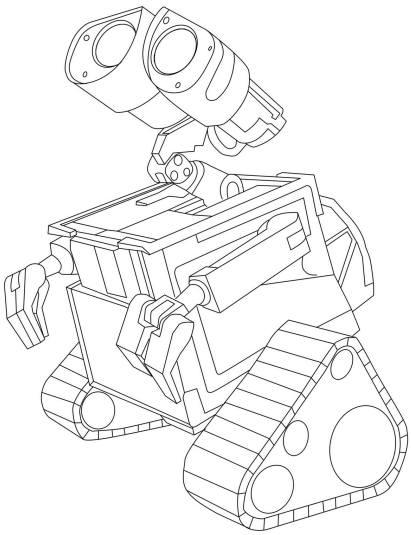
University of Nottingham
UK | CHINA | MALAYSIA

Exploring the Morphologies of High-Redshift Galaxies with Deep Learning

Clár-Bríd Tohill



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

May 2024

"Space is big. You just won't believe how vastly, hugely, mind-bogglingly big it is. I mean, you may think it's a long way down the road to the chemist's, but that's just peanuts to space."...

– Douglas Adams (1979), 'The Hitchhiker's Guide to the Galaxy'

Supervisors:	Dr Steven Bamford	
	Prof. Christopher Conselice	
Examiners:	Jim Geach	(University of Hertfordshire)
	Helen Russell	(University of Nottingham)
Submitted:	24 May 2024	
Examined:	2 July 2024	
Final version:	TBD	

Contents

Abstract

This thesis explores different machine learning techniques for the study of galaxy morphology, morphological classification, and the morphological evolution of galaxies. We utilise data from all of the CANDELS fields imaged with HST, as well as data from the CEERS program imaged with JWST.

In Chapter 2 we train convolutional neural networks to predict the non-parametric, concentration and asymmetry measurements from individual images of $\sim$150,000 galaxies at $0 < z < 7$ in the CANDELS fields. We apply a Bayesian hyperparameter optimisation to select suitable network architectures for this problem. Our resulting networks accurately reproduce measurements compared with standard algorithms. Furthermore, using simulated images, we show that our networks are more stable than the standard algorithms at low signal-to-noise. While both approaches suffer from similar systematic biases with redshift, these remain small out to $z \sim 7$. Once trained, measurements with our networks are $> 10^3$ times faster than previous methods. Our approach is thus able to reproduce standard measures of non-parametric morphologies and shows the potential of employing neural networks to provide superior results in substantially less time. This will be vital for making best use of the large and complex datasets provided by upcoming galaxy surveys, such as Euclid and Rubin-LSST.

In Chapter 3, we employ variational auto-encoders to perform feature extraction on galaxies at $z > 2$ using JWST/NIRCam data. Our sample comprises 6869 galaxies at $z > 2$, including 255 galaxies with $z > 5$ (when the Universe was $\sim 9\%$ of its current age), which have been detected in both the CANDELS/HST fields and CEERS/JWST, ensuring reliable measurements of redshift, mass, and star formation rates. To address potential biases, we eliminate background sources within our galaxy images, allowing our model to focus on the target galaxy. In order to avoid complicating the learned feature space, we eliminate galaxy orientation prior to encoding the galaxy features. We also normalise the apparent size of our sources prior to feature extraction, thereby constructing a physically meaningful feature space. By clustering the resulting feature space, we identify 11 distinct morphological classes that exhibit clear separation in various structural parameters, such as CAS-M20, Sérsic indices, specific star formation rates, and axis ratios. We observe a decline in the presence of spheroidal-type galaxies with increasing redshift, indicating a dominance of disk-like galaxies in the early universe. We demonstrate that conventional visual classification systems are inadequate for high-redshift morphology classification and advocate the need for a more detailed and refined classification scheme. Leveraging machine-extracted features, we propose a solution to this challenge and illustrate how our extracted clusters align with measured

parameters, offering greater physical relevance compared to traditional methods.

In Chapter 4, we expand upon the work carried out in Chapter 3 by incorporating the redshift information of our galaxy sample into our network. By conditioning our network on the redshift of our galaxy sample, we allow the network to encode a more efficient representation of our dataset, whilst simultaneously creating a tool to explore morphological variations with redshift. We introduce a novel technique of β-decay to our network, which forces the network to make good use of the redshift condition, whilst offering a method to determine the dominant features within a dataset. This work is on-going and we conclude with the next steps that this project will entail.

We finish with a brief overview of each research project, along with how the work in this thesis can be beneficial to future 'Big Data' surveys. We detail some improvements that can be made to this and other works, to further the study of galaxy morphology.

Acknowledgements

There are a great many people I want to thank for getting me to where I am today. I hope they know I am eternally grateful, and thankful, for the friendship, support, and memories, they have provided me with throughout this journey.

Firstly I would like to thank Steven and Chris for their continuing support and supervision during my PhD. I would not be the researcher I am without your guidance, and encouragement. During both Covid and my time working abroad, you made sure that I never lost sight of the goal, and were always supportive of my decisions, and for this I am grateful.

I would like to thank everyone in CAPT for creating an amazing environment to work in everyday. I would especially like to thank all of the support staff who have been incredible during the past 5 years. Phil Parry for solving every technical issue known to man in 5 minutes, Margaret for our morning chats that made sure I started every day off in good spirits, and Ella Batchelor for helping me get to some amazing conferences!

The Conselice group, especially Amy, Sunny, Leo, and Alex, who all made me feel so welcome when I joined the department. I would like to extend an extra thank you to Leo, you always went above and beyond to help me when I joined the group and I have enjoyed every one of our collaborations together!

To the Inflativerse crew: Roan, Amy, Tom, Kellie, Lizzie, Jimi, Jen, Fiona and Emma. You are all literal superstars and I want to thank you for being part of one of the best experiences of my PhD. Outreach has kept me going even when I wanted to give up, by reminding me what I love about astronomy, and seeing the good it can do.

Thanks to the CC (and regular crossword) gang: Amy, Roan, Kellie, Mikey, Joe, and Lauren, for helping me to _____ (I ran spectator puzzles stall).

Thank you to the wider gals community, you are all amazing and I am so glad to have you all in my life. I look forward to many more BBs with the Barbies: Kellie, Karel, Lizzie, Molly, Sophie, Jen, Paris, Lauryn, Isabel, Kira and Rowann.

My thesis partner in crime Dan, for all the coffees at C&C (and Richard for the weekly caffeine boosts) that have helped make this thesis possible, and the week of celebrations that is to come!

The amazing residents of Thyra Grove: Helena, Sam, Finn and Skye for the keeping me sane during lockdown, and for all the crab raves along the way!

I want to thank the DA: Mick, Roan and Baresca, for hosting many a party, and being so welcoming when I first came to Nottingham. For all the regular DA attendees: Me Ol Mucka Sukhi, Tomáš, Simon, Matt, Charutha, Swagat, Steph, Brad, Jessica, and Tom C.

All of the INT squad who made my time in La Palma all the more memorable: Rosa, Dimitris, George, Paige, Pablito, Connor, and Monu. Ya te estoy extrañando!
Thank you to my family, Mum, Dad, Niamh, my late Granda, Finn, and the best boi in the world, Bob. Thank you for your un-faltering support, even if you mistakenly tell people that I am an astrologist mum. You have always supported my decisions, even when it was difficult for you, and I would not be where I am without you.
Finally I'd like to thank Mick for being my biggest supporter, you are a constant source of joy and silliness in my life, and when I'm packing for another planet, you'll be top of the list.

Published works

The majority of the material presented in this thesis has already been published in — or has been submitted to — a journal as the following works.

i **Tohill C.**, Ferreira, L., Conselice, C. J., Bamford, S. P., Ferrari, F., 2021, ApJ, 916: *'Quantifying Non-parametric Structure of High-redshift Galaxies with Deep Learning'*.

ii **Tohill C.**, Bamford, S. P., Conselice, C. J., 2023, MemSAIt, : *'Exploring the Morphologies of High Redshift Galaxies with Machine Learning'*.

iii **Tohill C.**, Bamford, S. P., Conselice, C. J., Ferreira, L., Harvey, T., Adams, N., Austin, D., 2024, ApJ, 962: *'A Robust Study of High-Redshift Galaxies: Unsupervised Machine Learning for Characterising morphology with JWST up to $z \sim 8$'*.

Chapter 2 is based on material published in Paper i, Chapter 3 is based on material published in Paper ii and iii.
The vast majority of work presented here was carried out by the author, with advice from the paper authors above. Where the work contains the product of larger collaborations, this is mentioned in the relevant chapter.

Chapter 1

Introduction

1.1 A galaxy by any other name...

Throughout recorded history, humans have looked to the skies and attempted to make sense of the cosmos. Even before we even knew what galaxies were, we had begun to classify them. The earliest galaxy classification dates back to 964 AD, when Persian astronomer Abd al-Rahman al-Sufi described the Andromeda galaxy as a 'nebulous smear' (Hafez, 2010). Despite the limited understanding of the nature of these objects, they were clearly distinct from the point source stars that surrounded them. The telescope's invention in the early 1600s allowed more detailed study, revealing previously unseen 'nebulous' objects. French astronomer Charles Messier's 1781 catalogue marked an early effort, listing 110 such objects (Messier, 1781). Not long after, William and Caroline Herschel's systematic survey in the late 18th century identified over 5000 nebulous objects (Herschel, 1786), forming the basis for the New General Catalogue of Nebulae and Clusters of Stars (NGC) (Dreyer, 1888), which is still widely used by astronomers today.

As the light collecting power of telescopes continued to improve, astronomers were able to chip away at these illusive cosmic bodies. In the mid-1800s, Lord Rosse's observations revealed a distinctive spiral structure within some of these formerly obscure 'nebulous' objects, hinting at uncharted complexities within (Rosse, 1850).

His sketch of the Whirlpool Galaxy can be seen in Fig.1.1. This, amongst other sketches, showed that these nebulae were distinct from each other, each possessing their own structure and features.

As astronomy continued to advance with better telescopes and instrumentation, photography became the most up to date method for studying these objects. The most

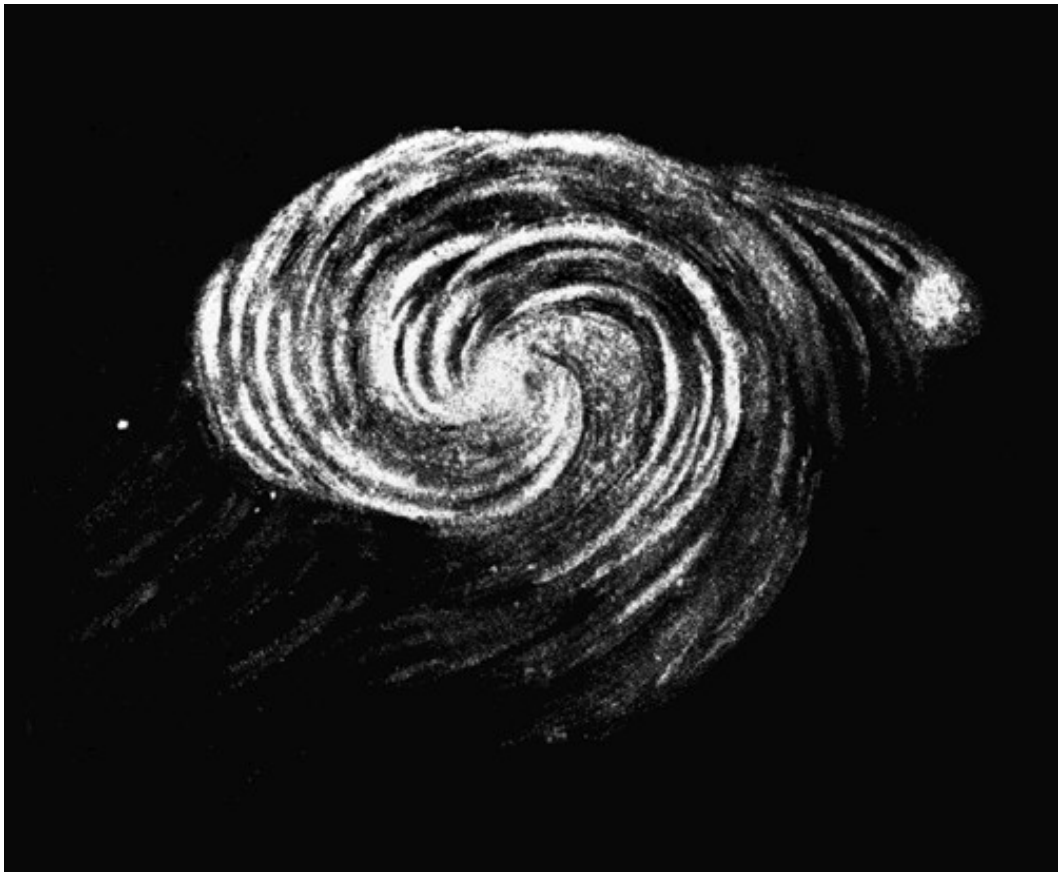


Figure 1.1: Lord Rosse's sketch of the Whirlpool Galaxy M51 which he observed from his observatory at Birr Castle, Ireland in 1845.

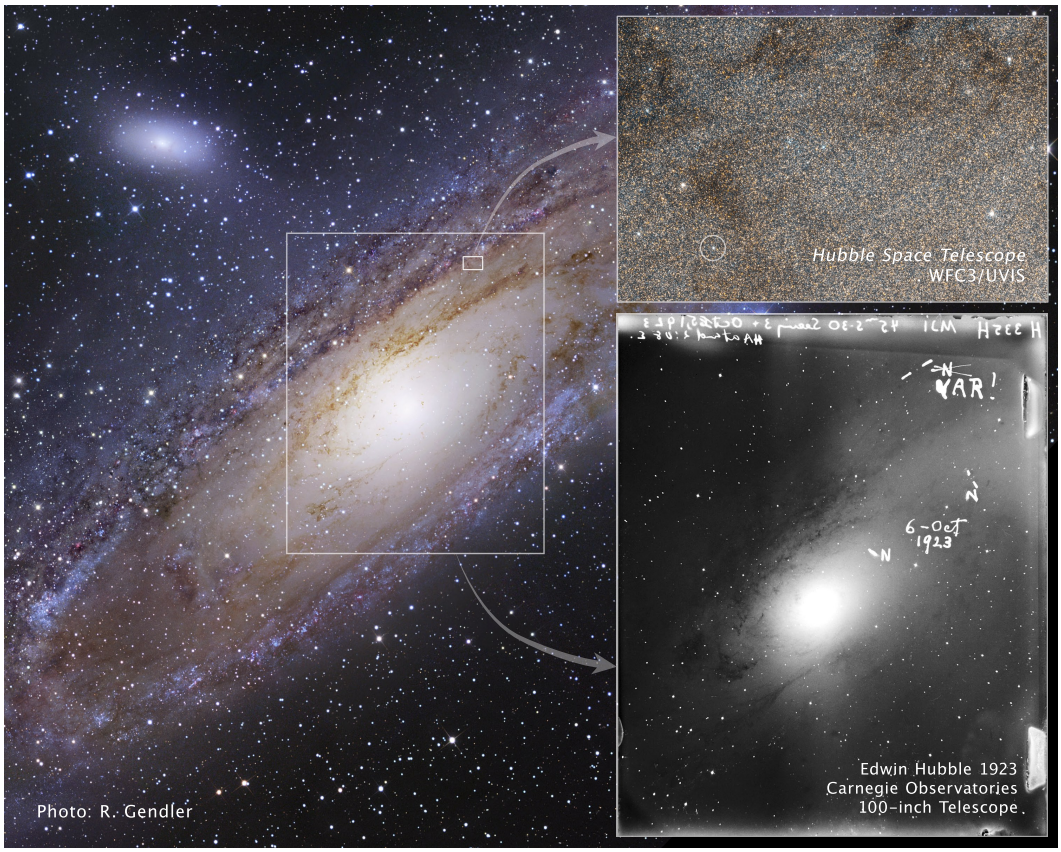


Figure 1.2: Taken 90 years apart, the Andromeda galaxy as imaged with the Hubble Space Telescope's Wide Field Camera 3, compared to the photographic plate of the Andromeda galaxy upon which Edwin Hubble first noted the presence of a Cepheid variable star within the galaxy. Credit NASA, ESA, and the Hubble Heritage Team (STScI/AURA); Illustration: NASA, ESA, and Z. Levay (STScI).

ground-breaking revelation came in the form of Edwin Hubble's work on the Andromeda nebula. Within the Andromeda nebula he noted stars, specifically one that varied in brightness over time, a Cepheid Variable star (see Fig.1.2). Cepheid Variable stars, discovered by Henrietta Leavitt (Leavitt & Pickering, 1912), are known to vary in brightness over distinct periods of time. This way they can be used as cosmic distance rulers (Hertzsprung, 1926). Using Cepheid Variable stars within the Andromeda nebula, he concluded its distance to be 900,000 light years from us, finally naming it a galaxy in its own right and forever changing our view of the Universe, and our place within it (Hubble, 1929).

1.2 Galaxy Formation and Evolution Models

As our understanding of the universe has deepened, we have begun to model how it both formed and evolved, which provides the basis for galaxy formation and evolution studies. Our current understanding of our Universe follows what we call the Λ-CDM cosmological model. This model describes an expanding universe that originated from a single point of origin, known as the Big Bang. Our best estimates, from both observations and these models, put the Universe at around 14 billion years old. Comprised of ~ 30% matter component, with the remainder made up of an unknown component known as dark energy. Of the 30% matter component, only 1/6 of this is baryonic matter, the stars, gas, and dust that we observe in the form of stars and galaxies etc. The other ~ 25% is made up of dark matter, an invisible substance that only interacts via gravity.
While there is still some disagreement if this is the correct model for the formation of the Universe, it does agree with observational evidence such as the recession velocity of galaxies (Hubble, 1929), the cosmic microwave background (Alpher & Herman, 1948; Penzias & Wilson, 1965), and elemental abundances observed in the universe (Font et al., 2006).
In the Λ-CDM model, structure is generally thought to arise through a growth of density perturbations which originate in the early universe. These density perturbations are then thought to grow linearly until they reach a critical density and collapse to form gravitationally bound dark matter halos. These halos continue to grow in mass (and size), either through accretion of material from their immediate neighborhood, or by merging with other halos. As the universe expanded, it cooled, allowing for baryonic matter to condense in these dark matter halos, creating the structure we observe in the universe. It is thought that galaxy formation occurs through the hierarchical clustering of these halos (White & Rees, 1978), leading to the large-scale structures we observe today.

1.2.1 Numerical simulations

While we cannot directly observe the beginnings of our universe, we can use known physics, along with the Λ-CDM model, to create numerical simulations capable of modelling how structure first formed in the universe (White & Frenk, 1991). They are divided into dark matter-only simulations, such as N-body simulations (Toomre & Toomre, 1972), and dark matter plus baryonic matter simulations, such as hydrodynamical simulations (Vogelsberger et al., 2014; Nelson et al., 2019). Studying baryonic

physics through hydrodynamical simulations, however, is computationally expensive compared with N-body simulations. An alternative approach is to model baryonic physics on top of N-body simulations through analytic models. This combination of numerical, dark matter-only simulations, and analytic models for the prescription of baryonic physics, is known as semi-analytic modelling (Baugh, 2006).
Through decades of work, researchers have refined these models, altering the background cosmology, in order for the simulations to reproduce the large-scale structure observed in the Universe. The key improvement to the models came in the form of dark energy, which was incorporated into the models to describe a universe expanding at an accelerated rate. Simplified prescriptions for unresolved processes such as star formation (Katz, 1992), and AGN activity (Sijacki et al., 2007), have also been developed in order to produce galaxy properties close to those observed (Huertas-Company et al., 2019). By studying how the properties of these galaxies vary with cosmic time we can gain a better insight into how galaxies evolve to that which we observe in the local universe. These simulations complement observational studies well, as they can be created as if observed with a specific instrument, allowing direct comparisons to be made between them and observational results.

1.2.2 Challenges in our understanding of galaxy evolution

Despite significant advancements, several key gaps remain in our understanding of galaxy evolution. For example, the impact of mergers on the physical processes of galaxies is not fully understood, as well as the role they play in the morphological transformation of galaxies. Additionally the mechanisms behind star formation enhancement and quenching within galaxies involve various factors such as feedback mechanisms, and environmental effects, however the physical processes behind these effects can vary between galaxies. Morphological classification of galaxies is a powerful tool to help address these gaps. By classifying galaxies based on their morphology, researchers have discovered that the dominant feedback processes depend on the morphological type, with star formation quenching being dominated by supernova feedback in dwarf and irregular galaxies (Dekel & Silk, 1986; Dashyan & Dubois, 2020), compared to AGN feedback in larger, more evolved galaxies, such as elliptical and disk galaxies (Binney & Tabor, 1995; Choi et al., 2015). Morphological classification of galaxies also allows researchers to investigate galaxies that have recently undergone a merger by selecting galaxies with disrupted morphologies, or tidal features (Conselice et al., 2003;

Lotz et al., 2008b). Studying the properties of these galaxies has revealed that mergers can cause an enhancement of star-formation (Duc et al., 1997; Sanders et al., 1988; Bell et al., 2006), as well as explain the morphological transformation of disk galaxies into ellipticals after a major merger event (Toomre & Toomre, 1972; Negroponte & White, 1983). These are just some examples of how morphological studies can be utilised to help advance our understanding of galaxy evolution. Morphological classifications are discussed in detail in §1.3.

1.2.3 Observational evidence

Galaxy evolution can be studied through observational evidence by tracking how physical properties of galaxies vary over cosmic time, such as star formation rates (Sandage, 1986; Madau & Dickinson, 2014), merger histories (Conselice, 2014), age of stellar populations (Haywood et al., 2013), molecular gas content (Bouché et al., 2010), amongst others. All of these affect the structure and morphology of a galaxy, which has in turn been shown to correlate with its internal kinematics. Spectroscopic surveys such as SAURON and MaNGA (de Zeeuw et al., 2002; Bundy et al., 2015) can probe the stellar velocity and velocity dispersion within galaxies, which have shown that elliptical galaxies are supported by high-velocity dispersion, while disk galaxies are supported by rotation (Cappellari et al., 2007, 2011).

Galaxy morphology has also been shown to depend on the environment in which they are situated. In dense environments, like galaxy groups or clusters, elliptical and lenticular type galaxies dominate the morphology, while spiral galaxies are uncommon. This dependence on environment is known as the Morphology-Density relation (Dressler, 1980), and is shown in Fig.1.3.

As the morphology of a galaxy holds information about its internal kinematics and physical properties, another method of studying galaxy evolution is to investigate how morphology varies over cosmic time. In order to study how morphology varies, we first need to have a good understanding of the range of morphologies we observe both in the local, and in the distant Universe. To do this, we use classification schemes to sort galaxies into different morphological types, which we can then compare to learn more about which physical properties dominate the morphological evolution of galaxies.

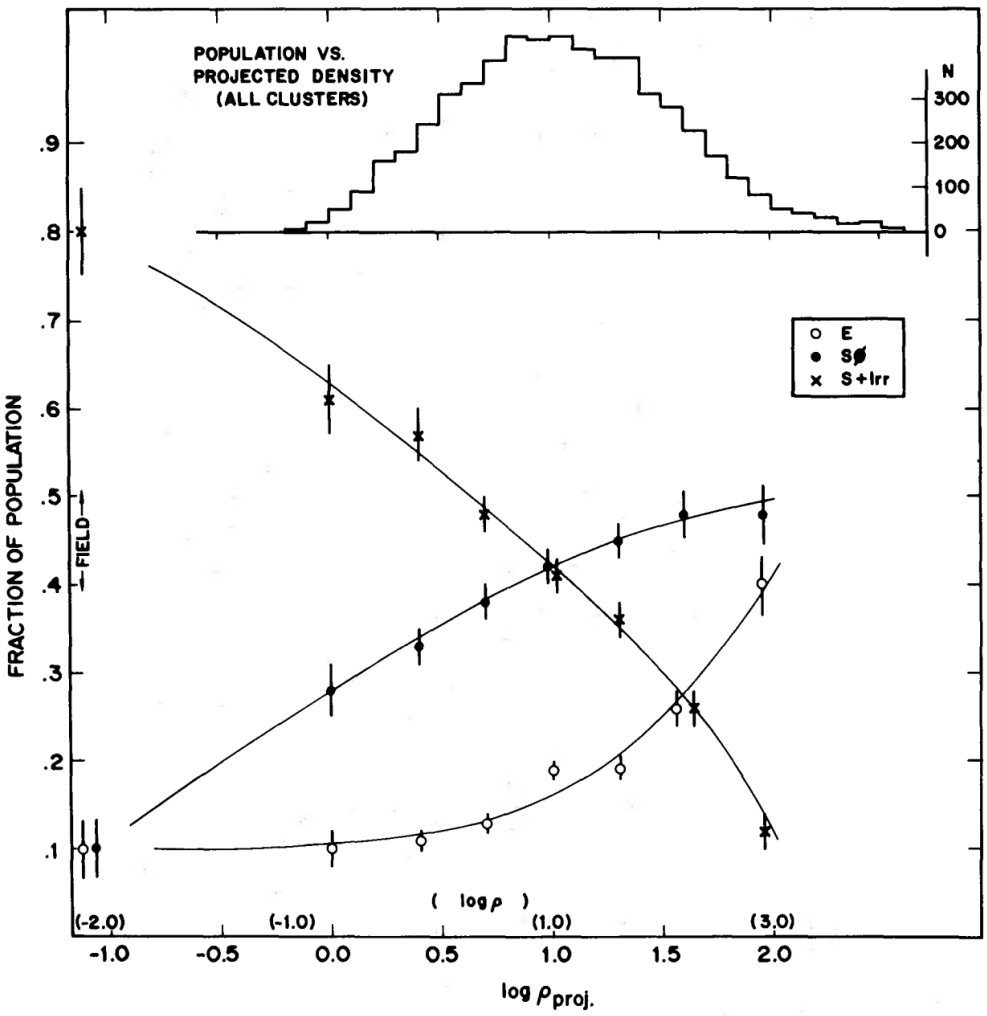


Figure 1.3: Fraction of Elliptical (E), S0, and Spiral and Irregular galaxies (S+Irr), as a function of the projected density in galaxies Mpc^{-2} (Credit: Dressler (1980)).

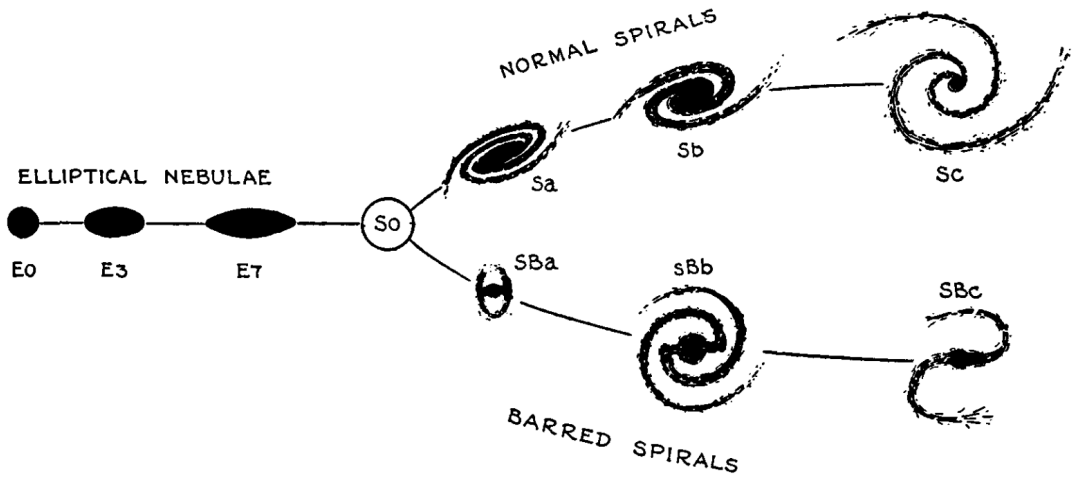


Figure 1.4: Schematic of Edwin Hubble's tuning fork of galaxy morphology. Galaxies are split into Elliptical and Spiral type with the Spiral galaxies further split based on the presence of a central bar (Credit: Hubble (1936)).

1.3 Visual classifications

While we know a lot more about what galaxies are today, the idea of classifying galaxy morphology by eye has not changed. From the blurry smudge on the sky to the high resolution images from space based telescopes like Hubble and JWST, we can now classify these objects in full detail. With this increase in resolution, morphological classification schemes have changed over time to better account for the wide variety of galaxy morphologies observed.

1.3.1 Low Redshift Morphology

Morphological classification of galaxies is still widely used to describe and study galaxies. The most commonly used classification system is that of Edwin Hubble's 'Galaxy Tuning Fork'. Devised in 1926 (Hubble, 1926), the tuning fork classification system splits nearby galaxies into 2 main groups, Elliptical galaxies, and Spiral galaxies. Spiral galaxies are further split into normal spirals and barred spirals. A schematic for this system can be seen in Fig.1.4.

This classification system has been shown to correspond to physical properties of the galaxies. Elliptical galaxies were found to be dominated by older stars, indicating that they were older and more evolved than the spiral type galaxies (Morgan, 1958). On

the other hand, spiral galaxies were found to possess much younger stellar populations, and in most cases are still undergoing continuous star formation (Kennicutt, 1989). This bimodality of morphology is typically referred to as 'early' (elliptical) and 'late' (spiral) type galaxies. It should be noted that while many people incorrectly assume Hubble meant this as a morphological evolution, it is instead in reference to the stellar populations each galaxy is made up of. 'Early' and 'late' type refer to the positions in the spectral sequence (Kirkpatrick et al., 1999), and hence are a spectral classification, not a physical stage of evolution (Baldry, 2008). Hubble also included another type of galaxy in this classification scheme which he denoted as S0 galaxies (see Fig.1.4). He believed there to be an intermediate population between the two main classes, Spirals and Ellipticals, which are now referred to as lenticular galaxies. This inclusion to the Hubble Tuning fork was expanded by Allan Sandage in his work Sandage (1961). There have since been many revisions and updates to the Hubble Tuning fork (see Morgan (1958); de Vaucouleurs (1963); Sandage (1986); Buta (1995)) which include even more morphological types such as ring, irregular and lens galaxies.

Whilst there have been many revisions to the Hubble classification system, the main framework has remained since its creation almost 100 years ago. The simplicity of this system, and the correlation with intrinsic physical parameters of each galaxy has proven invaluable and has consequently been utilised in many different morphological studies. One of the largest morphological studies that take inspiration from the Hubble sequence is Galaxy Zoo (Lintott et al., 2008). Galaxy Zoo is one of the largest citizen science projects in astronomy and asks participants to judge the morphology of a galaxy using a number of questions about its dominant structure. In total, this project amassed over 50 million classifications in the first year alone. These classifications proved to be as accurate as a professional astronomer, thus allowing astronomers to study a much larger sample of galaxies than originally thought possible. The accuracy of the classifications was found to decrease when moving to more distant galaxies however, steps were taken to try to remove these biases from the classifications (Willett et al., 2017).

1.3.2 Galaxy Morphology z > 1

The faintness and small angular size of galaxies at high redshift ($z \gtrsim 1$) makes them difficult to classify in the same manner as those nearby. Cosmological dimming causes more subtle features, such as spiral arms, to rapidly disappear with increasing redshift, leaving only the brightest galaxy components detectable (Barden et al., 2008). The

resolution is worsened when we add in the atmospheric distortions when observing these objects with ground based telescopes. To make observing these galaxies even more challenging, as they are moving away from us their Spectral Energy Distribution (SED) is shifted towards redder wavelengths as a function of $\lambda(1 + z)$. This causes the light from the galaxy to be shifted towards redder wavelengths to a point where the rest-frame UV is shifted into the optical (Blanton et al., 2003). This is an issue for morphological studies as the UV light is dominated by young stellar populations and hence, will alter the morphology from that observed in the optical.

Space-based telescopes like HST can combat the issue of angular resolution as there is no atmosphere, hence the distortions are removed. While the Earth's atmosphere is transparent in some regions of the NIR/FIR, the sky is still very bright and variable which limits our ability to measure low surface brightness features. By observing from space we also remove this sky background from being an issue, allowing for deeper, more resolved images. However, despite the advantages, space based telescopes like HST are still limited by their wavelength coverage, able to probe rest-frame optical up to a maximum of $z = 3$.

At early times, higher rates of star-formation and merging also increase the prevalence of more varied and irregular morphologies (Abraham et al., 1996a; Elmegreen et al., 2005a; Conselice & Arnold, 2009; Mortlock et al., 2013). Consequently, the Hubble sequence is of limited applicability at high redshift as most galaxies would end up in this irregular category. Work carried out in the past by Elmegreen et al. (2005b) classified high redshift galaxies in the Hubble Ultra Deep Field into 6 main groups; chain, clump cluster, double clump, tadpole, spiral and elliptical. These groups were determined by eye to match previous classifications by others (Cowie et al., 1995; van den Bergh et al., 1996). From these early results it was not clear if the intrinsic galaxy structure was very different, or if the different morphologies resulted from the above mentioned observation effects. To determine if this change in morphology was due to observation effects alone, Bunker et al. (2000) compared a variety of galaxy morphologies at the same rest-frame wavelength and found that morphological k-corrections are only important in a minority of cases, and that galaxies are intrinsically more peculiar at high redshift. From these works it is now clear that galaxies underwent a significant change in their morphology as part of their evolution and thus, to better understand galaxy evolution, we need a clearer understanding of high-redshift galaxy morphology.

With the launch of JWST in 2022 (Gardner et al., 2006; Menzel et al., 2023), we are now able to observe these distant galaxies in more detail. JWST is able to observe

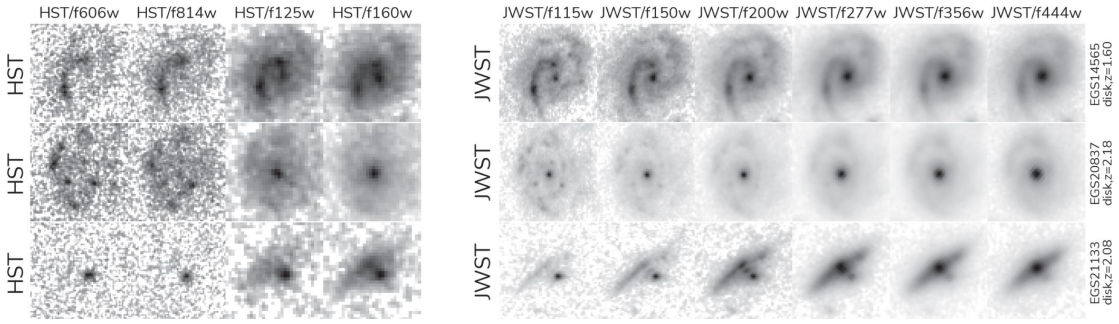


Figure 1.5: Comparison of a sample of galaxies observed in 4 filters with HST vs 6 filters with JWST (Modified from Ferreira et al. (2023)). Both the increase in resolution, and the ability to probe rest-frame optical allows for a much clearer view of these galaxies. The morphological classification from Ferreira et al. (2023) can be seen to the right of the galaxy images.

rest-frame optical up to $z \sim 10$, allowing us to explore the optical morphologies of these distant galaxies for the first time (Pontoppidan et al., 2022). The improvement in both resolution and wavelength coverage can be seen in Fig.1.5. This regime now requires its own, more suitable, morphological classification system. This issue is addressed in Chapter 3.

1.4 Galaxy structural measurements

A galaxy's morphology is a useful indicator of its assembly, interaction and star-formation history. Morphological studies have therefore proven invaluable for tracing the evolution of the galaxy population over cosmic time. However, in order to study galaxy populations in more detail than visual classifications, a more quantitative approach to studying galaxy morphology was needed.

1.4.1 Parametric measurements

One way in which we can quantify a galaxy's morphology is with parametric measurements. These are model-dependent measurements. In 1963 José Luis Sérsic proposed a function (Sérsic, 1963) which describes how the intensity of light measured in a galaxy varies from the center of the galaxy itself, given by:

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

where n is the Sérsic index that defines the slope of the distribution, b_n is a function of n, and is defined so that, I_e, corresponds to the intensity at the half-light radius R_e. This was a generalisation of De Vaucouleurs Law (de Vaucouleurs, 1948), where setting $n = 4$ corresponds to the light distribution of elliptical galaxies. This one parameter was found to be able to separate spiral and elliptical galaxies with reasonable accuracy, as the light within elliptical galaxies is more concentrated in the central region than spiral galaxies, leader to larger values of n. It is also possible to fit a Sérsic profile to the bulge and disk component of a galaxy separately, van der Kruit & Searle (1981) found that disks have exponential profiles once the bulge component was removed. This correlation of Sérsic index with general morphology allows it to be used as an automated method of classifying galaxies in the Hubble sequence.

As parametric measurements assume that the light distribution in a galaxy will follow a particular shape they do not provide any information about finer structure such as spiral arms or bars. This can lead to galaxies with very different morphologies having overlapping indices. Most fitting algorithms used today apply three component fits to galaxies to model the disk, bulge, and bar components separately to overcome this issue (Peng et al., 2002; Laurikainen et al., 2005; Weinzirl et al., 2009).

1.4.2 Non-Parametric measurements

While parametric measurements can be a powerful tool, they break down for more irregular, peculiar-type galaxies, as they typically assume a smooth light distribution. Non-parametric methods make weaker assumptions. They are therefore more applicable to the variety of galaxies seen in the more distant universe, such as those with 'clumpy' morphologies (Abraham et al. 1994; Noguchi 1998; Bershady et al. 2000). Motivated by these considerations, a number of authors, including Abraham et al. (1994, 1996b), Schade et al. (1995), and Conselice (1997) focused on two such non-parametric parameters, the concentration (C) and asymmetry (A) of a galaxy's light distribution. It has been shown that the concentration parameter correlates with the bulge-to-disk ratio (B/D) of a galaxy, while the asymmetry parameter is a good indicator of the merger history of the galaxy (Conselice 2003; Lotz et al. 2008a; Nevin et al. 2019). Using these parameters they were able to separate galaxies into their morphological type based on their position in this $C - A$ plane. Conselice (2003) expanded on this by introducing a third parameter, the smoothness (S) of a galaxy's light distribution, creating the CAS system which has become one of the most common non-parametric measures of galaxy

structure. This system has since been used in many investigations of galaxy structure across a wide range of redshifts (e.g. Yagi et al. 2006; Hoyos et al. 2012).

Another set of non-parametric measurements is the Gini-M20 system introduced by Lotz et al. (2004b). These statistical measures describe how the light from a galaxy is distributed in terms of pixel flux values. It was shown that utilising these along with the CAS system can allow for an even better separation in galaxy morphology, and in particular, is adept at locating merging galaxies due to the large variations in the light profile for most of these systems.

While non-parametric measurements are more suitable than parametric measurements for quantifying galaxy morphology in the high redshift universe they still have some issues. A lot of these measurements, like the Gini-M20 system and the smoothness parameter, require a sufficient number of resolution elements covering the galaxy which can be an issue for a lot of galaxies in the distant universe because of their small apparent size, due to both angular scaling and the fact that galaxies were physically smaller in the past (Whitney et al., 2019). These measurements, along with parametric measurements, also require many pre-processing steps, which can introduce biases, for example, when creating segmentation masks, fainter regions of a galaxy can be removed, which will impact the final measurement. These pre-processing steps are also time-consuming, which is issue with the amount of data we are now collecting every day. A potential solution to these issues, and a method that has become popular in astronomy recently, is to use machine learning (ML) to quantify galaxy morphology. This is explored in Chapter 2. Beforehand, an overview of applications of ML for galaxy morphology and structural parameters is detailed below.

1.5 Machine Learning for Galaxy Classification

The concept of machine learning has deep roots in astronomy, spanning from its initial, simpler applications at the end of the last century to the present-day utilisation of large-scale, deep learning methodologies. While our focus in this work is solely on the applications to galaxy morphology and classification, it's worth noting that this is just one area of astronomy that has benefited from machine learning. We begin with an overview of the main ideas behind neural networks, and how they have been applied to different problems in astronomy.

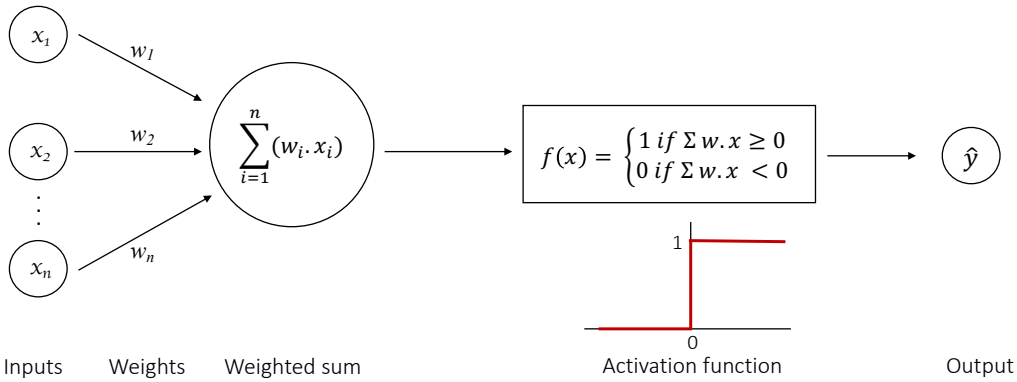


Figure 1.6: A single artificial neuron. Inputs (x_i) are multiplied by a corresponding weight (w_i), and the weighted sum is passed to the activation function ($f(x)$), which provides an output ($\hat{y}$).

1.5.1 Neural networks

Neural networks are models that were designed to recognise patterns and solve problems akin to how the human brain processes information. They are built up of artificial neurons, designed to process information similarly to neurons in the brain. Artificial neurons are mathematical functions that take an input (x_i), e.g a pixel value from an image, and multiply this input by a corresponding weight (w_i), to signify the importance of each input. These are then summed together to give a weighted sum of your input values. The artificial neuron then needs to make a decision on if it should 'fire' or activate based on the input. This is done using an activation function. An activation function can be something as simple as a threshold, e.g only 'fires' when the input is above a certain value, as shown in Fig:1.6, or more flexible like a sigmoid function. The output from the activation function is then used as your prediction ($\hat{y}$) in your model. In order to train your model, you need to utilise a loss function $\mathcal{L}(y, \hat{y})$, to compare how your models prediction compares to the ground truth you are trying to predict (y). The loss function is any differentiable function, common examples are the mean squared error, or cross-entropy/log loss. The main idea behind a loss function is that you want to minimise it, so that your model can provide more accurate predictions. This is done by adjusting the weights within your neuron, and recalculating the loss. This process is repeated multiple times as your model trains, improving your models predictions.

A typical neural network is made up of lots of these artificial neurons all grouped together. Instead of one layer of neurons outputting your model prediction, the output of

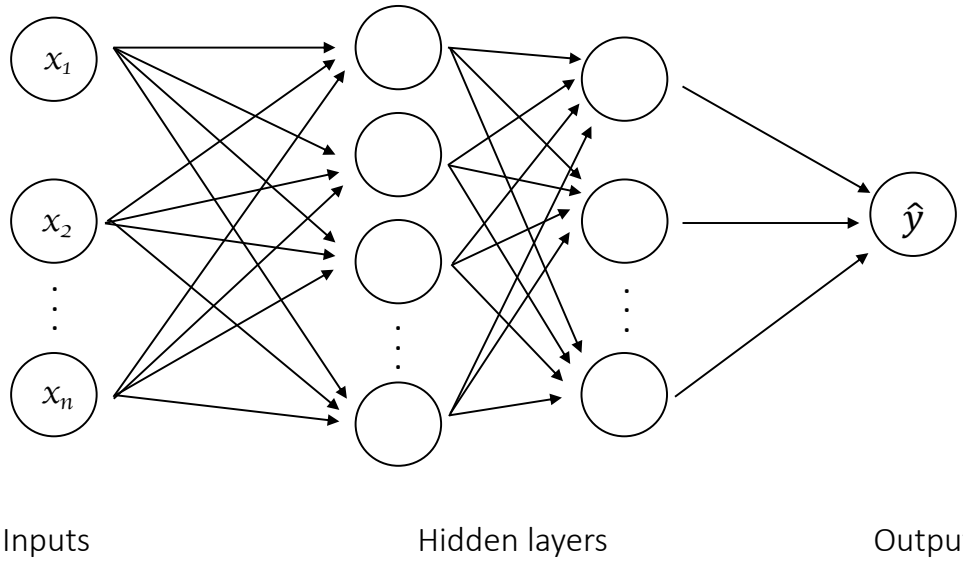


Figure 1.7: Architecture of a multi-layer perceptron (also known as an artificial neural network). In this model there is an input layer, two hidden layers, and an output layer.

a neuron is passed to another layer (or multiple layers) of neurons, often referred to as a 'hidden layer', before outputting it's prediction. This type of model is called a multi-layer perceptron (MLP), or an artificial neural network (ANN). A MLP is trained similarly to a single artificial neuron however, in order to determine which weights are updated within the model, a process known as backpropagation is utilised. Backpropagation works by stepping back through each layer of the network, from the output layer to the input layer, to determine how much each neuron contributed to the overall loss. It does this by calculating the gradient of the loss with respect to the weights, which shows the impact of each weight on the total error. The weights are then updated via gradient descent, which adjusts each weight in a direction that reduces the overall loss. This process repeats, reducing the loss over time until it converges on a minimum value. These networks can be used in both classification and regression problems, and we discuss their uses within astronomy in the next section.

1.5.2 Applications in astronomy

Machine Learning (ML) in astronomy started to grow rapidly in the 1990s with the arrival of large sky surveys, such as the Second Palomar Observatory Sky Survey (Poss-II, Reid et al. 1991), which saw the demand for automated tools to handle the vast amount of data

now available. Classification tasks, such as star-galaxy separation, and morphological classification of galaxies, that were previously analysed by humans, were now completed using ML algorithms such as, decision-trees (Weir et al., 1995), and artificial neural networks (ANNs) (Storrie-Lombardi et al., 1992; Naim et al., 1995; Odewahn et al., 2004). ANNs have also been applied to regression problems such as estimating physical parameters from stellar spectra (Bailer-Jones et al., 1997), and photometric redshift estimates (Collister & Lahav, 2004). These are examples of supervised ML tasks, in that, they require a labelled dataset in order to train the networks.

While ANNs have proven to be a valuable tool in astronomy, they fall down when it comes to image classification problems as they are not able to exploit the spatial structure of the data. Each pixel in an image is treated as an independent vector, regardless of its position within an image. A more suitable network for image classification problems are convolutional neural networks (CNNs).

CNNs are made up of convolutional layers, which are able to extract features from images by applying multiple filters (convolutional kernels) to the image. Individually, these filters can detect simple features. However, successive layers act hierarchically, identifying increasingly complex patterns. One major advantage of CNNs for image classification problems is the fact that they are able to exploit the spatial structure of the data which in turn reduces the number of parameters, and allows the recognition of location invariant features. CNN's were first popularised for image recognition/classification problems with the creation of LeNet-5 (Lecun et al., 1998), a network trained to classify handwritten digits. From this, CNNs have been applied in a range of fields, addressing a number of different problems and are becoming increasingly popular in astronomy. CNNs were first utilised for galaxy classification by Dieleman et al. (2015) using data from the Galaxy Zoo project (Willett et al., 2013). The advent of CNNs provided a technique for efficiently extracting high-quality information directly from images. CNNs have since seen wide usage in extra-galactic astronomy, including morphological classification (e.g., Domínguez Sánchez et al. 2018; Cheng et al. 2020b; Barchi et al. 2020), performing photometry (Tuccillo et al. 2018; Boucaud et al. 2020), and estimating merger rates (Ferreira et al. 2020). We explore and discuss the use of CNNs for estimating galaxy structural parameters in Chapter 2.

1.5.3 Unsupervised Machine Learning

While supervised ML has had many successes in astronomy, it is not suitable for future 'Big Data' surveys, as these will result in a large amount of unlabelled data that needs to be analysed efficiently. A solution to this is to turn to unsupervised ML (UML), which as the name may suggest, does not require labels to train. These types of networks are suitable for problems where the truth is unknown, there are too much data to label, or there are known biases with the labels that we want to avoid. Due to the unsupervised nature of these networks, they are also capable of finding previously unknown patterns within data, which is not possible with supervised ML.

UML can be split into three main tasks, clustering techniques, association rules, and dimensionality reduction. As we utilise both dimensionality reduction, and clustering techniques in our work we will discuss these in more detail below.

Dimensionality reduction is a powerful tool that allows important information to be extracted from a dataset whilst minimising the information lost in the process. This can allow data to be more easily interpreted, and can indicate which features in a dataset are the most important. This process can also be referred to as a feature extraction process. Typical UML networks used for this process are autoencoders (AEs) (including variational autoencoders (VAEs, Kingma & Welling 2013), which will be discussed in detail in Chapter 5) and principal component analysis (PCA, Jolliffe 1986).

Clustering UML methods aim to group data points together based on their similarity. This similarity is normally calculated based on a distance metric such as the Euclidean distance, and results in the dataset being split into different clusters or groups that optimise the separation in this space. Certain clustering methods such as K-means clustering (Likas et al., 2003), require some level of human input such as the number of clusters which can be argued as a partly supervised task, however there are other algorithms such as Gaussian mixture models (GMMs) and hierarchical clustering (HC, Johnson 1967) that avoid this.

When applying UML to galaxy classification problems, features are extracted from the galaxy images via a dimensionality reduction process, and clustering algorithms can be used to group galaxies with similar features together, resulting in an unsupervised classification scheme. Previous works have utilised these types of networks to investigate galaxy morphology at both low-redshift ($z < 1$) (Martin et al., 2020; Spindler et al., 2021; Cheng et al., 2021) and high redshift ($z < 4$) (Hocking et al., 2018) for classification problems. While promising, many of these works find hundreds of different morphological classes, which becomes challenging when trying to explore evolutionary

trends.

Another challenge unveiled through these works is the impact of geometrical effects within the data, such as the apparent size of the galaxy, its position angle, and if there are any background or foreground objects within the image data. These effects are imposed by us on the data from our viewing position on Earth and do not have any physical meaning for the galaxy itself. This is a big problem as these networks will assume these are important features, as encoding them will minimise the information loss within the network, possibly resulting in clusters that are not morphologically similar. This was shown in work by Spindler et al. (2021), who found that half of the encoded features from their network corresponded to geometric effects, and not morphological features of their target galaxies as desired. Other authors such as Cheng et al. (2021) address this rotation issue after feature extraction. During their clustering of the extracted features, they use the rotation of each galaxy as a feature to define the clusters, thus avoiding galaxy orientation as a feature. This method was successful, however one downfall is that other structural features could be unintentionally excluded from the clustering method as some of their encoded information was used to encode this rotation. Addressing this rotation issue manually also means that this method is no longer totally unsupervised. We aim to address this in our work by normalising geometric effects within our training data to remove these from being encoded as features, allowing for intrinsic morphological features to be encoded instead, resulting in more meaningful clusters of galaxies that are morphologically similar.

We address this issue in Chapter 3 by exploring the use of different unsupervised machine learning models for classifying high redshift morphology in a robust and more easy to interpret way. In this high-redshift regime, where the full spectrum of galaxy morphology remains unknown, these networks prove advantageous as they do not require labels to train the model. In particular we are also interested in the ability to extract global trends in morphology using these types of networks by investigating how the morphology varies with cosmic time. This is explored in the work presented in Chapter 4.

Chapter 2

Quantifying Non-parametric Structure of High-redshift Galaxies with Deep Learning

2.1 Introduction

With the imminent arrival of large imaging surveys from new facilities, such as the Euclid, Rubin and Roman telescopes, it is of paramount importance to look into the efficacy of existing methods for measuring galaxy structure. For example, parametric structural measurements are often very time-consuming to apply to large surveys. Non-parametric measurements are generally faster, but the algorithms are still typically applied to individual galaxies in series. While the problem is 'embarrassingly parallel', significant computational resources are required to measure large numbers of galaxies in a timely fashion. With the future of extragalactic astronomy moving to extremely large surveys, it is useful to explore more computationally efficient approaches.

In this work we therefore create neural networks capable of predicting concentration and asymmetry parameters from a galaxy's image. (For now we neglect the smoothness parameter as it is more difficult to measure at high redshifts and needs a separate treatment.) We show that our networks are consistent with conventional algorithms in their output, and demonstrate reliable behaviour down to very low signal-to-noise ratios. Furthermore, we find that our trained network is able to analyse $\sim 10,000$ galaxies in under 1.5 seconds, much faster than convention methods, making it well-suited to the large number of galaxies in future surveys.

This chapter is organised as follows. In §2.2 we introduce the imaging data used in this work and describe how the conventional CAS parameters are measured using the MORFOMETRYKA software (Ferrari et al., 2015). The pre-processing of the data and all data augmentation is detailed in §2.3.1 and 2.3.2. In §2.3.3 and 2.3.4 we describe the architecture and optimization of our neural networks. The resulting performance of these networks is demonstrated through a number of tests in §2.4, concluding with a brief summary in §2.5.

2.2 Data

2.2.1 CANDELS Fields

All of the images used in this project were taken with the Wide Field Camera 3 (WFC3) of the Hubble Space Telescope (HST) as part of the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS). We use data from all 5 CANDELS fields: the Great Observatories Origins Deep Survey (GOODS)-North and GOODS-South fields, COSMOS, Extended Groth Strip (EGS) and Ultra-Deep Survey (UDS).
The CANDELS/Deep survey (5σ point-source limit $H = 27.7$ mag) covers an area of ~ 125 arcmin^2 with a resolution of 0.06 arcsec per pixel (Grogin et al., 2011; Koekemoer et al., 2011). In total we have $\sim 150,000$ galaxy postage-stamp images. These galaxies have photometric redshifts covering $z = 0$–7, with many parameters already calculated, including star formation rates (SFR) (Duncan et al., 2019) and CAS values. The apparent magnitude–size distribution of our sample is shown in Fig. 2.1. In this work we use imaging from the H-band ($F160$W), as it provides the most complete deep-coverage over all five CANDELS fields.

2.2.2 Concentration and asymmetry

As mentioned in the introduction, non-parametric methods have been used for many years to analyse the light distributions of distant galaxies, in order to better understand their structure (Conselice 2003; Lotz et al. 2004a; Sazonova et al. 2020). Such methods make very few assumptions, and so can be applied to peculiar and irregular galaxies as well as to classic Hubble types.
In this work we utilise a subset of the CAS (Concentration, Asymmetry and Clumpiness) system as defined in Conselice (2003). This is a robust, non-parametric method for classifying galaxy structure, in a manner that is sensitive to their ongoing and past

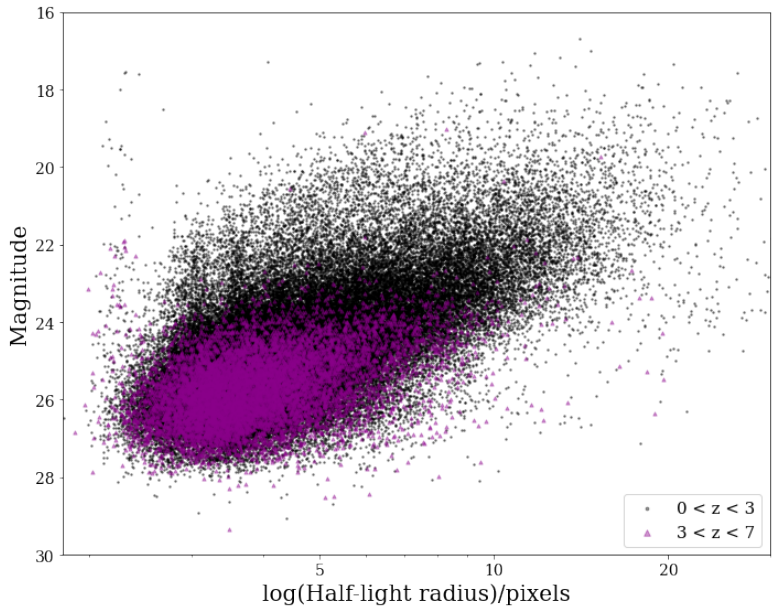


Figure 2.1: Distribution of the apparent magnitude in the H-band (F160W) vs half light radius, with a resolution of 0.06 arcsec per pixel, for the CANDELS galaxy sample used in this work.

formation modes. In this work, only concentration and asymmetry are considered. The concentration (C) is based on the measurement first established by Bershady et al. (2000), which was found to correlate with both galaxy bulge-to-disk ratio (B/D) and the effective radius of the bulge. This quantity is defined as

$$C = 5 \log_{10} \left(\frac{r_{80}}{r_{20}} \right). \tag{2.1}$$

where r_{20} and r_{80} are the radii containing 20% and 80% of the total light of the galaxy, respectively. The value of C is simply a measure of how concentrated the light in the central region is relative to the galaxy's overall size. Galaxies with higher concentrations are typically ellipticals, early-type disks, and edge-on disks. In this manner, it shares similarities with the Sérsic index (Graham et al., 2005).

Galaxy asymmetry was first used in a basic form by Schade et al. (1995), when trying to classify distant galaxies imaged with HST. Asymmetry (A) is determined by rotating a galaxy 180° about its center and then subtracting from the original image. The centre of rotation is determined by an iterative process that finds the minimum asymmetry. Further algorithmic details are described in Conselice (2000) and Conselice (2003). The absolute values of the residuals are summed and normalized by the original galaxy

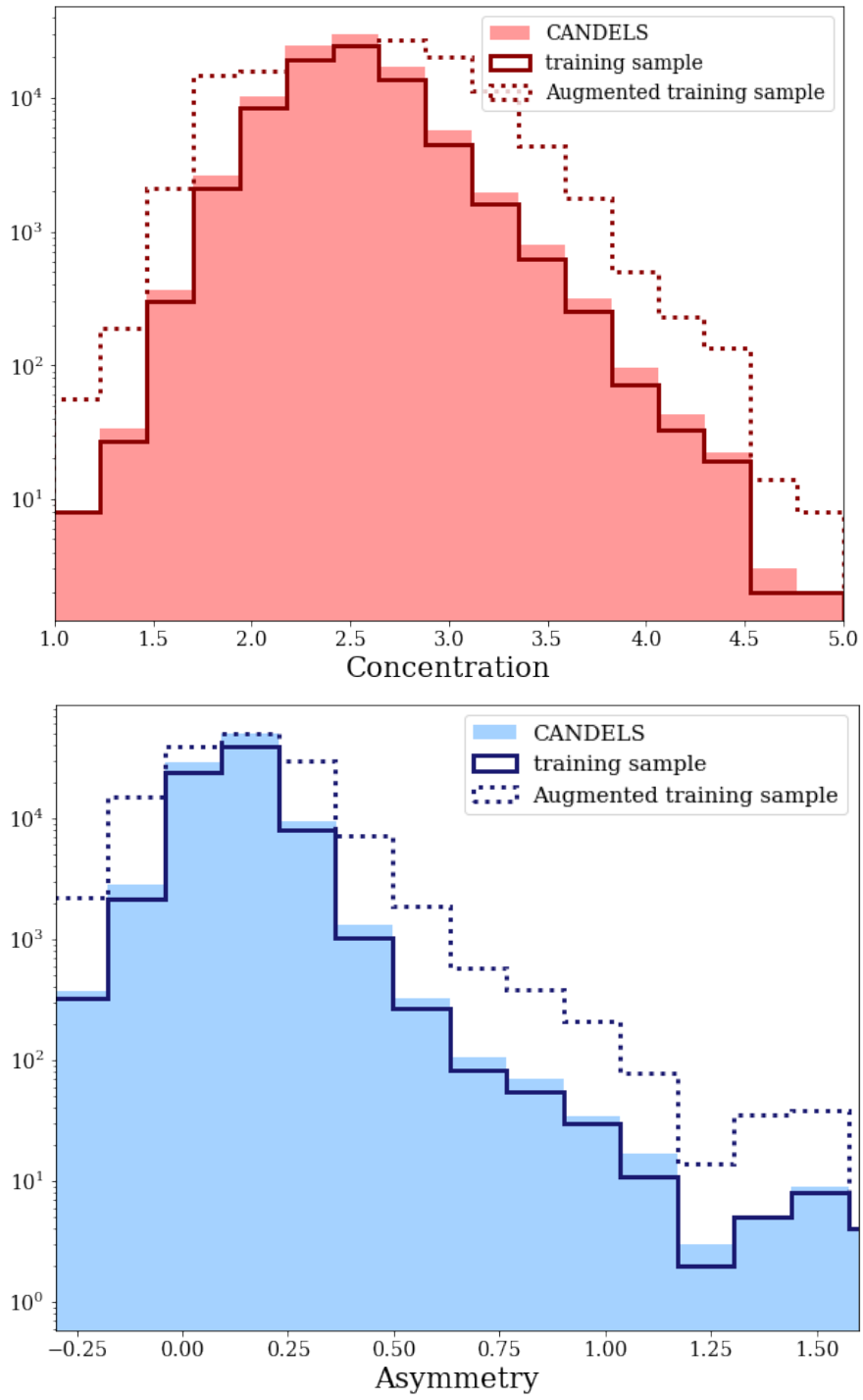


Figure 2.2: Distributions of asymmetry and concentration values for our selected sample of CANDELS galaxies, measured with MORFOMETRYKA. The solid line shows the selected training sample and the dotted line shows the training sample after our augmentation process.

flux. The resulting asymmetry contains a contribution from the background noise. This is accounted for by subtracting a background term, determined by computing the asymmetry for small areas of sky near the galaxy. The basic calculation for the asymmetry is therefore given by

$$A = \frac{\sum |I - I_{180}|}{I} - A_{\text{bkg}}\,, \tag{2.2}$$

where I is the original galaxy image, I_{180} is the rotated galaxy image, and A_{bkg} is the background asymmetry (discussed further below).

Asymmetry can be used to identify a number of interesting galaxy classes, such as mergers and starburst galaxies (Conselice 1997, Conselice 2000, Bluck et al. 2012). These types of galaxies have a higher A value than regular ellipticals and disk galaxies, due to distributed areas of increased star formation.

2.2.3 Morfometryka

CAS measurements were originally obtained using IRAF. However, a more modern implementation, in Python, is provided by Morfometryka[1] (Ferrari et al., 2015). Morfometryka extracts a number of features from astronomical images, such as non-parametric morphology (including the CAS parameters) and Sérsic profiles. Full details of the software can be found in Ferrari et al. (2015), however we will briefly describe how the parameters used in this work were calculated.

Morfometryka calculates the concentration, C, as explained in 2.2.2, with the exception that the factor of 5 in Eq. 2.1 is omitted. However, in order to remain consistent with previous studies, this factor was re-applied to our concentration values.

The asymmetry, A, is also determined as described in Section 2.2.2, by applying Eq. 2.2 within a 1.5 × Petrosian-radius elliptical aperture centred on the galaxy. The Petrosian-radius is defined to be the radius at which the surface brightness at a given radius is a fraction (η) of the total surface brightness within that radius. We define (η) to be:

$$\eta = \frac{I(r)}{\langle I(r)\rangle} \tag{2.3}$$

where $I(r)$ is the surface brightness at radius r and $\langle I(r)\rangle$ is the mean surface brightness within that radius. In Morfometryka the Petrosian-radius is calculated using $\eta = 0.2$. However, the background term in Eq. 2.2 is computed in a way that slightly

[1]The results in this work are based on Morfometryka version 8.2

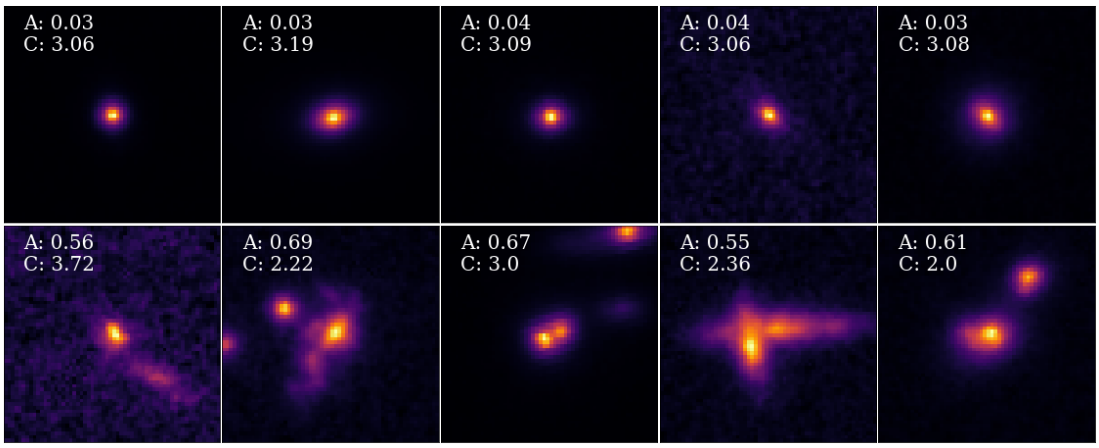


Figure 2.3: *Top row:* Images of the galaxies with high concentration values. These galaxies appear to be compact, spheroidal and have no close neighbours. These galaxies also typically have low asymmetry values, although there are some objects with high concentration and asymmetry. *Bottom row:* Images of galaxies with high asymmetry values. Many of these galaxies appear to be undergoing mergers and have tidal effects present, although there are occasional cases of line-of-sight projection. The concentration (C) and asymmetry (A) are indicated above each galaxy stamp.

deviates from the original CAS implementation. The standard approach utilises a single background region. Originally MORFOMETRYKA did not include the background asymmetry correction term. For our measurements, we construct a 10×10 pixel grid over the image area outside the galaxy segmentation map. We then measure the asymmetry for each cell in the grid, according to the first term of Eq. 2.2. Finally, we select the median asymmetry across all the cells as our background term, $A_{\rm bkg}$. This ensures a robust and accurate background correction, improving upon the original background subtraction by eliminating the bias inherent in choosing only one background area. This is now incorporated into MORFOMETRYKA.

The errors on the concentration values are derived from those of the individual size measurements, which assume Poisson distributed fluxes. The typical error on C is ±0.23. The error on the asymmetry values were calculated using the method described in Conselice (2003). We find that the typical error on A is ±0.072 for our sample.

We applied MORFOMETRYKA to all of the images in our dataset. We then select suitable galaxies for our analysis based upon the steps described in §2.3.1. The distributions of concentration and asymmetry values for our selected sample are shown in Fig. 2.2.

A subset of galaxy images were inspected to check that the measurements correspond to visual expectations. As can be seen from the top row of Fig. 2.3, galaxies with

high C values appear compact and spheroidal. Such galaxies typically have low A values, reflecting a broad anti-correlation between C and A for normal Hubble types. Galaxies with high asymmetries are shown in the bottom row of Fig. 2.3. The contrast between the two sets of galaxies is clear, with high A galaxies appearing disrupted, or possessing features associated with merging, such as tidal tails and multiple bright sources. Note that high asymmetry galaxies span a range of concentrations. This is a reassuring reconfirmation of how these parameters have been seen to behave in past studies (Conselice et al., 2008, 2011).

2.3 Method

2.3.1 Pre-Processing

The initial images used in this analysis are 101×101 pixel cutouts, with the target galaxy in the center of each stamp. As we are only interested in training the network to predict the A and C values for the target galaxy, we need to remove any other sources. In order to remove neighbouring sources from the cutouts, the galclean algorithm (Ferreira et al., 2018) was utilised. This algorithm removes any non-central sources at a certain threshold above the background level. These masked areas are replaced with values sampled randomly from the background distribution to ensure they do not leave shapes which could be picked up by the network.

The majority of our galaxies have a half light radius of ~ 10 pixels. For computational efficiency, the individual galaxy images are therefore further reduced in size to 60×60 pixels, centered on the galaxy.

Since we are interested in measuring structure irrespective of overall galaxy brightness, we individually normalize our images. The pixel values of each image are rescaled so that the maximum pixel value of each image has a value of 1. This is also a standard pre-processing procedure for deep learning. It improves learning efficiency by ensuring that the inputs to the networks are compatible with the domain of the activation functions used within the model.

We wish to consider only reliable galaxy detections, for which structural parameters can reasonably be obtained. We therefore limit our sample to galaxies above a minimum signal-to-noise. We define the average signal-to-noise per pixel for each galaxy as

$$SNR_p = \frac{L_{\rm tot}}{q\pi R_p^2 \sigma_{sky}} \tag{2.4}$$

where $L_{\rm tot}$ is the total integrated flux within the Petrosian region (with semi-major axis R_p), q is the axis ratio measured from the intensity distribution using the image moments, and σ_{sky} is the standard deviation of the sky background. By visual inspection we define a selection for our galaxy sample of $SNR_p > 2$. We also limit our sample to galaxies with $R_p > 5$ to ensure they are properly resolved.

Once these steps have been completed, we are left with 94,192 galaxy images with a median $SNR_p \sim 4.5$. These images were then split randomly into training (80%), testing (10%) and validation (10%) datasets to apply to our machine learning methods. With over 9,000 galaxies in each of our testing and validation sets, our performance estimates will be both accurate and precise.

2.3.2 Data Augmentation

Unbalanced datasets, whereby there are many more galaxies at one particular value compared with others, can cause issues when dealing with both regression and classification problems in machine learning. The relative frequency of classes in the training set acts as a prior; the network may therefore be biased against identifying rare cases. In extreme circumstances, the network may fail to learn to identify rare cases at all. One way to combat this issue is by data augmentation (Shorten & Khoshgoftaar, 2019).

Data augmentation is primarily used as a way of creating a larger training sample, which more finely samples the space of possible inputs. It is a form of regularisation and hence helps to prevent overfitting. By selectively expanding the size of the potential training set, augmentation can also help to balance the prevalence of different classes, while still using all of the input data.

Looking at Fig.2.2, there is a large imbalance in the CAS values for our sample, such that very high asymmetries are not common, nor are very low or high concentrations. As we want our model to be accurate across all concentrations and asymmetries, we selectively apply augmentation to create a more balanced training set. That is, we need to supplement the images that occupy the parameter space where there are few galaxies. For the range of C or A values where there are around half the number of images compared to the median value, we rotated each image by 90° once. Where there are relatively fewer images, we apply a greater variety of augmentations: rotating by 90° 3 times and mirroring along both axes. These images were then shuffled and added to the training set. After data augmentation, our training sample increases in size from 75,353 to 141,453 images.

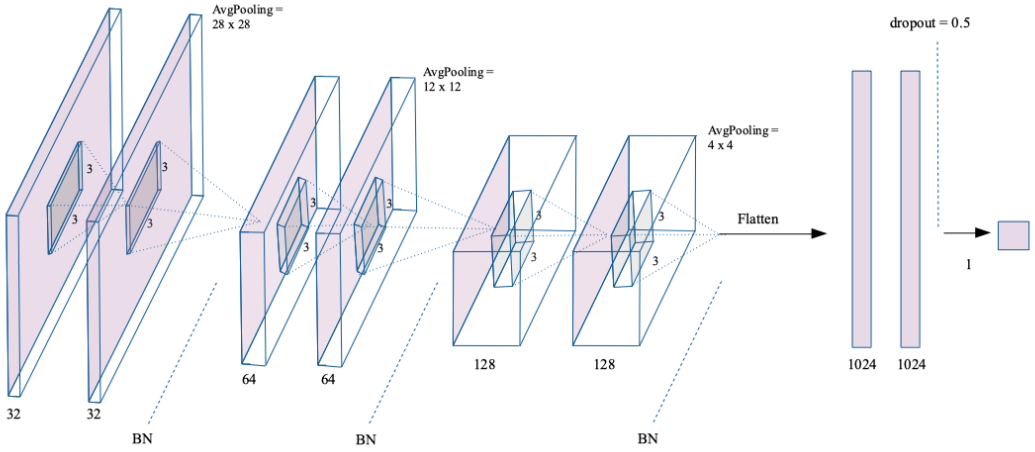


Figure 2.4: Architecture of the asymmetry network after optimization of its hyperparameters. This network takes input images of size 60×60 pixels, followed by 3 blocks, each containing 2 convolutional layers with 32, 64 and 128 features, respectively. Each block is followed by an average pooling layer of size 2 and a batch normalisation layer. Finally there are 2 fully-connected layers of size 1024 each, which is followed by dropout of 0.5 before the output value, i.e the network's prediction of the asymmetry.

2.3.3 Convolutional Neural Networks

The purpose of this project is to efficiently and robustly predict CAS values of a galaxy from an image. We chose to implement a Convolutional Neural Network (CNN), as these are known to perform well when dealing with spatial structured data (for details see §1.5.2).

There are many factors to consider when choosing the optimum architecture for a network. Many early studies based their architecture on previous studies (Huertas-Company et al. 2015; Domínguez Sánchez et al. 2018; Aniyan & Thorat 2017), trial-and-error (Dieleman et al. 2015; Feinstein et al. 2020), and arbitrary choices. However, there are a number of optimisation techniques that allow these choices to be optimised in a more satisfactory manner for the problem at hand. The variety of network architectures we consider, and our method for selecting from these, are described in the following section.

To evaluate how well our networks are performing we compute the mean absolute error (MAE) and root mean squared error (RMSE) of the network's predictions. The RMSE metric also serves as our loss function. The MAE is simply a measure of the average

magnitude of error between the network's prediction and the expected result,

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|\,, \tag{2.5}$$

where n is the number of samples, y_i is the expected value and $\hat{y}_i$ is the network's prediction. The RMSE is similar to the MAE, but it is more sensitive to large errors and so can indicate if there are many outliers present. It is calculated as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}\,. \tag{2.6}$$

2.3.4 Bayesian Optimization

The various choices that must be made before training a network can be considered as hyperparameters. These include aspects of the network architecture, such as the number of convolutional layers and number of filters in each layer, and of the training, such as the update algorithm, learning rate and batch size. Varying these choices can significantly alter the performance of the trained network. The problem of determining which combination of hyperparameters will be best suited to a given problem typically involves a trial and error process, which is often only partially explored, or entirely neglected, resulting in a non-optimal solution.

To avoid this, many optimization techniques have been developed, from simplistic random or grid-based searches (Bergstra et al., 2011), to more advanced techniques such as random forests (Hutter et al., 2011). The aim of these techniques is to find the optimum hyperparameters that will minimise the average loss. Traditionally, these techniques can be computationally expensive, as each variation in the hyperparameters results in a new version of the network which must be trained and then evaluated. Bayesian Optimisation (Snoek et al., 2015) provides a more efficient solution: a record of past evaluation results are kept and used to form a probabilistic model, which the method builds upon, reducing the time to converge on a optimal model.

Our networks comprise a number of convolutional blocks, between 1 and 3, with each block having either 1 or 2 convolutional layers. Each convolutional layer in a block has the same number of filters between 8 and 256 in powers of 2. The kernel sizes are all fixed to 3×3. Each convolutional block is followed by a `BatchNormalisation` layer and an `AveragePooling` layer of fixed ($2\times$ 2) size. Originally we started with `MaxPooling` layers, however we found that the networks' performance improved when

using `AveragePooling` layers. (Similar behaviour was found by Pasquet et al. 2019 when analysing SDSS images.) Following the convolutional blocks we add some fully connected layers, with their number and size as hyperparameters. The number of fully connected layers ranges from 1 to 4, with each layer having the same number of filters between 128 and 1024 in powers of 2. We include a dropout layer before our output layer as a form of regularisation, allowing the dropout rate to vary as another hyperparameter. The dropout rate is allowed to vary continuously between 0.25 and 0.60. The activation function is fixed to the common `ReLu` (Nair & Hinton, 2010) non-linearity.

When training a network, an optimization algorithm adjusts the weights to minimise the cost function. With a plethora of optimizers now available, we have included the choice as a hyperparameter, selecting from a pool of those most commonly used, we include Adam, Adadelta, RMSprop, SGD and Adamax. We also set the learning rate as a hyperparameter, where we evaluate 5 values, 0.001, 0.005, 0.01, 0.05 and 0.1.

The parameters we defined as hyperparameters and their optimised values are displayed in Table (2.1).

Each network was trained for a maximum of 300 epochs, but we applied "early stopping" to halt the training when the validation loss had converged, which was typically after $\sim$ 100 epochs.

Our Bayesian Optimization was carried out using the GPyOpt python package (The GPyOpt authors, 2016), with the aim to minimise the RMSE of the networks. Each network created during the optimization was trained and validated using the samples defined in §2.2. The MAE, RMSE and the Pearson coefficient were monitored for each iteration in the optimization. The network that had the lowest MAE and RMSE was selected as the optimum architecture for our network.

The architecture of the CNN selected for our asymmetry network is shown in Fig.2.4. To ensure that the choice of optimum architecture is robust, we retrain multiple times, and compare the variation in the loss to the variation observed between different networks. The variation in MAE for the optimum asymmetry network is quite stable and varies by $\sim$ 0.001. Comparing the top 10 network architectures, we find that the MAE varies by $\sim$ 0.002. The hyperparameters of these networks are quite similar, although the number of fully connected layers varies between 1 and 2, the dropout rate between 0.47 and 0.56, and the optimizer varies between Adamax and Adadelta. These parameters are not as significant in determining the optimum network.

The selected concentration network has a similar architecture, with some slight variations. The MAE loss variation across different training runs is $\sim$ 0.002, and the MAE of

Hyperparameter	Optimum value	
	Asymmetry	Concentration
batch size	512	512
convolutional blocks	3	3
conv. layers per block	2	2
fully-connected layers	2	2
fully-connected layer size	1024	512
number of filters	32	64
optimization	Adamax	Adam
learning rate	0.001	0.001
dropout	0.50	0.55

Table 2.1: Summary of the hyperparameters selected by the Bayesian Optimization technique.

the top 10 architectures vary by ~ 0.003, very similar to above. Looking at the variation in the architectures which give equivalent performance, we see that the number of fully connected layers varies between 1 and 2 layers, the batch size between 256 and 512, the dropout rate from 0.3 to 0.6, and the number of convolutional blocks varies between 2 and 3.

Following our use of Bayesian Optimisation and the above tests, we can be confident that our final selected networks are well-optimised. However, it is also reassuring that the performances we report below are robust to minor variations in network architecture and training.

2.4 Results

As explained above, we train our networks on a subset (80%) of the images and select our optimal model by its performance on a validation set (10%). To then evaluate our selected, trained network, we use an additional independent test set (10%; 9,420 galaxies). This ensures that the metric used to evaluate the network's performance is not biased by over-fitting the hyperparameters. We find that the networks perform similarly on both the test and validation sets: another indication that the selected network architecture is robust.

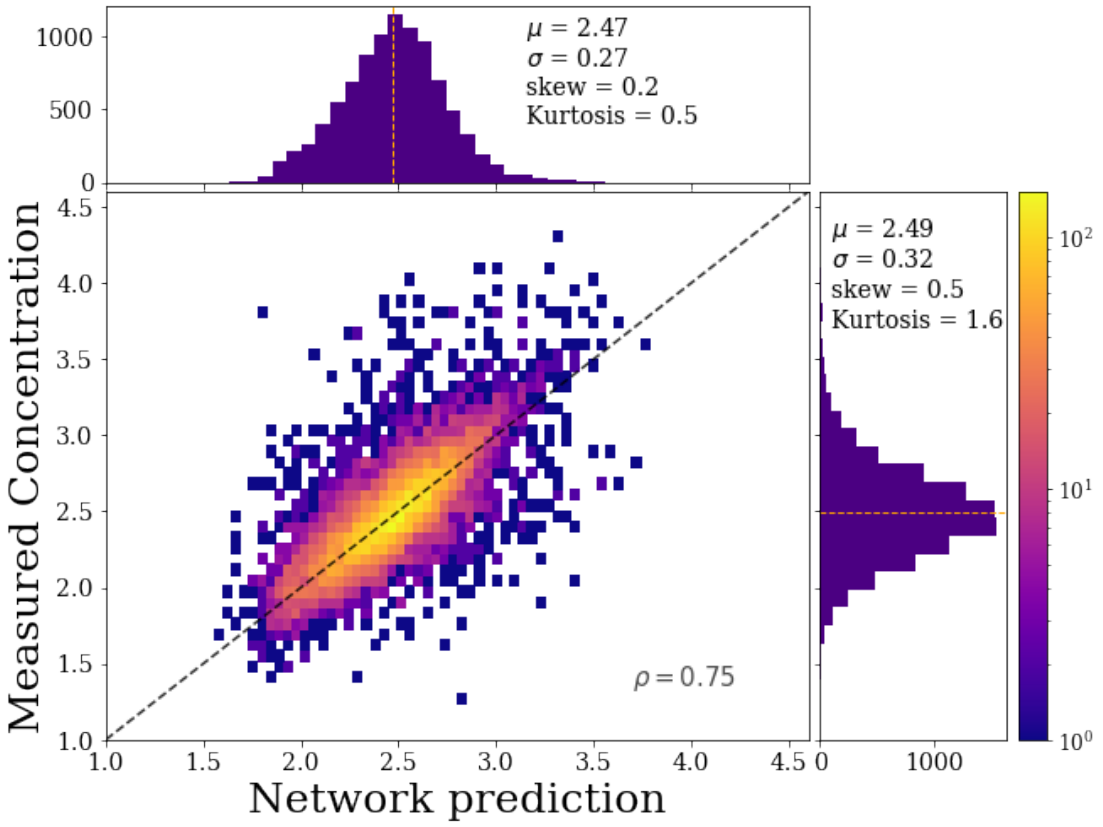


Figure 2.5: Our final network's predictions of concentration values for the test set versus those measured by Morfometryka. The network does not see any of the images used in this comparison during its training. The MAE of the network was 0.15, with a RMSE of 0.21 (see §2.4). The Pearson correlation coefficient of 0.75 indicates that there is a strong correlation between the two.

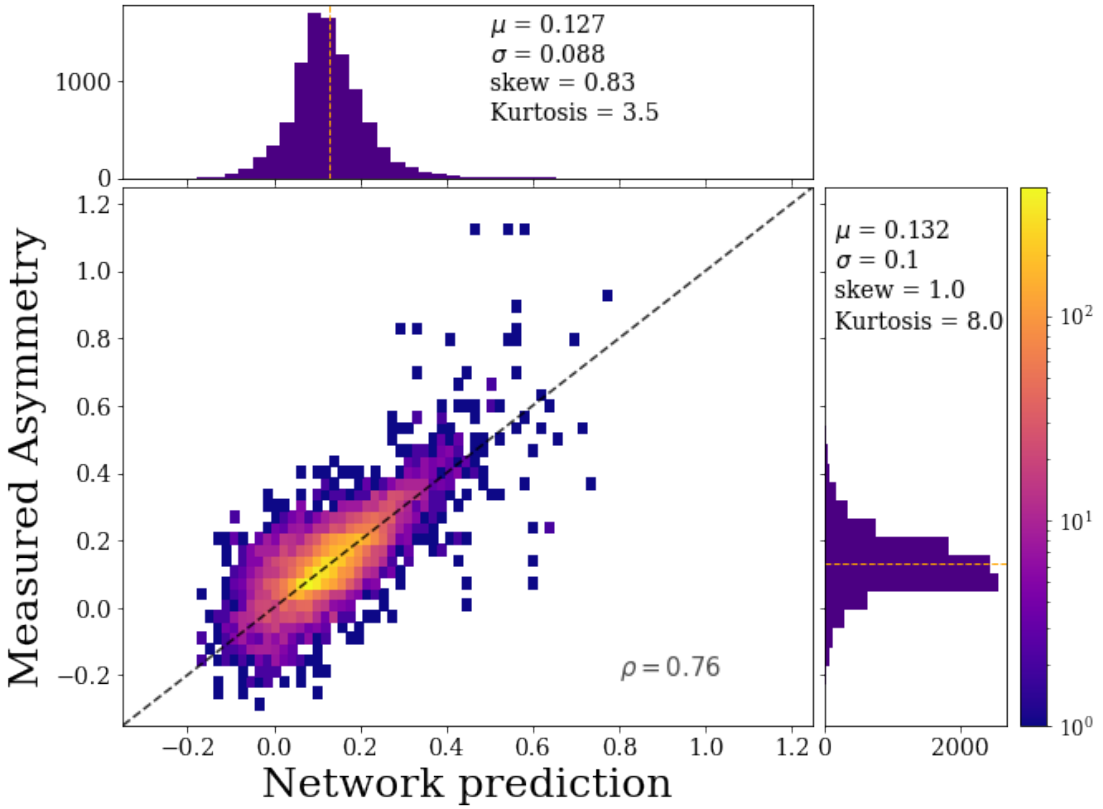


Figure 2.6: Our final network's predictions of the asymmetries for the test set versus those measured by MORFOMETRYKA. The network did not use any of these images during its training. The MAE of the network was 0.045, with a RMSE of 0.065 (see §2.4). The Pearson correlation coefficient of 0.76, indicates that there is a strong correlation between the two.

2.4.1 Model performance

The basic results of the concentration network after the hyperparameter optimisation can be seen in Fig.2.5. The network's predictions correlate strongly with the Morfometryka measurements, with a MAE value of 0.15 with a RMSE of 0.21. This error is lower than the average error on the concentration measurements. This shows that our machine learning regression can measure these parameters just as well as the direct measurement method. Hence, the values from the network can be utilised with a similar level of confidence as the original algorithm However, the scatter does get larger at parameter values where there are fewer galaxies.

The results for the asymmetry network, again after the hyperparameter optimisation, can be seen in Fig.2.6. The network's predictions for the asymmetries have a MAE of 0.045 with a RMSE of 0.065. As before, this error is lower than the average error on the asymmetry measurements, showing that our networks can be used to reliably measure both the concentration and asymmetry values for a galaxy.

Overall, both networks perform well, achieving low residuals between the measured values and the network's predictions. Looking at the images of galaxies where there was a large difference between our networks and the Morfometryka-measured CAS values, we find that they are quite noisy, with $SNR_p < 4$. From this we decided to further investigate the impact of noise on both our network predictions and the directly-measured CAS values.

2.4.2 Impact of noise

Our networks' ability to accurately predict CAS values is potentially dependent on the noise level in a given galaxy image. To investigate this, we consider how the residuals (network prediction − Morfometryka value) of each network depend on the signal-to-noise per pixel, SNR_p. This is shown in Fig.2.7. Confirming the results from Figs.2.5 & 2.6, we see that for galaxies with moderate and high SNR_p the residuals are close to zero. The random scatter is also fairly constant with SNR_p, indicating that our networks are reliable across a broad range of SNR_p. There are a small number of galaxies with large deviations at the higher SNR_p however, when inspecting these images we find that most contain another source in the image that was not removed by the galclean algorithm. This could explain why the measurements for these galaxies from Morfometryka and our networks varied. Within the low SNR_p regime we find a slight bias where the networks, on average, under-predict the values measured by standard algorithms.

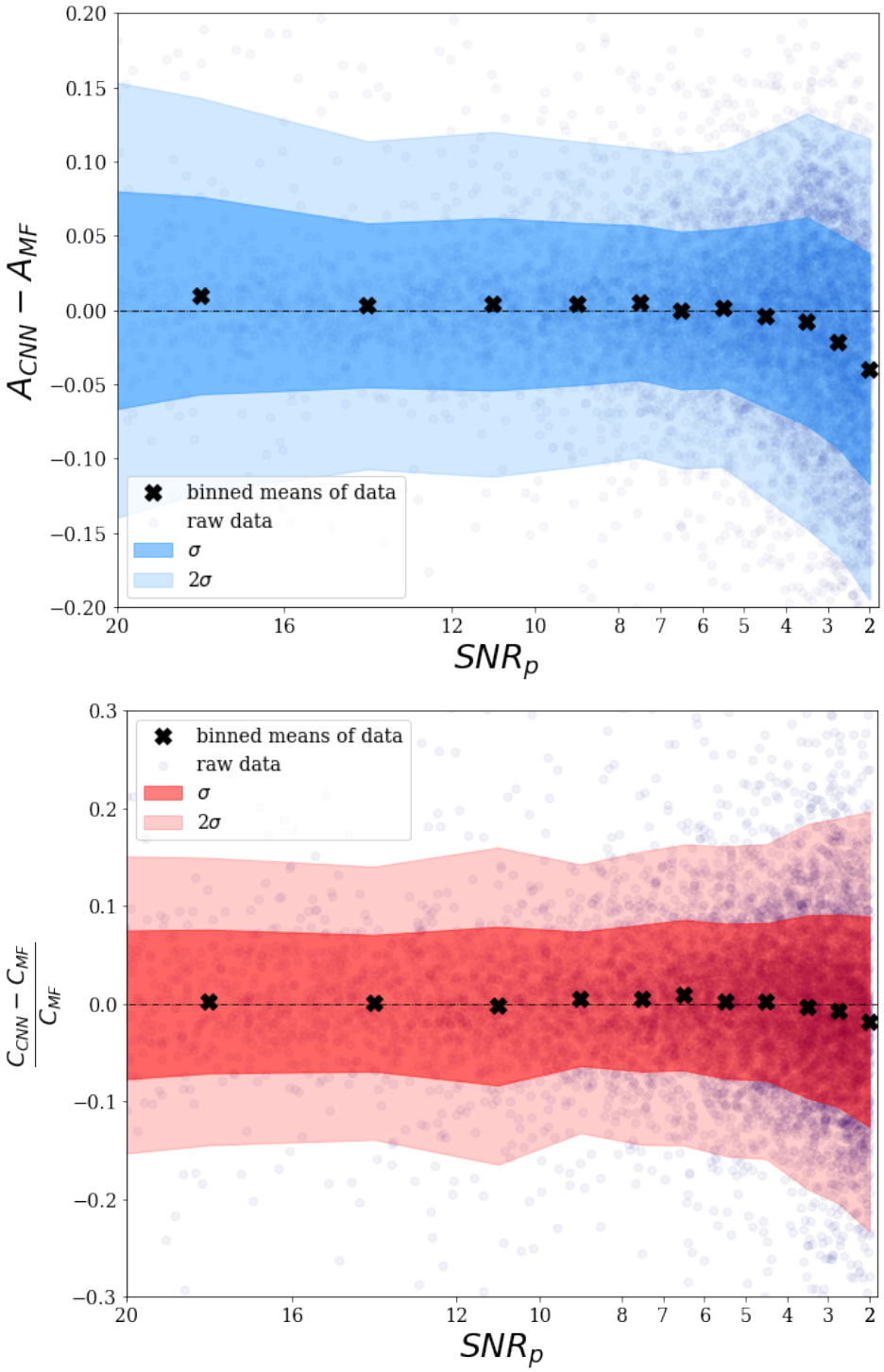


Figure 2.7: *Top:* The residuals of the asymmetry network and the measured values from MORFOMETRYKA versus the SNR_p of each galaxy image. *Bottom:* The fractional residuals between the concentration network and the measured values from MORFOMETRYKA versus the SNR_p of each galaxy image. In both panels, black points represent the means plotted for each bin with the darker shaded region representing ±1 standard deviation and the lighter shaded region shows the area containing 2σ. The trend indicates that there is a bias at low SNR_p, where the networks will, on average, slightly under-predict the values measured by the standard algorithms.

The origin of this systematic trend at low signal-to-noise is interesting. Our networks have been trained to reproduce the measured values, and are clearly doing so in the majority of cases. So why the deviation at low SNR_p? This could be seen as a failure of our model to capture the details of the measurements. On the other hand, we apply regularisation and optimise the hyperparameters to avoid over-fitting, with the aim of producing a generally applicable model, capable of accurate measurements for a wide variety of images. One optimistic possibility is that our networks are able to learn a model which is more robust than the regular methods. This is not inconceivable, since the regular methods must make a series of algorithmic 'decisions' (masking, fitting elliptical isophotes, recentering, etc.). The networks, instead, consider all of these issues within a single 'holistic' calculation.

In order to determine if this low signal-to-noise trend is a bias in our networks or in the standard algorithm (as implemented in Morfometryka), we investigate how noise impacts these two approaches in an independent manner. For this test, we select a sub-sample of 622 high SNR_p galaxies ($SNR_p > 10$), with low asymmetry residuals ($|\Delta A| < 0.01$) from the validation set. These galaxies also have low residuals in their concentration values. These galaxies are chosen as both the network and Morfometryka predicted these galaxies to have similar parameters, and hence we can assume these to be the true values for the purpose of this test.

We then produce versions of each galaxy image with varying SNR_p values. To do so, we first measure the mean and standard deviation in the background of the original galaxy image, then create an image with corresponding Gaussian noise. To this simulated background image, we add the original image with the overall flux scaled, such that we achieve our desired SNR_p. Finally, the image is normalized in the usual manner (§2.3.1). Except for the variation in SNR_p, each galaxy image remains identical to its original version. An example of these simulated noisy images can be seen in Fig.2.8.

The asymmetry and concentration values for these galaxies are then re-measured at each SNR_p, using both Morfometryka and our trained networks. The variation from the values measured in the $SNR_p = 10$ image is plotted as a function of decreasing SNR_p in Figs.2.9 & 2.10.

For asymmetry, it can be seen that at both high and moderate SNR_p the values recovered by our network are very similar to the 'true' values. For $SNR_p > 5$ the recovered values vary with an average standard deviation of 0.025, reflecting the uncertainties due to shot noise.

Furthermore, this scatter is significantly lower for our network than the standard algo-

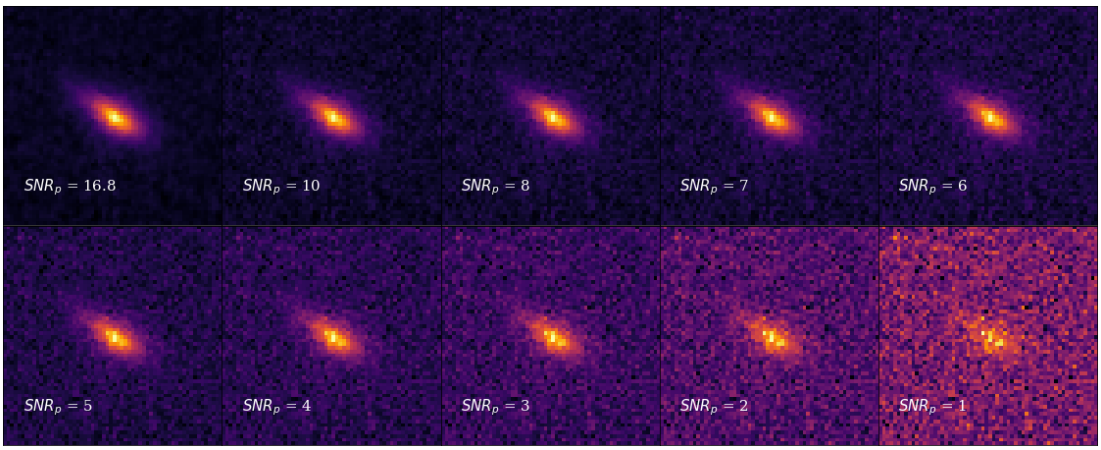


Figure 2.8: An example of our simulated noisy galaxy images. The top-left panel shows the original image, while the remainder show the same galaxy at different simulated SNR_p.

rithm. The average scatter in the asymmetry measurements, at $SNR_p \geq 5$, is 0.025 compared to 0.037 for Morfometryka. Since we have already seen that our networks accurately recover Morfometryka measurements, this suggests that our network is using information in these moderately-noisy images that is not utilized by the Morfometryka algorithm.

At low SNR_p, we find a bias present in both the network and Morfometryka, such that the A values are, on average, overestimated. However, it can be seen from Fig.2.9 that Morfometryka has a larger bias at these low SNR_p, indicating that our network is slightly more accurate in the low SNR_p regime. At $SNR_p = 3$ the bias in the network's asymmetry values is 0.016 compared to 0.028 for Morfometryka.

While there is scatter in the individual measurements, on average our network is able to accurately estimate asymmetry, with little bias from the 'true' value, at a lower SNR_p than the original algorithm. This is useful for merger fraction estimates, especially at high redshift, as we can now include galaxy images down to a SNR_p as low as 3 while still retrieving unbiased measurements with our network. This means that we are able to measure reliable CAS parameters for more galaxies using deep learning and then we can with a direct measurement.

For concentration, plotted in Fig.2.10, we again see a a difference in the variation of the measurements for moderately-noisy images, with our network producing a significantly lower scatter than the standard algorithm. The average scatter in the concentration measurements at $SNR_p \geq 5$ are 0.04 for the network compared to 0.07 for Morfome-

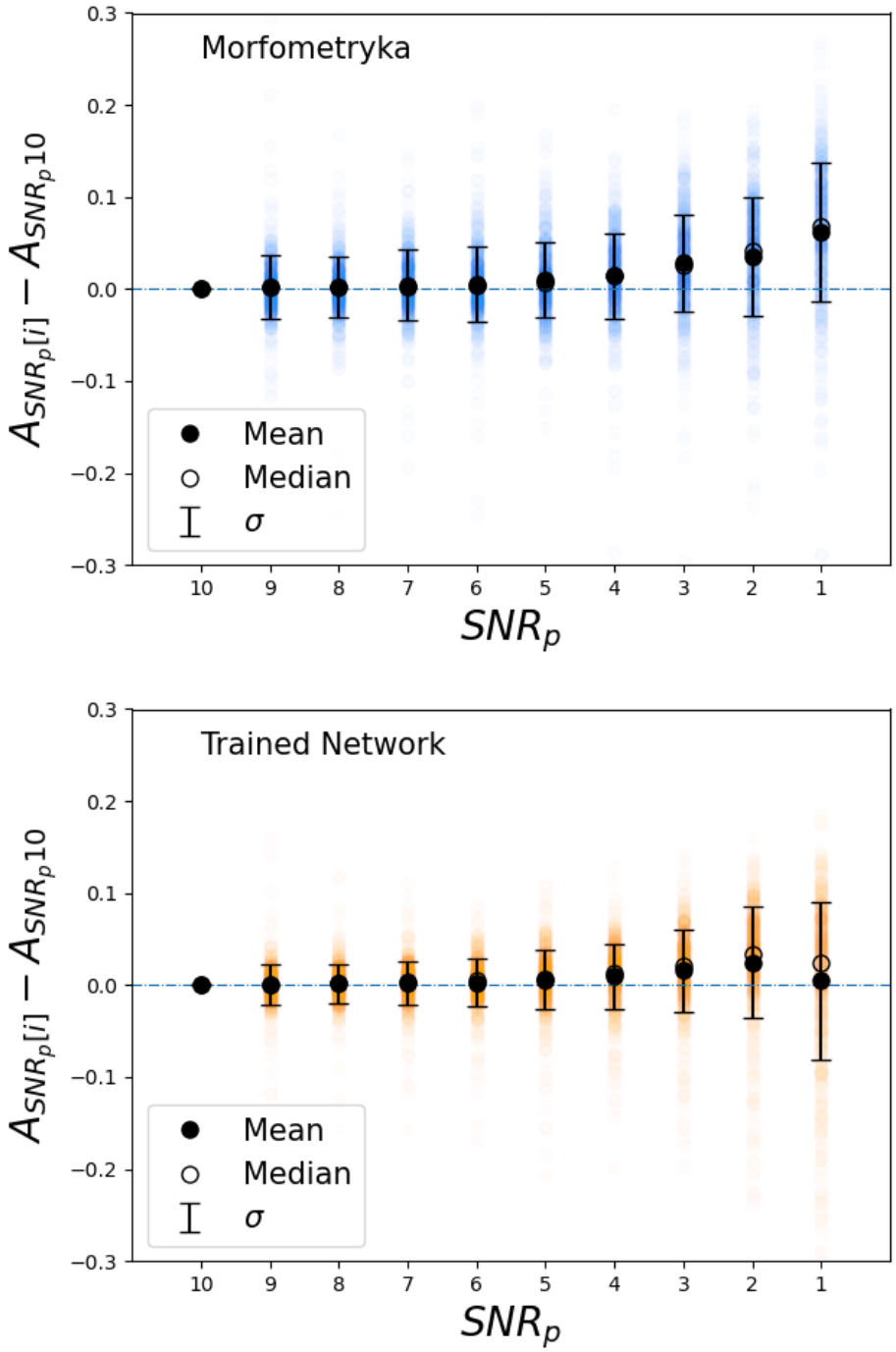


Figure 2.9: *Top:* Deviation in the asymmetry measurements from $SNR_p = 10$ for Morfometryka as the signal-to-noise is decreased in bins of SNR_p. At $SNR_p \leq 5$ there is a consistent bias, such that, on average, Morfometryka overestimates the values. Black points indicate the mean residual at each bin, with error-bars showing ±1 standard deviation. The median deviations are shown by open points. *Bottom:* As the top panel, but for our trained network. A similar trend can be seen at low SNR_p, but the scatter and systematic deviation is lower. This shows that the network is more stable and accurate than Morfometryka at these low SNR_p values.

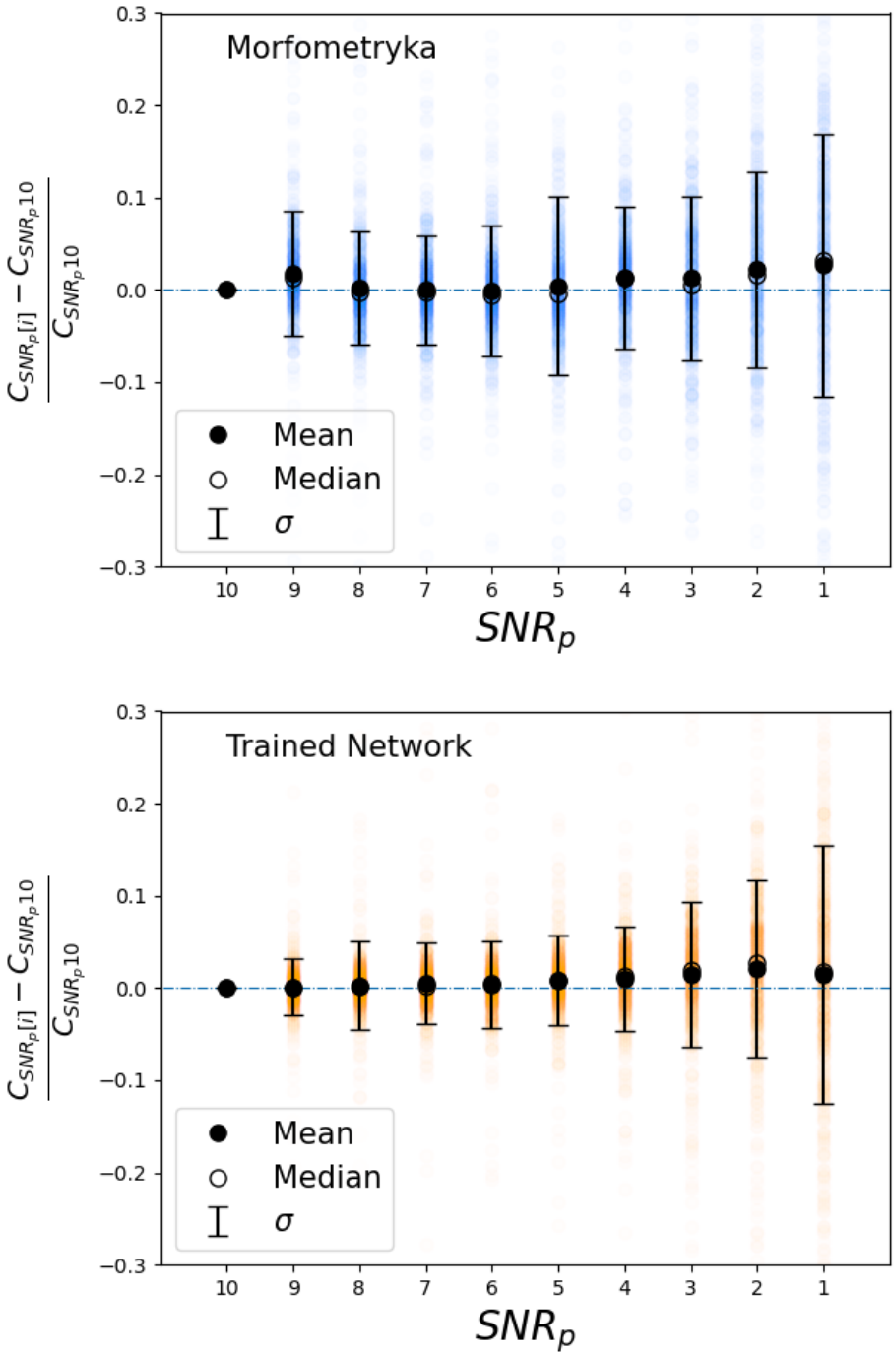


Figure 2.10: *Top:* Deviation in the concentration measurements from $SNR_p = 10$ for Morfometryka as the signal-to-noise is decreased in bins of SNR_p. Black points indicate the mean residual at each bin, with error-bars showing ±1 standard deviation. The median deviations are shown by open points. *Bottom:* As the top panel, but for our trained network. At $SNR_p \leq 5$ there is a slight bias, such that, on average, Morfometryka and the network overestimate the values of the sample. It can be seen that the C measurements are quite stable for $SNR_p \geq 3$. It can also be seen that the network produces a significantly lower scatter than Morfometryka

.

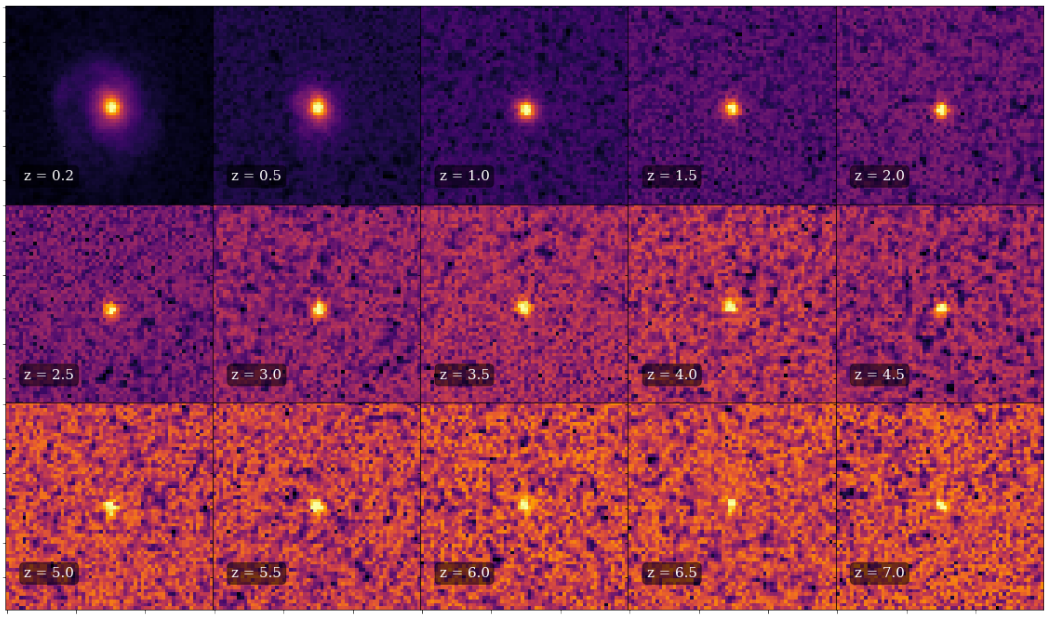


Figure 2.11: An example of our redshifted Frei galaxy images. In our redshifting technique we include effects due to cosmological dimming, luminosity evolution, size evolution and geometric scaling. Top left panel shows the reference image used in our tests. We start at z = 0.2 so that the whole galaxy fits within the 60×60 crop which is the input for our networks. The simulated redshift is indicated in each image.

TRYKA. Both the network and MORFOMETRYKA slightly overestimate the concentration at $SNR_p \leq 5$.

We also investigate the 'catastrophic' fraction (f_c) of both MORFOMETRYKA and our network, that is the number of galaxies that fall outside of 2 sigma deviation from the mean. Again the network performs marginally better, with f_c at $SNR_p \geq 5$ being 4.3% for MORFOMETRYKA compared to 3.3% for our network.

Based on these results, we conclude that our deep-learning approach is performing at least as well as traditional measurements of non-parametric structure.

2.4.3 Impact of redshift effects

While the previous test examined how our networks fare with respect to noise alone, here we combine the effects of signal-to-noise and resolution to determine our networks' performance for galaxies at high redshifts.

It is known that at higher redshifts, cosmological dimming and decreasing apparent size result in galaxies appearing more symmetric and less concentrated than they otherwise would (Conselice, 2003). However, through the use of simulations, we can model and

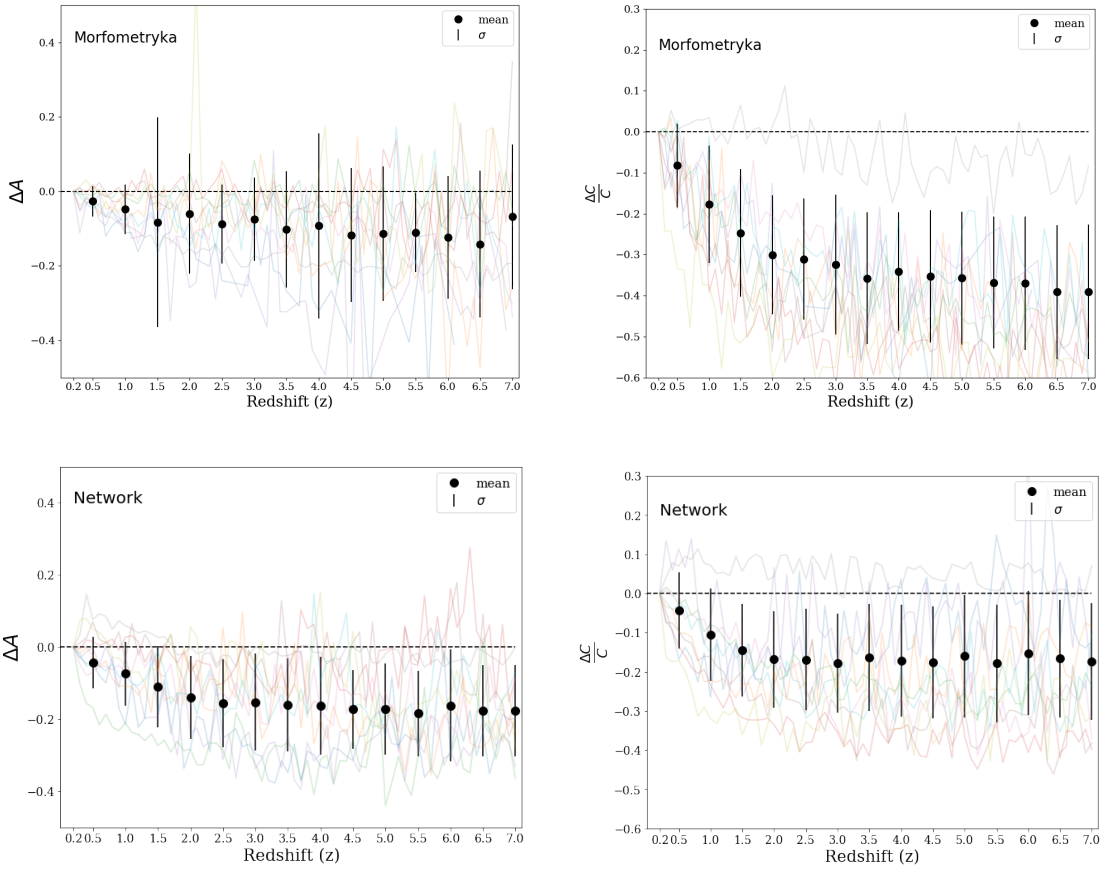


Figure 2.12: Variations in the C and A measurements from our trained networks and MORFOMETRYKA as a function of simulated redshift. Our simulations include geometric scaling (with accompanying reduction in resolution), size evolution, cosmological dimming and luminosity evolution. The original asymmetry measurements range from $0.01 < A < 0.48$, while the concentration measurements vary from $2.4 < C < 5.0$. There are a total of 100 galaxies from the Frei sample in this analysis. The variations for a sample of individual galaxies in the sample are shown by faint lines. These span a range of morphological types and initial asymmetry and concentration values. Black points indicate the mean variation and the error bars show the standard deviation of the individual measurements.

correct for this variation. We thus quantify the extent to which C and A values estimated by our networks are biased by these issues, and hence the level of any correction which should be applied when comparing galaxies at different redshifts.

To investigate how the performance of our networks varies with redshift, we take a sample of nearby galaxies, with reliably measured concentration and asymmetry values, and simulate how they would appear at higher redshift. For this test, we selected objects from the Frei catalogue of nearby galaxies (Frei et al., 1996). These are a well studied sample of regular, nearby galaxies, containing all Hubble types and with previously measured A and C values (Conselice, 2003). We simulate the appearance of these galaxies as if they were observed at a range of redshifts, from $z = 0.1$–7.

There are a number of effects that need to be considered when artificially redshifting galaxies. The first effect we address is geometric scaling, whereby the apparent size of the galaxy will decrease when viewed at a higher redshift. We follow the same procedure as described in Conselice (2003) and de Albernaz Ferreira & Ferrari (2018) to reduce the sizes of the galaxies to how they would appear at higher redshifts.

The angle that a galaxy with true size d will subtend varies with redshift following;

$$\theta_z = \frac{d}{A_z} = d\frac{(1+z)^2}{L_z} \tag{2.7}$$

where A_z is the angular diameter distance and L_z is the luminosity distance. In order to generalise this equation for any instrument we can say that the angle a galaxy subtends in pixels (n) is $\theta_z = n_z p_z$, where p_z is the pixel scale of the instrument. A galaxy imaged at a higher redshift (z_2) will therefore appear as n_{z_1}/n_{z_2} times smaller than the original image, resulting in;

$$\frac{n_{z_1}}{n_{z_2}} = \frac{L_{z_2}}{L_{z_1}}\frac{(1+z_1)^2}{(1+z_2)^2}\frac{p_{z_2}}{p_{z_1}} \tag{2.8}$$

Previous simulation work has kept the physical size of the galaxies constant. However, it is well known that galaxies of a given stellar mass are intrinsically smaller at higher redshifts, reducing in size by a factor of ~ 5 between $z = 0$ and $z = 3$ (Trujillo et al., 2007; Buitrago et al., 2008). We therefore introduce size evolution to better represent the properties of high-redshift galaxies.

We use the size evolution determined by Whitney et al. (2019), which is based on the same H-band data from CANDELS GOODS North and South fields that we use in our training sample. They measured how the average physical Petrosian radius, R_p, changes with redshift for a mass selected sample, finding that it varies according to

$R_p(z) = \frac{R_p(z=0)}{(1+z)^\beta}$, with $\beta = -0.97$. We therefore multiply the geometric scaling factor by this value to correct for the size evolution in our simulation.

After the (flux-preserving) geometrical scaling, we apply cosmological dimming Tolman (1930), according to

$$I(z) = \frac{I(z=0)}{(1+z)^4} \tag{2.9}$$

where $I(z)$ is the observed intensity. This is one of the major issues when detecting high redshift galaxies, as it introduces a bias such that only the brightest, most compact galaxies are detectable. The intrinsic brightness of galaxies, with a given stellar mass, varies with redshift. We therefore implement an evolution in the surface brightness of the galaxies as outlined in Whitney et al. (2020). They found that the correction for the intrinsic surface brightness follows

$$\mu(z) = \mu(z=0)(1+z)^\alpha \tag{2.10}$$

where $\mu = -2.5log(I)$ and $\alpha = -0.13$. The value of α was found to vary from -0.09 to -0.18, but this will not result in much variation in our results. The value of -0.13 was the value found for their size corrected sample.

To complete our simulations, the galaxies are convolved with the HST PSF in the H_{160}-band filter and placed in an actual CANDELS background.

We do not account for morphological or magnitude k-corrections. Instead, we test how a galaxy image would vary in restframe optical wavelengths, i.e. choosing appropriate observed filters for different redshifts. While we are currently only able to probe restframe optical up to $z = 3$ with HST, future surveys, such as JWST, will be able to probe up to $z = 7$. Furthermore, it has been found that the CAS parameters do not vary much between the UV and optical for star forming galaxies (Conselice, 2003).

We select the brightest galaxies for this test, such that they are above $SNR_p > 2$ in all images out to $z = 7$, as this was the cut off used in our training sample. This leaves us with 100 out of the original 112 galaxies in the Frei sample. The original asymmetry values of the sample range from $0.01 < A < 0.48$, while the concentration measurements vary from $2.4 < C < 5.0$. The A and C parameters of each galaxy were remeasured at each redshift. We only consider $z = 0.2$ onward, since we need the whole galaxy to fit within a 60×60 pixel image, for input to our networks. An example of one of the redshifted galaxies is shown in Fig.2.11.

The variations of the CAS values measured by our networks and Morfometryka are plotted against redshift in Fig.2.12, with the full distributions at a sample of redshifts

shown in Fig.2.13 for comparison. As expected, at higher redshifts both methods measure the galaxies to be more symmetric and less concentrated than at $z = 0.2$. As the outer regions of a galaxy fade below the background noise, and their apparent size approaches the resolution limit of the PSF, they appear more symmetric, as has been found previously (Conselice, 2003).

While the Frei sample of galaxies appear somewhat different to those the networks were trained on, we see that the networks still perform well, measuring values similar to MORFOMETRYKA. The average change in asymmetry at $z = 1.0$ compared to $z = 0.2$ is 0.074, which is similar to the average error on the high-z asymmetry measurements. This is important for merger estimates, as this variation is small enough to avoid a merger appearing as a non-merger and vice versa. At redshifts higher than $z = 1$ the average variation is around twice the average error on the measurements. While we cannot be sure what the equivalent $z \sim 0$ A value of an individual galaxy would be, if investigating the galaxy population at high redshift, the average A value could be corrected.

We have included the equivalent results from Conselice (2003) in Fig.2.13. In their redshift test they included 82 galaxies from the Frei sample and investigated redshift effects up to $z = 3$. They implement no size correction in their redshifting procedure. Nevertheless, we see similar behaviour in both the bias and scatter, illustrating that the qualitative behaviour is insensitive to the details of our simulations.

These results indicate that our networks' measurements at high redshift are better behaved than MORFOMETRYKA, with MORFOMETRYKA having a broader range of values. While both methods show systematic biases, the reduced scatter and lower prevalence of outliers suggests one could more confidently correct high-z C and A values based on the trained networks.

The differences seen in this section are greater than might be anticipated purely due to SNR from the results in Sec. 2.4.2. This indicates that there are other factors affecting MORFOMETRYKA more than the networks. In the SNR tests, simple uncorrelated Gaussian noise was added to the original images. However, in the artificial-redshifting procedure, the galaxies are placed in an apparently empty region of a real CANDELS image. The resulting background is more realistic, containing low-level structure due to pixel covariances introduced during the reduction and faint background galaxies. This could affect the asymmetry and background calculations by MORFOMETRYKA, especially at higher redshifts where the background subtraction becomes more significant. Our networks, which we have shown to be less susceptible to noise, appear to be less sensitive to these effects. The result is more stable measurements, which may be applied

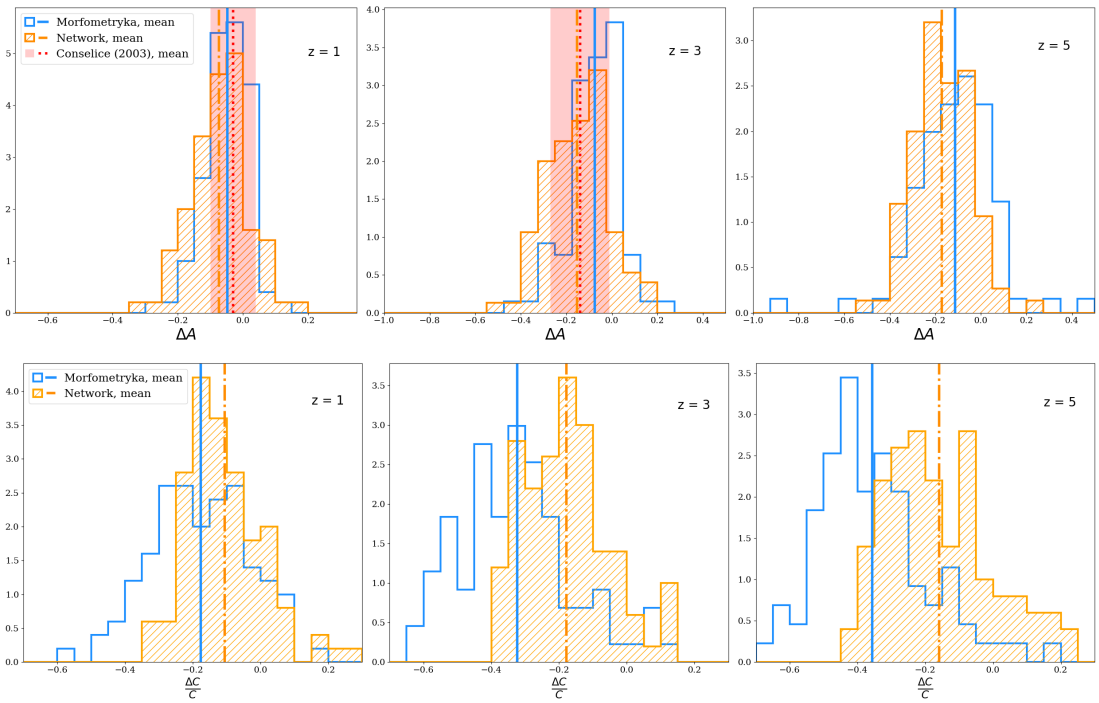


Figure 2.13: Investigating the effect of redshift on the asymmetry (*top*) and concentration (*bottom*) measurements from both our networks and MORFOMETRYKA. Plots show the distributions of the deviations in the measurements for a sample of redshifts. It can be seen that there are some extreme outliers in the measurements made by Morfometryka, especially with the asymmetry measurements. The range in the first asymmetry plot is reduced to better see both distributions. The original asymmetry measurements range from $0.01 < A < 0.48$, while the concentration measurements vary from $2.4 < C < 5.0$. There are a total of 100 galaxies from the Frei sample in this analysis. We also show the results from Conselice (2003) (the shaded region) demonstrating that our trends agree well with those found previously.

at high-redshift with greater confidence.

Looking at the concentration measurements, we see that the network performs significantly better than the standard algorithm at recovering the original C values. The network variations are, on average, a factor of two lower than those measured by MORFOMETRYKA. The average change in the concentration measurements between $z = 1$–3 is 15% compared to 26% for MORFOMETRYKA. However, as mentioned above, such systematic trends could be corrected. More importantly, the scatter in the network's measurements is somewhat lower and more consistent than those measured by the standard algorithm.

Our networks have been trained on CANDELS data, but successfully applied to data with a simple noise-degradation, and to artificially-redshifted, ground-based data. The individual images cover a very wide variety of appearances, and yet we recover reliable measurements across our test set. This indicates that our networks are not particularly sensitive to the details of the observations. They can be applied to roughly comparable datasets with similar performance to standard methods, without the need for retraining. The primary reason for this flexibility, is that both C and A measurements are determined from an image alone, without requiring any other information, such as the PSF, noise characteristics, etc. The network has learned to calculate a statistic from the image pixel values, irrespective of the observational details. It is therefore expected, but still pleasing to see, that the networks remain accurate when applied to a wide variety of images.

2.4.4 Computational efficiency

We now briefly turn to the efficiency of our CNNs, compared to CAS measurements using conventional measurements. Running both our trained networks and MORFOMETRYKA on a single computational core, for comparison, our CNNs are able to produce measurements $\sim 3,000$ times faster. However, for a modern workstation, containing a single high-end consumer GPU (e.g. an NVIDIA GeForce GTX 1080 Ti) and 16 CPU cores, the results are even more striking. On such a system, our trained networks can analyse $\sim 10,000$ galaxies in under 1.5 seconds, while it would take 2 hours to perform these measurements using the MORFOMETRYKA code. Thus, our networks could measure all 1.5 billion resolved galaxies in the Euclid survey (Laureijs et al., 2011) on a single machine in a little over an hour. To do the same with MORFOMETRYKA would take several weeks on a 1000-CPU cluster! Even with highly-optimized software, using

conventional algorithms would require significant time on a computing cluster.

In the previous section we have argued that our networks may be applied to other datasets without needing to retrain, providing the data characteristics are reasonably similar (which will be the case for any intermediate- to high-redshift galaxy surveys). However, should retraining be deemed necessary, this need not be an onerous process. In Sec. 2.3.4 we show that the network performance is consistent for moderate variations around the optimum. We expect that the selected hyperparameters will be a suitable choice for a variety of datasets. There should be no need to rerun the Bayesian Optimisation process again.

Given the performance we see for our networks, only a few tens of thousands of galaxies would be required to retrain the network. Using the optimal architecture, fully training the network with 75,000 training examples takes only around 30 minutes. Transfer learning is also a possibility, but the training time should be no longer. In any case, the network training time is short compared to that required to prepare the training set, which itself is substantially faster than applying conventional methods to a large dataset.

A more general argument in favour of moving towards deep learning techniques for these kinds of calculations, is that there is potential for many, currently required, preparatory steps to be avoided. Pre-processing steps such as creating segmentation maps and cleaning neighbouring objects could, in principle, be performed by the network itself. In this work we have not explored this, and have instead applied our networks to the data prepared for Morfometryka. However, an indication of the networks' robustness is provided by its stability when applied to artificially-redshifted galaxies. This could significantly reduce the computational and human time spent preparing the data to run these measurements.

It should be noted that Morfometryka performs a number of additional measurements that are complementary to those discussed in this work. However, one could train a network, in the same manner presented here, to predict these parameters. Indeed, this has already been done for Sérsic profiles (Tuccillo et al., 2018).

A further outstanding issue is that of uncertainties. We have not attempted to produce uncertainties on individual measurements output by our networks, beyond examining the scatter relative to Morfometryka. However, Pearson et al. (2021) give a detailed explanation of how the estimation of uncertainties can be incorporated into CNNs.

Extending our network to measure a wider variety of parameters, with uncertainties, is beyond the scope of this present work. However, we hope that this work demonstrates the benefits for upcoming 'Big Data' surveys, such as Euclid and Rubin-LSST. Deep-

learning has the potential to improve over conventional approaches in terms of the efficiency, accuracy and flexibility with which the next generation of surveys can be analysed.

2.5 Summary

In this work we trained two convolutional neural networks to perform concentration (C) and asymmetry (A) measurements based on individual galaxy input images. Our trained networks reproduce measurements by standard algorithms with an average absolute error on the C and A values of 0.15 and 0.045, respectively. These are lower than the average uncertainties on those measurements using conventional methods. Our networks can therefore be used to measure these quantities with a similar level of confidence to existing algorithms. Analysing these quantities for large samples of galaxies can provide an estimate of the merger fraction, and help us understand the transition from peculiar/irregular galaxies at high redshift to the well-defined Hubble sequence we observe locally.

We have shown how both our networks' and MORFOMETRYKA's measurements are impacted by noise, but find that our networks' estimates are more stable in the low signal-to-noise regime, in terms of both lower scatter and systematic bias. By artificially-redshifting a sample of local galaxies from the Frei catalogue, we investigate trends in the measurements due to redshift effects. Again, we find that our networks produce measurements with a lower level of random variation, compared to the conventional algorithms. While the measured A and C values are slightly biased at high-redshift, our networks and MORFOMETRYKA are both affected in similar manner, and consistent with behaviour seen previously (Conselice, 2003). Furthermore, the systematic offsets are comparable to the random uncertainty on individual galaxy measurements, and so relatively minor.

Our trained networks are up to several thousand times faster than previous non-parametric measurement algorithms, presenting a substantial advantage for upcoming surveys. Our trained networks are made public with this work [2]. CNN-based approaches are more efficient and, as we have shown for measuring CAS parameters, can be more accurate and reliable than traditional measurements. Measuring non-parametric morphologies in upcoming galaxy surveys, including those by the Euclid, Rubin, and Roman observatories, will greatly benefit from the methods presented in this work. In addition,

[2] https://github.com/cbtohill/CASNET

the high accuracy of our CNN-based measurements make them equally suitable for use on smaller samples from deeper surveys, such as those by JWST.

Chapter 3

A Robust Study of High-Redshift Galaxies: Unsupervised Machine Learning for Characterising morphology with JWST up to $z \sim 8$

3.1 Galaxy Morphology

With the successful launch of the James Webb Space Telescope (JWST) we have access to the highest resolution imaging of these distant galaxies, allowing us to explore the high-redshift regime in the greatest depth, and detail, to date. This opens a window into better understanding the formation of the first galaxies and their evolution over cosmic time. There have already been a number of studies investigating the morphologies of these most distant objects (Ferreira et al., 2023; Huertas-Company et al., 2023; Guo et al., 2023). While these studies focus on the morphology of these distant galaxies, they still characterise them using nomenclature used at low-redshift, investigating fractions of spheroid, disk and irregular type morphologies. While some galaxies at high-redshift fall into one of these groups, it is not known if this is applicable to all galaxies we observe in the distant universe. Nor is it known what features are important in concluding what morphological group a galaxy falls into.

The problem that we investigate is how do we robustly classify these distant galaxies into self-similar types? How do we determine what features of a galaxies morphology are most important in its characterisation? Previous attempts to solve this issue involve

citizen science projects such as Galaxy Zoo (Willett et al., 2013). These aim to amass a large number of visual classifications by asking the public to answer a number of questions about a galaxy's shape, color etc. (Bamford et al., 2009; Cardamone et al., 2009; Schawinski et al., 2014). However, there are a finite number of questions and features that each classifier is able to select. This functions well for galaxies in the local universe and up to z∼1, where the majority of galaxies fall into broad classifications of spiral, elliptical and irregular, however, this breaks down at the higher redshifts where the majority of galaxies lie in this irregular group. In order to better classify galaxies at high-redshift using this method there needs to be new questions and features available for each classifier to choose. The issue is these features are unknown, as there is no robust classification scheme in the distant universe. There is solution to this problem, and one that has become popular in recent years – machine learning. Work has been carried out by Walmsley et al. (2022a) that combines these visual classifications from Galaxy Zoo with machine learning, allowing for many more classifications, and also enabling researchers to locate anomalies within their datasets (Walmsley et al., 2022b). While these techniques have proven very successful, they still require labels to initially train the networks. By using unsupervised machine learning we can remove the need for any classification or labels.

In this work we utilise an unsupervised deep learning algorithm to extract the most dominant morphological features for distant JWST galaxies and separate them into self-similar types, allowing for a new broad classification system. The intrinsic properties of the galaxies within each group are investigated with redshift, mass, and star formation rates to provide new insights into the evolution of galaxy structure since $z < 8$.

This chapter is organised as follows. In §3.2 we introduce the imaging data and survey used in this work, along with our selection criteria. In §3.3 we detail the various architectures we explore in this work as well as the clustering algorithm used. We discuss our data standardisation process in §3.4 and the optimisation of our networks in §3.5. The clustering algorithm used in this work is detailed in §3.6. The resulting optimised network and results are included in §3.7, which also includes information about the extracted morphologies and clusters. We conclude with a brief summary of our main results in §3.8.

3.2 Data

3.2.1 JWST data

All of the images used in this project are from the Cosmic Evolution Early Release Science Survey (CEERS; PI: Finkelstein, ID=1345, Finkelstein et al. (2023)) public release fields (Bagley et al., 2023), imaged with the NIRCam instrument on the James Webb Space Telescope (JWST). NIRCam offers wavelength coverage from $0.6 - 2.3\mu m$ with a resolution of 0.031″/pixel, and from $2.3 - 5\mu m$ with a resolution of 0.063″/pixel.

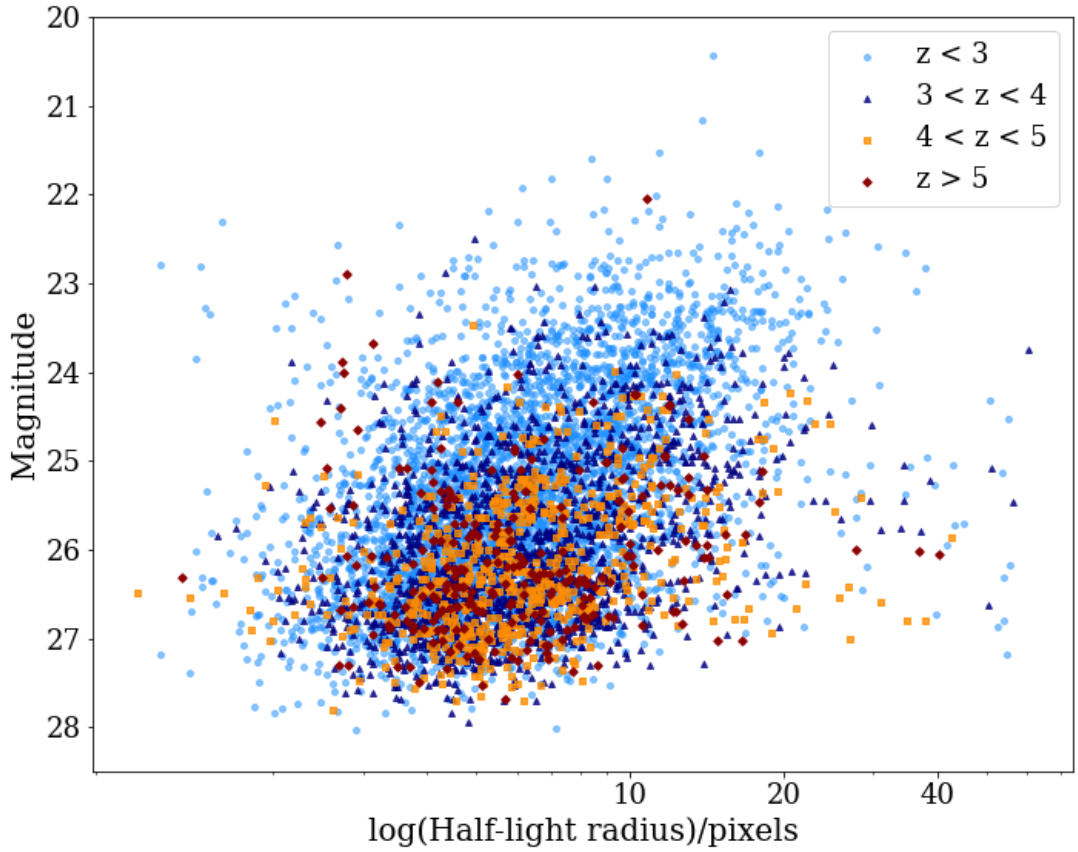


Figure 3.1: Distribution of the apparent H_{160} magnitude of our galaxy sample vs. half-light radius in pixels. The resolution from NIRCam is 0.03″ and 0.063″ per pixel for $0.6 - 2.3\mu m$ and $2.3 - 5\mu m$ respectively. Our sample is measured from the JWST CEERS imaging.

The data has been reduced using the pipeline mentioned in Ferreira et al. (2023). This is a modified version of the JWST official pipeline 1.6.2, see Ferreira et al. (2023); Adams et al. (2023) for more detail. We select galaxies that overlap with the Cosmic

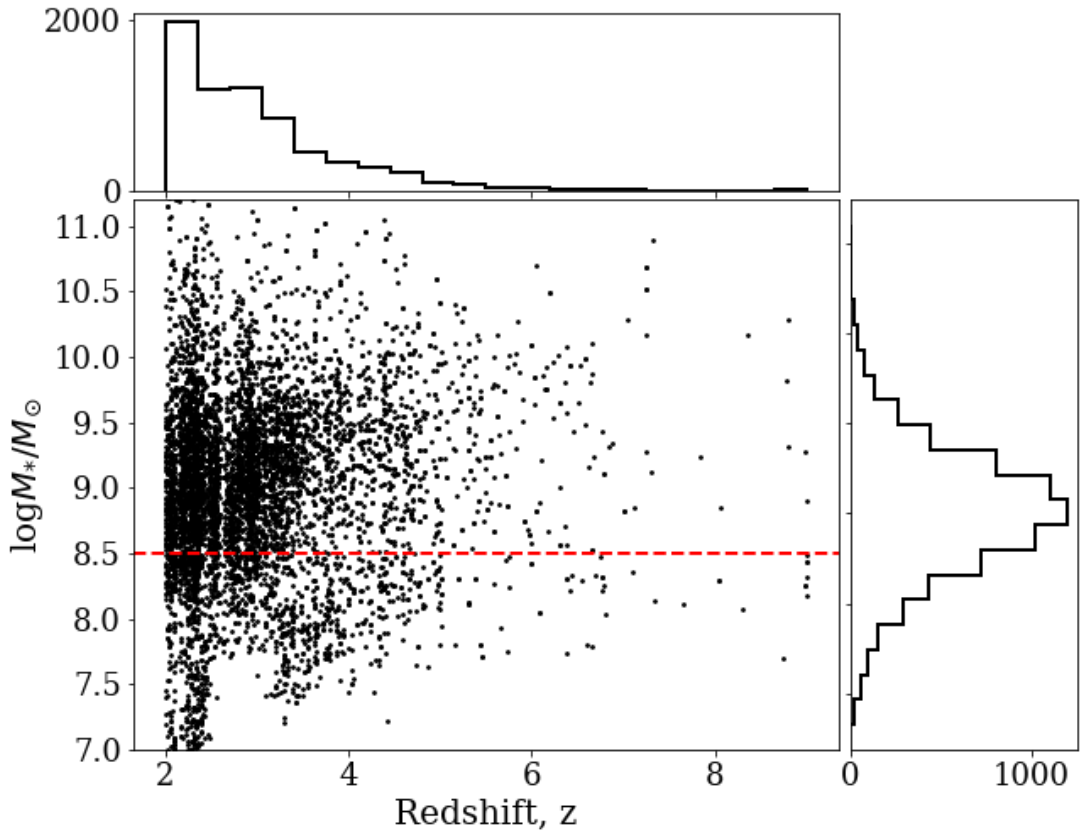


Figure 3.2: Distribution of the stellar masses of galaxies within our matched sample vs redshift. Red line shows mass cut at $10^{8.5} M_\odot$. Galaxies above this mass are utilised in this work which is above the completeness limit at all redshifts.

Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. (2011); Koekemoer et al. (2011)) for this work. In total we have ~ 10,000 galaxies from CEERS that overlap with sources in the CANDELS field across all redshifts. The reason for matching our JWST sample to HST imaging is so that we can use the reliable and well tested redshifts, photometry, mass and star formation rate (SFR) measurements that have been derived and utilised in previous works (Duncan et al., 2019; Whitney et al., 2019). Known AGN have also been removed. In total we select 6869 galaxies with $z > 2$ that have a match in the CANDELS survey. The apparent magnitude-size distribution of our sample is shown in Fig.3.1. While we will be using the redshift, SFR, mass and other measurements from the original CANDELS galaxies, we re-measure the non-parametric morphology with Morfometryka to have updated and more accurate CAS, Sérsic, Gini-M20 etc values. We expect these to change from the values measured off the HST images due to the increase in resolution and signal-to-noise of the data. It

should be noted that these measurements were performed on the original galaxy stamps from JWST before any standardisation, discussed in §3.4, was applied. As we want to probe the rest frame optical wavelength for all of the galaxies in our sample, we use imaging from the F150W, F200W, F277W, F356W, F410M, F444W bands and match to the redshift of each galaxy, see Table 3.1 for redshift cuts.

Redshift	**Band**	**Total**	λ_{pivot} (μm)
$2.0 \leq z < 3.0$	F150W	4186	1.501
$3.0 \leq z < 4.0$	F200W	1672	1.990
$4.0 \leq z < 5.0$	F277W	644	2.786
$5.0 \leq z < 6.0$	F356W	142	3.563
$6.0 \leq z < 7.0$	F410M	71	4.092
$z \geq 7.0$	F444W	42	4.421

Table 3.1: Summary of the bands utilised from JWST NirCam with the associated redshift that aligns with the optical rest-frame.

In order to ensure that the conclusions we draw in this work are representative of the sample we make a mass cut at $M_* > 10^{8.5} M_\odot$ to ensure we are complete at the highest redshifts. This can be seen in Fig.3.2. Galaxies above this limit are used for the analysis in this work.

3.3 Method

3.3.1 Machine Learning

In recent years Machine Learning (ML) has proven to be very successful in astronomy and has been applied to many different problems. ML has been utilised to predict many morphological parameters of galaxies from parametric measurements such as the Sérsic index (Tuccillo et al., 2018; Li et al., 2022; Tortorelli & Mercurio, 2023), to non-parametric structural measurements such as the *CAS* system (see Chapter 2). Supervised ML has also been successfully applied to visual classifications such as mergers (Ferreira et al., 2020), anomaly detection (Walmsley et al., 2022b), and Hubble type classifications (Dieleman et al., 2015; Domínguez Sánchez et al., 2018; Cheng et al., 2020b). More recently Robertson et al. (2023) has used deep learning to uncover the abundance of 'disky' objects at high-redshift with JWST. While these studies have proven to be successful, they require prior knowledge of the data in order to have labels to train your network on. This comes with its own issues, firstly you need to amass

enough labels to train your networks which is possible through citizen science projects such as Galaxy Zoo (Lintott et al., 2008). However, as these classifications are made by humans they come with their own intrinsic biases due to the subjective nature of the classifier. When we use these labels to train ML networks we are propagating this bias forward into any future classifications as well. With the future of astronomy consisting of "Big Data" surveys, it will take hundreds of people years to classify all of the 1.5 billion resolved galaxies in the Euclid survey (Laureijs et al., 2011). One solution to remove this bias and to improve the efficiency of these classifications is to move towards using unsupervised machine learning techniques.

3.3.2 Unsupervised Machine Learning

As the name suggests, unsupervised machine learning techniques require no labels to train but use only the data that you are interested in investigating as an input. For this reason unsupervised methods can be a more robust and unbiased method for data analysis. There have been studies in recent years that have applied unsupervised techniques to different problems in astronomy such as strong gravitational lenses (Cheng et al., 2020a), anomaly detection (Baron & Poznanski, 2017; Margalef-Bentabol et al., 2020) and galaxy morphology (Hocking et al., 2018; Martin et al., 2020; Cheng et al., 2021). These studies work by training a network to perform feature extraction on input data to recover the main features of your image. These features can then be explored and analysed, allowing you to perform different tasks such as grouping similar images together (a classification type analysis), finding outliers in the data (anomaly detection), looking for correlations between features and physical properties (morphology studies), etc. As the issue we are trying to address is a classification-type analysis we will need to group similar extracted features together. To perform our feature extraction we explore the use of variational autoencoders (VAEs) (Kingma & Welling, 2013) which we describe in detail below.

3.3.3 Variational Autoencoders

In this work we utilise a type of autoencoder (AE) network. The main idea behind an autoencoding network is that of dimensionality reduction. Dimensionality reduction is the process by which the number of features needed to describe some data are reduced. An AE is composed of two main components, the encoder and the decoder (see Fig.3.3). The encoder takes an input, which in this example is an image, and

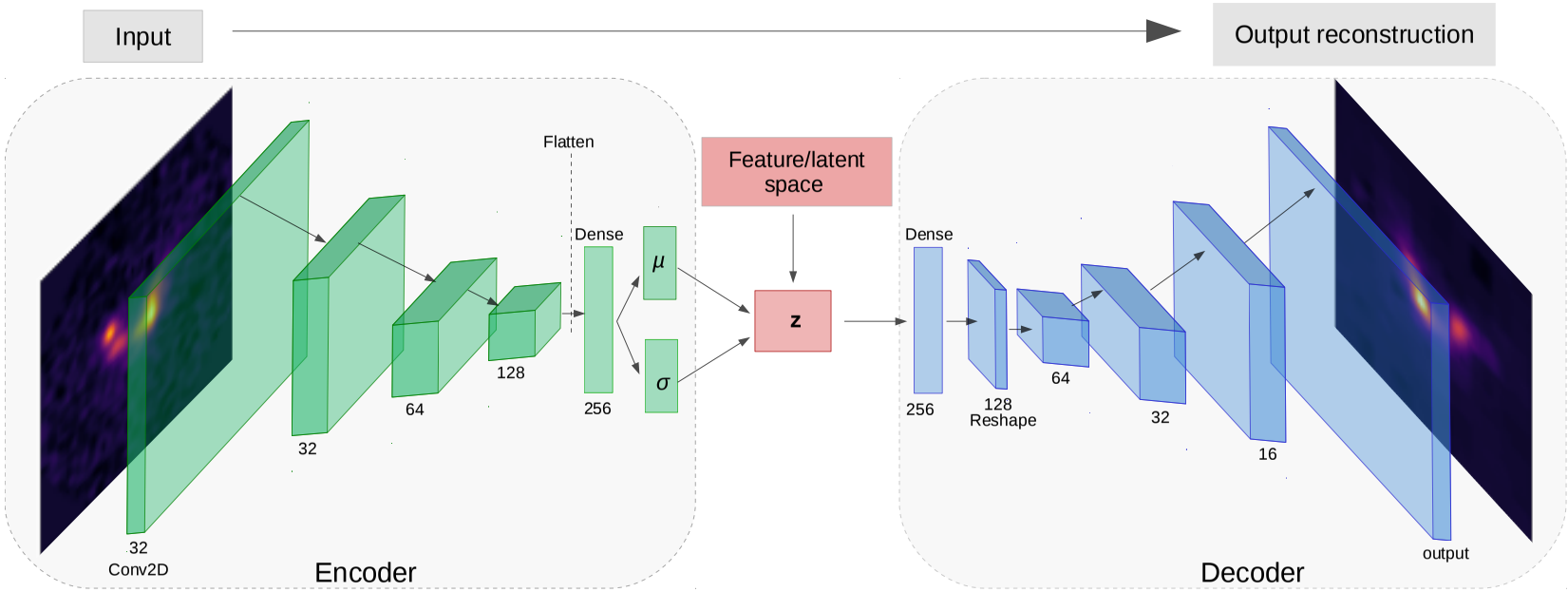


Figure 3.3: Architecture of our MMD-VAE network. On the left is the encoder, this takes the input data and compresses it into a smaller number of dimensions represented by the feature space (z). The decoder then samples from this feature space to create a reconstruction of the input. The aim of this work is to cluster this feature space to extract groups of galaxies with similar morphologies.

encodes the information into a lower dimensional representation of the data. This lower dimensional representation is stored in the latent space which is also referred to as the feature space. These will be used interchangeably throughout this work. The decoder then samples from this latent space to create a reconstruction of your input. The AE is trained to compress your input data whilst minimising the reconstruction loss between it and the output, decoded image. One downfall to AEs is that they are prone to overfitting as there is no regularisation of this latent space. In order to combat this, we can use a VAE (Kingma & Welling, 2013). The main idea behind a VAE is that instead of encoding each extracted feature into one number, it is instead encoded as a distribution that the decoder can sample from to recreate your input. This forces the latent space to be smoother which can also allow for generative processes (i.e creating mock galaxy images). During the training of these networks the reconstruction loss is minimised the same way as before, however, there is an extra penalty on the network if the latent space diverges from a standard normal distribution. This is called the Kulback-Leibler (KL) divergence. While the traditional VAE has been used in many studies with success (Thorne et al., 2021; Xu et al., 2023), it has been shown that the extracted features can be entangled and difficult to separate into distinct, differentiable features. This is an issue for our work as we want to be able to compare the network's features to known and well established morphological properties e.g the concentration of light, close pairs, asymmetry etc. There are a number of variations to the VAE that aim to address this issue of entanglement. Two of the more successful variations are the β – VAE (Higgins et al., 2017), and the MMD-VAE (Zhao et al., 2017). While Locatello et al. (2018) state that perfect disentangled representations are theoretically impossible, imperfect disentangling of features is an extremely useful concept and has been shown to be a very powerful tool in many disciplines (Chen et al., 2016; Higgins et al., 2017; Eastwood & Williams, 2018). We investigated both network architectures in our work to determine the optimal structure for our problem which we explain in detail below.

3.3.3.1 β-VAE

The β-VAE architecture was first introduced by Higgins et al. (2017). This type of VAE incorporates a weight on the KL loss to improve the disentanglement of features in the latent space. A value of $\beta = 1$ represents the original VAE, a $\beta > 1$ forces a stronger constraint on the latent space to learn a more efficient latent representation of the data. The idea is that if there are some features of the data that are independent of each other then the network will be able to better disentangle them, leading to a more

robust representation of the data. The loss is defined as;

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{KL}} \tag{3.1}$$

where

$$\mathcal{L}_{\text{recon}} = \mathbb{E}_{q_\phi(z|x)} \log p_\theta(\hat{x}|z) \tag{3.2}$$

$$\mathcal{L}_{\text{KL}} = \beta D_{\text{KL}}[q(z|x)||p(z)] = \beta \frac{1}{2} \sum_j 1 + (\log \sigma_j)^2 - \mu_j^2 - \sigma_j^2 \tag{3.3}$$

Where $p_\theta(\hat{x}|z)$ is the likelihood of the data $\hat{x}$ given the latent space z and $q_\phi(z|x)$ is the posterior distribution of your latent space. The aim is that the network will reduce the reconstruction loss between the input and the decoded data, while at the same time the KL divergence encourages the posterior to follow a distribution, normally a unit Gaussian. The reconstruction loss ($\mathcal{L}_{\text{recon}}$) in this work is simply the mean square error (MSE) between the reconstructed image and the input data. The network will be penalised for diverging from either of these conditions, as is the case with the traditional VAE. In the β – VAE there is an additional adjustable hyperparameter β that is introduced to the KL divergence term to balance this with the reconstruction loss. Higgins et al. (2017) showed that this additional hyperparameter is able to moderate the latent information and force the network to learn a more efficient representation of the data that is also disentangled. As we are interested in comparing the extracted features to known morphological features this disentanglement is important, however while the KL divergence can be moderated it still penalises the latent space diverging from a unit Gaussian. While this is useful for generative purposes as it creates a smooth latent space to sample from, it may not be best suited for our proposes as we are interested in retrieving distinct sub-clusters within the feature space in order to obtain a robust separation of galaxy types. With this in mind we explore another variation of the VAE known as the MMD-VAE.

3.3.3.2 MMD-VAE

The second network we investigate is the MMD-VAE. This type of network differs from the β – VAE in that it does not exploit the KL divergence but instead, finds the maximum mean discrepancy (MMD) (Gretton et al., 2008) between your prior distribution $p(z)$ and the posterior $q(z)$. The MMD of two distributions is minimised when they are identical. Instead of comparing the overall distributions like the KL divergence, it

samples from each distribution and compares the means of each sample. If these are very different it is unlikely that the two samples are from the same distribution. In order to sample from each distribution it uses the kernel trick. This allows non-linear data to be projected onto a higher dimensional space where it can be linearly divided by a plane. The MMD loss is thus defined as:

$$\begin{aligned}\mathcal{L}_{MMD} = \mathbb{E}_{p(z),p(z')}[k(z,z')] + \mathbb{E}_{q(z),q(z')}[k(z,z')] \\ -2\mathbb{E}_{p(z),q(z')}[k(z,z')]\end{aligned} \tag{3.4}$$

Where $k(z,z')$ can be any universal function. The most common being a Gaussian kernel which we utilise in this work. The reconstruction loss is the same as before and so we end up with the total loss function in our MMD-VAE:

$$\mathcal{L}_{\text{MMD}-\text{VAE}} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{MMD}} \tag{3.5}$$

The advantage MMD-VAE has over $\beta-$ VAE is that it is not penalised for diverging from a Gaussian distribution as the loss is defined by the moments of the distributions and not the density. This is better suited to our problem as we ideally want a latent space that is easy to separate into clusters which is more achievable when the latent space is less compact/dense as it is when you have a high β value in the $\beta-$ VAE network.
To fully compare both networks we optimise both and compare how the reconstruction loss varies between then for the same number of latent variables. This comparison is discussed in §3.5.

3.4 Data pre-processing

3.4.1 Observational bias - rotation invariance

One common issue that can arise with feature extraction is the fact that the network is trying to reproduce the input images with as few features as possible. This causes features such as shape, orientation, size and position to be encoded first as these will result in a smaller reconstruction loss than more finer details, as discussed in §1.5.3. In our work we want to address and remove these observational biases before trying to cluster our galaxies, thus allowing the feature space to be physically meaningful and without the risk of missing any other subtle features of the galaxies.

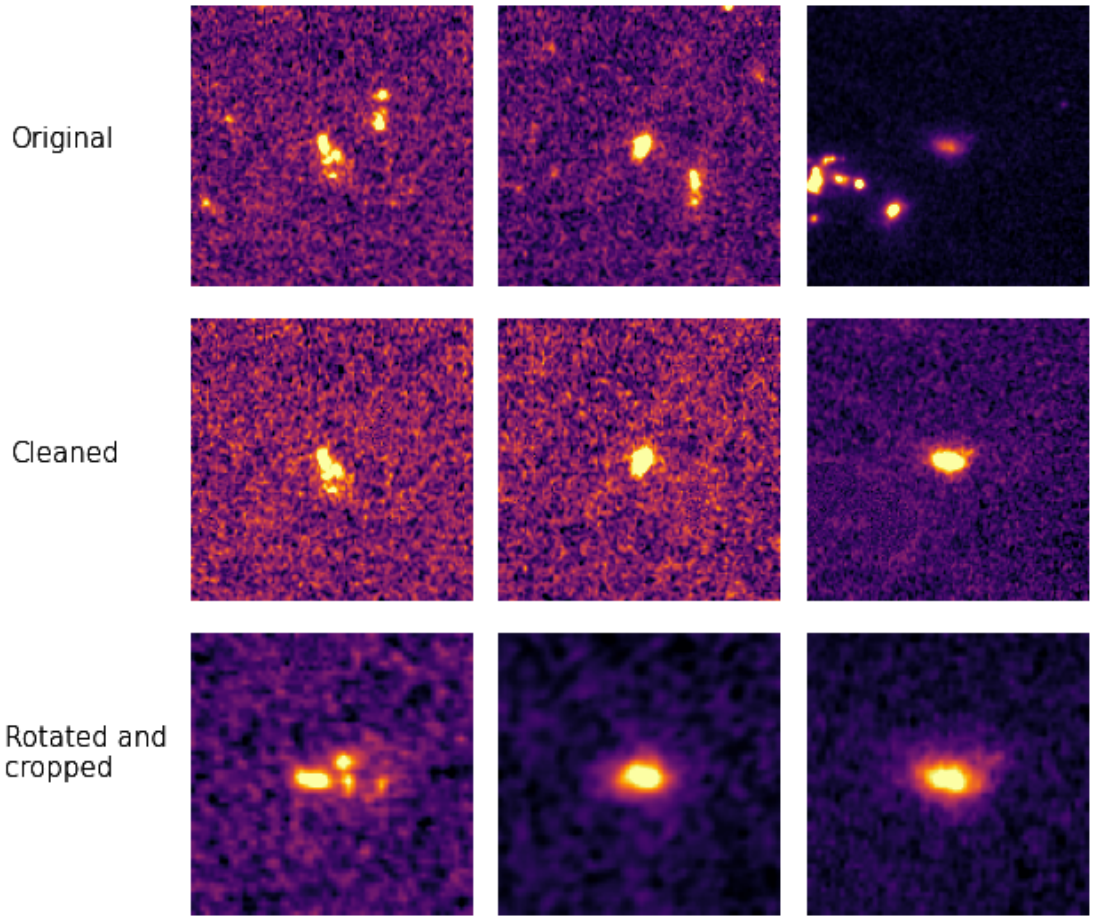


Figure 3.4: An example of our image standardisation process. *Top:* Original CEERS JWST images. *Middle:* This shows the galaxy images after they have been cleaned of background sources. *Bottom:* Re-scaled and rotated images.

3.4.2 Image Standardisation

To address these observational biases in our work we pre-process our data before training the network to prevent these features from being an issue, standardising our galaxy sample.

An example of this is shown in Fig.3.4. In the top row we have our original JWST images with the target galaxy in the center however, as you can see there are quite a number of background sources that as we have seen previously, a network will try to encode as a feature. We first remove these sources from our images using the galclean (de Albernaz Ferreira & Ferrari, 2018) algorithm. This algorithm removes any non-central sources at a certain threshold above the background level. These masked areas are replaced with values sampled randomly from the background distribution to ensure they do not leave shapes which could be picked up by the network. The clean images can be seen in the middle row. The next issue we address is the orientation of the galaxies. As it has been shown in previous works this is one of the dominant features to a network and so we rotate all of our galaxies by their position angle to prevent this from becoming an issue. The position angle is measured on the JWST images to ensure that faint objects do not bias the standardisation process. The last feature we address is the apparent size of the galaxies. We re-scale all of our images to the average Petrosian radius (R_p) of 15 pixels if they are smaller than this. This will allow the network to focus on the finer details of the images instead of wasting information encoding the size of the galaxies. We do not downscale our images as we do not want to lose any resolution as this will take away from the feature extraction and some finer features could be lost. As our galaxies are all at high redshift we also crop our images to remove as much of the background as possible. An example of this can be seen in the bottom row of Fig.3.4. We then use these cleaned images to train our network.

3.5 Model training and optimisation

Depending on the chosen network architecture, there will be a varying number of hyperparameters that need to be optimised. There are various methods to determine what set of hyperparameters will provide a suitable architecture for the problem being addressed. These range from the more basic random or grid search methods (Bergstra et al., 2011), to more advanced techniques such as random forest (Hutter et al., 2011). These methods, whilst being used successfully in the past, are computationally expensive

Hyperparameter	Optimum value	
	MMD-VAE	$\beta - VAE$
batch size	32	32
fully-connected layer size	256	128
number of filters	32	64
optimization	Adam	Adamax
β	N/A	0.01
λ	10	N/A

Table 3.2: Summary of the hyperparameters selected by the Bayesian Optimization technique.

and can take a while to converge on an optimum value. A more efficient approach to optimising network architectures is Bayesian Optimisation (Snoek et al., 2015). Bayesian Optimisation builds upon previously evaluated models to create a probabilistic model which is built upon to more efficiently converge on an optimum solution. The hyperparameters within our network are as follows: the batch size fed to the network during training, the initial number of filters in the first layer of the network, the number of dense filters in the dense layer on the encoder, the optimiser used, the value of β and λ depending on the network being trained, and the number of latent dimensions used to encode our data. The latter we will address separately, as simply by increasing the number of latent dimensions the loss from our network will decrease, so to force the optimisation process to focus on the architecture of the network we keep the number of latent dimensions fixed at five. We chose this value as it is large enough to let the network encode the main features of each image so to have a reasonable loss to optimise the network on, but not too large that we risk encoding noise that would cause variations each time the network is trained. The optimum network hyperparameters are shown in Table.3.2. For all networks trained the learning rate was reduced when the validation reconstruction loss had plateaued for a set number of epochs which is referred to as the 'patience'. The patience for our learning rate was 20 epochs and if no improvement was seen after 50 epochs (the patience for the reconstruction loss) the training was stopped. All networks were allowed to run until there was no improvement seen in the validation reconstruction loss.

3.5.1 Dimensionality of latent space

The main principle of a VAE is that of dimensionality reduction. Determining the optimum number of latent dimensions to encode the data into is another hyperparameter.

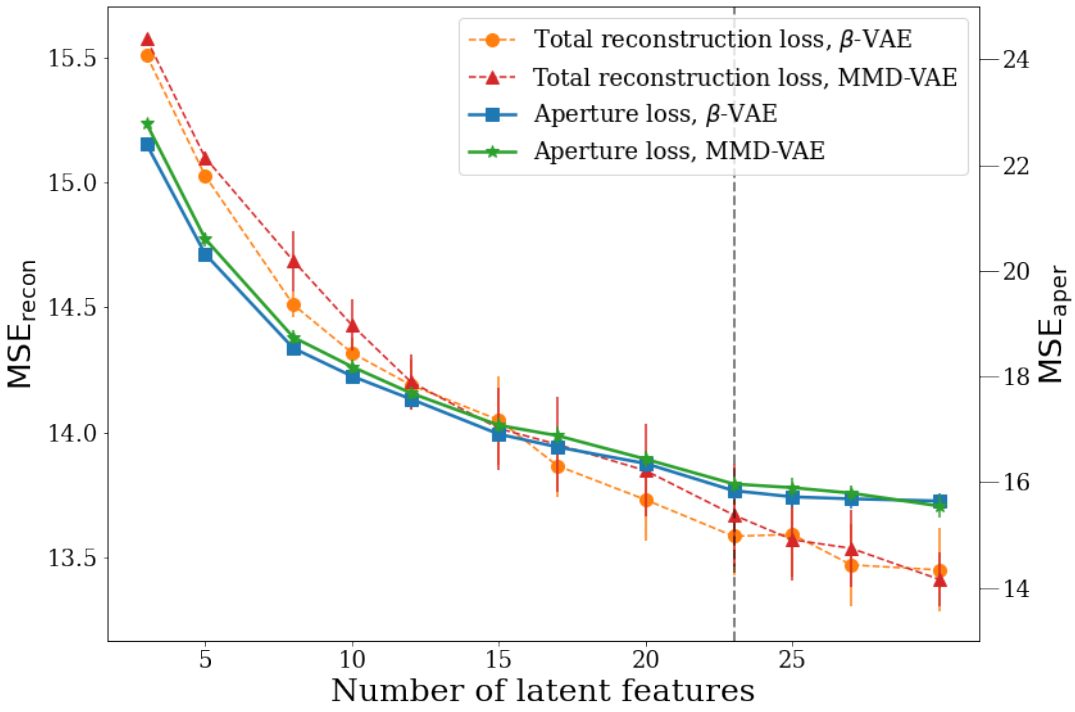


Figure 3.5: Variation of the reconstruction loss, both total and within an aperture, of the validation set from our networks vs the dimensionality of the latent space. Solid points show the average MSE for the last 50 epoches (our patience value, see text for details) and error bars show 1σ of the variation. It can be seen that after 23 latent dimensions there is no improvement in aperture loss within the errors, whilst the total reconstruction loss continues to decrease.

A small number of latent features and you do not adequately capture the morphological information stored in each galaxy image. Too high, and you are left with a latent representation that is large and hard to interpret, making the idea of dimensionality reduction meaningless. To determine a good balance between these extremes we test 12 cases varying the dimensionality of the latent space from 3 to 30. We use the reconstruction loss from our validation sample, which is simply the MSE between the reconstructed image from the decoder and the input image to determine the optimum number of latent features. We separate this loss into total reconstruction loss (i.e the whole image) $\mathcal{L}_{\rm recon}$, and the loss within an aperture of $R_p = 15$ pixels, $\mathcal{L}_{\rm aper}$, the average size of all our galaxy sample. The idea is that when the network has encoded the main features within each galaxy it will start to use information to encode the noise in the images. By exploring where the aperture loss plateaus we can select this to be the optimum number of latent dimensions to encode the morphological information of

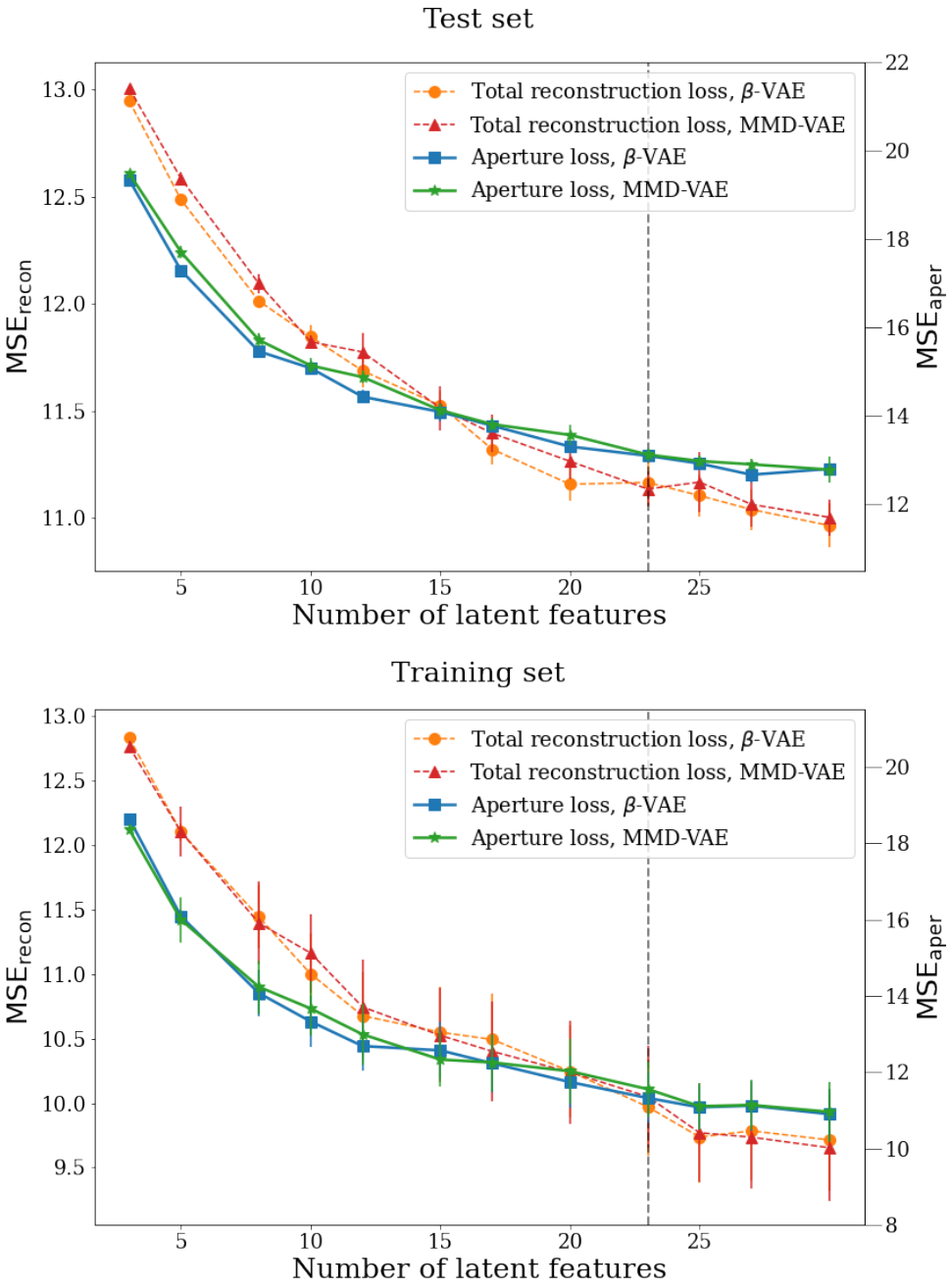


Figure 3.6: Variation in the total reconstruction loss and the reconstruction loss within an aperture (see §3.5.1 for details) for the test and training set (same as in Fig.3.5). Solid points show the average MSE for the last 50 epochs and error bars show 1σ of the variation. The dashed line shows the chosen dimensionality of the latent space for our analysis.

our sample of galaxies. The average variation in the loss for both networks is shown in

Fig.3.5. The average loss is calculated on the validation set during the last 50 epochs of training each network which is the patience value for our training (i.e. when the training plateaus and there is no more improvement in the loss), and the error bars show the 1σ variations in this loss. Both networks perform similarly both at encoding the whole galaxy image and within $1R_p$ for each galaxy, see Fig.3.6 for the training and test sets. For all latent dimensions investigated the total reconstruction loss continues to improve from a $\mathrm{MSE_{recon}}$ of ~15.5 to a $\mathrm{MSE_{recon}}$ of ~13.5, and continues to improve. However, it can be seen that after 23 latent dimensions there is no improvement within the aperture loss for either network. The $\mathrm{MSE_{aper}}$ of the reconstructions within the aperture plateau at around ~ 15.8 for both networks. This indicates that the network is utilising any extra dimensions to encode the noise in the image or background sources that might have been missed from the image cleaning process, which we also do not care about. This is also reflected in the larger error bars when we get to these number of latent dimensions. Thus for the analysis in this work we have a dimensionality of 23 for our feature space. We expected a smaller number of features would be needed to encode the main morphological structure of our sample as one of the main aims of this work is to minimise the latent space in order to ensure it is interpretable and physically meaningful. In earlier iterations of this work we utilised 5 latent dimensions to encode the data and found that, while the overall reconstructions were reasonable, the network tended to use individual latent features to encode a variety of physical features. This indicated that limiting the number of features to very low dimensionality limited both the networks reconstructions, and also limited us from mapping individual latent features to physical features. The fact that 23 latent dimensions are needed to well reconstruct our galaxy sample reflects the diversity of morphologies observed in the early universe at the resolution of JWST. This number of latent dimensions, whilst higher than first expected, is much lower than previous works that allow their feature space to get very large in order to achieve the best reconstructions possible, thus rendering their feature space un-interpretable and difficult to map to physical features. For the rest of the analysis in this work we utilise the MMD-VAE network as both networks have a similar performance however for the reasons stated earlier in §3.3.3.2, MMD-VAE has been shown to be better at disentangling features and puts less constraint on the distribution of the latent features, which is better for our problem.

3.6 Clustering

The aim of our work is to be able to separate our galaxies into different clusters based on their intrinsic features determined by our network. This is not a simple task, as we do not know the true number of clusters that exist in our data and so the question is, how do we determine how many different groups of galaxies exist in our data? To address this issue we explore a method known as hierarchical clustering, as no prior knowledge of how many 'true' clusters there are within the data is required.

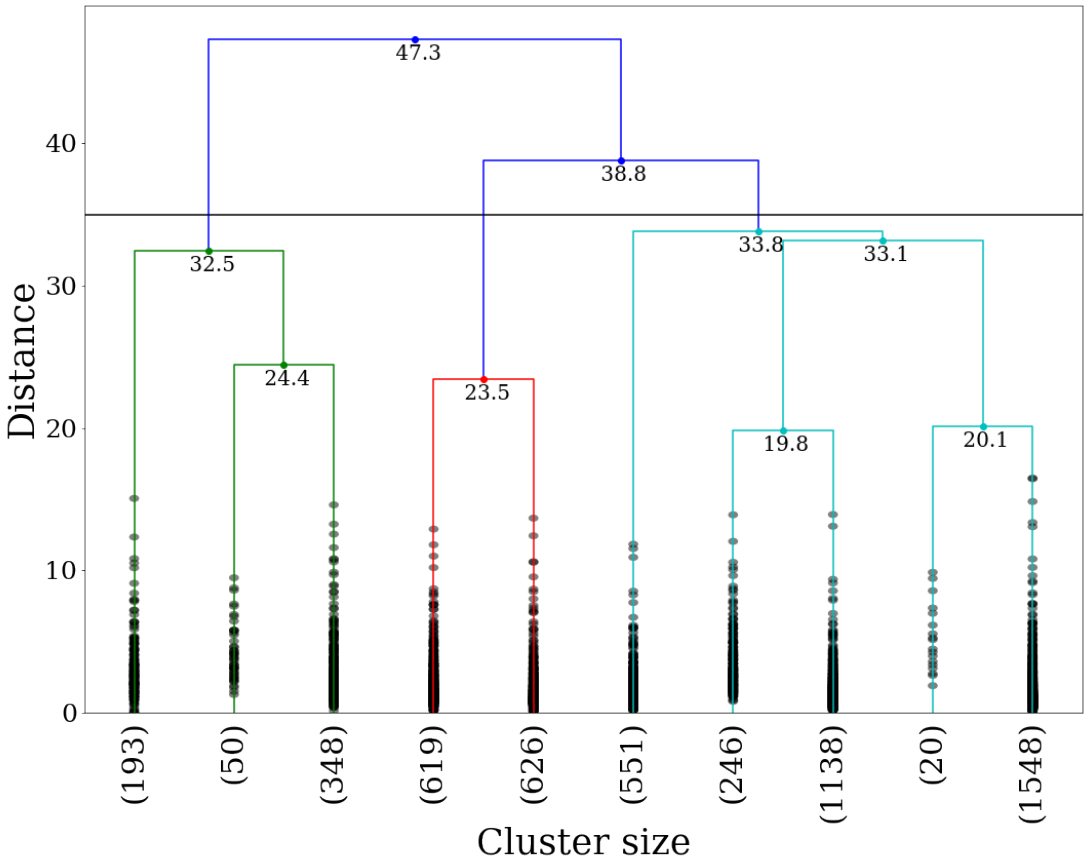


Figure 3.7: Dendrogram of the HC algorithm. The y axis shows the similarity or distance between the points while the x axis shows the number of data points in the smallest clusters. The horizontal black line shows the cut-off point in similarity that would result in 3 clusters.

3.6.1 Hierarchical clustering

Within this work we utilise the hierarchical clustering algorithm (HC) (Johnson, 1967). We focus specifically on agglomerative HC which can be thought of as a 'bottom up' clustering approach. This technique initially assumes that each point is its own cluster and will merge similar clusters together after every iteration until all data points are contained within one cluster. Clusters are merged or considered similar if they are close together in the feature space. This method allows for uneven cluster sizes and also uneven shaped clusters. This allows more freedom in the latent space which is a feature of the MMD-VAE that we utilise in this work. An example of a HC dendrogram is shown in Fig.3.7. This method has been used in similar studies for classifying different morphological classes at low redshift (Cheng et al., 2021).

In order to measure how similar two clusters are the Ward's linkage method is used. This method measures the variance within each cluster by means of the sum of squares within them, and aims to minimise this when grouping clusters together. The distance computed is thus the increase in the sum of squares when two clusters are merged. As this distance is minimised, the resulting clusters are created by grouping the closest points in our feature space together. In order to select clusters using this method, traditionally a single distance is used as a cut off point, selecting all clusters above this threshold. However, this is not applicable to galaxy morphology studies as different morphological types require more features to describe them than others, for example, spheroids require less information than mergers or spiral galaxies. In order to ensure that the clusters we extract are well separated and that we do not miss any due to a restrictive cut we explore splits down each branch in the HC tree until the split is due to the signal-to-noise of the sample or if the cluster has less than 50 samples, as this is $\sim 1\%$ of our sample and would not allow any meaningful analysis to be conducted in terms of redshift evolution or different SFRs. The extracted clusters from our trained feature space are discussed in §3.7.2.

3.7 Results

3.7.1 Image Reconstruction and feature extraction

We train our MMD-VAE network on a subset (80%) of the images in our sample and use a validation set (10%) to monitor the training to ensure the network is not over-fitting. We then compare the reconstruction loss between the validation data and a further

independent test sample (10%) to evaluate the performance of the network. We find a similar loss between these sub-samples indicating that the network architecture we have trained is robust. An example of some reconstructions from the network can be seen in Fig.3.8. The top row shows the input augmented JWST images, the middle shows the reconstructions from the network and the final row is the residuals between the input and the reconstructions. It can be seen that the network is able to encode the general morphology of each galaxy including shape, ellipticity, concentration and asymmetry of the light distribution, pairs and the clumpiness. This is a good sign as we do not see any orientation effects taking up any of the encoded information. We can also see the effect of information loss, our reconstructions are very smooth and have removed the noise from the images as well.

Looking into the latent dimensions individually we can investigate which features the network is encoding. We have 23 latent dimensions in total that represent the feature space. In order to better understand the encoded space we investigate if there are any correlations between known galaxy parameters, both parametric and non-parametric, and individual latent features. We compute the Spearman's correlation co-efficient between each latent feature and our measured morphology to better understand what is being encoded. The highest ranked latent feature and the corresponding measurement can be seen in Table.3.3. The full correlation matrix can be seen in Fig.3.9.

In order to better visualise this correlation we plot each correlated feature vs the latent dimension with which it had the highest Spearman's co-efficient. We split each feature into bins to better visualise where different galaxy types lie along each feature. These can be seen in Fig.3.10. Looking at the parametric and non-parametric properties we can see that they are well separated in each of these features which is important as it shows us that the network is learning features that are physically meaningful to our galaxy sample. We expect this from looking at the reconstructions from the network as they capture the overall morphology of our galaxies well. As a further demonstration, we show generated images from our learned latent space in Fig.3.11. The generated images show a smooth transition in morphology along each latent feature as expected with a VAE. We include examples for the latent features that had the largest Spearman correlation coefficients with measured physical and structural properties as shown in Fig.3.10.

We can see that while the sSFR does correlate with latent dimension 8 it is not as well separated as the other morphological measurements. This is to be expected as not all galaxies with high sSFRs will appear morphologically similar. It has been observed

Figure 3.8: *Top:* Input images to encoder. *Middle:* Reconstructed images from the test set using 23 encoded features. *Bottom:* Residuals showing how the reconstructions have encoded the galaxy light and not noise. It can be seen that the network is capturing the overall morphology of the galaxy sample well. Note that, whilst the main morphology of each galaxy is well reconstructed, some finer details are missed, such as the brightness of each clump. This is good for this work as we want to robustly classify different galaxy morphologies e.g. clumpy vs. not-clumpy morphology and not end up with many more groups split by small variations in the location of clumps as some previous works have been subject to. This will allow us to have a broader, more robust, separation of galaxy morphology.

Latent Feature	Correlated feature	Spearmans rank
Latent 1	Axis ratio	0.43
Latent 8	Asymmetry	0.49
Latent 7	Sérsic index	0.38
Latent 8	sSFR	0.19
Latent 17	Concentration	0.32

Table 3.3: The highest correlation value and corresponding galaxy property for each latent feature.

that in the high-z universe high star-forming galaxies can be quite compact, and do not always resemble the classic star-forming clumpy morphology we think of traditionally. This makes it difficult to separate out one feature to describe how star formation affects galaxy morphology and makes visually identifying high star-forming galaxies biased and incomplete. By clustering galaxies based on a combination of all their extracted features we can avoid any pre-defined assumptions and uncover high star-forming galaxies with many different morphologies. We explore the morphology of high sSFR galaxies in §3.7.7.

Another method for visualising how our latent space correlates with the measured properties of the galaxies is to apply further dimensionality reduction to our 23 dimension space and represent it on a 2D plane. To do this we utilise the Uniform Manifold Approximation and Projection algorithm (UMAP, McInnes et al. (2018)). The UMAP algorithm seeks to produce an embedding, which is a low dimensional projection of the data, that best preserves the topological structure of the manifold. The 2D projection of our latent space is used to help visualise how the network is learning different structures in our data but it is important to remember that the coordinates of the UMAP plane have no physical meaning and are not used for any clustering purposes in this work. The UMAP representation of our 23 dimension latent space for our galaxy sample is shown in Fig.3.12. Each UMAP visualisation is coloured by the median value of a different measured galaxy parameter. It can be seen that as you move across the UMAP space from right to left there is a clear correlation with increasing asymmetry and decreasing concentration and axis ratio. From this we can see that our encoded representation space correlates with well know measured galaxy properties as expected and hence can be use to group structurally similar galaxies together which is the aim of this work. This correlation is similar to that found by Vega-Ferrero et al. (2023). In their work they used a UMAP representation of their 1024 dimension space and showed how their space also correlates with many physical properties and how this can be beneficial when separating

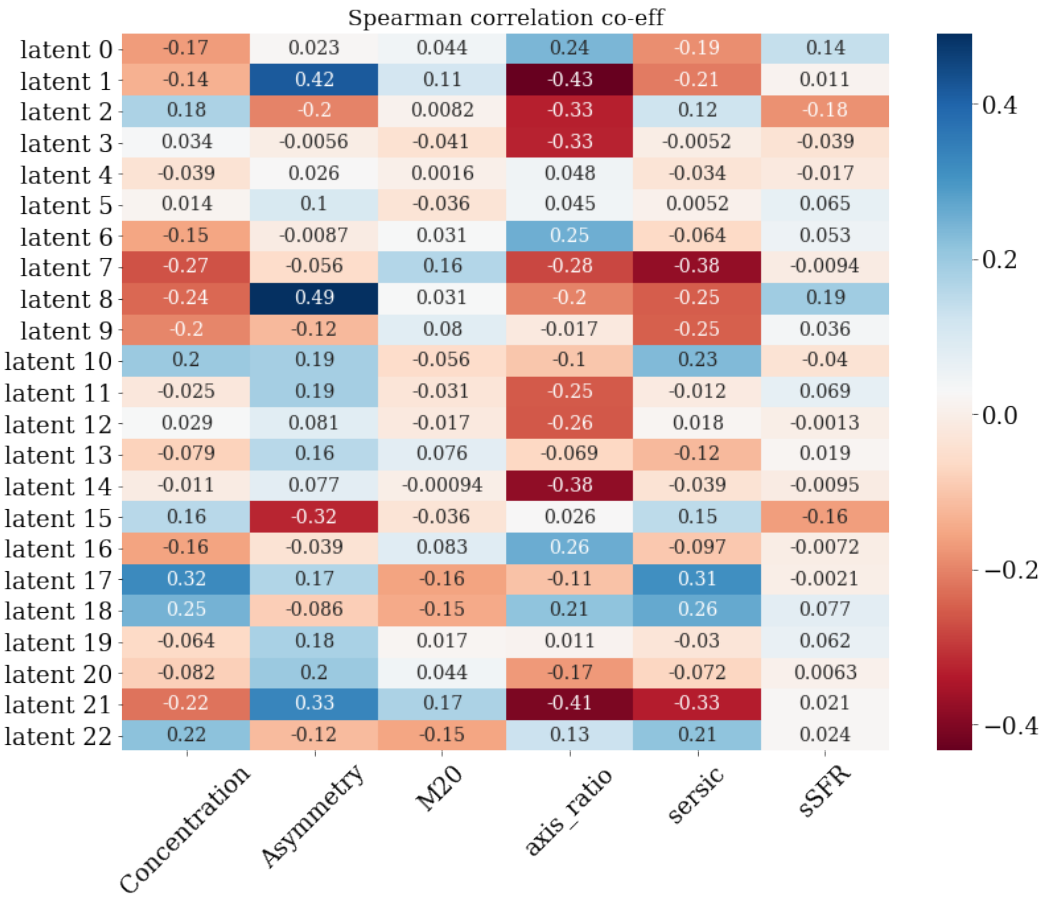


Figure 3.9: Distribution of the Spearman rank correlation coefficients for each latent dimension in our trained MMD-VAE and each measured galaxy property. Note that each latent variable is sensitive to a particular combination of particular attributes, for example latent feature 1 correlates with both asymmetry and axis ratio but not with concentration or M20. We can see that it is not a one to one correlation between each latent dimension and each physical parameter indicating that the network is using multiple latent features to encode each physical property. The fact that there are combinations of correlations between features and latent dimensions and not just random noise shows that the network is learning the different features of each galaxy, thus allowing the clustering of the latent space to provide different morphological types of galaxies.

different types of galaxies.

3.7.2 Extracted clusters

Running our HC clustering algorithm on our trained feature space we can investigate how the network has separated galaxies based on their morphological features and physical properties. We run the HC algorithm as explained in §3.6. We find a total of 11 clusters

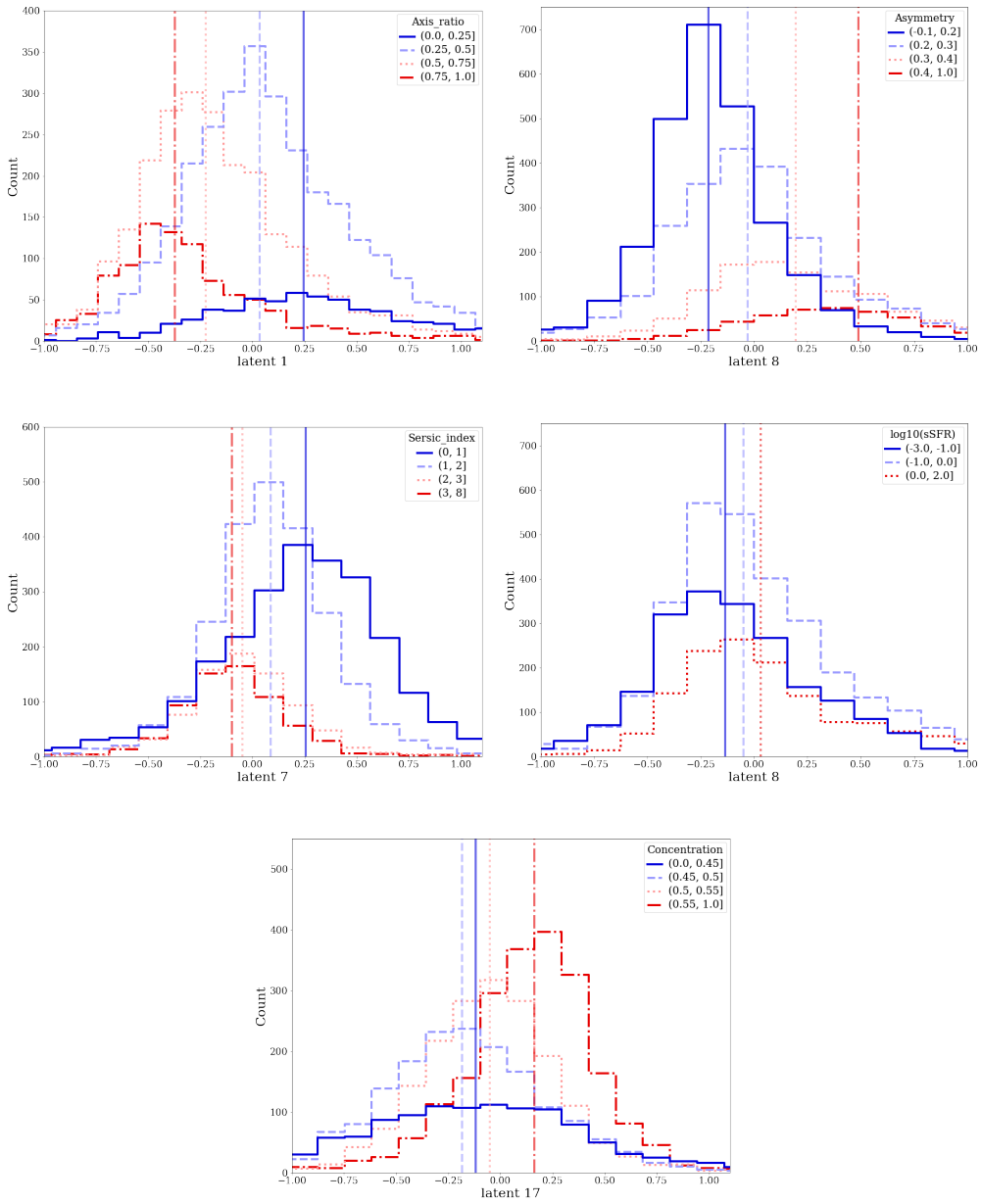


Figure 3.10: Correlation between features in the latent space and known measured properties for our galaxy sample. Each measured property is split into bins and plotted against the latent feature with which it had the largest Spearman's correlation coefficient. The median value for each selected range is also plotted. The ranges selected for each parameter can be seen in the legend. It can be seen that the features that the network is learning are physically meaningful and correlate with well known galaxy properties.

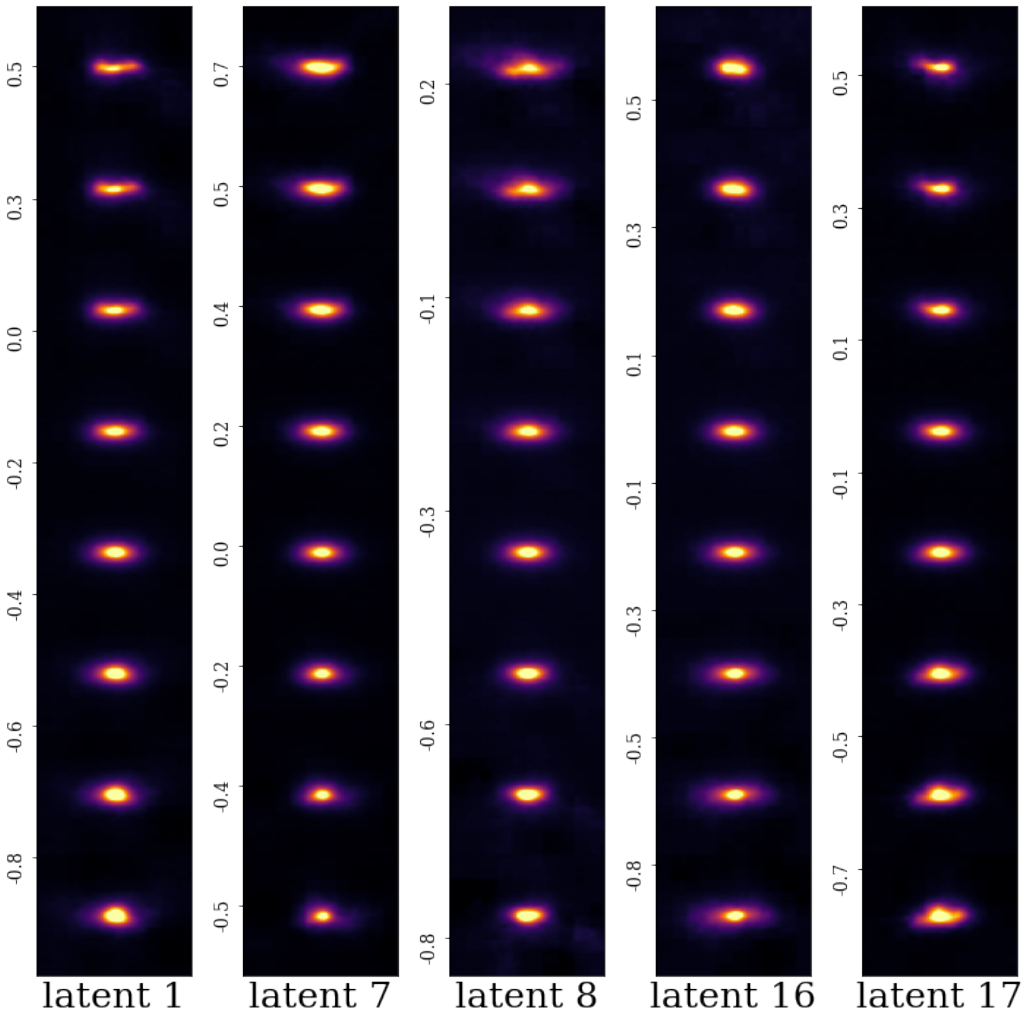


Figure 3.11: Generated images from the trained latent space. Here we show examples of generated galaxy images along different dimensions in the latent space to show the smooth transition between morphology. The latent dimensions shown are those with the highest correlation to measured properties (see Fig.3.10 for more detail.) We calculate the mean of each feature and linearly interpolate between $\pm 2\sigma$ whilst keeping all other latent features constant to show the morphology encoded by each dimension. Visually we can see how the first latent feature correlates well with axis ratio, while feature 7 correlates well with Sérsic index. Another important morphological feature is the bulge to disk ratio which is well encoded in feature 16.

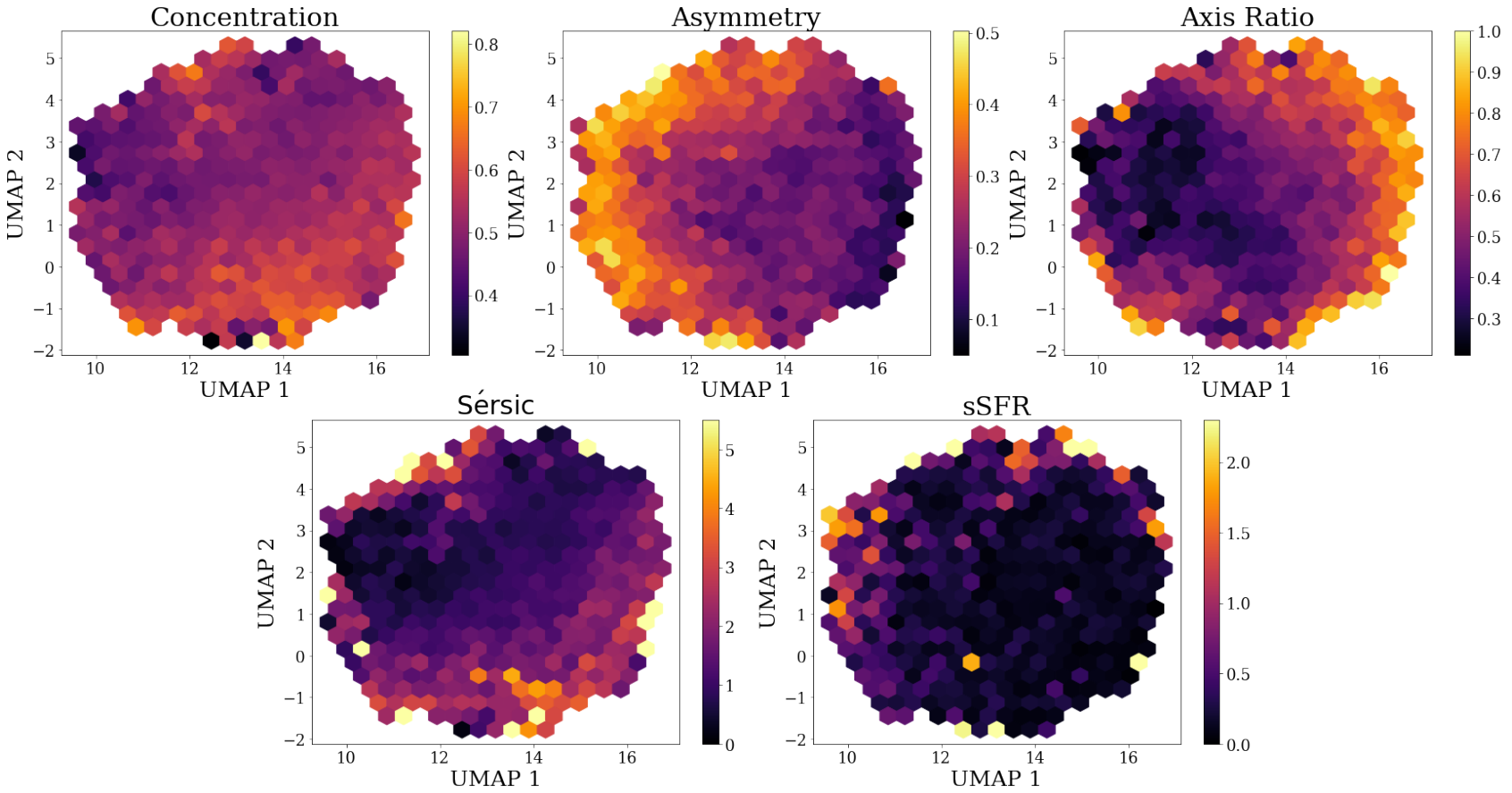


Figure 3.12: UMAP visualisation of our 23 latent features for our entire galaxy sample coloured by different measured morphological parameters. Each plot is coloured by the median value in each hexagonal bin in the 2D UMAP plane. From left to right shows the distribution of the concentration, asymmetry and axis ratio, Sérsic index, and the sSFR for our galaxy sample.

after removing the clusters that extracted the lowest SNR galaxies. These low SNR galaxies accounted for 5% of the sample and had an average signal-to-noise per pixel (SNR_p) of $SNR_p \sim 2$ (see §2.3.1 for calculation of SNR_p). The median SNR_p for the rest of our groups was $\sim$ 10 for comparison. Randomly selected cutouts of galaxies from the 11 clusters found can be seen in Fig.3.13. Visually inspecting the galaxies within each cluster we can already see how they are visually distinct from each other. Each cluster label indicates which main/parent branch it splits from in the HC tree. This parent branch is labelled as a super-group in Fig.3.13 along with the dominant morphology of galaxies along this branch. The super-group label is not designed to be a morphological classification as each super-group is composed of galaxies with distinct structural features however, it is simply one global feature that galaxies within this group have in common. This shows how the clustering algorithm has separated out different areas in the latent space as having very separate overall shape. Super-group 2.2 for example is labelled as *Centrally-concentrated* as each of the morphological classes within this group have a bright bulge like region however, group 2.2.1 is dominated by galaxy pairs or double cored so will possess structural parameters more associated with irregular type galaxies.

Comparing our machine extracted clusters to previous works that also utilise unsupervised machine learning to extract morphological clusters, Hocking et al. (2018) found 200 clusters when trying to classify morphologies from the CANDELS fields. Our method results in much fewer clusters allowing for a more broad classification on the main features each galaxy image possesses. It also allows us to explore the evolution in morphology with redshift and what the main morphological features are for each group more easily than comparing 200 separate clusters. It should be noted that their work does not include as high redshifts as ours, Hocking et al. (2018) includes galaxies up to z$\sim$4 and also includes low-redshift galaxies which will require more information to describe their morphological features simply due to the increase in resolution. This could account for the greater number of clusters they obtain. The evolution of our machine selected clusters as a function of redshift will be explored in a follow up paper.

Whilst our machine found clusters appear to be different galaxy types by eye we want to investigate their structural and physical properties to determine if this also correlates with their apparent morphology.

Group	No.samples	Morphological description	$C_{1/2}$	$A_{1/2}$	$M20_{1/2}$	$sSFR_{1/2}$	$Sérsic_{1/2}$
1.1	58	Edge on, multiple systems, chain, clumpy	0.44	0.29	-1.02	0.55	0.60
1.2	135	Face on multiple systems, clump clusters, mergers	0.44	0.37	-1.02	1.05	0.54
1.3	365	Clumpy disk-like, face on disk	0.48	0.26	-1.46	0.50	0.88
2.1.1	619	Edge on disks, clumpy, single objects	0.48	0.22	-1.40	0.14	0.77
2.1.2	171	Disturbed disks, tidal features	0.50	0.33	-1.32	0.59	1.06
2.1.3	455	Tadpole galaxies, asymmetry along semi-major axis	0.52	0.33	-1.74	0.29	1.90
2.2.1	112	Close pairs, doubles	0.47	0.37	-1.13	0.39	0.84
2.2.2	439	Disks with tail/tidal disruption, tadpole	0.50	0.32	-1.56	0.24	1.21
2.2.3	1138	Smooth light distribution, spheroidal, elongated	0.52	0.19	-1.58	0.19	1.45
2.2.4	922	Spheroidal, bulge dominated, centrally concentrated	0.57	0.18	-1.76	0.15	2.25
2.2.5	626	Bulge and disk component	0.56	0.21	-1.70	0.17	1.50

Table 3.4: Brief description of each group's morphology, number of samples in each cluster and median values for structural measurements. The horizontal lines show the super-group splits.

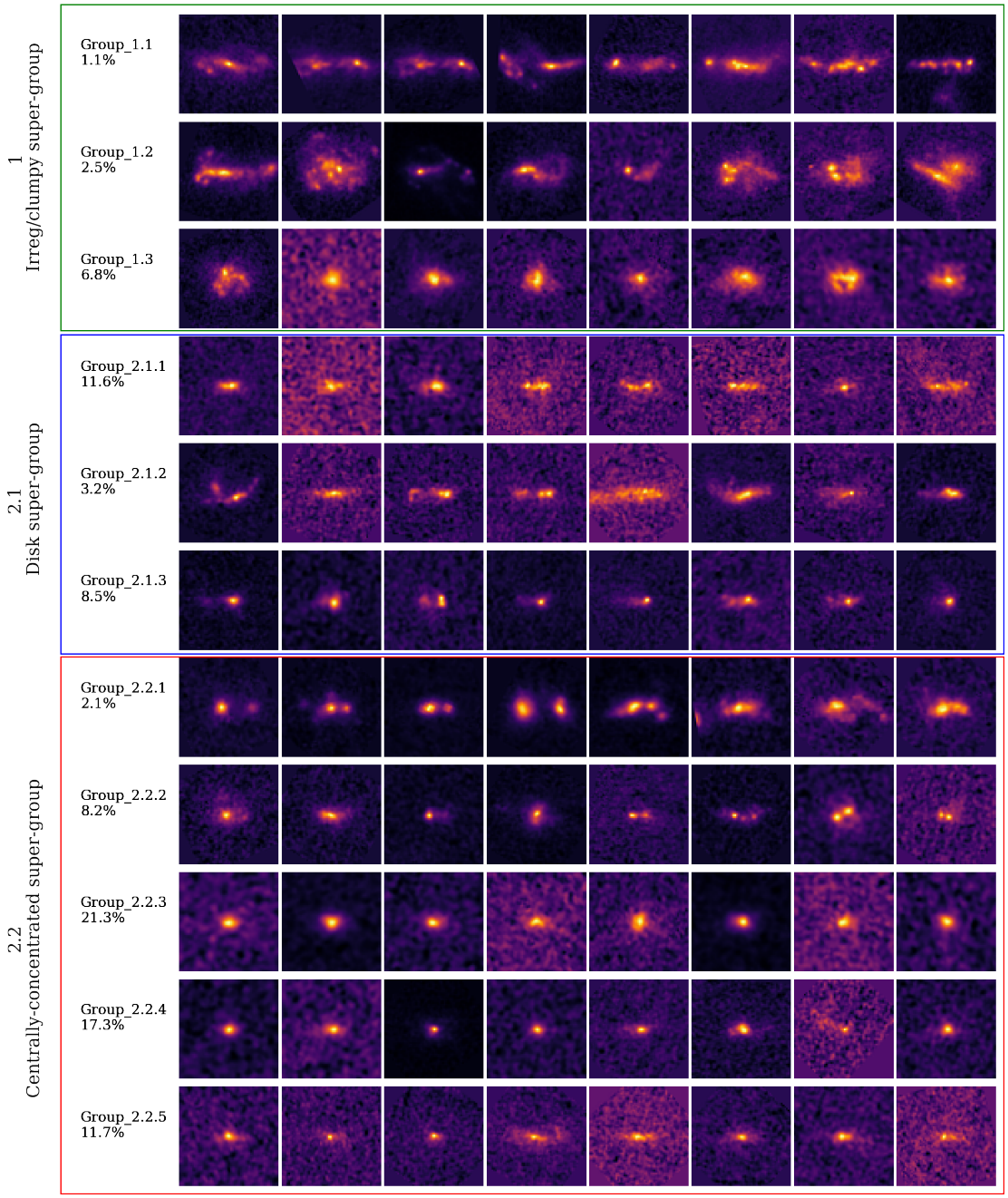


Figure 3.13: Examples of galaxies from each group found by the HC clustering algorithm. It can be seen that each group is visually distinct from each other. Group names indicate which parent branch they belong to from the HC tree, parent branches are labelled as super-groups along with the dominant morphology along that branch.

3.7.3 Comparison to Structural and Physical Properties

We use specific star formation rates (sSFRs), masses and redshifts from the original CANDELS galaxies as explained in Duncan et al. (2019) to ensure that the classifications are reliable, however we remeasure both the parametric and non parametric morphological parameters of each galaxy using MORFOMETRYKA (Ferrari et al., 2015) as these will benefit from the higher resolution offered by JWST allowing more accurate measurements compared to those measured on the HST images. MORFOMETRYKA uses the un-standardised images, along with the associated PSF images, to calculate these parameters. Investigating the distribution of these structural measurements will give a better insight into the physical properties of galaxies within each group, seen in Fig.3.14.

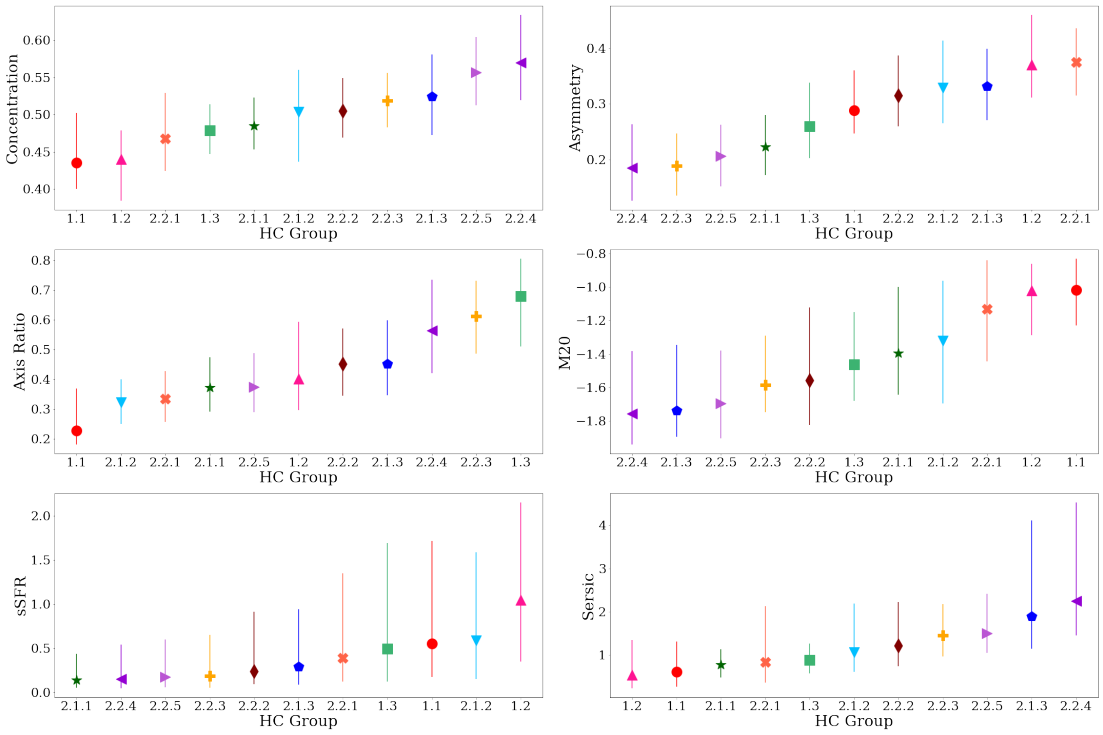


Figure 3.14: Distributions of the properties of the galaxies within each group determined from the hierarchical clustering algorithm. Each group is distinct from each other in both parametric and non-parametric morphology measurements and also physical properties such as the sSFR. Groups are ordered in increasing value within in measurement. Coloured points and symbols are the median and match across all figures. The error bars show 25th and 75th percentile of the distribution of values within each cluster. Note that the error on the mean is $\sim$ 20 times smaller, hence the trends we see are of high significance.

Fig.3.14 shows the median values of different physical and structural parameters for each

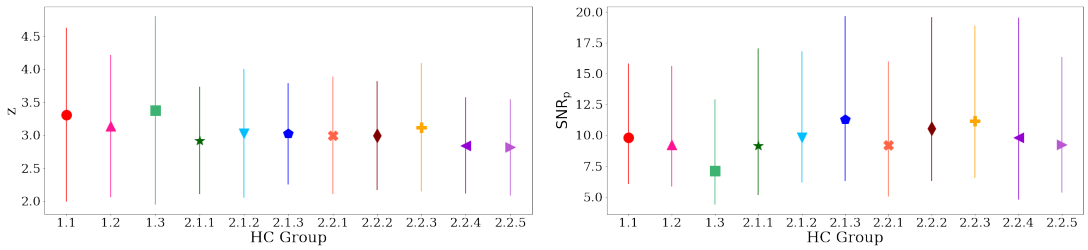


Figure 3.15: Left: Distributions of the redshifts of galaxies within each morphological cluster. It can be seen that the network does not split groups based on their redshift which further supports the claim that our standardisation approach successfully removes many observational biases such as apparent size/apparent brightness. This also allows further study on the evolution of galaxies within each morphological class with cosmic time. Right: Distributions of the SNR_p of galaxies within each morphological cluster. It can be seen that the network does not split groups based on their SNR_p. Coloured points and symbols are the same as the other figures in this work. Solid points show the mean and error bars show the standard deviation within each group.

of our machine defined clusters. The error bars show the 25th and 75th percentile of the distribution of values within each group. It can be seen that each of our morphological groups possess distinct physical and structural properties, even though these parameters were not used to perform the clustering. While the separation between the groups can be small considering the distribution within each group, there is still a clear transitions between the properties of the galaxies as you move between different morphological classes. Note that the error on the mean is $\sim$ 20 times smaller than the interquartile range. The groups have been ordered in increasing value for each parameter to better show this smooth transition. For example, looking at group 2.2.4, galaxies in this group possess the highest concentration of light from all groups, most symmetric light distributions, low sSFR and high Sérsic indices agreeing with what we would classify visually as dominated by spheroidal, compact, bulge dominated systems. The same can be said for Group 1.1 and 1.2 which possess almost the opposite properties, with asymmetric light distributions, low central concentration, and the highest sSFR of the sample, agreeing with our knowledge that star forming clumps would increase the asymmetry of the galaxy's light distribution. We also find no separation between our groups with redshift which indicates that the network is not splitting galaxies based on any redshift effects. This reinforces the claim that our standardisation process has removed the majority of observational effects from the latent space, allowing for a more robust classification scheme. The redshift and SNR_p distributions of our groups are shown in Fig.3.15.

Fig.3.16 shows where each cluster lies in the $C - A$ plane as well as the distribution in Sérsic index with asymmetry. Each coloured point shows the median value for each of our machine found clusters and the error bars show the interquartile ranges of the objects within each cluster. Clusters are well separated in the $C - A$ plane and show less, but still clear, separation with Sérsic index. Similar results have been found by Kartaltepe et al. (2023) when investigating the separation of their visually defined galaxy types. In their work they showed that defining galaxy morphology by a cut in Sérsic index would lead to a high misclassification rate between spheroids and disk galaxies due to the overlap in these measurements between different morphological types. This further emphasises the need for a more refined and robust classification scheme based on the morphological features of a galaxy and not on model-dependent measurements. Nevertheless Fig.3.16 shows a clear correlation between our clusters and the measured structural parameters of our galaxy sample. By using the average structural properties and visually inspecting galaxies from each group we can associate a morphological label to each which are listed in Table.3.4.
Below we give a brief description of the main morphological features of each cluster and compare these to previous studies of high-redshift morphology.

3.7.3.1 Group 1.1 - Chain galaxies

These galaxies resemble the 'chain' type morphologies first introduced by Cowie et al. (1995). They are very elongated structures with low axis ratios, dominated by a very elliptical shape that lacks a clear central bulge. Cowie et al. (1995) stated that these extreme ellipticities argue against the possibility that these are simply galaxies viewed from edge-on and are in fact their own class of peculiar objects. More recently however, work by both van der Wel et al. (2014) and Zhang et al. (2019) suggest that there is a population of star forming galaxies that are intrinsically more elongated/prolate at higher redshifts that transition to more oblate structures later on in their evolution. Zhang et al. (2019) also compare their results to simulations and find this to be consistent with predictions by Ceverino et al. (2015) and Tomassetti et al. (2016).

3.7.3.2 Group 1.2 - Clump clusters

Group 1.2 has similar properties to our chain galaxies (group 1.1), it however possess larger axis ratios giving rise to the argument that these galaxies are simply the face-on view of chain galaxies. This argument was first put forward by Elmegreen et al. (2004)

who named these 'Clump Clusters' and stated that the distribution of axis ratios agrees with what we see for normal disk galaxies. Comparing the properties of this group with our chain galaxies (group 1.1) we see that they possess very similar distributions except for their axis ratio and asymmetry. This increase in asymmetry is consistent with the fact that these objects are face on and hence have a larger projected area resulting in a larger asymmetry. Both groups also occupy a similar redshift range and mass distribution with a median mass for group 1.1 of $\log(\text{Mass}[M_\odot]) = 9.60^{+0.26}_{-0.38}$ and $9.64^{+0.50}_{-0.37}$ for group 1.2 (where the errors denote the 25th-75th percentile range of the distribution).

3.7.3.3 Group 1.3 - Clumpy disks

These galaxies, whilst possessing a slightly clumpy morphology, are disk-dominated, with a central bulge-like region. They possess the highest axis-ratios, resembling face-on disks, but have a higher concentration than the clump clusters (group 1.2) hinting at a more evolved morphology. These systems have intermediate asymmetries possibly due to the fact they have a high sSFR which leads to some clumpy features.

3.7.3.4 Group 2.1.1 - Edge-on Disks

Resembling edge-on disks with no central concentrated bulge region these galaxies are quite symmetric in their light distribution with intermediate concentration indicating no clear central region. They possess the lowest sSFR of our sample with low Sérsic indices. This lack of on-going star formation could indicate that these galaxies are more evolved with an outside-in formation that will perhaps form their bulge later in their evolution via means other than SF.

3.7.3.5 Group 2.1.2 - Disrupted disks

These objects are disk dominated, single object systems that possess disrupted morphologies that could be the result of possible mergers or tidal interactions. These objects have very low axis ratios with intermediate concentrations indicating that the central galaxy is disk like, however they have very high asymmetries and M20 values due to the tails and plumes caused by some interaction. As there is no companion visible, it is a possibility that these galaxies have been caught after a recent merger which could account for the disturbed morphology.

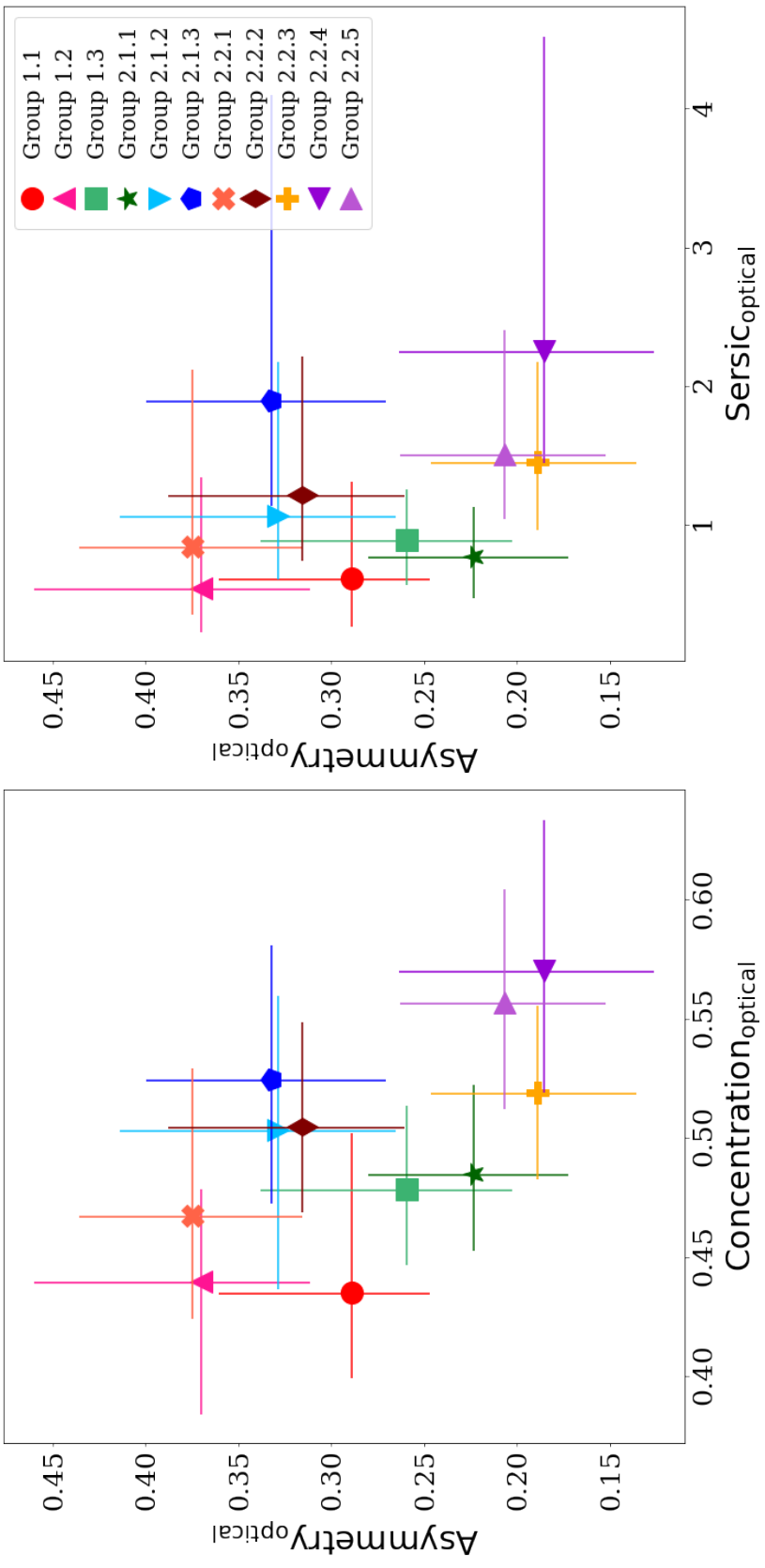


Figure 3.16: Distributions of the structural properties of the galaxies within each group determined from the hierarchical clustering algorithm. Each point represents one of our groups from our unsupervised machine learning clustering. Points show the median value and the error bars show the 25th and 75th percentiles of the values within each cluster. Symbols and colours as in Fig.3.14.

3.7.3.6 Group 2.1.3 - Tadpole galaxies

Closely resembling the 'Tadpole' type galaxy (van den Bergh, 1998) with an offset nucleus with a tail resembling the shape of a tadpole. These galaxies are very asymmetric along their semi-major axis with a range of axis ratios. The origin of this structure is not fully understood, some cases could be an offset burst of star-formation, a tidal interaction, ram-pressure stripping or accretion of cosmic gas (Elmegreen & Elmegreen, 2010).

3.7.3.7 Group 2.2.1 - Close pairs/Double clump

Visually dominated by 2 close objects or 2 main clumps these are the most asymmetric galaxies in our sample. Similar in morphology to many previous studies such as the 'Double nuclei in a common envelope' by Toomre & Toomre (1972) and 'Double-core' galaxies in Elmegreen et al. (2004).

3.7.3.8 Group 2.2.2 - Tail and Tadpole galaxies

This group possesses structural parameters very similar to our Tadpole group (Group 2.1.3) however the bright nuclei in this sample is to the left of the main object as opposed to the right with the previous group. This is due to our image standardisation process whereby each galaxy is rotated according to its position angle. This results in some tadpole galaxies being aligned 180° opposite to others. While this removes random orientations from our network to allow meaningful features to be encoded, we cannot avoid the separation of these clusters in the feature space even though they have the same distribution of physical and structural properties. This issue will be addressed in more detail in a follow up paper.

3.7.3.9 Group 2.2.3 - Elongated spheroids

This class is dominated by symmetric, high axis ratio, low sSFR galaxies with intermediate Sérsic indices. They are spheroidal in shape but elongated, closely resembling smooth disky objects. While some may classify these objects as disks visually, their physical parameters are more in-tune with what we would expect for spheroidal galaxies. This is discussed in more detail in §3.7.4.

3.7.3.10 Group 2.2.4 - Compact Spheroids

The most concentrated and symmetric light distributions of our whole sample, these galaxies match the classic compact spheroidal type galaxies as expected from visual assessment. These galaxies are similar in properties to our elongated spheroids (group 2.2.3), they however have higher Sérsic indices, and are more centrally concentrated. These also match well with the visual classifications which are discusses in §3.7.4 in more detail. They make up a total of 17.3% of our sample.

3.7.3.11 Group 2.2.5 - Bulge and Disk components

The final class of galaxy found by our clustering technique are edge-on disks with a clear bulge component. These are separate to the other edge-on disks in our sample as they have a centrally concentrated bulge component and a clear disk component as well. The fact we see these types of galaxies at high-z is an indicator that bulge and disk formation is already in place very early on.

3.7.4 Comparison to Visual Classifications

While the aim of this work is to provide a robust, non-biased approach to classifying galaxy morphology we want to compare how the networks classification system holds up against visual classifications. We have a total of 2619 visual classifications for our sample taken from Ferreira et al. (2023). While this is only $\sim$ 50% of our sample it should still give us an indication of the average galaxy type from each HC cluster. These can be seen in Fig.3.17, we also show the visual classifications within each super-group in Fig.3.18. While there is by no means a clear correlation between the visual classifications and our clusters, this is to be expected, as the classifiers only had a limited number of labels that each galaxy had to be classified into. These are the classic spheroid, disk, peculiar and PSF labels that have been utilised in many low-z studies. There was also an option for the classifiers to choose ambiguous if they felt that no label represented the object or if they were unsure, which account for 15% of the classifications in our sample.

Whilst we can see some correlations with the visual classifications, such as 70% of spheroid classifications reside in clusters 2.2.3 and 2.3.4, which agree with our assessment of these clusters and their structural parameters, and groups 2.1.1 and 2.2.5 are disk dominated which also agrees with our expectations there are few other strong correlations. The aim of this work is to show that these traditional 'morphological types'

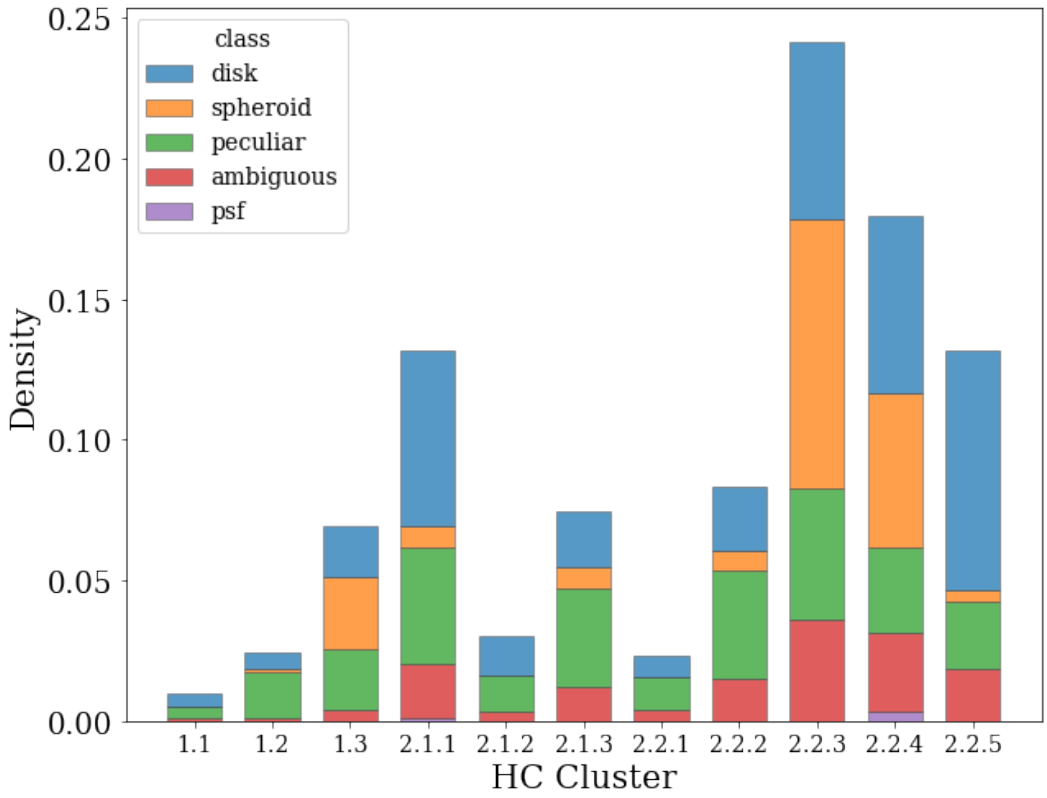


Figure 3.17: Barplot of visual classifications from Ferreira et al. (2023) within each of our HC found clusters. We have a visual classification for ~ 50% of our sample.

are not representative of the wide variety of galaxy morphology and structure we see in the high-z universe and that we need to better understand which features are important at better separating galaxies in various stages of their evolution. Ferreira et al. (2023) also state in their work that there can be issues with misclassification of face on disks with spheroids and so suggest a combination of visual classifications and structural parameters would help to resolve this issue which we are combining in the feature extraction process as the network has all information about the pixel by pixel light distribution of each galaxy. In recent work, Vega-Ferrero et al. (2023) found that a large proportion of visually classified disks perhaps lie in a region of representation space populated with spheroids. They compared their results with galaxies from TNG50 and found these regions are 'occupied by objects with low stellar specific angular momentum and non-oblate structure'. Looking at our clusters we find that group 2.2.3, while having ~ 45% of the spheroid classifications, also had a large proportion of disk like classifications. This could be due, in part, to the reasons stated by Vega-Ferrero et al. (2023) above. The properties of this group more closely resemble spheroid-like galaxies with high

concentration of light, low sSFR, low asymmetry and larger Sérsic indices.

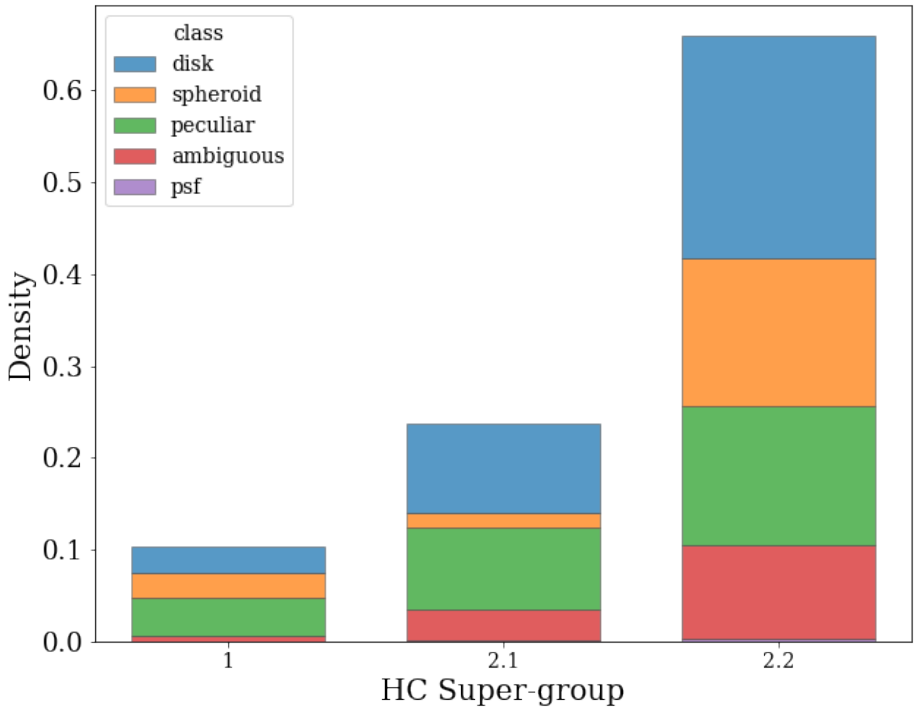


Figure 3.18: Barplot of visual classifications from Ferreira et al. (2023) within each of our HC found super-groups. We have visual classifications for ~ 50% of our sample. As it can be seen there are no strong correlations except for the centrally concentrated super-group (2.2) possessing most of the spheroid classifications. This further supports the argument that the visual classification systems used at low redshift are not suitable for the high redshift regime.

There is an argument for moving away from visual classifications all together as we now have more advanced techniques to separate galaxies based solely on their physical and measured features alone. With the increase in the amount of data being collected every day, visually inspecting each galaxy is also inconceivable. Measuring the properties of each galaxy accurately however, requires detailed spectral analysis which is time consuming and again would be inconceivable for the amount of data now available. Performing feature extraction on galaxy images unifies these two methods together and can be trained on a small sub-sample of images with detailed measurements and then applied on a much larger, un-labelled dataset, ideal for the future of galaxy surveys and galaxy evolution studies.

3.7.5 Impact of noise

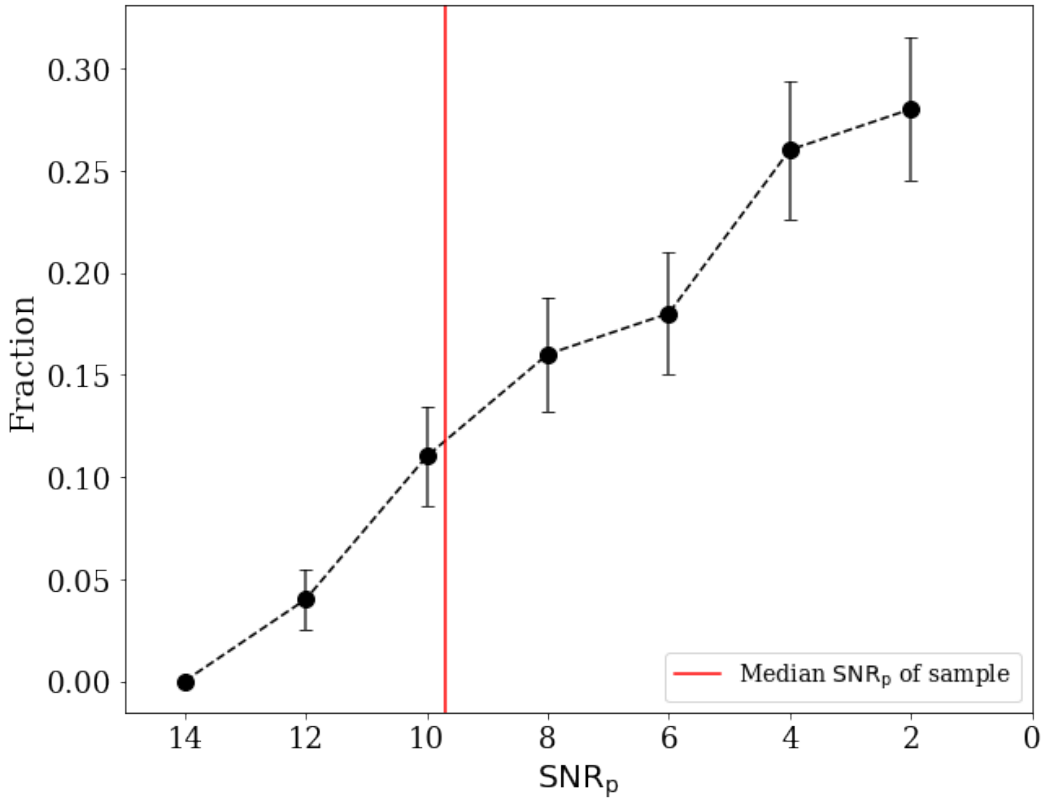


Figure 3.19: Fraction of galaxies that change classification in our machine-defined clusters vs the SNR_p of the galaxy. The errors are given as 1σ of the binomial distribution. The red vertical line shows the median SNR_p of our full sample.

One common issue with visual classifications and both parametric and non-parametric measurements is that they are sensitive to noise (van den Bergh et al., 2002; Ferrari et al., 2015; Gullberg et al., 2019). This is another motivation behind utilising UML to extract different morphological features, and why we optimise our latent space to focus on encoding the galaxy light and not the background noise (see §3.5.1). While the network is performing well at separating out distinct morphological clusters of galaxies, we want to evaluate how robust these clusters are to the signal-to-noise level of our images. To test this we select a sample of 15 high signal-to-noise ($14 < SNR_p < 20$) galaxies from each machine-defined cluster and simulate each galaxy at various SNR_p levels from 12 down to 2 (for details of how these images are simulated see §2.4.2). An example of the simulated noisy images can be seen in Fig.3.20. We assume for this test that these high SNR_p images are in their 'true' clusters, as the majority of their features

should be detectable by the network. We then run the simulated images through the trained network and the HC algorithm to check what fraction of galaxies remain in the same morphological cluster. The results are plotted in Fig.3.19. The black points show the fraction of galaxies where their classification changed and the error bars show 1σ of the binomial distribution. The first point is plotted at SNR_p of 14, and is at zero as we are assuming these galaxies to be in the correct cluster. The median SNR_p of our entire galaxy sample is shown by the red vertical line. It can be seen that at the median SNR_p of our sample, 87-91% of our galaxies remain in their original machine-defined cluster. This means that per group we would expect $\sim$ 1% to change class at this SNR_p level. As the SNR_p in the images continues to decrease we see an increase in the fraction of galaxies that change classification. Groups 2.1.1 and 2.1.2 had the largest fraction of galaxies change class, while group 2.2.3 remained the most consistent. This is to be expected as when the noise in an image increases you start to lose fainter features from the images rapidly. This means that, for example, galaxies defined as having tidal features in our classifications (group 2.1.2) would no longer be differentiable from the background noise by the network, and end up in a different morphological category defined by the remaining detectable features. This is not something that can be fully addressed in any classification scheme as this would also mean these features would no longer be picked up in a visual classification, parametric, or non-parametric measurement system. A good example of this can be seen in Fig.3.20, where the fainter structures in the galaxy become indistinguishable from the background noise around SNR_p=5. From this we can argue that the different morphological clusters found by our method are robust to various levels of noise however, below a $SNR_p < 6$ there is less certainty of the morphological classification from this method. For reference, 82% of our sample has a $SNR_p > 6$ so we can be confident in the majority of our morphological classifications.

3.7.6 Evolution of Massive Galaxies

As this work covers a large redshift range we can start to explore the evolution of the different clusters with redshift. To investigate these trends we plot the evolution in the fraction of galaxies in each cluster splitting into 4 bins: $2 < z < 3$, $3 < z < 4$, $4 < z < 5$, and $z > 5$. In order to ensure any conclusions drawn are not affected by incompleteness we look at the most massive galaxies ($\log M_*/M_\odot > 10$) within our sample across the whole redshift range. This can be seen in Fig.3.21. The groups are split by their HC super-group for ease of visualisation and so it is easier to compare

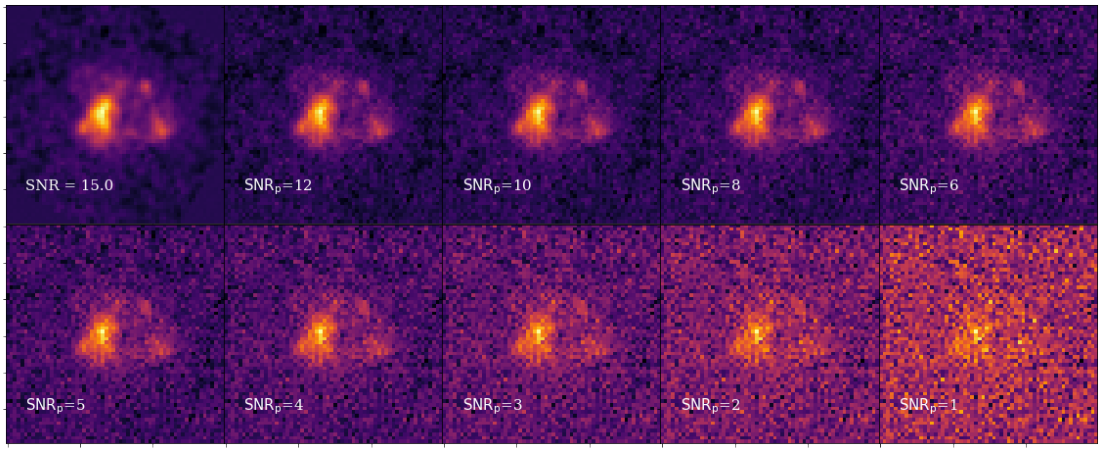


Figure 3.20: An example of our simulated noisy galaxy images. In the top left panel we show the original JWST cutout followed by the same galaxy at different simulated SNR_p levels.

different groups. We have a total of 507 galaxies above this mass and so at the highest redshift bins we can be affected by small sample statistics which is reflected in the errors at these values. Consequently the last bin includes all galaxies above $z > 5$ as there are not enough objects at this redshift to split this further. For this reason there is less certainty of the evolution after $z > 5$ and will need larger samples to study this in depth which will be possible in the near future with more data from JWST. Whilst there are not many galaxies at these masses we can still see some clear trends especially with group 2.2.4 (our spheroid dominated class) decreasing with redshift which agrees with results found in other works using similar data from JWST. In Huertas-Company et al. (2023) they found a decrease in early type/bulge dominated systems with redshift as did Kartaltepe et al. (2023) who found a decrease in spheroid only classified galaxies at the high masses (although they note this could be due to faint features being missed or small number statistics). When we reduce our mass cut to match the analysis in Kartaltepe et al. (2023) of $> 10^9 M_\odot$ and investigate our 2 spheroid dominated classes (2.2.3 and 2.2.4) we find that it decreases from $\sim 38\%$ at $z = 3-4$, to $\sim 32\%$ at $z > 5$, similar to their findings that their spheroid only class falls from $\sim 42\%$ at $z = 3$, to $\sim$ 30-40% at $z > 5$. These results are fairly similar even though our sample size is more than double the sample used by Kartaltepe et al. (2023) at this mass range. We also see an increase in groups 1.3 and 2.1.2, which are various types of disk dominated galaxies in our classification system but are distinct from each other in terms of being face-on disks and disturbed disks respectively. This result agrees with many works that find disks to

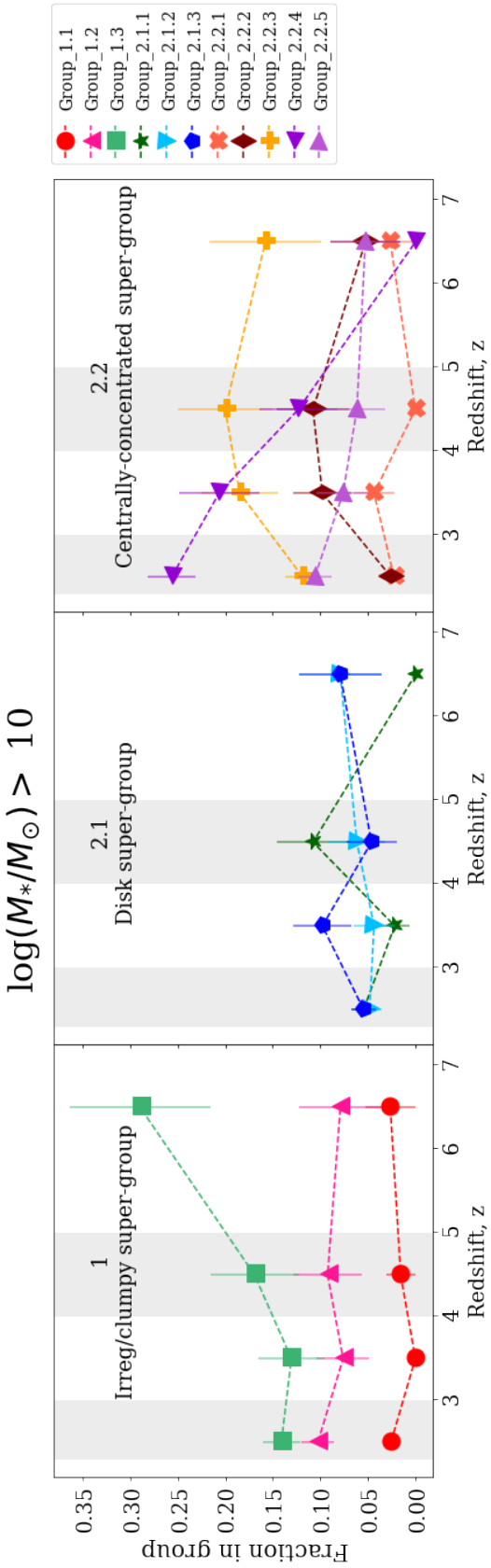


Figure 3.21: Evolution of the most massive galaxies within each morphological cluster with redshift. The errors are given as 1σ of the binomial distribution. Colours and symbols as in Fig.3.14 and Fig.3.16. The groups are divided by their HC super-group for ease of visualisation and comparison. The shaded regions show the bin widths for each point. The last bin includes all galaxies above $z > 5$. See text for details.

be dominant or already in place at high-z (Ferreira et al., 2023, 2022; Huertas-Company et al., 2023). We also see a growth in group 2.2.5 which are galaxies with a distinct bulge and disk component indicating that the process of forming a bulge with a disk component is already in place very early on agreeing with some of the main findings from Huertas-Company et al. (2023). We acknowledge that these findings are subject to small sample statistics and can be further explored and solidified with future surveys and more data especially at the high-z end but it is a good sign that even with the variety of morphologies that are found by exploring the networks extracted features agree with various other works using very different methods. We will study the evolution of these galaxies with redshift in more depth in a follow up paper.

3.7.7 Morphology of High sSFR Galaxies

We analyse the morphology of high star forming galaxies within our machine found clusters by making a cut at $\log \mathrm{sSFR} > 0.5$ as this is representative of our galaxy population in the high-z regime. We define sSFR as the $\mathrm{SFR}[M_{\odot} yr_{-1}]/\log(M^*/M_{\odot})$. This cut can be seen in Fig.3.22. While we are most likely missing a lot of the fainter and lower mass galaxies at high-z, we expect to be more complete for the high sSFR galaxies as this high star-formation rate is what allows us to detect even the most distant galaxies. We find that 4 out of our 11 machine-defined clusters have a median sSFR above this cut and so we can begin to investigate the variation in morphology seen for these high star-forming systems. The 4 groups are the Chain galaxies (group 1.1), the Clump Clusters (group 1.2), Clump Disks (group 1.3), and Disrupted Disks (group 2.1.2). Both the Chain and Clump Cluster groups possess galaxies with the highest sSFRs in our sample which is to be expected as it is believed the bright knots dominating their morphology are areas undergoing intense star formation (Cowie et al., 1995; Elmegreen et al., 2004). These star-forming clumps are picked up easily by structural parameters as these groups also possess the lowest concentration of light and Sérsic indices, and possess high asymmetry values. On top of that they are also the two highest M20 groups, showing that these systems could be picked up easily using traditional methods. The Clump Disk galaxy type in our classification scheme appear visually to be slightly messy, containing some clumpy morphology, which could indicate they are undergoing intense star-formation, however looking at the non-parametric measurements alone we find this group to have intermediate CAS and M20 values and so could be missed if selection cuts were to be made using these structural measurement systems. They are

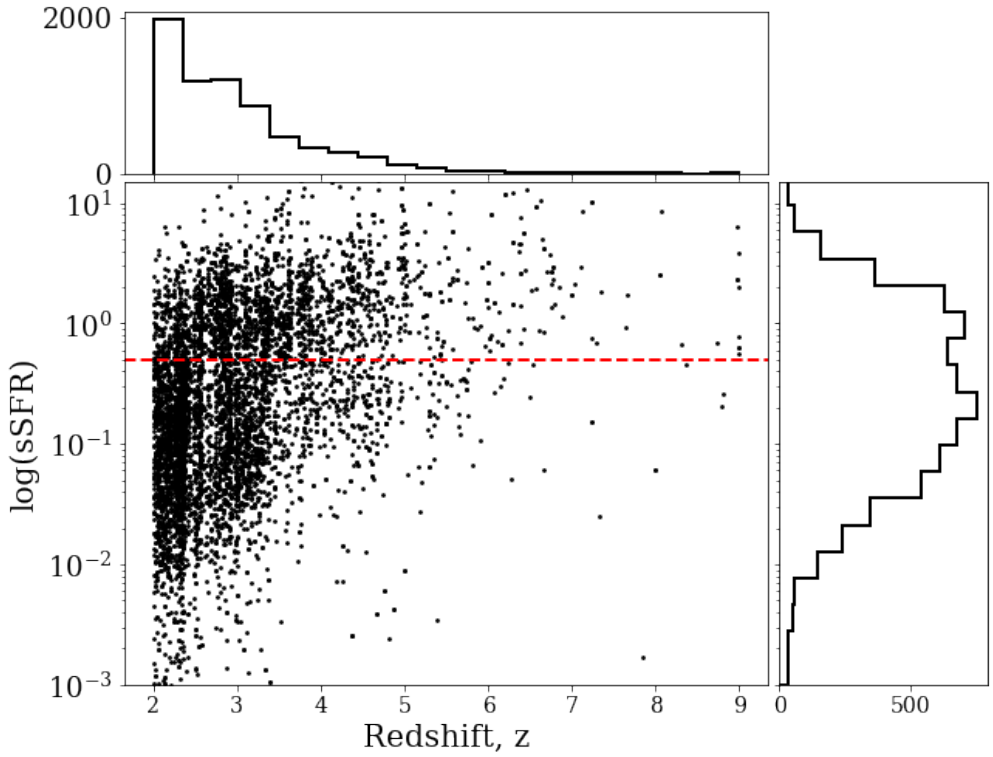


Figure 3.22: Distribution of the sSFR of galaxies within our matched sample vs redshift. Red line shows sSFR cut at $\log(\text{sSFR}) > 0.5$.

however found by our classification system, showing again that our network is able to pick up on features that are missed with current measurements. The final group we look at is the Disturbed Disks which, like our Clump Disk group, are asymmetric but possess intermediate concentration of light corresponding to a bulge component indicating that these systems are more evolved than the Chain and Clump cluster systems. These systems do not possess the classic star-forming clumpy morphology but instead have an asymmetry in the form of a tail or disturbed region. As these are all individual systems with no clear neighbour, it could be argued these galaxies have recently undergone a merger which could account for both the disturbed morphology, and corresponding increased star-formation. From these groups we can already see how diverse and varied the morphologies of these high sSFR galaxies are and why it is important to understand the processes that lead to this diverse structure if we want to better constrain their formation and evolution. We acknowledge that the classifications presented in this work do not allow us to confirm the formation history of these systems, however the ability to separate out these high sSFR systems with different morphologies for future, more detailed analysis and observations allows for an unbiased and robust selection

process. Again it should be noted that more detailed analysis of the evolution of these morphological groups with redshift will be carried out in a follow up paper.

3.8 Summary

We present our work utilising unsupervised machine learning to perform feature extraction on high-redshift galaxies imaged with JWST. We apply a hierarchical clustering algorithm to extract separate, self-similar, morphological classes of galaxies; resulting in a robust, more meaningful classification system of these objects. This is the highest redshift sample to date using this technique. Applying our method to optical rest-frame galaxy images imaged with NIRCam on JWST we find 11 separate morphological clusters that possess different morphological features, physical properties and structural measurements e.g. sSFR, CAS-M20 parameters, Sérsic indices and axis ratios. Our resulting clusters are devoid of human biases and would not be as well separated if classified with traditional nomenclature. We compare our findings with visual classifications and find that only spheroids are well separated in this traditional classification system and that disks and peculiar type galaxies need much more detailed descriptions. We improve upon previous studies using similar methods in multiple ways;

- We remove the observational biases imposed on the images by standardising our sample before performing any feature extraction. This allows the network to focus on the morphological features of the target galaxy without wasting information encoding the position angle of the galaxy, background sources, and noise, leading to a more physically meaningful feature space.

- Our method results in many fewer, well separated, morphological classes that can be investigated in detail which is not possible in some previous studies when hundreds of clusters are extracted.

- We have a relatively small feature space that can be investigated and linked to individual structural properties leading to clusters that are well separated in both parametric and non-parametric parameters.

- We explore the highest redshift sample to date utilising unsupervised machine learning. Thanks to JWST we also have access to rest-frame optical imaging across all redshifts so our classification system is not biased by variations between UV and optical morphologies.

Due to the wide redshift range covered in this work we have access to a wide span of early history of galaxy formation which allows us to investigate various trends with cosmic time. Our main findings are summarised below.

- We find that there is a wide variety of galaxy morphology already in place at high-z. In total we find 11 distinct morphological types for our sample.
- We confirm that our unsupervised machine-defined clusters support work to construct a visual classification scheme suitable for high-z, while side-stepping the issue of applying pre-defined categories to new observational regimes.
- We find a decrease in concentrated spheroidal type galaxies with redshift as found by others, and find that disk-like galaxies dominate at high-z, though these are typically clumpy and/or disturbed in morphology.
- Unsupervised methods allow us to establish which morphological features are important and have an impact on the physical properties of the galaxies themselves. The resulting extracted features will provide a more detailed and better suited classification system.

As mentioned in this work, we plan to carry out more detailed studies with redshift evolution and link the morphological classes found in this work to the low redshift universe. With the accumulation of data from JWST we expect our view of the distant universe to continue to expand and improve with more and more observations and detailed analysis, all of which will help improve galaxy evolution studies such as the work carried out in this chapter. Such approaches to galaxy morphology classification are also required to handle the amount of data expected with the future of JWST and similar surveys.

Chapter 4

Evolution of galaxy morphology with redshift

4.1 Morphology Evolution

Since the era of the Hubble Space Telescope began, we have begun to uncover more information about the distant universe and the morphology of galaxies at cosmic dawn. Galaxy morphology is known to be clumpier and more irregular at these early times compared to the local universe (Noguchi, 1998; Conselice et al., 2005). This has been linked to an increase in the number of mergers at this time (Conselice et al., 2013; Whitney et al., 2019), and also to instabilities within gas-rich, turbulent disks (Bournaud & Elmegreen, 2009). It is thought that present day spiral galaxies would have formed through a clumpy phase in their evolution. It has also been found that galaxies in the distant universe possess higher asymmetries (Conselice & Arnold, 2009), which is to be expected from the clumpy morphology observed.

Galaxies, both star forming and quiescent, are also observed as being more compact at a fixed stellar mass (Buitrago et al., 2008; Cassata et al., 2013). Work carried out by van Dokkum et al. (2010) investigating these compact galaxies, showed that they increased their stellar mass by a factor of 2 since $z \sim 2$, explaining how massive galaxies build up their outer regions later on. This inside-out formation theory explains one evolutionary scenario we can observe through studying galaxy morphology.

What is still not understood however, is how the morphology we observe at high-redshift evolves into the galaxies which we observe in the local universe. Is there a global trend within galaxy morphology evolution that we can learn from studying a wide variety of galaxy morphology? In this work we expand upon the work presented in Chapter 3

by incorporating the redshift information into our network during training. We utilise the same dataset as presented in §3.2 for this project, making use of the increased resolution offered by JWST. The aim of this work is to explore if we can uncover a global evolution of morphology both within our dataset as a whole, and within each extracted morphological type. By providing our network with both the morphological information via the galaxy images, and the redshift information, we allow our model to learn a more accurate representation of our dataset. This allows our model to learn trends and changes within different extracted morphological features separate from redshift effects.

This work is organised as follows. In §4.2.1 we detail the update to the architecture we explore in this work. We present a network test using the fashion-MNIST digits in §4.3, to show how our updated network is performing, before applying to our galaxy sample in §4.4. We present our novel technique of β-decay in §4.5, and explore how this affects our network's representation of our dataset. We begin to explore if there is any evolution within our extracted clusters with redshift in §4.6.1. We conclude with a brief summary of the next steps in this work in §4.7.

4.2 Method

As mentioned in §3.3 unsupervised machine learning (UML), specifically variational autoencoders (VAEs), have been used in astronomy for a number of different tasks, including galaxy morphology classification (Spindler et al., 2021; Cheng et al., 2021). In previous works the morphological features and classes were fixed at all redshifts, and they studied how the distribution of galaxies over the classes varied with redshift. Here, we allow the morphological features of interest, and resulting classes, to themselves smoothly vary with redshift. We can both study how galaxies are distributed between the classes, but potentially also consider the evolution within individual classes. For example, spiral galaxies are found at all redshifts, but they are typically clumpier at high redshift (Noguchi, 1998). In previous approaches, one might find two classes, smooth spirals and clumpy spirals, and measure that galaxies move between the two classes as a function of redshift. Alternatively, one could apply the fixed morphological classes to bins in redshift, in which case you might find a single clumpy-spiral class at high-z, a single smooth class at low-z, but two classes at intermediate-z. However, a redshift-conditioned approach applied to the whole dataset might only identify one spiral class, but the morphological properties of that class would vary continuously from smooth to

clumpy as one changes the redshift. Arguably, the latter approach has learned a more meaningful model of how galaxy morphology evolves.

In this project we aim to explore this question, is there a global evolution in morphology that we can uncover through extracting the morphological features of our galaxy sample? Instead of training a model on all of the galaxy images and then exploring how the extracted clusters vary with z, can we train a model with the redshift information so that the model can learn a more accurate representation of galaxies, at the same time providing us with a tool to explore how galaxies vary with redshift in a generative way. As with the work in Chapter 3, we want to ensure that we are not imposing any preconceptions on the results and so we use unsupervised machine learning techniques to address this problem, specifically VAEs. As we want to explore how morphological type varies throughout cosmic time, we need to provide the VAE with additional information about the redshift of each galaxy, as well as the image data, to allow the network to learn any trends of morphological evolution with cosmic time. To allow this we use a conditional VAE (CVAE, Sohn et al. 2015), which is a semi-supervised network as we provide a label along with each image to help the network learn a specific representation of our data. The specific alterations to our VAE architecture are described below.

4.2.1 β-CVAE

For our CVAE we start with the same β-VAE architecture as described in §3.3.3.1. We use the optimised hyperparameters found in §3.5. To provide our network with the redshift of each galaxy we pass both the image data and the label (i.e., the redshift) to the encoder. We use the value of $(1 + z)^{-1}$ as the label for each galaxy, as this gives the expansion factor of the universe at that redshift, which is on the range $0 - 1$. We explored different architectures, varying where the condition was included in the model, to ensure the condition is being utilised to its full capacity.

Originally, the condition was appended to the sampled latent space as an extra dimension. However, we found that the model was ignoring the conditional dimension as the other latent features allowed it more freedom to encode different morphological features, thus minimising the overall loss. In order to try to encourage the network to use the condition, before we sample the reduced image data, we append the condition to the data and then pass it through a hidden layer within the encoder. We then sample from this modified representation to end up with our N latent dimensions (N was set to 23 in our previous work). Before the latent features are passed to the decoder, we also append the redshift

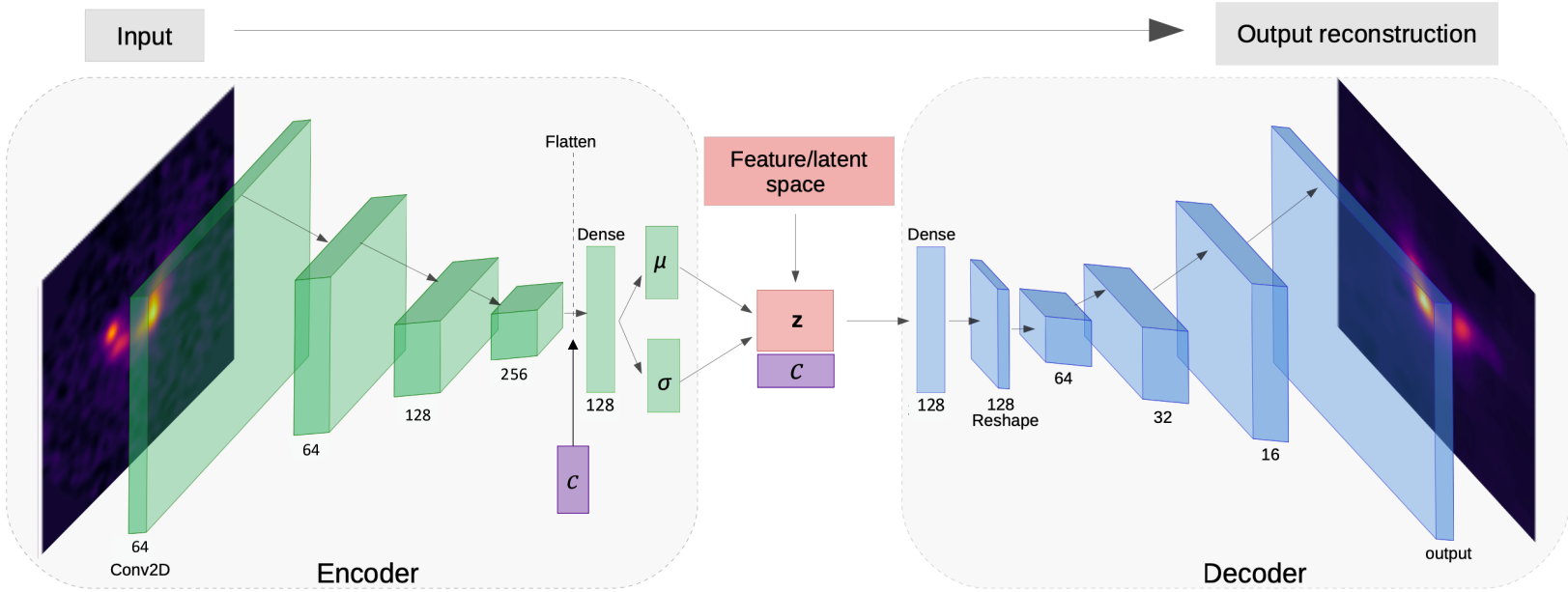


Figure 4.1: Architecture of our β-CVAE network. Same as in Fig.3.3 however, we now append our condition (c, highlighted in purple) to the encoded features before the hidden layers in the network, as well as including it as an extra dimension to our feature space. This results in a feature space of $N + 1$ dimensions (N feature dimensions, 1 conditional dimension).

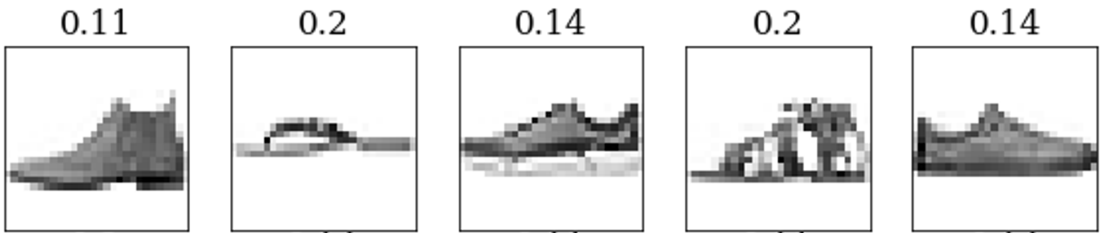


Figure 4.2: A sample of shoe images from the Fashion MNIST dataset. The label of each category is shown above each image. This value is the reciprocal of the category number.

of each galaxy as an extra latent dimension, resulting in $N + 1$ dimensions, to allow the network to learn a representation of the average galaxy morphology at different cosmic times. This inclusion of the condition before the hidden layers in our network, allows the encoder to use the information held within the label of each galaxy, and its inclusion as an additional input to the decoder also allows us to generate images with varying labels, to visualise exactly what the network is learning. The updated architecture is shown in Fig.4.1.

4.3 Network Test

Before training our network on our galaxy images, we want to demonstrate the concept of having a condition within the latent space, to ensure that our network is working as we expect it to. To do this we created a test using the Fashion-MNIST dataset, a sample dataset composed of 10 fashion categories and 70,000 grey scale images (Xiao et al., 2017). As we want to create a test dataset that is similar in variation to our galaxy images we only use 3 of the 10 categories that correspond to different types of shoes: sandals, boots, and trainers. We want to determine if the network is able to learn a smooth transition between these three categories using their labels, akin to what we hope to see for our galaxy sample.

4.3.1 Fashion-MNIST data

As mentioned above, we use the 3 shoe categories from the fashion MNIST dataset in our test, sandals, boots, and trainers. An example of these images are shown in Fig.4.2. As we are using a CVAE we need a label for each image to use as a placeholder for the redshift of our galaxies. As each different type of shoe is its own category (from

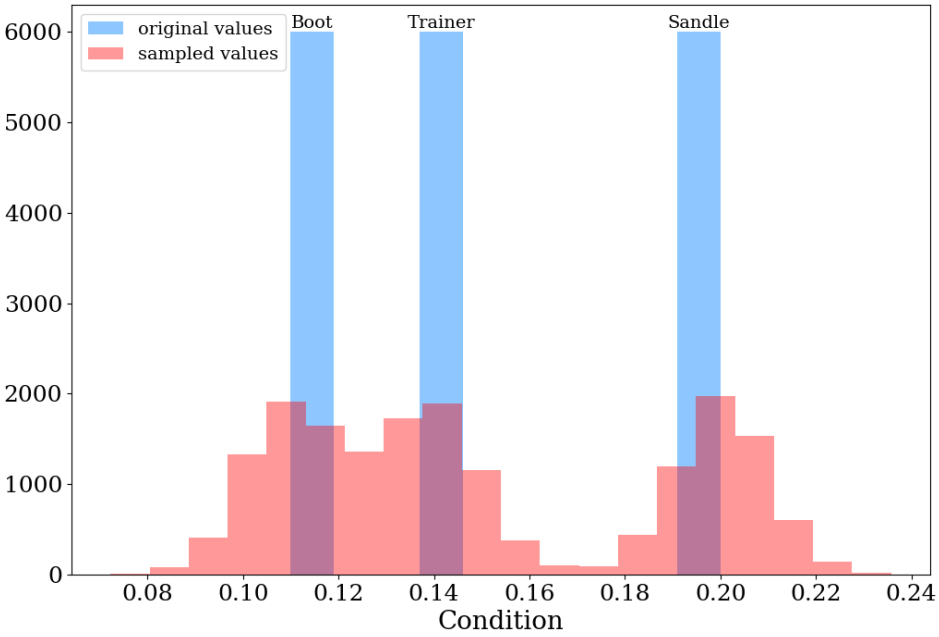


Figure 4.3: Distribution of the labels for our Fashion MNIST dataset before and after our augmentation process. See text for details.

1-10) we thus use the reciprocal of the category as the label, to ensure the labels are compatible with the domain of the activation functions used within the model. These values are shown above each of the images in Fig.4.2.

As the sandal category is more detailed compared to the boot category we can make a rough comparison in this test that the boot images correspond to our high redshift sample, trainers can be thought of as an intermediate redshift sample, and sandals can be thought of as our low redshift sample. We then create an alternative label for each image that allows some variation and overlap between labels to better represent the redshift distribution of our galaxy set. To do this we used a sampled label for each based on a Gaussian distribution with a mean equal to the original label with a σ=0.01. The distribution of the labels before and after this augmentation is shown in Fig.4.3. We will now explore how our network performs with these sampled labels and the fashion-MNIST dataset.

4.3.2 Fashion-MNIST Results

In this section, we will explore how our network learns the conditional representation of the sampled Fashion-MNIST shoe images. By conducting this test, we aim to confirm whether our network is functioning as intended. This validation is crucial for interpreting results when we apply our network to our galaxy dataset. Moreover, it establishes a benchmark for assessing how effectively our network can utilise the conditional information within the feature space. Understanding this fully will aid us in drawing accurate conclusions from our galaxy sample, where the variations due to the condition my be difficult to detect.

4.3.3 Learned conditional representation

We train our β-CVAE network on the Fashion-MNIST data using our sampled labels. In order to keep training times low, we only use 5 latent dimensions for this test, plus the one conditional dimension. We found in the previous chapter that with 5 latent dimensions, much of the variation in the data was captured, thus this should be a sensible selection for this test. After training our network we can look at each individual latent dimension, including the conditional dimension, to determine if the network has learned a representation that can smoothly transition between each category. Each latent dimension, along with the conditional dimension, has been plotted in Fig.4.4.
In Fig.4.4 latent dimensions 0 to 4 correspond to the features learned by the network from the image structure, while latent 5 corresponds to the information learned from the conditional label, i.e the type of shoe. It can be seen that the network is able to smoothly transition between a boot at low values of this conditional dimension, to a sandal at higher values, as learned through their labels. As there is some variety within a shoe type, the network has learned the average object at each conditional value. As there was more overlap between the boot and trainer category, it can be seen that these two objects are not as distinct as the sandal, nevertheless, the network has clearly used this conditional value to encode a lot of information about each object. The idea is that when we apply our network to our galaxy dataset, the network will be able to extract some information about the average galaxy at each redshift, allowing the model to encode a latent space that can both evolve with redshift, and separate redshift evolution effects from other morphological variations.

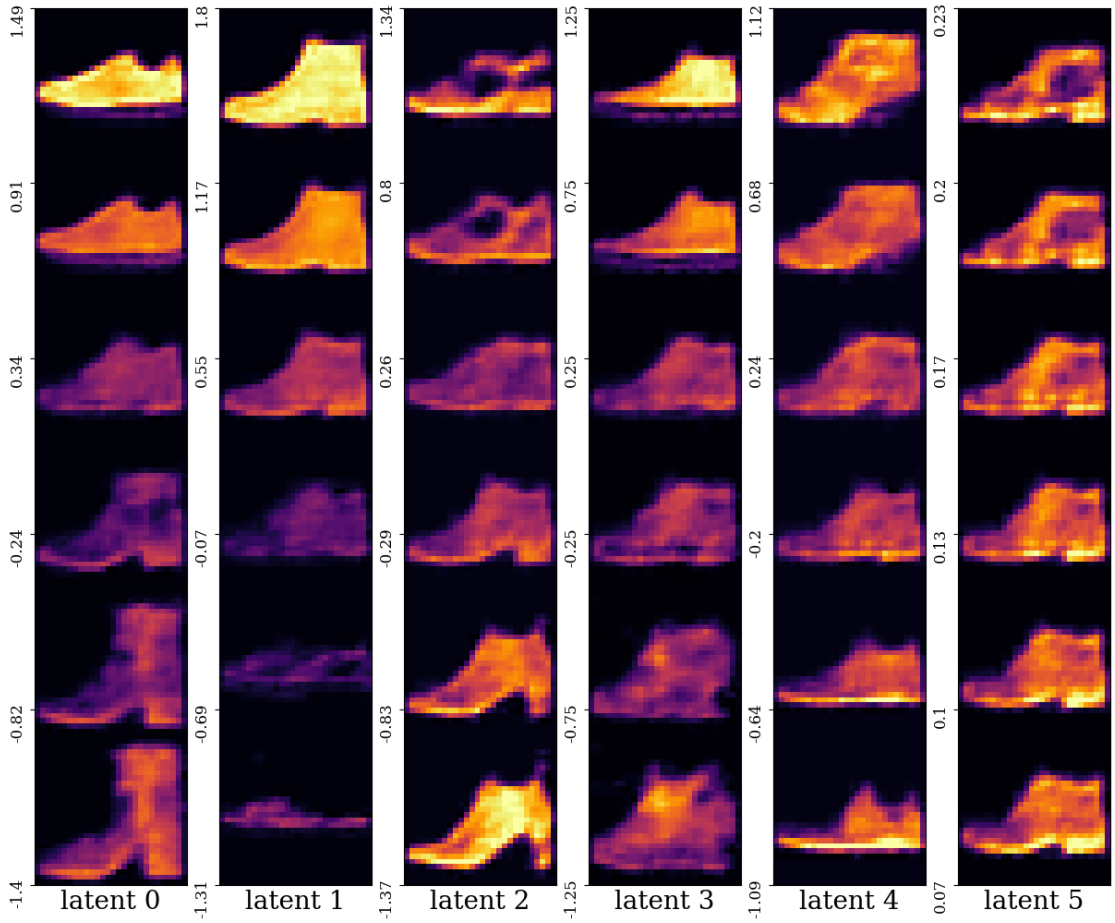


Figure 4.4: Generated images from the trained latent space. Here we show examples of generated fashion-MNIST images along each dimension in the latent space plus the conditional dimension. For feature dimensions we calculate the mean of each feature and linearly interpolate between $\pm 2\sigma$ whilst keeping all other latent features constant to show the features encoded by each dimension. For the conditional dimension we keep all other latent features constant and vary between the min and max label values. It can be seen that the network has used the conditional dimension (far right column) to encode how the images vary based on their labels, transitioning from a boot at low values of this condition, to a sandal at higher values.

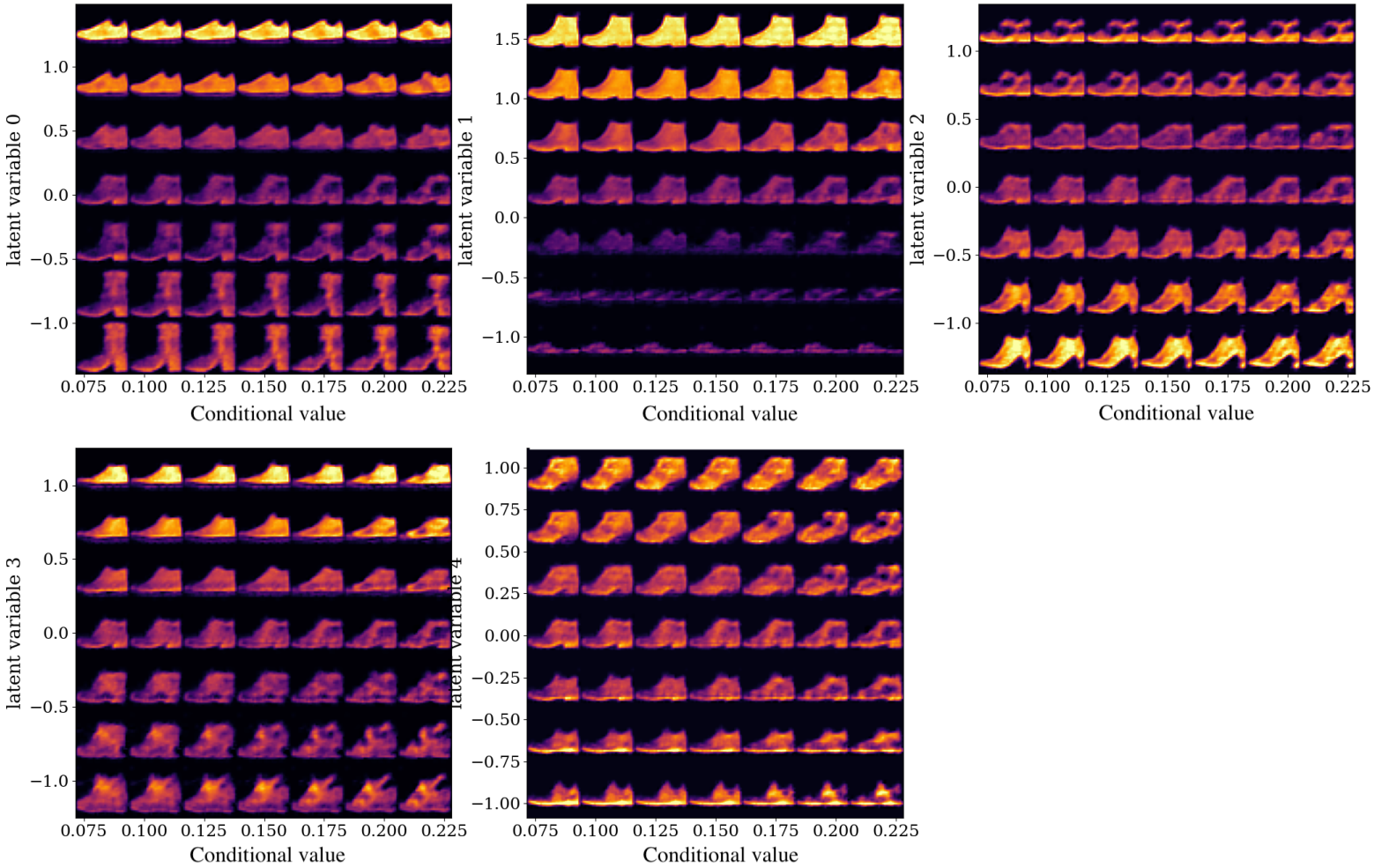


Figure 4.5: Generated images from the trained latent space, in each plot images are generate by varying both the target latent dimension and the conditional dimension, whilst keeping everything else at the average values, to show how the learned features depend on the condition value. Within each plot the condition is varied from low values (boots), to high values (sandals), along the x axis.

4.3.4 Latent features vs condition

We can now look at how each individual latent dimension varies with this condition, this is shown in Fig.4.5. As the network has encoded a lot of information within the conditional dimension, each learned feature is now able to encode features that are able to vary between each class of object. Looking at Fig.4.5 it can be seen that the network's learned features vary with the input condition as desired. For example, looking at latent variable 0, it can be seen that it it is encoding whether the object has a high-top component to the shoe or not. As we move across this latent dimension to the right, we can see that the object varies from a boot, to a sandal smoothly. For example, at low values of latent variable 0, $-1.0 < c < -0.5$, it can be seen that this encodes knee high boots, and as we move to higher values of the condition, this smoothly transitions to 'strappy' high heels. This shows that by providing the network with the conditional values, it has more freedom when encoding features as it can use the same latent feature to describe a wide variety of objects. When we apply our network to our galaxy sample this is the behaviour we want as it will allow us to cluster galaxies into different morphological types as before however, should allow variation due to redshift and evolution effects within each group.

It is important that the network works as expected with the sampled labels, as these better represent the variations and overlap seen in the galaxy data, and the fact our network can still learn a condition dependent representation showcases the power of our model. In Sec.4.4 we apply the network to our galaxy sample, using the scale factor as the conditional values.

4.3.5 Dependence on β

The freedom within our network is dependent on the value of β, as a higher value will allow for better disentanglement of features, forcing the network to learn a more general representation of your data. This improvement of feature disentanglement however, comes at a cost of a diminished reconstruction of your dataset (Higgins et al., 2017). This means that our network's ability to utilise the conditional value will be dependent on β, and too low a value could result in the network not using this information. While a value of 0.01 was suitable for the fashion-MNIST dataset, it may not be suitable for our galaxy dataset. We explore how changing beta affects our model using our galaxy dataset below.

4.4 Galaxy sample

We next apply our galaxy sample and labels to our network, keeping all the hyperparameters the same as the fashion-MNIST test. When we did this we found that the network did not use the conditional dimension, as there was no variation across this dimension in our latent space, thus the results were similar to previous chapter. One hypothesis for why our network worked well for the fashion-MNIST data and not our galaxy sample is possibly due to the value of β, which we briefly addressed in §4.3.5. As our galaxy sample is noisier, and has a wider variety in morphology for each label than our shoe sample did, it may find that by using the feature dimensions only, it can encode most of the variation within our images, and ignores the conditional dimension. We thus try a higher value of β, increasing it from 0.01 to 10, to try to force our network to utilise the redshift information within the images.

4.4.1 Learned conditional representation

When we applied this higher β value to the network, we found that our model began to utilise the conditional dimension, as shown in Fig.4.6. In the top plot we show the learned 5 features for our galaxy sample as well as the conditional dimension (latent 5). While the learned conditional representation of our galaxies is not as clear as the fashion-MNIST dataset, we plot the residuals from each column to better show any differences. This plot is made by subtracting the generated image with a value of 0 in columns $0-4$, and 0.26, i.e the central value, for our conditional dimension, column 5, and plotting the residuals. It can be seen that as we move towards lower values of the condition (i.e., to higher redshift), the light is more diffuse, and less centrally concentrated than at higher values of the condition (i.e., lower redshift galaxies). This is possibly due to resolution effects, or the PSF smoothing effects, which would cause a decrease in the concentration due to the smaller angular size of high-redshift galaxies. As we normalise the sizes of our sources, the angular scale varies from image to image. Images containing small galaxies are increased in size by oversampling the pixel scale. As a result the effective smoothing due to the PSF is greater (in terms of numbers of pixels and fraction of the images size) for images of small galaxies. However, recent work by Wang et al. (2024) found that this PSF smoothing effect is significant with HST data, but JWST data is much less affected, owing to the increase in resolution. While we cannot confirm this is an intrinsic difference, we can explore how the other learned features vary with this condition, to try to uncover if the network is learning any other

evolution effects.
Looking at the other learned features (latent 0 - 4), we see that the network is learning things like axis ratio, bulge-to-disk ratio, asymmetry, and whether there is one or two components within the image. While 5 feature dimensions is not enough to encode all the main morphological features of our dataset, as we showed in Chapter 3, it is reassuring to see that the network is still learning similar features as without the condition. We would hope that when we cluster our feature space we will have similar clusters to the previous work, but now with the ability to explore how these might vary with cosmic time.

4.4.2 Latent features vs condition

We can now look at how each learned feature varies with the conditional dimension as we did with the fashion-MNIST data, this is shown in Fig.4.7. Looking at the how each latent feature varies, we can see that as we move to the right in each figure (which corresponds to moving to lower redshifts in our sample), the central light in each generated galaxy image appears more concentrated. For example, looking at low values of latent 3, which encodes if there is are multiple components/sources within a galaxy, as we move to lower z values the concentration of each source increases.
This trend can be see across all learned features, which is what we expected from Fig.4.6, and from our test with the fashion-MNIST dataset. We now want to explore if the network is able to make more use of the redshift condition. However, setting β much higher would mean that there would not be enough freedom for the other latent dimensions to learn any morphological features as the network will be heavily penalised by any latent feature diverging from a unit Gaussian due to the high KL loss. This forces each latent dimension to learn the 'average' galaxy which is similar to the image generated at the 0 value of latent features $0 - 4$ in Fig.4.6. This is not conducive to our problem as we want to extract morphological features from our dataset to be able to cluster galaxies by different morphological type similarly to Chapter 3, along with the redshift information encoded to allow us to explore changes in morphology with time. In this case, there is an argument to be made for training a network with a high β value to force the network to use the conditional dimension, then reducing β to allow the other latent features to be utilised. A similar approach was taken by Burgess et al. (2018). In their work they varied the target KL loss during training in order to allow the network to encode features more precisely, leading to a disentangled feature space. They

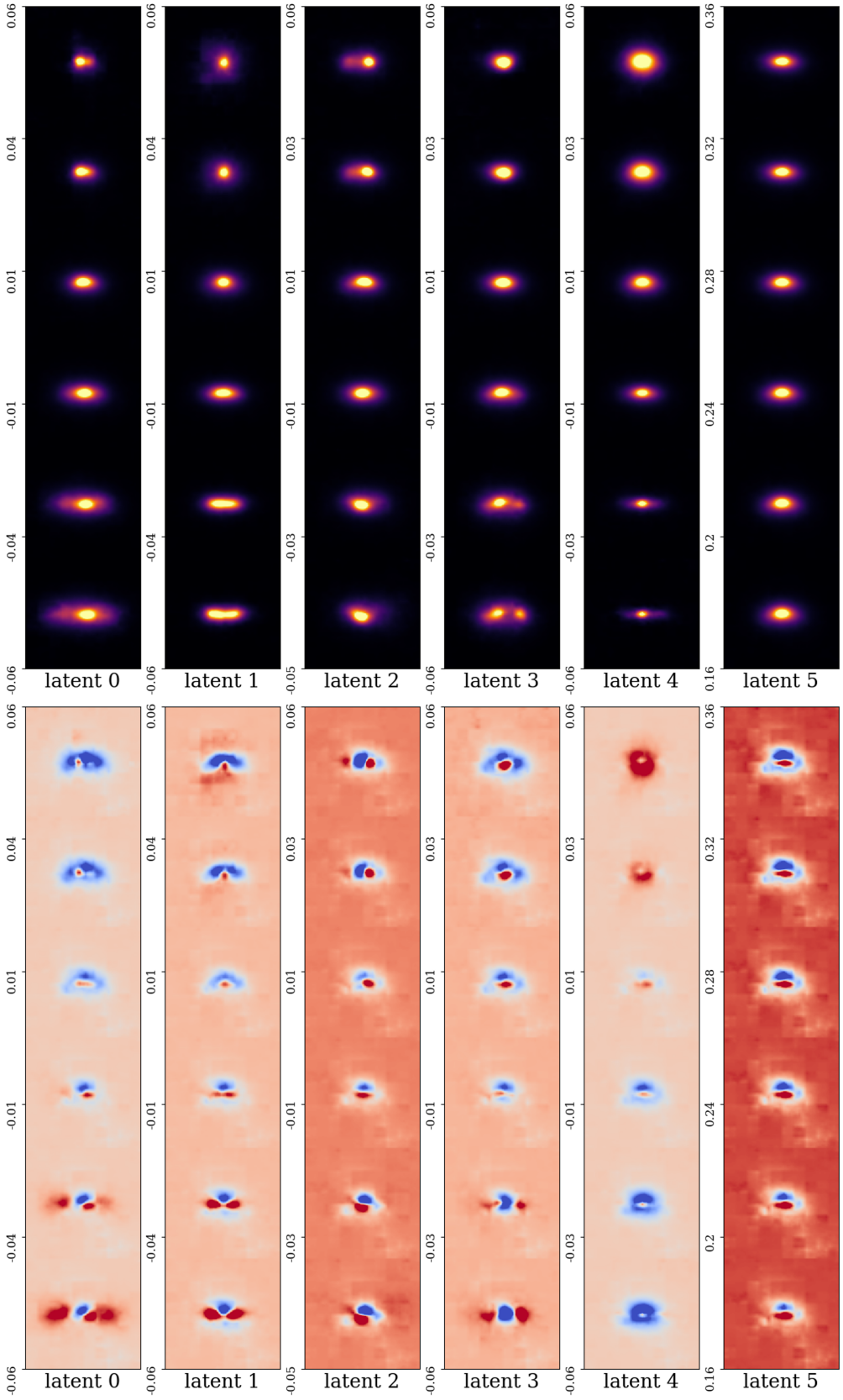


Figure 4.6: Top: Generated images from our trained latent space, same as in Fig.4.4, for our galaxy dataset. Bottom: Residuals when subtracting the generated image with a value of 0 in that feature. This allows us to better see what is changing along each dimension.

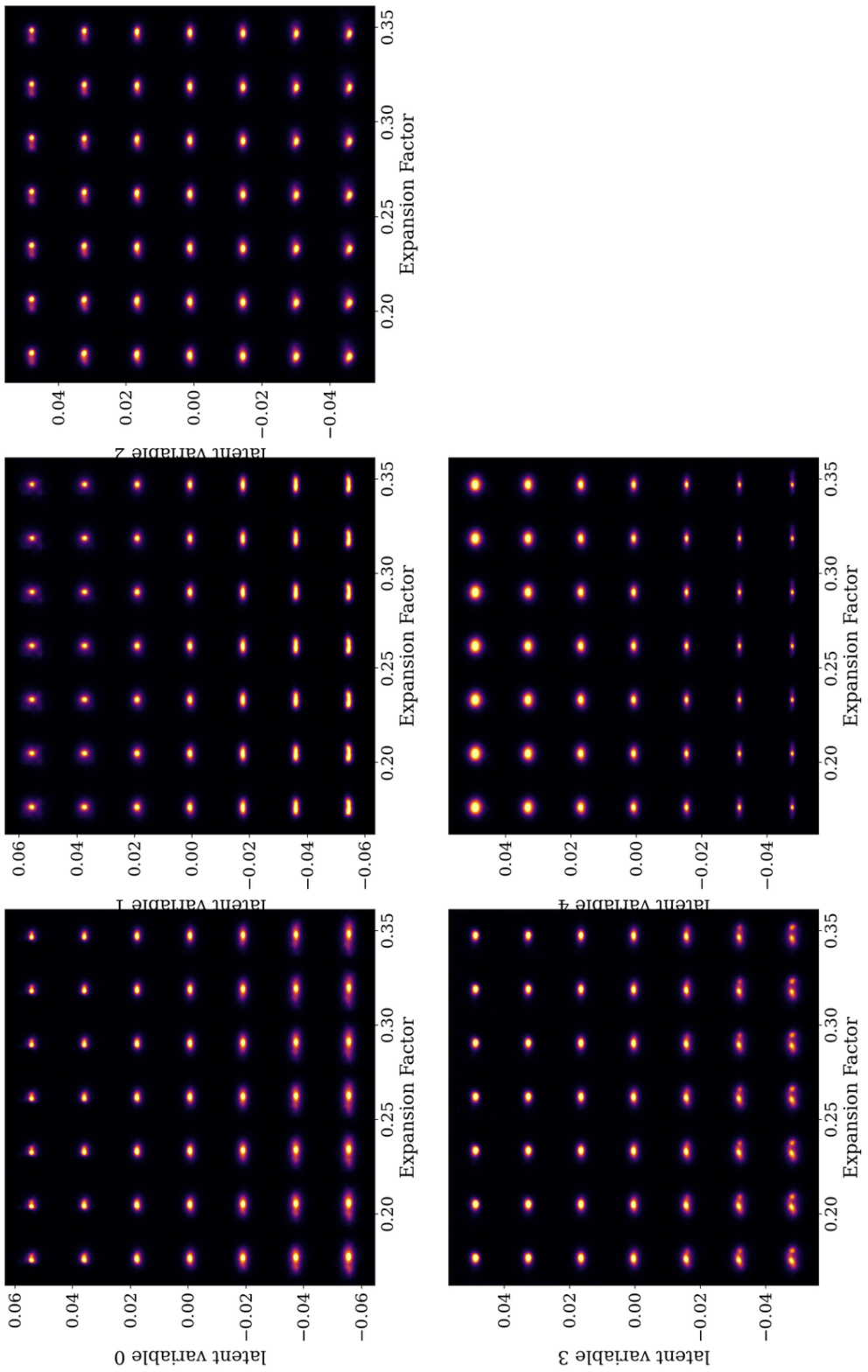


Figure 4.7: Generated images from the trained latent space, in each plot images are generate by varying both the target latent dimension and the conditional dimension, whilst keeping everything else at the average values, to show how the learned features depend on the condition value. Within each plot the condition is varied from low values (high redshift), to high values (lower redshift), along the x axis.

found that their network was more successful than both traditional VAE, and β-VAE, at disentangling features, without this trade-off of reduced reconstructions. We take a slightly different approach by varying β to update the KL loss, which has proven to successfully disentangle features, and lead to better reconstructions of your data (Fu et al., 2019; Shao et al., 2020; Sankarapandian & Kulis, 2021). We explain how we incorporated this into our network below.

4.5 β decay

As mentioned previously, our network's ability to learn a conditional representation of our dataset depends on the value of β. At very low values of β, the network's loss is dominated by the reconstruction loss and thus it will focus on encoding the morphological features within each galaxy image as this will improve the overall reconstruction the most. At very high values of β, the dominant loss within the network will be the KL loss, which will penalise any latent dimension from diverging from a unit Gaussian. This will mean that the network will be forced to use the conditional dimension as we have shown previously in both our test dataset, and our galaxy dataset. We therefore want to start training our network using a high β value, forcing it to learn the conditional representation of our data. We then want to reduce this value to allow the other latent features to be used, allowing us to cluster the feature space to extract the different morphological types as we have shown in Chapter 3. To do this, we have implemented our own class that allows the reconstruction loss within the network to be monitored, and will change the value of β based on when this loss plateaus.

We initialise our training as before, with the 5 latent dimensions, plus our conditional dimension, however we start at a value of $\beta = 10$ to force the network to learn the conditional representation of our data. We then decrease β once there has been no improvement to the overall loss within our validation set after a certain number of epochs, which we refer to as the patience value. We set the early stopping patience value so that it is 5 times greater than the patience value to decay β, allowing for β to decay a number of times before we stop training. This will ensure that β gets a chance to become low enough for the other latent features to switch on, and we do not risk ending the training before the network has had a chance to learn the important features within our dataset. We decrease β by a factor of 2 each time. We chose this value as we tested using a larger decay factor of 10, and found that this allowed multiple latent features to switch on once β had decayed. This meant that the network would focus on using these

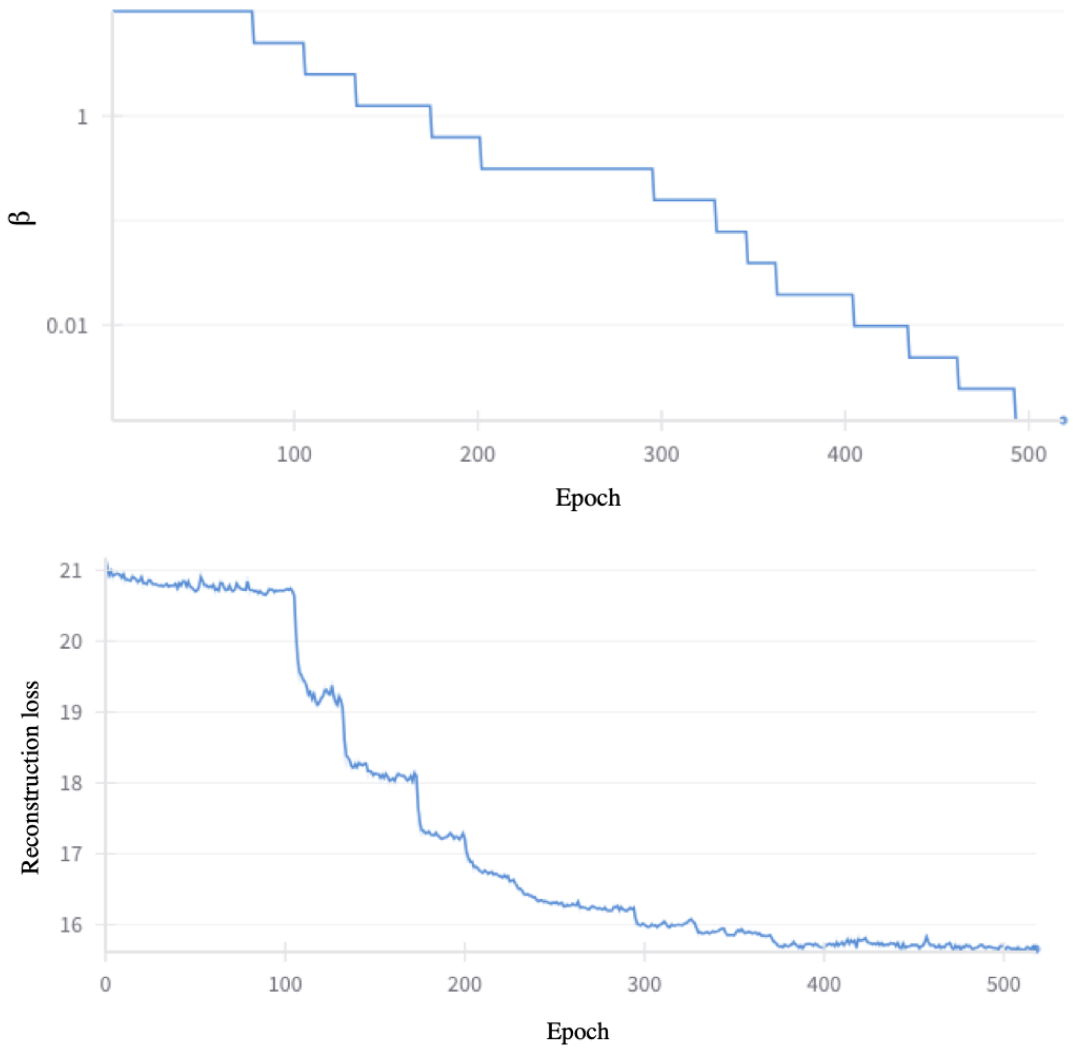


Figure 4.8: Top: Monitoring how the value of β is changing throughout the training of our network. Bottom: Change in the reconstruction loss of our network over the same training epochs. It can be seen that as the value of β is lowered, our network has more freedom to use the other latent features, and as each feature 'switches on', there is a corresponding drop in the overall reconstruction loss of our network.

features to improve the reconstruction loss, and not learn how to improve the conditional dimension. To avoid this, by decaying β only slowly, it allows the network more time to learn a more efficient representation of our dataset, and utilise this redshift information to the fullest. For these tests we keep the learning rate constant, but decaying the learning rate is something that we will incorporate into this work in the future.

4.5.1 Impact on network

In Fig.4.8 we show how β decay is implemented during the training of our network. The top figure shows how the value of β changes during training, and below shows the corresponding change to the reconstruction loss of our validation set. It can be seen that after β is reduced, the reconstruction loss typically undergoes a rapid improvement, then continues its descent at a more sedate pace. It can also be seen however that this loss does not improve with every change in β, which is what we expected as β needs to decrease enough for each latent feature to vary without impacting the KL loss of the network. For example, initially β is set to 10, and we see that the network is not able to learn anything until β decreases to 2.5, after which one of the latent dimensions 'switches on', allowing for some morphological information to be encoded. We now look at which features the network learns, and how these correlate with the learned conditional representation.

4.5.2 Encoded latent features

As before, we can look at the learned features as we decay β, in order to see what information each latent feature is encoding. We start by looking at what the network has learned after the first 30 epochs. We show this in Fig.4.9. As β is still high, the network can only use the conditional information to learn a representation of our dataset. It can be seen from the residuals that there is no variation along any other dimension in our latent space. As we continue to train, the reconstruction loss does not improve and so we decrease β. It is not until β reaches a value of 2.5 that we see an improvement in our loss indicating that one of the latent features has 'switched on', as β is low enough for some variation to be allowed within the network. To determine what our network has encoded, we can look at what latent features show any variation at epoch 120. This is shown in Fig.4.10.

Looking at latent 2 in Fig.4.10, it can be seen that there is now variation along this dimension, with this feature encoding the concentration. Notice how this is the only

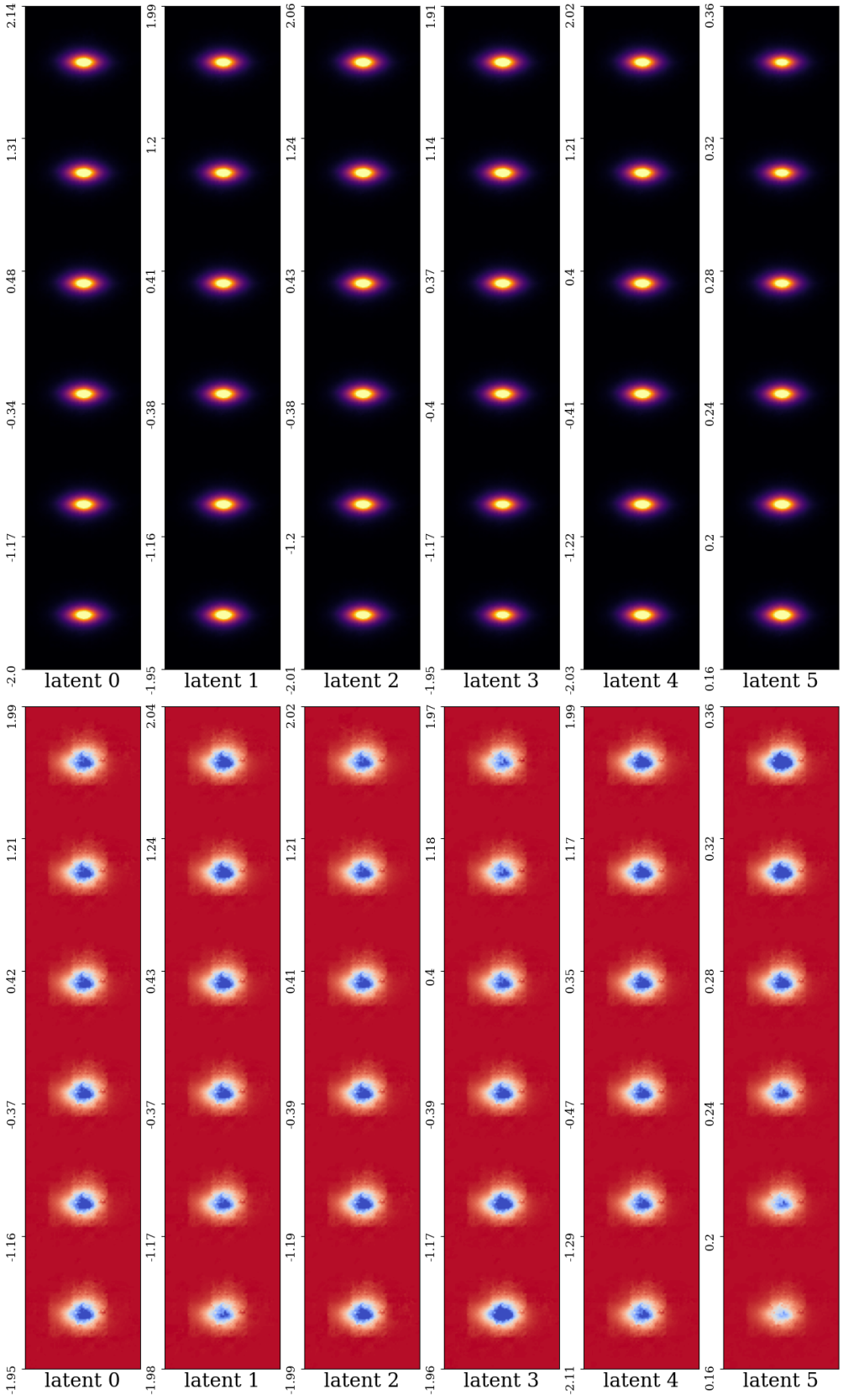


Figure 4.9: Learned features and conditional representation after training for 30 epochs. As β is still high, it can be seen that the network is only using the conditional dimension, as it does not yet have enough freedom to utilise the other latent features.

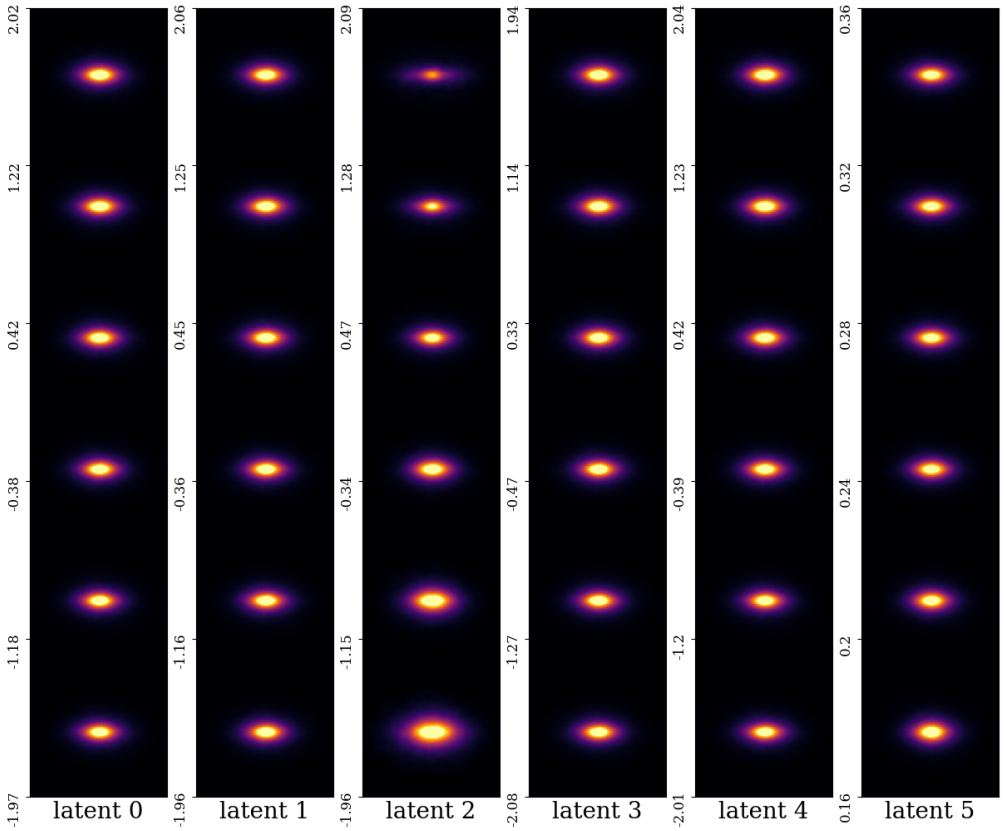


Figure 4.10: The latent features encoded by our network after training for 120 epochs. β has decreased to 2.5, allowing for some variation to be seen in other latent features. It can be seen that there is variation along latent 2, encoding the light profile across a galaxy.

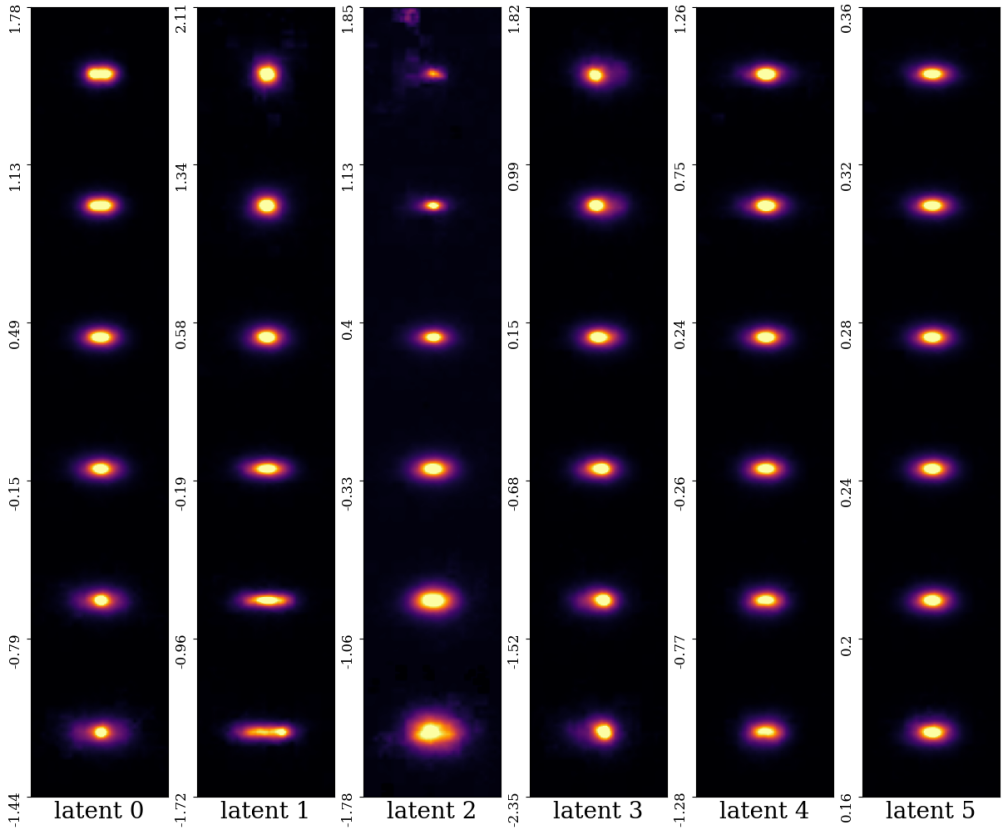


Figure 4.11: The latent features encoded by our network after training for 500 epochs. It can be seen that all latent features have been utilised by the network to encode different morphological features.

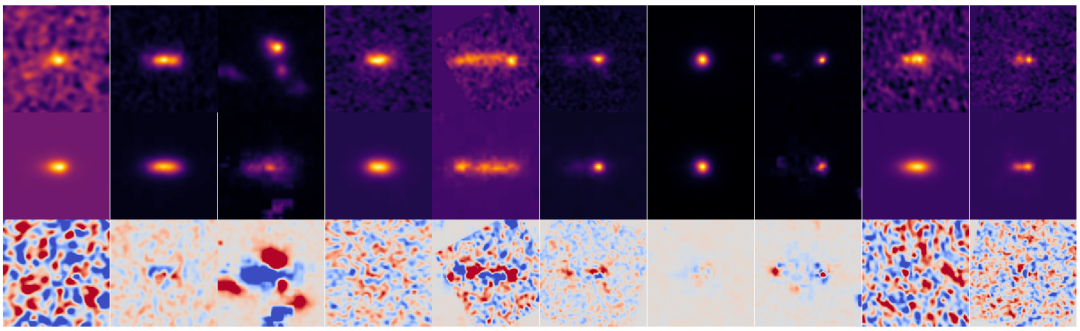


Figure 4.12: Our β-CVAE network's reconstruction of a sample of galaxies from our test set. Top: original images. Middle: our network's reconstruction. Bottom: residuals between input images and reconstruction.

latent feature that has any variation across it, except for the conditional dimension (latent 5). The value of β decays again at epoch ~130, and we again see a corresponding decrease in the reconstruction loss at this epoch. This improvement was caused by refining the encoded information along latent 2, and not by the network using another dimension. This is promising, as it shows that by restricting how much variation the network is allowed during training, can result in a more efficient encoding of your data. As our network continues to train, we see as β decays, the reconstruction loss of our network decreases as expected. In Fig.4.11 we show what the latent space is encoding at epoch ~500, as this is near the end of the training of our network. It can be seen that after 500 epochs, the network has utilised all 5 of the feature dimensions to encoded different morphological features of our dataset. After latent 2, the next feature that was encoded was latent 1, which encodes the axis ratio indicating that this is the next dominant feature in our dataset. The third feature to be encoded was latent 3, which encodes the asymmetry of a galaxy. Latent 0, which encodes the bulge-to-disk ratio, was next, followed by latent 4 which is encoding information about the central region. The order in which our network learned these features reflects their dominance within our dataset as a whole, as these features are the most important when reconstructing the galaxy images.

We can now look at how well our network is reconstructing our galaxy sample with 5 feature dimensions and the conditional representation, this is shown in Fig.4.12. It can be seen that the network is doing a reasonable job, despite only having 6 dimensions to encode information about the morphology of our dataset.

From this we can see that the results are similar, however not as good as in Chapter 3, as we allowed the network 23 latent dimensions to encode the morphological information in

Cluster	Dominant morphology	No. of samples
1	Face on disks, elongated, smooth light distribution	1014
2	tidally disrupted disks, tails	1182
3	edge-on clumpy disks, asymmetric objects, 2 core	1036
4	Major mergers, 2 pairs	107
5	Spheroidal, centrally concentrated	1248
6	Asymmetric, bright core on right side	752

Table 4.1: Dominant morphological type seen in each cluster found by the HC algorithm.

that work. One of the next steps within this work will be to explore using a higher number of latent dimensions, to allow for the majority of morphological variations to be encoded. As we want to fully explore how the condition is effecting the learned representations first, we continue to work with just the 5 latent features, and the conditional dimension for the rest of this test.

4.6 Clustering the β-CVAE latent space

In order to cluster the latent space of our CVAE, we utilise the same hierarchical clustering (HC) technique that we explored in Chapter 3 (see §3.6.1 for details). As we are using a small number of latent features in this test, we do not expect to get the exact same clusters as in Chapter 3, but part of the future work will be to optimise the number of latent dimensions our network needs to encode the variation in our galaxy sample, so we can directly compare the results of using the redshift information within the encoding.

We cluster the latent space without the conditional dimension, as we want to explore if the learned latent features have encoded redshift dependent features. Comparing how the clusters change when we include the conditional dimension is something we would like to investigate in the future. As we have restricted the amount of information that can be encoded, we look at the first main parent clusters, before there are too many splits in the tree to explore if we see any correlation within a group with redshift. When we expand this work to incorporate more latent dimensions we will fully explore all clusters as described in Chapter 3.

We find that there are 6 main clusters, that continue to branch into smaller sub-clusters, however, as the network is only able to utilise a small number of features we look at these major clusters as these should separate based on more global morphology. Looking at the 6 clusters we see that they separate out based on spheroids and disk, and signs of

merging or not. This is to be expected as the first features that were encoded were axis ratio, asymmetry and the light profile of a galaxy. We list the dominant morphology seen in each cluster in Table 4.1.

4.6.1 Variation with redshift

We can now look within each cluster to explore if we can see a variation with redshift. To do this we select a galaxy at random within each group at increasing values of the condition. We then plot these in increasing redshift to determine if we can detect any global changes apart from the concentration/PSF smoothing variation we saw in Fig.4.9. The variation within each group is plotted in Fig.4.13. Each column represents one of the HC clusters, going from bottom to top is increasing redshift within that cluster. The scale factor and the corresponding redshift is indicated in the top left of each figure. As we increase in redshift along each cluster, we can see the decrease in resolution, and the smoothing effect from the PSF. It is impressive however, that even with a very small number of latent features, the network is able to group similar morphologies together despite this decrease in resolution. It is difficult to determine with a small sample from each group if there are any other trends that might also be difficult to see by eye. In order to fully explore and understand if there are any other global trends, we need to investigate how the structural measurements, and physical properties vary within each group with redshift as well.
We can see how some of these clusters have multiple types of morphology within them, for example in cluster 3 we see both clumpy disks and tadpole galaxies, which we found to be two distinct classes in our previous work. This is due to the small number of clusters we look at here, and also due to the small number of latent features we use to encode our galaxy sample. This will impact our results, and will make it more challenging to determine if we are seeing any changes with redshift within each morphological class itself. Therefore, we need to optimise the number of latent features, and the number of clusters before we can fully analyse our results. The details of the next steps are explained below.

4.7 Future ideas

This is ongoing work, and more exploration is needed to fully understand if our network is able to extract more information about the morphological evolution of galaxies. We

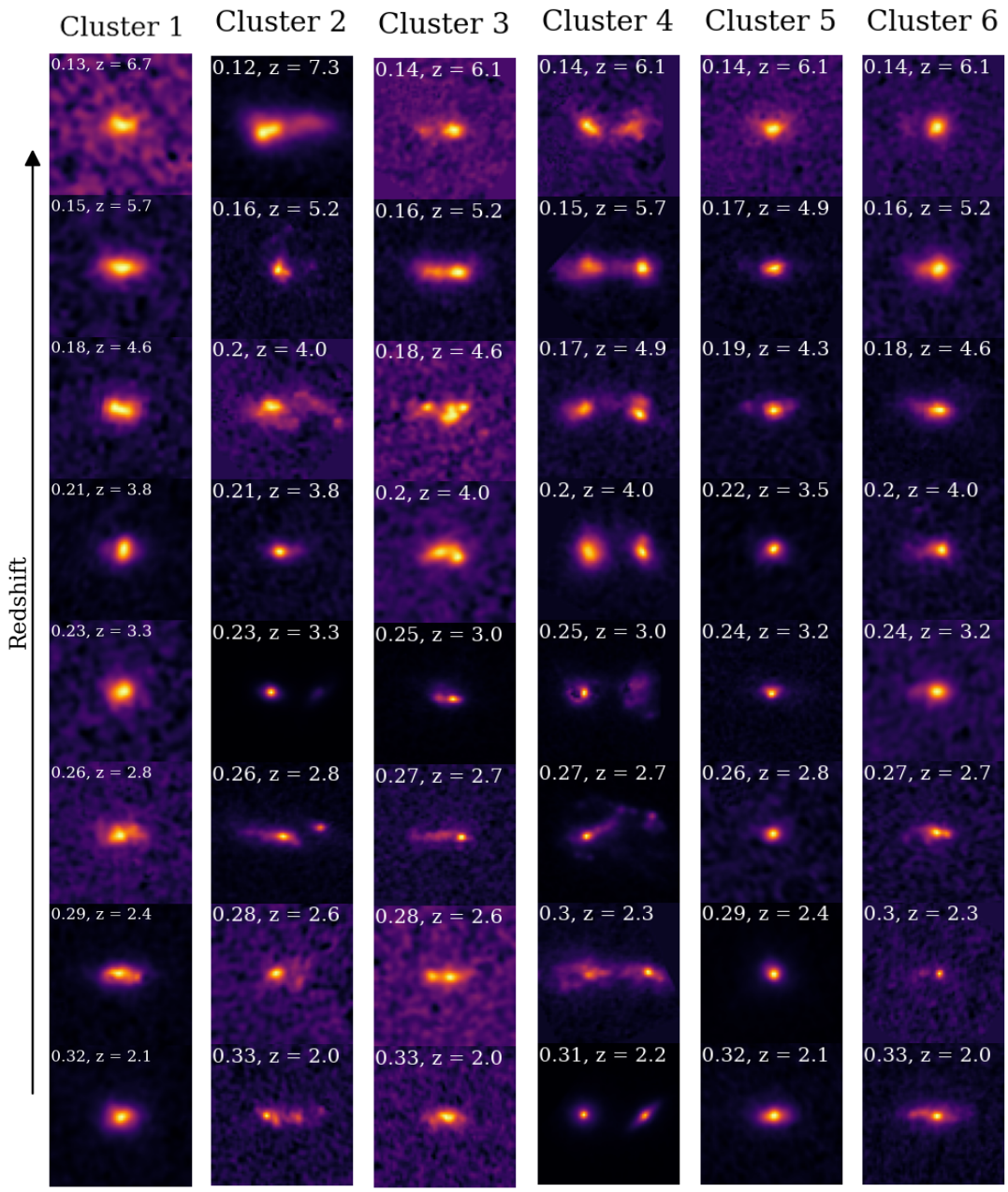


Figure 4.13: Stepping through each cluster found by the HC algorithm in increasing redshift. The scale factor and the corresponding redshift is indicated in the top left of each figure. Morphological descriptions for each cluster can be found in Table 4.1.

list the next steps that are necessary within this work:

- The number of latent features being utilised in the network needs to be optimised. We showed 5 latent dimensions in this Chapter as a proof of concept of how our network works, and showcased how we plan to extract the redshift information however, more dimensions are needed. We could apply a similar method as described in Chapter 3 however, our β-CVAE network has a more principled way of choosing this. As we have shown, each latent dimension will only be used when β is low enough to allow variation along that dimension. Therefore, if we allow the network to utilise as many dimensions as it wants, there will be a point during the training where no more dimensions 'switch on', or they will start encoding things like the background noise. When this point is reached, we can stop the training resulting in the optimum number of latent features the network needs to encode the morphology of our dataset.

- When we ran the HC algorithm, we only looked at the learned latent features, and ignored the conditional dimension. It would be interesting to compare how the clusters vary, and how the galaxies within each cluster vary, if we were to include the conditional dimension when running our HC algorithm.

- Once we have optimised our latent space and fully explored the clusters we extract, we can then compare them to the different morphological types found in Chapter 3. We would expect slight variations, as our β-CVAE network has been allowed to use the redshift information to encode a more accurate representation of our dataset, however, as we showed in the previous Chapter, each morphological type extracted has distinct features and structural parameters, hence, should also be found with our updated network.

- As we have begun to show, we want to fully explore any variations in morphology with cosmic time. Hence, along with the variation shown in Fig.4.13, we also want to explore how the structural parameters within each group itself varies with redshift. This should help us to understand if there are any other changes that our network has encoded, that might be difficult to determine by eye alone.

Chapter 5

Conclusions and Future Work

This thesis utilises imaging data from both HST and JWST to better understand the morphology and structure of galaxies from $0.1 < z < 8$. This was studied through a variety of deep machine learning architectures, along with structural measurements from Morfometryka, and visual classifications.
In Chapter 2 we utilised CNNs to predict non-parametric structural parameters from single band galaxy images and compared the performance of our networks to traditional algorithms with regards to both performance and efficiency. In Chapter 3 we developed a wholly unsupervised, morphological classification scheme, to classify high-redshift galaxies using a VAE, overcoming concerns with previous works by normalising geometric effects within our dataset. Finally in Chapter 4 we explore how morphology varies with cosmic time using a CVAE to extract redshift dependant features from our galaxy sample to learn how this impacts different morphological types.
Below I have summarised the main findings from each research chapter.

5.1 Galaxy structure with deep learning

The future of extragalactic astronomy consists of 'Big Data' surveys, which will image billions of galaxies. Current state of the art computational methods for analysing these surveys will become impractical due to the computational resources and time they need. While detailed analyses will be required for certain measurements, machine learning techniques can replace many current algorithms. We showed how this was possible in Chapter 2 by training CNNs to predict the asymmetry and concentration of a galaxy from a single band image. Our trained networks were able to reproduce the non-parametric structural parameters as measured by Morfometryka to a high accuracy, with lower

errors than conventional methods.
It is known that the CAS values will vary for a galaxy due to redshift effects such as cosmological dimming and angular scaling. To quantify the impact these effects have on our networks measurements, we artificially redshifted a sample of local galaxies, and measured how the networks predictions varied compared to MORFOMETRYKA. We found that our trained networks produce measurements with a lower level of random variation compared to the conventional methods, and are consistent with behaviour seen in previous works (Conselice, 2003). We also showed that our networks are less impacted by noise than MORFOMETRYKA, and are also more stable in the low signal-to-noise regime, in terms of both lower scatter and systematic bias.
Regarding the computational efficiency, our trained networks are thousands of times faster than conventional methods, and along with the improvements with noise and redshift effects, they are much more suited to 'Big Data' surveys.

5.2 Unsupervised Machine Learning for characterising morphology with JWST up to $z \sim 8$

In the distant universe, galaxies display a wide variety of morphologies that are intrinsically different to those we observe in the local universe. The classic Hubble type morphology scheme cannot describe the range of morphologies observed, neither should it be assumed that galaxies should fit into these general categories as this risks biasing our view and understanding of this regime. In order to study these distant galaxies without bias, we need to remove any assumptions we have, thus we turn to unsupervised methods. In Chapter 3 we explore the use of UML to study the morphologies of galaxies $2 < z < 8$ in order to create a meaningful classification scheme that is based on the morphological features of our galaxy sample alone. We used rest-frame optical imaging from JWST to ensure our results would not be biased by variations between UV and optical morphologies.
To extract the morphological features from our dataset we trained a variational autoencoder, selecting the appropriate architecture to ensure that the extracted features are disentangled from each other, resulting in a more easily interpretable feature space. We showed how this encoded feature space correlated with morphological properties of our sample such as the CAS parameters, Sérsic index, and axis ratio, resulting in clusters with distinct morphologies. Improving upon previous works, we normalised geometric

effects in our dataset such as position angle, and apparent size of our galaxies, to ensure that we were extracting meaningful features, and that our resulting clusters were dependent upon morphological differences only and not observational effects that could lead to biased results.

In total we found 11 distinct morphological types for our sample that were well separated in both parametric, non-parametric, and physical properties, indicating that a wide variety of galaxy morphology was already in place in the distant universe. Investigating how the fraction of galaxies in each class varied with cosmic time, we showed a decrease in concentrated spheroidal type galaxies, agreeing with previous results. We also found that disk-like galaxies are dominant in the distant universe however, these are typically clumpy and/or disturbed in morphology.

This work allows us to better understand which morphological features are important and have an impact on the physical properties of the galaxies themselves. The extracted features can be used to create a more detailed, better suited classification system for this high-redshift regime allowing for galaxies to be studied in more detail. Trends with cosmic time can thus be explored and help us to better understand the evolution of galaxies and how this varies between morphological type.

5.3 Evolution of galaxy morphology

Galaxy evolution can be studied in a number of ways, one of which is how the morphology of a galaxy changes with cosmic time. This is typically studied by observing how the fraction of different morphological types vary with redshift. However, this is more difficult when we look to higher redshift as many galaxies do not fall within the traditional classification schemes. We showed in Chapter 3 how the morphology of these distant galaxies can be classified and looked at how the fraction of galaxies within each class varied with redshift. This is able to tell us what types of morphology are dominant at certain times, but does not allow us to explore if there are variations within each morphological type itself. We therefore looked to train a network that was able to utilise the redshift information of a galaxy, in order to encode a more meaningful representation, and allow us to explore variations with time in a generative way.

Preliminary work has shown how our network can learn the change in the apparent concentration of a typical galaxy with redshift. This could be due in part to intrinsic changes, or could be due to resolution effects such as the PSF smoothing we observe at high redshift due to the smaller apparent size of galaxies. More work needs to be done

to explore if there are any other changes with redshift within different morphological types of galaxies. In order to do this thoroughly we need to explore how the different structural and physical properties of galaxies vary with time within each class. This will allow us to better understand if there are any global changes that can be detected by our network.

5.4 Future work

With the launch of JWST we now have access to the optical morphologies of galaxies up to $z \sim 10$, allowing us to probe the morphological history of galaxy evolution. Work carried out in this thesis begins to make use of these data however, our work incorporating the redshift of galaxies is ongoing. It would be interesting to explore galaxies $z < 2$ along with our high-redshift sample, to allow us to link the morphology we encode at high-redshift with that in the local universe, revealing how galaxy evolution varies between morphological type.

Within this work we have shown the importance of removing geometric effects from datasets to allow for meaningful results to be extracted from data. This is something that I believe should be as instinctive, and widespread, as other pre-processing steps such as normalising your data when utilising any ML model, otherwise your results will be biased by observational effects. These effects are imposed on the data by us from our viewing position on Earth and so should be automatically removed before we perform any kind of morphological analysis. While observational geometry can be, at least partly, corrected via a pre-processing step, there is potential to include a geometric transformation within the model architecture itself in the form of spatial transformer networks (Jaderberg et al., 2015). These actively spatially transform feature maps during training, making your model spatially invariant, allowing it to learn the geometrical correction alongside the latent feature space.

Within the next few years, we are expected to have imaged billions of galaxies with observatories such as Euclid, JWST, Rubin, and Roman, which will require the power of ML to analyse. Utilising the networks and techniques we have presented in this thesis to study structural parameters, will thoroughly reveal the morphological evolution of galaxies across cosmic time.

Bibliography

Abraham, R. G., Tanvir, N. R., Santiago, B. X., et al. 1996a, MNRAS, 279, L47, doi: 10.1093/mnras/279.3.L47

Abraham, R. G., Valdes, F., Yee, H. K. C., & van den Bergh, S. 1994, ApJ, 432, 75, doi: 10.1086/174550

Abraham, R. G., van den Bergh, S., Glazebrook, K., et al. 1996b, ApJS, 107, 1, doi: 10.1086/192352

Adams, N. J., Conselice, C. J., Ferreira, L., et al. 2023, MNRAS, 518, 4755, doi: 10.1093/mnras/stac3347

Alpher, R. A., & Herman, R. 1948, Nature, 162, 774, doi: 10.1038/162774b0

Aniyan, A. K., & Thorat, K. 2017, ApJS, 230, 20, doi: 10.3847/1538-4365/aa7333

Bagley, M. B., Finkelstein, S. L., Koekemoer, A. M., et al. 2023, ApJ, 946, L12, doi: 10.3847/2041-8213/acbb08

Bailer-Jones, C. A. L., Irwin, M., Gilmore, G., & von Hippel, T. 1997, MNRAS, 292, 157, doi: 10.1093/mnras/292.1.157

Baldry, I. K. 2008, Astronomy and Geophysics, 49, 5.25, doi: 10.1111/j.1468-4004.2008.49525.x

Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, MNRAS, 393, 1324, doi: 10.1111/j.1365-2966.2008.14252.x

Barchi, P. H., de Carvalho, R. R., Rosa, R. R., et al. 2020, Astronomy and Computing, 30, 100334, doi: 10.1016/j.ascom.2019.100334

Barden, M., Jahnke, K., & Häußler, B. 2008, ApJS, 175, 105, doi: 10.1086/524039

Baron, D., & Poznanski, D. 2017, MNRAS, 465, 4530, doi: 10.1093/mnras/stw3021

Baugh, C. M. 2006, Reports on Progress in Physics, 69, 3101, doi: 10.1088/0034-4885/69/12/R02

Bell, E. F., Phleps, S., Somerville, R. S., et al. 2006, ApJ, 652, 270, doi: 10.1086/508408

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. 2011, in 25th Annual Conference on Neural Information Processing Systems (NIPS 2011), Vol. 24 (Neural Information Processing Systems Foundation). https://hal.inria.fr/hal-00642998

Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, AJ, 119, 2645, doi: 10.1086/301386

Binney, J., & Tabor, G. 1995, MNRAS, 276, 663, doi: 10.1093/mnras/276.2.663

Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, ApJ, 594, 186, doi: 10.1086/375528

Bluck, A. F. L., Conselice, C. J., Buitrago, F. o., et al. 2012, ApJ, 747, 34, doi: 10.1088/0004-637X/747/1/34

Boucaud, A., Huertas-Company, M., Heneka, C., et al. 2020, MNRAS, 491, 2481, doi: 10.1093/mnras/stz3056

Bouché, N., Dekel, A., Genzel, R., et al. 2010, ApJ, 718, 1001, doi: 10.1088/0004-637X/718/2/1001

Bournaud, F., & Elmegreen, B. G. 2009, ApJ, 694, L158, doi: 10.1088/0004-637X/694/2/L158

Buitrago, F., Trujillo, I., Conselice, C. J., et al. 2008, ApJ, 687, L61, doi: 10.1086/592836

Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7, doi: 10.1088/0004-637X/798/1/7

Bunker, A., Spinrad, H., Stern, D., et al. 2000, arXiv e-prints, astro, doi: 10.48550/arXiv.astro-ph/0004348

Burgess, C. P., Higgins, I., Pal, A., et al. 2018, arXiv e-prints, arXiv:1804.03599, doi: 10.48550/arXiv.1804.03599

Buta, R. 1995, ApJS, 96, 39, doi: 10.1086/192113

Cappellari, M., Emsellem, E., Bacon, R., et al. 2007, MNRAS, 379, 418, doi: 10.1111/j.1365-2966.2007.11963.x

Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, MNRAS, 416, 1680, doi: 10.1111/j.1365-2966.2011.18600.x

Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, MNRAS, 399, 1191, doi: 10.1111/j.1365-2966.2009.15383.x

Cassata, P., Giavalisco, M., Williams, C. C., et al. 2013, ApJ, 775, 106, doi: 10.1088/0004-637X/775/2/106

Ceverino, D., Primack, J., & Dekel, A. 2015, MNRAS, 453, 408, doi: 10.1093/mnras/stv1603

Chen, X., Duan, Y., Houthooft, R., et al. 2016, arXiv e-prints, arXiv:1606.03657, doi: 10.48550/arXiv.1606.03657

Cheng, T.-Y., Huertas-Company, M., Conselice, C. J., et al. 2021, MNRAS, 503, 4446, doi: 10.1093/mnras/stab734

Cheng, T.-Y., Li, N., Conselice, C. J., et al. 2020a, MNRAS, 494, 3750, doi: 10.1093/mnras/staa1015

Cheng, T.-Y., Conselice, C. J., Aragón-Salamanca, A., et al. 2020b, MNRAS, 493, 4209, doi: 10.1093/mnras/staa501

Choi, E., Ostriker, J. P., Naab, T., Oser, L., & Moster, B. P. 2015, MNRAS, 449, 4105, doi: 10.1093/mnras/stv575

Collister, A. A., & Lahav, O. 2004, PASP, 116, 345, doi: 10.1086/383254

Conselice, C. J. 1997, PASP, 109, 1251, doi: 10.1086/134004

—. 2000, arXiv e-prints, astro. https://arxiv.org/abs/astro-ph/0012454

—. 2003, ApJS, 147, 1, doi: 10.1086/375001

—. 2014, ARA&A, 52, 291, doi: 10.1146/annurev-astro-081913-040037

Conselice, C. J., & Arnold, J. 2009, MNRAS, 397, 208, doi: 10.1111/j.1365-2966.2009.14959.x

Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, AJ, 126, 1183, doi: 10.1086/377318

Conselice, C. J., Blackburne, J. A., & Papovich, C. 2005, ApJ, 620, 564, doi: 10.1086/426102

Conselice, C. J., Bluck, A. F. L., Ravindranath, S., et al. 2011, MNRAS, 417, 2770, doi: 10.1111/j.1365-2966.2011.19442.x

Conselice, C. J., Mortlock, A., Bluck, A. F. L., Grützbauch, R., & Duncan, K. 2013, MNRAS, 430, 1051, doi: 10.1093/mnras/sts682

Conselice, C. J., Rajgor, S., & Myers, R. 2008, MNRAS, 386, 909, doi: 10.1111/j.1365-2966.2008.13069.x

Cowie, L. L., Hu, E. M., & Songaila, A. 1995, AJ, 110, 1576, doi: 10.1086/117631

Dashyan, G., & Dubois, Y. 2020, A&A, 638, A123, doi: 10.1051/0004-6361/201936339

de Albernaz Ferreira, L., & Ferrari, F. 2018, MNRAS, 473, 2701, doi: 10.1093/mnras/stx2266

de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247

—. 1963, ApJS, 8, 31, doi: 10.1086/190084

de Zeeuw, P. T., Bureau, M., Emsellem, E., et al. 2002, MNRAS, 329, 513, doi: 10.1046/j.1365-8711.2002.05059.x

Dekel, A., & Silk, J. 1986, ApJ, 303, 39, doi: 10.1086/164050

Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441, doi: 10.1093/mnras/stv632

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, MNRAS, 476, 3661, doi: 10.1093/mnras/sty338

Dressler, A. 1980, ApJ, 236, 351, doi: 10.1086/157753

Dreyer, J. L. E. 1888, Mem. RAS, 49, 1

Duc, P. A., Brinks, E., Wink, J. E., & Mirabel, I. F. 1997, A&A, 326, 537

Duncan, K., Conselice, C. J., Mundy, C., et al. 2019, ApJ, 876, 110, doi: 10.3847/1538-4357/ab148a

Eastwood, C., & Williams, C. K. I. 2018, in International Conference on Learning Representations. https://api.semanticscholar.org/CorpusID:19571619

Elmegreen, B. G., & Elmegreen, D. M. 2010, ApJ, 722, 1895, doi: 10.1088/0004-637X/722/2/1895

Elmegreen, D. M., Elmegreen, B. G., & Hirst, A. C. 2004, ApJ, 604, L21, doi: 10.1086/383312

Elmegreen, D. M., Elmegreen, B. G., Rubin, D. S., & Schaffer, M. A. 2005a, ApJ, 631, 85, doi: 10.1086/432502

—. 2005b, ApJ, 631, 85, doi: 10.1086/432502

Feinstein, A. D., Montet, B. T., Ansdell, M., et al. 2020, AJ, 160, 219, doi: 10.3847/1538-3881/abac0a

Ferrari, F., de Carvalho, R. R., & Trevisan, M. 2015, ApJ, 814, 55, doi: 10.1088/0004-637X/814/1/55

Ferreira, L., Conselice, C. J., Duncan, K., et al. 2020, ApJ, 895, 115, doi: 10.3847/1538-4357/ab8f9b

Ferreira, L., Ferrari, F., & Griffiths, A. 2018, galclean: v1.0.0, v1.0.0, Zenodo, doi: 10.5281/zenodo.4004571

Ferreira, L., Adams, N., Conselice, C. J., et al. 2022, ApJ, 938, L2, doi: 10.3847/2041-8213/ac947c

Ferreira, L., Conselice, C. J., Sazonova, E., et al. 2023, ApJ, 955, 94, doi: 10.3847/1538-4357/acec76

Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, ApJ, 946, L13, doi: 10.3847/2041-8213/acade4

Font, A. S., Johnston, K. V., Bullock, J. S., & Robertson, B. E. 2006, ApJ, 638, 585, doi: 10.1086/498970

Frei, Z., Guhathakurta, P., Gunn, J. E., & Tyson, J. A. 1996, AJ, 111, 174, doi: 10.1086/117771

Fu, H., Li, C., Liu, X., et al. 2019, arXiv e-prints, arXiv:1903.10145, doi: 10.48550/arXiv.1903.10145

Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, Space Sci. Rev., 123, 485, doi: 10.1007/s11214-006-8315-7

Graham, A. W., Driver, S. P., Petrosian, V., et al. 2005, AJ, 130, 1535, doi: 10.1086/444475

Gretton, A., Borgwardt, K., Rasch, M. J., Scholkopf, B., & Smola, A. J. 2008, arXiv e-prints, arXiv:0805.2368, doi: 10.48550/arXiv.0805.2368

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35, doi: 10.1088/0067-0049/197/2/35

Gullberg, B., Smail, I., Swinbank, A. M., et al. 2019, MNRAS, 490, 4956, doi: 10.1093/mnras/stz2835

Guo, Y., Jogee, S., Finkelstein, S. L., et al. 2023, ApJ, 945, L10, doi: 10.3847/2041-8213/acacfb

Hafez, I. 2010, PhD thesis, James Cook University, Australia

Haywood, M., Di Matteo, P., Lehnert, M. D., Katz, D., & Gómez, A. 2013, A&A, 560, A109, doi: 10.1051/0004-6361/201321397

Herschel, W. 1786, Philosophical Transactions of the Royal Society of London Series I, 76, 457

Hertzsprung, E. 1926, Bull. Astron. Inst. Netherlands, 3, 115

Higgins, I., Matthey, L., Pal, A., et al. 2017, in International Conference on Learning Representations. https://openreview.net/forum?id=Sy2fzU9gl

Hocking, A., Geach, J. E., Sun, Y., & Davey, N. 2018, MNRAS, 473, 1108, doi: 10.1093/mnras/stx2351

Hoyos, C., Aragón-Salamanca, A., Gray, M. E., et al. 2012, MNRAS, 419, 2703, doi: 10.1111/j.1365-2966.2011.19918.x

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168, doi: 10.1073/pnas.15.3.168

Hubble, E. P. 1926, ApJ, 64, 321, doi: 10.1086/143018

—. 1936, Realm of the Nebulae

Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, ApJS, 221, 8, doi: 10.1088/0067-0049/221/1/8

Huertas-Company, M., Rodriguez-Gomez, V., Nelson, D., et al. 2019, MNRAS, 489, 1859, doi: 10.1093/mnras/stz2191

Huertas-Company, M., Iyer, K. G., Angeloudi, E., et al. 2023, arXiv e-prints, arXiv:2305.02478, doi: 10.48550/arXiv.2305.02478

Hutter, F., Hoos, H. H., & Leyton-Brown, K. 2011, in Learning and Intelligent Optimization (Springer Berlin Heidelberg), 507–523, doi: 10.1007/978-3-642-25566-3_40

Jaderberg, M., Simonyan, K., Zisserman, A., & Kavukcuoglu, K. 2015, arXiv e-prints, arXiv:1506.02025, doi: 10.48550/arXiv.1506.02025

Johnson, S. C. 1967, Psychometrika, 32, 241

Jolliffe, I. T. 1986, Principal component analysis

Kartaltepe, J. S., Rose, C., Vanderhoof, B. N., et al. 2023, ApJ, 946, L15, doi: 10.3847/2041-8213/acad01

Katz, N. 1992, ApJ, 391, 502, doi: 10.1086/171366

Kennicutt, Robert C., J. 1989, ApJ, 344, 685, doi: 10.1086/167834

Kingma, D. P., & Welling, M. 2013, arXiv e-prints, arXiv:1312.6114, doi: 10.48550/arXiv.1312.6114

Kirkpatrick, J. D., Reid, I. N., Liebert, J., et al. 1999, ApJ, 519, 802, doi: 10.1086/307414

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36, doi: 10.1088/0067-0049/197/2/36

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193. https://arxiv.org/abs/1110.3193

Laurikainen, E., Salo, H., & Buta, R. 2005, MNRAS, 362, 1319, doi: 10.1111/j.1365-2966.2005.09404.x

Leavitt, H. S., & Pickering, E. C. 1912, Harvard College Observatory Circular, 173, 1

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proceedings of the IEEE, 86, 2278

Li, R., Napolitano, N. R., Roy, N., et al. 2022, ApJ, 929, 152, doi: 10.3847/1538-4357/ac5ea0

Likas, A., Vlassis, N., & J. Verbeek, J. 2003, Pattern Recognition, 36, 451, doi: 10.1016/S0031-3203(02)00060-2

Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179, doi: 10.1111/j.1365-2966.2008.13689.x

Locatello, F., Bauer, S., Lucic, M., et al. 2018, arXiv e-prints, arXiv:1811.12359, doi: 10.48550/arXiv.1811.12359

Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2008a, MNRAS, 391, 1137, doi: 10.1111/j.1365-2966.2008.14004.x

Lotz, J. M., Madau, P., Giavalisco, M., & Primack, J. 2004a, in American Astronomical Society Meeting Abstracts, Vol. 205, American Astronomical Society Meeting Abstracts, 163.07

Lotz, J. M., Primack, J., & Madau, P. 2004b, AJ, 128, 163, doi: 10.1086/421849

Lotz, J. M., Davis, M., Faber, S. M., et al. 2008b, ApJ, 672, 177, doi: 10.1086/523659

Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415, doi: 10.1146/annurev-astro-081811-125615

Margalef-Bentabol, B., Huertas-Company, M., Charnock, T., et al. 2020, MNRAS, 496, 2346, doi: 10.1093/mnras/staa1647

Martin, G., Kaviraj, S., Hocking, A., Read, S. C., & Geach, J. E. 2020, MNRAS, 491, 1408, doi: 10.1093/mnras/stz3006

McInnes, L., Healy, J., & Melville, J. 2018, arXiv e-prints, arXiv:1802.03426, doi: 10.48550/arXiv.1802.03426

Menzel, M., Davis, M., Parrish, K., et al. 2023, PASP, 135, 058002, doi: 10.1088/1538-3873/acbb9f

Messier, C. 1781, Catalogue des Nébuleuses et des Amas d'Étoiles (Catalog of Nebulae and Star Clusters), Connoissance des Temps ou des Mouvements Célestes, for 1784, p. 227-267

Morgan, W. W. 1958, PASP, 70, 364, doi: 10.1086/127243

Mortlock, A., Conselice, C. J., Hartley, W. G., et al. 2013, MNRAS, 433, 1185, doi: 10.1093/mnras/stt793

Naim, A., Lahav, O., Sodre, L., J., & Storrie-Lombardi, M. C. 1995, MNRAS, 275, 567, doi: 10.1093/mnras/275.3.567

Nair, V., & Hinton, G. E. 2010, in Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10 (Madison, WI, USA: Omnipress), 807–814

Negroponte, J., & White, S. D. M. 1983, MNRAS, 205, 1009, doi: 10.1093/mnras/205.4.1009

Nelson, D., Springel, V., Pillepich, A., et al. 2019, Computational Astrophysics and Cosmology, 6, 2, doi: 10.1186/s40668-019-0028-x

Nevin, R., Blecha, L., Comerford, J., & Greene, J. 2019, ApJ, 872, 76, doi: 10.3847/1538-4357/aafd34

Noguchi, M. 1998, Nature, 392, 253, doi: 10.1038/32596

Odewahn, S. C., de Carvalho, R. R., Gal, R. R., et al. 2004, AJ, 128, 3092, doi: 10.1086/425525

Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., & Fouchez, D. 2019, A&A, 621, A26, doi: 10.1051/0004-6361/201833617

Pearson, J., Maresca, J., Li, N., & Dye, S. 2021, MNRAS, 505, 4362, doi: 10.1093/mnras/stab1547

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266, doi: 10.1086/340952

Penzias, A. A., & Wilson, R. W. 1965, ApJ, 142, 419, doi: 10.1086/148307

Pontoppidan, K. M., Barrientes, J., Blome, C., et al. 2022, ApJ, 936, L14, doi: 10.3847/2041-8213/ac8a4e

Reid, I. N., Brewer, C., Brucato, R. J., et al. 1991, PASP, 103, 661, doi: 10.1086/132866

Robertson, B. E., Tacchella, S., Johnson, B. D., et al. 2023, ApJ, 942, L42, doi: 10.3847/2041-8213/aca086

Rosse, T. E. O. 1850, Philosophical Transactions of the Royal Society of London Series I, 140, 499

Sandage, A. 1961, The Hubble Atlas of Galaxies

—. 1986, A&A, 161, 89

Sanders, D. B., Soifer, B. T., Elias, J. H., et al. 1988, ApJ, 325, 74, doi: 10.1086/165983

Sankarapandian, S., & Kulis, B. 2021, arXiv e-prints, arXiv:2107.10667, doi: 10.48550/arXiv.2107.10667

Sazonova, E., Alatalo, K., Lotz, J., et al. 2020, ApJ, 899, 85, doi: 10.3847/1538-4357/aba42f

Schade, D., Lilly, S. J., Crampton, D., et al. 1995, ApJ, 451, L1, doi: 10.1086/309677

Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, MNRAS, 440, 889, doi: 10.1093/mnras/stu327

Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Shao, H., Lin, H., Yang, Q., et al. 2020, arXiv e-prints, arXiv:2009.06795, doi: 10.48550/arXiv.2009.06795

Shorten, C., & Khoshgoftaar, T. 2019, Journal of Big Data, 6, 1, doi: 10.1186/s40537-019-0197-0

Sijacki, D., Springel, V., Di Matteo, T., & Hernquist, L. 2007, MNRAS, 380, 877, doi: 10.1111/j.1365-2966.2007.12153.x

Snoek, J., Rippel, O., Swersky, K., et al. 2015, arXiv e-prints, arXiv:1502.05700. https://arxiv.org/abs/1502.05700

Sohn, K., Lee, H., & Yan, X. 2015, Advances in neural information processing systems, 28

Spindler, A., Geach, J. E., & Smith, M. J. 2021, MNRAS, 502, 985, doi: 10.1093/mnras/staa3670

Storrie-Lombardi, M. C., Lahav, O., Sodre, L., J., & Storrie-Lombardi, L. J. 1992, MNRAS, 259, 8P, doi: 10.1093/mnras/259.1.8P

The GPyOpt authors. 2016, GPyOpt: A Bayesian Optimization framework in Python, http://github.com/SheffieldML/GPyOpt

Thorne, B., Knox, L., & Prabhu, K. 2021, MNRAS, 504, 2603, doi: 10.1093/mnras/stab1011

Tolman, R. C. 1930, Proceedings of the National Academy of Science, 16, 511, doi: 10.1073/pnas.16.7.511

Tomassetti, M., Dekel, A., Mandelker, N., et al. 2016, MNRAS, 458, 4477, doi: 10.1093/mnras/stw606

Toomre, A., & Toomre, J. 1972, ApJ, 178, 623, doi: 10.1086/151823

Tortorelli, L., & Mercurio, A. 2023, Frontiers in Astronomy and Space Sciences, 10, 51, doi: 10.3389/fspas.2023.989443

Trujillo, I., Conselice, C. J., Bundy, K., et al. 2007, MNRAS, 382, 109, doi: 10.1111/j.1365-2966.2007.12388.x

Tuccillo, D., Huertas-Company, M., Decencière, E., et al. 2018, MNRAS, 475, 894, doi: 10.1093/mnras/stx3186

van den Bergh, S. 1998, Galaxy Morphology and Classification

van den Bergh, S., Abraham, R. G., Ellis, R. S., et al. 1996, AJ, 112, 359, doi: 10.1086/118020

van den Bergh, S., Abraham, R. G., Whyte, L. F., et al. 2002, AJ, 123, 2913, doi: 10.1086/340355

van der Kruit, P. C., & Searle, L. 1981, A&A, 95, 105

van der Wel, A., Chang, Y.-Y., Bell, E. F., et al. 2014, ApJ, 792, L6, doi: 10.1088/2041-8205/792/1/L6

van Dokkum, P. G., Whitaker, K. E., Brammer, G., et al. 2010, ApJ, 709, 1018, doi: 10.1088/0004-637X/709/2/1018

Vega-Ferrero, J., Huertas-Company, M., Costantin, L., et al. 2023, arXiv e-prints, arXiv:2302.07277, doi: 10.48550/arXiv.2302.07277

Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518, doi: 10.1093/mnras/stu1536

Walmsley, M., Lintott, C., Géron, T., et al. 2022a, MNRAS, 509, 3966, doi: 10.1093/mnras/stab2093

Walmsley, M., Scaife, A. M. M., Lintott, C., et al. 2022b, MNRAS, 513, 1581, doi: 10.1093/mnras/stac525

Wang, J.-H., Li, Z.-Y., Zhuang, M.-Y., Ho, L. C., & Lai, L.-M. 2024, arXiv e-prints, arXiv:2403.02094, doi: 10.48550/arXiv.2403.02094

Weinzirl, T., Jogee, S., Khochfar, S., Burkert, A., & Kormendy, J. 2009, ApJ, 696, 411, doi: 10.1088/0004-637X/696/1/411

Weir, N., Fayyad, U. M., & Djorgovski, S. 1995, AJ, 109, 2401, doi: 10.1086/117459

White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52, doi: 10.1086/170483

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341, doi: 10.1093/mnras/183.3.341

Whitney, A., Conselice, C. J., Bhatawdekar, R., & Duncan, K. 2019, ApJ, 887, 113, doi: 10.3847/1538-4357/ab53d4

Whitney, A., Conselice, C. J., Duncan, K., & Spitler, L. R. 2020, ApJ, 903, 14, doi: 10.3847/1538-4357/abb824

Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835, doi: 10.1093/mnras/stt1458

Willett, K. W., Galloway, M. A., Bamford, S. P., et al. 2017, MNRAS, 464, 4176, doi: 10.1093/mnras/stw2568

Xiao, H., Rasul, K., & Vollgraf, R. 2017, arXiv e-prints, arXiv:1708.07747, doi: 10.48550/arXiv.1708.07747

Xu, Q., Shen, S., de Souza, R. S., et al. 2023, arXiv e-prints, arXiv:2303.08627, doi: 10.48550/arXiv.2303.08627

Yagi, M., Nakamura, Y., Doi, M., Shimasaku, K., & Okamura, S. 2006, MNRAS, 368, 211, doi: 10.1111/j.1365-2966.2006.10144.x

Zhang, H., Primack, J. R., Faber, S. M., et al. 2019, MNRAS, 484, 5170, doi: 10.1093/mnras/stz339

Zhao, S., Song, J., & Ermon, S. 2017, arXiv e-prints, arXiv:1706.02262. https://arxiv.org/abs/1706.02262