities $50\% < P(PM) \wedge P(SF) < 60\%$ are not considered in any class. These represent 2125 galaxies ($\approx 15\%$) of the sample with visual classifications. Figure 3.13 showcases some examples of galaxies in the CANDELS fields separated by the classification of our models. Post-mergers in the left panel and star forming galaxies in the right panel.

To investigate how the relative number of post-mergers and star forming galaxies changes over cosmic time, we divide the CANDELS sample in bins of $\Delta z = 0.25$. Figure 3.14 shows the change of class fractions change with redshift. We do this analysis in two mass regimes: low mass galaxies with $9.5 < \log(M_*/M_\odot) < 10.0$ (left panel) and high mass systems with $\log(M_*/M_\odot) > 10.0$ (right panel). Non-interacting star forming galaxies are shown in blue circles while post-mergers are

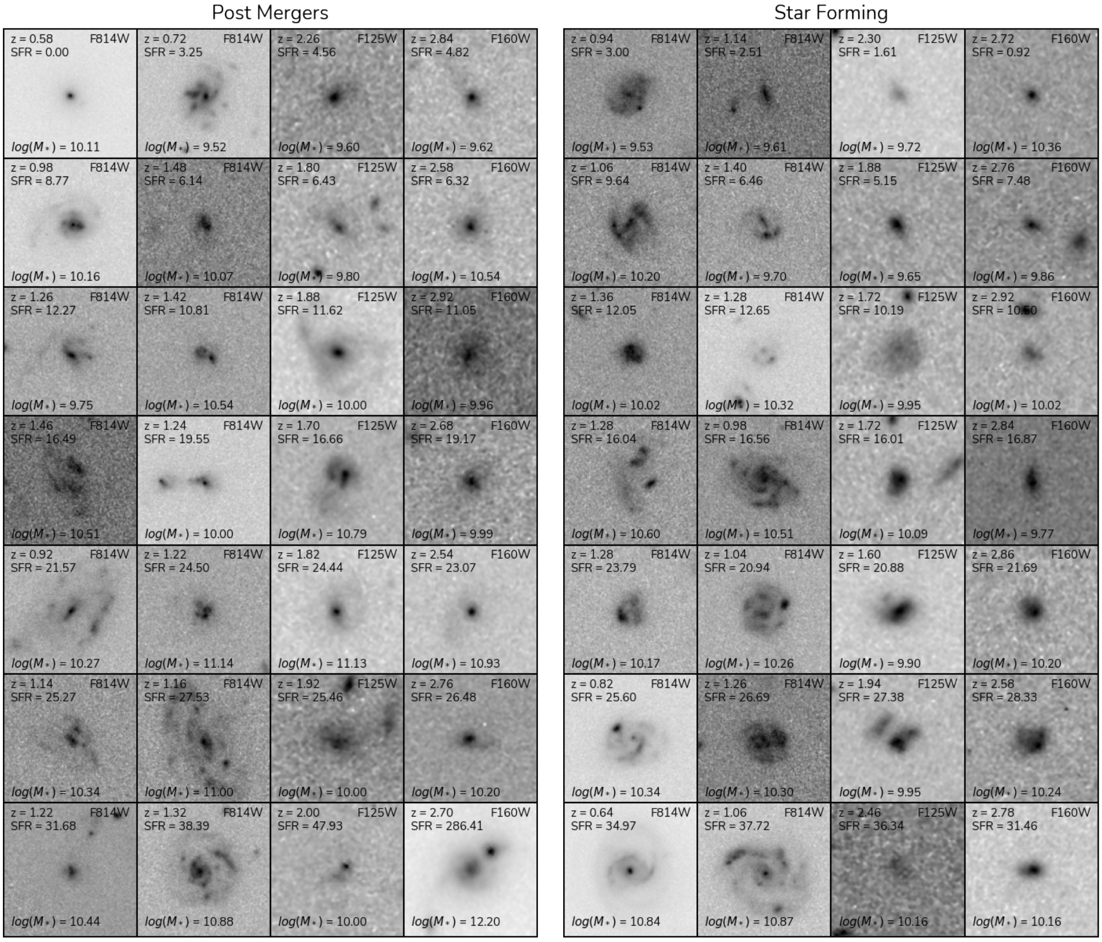


Figure 3.13: Examples of CANDELS galaxies with $A > 0.1$ classified by our models into Post-mergers (left) and star forming galaxies (right), with their redshifts, SFRs, and stellar masses. Images are ranked from left to right with increasing redshift and top to bottom with increasing SFR. All stamps use a square-root normalization.

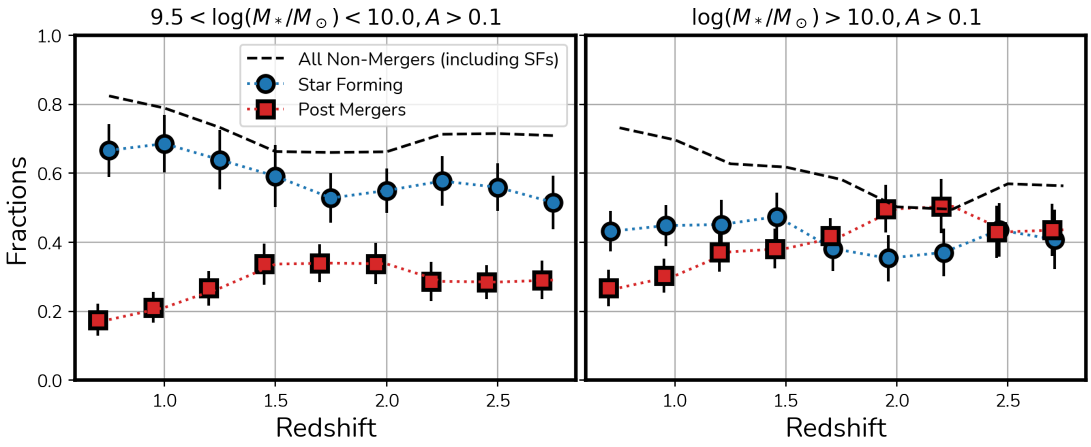


Figure 3.14: Relative class fractions for post-mergers and star forming galaxies vs. redshift for real galaxies in the CANDELS fields. The fraction of post-mergers increases from 30% at $z \sim 0.75$ to 50% by $z \sim 2$. Errorbars are drawn from bootstrapping the samples and applying the underlying uncertainty associated with the performance of our models, which decreases with redshift.

shown in red squares. The upper dashed line displays the fraction of galaxies that are not mergers, including the star forming galaxies and other low probability cases not included in any class.

There is no direct comparison to the fraction of post-mergers reported in Chapter 2. We are showing in Figure 3.14 only the cases with high asymmetries, and as a consequence our fractions here are intrinsically higher. We produce a smaller sample by removing all the regular galaxies with $A < 0.1$ – the majority of the sample in Chapter 2 – and thus end up with fractions that are 5 to 10 times higher. Additionally, our timescales here are also longer, which increase the fractions as well.

For the lower mass post-mergers we see an upward trend from $\sim$ 15% at $z = 0.5$ to $\sim$ 35% at $z = 2$, then a slight decrease beyond $z = 2$. This is still consistent with a $\sim$ 35% fraction within the error-bars. The star forming galaxies behave in the opposite way, decreasing from $\sim$ 70% at $z = 0.5$ to around $\sim$ 55% at

$z = 2$. This suggests that among asymmetric galaxies of this mass range, there is an exchange between the classes as we go to higher redshifts up to $z = 2$. This once again emphasizes that classifications of local galaxies that are purely based on the asymmetry (A) are highly contaminated with non-interacting star forming galaxies. However, this is mitigated at higher redshifts where we find more post-mergers. Nevertheless, samples selected based on A are still dominated by star forming galaxies, albeit to a lesser extent.

Trends for higher mass galaxies are substantially different (right panel of Figure 3.14). While the post-mergers exhibit a similar but more steep upward trend from $\sim 20\%$ at $z = 0.5$ fraction to $\sim 50\%$ at $z = 2$, the relative fraction of star forming galaxies show a constant value of $\sim 50\%$ at $0.5 < z < 3$, while the fraction of the rest of the sample (dashed line) goes from $\sim 75\%$ to $\sim 50\%$.

At the highest redshift bins, the error bars are large, indication for our networks to perform less accurately above $z = 2$. The fraction of post-mergers changes from 30% to around 50% at $z = 2$. We therefore attribute the downtrend in post-mergers beyond $z = 2$ to the poor performance of our models at high redshifts and do not take this to imply a real evolutionary effect.

We know that mergers are more common in the past (e.g, Mundy et al., 2017; Duncan et al., 2019; Ferreira et al., 2020; Whitney et al., 2021), and here we find further evidence that this is also the case for peculiar galaxies, indicating that the nature behind these disturbed morphologies at earlier times can be attributed to merging. To further investigate this, we select all galaxies from Kartaltepe et al. (2015) that are classified as an irregular / peculiar with `f_Irr` > 0.75, i.e., cases where more than 50% of the visual classifiers agree on the classification, and check how our networks perform on this subset. We observe similar trends with redshift, with the fraction of post-mergers increasing by $\sim 20\%$ from $z = 0.5$ ($\sim 30\%$) to

$z = 2$ ($\sim 50\%$), which agree with the results for the complete sample. Furthermore, our methods classify $\sim 50\%$ of the galaxies visually classified as potential mergers (`f_merger` > 0.75) in Kartaltepe et al. (2015) as post-mergers. This is higher than random, but does show the difficulty of obtaining exact matches between mergers determined visually compared with a quantitative process.

Visual representation of the classification

As a way to visualize how our networks organize the features extracted from the images to produce the final classification, we generate a 2D representation of the final dense layer of the network corresponding to 128 neurons (128 dimensions) using an UMAP (Fig. 3.15). The color code of the points expresses their respective labels, red for post-mergers and blue for star forming galaxies. Then, we overplot the positions assigned by the network for unlabeled CANDELS galaxies. We also include some examples of images of CANDELS galaxies close to their original position in this manifold as a way to visualize how the morphologies change with its position. Each region in the parameter space of this diagram is directly related to a probability. The maximum probability is found in the extreme regions further away from the center, which represents how different these objects are for the network. Images of galaxies the network struggles to identify are mixed in the bottom middle, representing the region where both probabilities are similar $P(PM) \sim P(SF)$.

3.6 Implications

Making use of the classifications from our deep learning models, we first explore the impact of major-mergers on classifications above the Star Forming Main Sequence

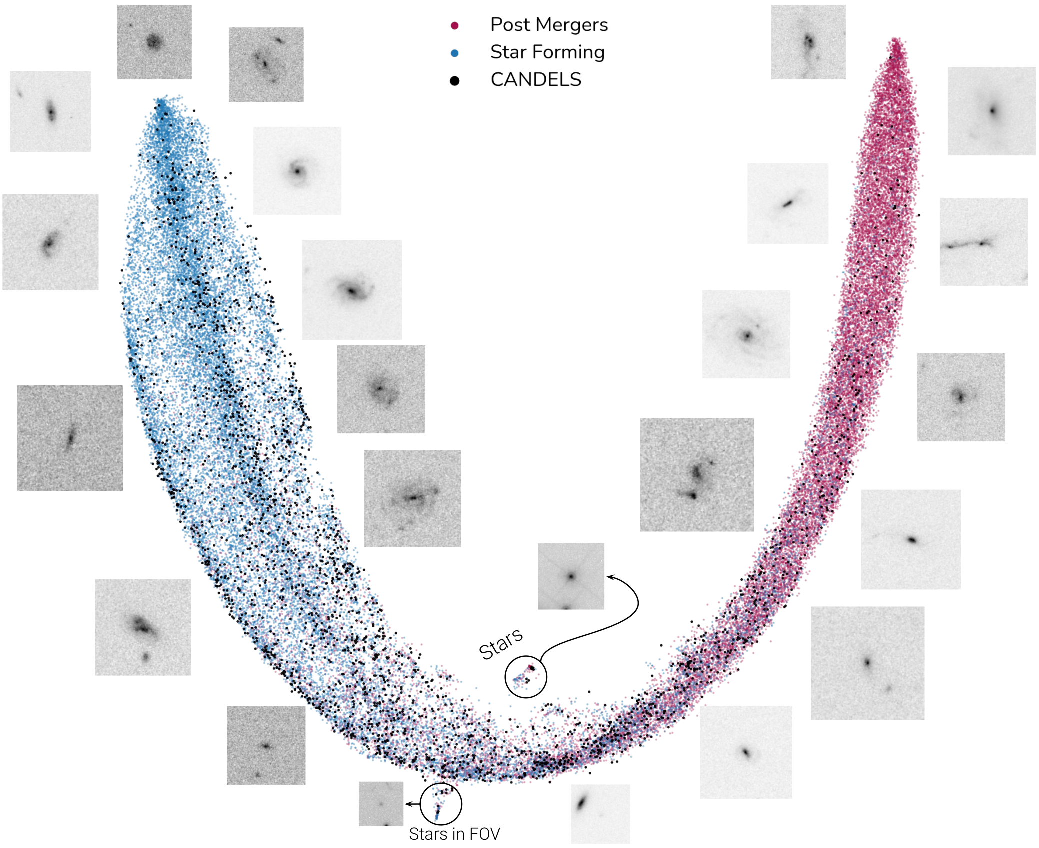


Figure 3.15: UMAP representation of the output from the last dense layer of the network. This representation shows the parameter space used for the network to generate the final probability. Probabilities are highest in the extremes at the top, and uncertainty increases due to increased contamination as we go along this structure towards the middle. Same random examples of CANDELS galaxies are placed close to their points in this manifold. Small regions identified by circles show the clustering of non-galactic detections in this parameter space, located close to the region of uncertain classifications at the bottom of the UMAP. "Stars" are stars in the center and "Stars in FOV" correspond to stars at the edge of the stamps.

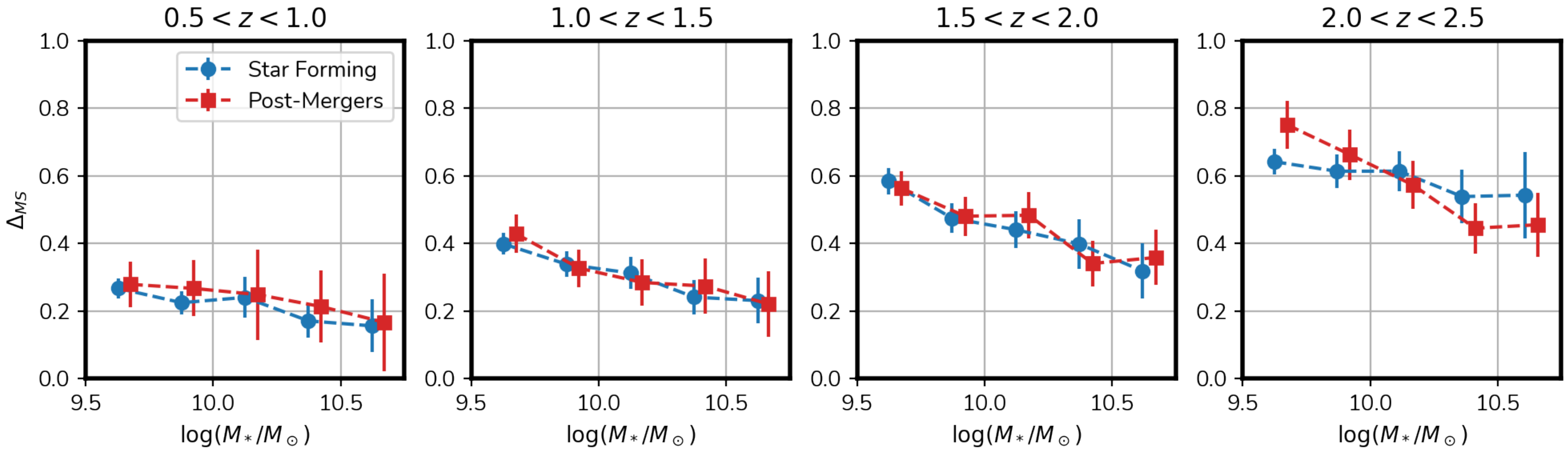


Figure 3.16: Mean distance to the star formation main sequence (Δ_{MS}) vs. log stellar mass in bins of redshift for CANDELS galaxies above the star formation main sequence. Post-mergers are plotted as red squares, star forming galaxies as blue circles. Errorbars are estimated using bootstrapping and show $\pm 1\ \sigma$. The classes are indistinguishable between $0.5 < z < 2$, both increase similarly in Δ_{MS} as we increase redshift. This represents the increase in scatter above the main sequence. All galaxies included in this diagram lie above the SFMS, as we are only interested in exploring the scattering above the SFMS. For the last redshift bin ($2 < z < 2.5$), there is a significant difference between the two classes both at the low mass end and at the high mass end. At low masses ($\log(M_*/M_\odot) < 10.0$), post-mergers scatter higher than star forming galaxies with a difference of $\Delta_{MS} \sim 0.1$ dex. At high masses ($\log(M_*/M_\odot) > 10.0$) the trend reverses and star forming galaxies scatter higher with a Δ_{MS} difference of ~ 0.1 dex. However, care needs to be taken in the interpretation of this trend as it could be spurious or insignificant given the errorbars and the performance metrics of our models at high redshift (Fig. 3.12).

as parametrized by Schreiber et al. (2015) (§3.6.1). We then discuss the structure of the two galaxy classes using Sérsic profile measurements (§3.6.2). In §3.6.3, we update classifications from Chapter 2 with our new specialized model, thus increasing certainty for previously undefined classifications. We then add to the discussion proposed by Bickley et al. (2021) regarding the Bayesian limitations of classifying post-mergers by considering an evolving merger fraction. We finish with §3.6.5, in which we compare extracted features from real CANDELS galaxies to features extracted from IllustrisTNG galaxies, as a way to address the challenges of transferring the model from simulations to real observations.

3.6.1 Classifications above the star forming main sequence

The influence of merging on the structure of peculiar / irregular galaxies at intermediate redshifts ($0.5 < z < 3.0$) is directly related to the question of whether merging galaxies can induce more starbursting episodes than galaxies evolving secularly. Enhanced star formation can then lead to more clumpy and asymmetric structures, and thus can impact the morphological appearance of galaxies greatly. By examining the star formation main sequence of galaxies, one can investigate the nature of galaxies with unusually high SFRs and the formation path that resulted in this physical effect.

In order to investigate this, we select only galaxies in our CANDELS fields sample that lie above the star forming main sequence as parametrized by Schreiber et al. (2015). We separate these sources by stellar masses, redshifts, and their post-merger/star forming classification, measuring the mean distance to the star formation main sequence (Δ_{MS}), as:

$$\Delta_{MS} = \log(\mathrm{SFR}) - \log(\mathrm{SFR_{MS}}), \tag{3.7}$$

where log(SFR) is the log star formation rate of a particular galaxy and log($\mathrm{SFR_{MS}}$) is the parametrization from Schreiber et al. (2015). The SFRs and stellar masses used here for CANDELS galaxies were compiled by Duncan et al. (2019) through SED fitting. We refer the reader to this publication for further details. In Figure 3.16 we show the mean value of each stellar mass bin, for four redshift ranges (one in each panel), separated into star forming and post-merger galaxies by our classifications. For the $0.5 < z < 2.5$ redshift range (panels A, B and C), we do not find any impactful difference between the classes and Δ_{MS}, with all offsets well within the errorbars. However, for redshifts $2.5 < z < 3.0$ post-mergers with $\log(M_*/M_\odot) < 10.0$ are on average ~ 0.1 dex higher than star forming galaxies of the same mass. The opposite is found for $\log(M_*/M_\odot) > 10.0$, however uncertainty is higher here. Additionally, Δ_{MS} increases with redshift in all cases, which describes a larger scatter above the star forming main sequence. However, given the performance metrics of our models at high redshift (Figure 3.12), we cannot claim that this is a real effect. We stress that in Figure 3.16 we only select galaxies above the SFMS, which is why the distance is always positive.

In summary, locations of post-mergers and non-interacting galaxies in the star forming main sequence diagram are comparable, with the possible exception at the highest redshifts. This suggests one of the following: within our sample of CANDELS galaxies, major-merging is not playing a major role in enhancing star-bursting episodes; or the timescale probed by our method is too large and the SFR enhancement from the captured post-mergers is short lived.

A relevant result was discussed in Hani et al. (2020), who investigated TNG300-1 post-mergers at $0.0 < z < 1.0$. They showed that post-mergers have enhanced specific star formation rates by a factor of ~ 2, but that this effect decays in timescales of ~ 0.5 Gyr, which can be driven in part by minor-mergers. Although we do not find evidence for an enhancement in starbursts due to major mergers, we do not rule out the importance of minor mergers to this effect. We trained our models without the presence of minor mergers, but we can not be sure that the star forming galaxies classified by our models are not in some cases triggered by minor mergers.

3.6.2 Structure and light profiles

Our deep learning classifications relate to two different formation pathways. These formation scenarios could result in structures that differ for post-mergers and star forming galaxies. To verify if in fact their structures are diverse from one another, we investigate light profile fitting by using Sérsic profiles measured by MORFOMETRYKA.

Figure 3.17 shows the distribution of Sérsic indices for post-mergers in red, and star forming galaxies in blue. In general, each class presents very distinct distributions: the post-mergers have a mean Sérsic index $n \sim 1.8^{+0.7}_{-0.6}$ roughly representative of a transition from disks to spheroids; star forming galaxies have systematically lower Sérsic indexes with $n \sim 1.1^{+0.5}_{-0.5}$, which is more consistent with disk dominated galaxies. This offset of ∼ 1 dex increases for classification thresholds at higher values. The average Sérsic profile (n) of post-mergers increases while the distribution for star forming galaxies continues with a similar shape. This is quantitative evidence that 1) post-mergers with higher light concentrations are more easily separable from non-interacting star forming galaxies, and 2) these types of galaxies are intrinsically different from each other.

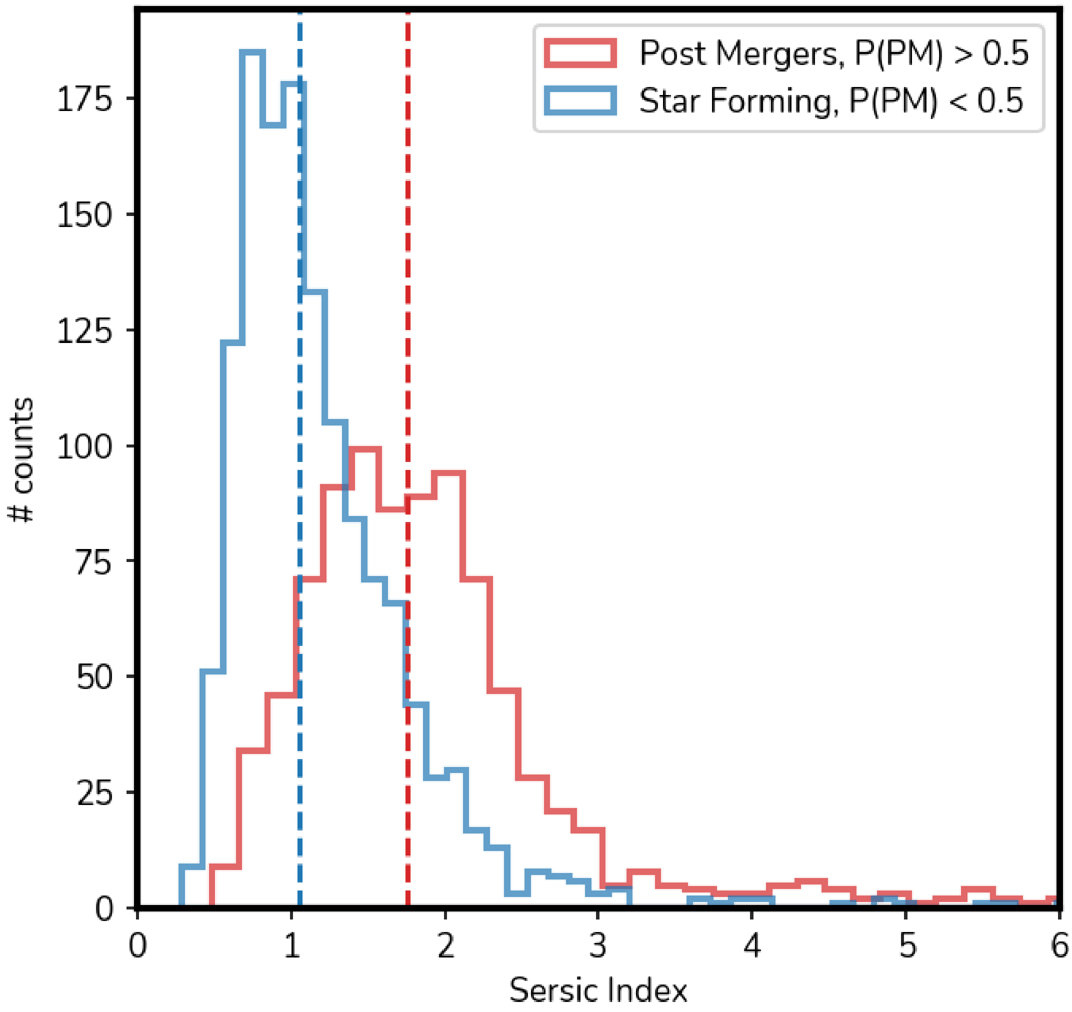


Figure 3.17: Sersic index distribution for post-mergers and star forming galaxies, in red and blue respectively. Post-mergers display more concentrated light distributions with $n \sim 1.8^{+0.7}_{-0.6}$ while the star forming galaxies have $n \sim 1.1^{+0.5}_{-0.5}$ consistent with disk dominated galaxies.

3.6.3 Merger Fractions and Rates

By using the new classifications from this work we can update classifications from Ferreira et al. (2020) for cases where the previous method had ambiguous probabilities for some major-mergers and non-mergers.

Our new dataset accounts for the effects of dust, it is not limited by orientation and probes the rest-frame optical. Thus we can check if any major merger classifications in the previous work can be attributed to non-interacting star forming galaxies or if any non-mergers can be re-classified as post-mergers. This is done by comparing the probabilities for major-mergers and non-mergers, $P(MM)$ and $P(NM)$, respectively, from Chapter 2 to the new probabilities $P(PM)$ and $P(SF)$. We update a non-merger classification to post-merger if

$$P(PM) > P(NM),$$

and update the major-merger classifications to non-merger if

$$P(SF) > P(MM).$$

In other words, we reclassify galaxies from the previous sample where our new method is more certain about its classification than the previous one. This leads to $\sim 5\%$ of major-mergers reclassified as star forming non mergers, which lowers the overall merger fractions at lower redshifts and keeps it similar at higher redshifts. In Figure 3.18 we compare the new merger fraction measurements, in green, to the results from Chapter 2, in gray.

The updated fit of the cosmic evolution of the merger fraction, $f_m(z)$

$$f_m(z) = 0.011 \pm 0.002 \times (1+z)^{2.71 \pm 0.31}, \tag{3.8}$$

with errors estimated with boostraping, agrees with the previous measurement in Ferreira et al. (2020) within errors. To measure the galaxy major-merger rate ($\mathcal{R}$), we combine the timescale ($\tau_m = 0.5$ Gyr) used in our selection (§3.3.1) with this merger fraction through

$$\mathcal{R} = \frac{f_m}{\tau_m}. \tag{3.9}$$

The updated galaxy major merger rate is

$$\mathcal{R} = 0.022 \pm 0.006 \times (1+z)^{2.71 \pm 0.31}. \tag{3.10}$$

We emphasize that this correction is a minor adjustment to the galaxy major-merger rates presented in Chapter 2, which remain broadly consistent with each other.

3.6.4 Bayesian Analysis of Mergers

We now investigate the possible contamination in merger samples that are selected through our method. This approach is fairly direct and based on Bayesian statistics, and relies on some understanding of the true intrinsic merger fraction and how it evolves with time. It also requires that we have a good understanding of the fraction of contamination in merger samples (Bickley et al., 2021). The basic Bayesian formula to understand this is given by the following.

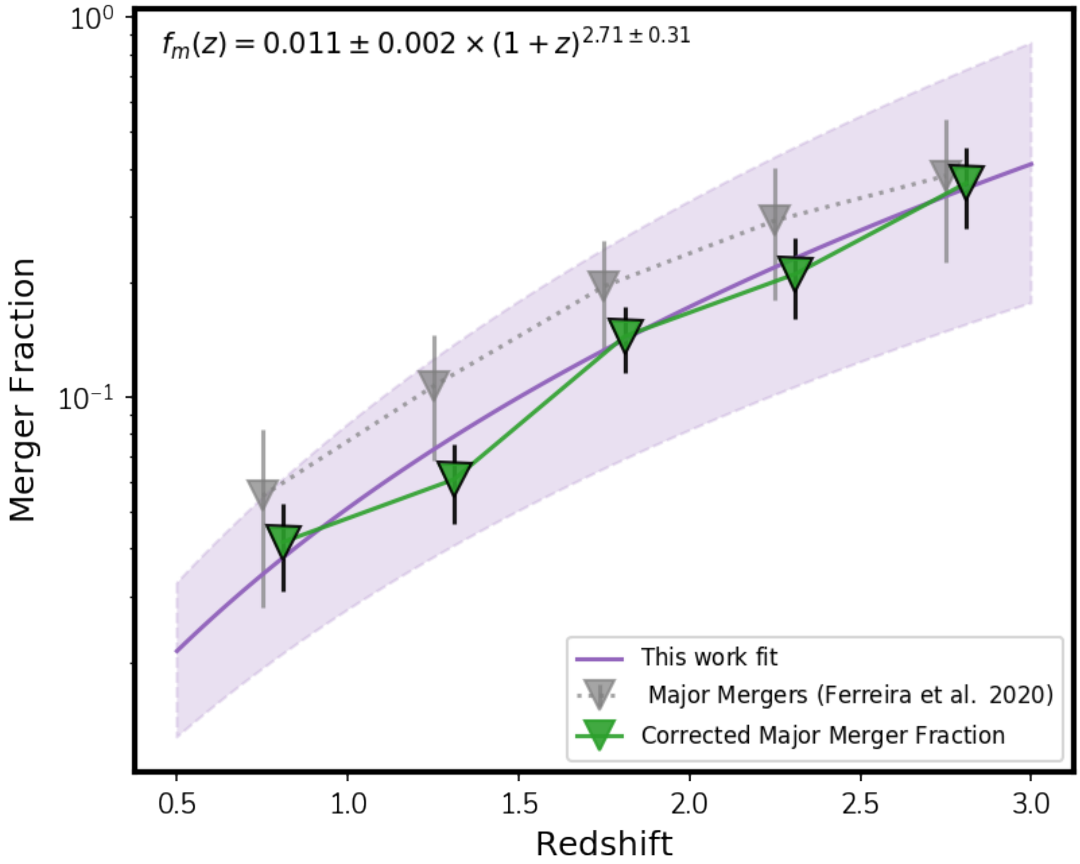


Figure 3.18: Major-merger fractions as a function of redshift. We show corrected merger fractions from Chapter 2 by re-classifying galaxies with our new method in mergers and non-mergers, shown in green. The original estimates are shown in gray.

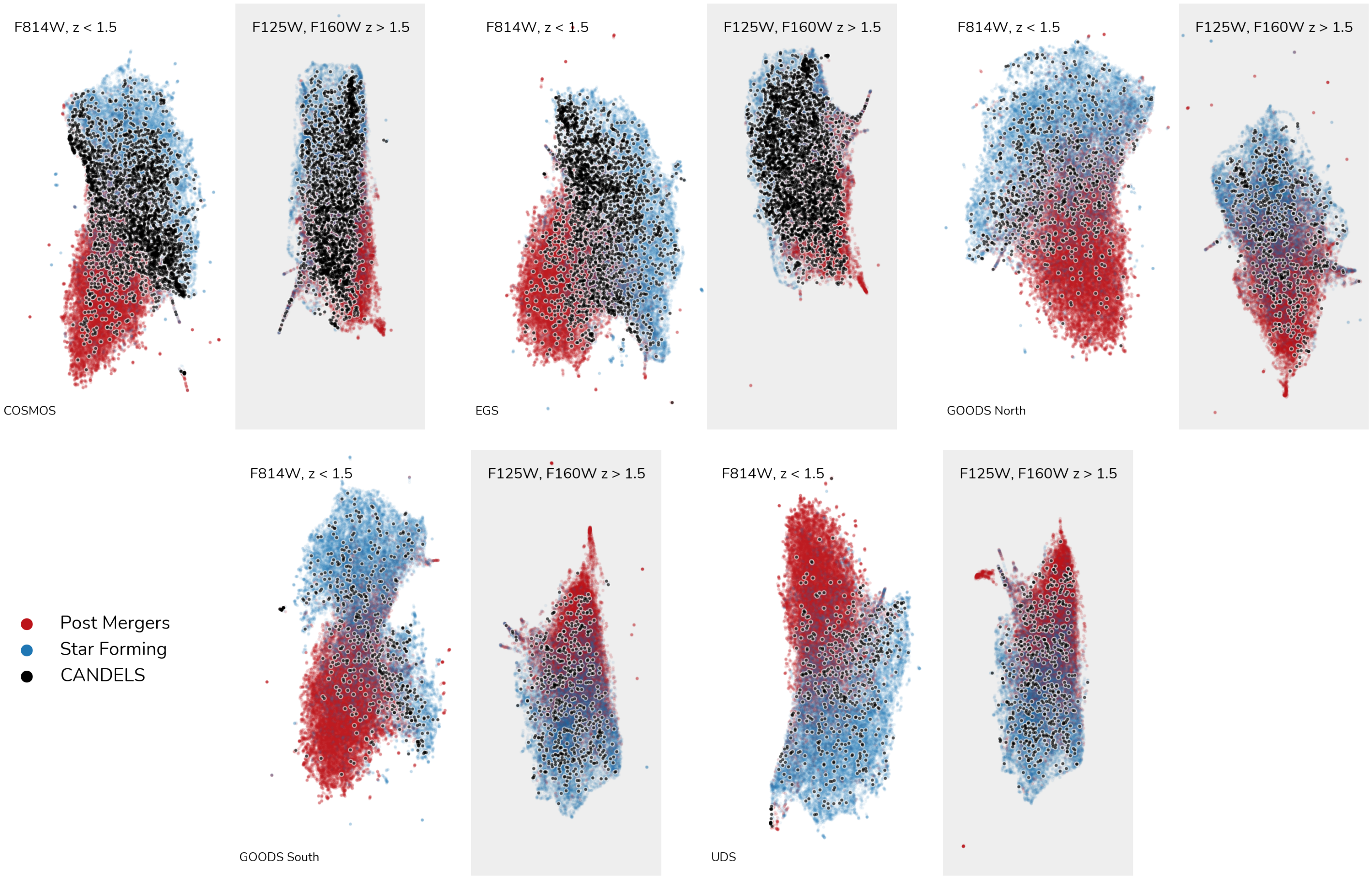


Figure 3.19: Extracted features by our networks in a UMAP 2D representation. For each model in our ensamble we generate a UMAP from the extracted features from the last convolutional layer of the trained networks, both when applied to the Illustris galaxies, color-coded by the class, and to the unlabeled CANDELS galaxies shown in black dots. Both Illustris and CANDELS extracted features populate the same region of this representation, showing that the features used by the network to then perform the classification task are in general domain invariant. Additionally, both classes – post-mergers and star forming galaxies – form separated clusters with some overlapping. Classification could be done in this representation alone, but it is then better organized by the fully connected layers that combine these features to produce the final output probability.

$$P(\mathbf{M}|\mathbf{S}) = \mathrm{P}(\mathbf{S}|\mathbf{M}) \times \frac{\mathrm{P}\,(\mathbf{M})}{\mathrm{P}\,(\mathbf{S})}, \tag{3.11}$$

where $P(\mathbf{M}|\mathbf{S})$ is the probability of a merger, given that a method used to select mergers, $(\mathbf{S})$, identifies it as such. The value of $\mathrm{P}(\mathbf{M})$ is the probability that an object is a merger before a selection of merger is made. $\mathrm{P}(\mathbf{S})$ is the probability that a galaxy is selected as a merger, whether a real merger or a false-positive. Because of the results of this Chapter, we know that this last number is very likely not equal to unity. It in fact can depend on various factors and methods of finding mergers. We can write the probability $\mathrm{P}(\mathbf{S})$ as:

$$P\,(\mathbf{S}) = P(\mathbf{S}|\mathbf{M}) \times \mathrm{P}(\mathbf{M}) + \mathrm{P}(\mathbf{S}|\mathbf{NM}) \times \mathrm{P}(\mathbf{NM}) \tag{3.12}$$

where **NM** standard for non-mergers, where $\mathrm{P}(\mathbf{S}|\mathbf{NM})$ is the probability of identifying correctly a non-mergers, and the value of $\mathrm{P}(\mathbf{NM})$ is the probability that the galaxy is not a merger. We can simplify this if we know, a priori, what the merger fraction is based on previous work. If we denote the merger fraction as f_{m}, and the machine learning probability of finding a merger/non-merger as p_{m} and p_{mn} then we can rewrite equation (3.11), as:

$$P(\mathbf{M}|\mathbf{S}) = \frac{\mathrm{p_m} \times \mathrm{f_m}}{(\mathrm{p_m f_m} + (1 - \mathrm{p_{nm}})(1 - \mathrm{f_m}))}. \tag{3.13}$$

Thus, for example, if the accuracy of a machine learning method for finding a merger is 0.9 and the accuracy for finding a non-merger is 0.9, and the merger fraction $f_{\mathrm{m}} = 0.1$, then the probability that a galaxy identified as a merger is actually a merger is $P(\mathbf{M}|\mathbf{S}) = 0.5$. This implies that even when the accuracy of finding mergers and non-mergers is 90%, at the lowest redshifts, where the merger fraction is low $\sim 10\%$, there is still a 60% chance that an identified merger is identified incorrectly as such. At higher redshifts, where the intrinsic merger fraction is higher, the probability of finding a merger correctly increases to $\sim 80\%$ when the merger fraction is as high as $f_{\mathrm{m}} \sim 0.3$.

We can generalise the equation for $P(\mathbf{M}|\mathbf{S})$, as a function of z, by considering how the merger fraction f_{m} evolves with redshift, such that:

$$P(\mathbf{M}|\mathbf{S},\mathrm{z}) = \frac{\mathrm{p_m(z)} \times \mathrm{f_m(z)}}{\mathrm{p_m(z) f_m(z)} + (1 - \mathrm{p_{nm}(z)})(1 - \mathrm{f_m(z)})},$$

$$f_m(z) \equiv f_0(1+z)^m,$$

which gives us a tool to understand how our classifications might be contaminated by sample imbalance effects with respect to redshift.

From this we can conclude that a significant fraction of individual galaxies within the CANDELS imaging may be incorrectly identified as either mergers or non-mergers. From our results here, our method effectiveness for correctly classifying mergers increases from $\sim 40\%$ at $z \sim 0.5$ to $\sim 70\%$ at $z \sim 3$. This is likely what can account for some of our misidentified galaxies as discussed in §3.5 when discussing the success of our method of separating star forming systems from those that are undergoing mergers.

These are conservative estimates which do not include the fact that we pre-select CANDELS galaxies based on their asymmetry. This should increase $P(\mathbf{M}|\mathbf{S},\mathrm{z})$ further since f_{m} is higher among galaxies with $A > 0.1$.

3.6.5 On Domain Adaptation Issues

There is a growing concern on the applicability of simulation trained deep learning models when applied to a intrinsically different domain. For us this is the case with going from cosmological simulations to real observations. When transferring from one domain to another, deep learning models might fail due to relying too much on domain specific features. Several techniques were developed to address this problem, focused on forcing neural networks to learn domain invariant features, leading to more robust models. Ćiprijanović et al. (2021) show that adopting techniques for domain adaptation could increase model performance when applying to the target domain by 20%.

In our case, the source domain is the IllustrisTNG galaxies and the target domain the CANDELS observations. To check if we need to apply domain adaptation techniques to this particular problem, we used UMAPs (Uniform Manifold Approximation and Projection, as described in Sec. 3.5.2) (McInnes et al., 2018) to reduce the high dimensional space generated by the features extracted by our network to a 2D-space that is easy to visualize[2]. Then, for each of our trained models we compare whether the features extracted by the network show similar distributions for Illustris and CANDELS galaxies. In Figure 3.19 we show UMAPs for each of the CANDELS fields models, for low redshift (left) and high redshift (right), color coded by their class in the case of Illustris and in black for real CANDELS galaxy images. As can be seen, these distributions of simulated galaxies and real observations are clustered together, with very few outliers not following the main cluster. Additionally, we can see that each class forms its own cluster, with overlapping regions, showing that features between classes are distinct and

---

[2] We also tested with t-SNEs with similar results.

in general not domain specific.[3]

We attribute the generalization success of our models to our mock data pipeline, which is tailored to mimic each individual CANDELS field with maximum fidelity – with their instrumental and observational features. Also augmentations with patches of the sky from CANDELS introduce real observations into our source domain, which not only make our training sets big (~ 140.000 images) but also help with domain confusion within the network. Thus, we do not include any domain adaptation process in our pipeline.

3.7 Summary

To shed light on the nature of peculiar/irregular objects at intermediate to high redshifts, we have constructed a framework based on forward-modelling of cosmological simulations with deep learning algorithms, that allows classifications with physically motivated labels based on the formation history of galaxies.

We used data from the IllustrisTNG TNG100-1 simulation to create realistic mocks of galaxies with CANDELS-like properties, including a full radiative transfer treatment with SKIRT for two specific classes of galaxies: post-mergers and non-merging star forming galaxies. These are selected so that their main difference is their formation history.

We produced a dataset of $\sim 160,000$ images of simulated IllustrisTNG galaxies with realistic visual properties that mimick CANDELS observations in the redshift range $0.5 < z < 3.0$. The images are used to train Deep Convolutional Neural Networks to distinguish between formation histories of post-mergers and star forming

---

[3] We also tested generating random noise images to check their position in this parameter space. As expected, they cluster away from the image regions, forming its own outlier region which is far from the main locus where galaxies are found.

galaxies. The main conclusions drawn from this work are summarized as follow:

- The classifier network combined with our new dataset produces classification models with a balanced performance of $\sim$ 80% accuracy, precision, and completeness when applied to a single-band imaging dataset, outperforming the asymmetry (A) by at least 25% within the simulated data. Additionally, for pristine images without any contamination and observational effects, the theoretical limit of our model is $\sim$ 91% accuracy. This is evidence that using the asymmetry (A) alone for ambiguous morphological cases might generate highly contaminated samples.
- We define two new contamination indicators, the overlapping percentage, Θ, and the average flux of the background sources, BG_{flux}, by leveraging how simulated galaxies are combined with true CANDELS background sky patches. Θ controls how sources overlap and are projected in the same stamp, while the BG_{flux} value probes the effect of the brightness of external sources on the classification of the central object. These allow us to explore in detail how deep learning classifications are impacted by contamination. We show that both crowded environments and projections and the relative brightness of external sources to the central galaxy negatively impacted our results. Based on this, we define quality control limits to our approach within the CANDELS fields as $\Theta \sim 10\%$ and $BG_{flux} < 10^{-3}$ e s^{-1} pix^{-1}. Although not universal, these limits provide guidelines for sample selection when applying our models to data.
- By applying our model to real CANDELS observations of galaxies with high asymmetries, we show that the relative fraction of post-mergers to star forming galaxies increases with higher redshift for two mass regimes. For

low mass sources ($9.5 < \log(M_*/M_\odot) < 10.0$), the post-merger fraction increases by $\sim 20\%$ within $0.5 < z < 2.0$, while the fraction of star forming galaxies decreases by $\sim 15\%$ in the same redshift range. In the high mass case ($\log(M_*/M_\odot) > 10.0$), the post-merger fraction increases by $\sim 25\%$ at $0.5 < z < 2.0$, while the fraction of star forming galaxies stays broadly constant.

- We explore the impact of major mergers on galaxies located above the star formation main sequence (SFMS) as parametrized by Schreiber et al. (2015). We separate CANDELS galaxies above the SFMS in the classes provided by our model and in bins of stellar mass. At $0.5 < z < 2.0$ we do not find any clear signs that major mergers play a critical role on the scattering above SFMS, with similar trends for post-mergers and star forming galaxies. However, in the highest redshift bin with good sample statistics ($2.0 < z < 2.5$) we see a post-merger driven SFR enhancement at lower masses of about ~ 0.1 dex.

- We show that the light distribution parametrized through Sérsic profiles of the CANDELS galaxies classified by our models as post-mergers are intrinsically distinct from those classified as star forming galaxies. The star forming galaxies sample is dominated by disk-like objects with an average Sérsic index of $n = 1.1^{+0.5}_{-0.5}$ while the post-mergers have more concentrated light profiles corresponding to higher central concentration with $n = 1.8^{+0.7}_{-0.6}$, with a long tail at higher Sérsic indices. Moreover, when we increase the probability threshold of our classifications to improve the purity of our selections, only the post-merger distribution display higher Sérsic indices. Evidently, our model predicts that post-mergers are more likely to be bulge-dominated galaxies.

- By using our updated data pipeline and models specifically tailored to distinguish between post-mergers and star forming galaxies, we revisit the merger fractions and merger rates from Chapter 2 by correcting ambiguous cases. This leads to updated galaxy merger rates that are slightly lower, but consistent with previously reported rates: $\mathcal{R} = 0.022 \pm 0.006 \times (1 + z)^{2.71 \pm 0.31}$.
- We show that our models use similar features to classify IllustrisTNG and real CANDELS galaxies, with no clear discrepancy between the two domains. Using the features extracted by the convolutional layers of our network, we generate UMAPs, which visualize the complex parameter space in 2D. Features of IllustrisTNG galaxies and CANDELS galaxies overlap for all the CANDELS fields. Although the CANDELS galaxies do not span the entire feature space of the IllustrisTNG galaxies used here, they are contained within that feature space.

Our machine learning driven approach provides a new way to investigate the formation history of galaxies with models that are informed by cosmological simulations. This includes the use of the models themselves, and the application of these models within accurate observing conditions.

Nevertheless, currently we are still limited to high-mass major merger cases due to resolution limitations from the simulations and mass completeness from the observations. In the upcoming years, combining the next generation of high resolution, small box simulations (e.g. TNG50-1, New Horizons) with observational data from the James Webb Space Telescope (JWST) and *Euclid* Telescope will open a new window to incorporate the effect of minor mergers and lower mass systems. Together, this will represent a major step towards uncovering unresolved questions of galaxy evolution.

Chapter 4

First Rest-frame Optical Observations of Galaxy Structure at z $>$ 3 with JWST in the SMACS 0723 Field

The content of this chapter has been accepted for publication (Ferreira et al., 2022b) in The Astrophysical Journal Letters. I acknowledge the help from Christopher Conselice and Elizabeta Sazonova for the visual classification effort reported here.

4.1 Abstract

We present early results regarding the morphological and structural properties of galaxies seen with the James Webb Space Telescope (JWST) at $z > 3$ in the Early Release Observations towards the SMACS 0723 cluster field. Using JWST we investigate, for the first time, the optical morphologies of a significant number of $z > 3$ galaxies with accurate photometric redshifts in this field to determine

the form of galaxy structure in the relatively early universe. We use visual morphologies and Morfometryka measures to perform quantitative morphology measurements, both parametric with light profile fitting (Sérsic indices) and non-parametric (CAS values). Using these, we measure the relative fraction of disk, spheroidal, and peculiar galaxies at $3 < z < 8$. We discover the surprising result that at $z > 1.5$ disk galaxies dominate the overall fraction of morphologies, with a factor of ~ 10 relative higher number of disk galaxies than seen by the Hubble Space Telescope at these redshifts. Our visual morphological estimates of galaxies align closely with their locations in CAS parameter space and their Sérsic indices.

4.2 Introduction

The James Webb Space Telescope (JWST) was launched on December 25, 2021 with its first operational image released to the public on July 11, 2022 by US President Joe Biden. This first image is a very deep image of the RELICS cluster SMACS J0723.3-732 (SMACS 0723, Ebeling et al., 2010; Repp & Ebeling, 2018; Coe et al., 2019). SMACS 0723 is massive cluster of galaxies at $z = 0.390$ which is also known to contain an extensive collection of strong gravitational arcs with a measured and modelled mass profile (e.g., Golubchik et al., 2022; Pascale et al., 2022). As this is the first JWST image to be released for a field where many objects have existing accurate photometric redshifts, it enables us to study the morphological evolution of galaxies with the earliest JWST data available.

Even before the release of the raw imaging, it was clear from the publicly released promotional color image that this cluster contained a collection of red and spiral galaxies that were not obviously present in the Hubble Space Telescope (HST) imaging. These observations provide the ideal resource for a first examination of the problem of how galaxy morphology changes from HST to JWST, and how rest-frame optical morphologies appear in the un-probed region $z > 3$. To understand this question better, we have undertaken an early analysis of this Early Release Observation (ERO) data released by the Space Telescope Science Institute (STScI) on July 13, 2022 to the public.

This Chapter is a first-look study of the morphological evolution of galaxies seen in the field around SMACS 0723, giving us our first look at how galaxy structure changes with redshifts up to $z = 8$. This is the first field where this analysis can be performed due to the limited available accurate photometric redshifts in other, early release observations.

Galaxy structure and morphology are one of the key aspects for understanding galaxy evolution, and will be a key measurement that JWST will make throughout its lifetime. Following the first servicing mission to Hubble, distant galaxies started to have their structure resolved in the mid-nineties. This revealed that faint, distant galaxies appear more peculiar and irregular than local ones, and cannot be easily classified on the Hubble sequence (Griffiths et al., 1994; Dressler et al., 1994; Driver et al., 1995; van den Bergh et al., 1996). Why that is the case has remained a major topic of discussion for almost three decades. These early observations, however, only showed that galaxies became more peculiar at fainter magnitudes, which did not necessarily correlate with further distances.

When redshifts became available, at first within the Hubble Deep Field, it was clear that galaxy structures evolve strongly and systematically with redshift, such that peculiar galaxies dominate the population at $z > 2.5$ (e.g., Conselice, 2003; Papovich et al., 2005; Elmegreen & Elmegreen, 2005; Dahlen et al., 2007; Buitrago et al., 2008; Conselice et al., 2008; Huertas-Company et al., 2009; Buitrago et al., 2012; Mantha et al., 2018). It is now well-established that galaxies as observed with HST become smaller and more irregular/peculiar at higher redshifts, and this has been accounted for by the merger process for a significant fraction (40-50%) of systems (Conselice et al., 2014).

However, although HST was revolutionary, morphological evolution measurements still suffer some limitations. First among these is that due to HST's limited red wavelength coverage, we have not measured the rest-frame optical light of galaxies within the first two Gyr after the Big Bang, that is at $z > 3$. Very few galaxies have been observed in the rest-frame optical bands at such redshifted wavelengths, and most of these utilise ground-based adaptive optics. This is due to the fact that the F160W band on HST only probes rest-frame optical light up to $z \sim 2.8$, whereas JWST permits us to obtain this information up to $z = 8$ with F444W, and even beyond with MIRI. Second, HST infrared imaging does not provide the necessary spatial resolution to resolve most high-redshift objects. Furthermore, we have found in previous observations (e.g., Conselice et al., 2005; Huertas-Company et al., 2009; Mortlock et al., 2013; Huertas-Company et al., 2016) that the number of galaxies that are classifiable as disks or spheroids (including ellipticals) declines quickly when observing systems at higher redshifts, up to $z = 3$.

Observations of galaxy structure and morphology at $z > 3$ do show that in the rest-frame UV, galaxies are peculiar and irregular (e.g., Elmegreen & Elmegreen, 2005; Conselice & Arnold, 2009). Galaxies at these redshifts are also often found to be clumpy, as seen with deep Wide Field Camera 3 (WFC3) data (Elmegreen & Elmegreen, 2005; Oesch et al., 2010; Margalef-Bentabol et al., 2018; Whitney et al., 2021; Margalef-Bentabol et al., 2022). Measurements of galaxies in pairs also demonstrates that the merger rate and the fraction of galaxies in mergers at $z \sim 6$ is as high as 50% (e.g., Conselice & Arnold, 2009; Duncan et al., 2019).

This implies that galaxy structure should likewise be distorted accordingly (e.g., Duncan et al., 2019; Shibuya et al., 2022). At the same time, we have believed for 30 years that the Hubble sequence is established quite early at $z \sim 1$ (e.g., Mortlock et al., 2013; Huertas-Company et al., 2016). However, all of these conclusions are based on HST imaging, which has now been superseded in significant ways by the redder bands, higher resolution, and better sensitivity of JWST.

Thus, in this Chapter we explore the morphological properties of the earliest galaxies through an approach based on galaxy classification and measurement. We demonstrate that these early galaxies have a more normal morphology than expected, with classifications showing that disk galaxies are much more common than previous observations suggested (e.g., Conselice et al., 2005; Huertas-Company et al., 2009; Conselice et al., 2014; **?**; Margalef-Bentabol et al., 2022). Overall, we argue that the formation of the Hubble sequence appears to be ongoing much earlier than we had anticipated based on HST observations.

We opt for a more classical approach to this novel dataset instead of the deep learning methods discussed in Chapter 3 and Chapter 4 due to the lack of visual classifications available, and the lack of cosmological simulations focused on the high redshift universe. This is the first step on moving towards the application of supervised methods to JWST data.

This Chapter is organized as follows: in § 4.3 we describe the data and our methods and outline. In § 4.4 we describe the morphological results of our study, § 4.5 is a short discussion of our results, and § 4.6 is an overall summary. Throughout this Chapter we assume a Λ cold dark matter cosmological model with $\Omega_\Lambda = 0.7$, $\Omega_\mathrm{M} = 0.3$ and $H_0 = 70$ km s^{-1} Mpc^{-1}. All magnitudes are given in the AB system (Oke, 1974; Oke & Gunn, 1983).

4.3 Data Reduction and Products

The data we use for this analysis originates from the Early Release Observations of SMACS 0723 (Pontoppidan et al., 2022) and include observations taken with the *Near Infrared Camera* (NIRCam; Rieke et al., 2005, 2008, 2015). The images were obtained on June 06, 2022 (PI: Pontoppidan; Program ID 2736) in the F090W, F150W, and F200W short-wavelength (SW) bands, and F356W, F277W, and F444W long-wavelength (LW) bands. The total integration time for this target is 12.5 hr. Figure 1 shows the combined color image of SMACS 0723 which we created from our own reduction.

We reprocess the uncalibrated lower-level JWST data products following a slightly modified version of the JWST official pipeline. This is because the initial release of the higher-level data products have been found to contain WCS alignment issues as well as sub-optimal background subtraction. The key differences are as follows: (1) We use version 1.5.2 of the pipeline as opposed to version 1.5.3, which was the most up-to-date version at the time of writing. This is because version 1.5.3 has a significant bug in the background subtraction step that led to sub-optimal performance.[1] (2) We apply the CEERS 1/F noise and flat field correction (Bagley et al. in prep) between stages 1 and 2 of the official pipeline. (3) We extract the `SkyMatchStep` from stage 3 and run it independently on each NIRCam frame, allowing for quicker assessment of the background subtraction performance and fine-tuning. (4) After Stage 3, we align the final science images onto a GAIA-derived WCS using `tweakreg`, part of the DrizzlePac python package[2]. We then pixel match the images with the use of `astropy reproject`[3]. Finally, we re-align the RELICS SMACS 0723 HST imaging to the GAIA DR2

---

[1] https://github.com/spacetelescope/jwst/issues/6920

[2] https://github.com/spacetelescope/drizzlepac

[3] https://reproject.readthedocs.io/en/stable/

catalog due to large $1''$ offsets, and match it with `astropy reproject` as well. We then apply astrometric corrections to the positions of sources available in the RELICS catalogs.

Overall, this data set allows us to probe the rest-frame optical images of galaxies out to $z = 8$. In Figure 4.2, we show the rest-frame wavelength probed by each individual filter and how they can be combined together for up to $z = 8$ optical rest-frame coverage. In addition to this, we combine our observations with HST Archival data in the WFC3 F160W band.

We employ two different approaches to the SMACS 0723 data: first a quantitative analysis using MORFOMETRYKA (Ferrari et al., 2015), where we measure non-parametric morphology estimates such as concentration, asymmetry, and smoothness (CAS; Conselice et al., 2003); Gini-M20 (Lotz et al., 2004), various sizes, as well as light profile fitting, which is described in detail §4.3.2. Second, we provide simple visual classifications for all sources with $S/N > 10$ in their optical rest-frame filters, described in detail in §5.3.3.

4.3.1 Photometric Redshifts

We use photometric redshifts derived through the Bayesian photo-z code (BPz) (Benítez, 2000; Benítez et al., 2004; Coe et al., 2006) by the RELICS program (Salmon et al., 2020), which used HST imaging in 7 bands for 41 clusters, including SMACS 0723, and archival *Spitzer* IRAC measurements to measure photometric redshifts of galaxies up to $z = 8$. The BPz code compares RELICS fluxes to 11 templates for ellipticals, spirals and starburst galaxies. The overall method for this is described in detail in Coe et al. (2019).

For our analysis, we selected 355 galaxies with RELICS photometric redshifts

Figure 4.1: JWST color image of SMACS 0723 showing the overall distribution of galaxy shapes and morphologies, including the lensing arcs. This image was produced from our reduced data products via a composite of data in 6 bands: F090W, F150W, F200W, F277W, F356W, and F444W. F090W and F150W were assigned blue colours, F200W and F277W green, and F356W and F444W orange and red respectively.

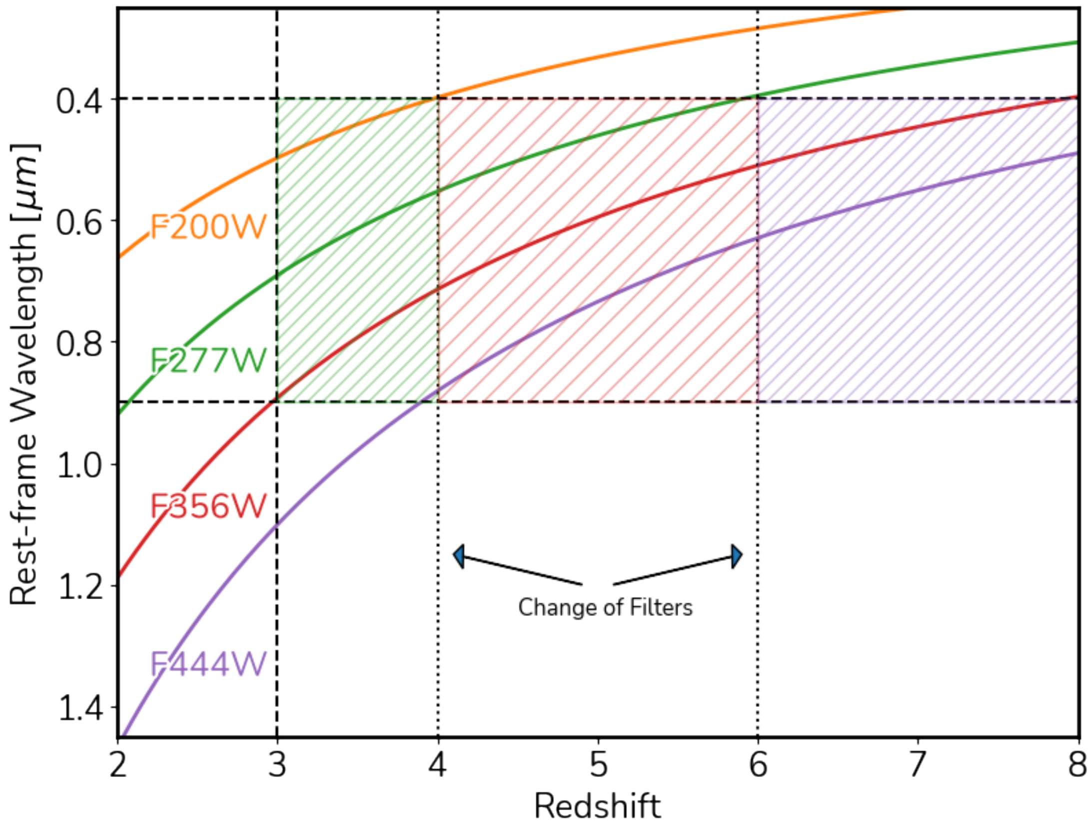


Figure 4.2: Rest-frame wavelength at a given redshift for the F200W, F277W, F356W, and F444W filters. The hatched regions show the areas where NIRCam filters probe the optical rest-frame for $z > 3$, with the color corresponding to the respective optimal filter for the redshift range. To follow galaxies' optical rest-frames we use F277W for $3 < z < 4$, F356W for $4 < z < 6$, and F444W for $6 < z < 8$.

from the JWST footprint of the SMACS 0723 field, restricting our sample to $1.5 \leq z \leq 8$. The distribution of redshifts across the SMACS 0723 field is shown in Figure 4.3. From these 355 sources 280 were considered classifiable, excluding stars and point sources.

As an additional sanity check of these redshifts, and to see if they could be improved upon, we conduct our own SED fitting procedure with the use of LePhare (Arnouts et al., 1999; Ilbert et al., 2006). Within LePhare, we use the COSMOS galaxy templates (Ilbert et al., 2009) which are based on the commonly used BC03 template set (Bruzual & Charlot, 2003). These templates are modified with dust attenuation up to $E(B-V) = 1.5$ (Calzetti et al., 2000) and attenuation from the IGM following Madau (1995). We initially run the SED fitting process on the original RELICS photometry and obtain strong agreement with their BPz based redshifts. We then add in photometry from an F200W selected NIRCam catalogue which is cross matched to the RELICS catalogue with a 0.5 arcsecond tolerance. To be consistent with the photometry derived in the RELICS catalogues, we use isophotal magnitudes as measured by SExtractor (Bertin & Arnouts, 1996) for our sources. We find that the photo-z's are consistent with the original RELICS estimations when the bluest NIRCam bands available (F090W and F150W) are added to the SED fitting procedure. However, when photometry from F200W and redwards are added, we find that some originally high redshift sources ($z > 3$) in the RELICS catalogue are given new solutions at $z < 1$, this is found to be the result of a lack of a strong Balmer break at $\sim 2\mu$m. Examining these sources in detail reveal them to be classified as star-forming disks in our later analysis. Estimations of the proper size and absolute magnitudes of these objects reveal many of the new low-z solutions to give extremely small (≤ 0.1 pkpc) and faint

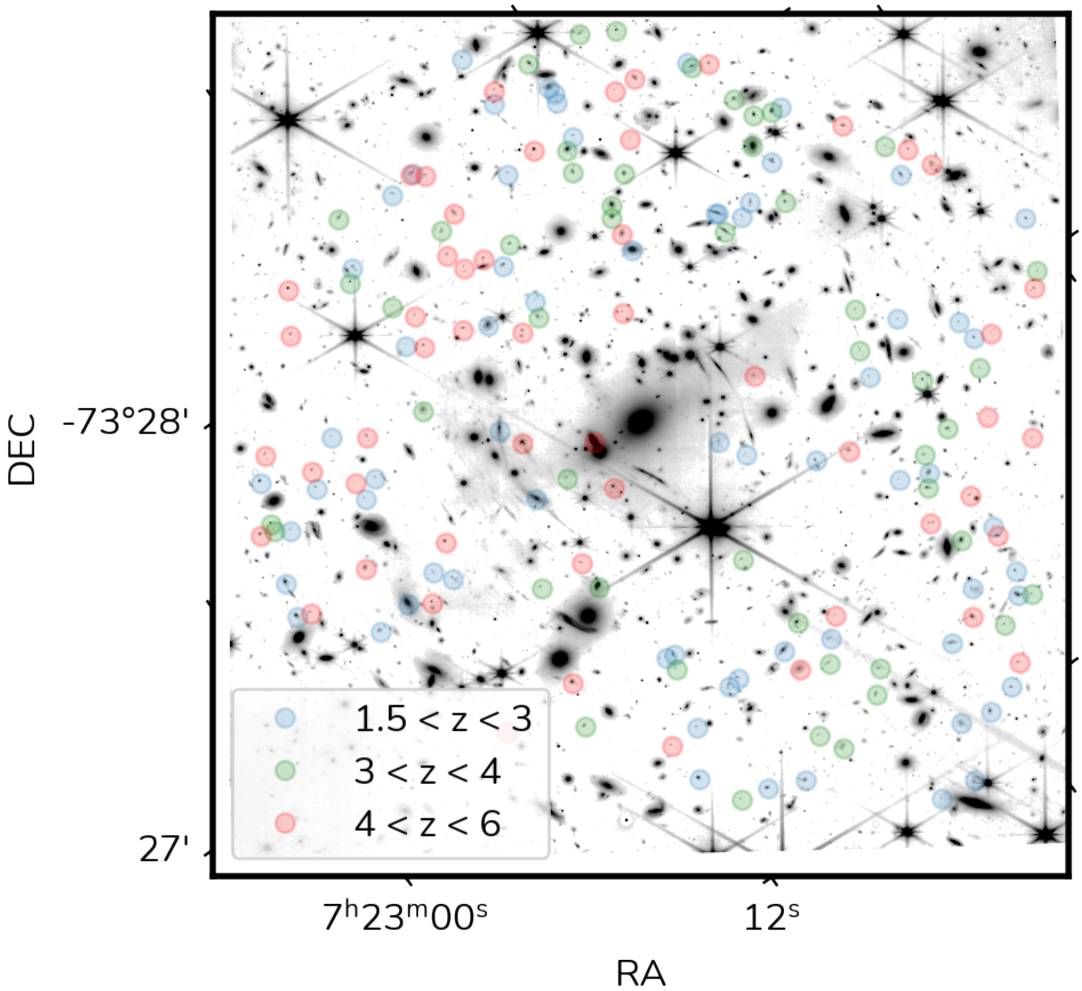


Figure 4.3: The distribution of redshifts across the SMACS 0723 field. Different redshift regimes, where we use our methodology to investigate the morphological evolution, are shown as differently-colored markers.

($M_{F415W} > -16$) properties for the sources[4]. Additionally, the current dataset was reduced prior to the calibration updates based on on-flight observations, which had large offsets on zero-points of up to ~ 0.4 mag (Adams et al., 2022; Rigby et al., 2022). This observation, combined with the subsequent disk classification lead us to proceed with the original RELICS BPz redshifts for all sources.

---

[4] The results presented here are robust against putting these sources at lower redshifts since they represent a small fraction of the overall sample.

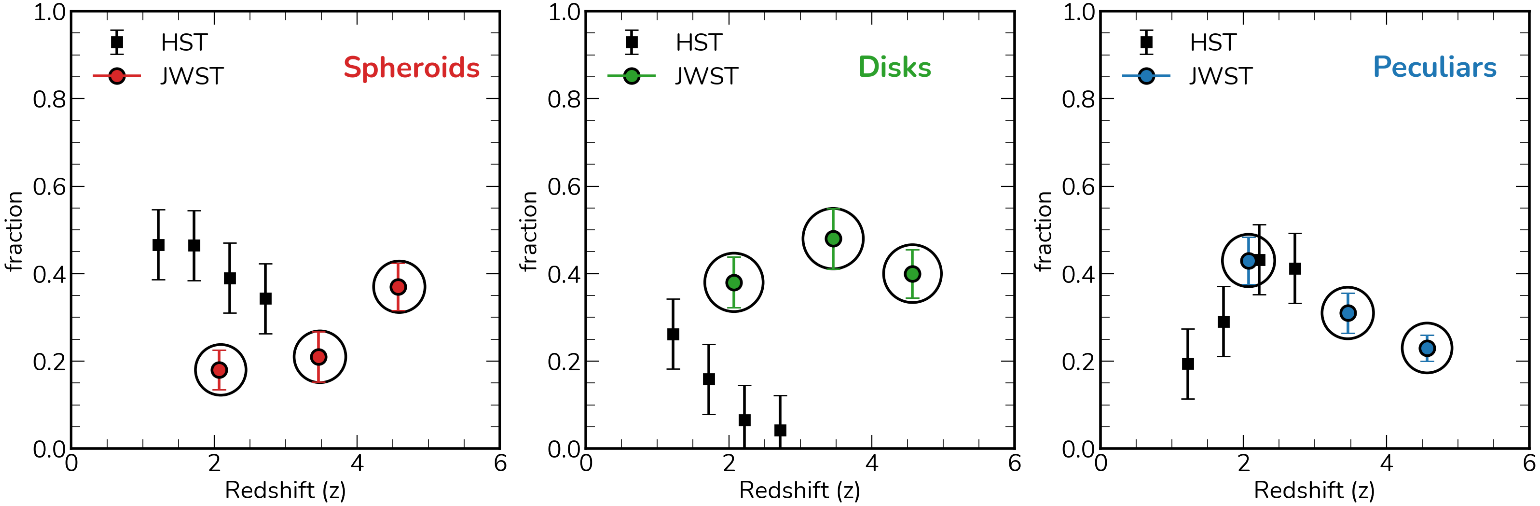


Figure 4.4: Plots showing the morphological evolution of the galaxies found in the SMACS 0723 field up to $z = 6$. These show the fraction of the total number of galaxies, within a given redshift bin, which has the given type as determined by visual morphologies. We also include the morphological evolution which has been derived from HST observations of the CANDELS fields in Mortlock et al. (2013). Circled markers denote the JWST observations at higher redshifts. We note that the increase in spheroids can be attributed to smaller sizes with increasing redshift z as discussed in § 4.4.1.

4.3.2 Quantitative morphologies: Morfometryka

MORFOMETRYKA was designed to perform several structural measurements on galaxy images, in an automatic non-interactive way (Ferrari et al., 2015; Lucatelli & Ferrari, 2019). It was devised mainly to measure non-parametric morphometric quantities, but also performs single-component Sérsic model fitting. It takes as input the galaxy and point spread function (PSF) images, estimates the background with an iterative algorithm, deblends the sources and defines which one is the target. Then, it filters out external sources using GalClean[5] (**?**). From the segmented region it calculates basic geometrical parameters (e.g. center, position angle, axial ratio) using image moments. Following this, it performs photometry, measuring fluxes in ellipses with the aforementioned parameters. Along the way, it masks point sources over the ellipse annulus with a sigma clipping criterion.

From the luminosity growth curve it establishes the Petrosian radius and the Petrosian Region, inside which all measurements are made. The 1D Sérsic fit is performed on the luminosity profile. For robustness, the 1D outputs are used as inputs for a 2D Sérsic fit done with the galaxy image and JWST PSF images generated with the official package `WebbPSF`[6]. MORFOMETRYKA uses the PSF to produce the Sérsic profiles and to mask an area of the size of the PSF FWHM from the central region of the source stamp for non-parametric morphology calculations. Even though these simulated PSFs are realistic, we note that deviations from the true PSF might exist. However, as we are interested in extended sources, effects of this type are neglible. Finally, MORFOMETRYKA measures several morphometric parameters (concentration; asymmetry; Gini; M20; entropy, spirality, and curvature, among others).

---

[5] GalClean – https://github.com/astroferreira/galclean

[6] `https://webbpsf.readthedocs.io/`

4.3.3 Visual Classification

All galaxies in the sample were classified by three co-authors of this Chapter with experience with galaxy structure and classification (CC, LF, ES). The visual classification scheme that we use is described in detail in Ferreira et al. (2022; in prep). Here we give a quick summary. In general, we use four categories for visual classifications, following Mortlock et al. (2013); these are defined as described below, and differ slightly from a traditional Hubble classification scheme, but are generally very similar.

- Class 0: Unclassifiable: Galaxies too small and/or too faint to classify, and images with artifacts.
- Class 1: Spheroids: These galaxies are resolved, symmetrically and centrally concentrated, with a smooth profile and are round/elliptical in shape.
- Class 2: Disks: This category includes galaxies that exhibit a resolved disk in the form of an outer area of lower surface brightness with a regularly increasing brightness towards the center of the galaxy. This classification does not depend on there being a spiral pattern in the system, although one can be present in this classification.
- Class 3: Peculiar. This class is for well-resolved galaxies with a morphology which is dominated by a disturbance or peculiarity and has no obvious disc or spheroid component.

Each galaxy is further classified as smooth or structured, where structured galaxies have features standing out from the smooth stellar envelope, such as star-formation clumps, tidal features, and merger signatures. Galaxies with distinct disk and bulge components were also classified as structured. Finally, the classifiers were able to provide additional notes on each source to aid in future analysis. We do not however use these further detailed morphologies in this Chapter.

These classifications were all carried out separately and then combined into an average, which we then use throughout this work. Our process was such that we only classified each galaxy in our sample in the wavelength which most closely matches the rest-frame optical wavelength of the observations. This allows us to match the classifications at different redshifts to determine how morphological evolution is occurring. We find that the classifications by the three classifiers agree 2/3 in $\sim 63\%$ (177) of the sources while perfect agreement 3/3 happens $\sim 33\%$ (87) of the time. Catastrophic classifications where all 3 classifiers disagree happens only in $\sim 5\%$ (16) of the cases. Table 4.1 shows the average of the fractions for each of the three classifiers for the three different types (sph, disk and peculiar). Also listed as the error-bar on these averages is the standard deviation of the fractions among the classifiers, showing that these are always relatively small and in some cases the agreement is to within a few percent.

4.4 Results

4.4.1 Distribution of Morphology with Redshift

One of the main questions that we can investigate with this early imaging from JWST is the distribution of morphological types with redshift. Given the redshifts we have from HST and the morphologies from JWST we can make the first measurement of the morphological distribution of galaxies up to $z < 8$.

In Figure 4.4 we plot the morphological distribution of our sample of galaxies

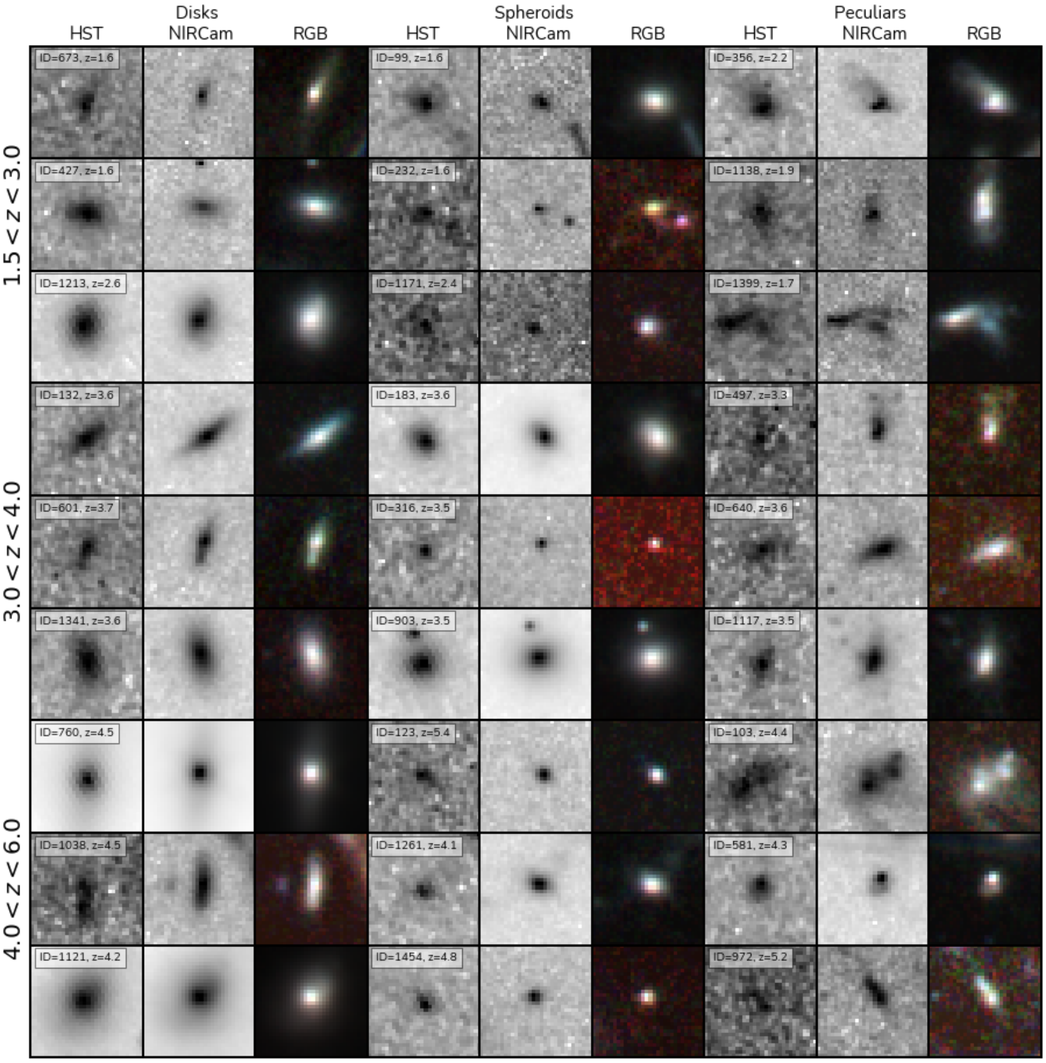


Figure 4.5: Comparison HST vs JWST images for 9 objects in each class within our sample. Left columns shows the HST F160W image, whilst the middle panel shows the JWST view at the same orientation, in the rest-frame optical. The far right panel shows the color image of this system as seen through the JWST NIRCam F277W, F356W and F444W filters, generated by TRINITY (Coe et al., 2012). The IDs refer to the original RELICs catalogs identification.

Table 4.1: The filters used in this study with the redshift ranges used within that filter and the average rest-frame wavelength in which that filter probes at that redshift. Also shown are the classification fractions for our three main types of galaxies; spheroids, disks, and peculiars. These are the averages of the three classifiers with the standard deviation listed as the value σ. Sérsic indices for each class and redshift bin is also provided, showing the mean of the distribution together with 15% and 85% percentile scatter.

Filter	z	<Rest-λ >	class (sph)	σ(sph)	class (disk)	σ(disk)	class (pec)	σ(pec)	n_{sph}	n_{disk}	n_{pec}
F090W	1.5-3.0	0.36-0.22 μm	0.18	0.02	0.38	0.10	0.43	0.12	$1.32^{1.6}_{1.14}$	$1.12^{1.62}_{0.58}$	$0.76^{1.54}_{0.37}$
F227W	3.0-4.0	0.69-0.55 μm	0.21	0.08	0.48	0.05	0.31	0.03	$1.22^{1.32}_{1.12}$	$1.04^{1.3}_{0.8}$	$1.04^{1.53}_{0.31}$
F356W	4.0-6.0	0.71-0.50 μm	0.37	0.01	0.40	0.02	0.23	0.03	$1.58^{1.81}_{1.08}$	$1.11^{1.29}_{0.81}$	$1.67^{1.09}_{1.26}$

with morphological classifications from the JWST imaging. As can be seen, we find a remarkable increase in the number of disk galaxies over what was thought to exist in previous analysis of the deepest HST imaging in the NIR, which found that that there were very few disk galaxies at $z > 1.5$, with a rapid decline in the numbers at higher redshifts (e.g., Conselice & Arnold, 2009; Mortlock et al., 2013; **?**).

More generally, a decline in spheroids was also seen, but the morphological change with redshift was not as pronounced as it was for the disk galaxies. Figure 4.5 shows some examples of different galaxy types and how they appear differently in the JWST vs. the HST imaging, revealing that morphologies are often much easier to make out within the JWST data. There is however, a propensity for these disks and spheroids to contain peculiar features, such as tidal features and clumpy regions, that can differentiate them from $z = 0$ examples. We however do not investigate these further in this Chapter.

Overall, we find that the disk galaxy population makes up about half of the galaxies that are identified within the field of SMACS 0723 at $z > 2.5$. This is a remarkable result, as it shows that galaxies such as the Milky Way could potentially have retained the same overall morphological state for over 12 billion years if these distant disk galaxies are similar to the ancestor galaxy of the Milky Way.

Morphology at $z > 4$

The morphologies of $z > 4$ galaxies, as probed by the F356W filter, differ from the rest-frame UV morphologies determined from HST imaging (Conselice & Arnold, 2009). Despite what might have been expected from HST observations,

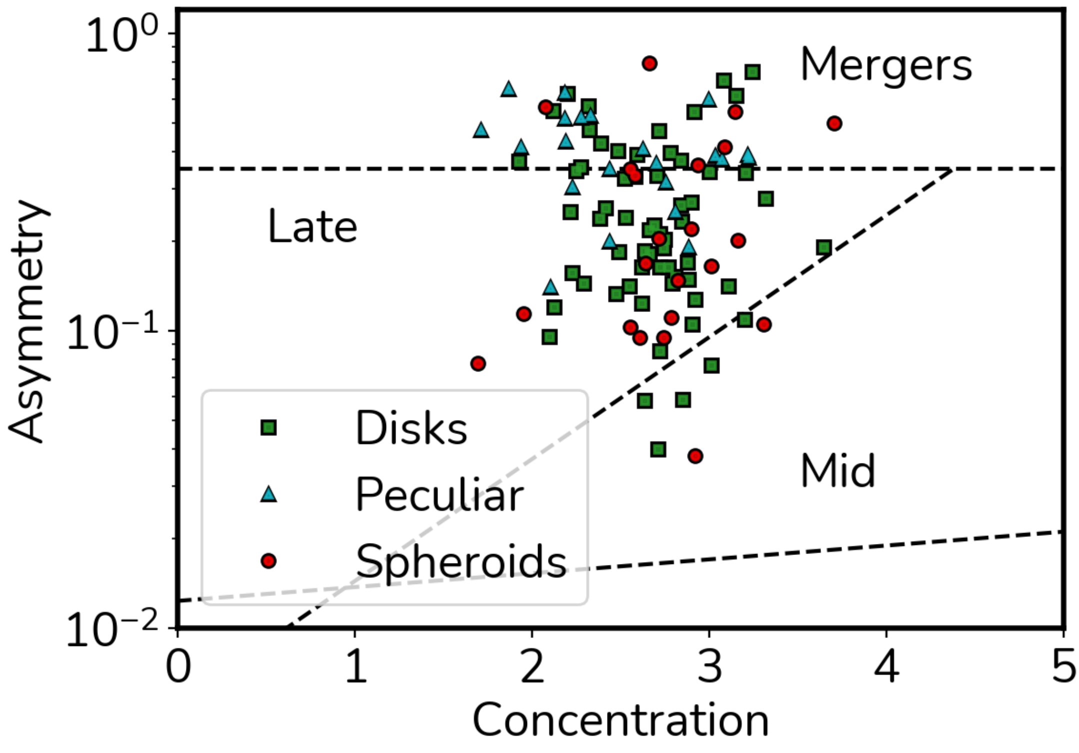


Figure 4.6: Concentration vs. Asymmetry. The abundance of disks results in their populating the area originally used for selecting late-type galaxies. Decision thresholds shown here are drawn from (Bershady et al., 2000).

the morphologies of at least the brightest galaxies are much less distorted than had been previously been thought based on HST observations. Also, the galaxies at this epoch are often very tiny, such that their size in a NIRCam image is just a bit larger, or within, the PSF of JWST. In fact, the larger number of spheroids/compact objects is in part due to the fact that so many of these systems are unresolved, an indication that their sizes are quite tiny. Future studies carefully measuring sizes with the use of the JWST PSF will examine these sizes and their evolution.

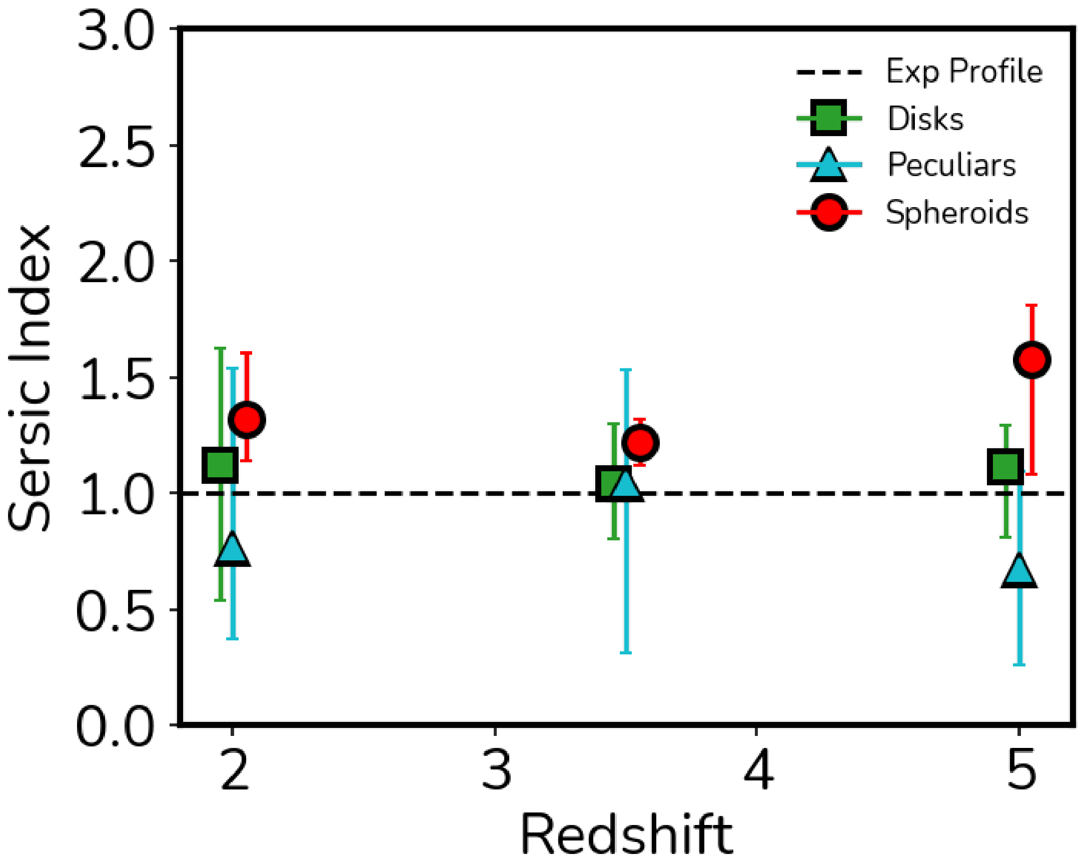


Figure 4.7: Plot showing the Sérsic index evolution in bins of redshift. Shown are the morphological types and their mean values for these indices alongside error bars representing the 15% and 85% percentiles of the distribution. We can see that the average hovers about $n = 1$, but that the spheroids appear to have higher values even at the higher redshifts. This is in contrast to low redshift Sérsic index distributions, where the populations of spheroids have higher Sérsic indexes with $n \sim 4$, while the disk galaxies generally.

4.4.2 Quantitative Morphologies

We quantify the structures of these galaxies based on the NIRCam imaging. This can be done in a number of ways, and will be the focus of future dedicated papers. We give a broad overview of quantitative morphology for our sample and leave it to future papers to elaborate on these issues.

First we show the concentration-asymmetry diagram, which has been used to classify galaxies and has a broad correspondence with galaxy types in the nearby universe (e.g., Conselice et al., 2000; Conselice et al., 2003; Conselice et al., 2008; Bluck et al., 2012; Whitney et al., 2021). We present the concentration-asymmetry diagram in the SMACS 0723 field in Figure 6 as measured by MORFOMETRYKA. What we find is that there is no great distinction between the disks and the spheroids, but we do find that the peculiars are in the region of high asymmetry where mergers are located (Conselice et al., 2000; Whitney et al., 2021). We also find that there are few galaxies with very high concentration values, consistent with previous work that found even massive galaxies to have low light concentrations (Buitrago et al., 2012).

Another avenue of investigation is the examination of the light profiles of our galaxies, which we have also measured. Previous work has shown that almost all massive galaxies at $z > 2$ have Sérsic indices which are $n \sim 1$, which differs for galaxies at lower redshifts where $n \sim 4$ for similar mass galaxies (e.g., Buitrago et al., 2012; Bluck et al., 2014). We show a basic view of the average Sérsic index evolution for our sample in Figure 5.7, which demonstrates that many of our galaxies contain indices with $n \sim 1$, with most disks around this value, as expected. We also find many spheroids at this Sérsic index, but on average these spheroids have a larger n value.The mean values together with 15% and 85%

percentiles of the distributions for each redshift bin and class are displayed in Table 4.1.

4.4.3 Formation of the Hubble Sequence

One of the primary goals of galaxy morphology and structural analysis is to determine when the Hubble sequence was established. By Hubble sequence we mean the establishment of spheroids (e.g., ellipticals) and spiral galaxies as we see in the case of the most massive galaxies in the nearby universe. We know for certain that there are fewer ellipticals and spirals at high redshift. However, an important question is: when did the first spheroids and disk galaxies form?

It is important to be clear about what we mean by this, as a definition of these galaxy types is neither trivial nor simple. By a 'spheroid' we mean a galaxy that exhibits a round or elliptical shape with a classical, steep light profile and a smooth structure. A 'disk galaxy' is one that is either a smooth, disk-like object, or something with visible spiral arms.

The trend of galaxy type with redshift has been measured by several different papers (Wuyts et al., 2011; Mortlock et al., 2013; Huertas-Company et al., 2015; Zhang et al., 2019). We include the analysis in Mortlock et al. (2013) as the basis for our understanding of the morphological evolution at $z < 3$. As Figure 4.4 shows, we have not reached the limit of where the first ellipticals and spheroids have formed. We will need to probe even higher redshifts to find when and if there are no spheroids or disk galaxies. Thus, at least some aspect of the Hubble sequence was in place at $z \sim 6$. It is, however, important to point out that these classifications are done purely by visual estimates in one band. We have not used colour or other features to classify galaxies, and we know from work with WFC3 that galaxy structure and physical properties of their stars becomes decoupled at higher redshifts $z > 1.5$ (Conselice et al., 2011). It remains to be seen how the physical properties of our 'Hubble types' here correlate with the underlying stellar masses in these systems.

4.5 Discussion

This is one of the earliest studies on the morphologies of galaxies at high redshift with JWST, and thus our conclusions will be revisited by others in the months and years to come. However, it does appear from an initial analysis that there are far more disk galaxies at high redshift than originally thought with HST. We in fact find that at the highest redshifts probed by the HST CANDELS results (Mortlock et al., 2013) there are up to 10 times more disk galaxies than we had thought, based on the JWST visual morphologies. Although some studies find higher fractions of disks among massive galaxies at high redshift (Wuyts et al., 2011; Zhang et al., 2019) in HST in comparison to Mortlock et al. (2013), they are often restricted to only the population of star-forming galaxies, which are generally dominated by disks or elongated shaped galaxies. Even with the sample mismatch, our fractions are still higher than the star-forming fractions from (Zhang et al., 2019) by about 10%, for example.

This implies that disk galaxies have existed in large numbers for quite a significant amount of time. This may mean that the morphologies of some disk galaxies, such as the Milky Way, have remained in their current form for over 10 billion years. This would challenge our ideas about mergers being a very common process, and it might be the case that mergers are only a dominant process for forming the stellar masses of certain types of galaxies, namely spheroids, which have a relatively constant merger fraction at $z > 2.5$ at around 10%. Although on average galaxies should go through multiple mergers over cosmic time (Duncan et al., 2019), it is not clear how these mergers would affect disk morphologies or if there are only certain galaxies that go through mergers multiple times while others, such as the disks we find here, do not undergo these mergers very often or at all at $z < 6$.

Alternatively, it is also possible that these high redshift disks undergo major mergers, but reform their disks after the disruptive event. This is a process that is found to happen in simulations of gas-rich mergers (e.g., Sparre & Springel, 2017; Peschken et al., 2020).

There are a few caveats with this study that future studies will be able to flesh out in much more detail. The first is that we only use the visual rest-frame optical morphology of a galaxy to determine whether or not it is a spheroid, disk, or a peculiar. These systems, however, are more obvious than they were in the HST imaging, implying that in the rest-frame optical we are seeing the underlying morphology in a much clearer way than we are in the rest-frame UV, despite strongly star-forming galaxies having a very similar appearance in the UV and optical, at least at $z < 3$ (Windhorst et al., 2002; Papovich et al., 2005; Taylor et al., 2015; Mager et al., 2018). It would appear that at least disk galaxies are not easily seen in the UV, and this is an indication that their stellar population and star formation histories are spatially segregated (old/young stars in bulges/disks, for example), just as they are at lower redshifts. Future studies will certainly be able to study these resolved structures in more detail to learn about the detailed process of disk formation, as done for HST observations at $z < 3$ (Huertas-Company et al., 2016; Margalef-Bentabol et al., 2022).

Finally, there is also the fact that this was conducted in a small field of view area of 2.2'x2.2', around a lensing cluster. In the future we will probe non-cluster regions and larger areas to allow for a more detailed comparison with previous results from HST imaging, such as CANDELS (Grogin et al., 2011; Koekemoer et al., 2011).

4.6 Summary and Conclusions

In this Chapter we present a morphological and structural analysis on some of the earliest galaxies imaged by the JWST, which has provided rest-frame optical morphologies and structures for a statistically significant number of galaxies at $z > 3$ for the first time. We also examine the structures of galaxies at $1.5 < z < 3$, where HST has not had the depth and resolution to infer galaxy morphology correctly. Three of the authors classified 280 galaxies visually at $1.5 < z < 8$ to determine basic morphological types - spheroid, disk, peculiar, at rest-frame optical wavelengths given by JWST. We also ran quantitative parametric and non-parameteric morphologies on these galaxies.

Our key findings are:

I. The morphological types of galaxies changes less quickly than previously believed, based on precursor HST imaging and results. That is, these early JWST results suggest that the formation of normal galaxy structure was much earlier than previously thought.

II. A major aspect of this is our discovery that disk galaxies are quite common at $z \sim 3-6$, where they make up $\sim 50\%$ of the galaxy population, which is over 10 times as high as what was previously thought to be the case with HST observations. That is, this epoch is surprisingly full of disk galaxies, which observationally we had not been able to determine before JWST.

III. Distant galaxies at $z > 3$ in the rest-frame optical, despite their appearance in the HST imaging, are not as highly clumpy and asymmetric as once thought. This effect has not been observed before due to the nature of existing deep imaging with the HST which could probe only ultraviolet light at $z > 3$. This shows the great power of JWST to probe rest-frame optical where the underlying mass of galaxies can now be traced and measured.

This study is the first examination of the problem of distant galaxy morphology with JWST, and specifically the formation of galaxy structure at $z > 3$. Our results suggest many directions for immediate future study. We have not included any new JWST galaxies that were not seen with HST, and have not examined the structural properties as a function of stellar mass or other physical properties. All of these will need to be fully examined in the future. Since the conclusion of this Chapter, many results based on newly available JWST are finding an abundance of disk galaxies at higher redshifts. For example, Kartaltepe et al. (2022) finds a similar fraction of disks up to redshift $z \sim 8$, while Nelson et al. (2022) discovers a population of red disk galaxies that are completely undetected in previous HST observations. In Chapter 5 we will touch on this new studies on the larger area dataset covered by the CEERS JWST survey.

The present study, however, shows the importance of JWST for understanding the structural evolution of galaxies, which is now open for detailed investigation.

Chapter 5

The JWST Hubble Sequence: The Rest-Frame Optical Evolution of Galaxy Structure at $1.5 < z < 8$

The content of this chapter has been submitted for publication in The Astrophysical Journal. I acknowledge the help from Christopher Conselice, Elizabeta Sazonova, Geferson Lucatelli, Joseph Caruara, Clar-Brid Tohill for the visual classification effort reported here.

5.1 Abstract

We present results on the morphological and structural evolution of a total of 4265 galaxies observed with JWST at $1.5 < z < 8$ in the JWST CEERS observations that overlap with the CANDELS EGS field, the biggest visually classified sample observed with JWST yet, ~ 20 times larger than previous studies. All sources were classified by six individual classifiers using a simple classification scheme

aiming at producing disk/spheroid/peculiar classifications, whereby we determine how the relative number of these morphologies evolves since the Universe's first billion years. Additionally, we explore structural and quantitative morphology measurements using MORFOMETRYKA, and show that galaxies at $z > 3.0$ are not dominated by irregular and peculiar structures as previously thought. We discuss the dominance of morphologically selected disk galaxies up to $z = 8$, a far higher redshift than previously thought possible. We compare our results to theory to show that the fraction of types we find is predicted by cosmological simulations, and that the Hubble Sequence was already in place as early as one billion years after the Big Bang. Additionally, we make our visual classifications public for the community.

5.2 Introduction

Since the discovery of galaxies, a principal aim of their study has been to characterize their structure and morphology. The very fact that galaxies appear to be extended, as opposed to point sources, already provides an elusive clue to their nature being different from that of the stars. In fact, it can be said that it was the extended nature of these objects that instigated the debate about whether they were external to our own galaxy, a problem solved through obtaining distances to these systems (Hubble, 1926). Even before then, however, the fact that the structure of galaxies holds important information had been known since at least the time of Lord Rosse and his discovery of spiral structure in nearby massive galaxies such as M51 (Rosse, 1850).

Since that time, galaxy structure, morphology, and how these properties evolve with time has remained a key aspect to understanding galaxy evolution (e.g., Delgado-Serrano et al., 2010; Lotz et al., 2004; Mortlock et al., 2013; Conselice et al., 2014; Ferreira et al., 2022b). The resolved structure of distant galaxies, in particular with the advent of the Hubble Space Telescope, clearly revealed that faint distant galaxies were more peculiar and irregular, and few fit into the Hubble sequence (e.g. Driver et al., 1995). Later, once redshifts became available, it became clear that galaxy structures evolve strongly, but systematically, with redshift, such that peculiar galaxies which dominate the population at $z > 2.5$ (e.g., Conselice et al., 2008; Conselice et al., 2014). However, because of the limited red wavelengths of Hubble, we still have not yet been able to trace the rest-frame optical light of galaxies back to within the first few Gyr of the Big Bang. The F160W band on the Hubble Space Telescope (HST) can only probe rest-frame visible light up to $z \sim 2.8$, but JWST permits us to obtain the same type of data out to $z \sim 9$ with F444W. Moreover, JWST's superior resolution and longer wavelength filter set allows galaxy structure to be better measured than with the lower resolution of HST.

Observations of galaxy structure and morphology at $z > 3$ do show that in the rest-frame UV galaxies are peculiar and irregular (e.g., Conselice, 2009). Moreover, galaxies are often very clumpy at these redshifts, as seen with deep WFC3 data (e.g., Oesch et al. 2010). Furthermore, observations of pairs of galaxies show that the merger rate and fraction of galaxies up to $z \sim 6$ is very high, and therefore that galaxy structure should likewise be affected significantly (e.g., Duncan et al., 2019). At the same time, we know that the Hubble sequence is established at $z < 1$ (e.g., Mortlock et al., 2013). However, whether the Hubble sequence already existed in the earlier Universe remains an open question. While earlier HST-based studies found that the dominating majority of galaxies at $z > 2$ are peculiar, recent JWST-based studies find a high number of regular disk galaxies at high redshift (e.g., Ferreira et al., 2022b; Nelson et al., 2022; Jacobs et al., 2022; Robertson et al., 2022), consistent with an even earlier emergence of the Hubble sequence.

Quantitative measures of galaxy structure and morphology also present stringent constraints for numerical simulations to reproduce. In recent years, full hydrodynamic simulations (Schaye et al., 2015; Nelson et al., 2019; Lovell et al., 2021; Marshall et al., 2022b) enable resolved morphologies to be predicted in a self-consistent manner, and recent novel simulation approaches allow these to be tested out to the highest redshifts (Roper et al., 2022). There are a number of difficulties when comparing morphologies between simulations and observations, however simple measures of the abundance of e.g. disk and elliptical galaxies can provide hints as to the underlying mechanisms leading to morphological evolution.

In this Chapter we explore the morphological properties of 4265 galaxies observed with JWST through visual galaxy classifications and quantitative morphology, from $z = 1.5$ to 8. We demonstrate that these early galaxies have predominantly disk morphologies, and that the Hubble sequence seems to be already established as early as $z \sim 8$.

The Chapter is organized as follows. In § 5.3 we describe the data products used, our reduction pipeline, the visual classification scheme adopted as well as our methods of quantitative morphology. § 5.4 describes the results from our classification effort and the quantitative morphology measurements of this sample. We follow with a discussion on the implications of these results in § 5.5, and finish with a summary of our main results in § 5.6.

5.3 Data and Methods

We use the public NIRCam JWST observations from the Cosmic Evolution Early Release Science Survey (CEERS; PI: Finkelstein, ID=1345, Finkelstein et al. in prep), that overlap with the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) on the Extended Growth Strip field (EGS). These data are reduced independently using a custom set-up of the JWST pipeline version 1.6.2 using the on-flight calibration files available through the CDRS 0942, an extensive description is given in § 5.3.1.

We select 4265 sources with $z > 1.5$ from the CANDELS catalogs which overlap with the area covered by CEERS. We take advantage of the robust photometric redshifts, star formation rates and stellar masses already derived for CANDELS in previous works (Duncan et al., 2014, 2019) to conduct this analysis. Neither morphological information is used for the selection of sources, and nor magnitude cuts are employed as we want to make sure that we include sources that might be faint in HST, but bright in JWST observations. This is also the case for morphology; we are also interested in sources that can show dramatic changes in morphology between the two instruments.

We employ two different approaches to these data: first we perform visual classifications for all sources, which is described in detail in §5.3.3. Second, we perform quantitative morphology through Morfometryka (Ferrari et al., 2015), where we measure non-parametric morphology estimates such as CAS, G-M20, H, Spirality, Sizes, as well as light profile fitting, which is described in detail §4.3.2.

5.3.1 Data Reduction

We reprocess all of the uncalibrated lower-level JWS data products for this field following our modified version of the JWST official pipeline. This is similar to the process used in Adams et al. (2022); Ferreira et al. (2022b) but with minor updates and improvements and can be summarised as follows: (1) We use version 1.6.2 of the pipeline with the Calibration Reference Data System (CRDS) version 0942 which was the most up-to-date version at the time of writing. Use of CRDS 0942 is essential for zero point issues we discuss in (Adams et al., 2022). (2) We apply the 1/f noise correction derived by Chris Willott on the resulting level 2 data of the JWST pipeline.[1] (3) We extract the sky subtraction step from stage 3 of the pipeline and run it independently on each NIRCam frame, allowing for quicker assessment of the background subtraction performance and fine-tuning. (4) We align calibrated imaging for each individual exposure to GAIA using `tweakreg`, part of the DrizzlePac python package.[2] (5) We pixel-match the final mosaics with the use of `astropy reproject`.[3] The final resolution of the drizzled images is 0.03 arcseconds/pixel. There is rapid development in the above procedure, and so we anticipate future studies to continue to make refinements to the JWST pipeline. Each one of the four June CEERS observations was processed into individual mosaics.

5.3.2 Photometric Redshifts and Stellar Masses

The photometric redshifts that we use in this Chapter originate from the redshifts calculated in Duncan et al. (2019) for EGS. These are based on the original

---

[1] `https://github.com/chriswillott/jwst`

[2] `https://github.com/spacetelescope/drizzlepac`

[3] `https://reproject.readthedocs.io/en/stable/`

CANDELS+GOODS WFC3/ACS imaging and data, Spitzer/IRAC S-CANDELS (Ashby et al., 2015) and ground based observations with CFHT (Stefanon et al., 2017). The overall method for this is described in detail in Duncan et al. (2019).

5.3.3 Visual Classification

As a way to define the morphologies of the galaxies in our sample of 4265 sources, we construct a simple classification scheme that yields a large amount of information with a small number of classification questions, as opposed to having a very detailed sub-classification scheme of structure sub-components. The classification scheme is summarized in the fluxogram in Fig. 5.1. At high redshift, fine structural details are often difficult to recover and in general are ambiguous, hence these questions capture the overall appearance of the source.

Our sample is the biggest visually classification sample observed with JWST yet, ~ 20 times larger than what is reported in previous JWST morphology results (Ferreira et al., 2022b; Nelson et al., 2022; Jacobs et al., 2022). A brief description of each possible resulting class is given below:

- Class 0: Unclassifiable: Galaxies not clearly visible, too faint to classify, and images with artifacts.
- Class 1: Point Sources: Sources that are smaller in angular size than the $\mathrm{PSF}_{\mathrm{FWHM}}$ or that present clear wings/spike patterns consistent with point-like objects but no extended component.
- Class 2: Disks: This category includes galaxies that exhibit a resolved disk in the form of an outer area of lower surface brightness with a regularly increasing brightness towards the center of the galaxy. This classification

does not depend on there being a spiral pattern in the system, although one can be present in this classification.

- Class 3: Spheroid: These galaxies are resolved, symmetrically and centrally concentrated, with a smooth profile and are round/elliptical in shape.
- Class 4: Peculiar. This class is for well-resolved galaxies with a morphology which is dominated by a disturbance or peculiarity, where the disturbance dominates any smooth components.

Each galaxy is further classified as smooth or structured, where structured galaxies have features standing out from the smooth stellar envelope, such as star-formation clumps, tidal features, and merger signatures. Galaxies with distinct disk and bulge components were also classified as structured. Finally, the classifiers were able to provide additional notes on each source to aid in future analysis.

Ultimately, after all classifications are aggregated we determine the final class of each object as the the one receiving the majority of the votes, as discussed in detail in Sec. 5.3.3. In the cases where classifiers disagreed, we included an `ambiguous` class.

Based on this classification scheme (Fig. 5.1), six volunteers classified all the 4265 sources. This effort produced a robust catalog where every galaxy has all six classifications combined in classification fractions, one for each individual question present in the scheme (Fig. 5.1).

To perform each classification, the classifiers were given access to a web application build with `flask`, `jinja` and `bootstrap` specifically tailored for this task. The volunteers were presented with the rest-frame image in the filter that corresponds to the source redshift (minimizing for $\lambda_{rest}/(1+z)$), a RGB image

(F277W+F356W+F444W) generated with TRILOGY (Coe et al., 2015), the PSF image of the respective filter together with the size of the PSF_{FWHM} and a questionnaire that reproduces Fig. 5.1. Results are stored in a `MySQL` database, that is then reduced and aggregated with `pandas` (Reback et al., 2022).

Each individual classifier results are combined in a single table using the following criteria.

First, we define how many votes there are for each source. That is, the votes that are not considered to be `unclassifiable` and `point source`. Then, if at least 50% (i.e., 3 or more votes are exclusively in these categories) of all votes are assigned to any of these individual classes, we consider the source to be unclassifiable or point-like, by the labels `n/a` and `ps`, respectively. For the point-like sources, we compare its size to the PSF_{FWHM}. If it is larger than the PSF_{FWHM}, we change its classification to `spheroid`. Same is done the other way around, sources smaller than the PSF_{FWHM} are changed programatically to `ps`.

Second, for all the rest of the sources that have more than 50% of the votes in disk, spheroid or peculiar categories, we average each individual classification decision in a class fraction. Hence, this class fraction is only based on the number of good votes (i.e., votes that are for classifiable and extended galaxies).

Third, to all galaxies that have a clear majority as $frac > 0.5$, we assign the given class as the final class. For all the remaining cases that the classifiers disagree on (e.g., 2 votes in each category), we define those sources to have an `ambiguous` class.

Finally, we include a structure index, the `smooth_fraction`, that is independent of the general appearance, designed as a large umbrella to capture sources with rich structures, sub-components, merging features.

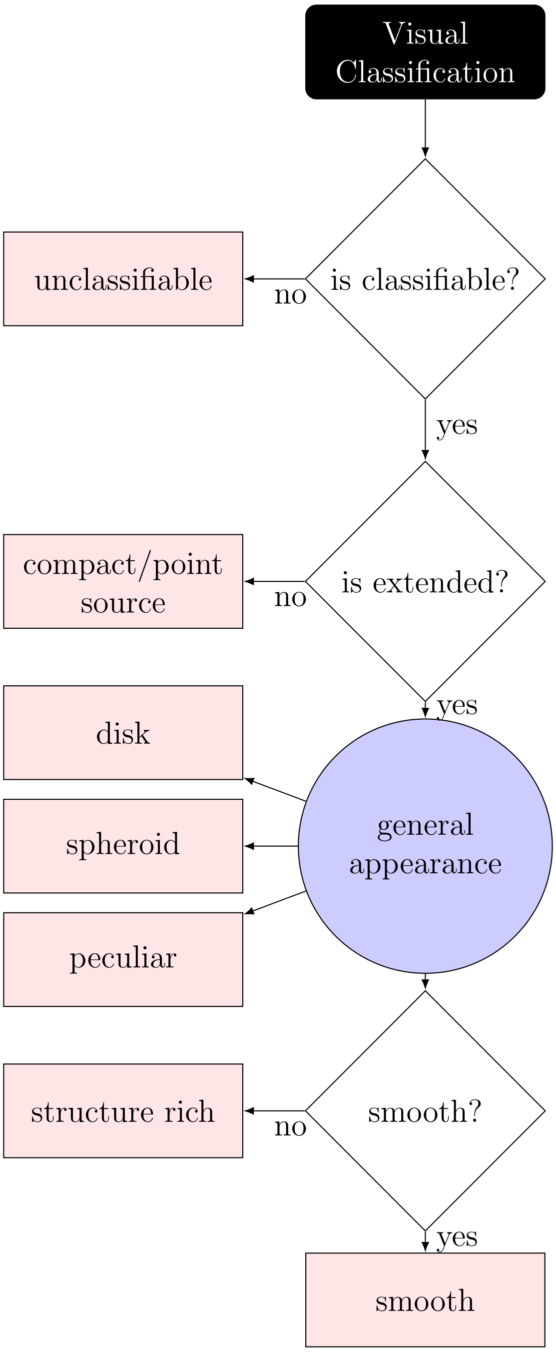


Figure 5.1: Fluxogram of our visual classification process. The classification is based on 4 basic questions that can produce a simple disk/spheroid/peculiar/compact classification and additional flags regarding whether the source is smooth or structurally rich.

This framework enables sources that might be ambiguous between two classifiers to have a more robust classification. The class fractions can also be used to control the purity of the samples, as higher agreement will represent a less contaminated dataset. We proceed, however, with the final classifications assigned by the majority of the classifiers.

We run Morfometryka for all filters available but only report results for the band that closest matches the rest-frame optical of the source ($\lambda = 0.5\mu m$-$0.7\mu m$).

5.3.4 BlueTides high-z Mocks

For a comparison to these observations, we also consider mock JWST images from the BlueTides Mock Image Catalogue (Marshall et al., 2022b,a). This is a catalogue of mock image stamps of $\sim 100,000$ galaxies from $z = 7$ to 12, with the particle distributions and SEDs of each galaxy taken from the BlueTides hydrodynamical simulation (Feng et al., 2015). The images are created with the NIRCam transmission curves and convolved with JWST model PSFs from WebbPSF (Perrin et al., 2015), to produce a realistic mock image for each galaxy. The BlueTides mock images have a pixel scale of 00155 for the NIRCam short-wavelength filters, half of our observed pixel scale, and 00315 for the long-wavelength filters, equivalent to our observed pixel scale. Here we use the F444W filters, the same used for this redshift restframe.

It is important to note that the images of the BlueTides galaxies are created in the 'face-on' direction, which is defined by the angular momentum of particles in the galaxy (see Marshall et al., 2020). However, studies suggest that the angular momentum of early galaxies does not correlate with their morphological structure (e.g. Park et al., 2022)—indeed, visual inspection of the BlueTides images shows that this direction does not necessarily correspond to the visual morphological 'face-on' direction. Thus, we do not expect this feature of the simulation to highly affect our comparison, although some biases could be present.

We select all galaxies that overlap with the redshift range probed here from $z = 7$ to 8. For each source stamp we add Gaussian noise to match the depth of the CEERS observations levels. Then, we run MORFOMETRYKA for all stamps together with the PSF generated with WEBBPSF.

5.4 Results

We report the morphology and structure evolution of the sample of 4265 galaxies (§ 5.3) based on visual classifications (§ 5.3.3) and in quantitative morphology measurements (§ 4.3.2). The aggregated classifications catalog based on 6 independent classifiers contains 1696 disks ($\sim$ 40%), 561 spheroids ($\sim$ 13%), 1112 peculiars ($\sim$ 26%), 434 ambiguous sources ($\sim$ 10%), 66 point sources ($\sim$ 1%) and 396 unclassifiable sources ($\sim$ 9%). Examples of each of these types are shown in Figure 5.2 in bins of increasing redshift. These visual classifications are the basis for the discussion in this section. In Section § 5.4.1 we detail the three base classes and the caveats from the visual classifications. We follow with a description of the quantitative morphologies of these sources and how they relate to the visual classifications in Section § 5.4.2. Finally, we compare these classifications with predictions from cosmological simulations in Section § 5.4.3.

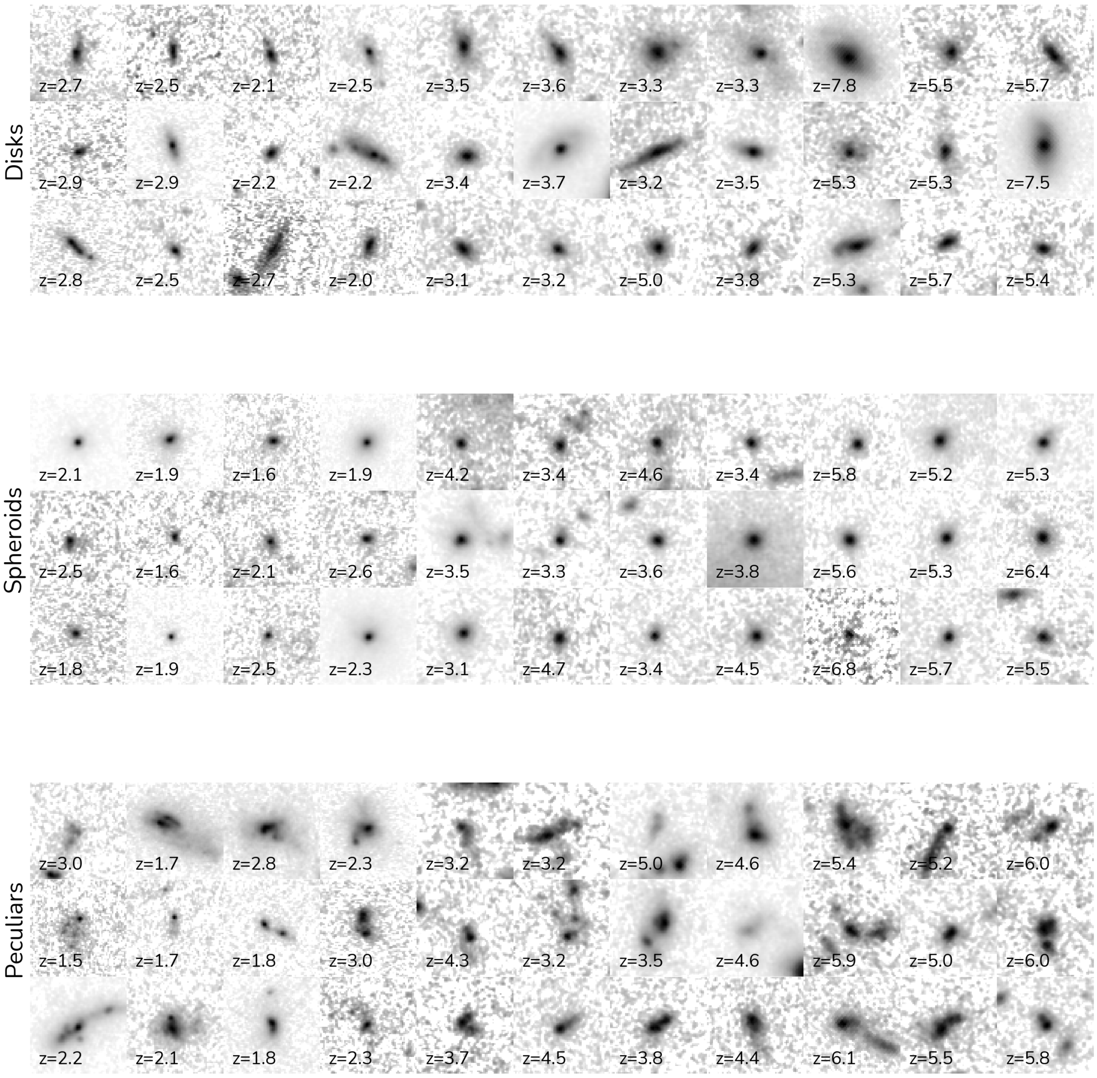


Figure 5.2: Rest-frame optical images for sources in our sample. The three panels show the three main classes `disks`, `spheroids` and `peculiars`, respectively. Galaxies are ordered horizontally by redshift, lowest redshifts in the left, highest redshifts to the right. Stamps are shown in square root normalization. Redshifts are from Duncan et al. (2019) based on the CANDELS fields.

5.4.1 Disks, Spheroids and Peculiars

Figure 5.2 displays examples randomly drawn from the catalog for the three main morphological classes: disks, spheroids and peculiars, respectively.

The visual distinction between these classes is clear, with the disks often showing two structural components in the form of a concentrated bulge and a disky envelope, while the spheroids are mostly single-profile, centrally concentrated sources, with some exhibiting PSF-like structure due to the central concentration or AGN emission. However, we note that for most cases, telling apart two types of light concentrations by eye is a difficult task, as sources at high redshifts do not show other clear, disk-like features such as spiral arms, bars and rings, and overall display lower concentrations (Buitrago et al., 2008; Buitrago et al., 2013). For a better distinction between face-on disks and spheroids, a quantitative approach such as Sérsic fitting might be used alongside visual classification.

Figure 5.3 shows three indicators for visual distinctions between the overall sample of `disks` and `spheroids`. `spheroids` are more compact with lower effective radius, higher axis ratios and lower information entropy indicating lack of structure. These distributions follow what is found for high redshift spheroids in previous studies, in that they are round and smaller than disks (Buitrago et al., 2013). However, the axis ratios found here are on the high end, with a lack of spheroids with intermediary axis ratios $0.4 < b/a \sim 0.7$. Ultimately, some biases might be present, such as the elongation/axis ratio causing some contamination, thus we possibly miss some elongated spheroids, but this is expected due to each classifier's subjective perspective on what defines these. We advise the user of the catalog to leverage the class fractions to control purity by only selecting sources with strong agreement.

The peculiars on the other hand vary wildly, from mild disturbances, to clear signs of galaxy merging, often with a companion nearby. Additionally, some high redshift disks and spheroids might end up being classified as peculiars due to them showing more asymmetric/disturbed morphologies than their low redshift counterparts. However, as discussed in § 5.4.2, for most peculiars, the quantitative morphology is consistent with disturbed morphologies.

5.4.2 Quantitative Morphology Evolution

Cross-examining the morphologies defined by eye using quantitative methods is essential for understanding how their appearance changes across cosmic time. We explore the visual morphologies with several quantitative morphology indicators, both non-parametric and parametric (§ 4.3.2).

Figure 5.4 shows the concentration (C) and asymmetry (A) plane based on MORFOMETRYKA measurements for 4 redshift bins. The mean values alongside the distribution's 15% and 85% percentiles for disks, spheroids and peculiars are plotted as blue squares, red circles and pink diamonds, respectively. Each class has its distributions positioned within the expected regions for high redshift galaxies, with peculiars occupying the top of the diagram, the disks the central region and the spheroids around the lower right, including a high overlap. The positions of each of the three classes remain fairly stable over all redshifts, but the spheroids have higher asymmetries and lower concentrations overall at higher redshift, with larger overlaps with the disks. Also displayed with solid lines is the merger criterion based on asymmetry, as

$$A > 0.35, \tag{5.1}$$

and the diagonal late-type vs. intermediate types boundary based on Bershady

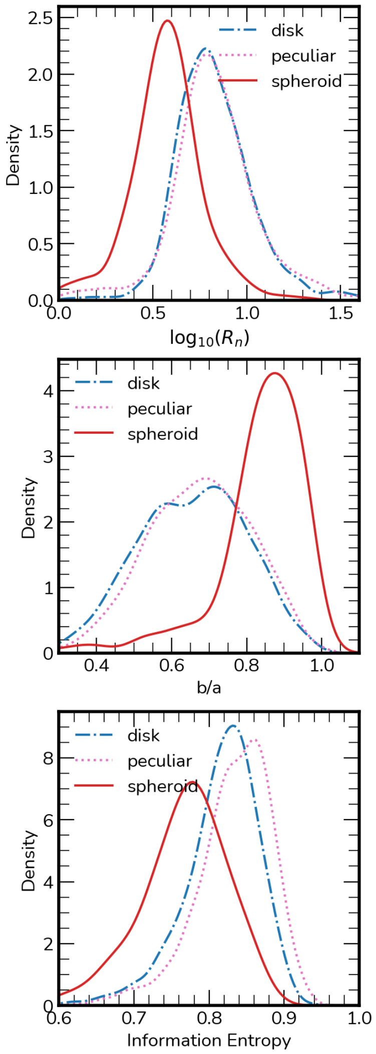


Figure 5.3: Effective radius (R_n) (top), axis ratio (b/a) (middle) and information entropy (H) (bottom). We show key measurements that clue to differences most used by the classifiers for the spheroid × other classes. Spheroids are defined by their lack of structure, low elongation and small sizes in general.

et al. (2000). We also explore the G-M_{20} plane (Lotz et al., 2004, 2008) but we do not find any clear separation between the types, apart from the distinction between sources with close companions or isolated, similar to what is reported in Rose et al. (2022).

For the highest redshift bin ($4 < z < 8$), we also show the contours for the distribution of the measurements for 50,000 galaxy mocks from the BlueTides simulations (Marshall et al., 2022b). The measurements from the real observations are in well agreement with the measurements for BlueTides mocks. We note that the effective spatial resolution of the BlueTides simulation is $1.5/h$ ckpc, which corresponds to 0.269 pkpc, or 005 at $z = 7$. This resolution may have an effect on the resulting galaxy morphologies, particularly the inner regions and thus their concentration, but no clear effect is seen here, different to what is found for high-z IllustrisTNG mocks (Whitney et al., 2021).

The spirality index, the standard deviation of the galaxy polar image (r, ϕ) gradient map, designed to measure the amount of non-radial structures in the galaxy, has proven to be very effective in discriminating different classes in this sample. If the galaxy is smooth, its polar image will consist of a single horizontal strip, which will imply a low value for σ_ψ. On the other hand, although we cannot resolve spiral arms in most cases, if the galaxy contains peripheral structures or companions, the polar image will be irregular with a corresponding high σ_ψ. Please see the Morfometryka (Ferrari et al., 2015) paper for a full description of the σ_ψ calculation.

In Figure 5.5 we show that combining σ_ψ with the asymmetry (A) warrants a reasonable quantitative separation of the overall classes, as the center of each class distribution is well separated in this plane, unlike in the C-A or G-M20 diagram. As an example as to what is captured by the σ_ψ measurement, we show three

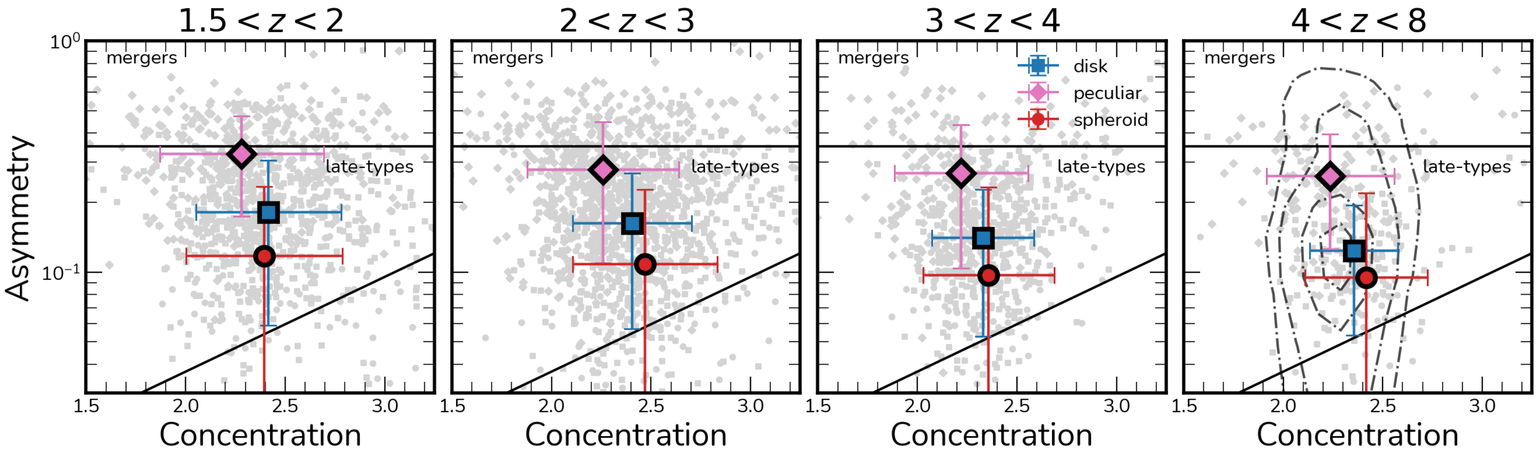


Figure 5.4: Concentration vs. Asymmetry diagrams. The evolution of the Concentration (C) and Asymmetry (A) in 4 different redshift bins, one for each panel. Peculiars are shown as pink diamonds, disks as blue squares and spheroids as red circles. The highest redshift panel shows contour lines based on the distribution of C-A measurements for BlueTides (Marshall et al., 2022b) galaxy mocks at $z \geq 7$ as discussed in § 5.3.4. The solid lines denote the merger selection threshold on the top, and the late type / early type separation on the lower diagonal line. Peculiars display high asymmetries when compared to other types. Disks display late type-like morphology, while spheroids are regular at lower redshifts, located on the bottom right of the plots, but move towards the center with increasing asymmetry and decreasing concentration with redshift. Galaxies overall get less concentrated and more asymmetric with increasing redshift. However, at high redshifts, sources display higher concentrations and asymmetries when compared to simulations. In comparison to what is expected from the local universe, classes are clustered more closely, while at $z \sim 0$ classes are more easily to distinguish, with wider concentrations and asymmetry ranges.

galaxies in Figure 5.6, one for each class, with their respective polar coordinate image and the gradient lines that are used to compute σ_ψ. The distinction between spheroids and disks is subtle, however it is very powerful when large non radial structure is present in the outskirts of the source.

Here in both Fig 5.4 and Fig 5.5 we see that we can distinguish the centre of the cluster for each class. However, this is not true in the non-parametric results reported in Chapter 4, where the spheroid population is almost indistinguishable from the disks, and the disks have higher overlap with the peculiars. We attribute this difference to a high contamination of point sources as spheroids in Chapter 4, which have intrinsically lower concentrations and can display higher asymmetries due to the JWST PSF, as opposed to extended sources. This is further corroborated by the distinct class fractions for spheroids reported later here in Fig 5.9 in comparison to Fig 4.4. This can also be seen in the Sersic profiles reported between the two Chapters. In the case of higher degree of mixture between peculiars and disks, we recognize that the cluster environment can affect these non-parametric measurements negatively due to both magnification and stray light from more neighboring sources. Galaxies in the sample of Chapter 4 are slightly distorted due to strong lensing (hence higher Asymmetries), and the measurement of the asymmetry itself can also suffer from strong gradients that might not be captured by the asymmetry of the background estimation. We do not do a systematic analysis of these effects here, as both this Chapter and Chapter 4 are focused on visual classifications. Nevertheless, we limit our comparisons between physical properties and the morphology of these galaxies only using the visual classifications.

Finally we explore the evolution of the Sérsic profiles (Sérsic, 1963) through the redshift evolution of the Sérsic index for galaxies with $M_* > 10^9 M_\odot$. In Figure 5.7 we report mean values together with the 15% and 85% percentile limits as error

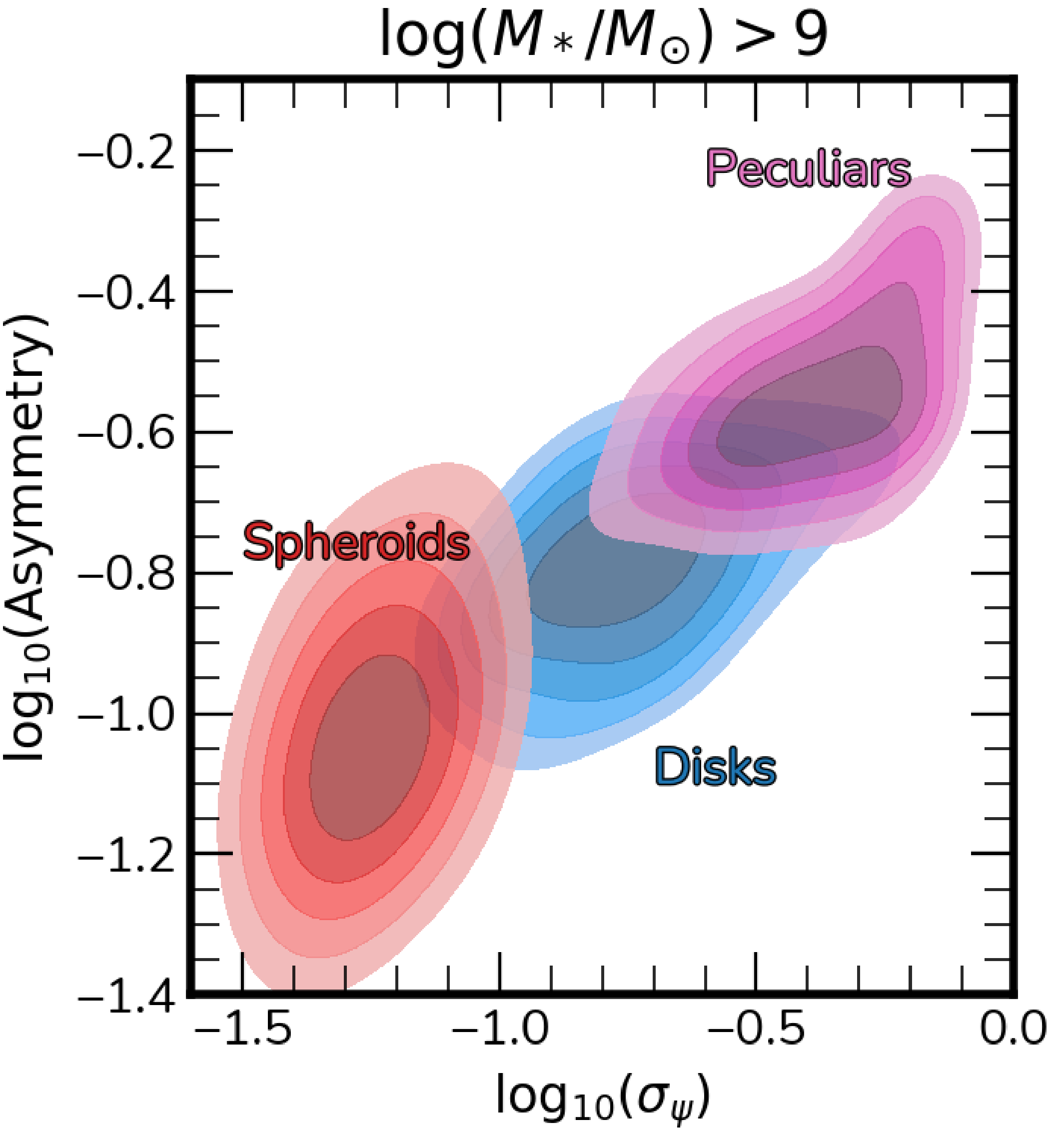


Figure 5.5: log Asymmetry (A) vs. log Spirality (σ_ψ) 2D distributions for each class. Kernel density estimation distributions for the top 50% of each morphological class in 5 bins of 10% fractions of the distribution. The asymmetry and σ_ψ correlates strongly, but are independent measurements as each classification distribution has a different slope. Spheroids show high diversity in A and low diversity in σ_ψ, while the contrary is true for disks. These two measurements form a parameter space capable of separating the classes in this sample relatively well compared to C-A.

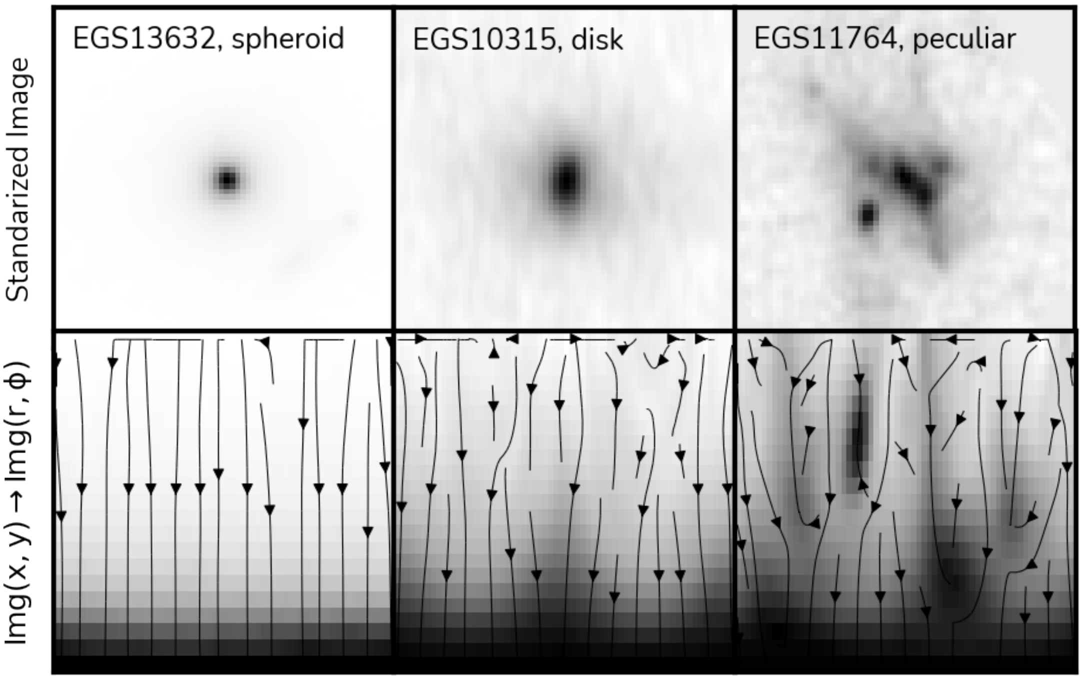


Figure 5.6: Three examples of the spirality σ_ψ measurement. Top row shows the standarized images of the sources ($q = 1$, $PA = 0$) while the bottom row displays the polar coordinates transformation (r, ϕ) of the above image. Black lines display the gradient field of the image. The σ_ψ measurement is based on the standard deviation of these field lines.

bars to represent the distributions at each redshift bin for each class. Disks and Peculiars exhibit similar Sérsic profiles, with $n \approx 1.0$ at $z \sim 6$ to $n \approx 1.3$ at $z \sim 1.5$. The spheroids, on the other hand, show higher Sérsic indicies at all redshifts, with $n \approx 1.8$ at $z \sim 6$ to $n \approx 2.5$ at $z \sim 1.5$. The distinction between the classes is clearer than what was reported in Ferreira et al. (2022b) as the CANDELS overlap allows us to quickly select high mass galaxies only. The slopes for each class are also different, with the spheroids increasing in Sérsic index more rapidly. The similarities among disks and peculiars suggest that the majority of these disturbed and merging systems are still disk dominated.

5.4.3 Predictions from Simulations

We now look at numerical simulation results for morphological evolution over a similar redshift range. Simulations that resolve galaxies self-consistently typically model mass elements either on a grid or as particles. Particle–based decomposition methods (e.g. Abadi et al., 2003; Crain et al., 2010; Thob et al., 2019; Irodotou & Thomas, 2021; Zana et al., 2022) have been extensively used in order to split galaxies into different morphological classes and facilitate a comparison between observed and simulated galactic properties (Tissera et al., 2012; Pillepich et al., 2015; Irodotou et al., 2019; Monachesi et al., 2019; Trayford et al., 2019; Rodriguez-Gomez et al., 2022). However, the true morphology of a system may not always be accurately captured, as particle–based methods can be sensitive to small perturbations in the distribution of particles, which become progressively more significant at lower stellar masses as these galaxies are resolved with fewer particles.

In this work, to ensure that both galaxies and their components are sufficiently

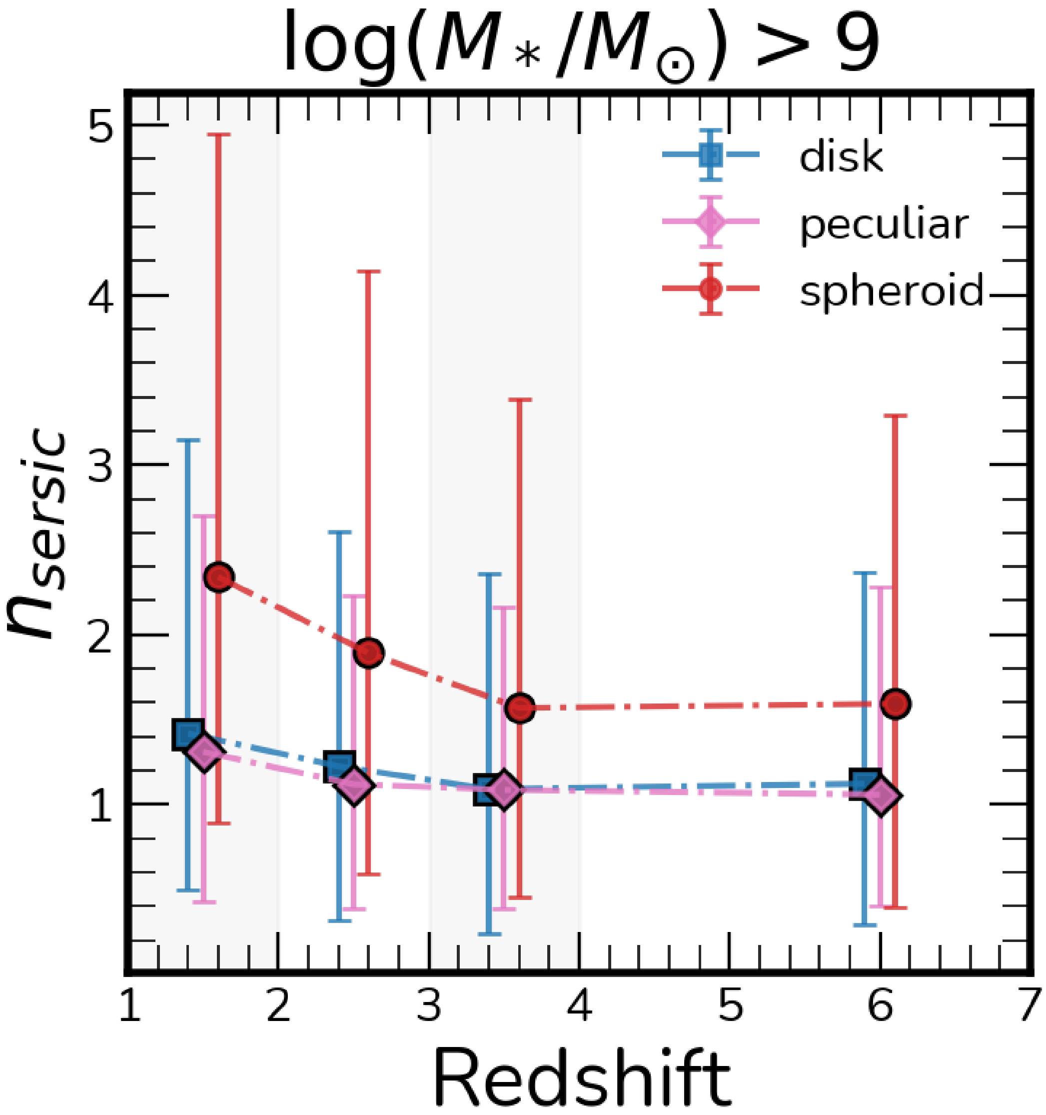


Figure 5.7: Sérsic index redshift evolution for each morphology class. Displayed as blue squares, red circles and pink diamonds are the means for disks, spheroids and peculiars, respectively. Error bars define the 15% and 85% percentile of the distributions.

resolved, and thus a particle–based decomposition is applicable, we use central and satellite galaxies from the EAGLE (Schaye et al., 2015; Crain et al., 2015, for $1.5 < z < 4$) and FLARES (Lovell et al., 2021, for $5 < z < 8$) simulations with stellar masses $\log(M_*/M_\odot) > 10$ (i.e. even for a galaxy with $B/T \sim 0.2$ the bulge is resolved with more than a thousand particles with mass of a few $\times\ 10^6$ each). We use the method developed in Irodotou & Thomas (2021) to decompose galaxies by firstly creating a Mollweide projection of the angular momentum map of each galaxy's stellar particles. Then, stellar particles are assigned to a disc or spheroid component based on their angular separation from the densest grid cell. This allows us to calculate bulge-to-total mass ratios (B/T) and use these to split galaxies into two morphological classes: (i) spheroids with $B/T > 0.75$, (ii) spirals with $B/T < 0.75$. The aforementioned B/T limits were calibrated at $z \sim 0$ in order for the EAGLE galaxies to match the morphological classes in the Conselice (2006) sample. In Figure 5.8 we show the comparison of these fractions with the relative fraction of disks and spheroids from the visual classifications for high mass galaxies with $M_* \geq 10^{10} M_\odot$, ignoring the peculiars, as we do not have a direct way to classify peculiars in the simulation dataset. The trends between simulations and visual classifications agree for $z > 3$, with the exception of a single anomalous visual classification redshift bin showing similar fractions of disks and spheroids in $5 < z < 6$. Fractions for $z < 3$ overall disagree, with an excess of 10% of disks in the visual classifications. It is worth noting, however, in this redshift range (i.e. $1.5 < z < 3$), (Lagos et al., 2018) showed that the fraction of dry major mergers in the EAGLE volume increases. Since this type of mergers can efficiently reduce the angular momentum of the remnant, this will translate in a negative correlation between our B/T values and redshift, as also seen in e.g. Fig. 4 of (Clauwens et al., 2018) for $M_* \geq 10^{10.5} M_\odot$.

Park et al. (2022) reports similar findings for the Horizon simulations, with the $z \geq 5$ fraction of disks dominating at around 75% by using quantitative morphology indicators as a proxy for morphology.

5.5 Implications

There are several implications from our visual classifications and quantitative morphology measurements. We explore the evolution of the Hubble Sequence in § 5.5.1, the evolution of the contribution of each morphological class to star formation and stellar mass in § 5.5.2, and in § 5.5.3 we briefly discuss the main differences between HST and JWST imaging that could explain some of the discrepancies from previous studies.

5.5.1 Evolution of the Hubble Sequence

One principal goal of looking at galaxy morphology and structure is to establish when and how the Hubble Sequence (Hubble, 1926) emerges in the context of the hierarchical assembly of the universe.

Here we report the redshift evolution of morphological classes that encompass the three main categories of the Hubble Sequence from $1.5 < z < 8$, from when the Universe was only ~ 0.6 Gyrs old up to ~ 4.2 Gyrs. In Figure 5.9 we show this evolution in two mass bins, with the left panel displaying sources with $M_* \leq 10^{9.5} M\odot$ while the right panel shows this evolution for $M_* > 10^{9.5} M\odot$. Spheroids are displayed by red circles, peculiars as pink diamonds, disks as blue squares, and the other three possible categories as `other` (point sources, ambiguous and unclassifiable) in gray crosses. The fraction of disks from Chapter 4 are shown as

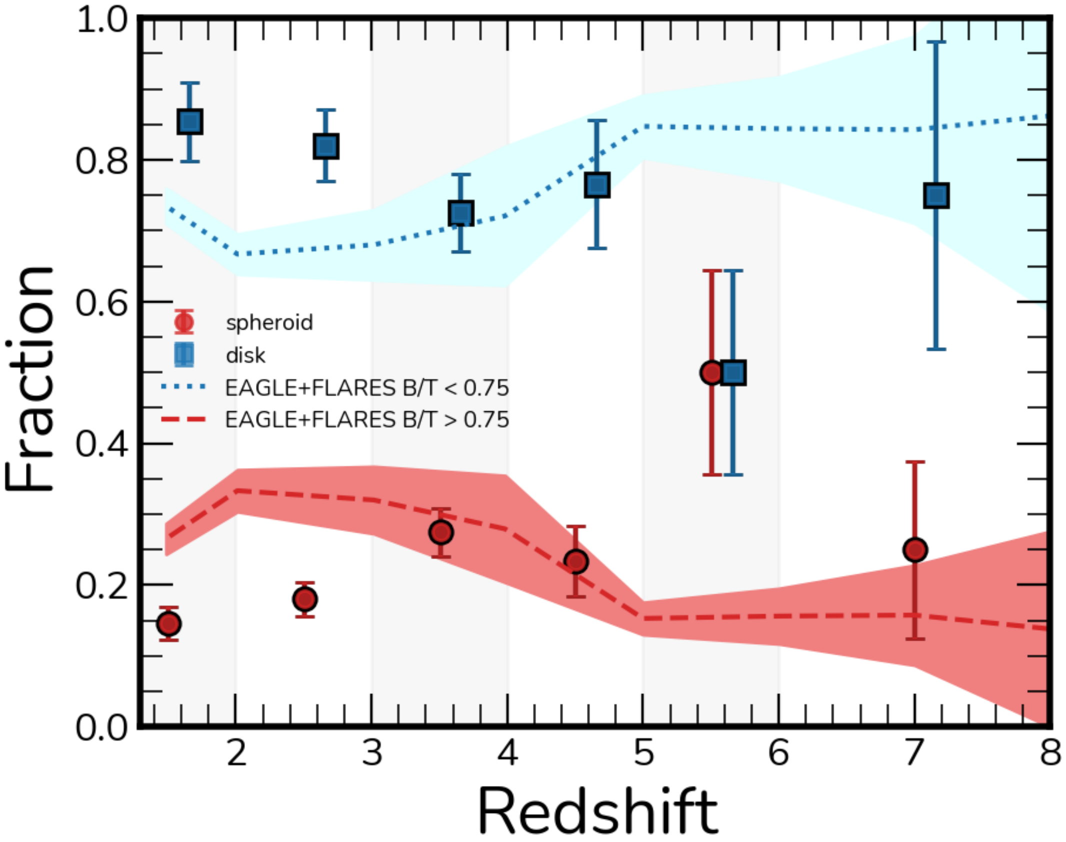


Figure 5.8: Morphology fractions compared to B/T morphological type selection in EAGLE+FLARES for massive galaxies ($M_* \geq 10^{10} M_\odot$). The relative fraction of disks and spheroids are shown as blue squares and red circles, respectively. The blue dotted line shows mean values for galaxies with $B/T < 0.75$ in EAGLE+FLARES. Dashed line shows galaxies with $B/T > 0.75$. Shaded regions represents ±2 sigma for each distribution. The visual classification fractions only account for disks and spheroids to allow a more direct comparison with the two thresholds in B/T in the simulations.

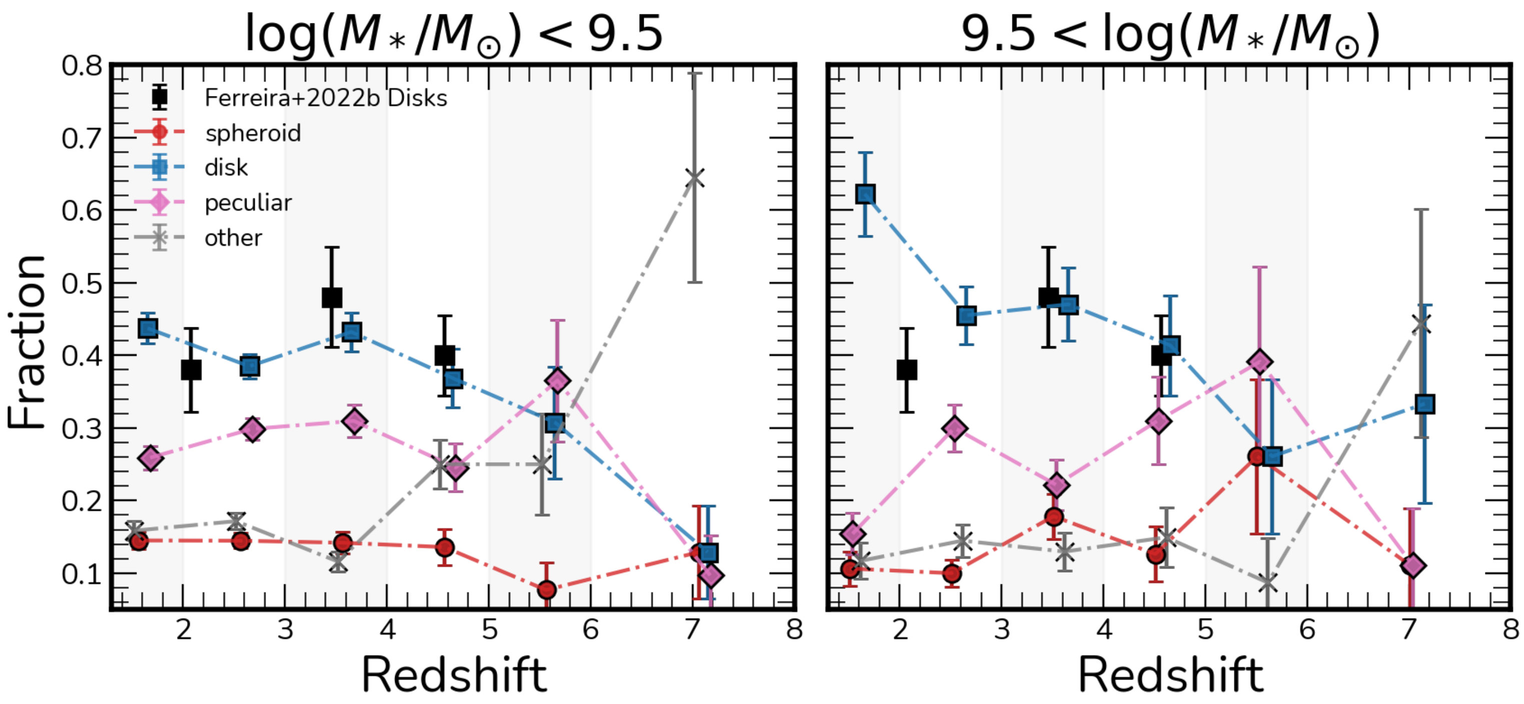


Figure 5.9: Morphology Fraction vs Redshift. The morphology fraction evolution with redshift for the main morphological classes of our classification framework (Fig. 5.1) in two mass bins, $\log(M_*/M_\odot) < 9.5$ (**left**) and $\log(M_*/M_\odot) > 9.5$ (**right**). Disks, spheroids, peculiars and other in blue squares, red circles, pink diamonds, and gray crosses, respectively. This `other` category aggregates the ambiguous, point source and unclassifiable sources. The black squares show the Disk fractions reported in Chapter 4.

black squares for comparison.

For low masses ($M_* \leq 10^{9.5} M\odot$) we do not find any systematic evolution with redshift, with all class fractions remaining fairly constant, with disks at $\sim$ 40%, peculiars at $\sim$ 25% and spheroids at $\sim$ 15%. The only exception is at the highest redshift bin $6 < z < 8$, where all fractions go down to $\sim$ 10% due to the sharp increase of faint and ambiguous sources, artifacts and unclassifiable cases, representing the limits on the depth of our dataset at the low mass range. The disk fractions between $2 < z < 5$ agrees well with what was previously reported in Chapter 4, but now with a 20 fold increase in sample size, as shown by the smaller error bars.

On the high mass cases ($M_* > 10^{9.5} M\odot$) we observe an evolution of the disk fraction from $\sim$ 55% at $z \sim 2$ to $\sim$ 35% at $z \sim 7$ while the fractions of peculiars increases from $\sim$ 20% to $\sim$ 30% in the same redshift range, and spheroids increase from $\sim$ 10% to $\sim$ 15% between $1.5 < z < 6$.

These trends suggest that the classic picture of morphology and structural evolution driven by merging might be only important for massive galaxies, where the low mass universe can be described broadly by a consistent Hubble Sequence in the range $1.5 < z < 6$.

The apparent abundance of disk galaxies in the high redshift universe is also reported by other recent JWST studies (Ferreira et al., 2022b; Nelson et al., 2022; Jacobs et al., 2022; Robertson et al., 2022).

5.5.2 Star Formation Evolution

One important goal of tracking morphologies across cosmic time is determining if different morphologies contribute to the star formation budget of the universe differently. In the previous section we explored the fractional evolution with redshifts, concluding that disk galaxies dominate the overall fraction of morphologies at $1.5 < z < 6$. In Figure 5.11 we show the class fractions in bins of specific star formation rates (sSFR) divided in four redshift bins. The overall fraction of galaxy types for each redshift panel can be seen in Fig. 5.9, while each bin in Fig. 5.11 shows these fractions for a given sSFR bin. Once more, the disks dominate the overall contribution in sSFR, with spheroids showing similar fraction to the overall fraction of spheroids in the sample. However, we see that at high sSFR bins, the contribution of peculiar galaxies increases, to become as important as disk galaxies despite their lower overall fractions in the sample. The same trend is shown for all redshift bins, as the fraction of peculiars increase with increasing sSFR. For the highest redshift bin ($4 < z < 8$), peculiars display roughly the same contribution as disks. Spheroid fractions are slightly higher at lower sSFR. This suggests that peculiar galaxies are important sites of star formation at all times in the Universe, and especially at higher redshifts.

In Figure 5.10 we show the contribution of each morphological class to the total stellar mass in each redshift bin. This shows that most of the mass in the early universe was located in peculiar galaxies, while a clear trend with redshift is evident for disk galaxies, such that for $z < 3$ most of the mass of the sample is distributed among disk galaxies. More massive overall individually, the spheroid galaxies hold just a small fraction of the total stellar mass in this sample ($f_m \sim 10\%$) as it is greatly outnumbered by the amount of mass in galaxies with disk morphologies. This is in contrast to what is found in the Hubble Deep Field (Conselice et al.,

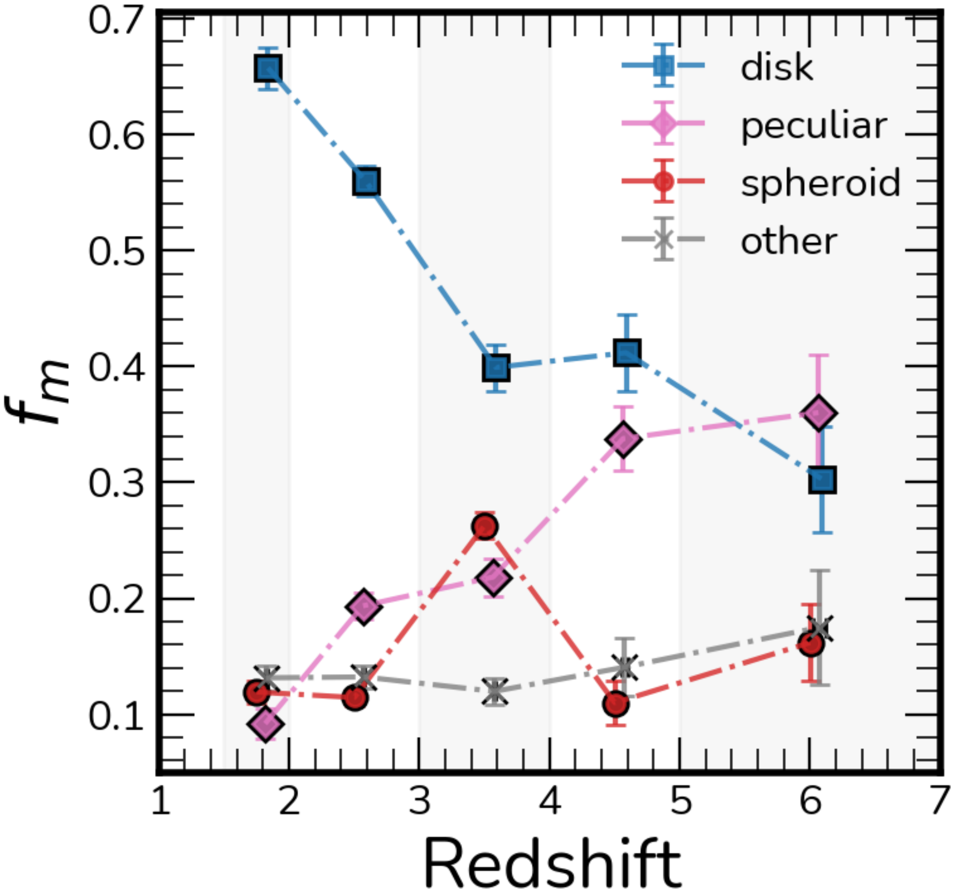


Figure 5.10: Fraction of stellar mass (f_m) in each morphology subsample vs Redshift. We show the contribution to total mass of each redshift bin from each morphological class. This represents how much each class contributes to the total mass in that redshift. Disks, spheroids, peculiars and other in blue squares, red circles, pink diamonds, and gray crosses, respectively. The `other` category aggregates the ambiguous, point source and unclassifiable sources. The contribution from disks shows a trade off with respect to peculiars and spheroids at higher redshifts.

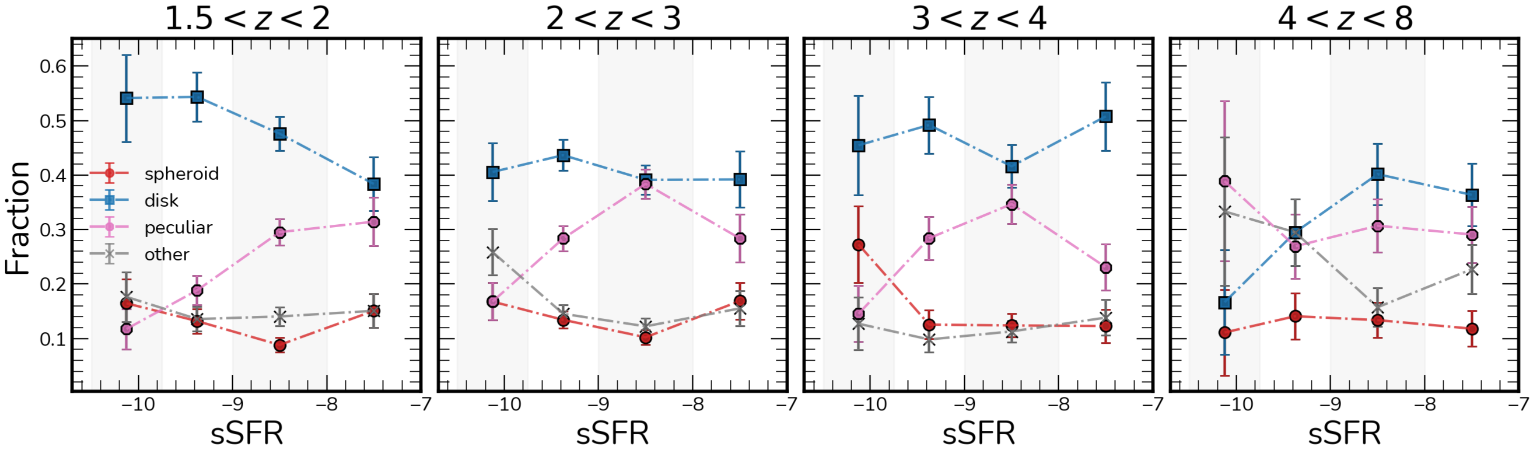


Figure 5.11: Morphology fractions vs average specific star formation. Disks, spheroids, peculiars and the other class are plotted in blue squares, red circles, pink diamonds, and gray crosses, respectively. Four redshift bins are shown and the `other` category aggregates the ambiguous, point source and unclassifiable sources. For all redshift bins and sSFR bins the disk galaxies dominate, with the exception of the highest redshift bin where disks and peculiar present similar contributions.

2005) due to the very different morphology fractions reported.

5.5.3 HST vs JWST

As discussed in Sec. 5.5.1, there is a stark difference between morphological classifications derived with HST and JWST observations. The new JWST data is challenging our understanding of galaxy evolution and structure formation in the early Universe. Here we discuss some differences between HST and JWST observations and analysis that contribute to this discrepancy.

A comparison for a select number of galaxies in our sample between the NIRCam stamps and the HST ACS and WFC3 stamps is shown in Figure 5.12. Many galaxies show very clear structures in NIRCam but ambiguous morphologies in HST. In some cases, such as EGS23205, only the central component is clearly seem in HST, while a disk, spiral arms and a bar pops up in the longer wavelength bands. In a few cases, such as EGS22543, the source is barely detected in the WFC3 and SW NIRCam images, while a clear disk is visible for the LW NIRCam stamps.

A detailed comparison with the CANDELS classifications (Kartaltepe et al., 2015) is beyond the scope of this Chapter as our classifications do not align perfectly with the scheme defined in Kartaltepe et al. (2015). However, we briefly discuss the modes for which the discrepancies between classifications based on HST imaging and NIRCam can be explained. First, many of the fine structure such as bars and spirals, are hard to resolve at high redshifts due to the WFC3 pixel scale, and can be mistaken by merging signatures or disturbances, such as the case of EGS14565. It shows clear spiral structure in JWST but not as clear in HST, which could be mistaken as a merger in HST. Secondly, we find the wavelength coverage to be critical, as many of the galaxies at high redshift in HST are probed in the blue side of the optical, and are prone to absorption from dust, giving rise to asymmetric looking structures. Moreover, the bluer bands probe the youngest stars, which have more irregular spatial distributions tracing sites of ongoing star formation but not the underlying mass distribution. Galaxies such as EGS522543 and EGS16559 are good examples of this. This is also expected to be an issue for galaxies at $z > 7$ in NIRCam images, as we start having the same issue as WFC3 for $2 < z < 3$.

Future morphology classification studies, with more detailed descriptions and covering larger datasets, such as the complete cycle of CEERS, PRIMER, CosmosWeb and NGDeep, together with the scope of citizen science projects like Galaxy Zoo (Lintott et al., 2011), will enable a detailed discussion on the main differences between HST and JWST morphology in the overlap region $2 < z < 3$.

5.6 Summary and Conclusions

We present results on the rest-frame optical morphologies and the structural evolution of JWST observed galaxies at $z = 1.5$ to 8 in a statistically significant sample of 4265 galaxies for the first time, using both visual classifications and quantitative morphology. We focus on galaxies observed by the CEERS program that overlap with the CANDELS fields, enabling us to use robust measurements of redshifts, stellar masses and star formation rates available in CANDELS.

Our major findings are:

- I. Distant galaxies at $z > 1.5$ display surprisingly regular disk morphologies at early times, such that for galaxies with $M_* < 10^{9.5} M_\odot$, the fraction of

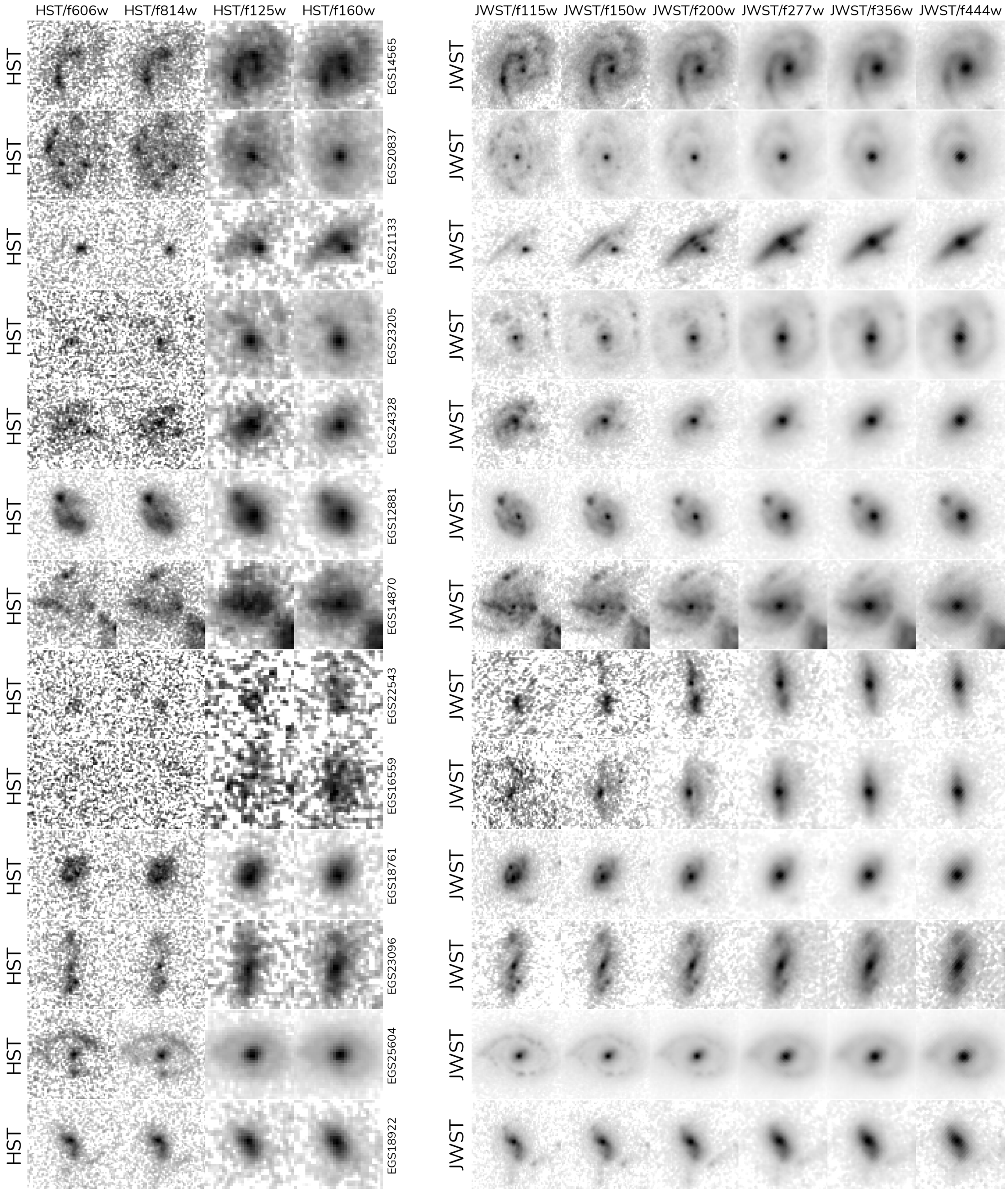


Figure 5.12: A HST vs. JWST comparison. We show 13 galaxies in this sample that have observations in the four main CANDELS filters (left panel) and SW and LW filters in JWST (right panel). Faint features in CANDELS are generally very clear in JWST. In some cases only the central core of the galaxy is visible with the HST imaging. Some merging signatures in HST are also resolved be to regular structures, such as disks and spirals.

disks/spheroids/peculiars seems to be constant for $1.5 < z < 6$, showing that the Hubble Sequence was already in place as early as one billion years after the Big Bang.

- II. For galaxies with higher masses $M_* > 10^{9.5} M_\odot$, tremendous evolution is observed in the fraction of disks and peculiars, suggesting that the role of mergers might be more important to the massive cases.

- III. Non-parametric morphology measurements agree well with visual classifications. However, a large overlap exists between classes in the usual CAS and G-M20 planes. We find that the spirality index (σ_ψ) when combined with the asymmetry (A) makes a powerful diagnostic to separate disks/spheroids/peculiars.

- IV. Comparisons with B/T studies from EAGLE and FLARES show that quantitative structures at high redshifts agree well with simulations, and are not unexpected from a theoretical standpoint, even if discrepant with previous morphological studies with HST.

- V. Galaxies with disk morphologies dominate both the low sSFR population of galaxies and high sSFR populations, fairly outnumbering spheroids. However, the peculiar contribution to the sSFR budget increases with increasing sSFR and redshift, such that at the highest redshifts, the majority of the highly star forming galaxy population has disturbed/peculiar morphologies.

- VI. The contribution to the total stellar mass of galaxies at high redshift is dominated by peculiar galaxies, while most of the stellar mass in the Universe at $z < 3$ is located in disk galaxies.

- VII. We report clear examples of galaxies whose morphologies are hidden in HST imaging, but become clear and unambiguous in the NIRCam observations. Spirals and bars are better resolved and clear in the LW NIRCam filters.

In addition to the morphology study presented in this Chapter, we release the first version of our catalog of aggregated classifications to the community. Our goal is that this large sample of visually classified galaxies will serve as a base for early studies on morphology and structure, and will help the community develop methods and tools to tackle scheduled larger area observations such as CosmosWeb while data releases from large citizen science classifications projects, such as GalaxyZoo, are not available. These classifications can be used, for example, as an early training dataset for deep learning methods, or as a transfer learning sample for already established models.

Chapter 6

Conclusions and Future Work

In this thesis we discussed the rest-frame morphological evolution of galaxies from $z = 0.5$ to 8 with ACS and WFC3 HST imaging from the CANDELS fields, NIRCam JWST imaging from the CEERS programs, and datasets of mock observations generated from the IllustrisTNG cosmological simulations, both from TNG300-1 and TNG100-1. This was done by combining several independent methodologies, ranging from visual classifications to quantitative morphologies with MORFOMETRYKA, and artificial intelligence deep learning models.

In Chapter 2 and Chapter 3 we combined cosmological simulations and deep learning to explore challenges regarding the nature of cosmic evolution of galaxy mergers in CANDELS. Chapter 4 and Chapter 5 are dedicated to the rest-frame optical morphological evolution of galaxies from $z = 1.5$ to 8 in the newly available observations from JWST with NIRCam, covering the wavelength range from $\lambda = 0.9\mu m - 4.4\mu m$, opening a complete new window to the resolved morphology of extremely high redshift galaxies.

Over the subsequent sections I will summarize the key results presented in this thesis, following with a brief discussion future avenues of exploration enabled by these developments.

6.1 Deep learning based galaxy merger rates since $z \sim 3$

Supervised deep learning based methods rely on previously defined labels. Common practice is the use of visual classifications to inform data-sets used for training. However, in the case of mergers their morphologies are so varied and rich, with strong dependence on timescales and their dynamical state, such that defining the stage of a merger through visual classifications is a very challenging task, as information about its evolution is not available directly from observations. In Chapter 2 we describe a novel way to combine deep learning and cosmological simulations to produce classification models that are physically motivated by the formation histories of galaxies extracted from merger trees. With a large F125W and F160W imaging mock dataset of mergers and non-mergers from IllustrisTNG 300-1, covering a wide range of states and configurations, we trained CNN models capable to classifying mergers with up to 90% accuracy, with only their images as input.

Applying these models to all the CANDELS fields enabled us to measure the galaxy merger rates up to $z = 3$. This is the first agreement between morphology and results based on galaxy pairs counts. We find that these rates follow a power law of the form $R(z) = 0.02 \times (1 + z)^{2.76 \pm 0.21}$, with the highest merger rates at the highest redshifts. Additionally, we show that the deep learning classifications follow similar cues to the ones that visual classifications are based on, and that miss-classifications are often due ambiguous visual morphologies.

6.2 The cosmic evolution of clumpy galaxies in CANDELS since $z \sim 3$

The successful approach described in Chapter 2 led us to use the same framework to explore other open questions regarding the high redshift morphology of merging galaxies. One such challenge is the difficult selection between galaxies that recently underwent a major merging event, and non-interacting galaxies that are highly star-forming. These two different populations are notorious for being ambiguous, both can be asymmetric, peculiar looking and clumpy, but for different physical reasons.

To uncover the nature of these clumpy galaxies, we refine the framework from Chapter 2, adapting it for the higher resolution simulation Illustris TNG100-1, and making adjustments to our mock imaging pipelines to include a full description of unresolved and resolved dust with the radiative transfer code SKIRT, a critical requirement for the modeling fidelity of clumpy star forming regions and dust lanes that contribute to the complexity of these morphologies. We generate a large sample of 160,000 galaxy mocks of non-interacting star forming galaxies and post-mergers. With these data, we design a training schedule with two target models for each one of the CANDELS fields, a model for low redshift ($z < 1.5$) and a model for high redshifts ($z > 1.5$), given the different sampling scale between the ACS and NIR-WFC3 cameras.

With these models, we classify all asymmetric galaxies in CANDELS ($A > 0.1$) in one of the two classes. By exploring their relative fractions on this sample we are able to trace the evolution of post-merger counts over $0.5 < z < 1.5$, showing that there is an increase in post-mergers among the peculiar galaxies from $z = 0.5$ to 3.

By visualizing the outputs of the layers of these models, we compare the features extracted from the simulated images and from real CANDELS galaxies, showing that they are very similar and occupy similar regions of a UMAP diagram. Moreover, we show that the asymmetry alone performs very poorly at this task, with almost 100% overlap between the distributions.

Finally, we show that recent major mergers have a similar impact to the scattering above the star forming main sequence of galaxies as non-interacting star-forming galaxies. Finally we discuss how to update the merger rates from Chapter 2.

6.3 Abundance of disk galaxies in the early universe

Following the first data release of the recently launched JWST, we quickly prepared to process the raw data products for the NIRCam observations covering $\lambda = 0.9\mu m$ to $4.4\mu m$. Within this wavelength coverage, we are able to resolve the rest-frame morphologies of galaxies at $z > 3$ for the first time. In particular, we investigate the morphologies of galaxies previously detected with HST in the ERO SMACS 0723 cluster observations in Chapter 4.

Through visual classifications of disks, spheroids and peculiars, we show that the fraction of disk galaxies in the JWST observations for $z > 1.5$ compared to HST observations (Mortlock et al., 2013) are 10 times higher than expected. By measuring the quantitative morphologies of these galaxies we also report that their concentrations, asymmetries and Sersic profiles are overall consistent with a dominance of disky features.

6.4 Evidences for the Hubble sequence at z > 3 with JWST

Based on the unexpected results we reported in Chapter 4, we follow up in Chapter 5 with an investigation using the early observations of the CEERS ERS JWST program that overlap with the EGS observations in the CANDELS fields. Using galaxies previously detected in the CANDELS fields, we select 4265 galaxies that are observed by both imaging programs, that have photometric redshifts, stellar masses and star formation rates.

We carry out an extensive classification effort to visually classify all the galaxies in this sample and investigate the evolution of the rest-frame optical Hubble sequence from $z = 1.5$ to 8. We show that the abundance of disks in the early universe is also present in this larger sample, covering a wider area, and the fraction of disk, peculiar and spheroid morphologies seem to be fairly constant over that redshift range for galaxies with $M_* < 10^{9.5} M_\odot$. For high mass galaxies with $M_* > 10^{9.5} M_\odot$, there is a strong evolution between disks and peculiar galaxies. However, these trends happen much earlier than what is predicted and observed by HST observations.

6.5 Future work

The pathway forward has a multitude of parallel branches of investigation on the morphological evolution of galaxies since $z = 8$, especially now with the start of operations of JWST. The most obvious next step based on the results presented in this thesis is a detailed study on the importance of galaxy mergers to morphological transformation at $z > 3$ leveraging JWST observations of the CEERS, PRIMER, and NGDEEP programs, extending the morphology-based galaxy merger rates to higher redshifts. Chapter 4 and Chapter 5 suggests that the role of merging on the morphology of galaxies seems to be very complex and to have started earlier than anticipated. In fact, the visual classifications that are part of Chapter 5 together with the still unexplored 'smooth' indicator, could provide merging classifications that could be used for this. These early JWST works are in preparation for that, and will help us understand and develop methods for larger programs, such as COSMOS-WEB.

On the other hand, the deep learning methods described in Chapter 2 and Chapter 3 will be soon available to be applied to JWST data, levering cosmological simulations especially dedicated to the early universe, such as the BlueTides simulations and the FLARES simulations. We still lack publicly available merger trees to make this possible, but this is possible to be produced from the current particle data.

The rest-frame optical size evolution of galaxies in this redshift range is now possible, given NIRCam observed sources as high as z=8 can be now resolved. Moreover, we now have new eyes to understand the SFR density of the universe.

Bibliography

Abadi M., Barham P., Chen J. et al., 2016. *TensorFlow: A system for large-scale machine learning.* arXiv e-prints, arXiv:1605.08695.

Abadi M.G., Navarro J.F., Steinmetz M. et al., 2003. *Simulations of Galaxy Formation in a Λ Cold Dark Matter Universe. I. Dynamical and Photometric Properties of a Simulated Disk Galaxy.* ApJ, 591, 2, 499–514.

Abraham R.G., Valdes F., Yee H.K.C. et al., 1994. *The morphologies of distant galaxies. 1: an automated classification system.* The Astrophysical Journal, 432, 75.

Ackermann S., Schawinski K., Zhang C. et al., 2018. *Using transfer learning to detect galaxy mergers.* Monthly Notices of the Royal Astronomical Society, 479, 1, 415–425.

Adams N.J., Conselice C.J., Ferreira L. et al., 2022. *Discovery and properties of ultra-high redshift galaxies (*$9 < z < 12$*) in the JWST ERO SMACS 0723 Field.* arXiv e-prints, arXiv:2207.11217.

Albernaz Ferreira L. & Ferrari F., 2018. *The Impact Of Redshift On Galaxy Morphometric Classification: Case Studies for SDSS, DES, LSST and HST with MORFOMETRYKA.* Monthly Notices of the Royal Astronomical Society, 473, 2, 2701–2713.

Allen M.G., Groves B.A., Dopita M.A. et al., 2008. *The MAPPINGS III Library of Fast Radiative Shock Models.* ApJS, 178, 1, 20–55.

Almeida J.S., Elmegreen B.G., Muñoz-Tuñón C. et al., 2014. *Star formation sustained by gas accretion.* Astronomy and Astrophysics Review, 22, 1, 1–60.

Alpher R.A. & Herman R., 1948. *Evolution of the Universe.* Nature, 162, 4124, 774–775.

Arnouts S., Cristiani S., Moscardini L. et al., 1999. *Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North.* MNRAS, 310, 2, 540–556.

Arp H., 1966. *Atlas of Peculiar Galaxies.* ApJS, 14, 1.

Ashby M.L.N., Willner S.P., Fazio G.G. et al., 2015. *S-CANDELS: The Spitzer-Cosmic Assembly Near-Infrared Deep Extragalactic Survey. Survey Design, Photometry, and Deep IRAC Source Counts.* ApJS, 218, 2, 33.

Atek H., Shuntov M., Furtak L.J. et al., 2022. *Revealing Galaxy Candidates out to* $z \sim 16$ *with JWST Observations of the Lensing Cluster SMACS0723.* arXiv e-prints, arXiv:2207.12338.

Baillard A., Bertin E., de Lapparent V. et al., 2011. *The EFIGI catalogue of 4458 nearby galaxies with detailed morphology.* A&A, 532, A74.

Barden M., Jahnke K. & Häußler B., 2008. *FERENGI: Redshifting Galaxies from SDSS to GEMS, STAGES, and COSMOS.* The Astrophysical Journal Supplement Series, 175, 1, 105–115.

Bell E.F., Wolf C., Meisenheimer K. et al., 2004. *Nearly 5000 Distant Early-Type Galaxies in COMBO-17: A Red Sequence and Its Evolution since z˜1.* ApJ, 608, 2, 752–767.

Benítez N., 2000. *Bayesian Photometric Redshift Estimation.* ApJ, 536, 2, 571–583.

Benítez N., Ford H., Bouwens R. et al., 2004. *Faint Galaxies in Deep Advanced Camera for Surveys Observations.* ApJS, 150, 1, 1–18.

Bershady M.A., Jangren A. & Conselice C.J., 2000. *Structural and Photometric Classification of Galaxies. I. Calibration Based on a Nearby Galaxy Sample.* The Astronomical Journal, 119, 6, 2645–2663.

Bertin E. & Arnouts S., 1996. *SExtractor: Software for source extraction.* Astronomy and Astrophysics Supplement, 117, 393–404.

Bertone S. & Conselice C.J., 2009. *A comparison of galaxy merger history observations and predictions from semi-analytic models.* MNRAS, 396, 4, 2345–2358.

Bickley R.W., Bottrell C., Hani M.H. et al., 2021. *Convolutional neural network identification of galaxy post-mergers in UNIONS using IllustrisTNG.* MNRAS, 504, 1, 372–392.

Blanton M.R., Brinkmann J., Csabai I. et al., 2003. *Estimating Fixed-Frame Galaxy Magnitudes in the Sloan Digital Sky Survey.* The Astronomical Journal, 125, 5, 2348–2360.

Bluck A.F.L., Conselice C.J., Buitrago F. et al., 2012. *THE STRUCTURES AND TOTAL (MINOR + MAJOR) MERGER HISTORIES OF MASSIVE*

GALAXIES UP TO $z \sim 3$ IN THE HST GOODS NICMOS SURVEY: A POSSIBLE SOLUTION TO THE SIZE EVOLUTION PROBLEM. The Astrophysical Journal, 747, 1, 34.

Bluck A.F.L., Mendel J.T., Ellison S.L. et al., 2014. *Bulge mass is king: the dominant role of the bulge in determining the fraction of passive galaxies in the Sloan Digital Sky Survey.* Monthly Notices of the Royal Astronomical Society, 441, 1, 599–629.

Blumenthal G.R., Faber S.M., Primack J.R. et al., 1984. *Formation of galaxies and large-scale structure with cold dark matter.* Nature, 311, 517–525.

Blumenthal K.A., Moreno J., Barnes J.E. et al., 2020. *Galaxy interactions in IllustrisTNG-100, I: The power and limitations of visual identification.* MNRAS, 492, 2, 2075–2094.

Bottrell C., Hani M.H., Teimoorinia H. et al., 2019. *Deep learning predictions of galaxy merger stage and the importance of observational realism.* MNRAS, 490, 4, 5390–5413.

Bottrell C., Hani M.H., Teimoorinia H. et al., 2021. *The combined and respective roles of imaging and stellar kinematics in identifying galaxy merger remnants.* MNRAS.

Bottrell C. & Hani M.H., 2022. *Realistic synthetic integral field spectroscopy with RealSim-IFS.* MNRAS, 514, 2, 2821–2838.

Bruzual G. & Charlot S., 2003. *Stellar population synthesis at the resolution of 2003.* MNRAS, 344, 4, 1000–1028.

Bruzual G. & Charlot S., 2003. *Stellar population synthesis at the resolution of 2003.* Monthly Notices of the Royal Astronomical Society, 344, 4, 1000–1028.

Buitrago F., Trujillo I., Conselice C.J. et al., 2008. *Size Evolution of the Most Massive Galaxies at $1.7 < z < 3$ from GOODS NICMOS Survey Imaging.* The Astrophysical Journal, 687, 2, L61–L64.

Buitrago F., Trujillo I., Conselice C.J. et al., 2012. *Early-type galaxies have been the predominant morphological class for massive galaxies since only z 1.* Monthly Notices of the Royal Astronomical Society, 428, 2, 1460–1478.

Buitrago F., Trujillo I., Conselice C.J. et al., 2013. *Early-type galaxies have been the predominant morphological class for massive galaxies since only $z \sim 1$.* MNRAS, 428, 2, 1460–1478.

Bundy K., Bershady M.A., Law D.R. et al., 2015. *Overview of the SDSS-IV MaNGA Survey: Mapping nearby Galaxies at Apache Point Observatory.* ApJ, 798, 1, 7.

Calzetti D., Armus L., Bohlin R.C. et al., 2000. *The Dust Content and Opacity of Actively Star-forming Galaxies.* The Astrophysical Journal, 533, 2, 682–695.

Camps P. & Baes M., 2015. *SKIRT: An advanced dust radiative transfer code with a user-friendly architecture.* Astronomy and Computing, 9, 20–33.

Camps P., Trayford J.W., Baes M. et al., 2016. *Far-infrared and dust properties of present-day galaxies in the EAGLE simulations.* MNRAS, 462, 1, 1057–1075.

Camps P. & Baes M., 2020. *SKIRT 9: Redesigning an advanced dust radiative transfer code to allow kinematics, line transfer and polarization by aligned dust grains.* Astronomy and Computing, 31.

Cappellari M., Emsellem E., Krajnović D. et al., 2011a. *The ATLAS^{3D} project - I. A volume-limited sample of 260 nearby early-type galaxies: science goals and selection criteria.* MNRAS, 413, 2, 813–836.

Cappellari M., Emsellem E., Krajnović D. et al., 2011b. *The ATLAS^{3D} project - VII. A new look at the morphology of nearby galaxies: the kinematic morphology-density relation.* MNRAS, 416, 3, 1680–1696.

Cardamone C., Schawinski K., Sarzi M. et al., 2009. *Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies.* MNRAS, 399, 3, 1191–1205.

Castellano M., Fontana A., Treu T. et al., 2022. *Early results from GLASS-JWST. III: Galaxy candidates at z$\sim$9-15.* arXiv e-prints, arXiv:2207.09436.

Cheng T.Y., Conselice C.J., Aragón-Salamanca A. et al., 2020. *Optimizing automatic morphological classification of galaxies with machine learning and deep learning using Dark Energy Survey imaging.* MNRAS, 493, 3, 4209–4228.

Chollet F. & others, 2015. *Keras.* \url{https://keras.io}.

Ćiprijanović A., Snyder G.F., Nord B. et al., 2020. *DeepMerge: Classifying high-redshift merging galaxies with deep neural networks.* Astronomy and Computing, 32, 100390.

Ćiprijanović A., Kafkes D., Downey K. et al., 2021. *DeepMerge - II. Building robust deep learning algorithms for merging galaxy identification across domains.* MNRAS, 506, 1, 677–691.

Clauwens B., Schaye J., Franx M. et al., 2018. *The three phases of galaxy formation.* MNRAS, 478, 3, 3994–4009.

Coe D., Benítez N., Sánchez S.F. et al., 2006. *Galaxies in the Hubble Ultra Deep Field. I. Detection, Multiband Photometry, Photometric Redshifts, and Morphology.* AJ, 132, 2, 926–959.

Coe D., Umetsu K., Zitrin A. et al., 2012. *CLASH: Precise New Constraints on the Mass Profile of the Galaxy Cluster A2261.* ApJ, 757, 1, 22.

Coe D., Bradley L. & Zitrin A., 2015. *Frontier Fields: High-redshift Predictions and Early Results.* ApJ, 800, 2, 84.

Coe D., Salmon B., Bradač M. et al., 2019. *RELICS: Reionization Lensing Cluster Survey.* ApJ, 884, 1, 85.

Cole S., Lacey C.G., Baugh C.M. et al., 2000. *Hierarchical galaxy formation.* MNRAS, 319, 1, 168–204.

Colless M., Dalton G., Maddox S. et al., 2001. *The 2dF Galaxy Redshift Survey: spectra and redshifts.* MNRAS, 328, 4, 1039–1063.

Conroy C., Gunn J.E. & White M., 2009. *The propagation of uncertainties in stellar population synthesis modeling. I. the relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies.* Astrophysical Journal, 699, 1, 486–506.

Conroy C. & Gunn J.E., 2010. *The propagation of uncertainties in stellar population synthesis modeling. III. Model calibration, comparison, and evaluation.* Astrophysical Journal, 712, 2, 833–857.

Conselice C.J., Bershady M.A. & Jangren A., 2000. *The Asymmetry of Galaxies: Physical Morphology for Nearby and High-Redshift Galaxies.* ApJ, 529, 2, 886–910.

Conselice C.J., Bershady M.A., Dickinson M. et al., 2003. *A Direct Measurement of Major Galaxy Mergers at z 3* . The Astronomical Journal, 126, 3, 1183–1207.

Conselice C.J., 2003. *the Relationship Between Stellar Light Distributions of Galaxies and Their.* ApJS, 147, 1–28.

Conselice C.J., Blackburne J.A. & Papovich C., 2005. *The Luminosity, Stellar Mass, and Number Density Evolution of Field Galaxies of Known Morphology from $z = 0.5$ to 3.* ApJ, 620, 2, 564–583.

Conselice C.J., 2006. *The fundamental properties of galaxies and a new galaxy classification system.* Monthly Notices of the Royal Astronomical Society, 373, 4, 1389–1408.

Conselice C.J., Rajgor S. & Myers R., 2008. *The structures of distant galaxies - I. Galaxy structures and the merger rate to z ~3 in the Hubble Ultra-Deep Field.* MNRAS, 386, 2, 909–927.

Conselice C.J. & Arnold J., 2009. *The structures of distant galaxies - II. Diverse galaxy structures and local environments at $z = 4$-6 implications for early galaxy assembly.* MNRAS, 397, 1, 208–231.

Conselice C.J., 2009. *The structures of distant galaxies - IV. A new empirical measurement of the time-scale for galaxy mergers - implications for the merger history.* Monthly Notices of the Royal Astronomical Society: Letters, 399, 1, 16–20.

Conselice C.J., Bluck A.F.L., Ravindranath S. et al., 2011. *The tumultuous formation of the Hubble sequence at $z > 1$ examined with HST/Wide-Field Camera-3 observations of the Hubble Ultra Deep Field.* MNRAS, 417, 4, 2770–2788.

Conselice C.J., Bluck A.F., Mortlock A. et al., 2014. *Galaxy formation as a cosmological tool - I. The galaxy merger history as a measure of cosmological parameters.* Monthly Notices of the Royal Astronomical Society, 444, 2, 1125–1143.

Conselice C.J., 2014. *The Evolution of Galaxy Structure Over Cosmic Time.* Annual Review of Astronomy and Astrophysics, 52, 1, 291–337.

Cook L.T., Zhu Y., Hall T.J. et al., 2000. *Bioelasticity imaging II: Spatial resolution.* Proceedings of SPIE - The International Society for Optical Engineering, 3982.

Cortijo-Ferrero C., González Delgado R.M., Pérez E. et al., 2017. *Star formation histories in mergers: the spatially resolved properties of the early-stage merger luminous infrared galaxies IC 1623 and NGC 6090.* MNRAS, 467, 4, 3898–3919.

Crain R.A., McCarthy I.G., Frenk C.S. et al., 2010. *X-ray coronae in simulations of disc galaxy formation.* MNRAS, 407, 3, 1403–1422.

Crain R.A., Schaye J., Bower R.G. et al., 2015. *The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations.* MNRAS, 450, 2, 1937–1961.

Dahlen T., Mobasher B., Dickinson M. et al., 2007. *Evolution of the Luminosity Function, Star Formation Rate, Morphology, and Size of Star-forming Galaxies Selected at Rest-Frame 1500 and 2800 Å.* ApJ, 654, 1, 172–185.

Darg D.W., Kaviraj S., Lintott C.J. et al., 2010. *Galaxy Zoo: The properties of merging galaxies in the nearby Universe - Local environments, colours, masses, star formation rates and AGN activity.* Monthly Notices of the Royal Astronomical Society, 401, 3, 1552–1563.

de Lapparent V., Baillard A. & Bertin E., 2011. *The EFIGI catalogue of 4458 nearby galaxies with morphology. II. Statistical properties along the Hubble sequence.* A&A, 532, A75.

de Vaucouleurs G., 1948. *Recherches sur les Nebuleuses Extragalactiques.* Annales d'Astrophysique, 11.

de Vaucouleurs G., 1961. *Integrated Colors of Bright Galaxies in the u, b, V System.* ApJS, 5, 233.

Delgado-Serrano R., Hammer F., Yang Y.B. et al., 2010. *How was the Hubble sequence 6 Gyr ago?* A&A, 509, A78.

Dickinson M., Hanley C., Elston R. et al., 2000. *The Unusual Infrared Object HDF-N J123656.3+621322.* ApJ, 531, 2, 624–634.

Donnan C.T., McLeod D.J., Dunlop J.S. et al., 2022. *The evolution of the galaxy UV luminosity function at redshifts z ~8-15 from deep JWST and ground-based near-infrared imaging.* arXiv e-prints, arXiv:2207.12356.

Dressler A., 1980. *Galaxy morphology in rich clusters: implications for the formation and evolution of galaxies.* ApJ, 236, 351–365.

Dressler A., Oemler Augustus J., Butcher H.R. et al., 1994. *The Morphology of Distant Cluster Galaxies. I. HST Observations of CL 0939+4713.* ApJ, 430, 107.

Driver S.P., Windhorst R.A., Ostrander E.J. et al., 1995. *The Morphological Mix of Field Galaxies to M I = 24.25 mag (b J approximately 26 mag) from a Deep Hubble Space Telescope WFPC2 Image.* ApJ, 449, L23.

Duncan K., Conselice C.J., Mortlock A. et al., 2014. *The mass evolution of the first galaxies: stellar mass functions and star formation rates at $4 < z < 7$ in the CANDELS GOODS-South field.* MNRAS, 444, 3, 2960–2984.

Duncan K., Conselice C.J., Mundy C. et al., 2019. *Observational Constraints on the Merger History of Galaxies since $z \approx 6$: Probabilistic Galaxy Pair Counts in the CANDELS Fields.* ApJ, 876, 2, 110.

Ebeling H., Edge A.C., Mantz A. et al., 2010. *The X-ray brightest clusters of galaxies from the Massive Cluster Survey.* MNRAS, 407, 1, 83–93.

Eisenstein D.J., Zehavi I., Hogg D.W. et al., 2005. *Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies.* The Astrophysical Journal, 633, 2, 560–574.

Ellison S.L., Patton D.R., Mendel J.T. et al., 2011. *Galaxy pairs in the Sloan Digital Sky Survey - IV. Interactions trigger active galactic nuclei.* MNRAS, 418, 3, 2043–2053.

Ellison S.L., Wilkinson S., Woo J. et al., 2022. *Galaxy mergers can rapidly shut down star formation.* arXiv e-prints, arXiv:2209.07613.

Elmegreen B.G. & Elmegreen D.M., 2005. *Stellar Populations in 10 Clump-Cluster Galaxies of the Hubble Ultra Deep Field.* ApJ, 627, 2, 632–646.

Euclid Collaboration, Merlin E., Castellano M. et al., 2022a. *Euclid preparation. XXV. The Euclid Morphology Challenge – Towards model-fitting photometry for billions of galaxies.* arXiv e-prints, arXiv:2209.12906.

Euclid Collaboration, Bretonnière H., Kuchner U. et al., 2022b. *Euclid preparation XXVI: The Euclid Morphology Challenge. Towards structural parameters for billions of galaxies.* arXiv e-prints, arXiv:2209.12907.

Faber S.M. & Jackson R.E., 1976. *Velocity dispersions and mass-to-light ratios for elliptical galaxies.* ApJ, 204, 668–683.

Feng Y., Di-Matteo T., Croft R.A. et al., 2015. *The BlueTides simulation: first galaxies and reionization.* MNRAS, 455, 3, 2778–2791.

Ferland G.J., Chatzikos M., Guzmán F. et al., 2017. *The 2017 release of Cloudy.* Revista Mexicana de Astronomia y Astrofisica, 53, 2, 385–438.

Ferrari F., de Carvalho R.R. & Trevisan M., 2015. *Morfometryka—A New Way of Establishing Morphological Classification of Galaxies.* ApJ, 814, 1, 55.

Ferreira L. & Ferrari F., 2018. *The impact of redshift on galaxy morphometric classification: case studies for SDSS, DES, LSST and HST with MORFOMETRYKA.* MNRAS, 473, 2, 2701–2713.

Ferreira L., Conselice C.J., Duncan K. et al., 2020. *Galaxy Merger Rates up to z {\ensuremath{\sim}} 3 Using a Bayesian Deep Learning Model: A Major-merger Classifier Using IllustrisTNG Simulation Data.* ApJ, 895, 2, 115.

Ferreira L., Conselice C.J., Kuchner U. et al., 2022a. *A Simulation-driven Deep Learning Approach for Separating Mergers and Star-forming Galaxies: The Formation Histories of Clumpy Galaxies in All of the CANDELS Fields.* ApJ, 931, 1, 34.

Ferreira L., Adams N., Conselice C.J. et al., 2022b. *Panic! At the Disks: First Rest-frame Optical Observations of Galaxy Structure at $z > 3$ with JWST in the SMACS 0723 Field.* arXiv e-prints, arXiv:2207.09428.

Font A.S., Johnston K.V., Bullock J.S. et al., 2006. *Chemical Abundance Distributions of Galactic Halos and Their Satellite Systems in a ΛCDM Universe.* ApJ, 638, 2, 585–595.

Frei Z., Guhathakurta P., Gunn J.E. et al., 1996. *A Catalog of Digital Images of 113 Nearby Galaxies.* AJ, 111, 174.

Gamow G., 1946. *Expanding Universe and the Origin of Elements.* Physical Review, 70, 7-8, 572–573.

Genzel R., Tacconi L.J., Rigopoulou D. et al., 2001. *Ultraluminous Infrared Mergers: Elliptical Galaxies in Formation?* ApJ, 563, 2, 527–545.

George K., 2017. *Structural analysis of star-forming blue early-type galaxies. Merger-driven star formation in elliptical galaxies.* A&A, 598, A45.

Golubchik M., Furtak L.J., Meena A.K. et al., 2022. *HST strong-lensing model for the first JWST galaxy cluster SMACS J0723.3-7327.* arXiv e-prints, arXiv:2207.05007.

Good I.J., 1952. *Rational Decisions.* Journal of the Royal Statistical Society. Series B (Methodological), 14, 1, 107–114.

Goodfellow I., Bengio Y. & Courville A., 2016. *Deep Learning.* MIT Press.

Griffiths R.E., Casertano S., Ratnatunga K.U. et al., 1994. *The Morphology of Faint Galaxies in Medium Deep Survey Images Using WFPC2.* ApJ, 435, L19.

Grogin N.A., Kocevski D.D., Faber S.M. et al., 2011. *Candels: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey.* Astrophysical Journal, Supplement Series, 197, 2.

Guo Y., Ferguson H.C., Bell E.F. et al., 2015. *Clumpy Galaxies in CANDELS. I. The Definition of UV Clumps and the Fraction of Clumpy Galaxies at 0.5 < z < 3.* ApJ, 800, 1, 39.

Guo Y., Rafelski M., Bell E.F. et al., 2018. *Clumpy Galaxies in CANDELS. II. Physical Properties of UV-bright Clumps at* $0.5 \leq z < 3$. ApJ, 853, 2, 108.

Hacohen G. & Weinshall D., 2019. *All Neural Networks are Created Equal.* pages 1–12.

Hani M.H., Gosain H., Ellison S.L. et al., 2020. *Interacting galaxies in the IllustrisTNG simulations - II: star formation in the post-merger stage.* MNRAS, 493, 3, 3716–3731.

Häussler B., McIntosh D.H., Barden M. et al., 2007. *GEMS: Galaxy Fitting Catalogs and Testing Parametric Galaxy Fitting Codes: GALFIT and GIM2D.* ApJS, 172, 2, 615–633.

Hogg D.W., 1999. *Distance measures in cosmology.* arXiv.org, astro-ph, 5116.

Hogg D.W., Blanton M.R., Eisenstein D.J. et al., 2003. *The Overdensities of Galaxy Environments as a Function of Luminosity and Color.* ApJ, 585, 1, L5–L9.

Hopkins P.F., Hernquist L., Cox T.J. et al., 2008. *A Cosmological Framework for the Co-Evolution of Quasars, Supermassive Black Holes, and Elliptical Galaxies. I. Galaxy Mergers and Quasar Activity.* ApJS, 175, 2, 356–389.

Houghton R.C.W., 2015. *Revisiting the original morphology-density relation.* MNRAS, 451, 4, 3427–3436.

Hubble E., 1929a. *A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae.* Proceedings of the National Academy of Science, 15, 168–173.

Hubble E.P., 1926. *Extragalactic nebulae.* apj, 64, 321–369.

Hubble E.P., 1929b. *A spiral nebula as a stellar system, Messier 31.* ApJ, 69, 103–158.

Hubble E.P., 1936. *Realm of the Nebulae.*

Huertas-Company M., Tasca L., Rouan D. et al., 2009. *A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images. II. Quantifying morphological k-correction in the COSMOS field at 1 < z < 2: Ks band vs. I band.* A&A, 497, 3, 743–753.

Huertas-Company M., Gravet R., Cabrera-Vives G. et al., 2015. *A catalog of visual-like morphologies in the 5 candels fields using deep learning.* Astrophysical Journal, Supplement Series, 221, 1, 8.

Huertas-Company M., Bernardi M., Pérez-González P.G. et al., 2016. *Mass assembly and morphological transformations since z ˜ 3 from CANDELS.* Monthly Notices of the Royal Astronomical Society, 462, 4, 4495–4516.

Huertas-Company M., Primack J.R., Dekel A. et al., 2018. *Deep Learning Identifies High-z Galaxies in a Central Blue Nugget Phase in a Characteristic Mass Range.* ApJ, 858, 2, 114.

Huertas-Company M., Rodriguez-Gomez V., Nelson D. et al., 2019. *The Hubble Sequence at $z \sim 0$ in the IllustrisTNG simulation with deep learning.* MNRAS, 489, 2, 1859–1879.

Huertas-Company M., Guo Y., Ginzburg O. et al., 2020. *Stellar masses of giant clumps in CANDELS and simulated galaxies using machine learning.* MNRAS, 499, 1, 814–835.

Ilbert O., Arnouts S., McCracken H.J. et al., 2006. *Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey.* A&A, 457, 3, 841–856.

Ilbert O., Capak P., Salvato M. et al., 2009. *Cosmos photometric redshifts with 30-bands for 2-deg 2.* ApJ, 690, 1236–1249.

Irodotou D., Thomas P.A., Henriques B.M. et al., 2019. *Morphological evolution and galactic sizes in the L-Galaxies SA model.* MNRAS, 489, 3, 3609–3624.

Irodotou D. & Thomas P.A., 2021. *Using angular momentum maps to detect kinematically distinct galactic components.* MNRAS, 501, 2, 2182–2197.

Jacobs C., Glazebrook K., Calabrò A. et al., 2022. *Early results from GLASS-JWST XIV: A first morphological atlas of the* $1 < z < 5$ *Universe in the rest-frame optical.* arXiv e-prints, arXiv:2208.06516.

Jiménez-Teja Y. & Benítez N., 2012. *A New Tool for Image Analysis Based on Chebyshev Rational Functions: CHEF Functions.* ApJ, 745, 2, 150.

Jogee S., Miller S.H., Penner K. et al., 2009. *History of Galaxy Interactions and Their Impact on Star Formation Over the Last 7 Gyr from GEMS.* ApJ, 697, 2, 1971–1992.

Kartaltepe J.S., Mozena M., Kocevski D. et al., 2015. *CANDELS Visual Classifications: Scheme, Data Release, and First Results.* ApJS, 221, 1, 11.

Kartaltepe J.S., Rose C., Vanderhoof B.N. et al., 2022. *CEERS Key Paper III: The Diversity of Galaxy Structure and Morphology at z=3-9 with JWST.* arXiv e-prints, arXiv:2210.14713.

Kelvin L.S., Driver S.P., Robotham A.S.G. et al., 2012. *Galaxy And Mass Assembly (GAMA): Structural Investigation of Galaxies via Model Analysis.* MNRAS, 421, 2, 1007–1039.

Koekemoer A.M., Faber S.M., Ferguson H.C. et al., 2011. *Candels: The cosmic assembly near-infrared deep extragalactic legacy survey - The hubble space telescope observations, imaging data products, and mosaics.* Astrophysical Journal, Supplement Series, 197, 2.

Kormendy J. & Bender R., 2012. *A Revised Parallel-sequence Morphological Classification of Galaxies: Structure and Formation of S0 and Spheroidal Galaxies.* ApJS, 198, 1, 2.

Krist J., Hook R. & Tim T., 2004. *The Tiny Tim User ' s Guide.* Changes, , June.

Krizhevsky A., Sutskever I. & Hinton G.E., 2017. *ImageNet Classification with Deep Convolutional Neural Networks.* Commun. ACM, 60, 6, 84–90.

Lagos C.d.P., Stevens A.R.H., Bower R.G. et al., 2018. *Quantifying the impact of mergers on the angular momentum of simulated galaxies.* MNRAS, 473, 4, 4956–4974.

Laureijs R., Amiaux J., Arduini S. et al., 2011. *Euclid Definition Study Report.* arXiv e-prints, arXiv:1110.3193.

Lintott C., Schawinski K., Bamford S. et al., 2011. *Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies.* MNRAS, 410, 1, 166–178.

Loshchilov I. & Hutter F., 2016. *SGDR: Stochastic Gradient Descent with Warm Restarts.* arXiv e-prints, arXiv:1608.03983.

Lotz J.M., Primack J. & Madau P., 2004. *A New Nonparametric Approach to Galaxy Morphological Classification.* The Astronomical Journal, 128, 1, 163–182.

Lotz J.M., Madau P., Giavalisco M. et al., 2006. *The Rest-Frame Far-Ultraviolet Morphologies of Star-forming Galaxies at z ˜1.5 and 4.* ApJ, 636, 2, 592–609.

Lotz J.M., Jonsson P., Cox T.J. et al., 2008. *Galaxy merger morphologies and time-scales from simulations of equal-mass gas-rich disc mergers.* Monthly Notices of the Royal Astronomical Society, 391, 3, 1137–1162.

Lovell C.C., Vijayan A.P., Thomas P.A. et al., 2021. *First Light And Reionization Epoch Simulations (FLARES) - I. Environmental dependence of high-redshift galaxy evolution.* MNRAS, 500, 2, 2127–2145.

LSST Science Collaboration, Abell P.A., Allison J. et al., 2009. *LSST Science Book, Version 2.0.* arXiv e-prints, arXiv:0912.0201.

Lu Z., Pu H., Wang F. et al., 2017. *The Expressive Power of Neural Networks: A View from the Width.* arXiv e-prints, arXiv:1709.02540.

Lubin L.M. & Sandage A., 2001. *The Tolman Surface Brightness Test for the Reality of the Expansion. IV. A Measurement of the Tolman Signal and the Luminosity Evolution of Early-Type Galaxies.* AJ, 122, 3, 1084–1103.

Lucatelli G. & Ferrari F., 2019. *Galaxy structural analysis with the curvature of the brightness profile.* MNRAS, 489, 1, 1161–1180.

Madau P., 1995. *Radiative Transfer in a Clumpy Universe: The Colors of High-Redshift Galaxies.* ApJ, 441, 18.

Madau P. & Dickinson M., 2014. *Cosmic Star-Formation History.* Annual Review of Astronomy and Astrophysics, 52, 1, 415–486.

Mager V.A., Conselice C.J., Seibert M. et al., 2018. *Galaxy Structure in the Ultraviolet: The Dependence of Morphological Parameters on Rest-frame Wavelength.* ApJ, 864, 2, 123.

Man A.W.S., Zirm A.W. & Toft S., 2016. *Resolving the Discrepancy of Galaxy Merger Fraction Measurements at z $\sim$ 0-3.* ApJ, 830, 2, 89.

Mantha K.B., McIntosh D.H., Brennan R. et al., 2018. *Major merging history in CANDELS. I. Evolution of the incidence of massive galaxy-galaxy pairs from z = 3 to z ∼ 0.* MNRAS, 475, 2, 1549–1573.

Margalef-Bentabol B., Conselice C.J., Mortlock A. et al., 2018. *Stellar populations, stellar masses and the formation of galaxy bulges and discs at z < 3 in CANDELS.* MNRAS, 473, 4, 5370–5384.

Margalef-Bentabol B., Conselice C.J., Haeussler B. et al., 2022. *Observations of the initial formation and evolution of spiral galaxies at 1 < z < 3 in the CANDELS fields.* MNRAS, 511, 1, 1502–1517.

Marinacci F., Vogelsberger M., Pakmor R. et al., 2018. *First results from the IllustrisTNG simulations: radio haloes and magnetic fields.* MNRAS, 480, 4, 5113–5139.

Marshall M.A., Ni Y., Matteo T.D. et al., 2020. *The host galaxies of z = 7 quasars: predictions from the BlueTides simulation.* MNRAS, 499, 3, 3819–3836.

Marshall M.A., Watts K., Wilkins S. et al., 2022a. *BlueTides Mock Image Catalogue.*

Marshall M.A., Watts K., Wilkins S. et al., 2022b. *The BLUETIDES mock image catalogue: simulated observations of high-redshift galaxies and predictions for JWST imaging surveys.* MNRAS, 516, 1, 1047–1061.

Marshall P.J., Lintott C.J. & Fletcher L.N., 2015. *Ideas for Citizen Science in Astronomy.* ARA&A, 53, 247–278.

Martin G., Kaviraj S., Hocking A. et al., 2020. *Galaxy morphological classification in deep-wide surveys via unsupervised machine learning.* MNRAS, 491, 1, 1408–1426.

Masters K.L., Mosleh M., Romer A.K. et al., 2010. *Galaxy Zoo: passive red spirals.* MNRAS, 405, 2, 783–799.

McElroy R., Bottrell C., Hani M.H. et al., 2022. *The observability of galaxy merger signatures in nearby gas-rich spirals.* MNRAS, 515, 3, 3406–3419.

McInnes L., Healy J. & Melville J., 2018. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.* arXiv e-prints, arXiv:1802.03426.

Melvin T., Masters K., Lintott C. et al., 2014. *Galaxy Zoo: an independent look at the evolution of the bar fraction over the last eight billion years from HST-COSMOS.* MNRAS, 438, 4, 2882–2897.

Minkowski R., 1962. *Internal Dispersion of Velocities in Other Galaxies.* volume 15, page 112.

Mo H., van den Bosch F.C. & White S., 2010. *Galaxy Formation and Evolution.*

Monachesi A., Gómez F.A., Grand R.J.J. et al., 2019. *The Auriga stellar haloes: connecting stellar population properties with accretion and merging history.* MNRAS, 485, 2, 2589–2616.

Mortlock A., Conselice C.J., Hartley W.G. et al., 2013. *The redshift and mass dependence on the formation of the Hubble sequence at* $z > 1$ *from CANDELS/UDS.* MNRAS, 433, 2, 1185–1201.

Mowla L.A., Dokkum P.v., Brammer G.B. et al., 2019. *COSMOS-DASH: The Evolution of the Galaxy Size–Mass Relation since* $z \sim 3$ *from New Wide-field WFC3 Imaging Combined with CANDELS/3D- HST .* The Astrophysical Journal, 880, 1, 57.

Mundy C.J., Conselice C.J., Duncan K.J. et al., 2017. *A consistent measure of the merger histories of massive galaxies using close-pair statistics - I. Major mergers at* $z < 3.5$*.* Monthly Notices of the Royal Astronomical Society, 470, 3, 3507–3531.

Naidu R.P., Oesch P.A., Setton D.J. et al., 2022. *Schrodinger's Galaxy Candidate: Puzzlingly Luminous at* $z \approx 17$*, or Dusty/Quenched at* $z \approx 5$*?* arXiv e-prints, arXiv:2208.02794.

Naiman J.P., Pillepich A., Springel V. et al., 2018. *First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium.* MNRAS, 477, 1, 1206–1224.

Nair P.B. & Abraham R.G., 2010. *A Catalog of Detailed Visual Morphological Classifications for 14,034 Galaxies in the Sloan Digital Sky Survey.* ApJS, 186, 2, 427–456.

Nelson D., Springel V., Pillepich A. et al., 2019. *The IllustrisTNG simulations: public data release.* Computational Astrophysics and Cosmology, 6, 1, 2.

Nelson E.J., Suess K.A., Bezanson R. et al., 2022. *JWST reveals a population of ultra-red, flattened disk galaxies at* $2<z<6$ *previously missed by HST.* arXiv e-prints, arXiv:2208.01630.

Oesch P.A., Carollo C.M., Feldmann R. et al., 2010. *The Buildup of the Hubble Sequence in the Cosmos Field.* ApJ, 714, 1, L47–L51.

Oke J.B., 1974. *Absolute Spectral Energy Distributions for White Dwarfs.* ApJS, 27, 21.

Oke J.B. & Gunn J.E., 1983. *Secondary standard stars for absolute spectrophotometry.* The Astrophysical Journal, 266, 1974, 713.

Opik E., 1922. *An estimate of the distance of the Andromeda Nebula.* ApJ, 55, 406–410.

Papovich C., Dickinson M., Giavalisco M. et al., 2005. *The Assembly of Diversity in the Morphologies and Stellar Populations of High-Redshift Galaxies.* ApJ, 631, 1, 101–120.

Park C., Lee J., Kim J. et al., 2022. *Formation and Morphology of the First Galaxies in the Cosmic Morning.* arXiv e-prints, arXiv:2202.11925.

Pascale M., Frye B., Diego J. et al., 2022. *Unscrambling the lensed galaxies in JWST images behind SMACS0723.* arXiv e-prints, arXiv:2207.07102.

Paszke A., Gross S., Massa F. et al., 2019. *PyTorch: An Imperative Style, High-Performance Deep Learning Library.* arXiv e-prints, arXiv:1912.01703.

Patton D.R., Wilson K.D., Metrow C.J. et al., 2020. *Interacting galaxies in the IllustrisTNG simulations - I: Triggered star formation in a cosmological context.* MNRAS, 494, 4, 4969–4985.

Pawlik M.M., Wild V., Walcher C.J. et al., 2016. *Shape asymmetry: a morphological indicator for automatic detection of galaxies in the post-coalescence merger stages.* MNRAS, 456, 3, 3032–3052.

Pearson W.J., Wang L., Alpaslan M. et al., 2019. *Effect of galaxy mergers on star-formation rates.* A&A, 631, A51.

Pearson W.J., Wang L., Trayford J.W. et al., 2019. *Identifying galaxy mergers in observations and simulations with deep learning.* Astronomy & Astrophysics, 626, A49.

Peng Y.j., Lilly S.J., Kovač K. et al., 2010. *MASS AND ENVIRONMENT AS DRIVERS OF GALAXY EVOLUTION IN SDSS AND zCOSMOS AND THE ORIGIN OF THE SCHECHTER FUNCTION.* The Astrophysical Journal, 721, 1, 193–221.

Penzias A.A. & Wilson R.W., 1965. *A Measurement of Excess Antenna Temperature at 4080 Mc/s.* The Astrophysical Journal, 142, 419.

Perrin M.D., Long J., Sivaramakrishnan A. et al., 2015. *WebbPSF: James Webb Space Telescope PSF Simulation Tool.*

Peschken N., Łokas E.L. & Athanassoula E., 2020. *Disc galaxies formed from major mergers in Illustris.* MNRAS, 493, 1, 1375–1387.

Pillepich A., Madau P. & Mayer L., 2015. *Building Late-type Spiral Galaxies by In-situ and Ex-situ Star Formation.* ApJ, 799, 2, 184.

Pillepich A., Nelson D., Hernquist L. et al., 2018a. *First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies.* MNRAS, 475, 1, 648–675.

Pillepich A., Springel V., Nelson D. et al., 2018b. *Simulating galaxy formation with the IllustrisTNG model.* MNRAS, 473, 3, 4077–4106.

Pillepich A., Nelson D., Springel V. et al., 2019. *First results from the TNG50 simulation: the evolution of stellar and gaseous discs across cosmic time.* MNRAS, 490, 3, 3196–3233.

Planck Collaboration, Aghanim N., Akrami Y. et al., 2018. *Planck 2018 results. VI. Cosmological parameters.*

Pontoppidan K., Blome C., Braun H. et al., 2022. *The JWST Early Release Observations.* arXiv e-prints, arXiv:2207.13067.

Powers D.M.W., 2011. *Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation.* Journal of Machine Learning Technology, 2, 1, 37–63.

Ravindranath S., Ferguson H.C., Conselice C. et al., 2004. *The Evolution of Disk Galaxies in the GOODS-South Field: Number Densities and Size Distribution.* The Astrophysical Journal, 604, 1, L9–L12.

Reback J., jbrockmendel, McKinney W. et al., 2022. *pandas-dev/pandas: Pandas 1.4.2.* Zenodo.

Reiman D.M. & Göhre B.E., 2019. *Deblending galaxy superpositions with branched generative adversarial networks.* Monthly Notices of the Royal Astronomical Society, 485, 2, 2617–2627.

Repp A. & Ebeling H., 2018. *Science from a glimpse: Hubble SNAPshot observations of massive galaxy clusters.* MNRAS, 479, 1, 844–864.

Rieke G.H., Wright G.S., Böker T. et al., 2015. *The Mid-Infrared Instrument for the James Webb Space Telescope, I: Introduction.* PASP, 127, 953, 584.

Rieke M., Kelly D., Horner S. et al., 2005. *The Near Infrared Camera (NIRCam) for the James Webb Space Telescope (JWST).* volume 207 of *American Astronomical Society Meeting Abstracts*, page 115.09.

Rieke M.J., Eisenstein D., Engelbracht C.W. et al., 2008. *Building for the James Webb Space Telescope: the Near-Infrared Camera.* volume 212 of *American Astronomical Society Meeting Abstracts*, page 79.01.

Rigby J., Perrin M., McElwain M. et al., 2022. *Characterization of JWST science performance from commissioning.* arXiv e-prints, arXiv:2207.05632.

Robertson B.E., Tacchella S., Johnson B.D. et al., 2022. *Morpheus Reveals Distant Disk Galaxy Morphologies with JWST: The First AI/ML Analysis of JWST Images.* arXiv e-prints, arXiv:2208.11456.

Rodriguez-Gomez V., Genel S., Vogelsberger M. et al., 2015. *The merger rate of galaxies in the Illustris simulation: A comparison with observations and semi-empirical models.* Monthly Notices of the Royal Astronomical Society, 449, 1, 49–64.

Rodriguez-Gomez V., Snyder G.F., Lotz J.M. et al., 2019. *The optical morphologies of galaxies in the IllustrisTNG simulation: a comparison to Pan-STARRS observations.* MNRAS, 483, 3, 4140–4159.

Rodriguez-Gomez V., Genel S., Fall S.M. et al., 2022. *Galactic angular momentum in the IllustrisTNG simulation - I. Connection to morphology, halo spin, and black hole mass.* MNRAS, 512, 4, 5978–5994.

Roper W.J., Lovell C.C., Vijayan A.P. et al., 2022. *First Light And Reionisation Epoch Simulations (FLARES) - IV. The size evolution of galaxies at* $z \geq 5$. MNRAS, 514, 2, 1921–1939.

Rose C., Kartaltepe J.S., Snyder G.F. et al., 2022. *Identifying Galaxy Mergers in Simulated CEERS NIRCam Images using Random Forests.* arXiv e-prints, arXiv:2208.11164.

Rosse T.E.O., 1850. *Observations on the Nebulae.* Philosophical Transactions of the Royal Society of London Series I, 140, 499–514.

Ruder S., 2016. *An overview of gradient descent optimization algorithms.* arXiv e-prints, arXiv:1609.04747.

Rumelhart D.E., Hinton G.E. & Williams R.J., 1986. *Learning representations by back-propagating errors.* Nature, 323, 6088, 533–536.

Salmon B., Coe D., Bradley L. et al., 2020. *RELICS: The Reionization Lensing Cluster Survey and the Brightest High-z Galaxies.* ApJ, 889, 2, 189.

Sánchez S.F., Kennicutt R.C., Gil de Paz A. et al., 2012. *CALIFA, the Calar Alto Legacy Integral Field Area survey. I. Survey presentation.* A&A, 538, A8.

Sandage A. & Tammann G.A., 1981. *A Revised Shapley-Ames Catalog of Bright Galaxies.*

Sandage A., 2005. *The Classification of Galaxies: Early History and Ongoing Developments.* ARA&A, 43, 1, 581–624.

Sazonova E., Alatalo K., Lotz J. et al., 2020. *The Morphology-Density Relationship in* $1 < z < 2$ *Clusters.* ApJ, 899, 1, 85.

Sazonova E., Alatalo K., Rowlands K. et al., 2021. *Are All Post-starbursts Mergers? HST Reveals Hidden Disturbances in the Majority of PSBs.* ApJ, 919, 2, 134.

Schade D., Lilly S.J., Crampton D. et al., 1995. *Canada-France Redshift Survey: Hubble Space Telescope Imaging of High-Redshift Field Galaxies.* ApJ, 451, L1.

Schawinski K., Urry C.M., Simmons B.D. et al., 2014. *The green valley is a red herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies.* MNRAS, 440, 1, 889–907.

Schaye J., Crain R.A., Bower R.G. et al., 2015. *The EAGLE project: Simulating the evolution and assembly of galaxies and their environments.* Monthly Notices of the Royal Astronomical Society, 446, 1, 521–554.

Schreiber C., Pannella M., Elbaz D. et al., 2015. *The Herschel view of the dominant mode of galaxy growth from* $z = 4$ *to the present day.* A&A, 575, A74.

Sérsic J.L., 1963. *Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy.* Boletin de la Asociacion Argentina de Astronomia, 6, 0.

Shibuya T., Miura N., Iwadate K. et al., 2022. *Galaxy morphologies revealed with Subaru HSC and super-resolution techniques. I. Major merger fractions of* L_{UV} *3-15* L^*_{UV} *dropout galaxies at z 4-7.* PASJ, 74, 1, 73–91.

Simons R.C., Kassin S.A., Snyder G.F. et al., 2019. *Distinguishing Mergers and Disks in High-redshift Observations of Galaxy Kinematics.* The Astrophysical Journal, 874, 1, 59.

Snoek J. & Larochelle H., 2017. *Practical Bayesian Optimization of Machine Learning Algorithms.* The Lancet Public Health, 2, 12, e540.

Snyder G.F., Torrey P., Lotz J.M. et al., 2015. *Galaxy morphology and star formation in the Illustris Simulation at* $z = 0$. Monthly Notices of the Royal Astronomical Society, 454, 2, 1886–1908.

Snyder G.F., Lotz J.M., Rodriguez-Gomez V. et al., 2017. *Massive close pairs measure rapid galaxy assembly in mergers at high redshift.* Monthly Notices of the Royal Astronomical Society, 468, 1, 207–216.

Snyder G.F., Rodriguez-Gomez V., Lotz J.M. et al., 2019. *Automated distant galaxy merger classifications from Space Telescope images using the Illustris simulation.* Monthly Notices of the Royal Astronomical Society, 486, 3, 3702–3720.

Sparre M. & Springel V., 2017. *The unorthodox evolution of major merger remnants into star-forming spiral galaxies.* MNRAS, 470, 4, 3946–3958.

Springel V., White S.D.M., Tormen G. et al., 2001. *Populating a cluster of galaxies - I. Results at [formmu2]z=0.* MNRAS, 328, 3, 726–750.

Springel V., Pakmor R., Pillepich A. et al., 2018. *First results from the IllustrisTNG simulations: matter and galaxy clustering.* MNRAS, 475, 1, 676–698.

Stefanon M., Yan H., Mobasher B. et al., 2017. *CANDELS Multi-wavelength Catalogs: Source Identification and Photometry in the CANDELS Extended Groth Strip.* ApJS, 229, 2, 32.

Storrie-Lombardi M.C., Lahav O., Sodre L. J. et al., 1992. *Morphological Classification of Galaxies by Artificial Neural Networks.* MNRAS, 259, 8P.

Strateva I., Ivezić Ž., Knapp G.R. et al., 2001. *Color Separation of Galaxy Types in the Sloan Digital Sky Survey Imaging Data.* AJ, 122, 4, 1861–1874.

Tan M. & Le Q.V., 2019. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.* arXiv e-prints, arXiv:1905.11946.

Taylor E.N., Hopkins A.M., Baldry I.K. et al., 2015. *Galaxy And Mass Assembly (GAMA): deconstructing bimodality - I. Red ones and blue ones.* Monthly Notices of the Royal Astronomical Society, 446, 2, 2144–2185.

The Dark Energy Survey Collaboration, 2005. *The Dark Energy Survey.* arXiv e-prints, astro-ph/0510346.

The GPyOpt A., 2016. *GPyOpt: A Bayesian Optimization framework in python.* \url{http://github.com/SheffieldML/GPyOpt}.

Thob A.C.R., Crain R.A., McCarthy I.G. et al., 2019. *The relationship between the morphology and kinematics of galaxies and its dependence on dark matter halo structure in EAGLE.* MNRAS, 485, 1, 972–987.

Thorp M.D., Bluck A.F.L., Ellison S.L. et al., 2021. *Towards robust determination of non-parametric morphologies in marginal astronomical data: resolving uncertainties with cosmological hydrodynamical simulations.* MNRAS, 507, 1, 886–903.

Tissera P.B., White S.D.M. & Scannapieco C., 2012. *Chemical signatures of formation processes in the stellar populations of simulated galaxies.* MNRAS, 420, 1, 255–270.

Tohill C., Ferreira L., Conselice C.J. et al., 2021. *Quantifying Non-parametric Structure of High-redshift Galaxies with Deep Learning.* ApJ, 916, 1, 4.

Toomre A. & Toomre J., 1972. *Galactic Bridges and Tails.* The Astrophysical Journal, 178, 623.

Trayford J.W., Camps P., Theuns T. et al., 2017. *Optical colours and spectral indices of $z = 0.1$ EAGLE galaxies with the 3D dust radiative transfer code SKIRT.* Monthly Notices of the Royal Astronomical Society, 470, 1, 771–799.

Trayford J.W., Frenk C.S., Theuns T. et al., 2019. *The star formation rate and stellar content contributions of morphological components in the EAGLE simulations.* MNRAS, 483, 1, 744–766.

Tully R.B., de Marseille O. & Fisher J.R., 1975. *A New Method of Determining Distances to Galaxies.* volume 7, page 426.

van den Bergh S., 1960. *A Preliminary Luminosity Clssification of Late-Type Galaxies.* ApJ, 131, 215.

van den Bergh S., 1976. *A new classification system for galaxies.* ApJ, 206, 883–887.

van den Bergh S., Abraham R.G., Ellis R.S. et al., 1996. *A Morphological Catalog of Galaxies in the Hubble deep Field.* AJ, 112, 359.

Vogelsberger M., Genel S., Springel V. et al., 2014. *Introducing the illustris project: Simulating the coevolution of dark and visible matter in the universe.* Monthly Notices of the Royal Astronomical Society, 444, 2, 1518–1547.

Vogelsberger M., Nelson D., Pillepich A. et al., 2020. *High-redshift JWST predictions from IllustrisTNG: dust modelling and galaxy luminosity functions.* Monthly Notices of the Royal Astronomical Society, 492, 4, 5167–5201.

Walmsley M., Smith L., Lintott C. et al., 2020. *Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning.* MNRAS, 491, 2, 1554–1574.

Walmsley M., Lintott C., Géron T. et al., 2022. *Galaxy Zoo DECaLS: Detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies.* MNRAS, 509, 3, 3966–3988.

Wang L., Pearson W.J. & Rodriguez-Gomez V., 2020. *Towards a consistent framework of comparing galaxy mergers in observations and simulations.* A&A, 644, A87.

White S.D.M. & Rees M.J., 1978. *Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering.* MNRAS, 183, 341–358.

White S.D.M. & Frenk C.S., 1991. *Galaxy Formation through Hierarchical Clustering.* ApJ, 379, 52.

Whitney A., Conselice C.J., Bhatawdekar R. et al., 2019. *Unbiased Differential Size Evolution and the Inside-out Growth of Galaxies in the Deep CANDELS GOODS Fields at $1 \leq z \leq 7$.* ApJ, 887, 2, 113.

Whitney A., Ferreira L., Conselice C.J. et al., 2021. *Galaxy Evolution in All Five CANDELS Fields and IllustrisTNG: Morphological, Structural, and the Major Merger Evolution to z 3.* ApJ, 919, 2, 139.

Williams R.E., Blacker B., Dickinson M. et al., 1996. *The Hubble Deep Field: Observations, Data Reduction, and Galaxy Photometry.* \aj, 112, 1335.

Windhorst R.A., Taylor V.A., Jansen R.A. et al., 2002. *A Hubble Space Telescope Survey of the Mid-Ultraviolet Morphology of Nearby Galaxies.* ApJS, 143, 1, 113–158.

Wuyts S., Förster Schreiber N.M., Lutz D. et al., 2011. *On Star Formation Rates and Star Formation Histories of Galaxies Out to z ˜3.* ApJ, 738, 1, 106.

Yan H., Ma Z., Ling C. et al., 2022. *First Batch of Candidate Galaxies at Redshifts 11 to 20 Revealed by the James Webb Space Telescope Early Release Observations.* arXiv e-prints, arXiv:2207.11558.

York D.G., Adelman J., Anderson John E. J. et al., 2000. *The Sloan Digital Sky Survey: Technical Summary.* AJ, 120, 3, 1579–1587.

Zana T., Lupi A., Bonetti M. et al., 2022. *Morphological decomposition of TNG50 galaxies: methodology and catalogue.* MNRAS, 515, 1, 1524–1543.

Zanisi L., Huertas-Company M., Lanusse F. et al., 2021. *A deep learning approach to test the small-scale galaxy morphology and its relationship with star formation activity in hydrodynamical simulations.* MNRAS, 501, 3, 4359–4382.

Zhang H., Primack J.R., Faber S.M. et al., 2019. *The evolution of galaxy shapes in CANDELS: from prolate to discy.* Monthly Notices of the Royal Astronomical Society, 484, 4, 5170–5191.

Zhou P., Feng J., Ma C. et al., 2020. *Towards Theoretically Understanding Why SGD Generalizes Better Than ADAM in Deep Learning.* arXiv e-prints, arXiv:2010.05627.