



A Robust Study of High-redshift Galaxies: Unsupervised Machine Learning for Characterizing Morphology with JWST up to $z \sim 8$

C. Tohill¹, S. P. Bamford¹, C. J. Conselice², L. Ferreira³, T. Harvey², N. Adams², and D. Austin²

¹University of Nottingham, School of Physics & Astronomy, Nottingham NG7 2RD, UK; clar-brid.tohill@nottingham.ac.uk

²Jodrell Bank Centre for Astrophysics, University of Manchester, Oxford Road, Manchester, UK

³University of Victoria, Astronomy Research Centre, Victoria, British Columbia, Canada

Received 2023 June 29; revised 2023 December 8; accepted 2023 December 19; published 2024 February 19

Abstract

Galaxy morphologies provide valuable insights into their formation processes, tracing the spatial distribution of ongoing star formation and encoding signatures of dynamical interactions. While such information has been extensively investigated at low redshift, it is crucial to develop a robust system for characterizing galaxy morphologies at earlier cosmic epochs. Relying solely on nomenclature established for low-redshift galaxies risks introducing biases that hinder our understanding of this new regime. In this paper, we employ variational autoencoders to perform feature extraction on galaxies at $z > 2$ using JWST/NIRCam data. Our sample comprises 6869 galaxies at $z > 2$, including 255 galaxies at $z > 5$, which have been detected in both the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey Hubble Space Telescope fields and the Cosmic Evolution Early Release Science Survey done with JWST, ensuring reliable measurements of redshift, mass, and star formation rates. To address potential biases, we eliminate galaxy orientation and background sources prior to encoding the galaxy features, thereby constructing a physically meaningful feature space. We identify 11 distinct morphological classes that exhibit clear separation in various structural parameters, such as the concentration, asymmetry, and smoothness (CAS) metric and M_{20} , Sérsic indices, specific star formation rates, and axis ratios. We observe a decline in the presence of spheroidal-type galaxies with increasing redshift, indicating the 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.

Unified Astronomy Thesaurus concepts: [Galaxy evolution \(594\)](#); [Convolutional neural networks \(1938\)](#); [High-redshift galaxies \(734\)](#); [Galaxy classification systems \(582\)](#)

1. Galaxy Morphology

The morphology of a galaxy is a record of its formation history. It has been shown that the morphology of a galaxy traces the spatial distribution of ongoing star formation and encodes the signatures of past and ongoing dynamical interactions, which can give us an indication of how galaxies evolved throughout cosmic time (Holmberg 1958; Dressler 1980; Kauffmann et al. 2003; Conselice et al. 2013). The Hubble classification scheme describes the morphologies of galaxies observed in the local Universe. This classification scheme not only describes the visual appearance of the galaxy, but it has been shown that morphological type correlates strongly with many intrinsic properties such as the star formation rate (SFR), age, number of past merging events, etc. (Sandage 1986; Lotz et al. 2008). However, while this system has recently been shown to describe galaxies up to very high redshifts of $z \geq 5$ (Ferreira et al. 2023; Huertas-Company et al. 2023; Jacobs et al. 2023; Kartaltepe et al. 2023), the number of irregular galaxies increases rapidly with redshift (Abraham et al. 1996; Mortlock et al. 2013; Ferreira et al. 2020), thus requiring a more detailed description.

Investigating galaxies at higher redshifts we observe much more peculiar and clumpy morphologies (Conselice et al. 2005).

Due to the variety of galaxy types we observe it becomes challenging to create a classification scheme for all galaxies akin to the Hubble tuning fork we apply at low redshifts. Relying on the nomenclature that has served us at low redshift risks biasing our understanding of this new regime. For example, work carried out in the past by Elmegreen et al. (2005) classified high-redshift galaxies in the Hubble Ultra Deep Field into six main groups: chain, clump cluster, double clump, tadpole, spiral, and elliptical. These groups were determined by eye to match matched previous classifications made by others (Cowie et al. 1995; van den Bergh et al. 1996). While many galaxies at high redshift fall into one of these categories, a more robust classification scheme is needed. The CAS parameter system (concentration, Bershadsky et al. 2000; asymmetry, Schade et al. 1995; and smoothness of a galaxy’s light profile), defined in Conselice (2003), is one way this can be achieved. For example, galaxies that have been classified as “tadpole” galaxies tend to have high asymmetries while “clump cluster” galaxies tend to have low concentration values. Similarly the Gini (G)– M_{20} nonparametric measurement system introduced by Lotz et al. (2004) showed that combining the CAS parameters with G and M_{20} was a more effective approach to classifying different morphologies. While these parameters can aid in distinguishing between certain classifications, it remains unclear if all high-redshift galaxies neatly fit into these morphological categories and how these relate to quantitative structure and other physical properties. Therefore, it is crucial to develop an efficient and robust method that groups



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

galaxies based on their intrinsic features, without imposing our own biases on the classification classes and criteria.

A classification scheme for early galaxies would allow us to probe deeper into the intrinsic properties of these galaxies and will help to better understand their evolution. We know that the SFR in the Universe peaked at around $z \sim 2$ (Madau & Dickinson 2014), meaning that galaxies would have had more star-forming regions, leading to more clumpy morphologies. We also know that the merger rates were higher in these earlier times (Patton et al. 2002; Duncan et al. 2019), meaning galaxies possess more disturbed morphologies with tidal disruptions, multiple cores, etc. However, what is not understood is how these galaxies evolved into those we observe today. For example, looking at galaxies, both star-forming and quiescent, at a fixed stellar mass, we observe those at higher redshift to be more compact than their lower-redshift counterparts (Wilman et al. 2020). How these evolved into what we see today could possibly be studied by investigating the evolution in morphological type with redshift. How to define morphological type is however not obvious at high redshift, which is the focus of this paper.

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; Guo et al. 2023; Huertas-Company et al. 2023). While these studies focus on the morphologies of these distant galaxies, they still characterize 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 when concluding into which morphological group a galaxy falls.

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 galaxy’s morphology are most important in its characterization? 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 $z \sim 1$, 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 classify galaxies at high redshift better 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 a solution to this problem, and one that has become popular in recent years: machine learning (ML). Work has been carried out by Walmsley et al. (2022a) that combines these visual classifications from Galaxy Zoo with ML, allowing for many more

classifications, and also enabling researchers to locate anomalies within their data sets (Walmsley et al. 2022b). While these techniques have proven very successful, they still require labels to train the networks initially. By using unsupervised ML we can remove the need for any classifications or labels.

In this work we utilize an unsupervised deep learning algorithm to extract the most dominant morphological features of 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 redshifts, masses, and SFRs to provide new insights into the evolution of galaxy structure since $z < 8$.

This paper is organized as follows. In Section 2 we introduce the imaging data and survey used in this work, along with our selection criteria. In Section 3 we detail the various architectures we explore in this work as well as the clustering algorithm used. We discuss our data standardization process in Section 4 and the optimization of our networks in Section 5. The clustering algorithm used in this work is detailed in Section 6. The resulting optimized network and results are included in Section 7, which also includes information about the extracted morphologies and clusters. We conclude with a brief summary of our main results in Section 8.

2. Data

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 NIRCcam instrument on JWST. NIRCcam offers wavelength coverage from 0.6 to 2.3 μm with a resolution of $0''.031 \text{ pixel}^{-1}$, and from 2.3 to 5 μm with a resolution of $0''.063 \text{ pixel}^{-1}$.

The data have 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 Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) for this work. In total we have $\sim 10,000$ 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 redshift, photometry, mass, and SFR measurements that have been derived and utilized in previous works (Duncan et al. 2019; Whitney et al. 2021). Known active galactic nuclei 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 Figure 1. While we will be using the redshift, SFR, mass and other measurements from the original CANDELS galaxies, we remeasure the nonparametric morphologies with MORFOMETRYKA to have updated and more accurate CAS, Sérsic, Gini– M_{20} , etc., values. We expect these to change from the values measured from the HST images due to the increase in resolution and signal-to-noise ratio (S/N) of the data. It should be noted that these measurements were performed on the original galaxy stamps from JWST before any standardization, discussed in Section 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,

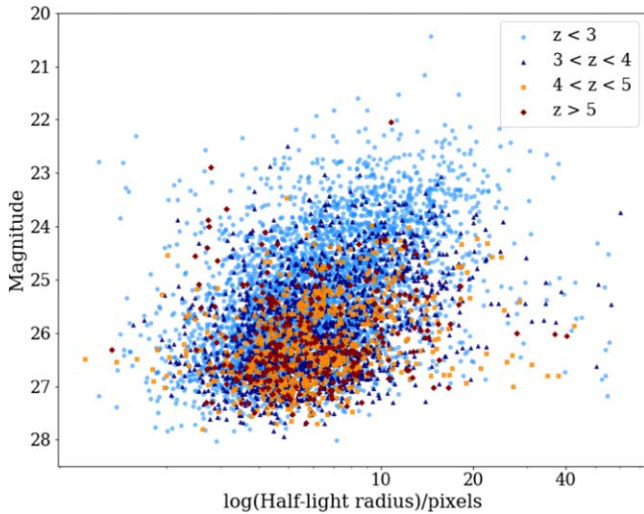


Figure 1. Distribution of the apparent H_{160} magnitude of our galaxy sample vs. half-light radius in pixels. The resolution from NIRCcam is $0''.03$ and $0''.063 \text{ pixel}^{-1}$ for $0.6\text{--}2.3 \mu\text{m}$ and $2.3\text{--}5 \mu\text{m}$, respectively. Our sample is measured from JWST/CEERS imaging.

F356W, F410M, and F444W bands and match to the redshift of each galaxy; see Table 1 for the redshift cuts.

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 Figure 2. Galaxies above this limit are used for the analysis in this work.

3. Method

3.1. Machine Learning

In recent years ML has proven to be very successful in astronomy and has been applied to many different problems. ML has been utilized 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 nonparametric structural measurements such as the CAS system (Tohill et al. 2021). 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) 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 with which to train the network. This comes with its own issues: first 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

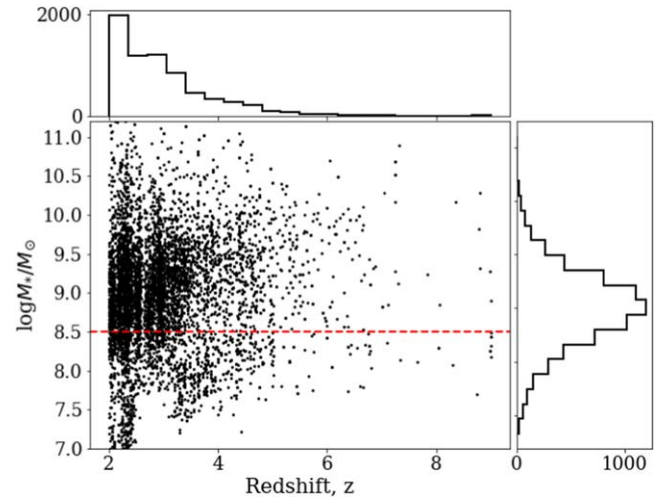


Figure 2. Distribution of the stellar masses of galaxies within our matched sample vs. redshift. The red line shows the mass cut at $10^{8.5} M_\odot$. Galaxies above this mass are utilized in this work, which is above the completeness limit at all redshifts.

Table 1

Summary of the Bands Utilized from JWST NIRCcam with the Associated Redshift That Aligns with the Optical Rest Frame

Redshift Range	Band	Total	$\lambda_{\text{pivot}} (\mu\text{m})$
$2.0 \leq z < 3.0$	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

classifications is to move toward using unsupervised ML techniques.

3.2. Unsupervised Machine Learning

As the name suggests, unsupervised ML 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 the images. These features can then be explored and analyzed, allowing the user 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.

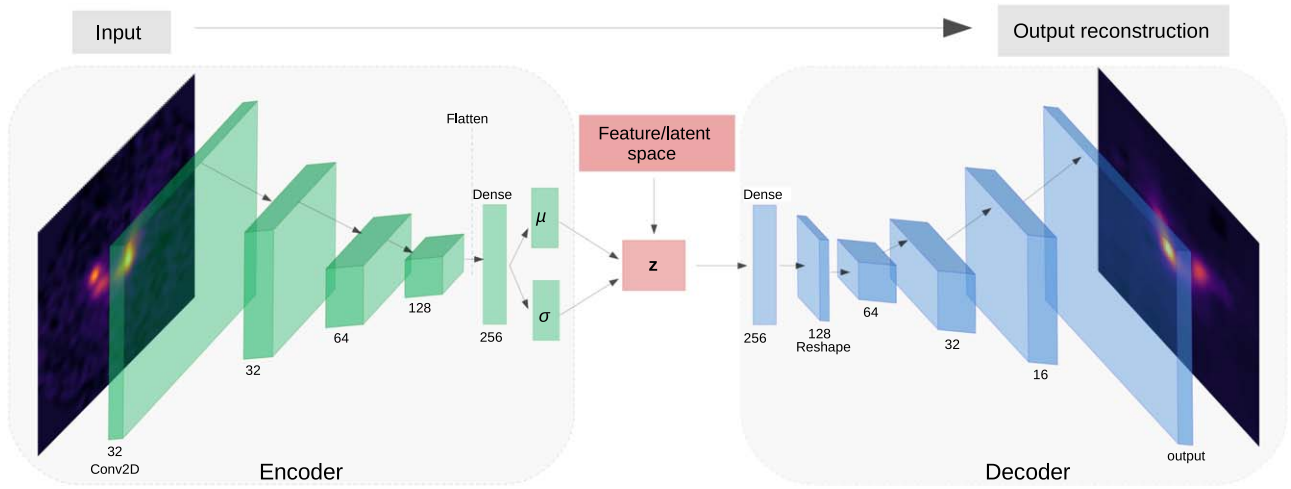


Figure 3. Architecture of our MMD-VAE network. On the left is the encoder, which takes the input data and compresses it into a smaller number of dimensions represented by the feature space (z). The decoder then samples from this feature space to create a reconstruction of the input. The aim of this work is to cluster this feature space to extract groups of galaxies with similar morphologies.

3.3. VAEs

In this work we utilize a type of autoencoder (AE) network. The main idea behind an AE 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 Figure 3). The encoder takes an input, which in this example is an image, and 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 terms will be used interchangeably throughout this work. The decoder then samples from this latent space to create a reconstruction of the input. The AE is trained to compress the input data while minimizing the reconstruction loss between it and the outputted, decoded image. One downfall to AEs is that they are prone to overfitting as there is no regularization 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 the 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 minimized 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 Kulback–Leibler (K-L) 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, asymmetries, etc. There are a number of variations to the VAE that aim to address this issue of entanglement. Two of the more success variations are β -VAE (Higgins et al. 2017), and the maximum mean discrepancy (MMD)-VAE (Zhao et al. 2017). While Locatello et al. (2018) state that perfect disentangled representations are theoretically impossible, imperfectly disentangling 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.

3.3.1. β -VAE

The β -VAE architecture was first introduced by Higgins et al. (2017). This type of VAE incorporates a weight on the K-L loss to improve the disentanglement of features in the latent space. A value of $\beta = 1$ represents the original VAE, while $\beta > 1$ 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 disentangle them better, leading to a more robust representation of the data. The loss is defined as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{KL}}, \quad (1)$$

where

$$\mathcal{L}_{\text{recon}} = \mathbb{E}_{q_{\phi}(z|x)} \log p_{\theta}(\hat{x}|z), \quad (2)$$

$$\begin{aligned} \mathcal{L}_{\text{KL}} = \beta D_{\text{KL}}[q(z|x) \| p(z)] = \beta \frac{1}{2} \sum_j 1 \\ + (\log \sigma_j)^2 - \mu_j^2 - \sigma_j^2. \end{aligned} \quad (3)$$

Here, $p_{\theta}(\hat{x}|z)$ is the likelihood of data \hat{x} given latent space z and $q_{\phi}(z|x)$ is the posterior distribution of the 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 K-L 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 squared error (MSE) between the reconstructed image and the input data. The network will be penalized for diverging from either of these conditions, as is the case with the traditional VAE. In β -VAE there is an additional adjustable hyperparameter β that is introduced to the K-L 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 K-L divergence can be moderated it still penalizes 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 purposes as we are interested in retrieving distinct subclusters 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 MMD-VAE.

3.4. MMD-VAE

The second network we investigate is MMD-VAE. This type of network differs from β -VAE in that it does not exploit the K-L divergence but instead finds the MMD (Gretton et al. 2008) between the prior distribution $p(z)$ and the posterior $q(z)$. The MMD of two distributions is minimized when they are identical. Instead of comparing the overall distributions like the K-L 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 the kernel trick is used. This allows nonlinear 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

$$\mathcal{L}_{\text{MMD}} = \mathbb{E}_{p(z), p(z')} [k(z, z')] + \mathbb{E}_{q(z), q(z')} [k(z, z')] - 2\mathbb{E}_{p(z), q(z')} [k(z, z')], \quad (4)$$

where $k(z, z')$ can be any universal function, the most common being a Gaussian kernel, which we utilize 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

$$\mathcal{L}_{\text{MMD-VAE}} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{MMD}}. \quad (5)$$

The advantage MMD-VAE has over β -VAE is that it is not penalized 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 β -VAE network.

To compare both networks fully we optimize both and compare how the reconstruction loss varies between them for the same number of latent variables. This comparison is discussed in Section 5.

4. Data Preprocessing

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. These features however are not intrinsic to the galaxy and are in fact observational biases that we have imposed on the data simply because of our observational position on Earth. The dominance of these features has been well demonstrated in Spindler et al. (2021). In their work they show how almost half of their latent information space encodes the orientation of their

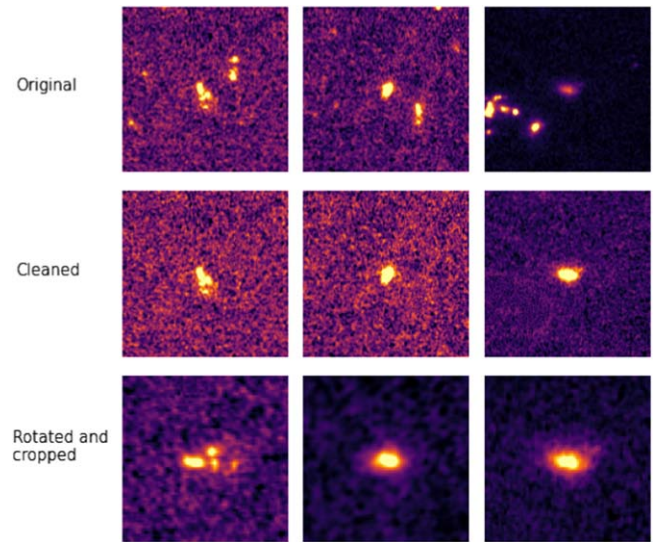


Figure 4. An example of our image standardization process. Top: original CEERS JWST images. Middle: this shows the galaxy images after they have been cleaned of background sources. Bottom: rescaled and rotated images.

galaxies and the positions of background sources. While they were able to produce generated galaxy images with their network, they show one of the main downfalls of unsupervised ML techniques. 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.

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.

4.2. Image Standardization

To address these observational biases in our work we preprocess our data before training the network to prevent these features from being an issue by standardizing our galaxy sample.

An example of this is shown in Figure 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, the 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 noncentral 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 in a network and so we rotate all of our galaxies by

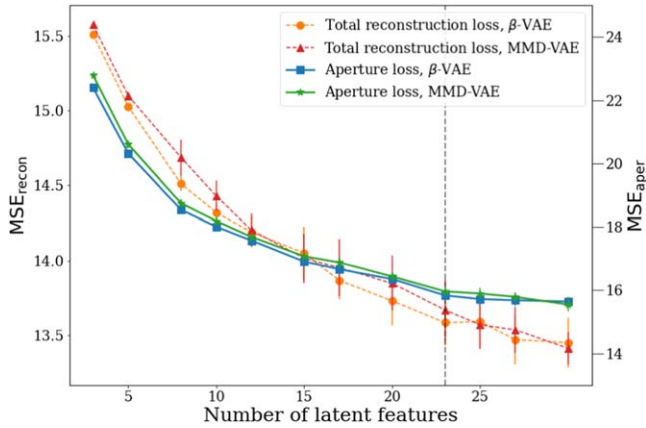


Figure 5. Variation of the reconstruction loss, both total and within an aperture, of the validation set from our networks vs. the dimensionality of the latent space. The solid points show the average MSE for the last 50 epochs (our patience value, see text for details) and the error bars show 1σ of the variation. It can be seen that after 23 latent dimensions there is no improvement in aperture loss within the errors, while the total reconstruction loss continues to decrease.

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 standardization process. The last feature we address is the apparent size of the galaxies. We rescale all of our images to an 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 Figure 4. We then use these cleaned images to train our network.

5. Model Training and Optimization

Depending on the chosen network architecture, there will be a varying number of hyperparameters that need to be optimized. 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, while being used successfully in the past, are computationally expensive and can take a while to converge on an optimum value. A more efficient approach to optimizing network architectures is Bayesian optimization (Snoek et al. 2015). Bayesian optimization builds upon previously evaluated models to create a probabilistic model which is built upon to converge on an optimum solution more efficiently. The hyperparameters within our network are 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 optimizer used, the values 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 optimization process to focus on the architecture of

Table 2
Summary of the Hyperparameters Selected by the Bayesian Optimization Technique

Hyperparameter	Optimum Value	
	MMD-VAE	β -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

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 optimize 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 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 of 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.

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. 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 the total reconstruction loss (i.e., the whole image) $\mathcal{L}_{\text{recon}}$ and the loss within an aperture of $R_p = 15$ pixels, $\mathcal{L}_{\text{aper}}$, i.e., 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 our sample of galaxies. The average variation in the loss for both networks is shown in Figure 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 $1 R_p$ for each galaxy (see Appendix Figure A1 for the training and test sets). For all latent dimensions investigated, the total reconstruction loss continues to improve from an $\text{MSE}_{\text{recon}}$ of ~ 15.5 to an $\text{MSE}_{\text{recon}}$ 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 MSE_{aper} of the reconstructions within the aperture plateau at around ~ 15.8 for both networks. This indicates that the networks are utilizing any extra dimensions to encode the noise in the image or background sources that might have been missed from the image cleaning process, about which we also do not care. This is also reflected in the larger error bars when we get to this many 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 minimize the latent space in order to ensure it is interpretable and physically meaningful. In earlier iterations of this work we utilized five 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, while higher than first expected, is much lower than previous works that allowed their feature space to get very large in order to achieve the best reconstructions possible, thus rendering their feature space uninterpretable and difficult to map to physical features. For the rest of the analysis in this work we utilize the MMD-VAE network as both networks have a similar performance; however, for the reasons stated earlier in Section 3.4, 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.

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 (HC), as no prior knowledge of how many “true” clusters there are within the data is required.

6.1. Hierarchical Clustering

Within this work we utilize the HC algorithm (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 utilize in this work. An example of an HC dendrogram is shown in Figure 6. This method has been used in similar studies for classifying different morphological classes at low redshift (Cheng et al. 2021).

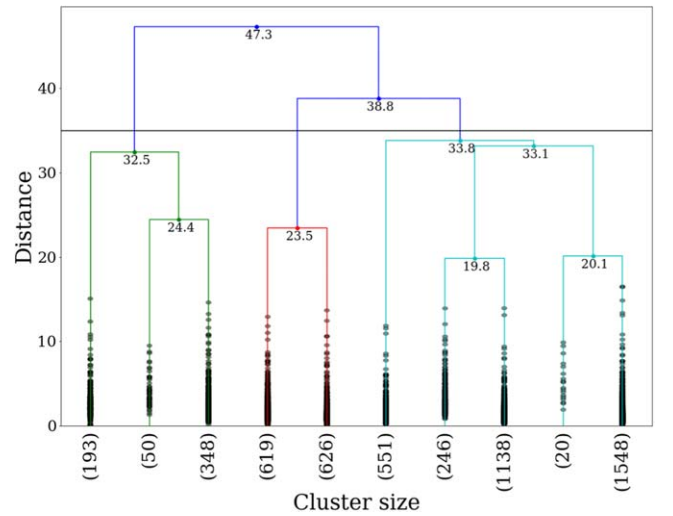


Figure 6. Dendrogram of the HC algorithm. The y-axis shows the similarity or distance between the points while the x-axis shows the number of data points in the smallest clusters. The horizontal black line shows the cutoff point in similarity that would result in three clusters.

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 minimize 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 minimized, the resulting clusters are created by grouping the closest points in the feature space together. In order to select clusters using this method, traditionally a single distance is used as a cutoff point, and then 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 S/N of the sample or if the cluster has fewer 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 Section 7.2.

7. Results

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 overfitting. 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 subsamples, indicating that the network architecture we have trained is robust. An example of some reconstructions from the network can be seen in Figure 7. The top row shows the input augmented JWST images, the middle shows the reconstructions from the network, and the final row shows the residuals between the input and the reconstructions. It can be seen that the network is able to encode the general morphology

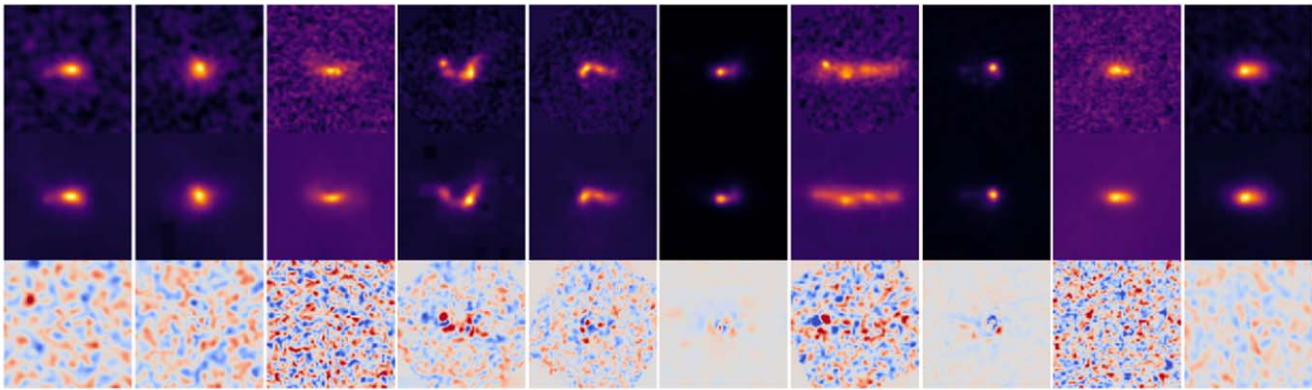


Figure 7. Top: images input to the encoder. Middle: reconstructed images from the test set using 23 encoded features. Bottom: residuals showing how the reconstructions have encoded the galaxy light and not noise. It can be seen that the network is capturing the overall morphologies of the galaxy sample well. Note that, while 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 classify different galaxy morphologies robustly, e.g., clumpy vs. nonclumpy morphology and not end up with many more groups split by small variations in the location of clumps, as in some previous works. This allows us to have a broader, more robust separation of galaxy morphologies.

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, as our reconstructions are very smooth and noise has been removed from the images.

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 understand the encoded space better we investigate if there are any correlations between known galaxy parameters, both parametric and nonparametric, and individual latent features. We compute the Spearman’s correlation coefficient between each latent feature and our measured morphology to understand better what is being encoded. The highest ranked latent feature and the corresponding measurement can be seen in Table 3. (The full correlation matrix can be seen in Appendix Figure A2.)

In order to visualize this correlation better we plot each correlated feature versus the latent dimension with which it had the highest Spearman’s correlation coefficient. We split each feature into bins to visualize better where different galaxy types lie along each feature. These can be seen in Figure 8. Looking at the parametric and nonparametric 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 Figure A3. 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’s correlation coefficients with measured physical and structural properties, as shown in Figure 8.

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 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

Table 3
The Highest Correlation Value and Corresponding Galaxy Property for Each Latent Feature

Latent Feature	Correlated Feature	Spearman’s 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

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 predefined assumptions and uncover high star-forming galaxies with many different morphologies. We explore the morphology of high-sSFR galaxies in Section 7.7.

Another method for visualizing 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 utilize the UMAP algorithm (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 visualize 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 Figure 9. Each UMAP visualization is colored 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, Sérsic index, and axis ratio. We can also see a trend in the sSFR with the lower-sSFR galaxies residing in the lower-right region of the UMAP. However, as mentioned above, the sSFR does not correlate as well as the other measurements. From this we can see that our encoded representation space correlates with

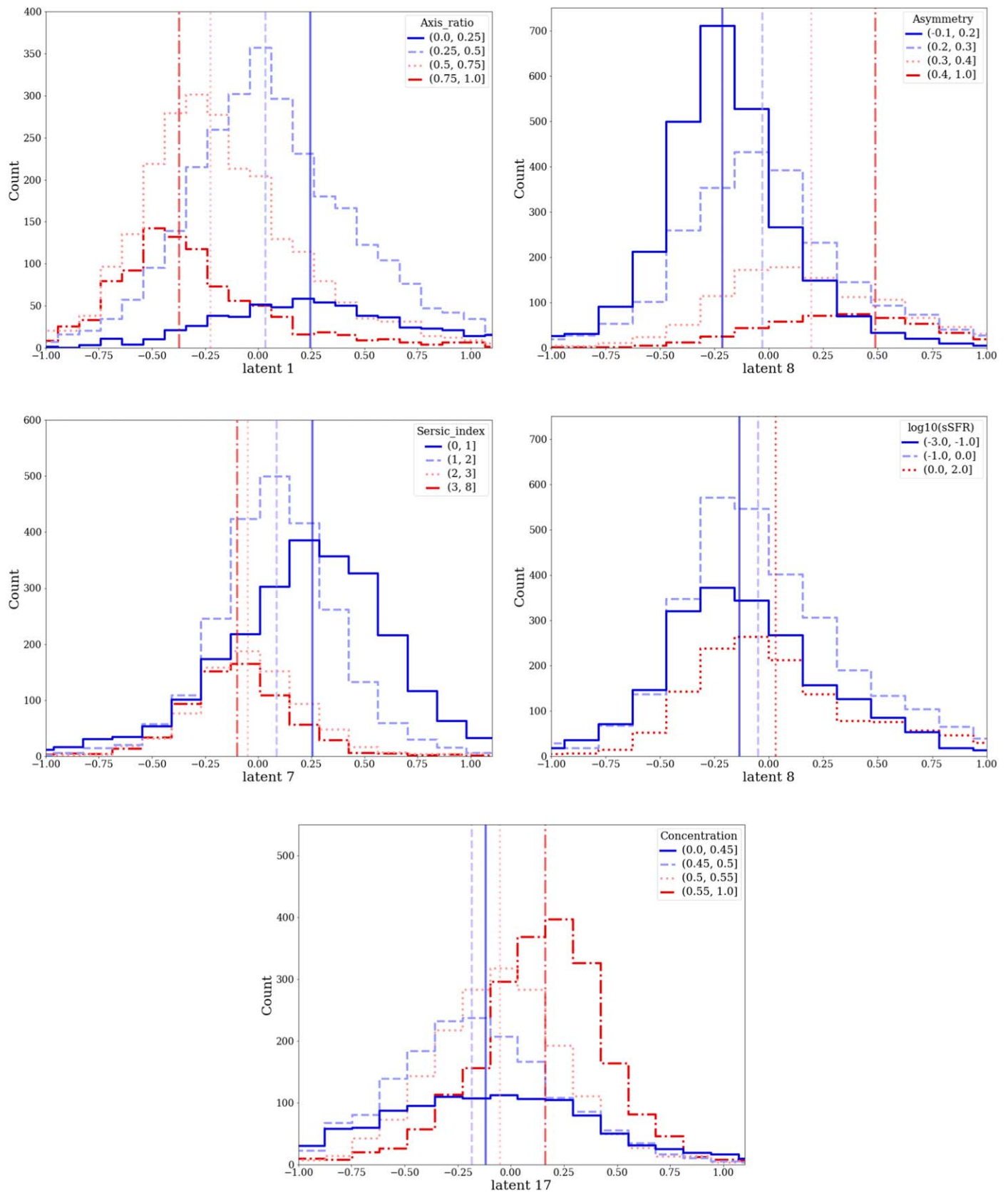


Figure 8. Correlation between features in the latent space and known measured properties for our galaxy sample. Each measured property is split into bins and plotted against the latent feature with which it had the largest Spearman’s correlation coefficient. The median value for each selected range is also plotted. The ranges selected for each parameter can be seen in the legend. It can be seen that the features that the network are learning are physically meaningful and correlate with well-known galaxy properties.

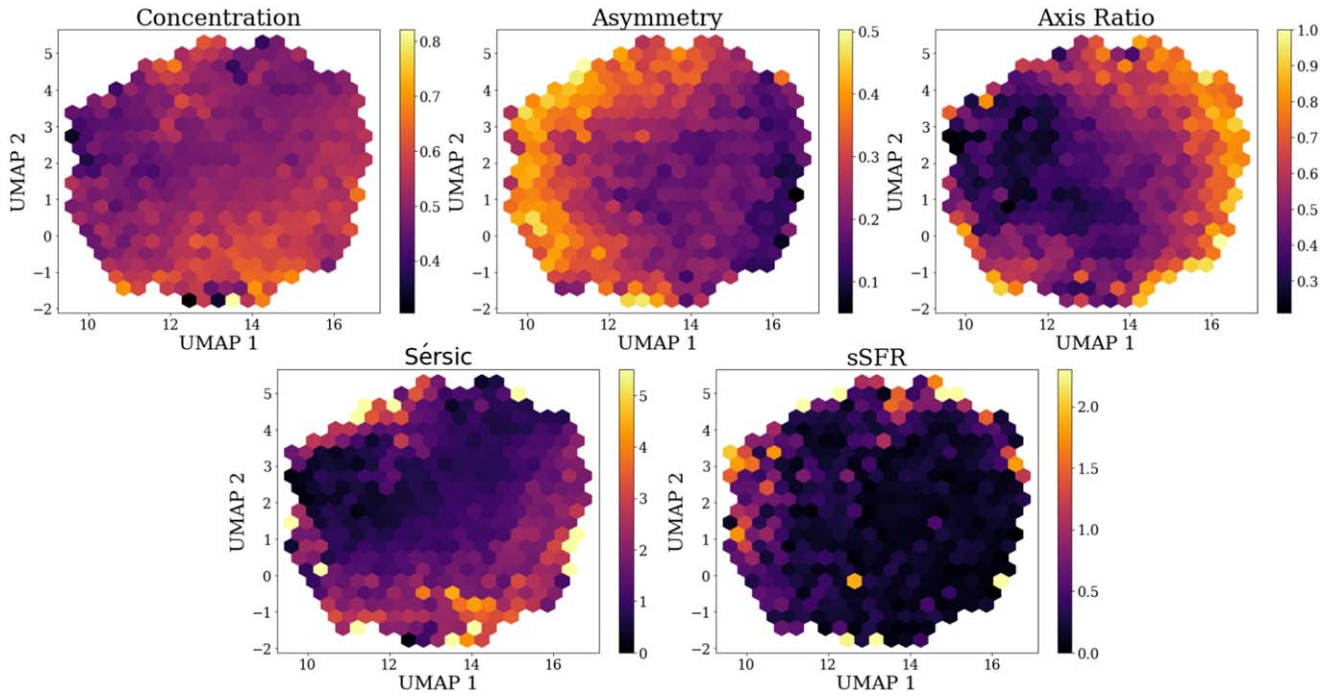


Figure 9. Uniform Manifold Approximation and Projection (UMAP) visualization of our 23 latent features for our entire galaxy sample colored by different measured morphological parameters. Each plot is colored by the median value in each hexagonal bin in the 2D UMAP plane. From left to right shows the distribution of the concentration, asymmetry, axis ratio, Sérsic index, and the specific star formation rate (sSFR) for our galaxy sample.

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. (2024). 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 different types of galaxies.

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 Section 6. We find a total of 11 clusters after removing the clusters that extracted the lowest-S/N galaxies. These low-S/N galaxies accounted for 5% of the sample and had an average S/N per pixel (S/N_p) of $S/N_p \sim 2$ (see Tohill et al. 2021 for calculation of S/N_p). The median S/N_p for the rest of our groups was ~ 10 for comparison. The S/N_p distributions for our groups can be seen in Appendix Figure A5. Randomly selected cutouts of galaxies from the 11 clusters found can be seen in Figure 10. 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 labeled as a supergroup in Figure 10 along with the dominant morphology of galaxies along this branch. The supergroup label is not designed to be a morphological classification as each supergroup is composed of galaxies with distinct structural features; 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. Supergroup 2.2 for example is labeled 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 possesses structural parameters more associated with irregular-type galaxies.

Next, we compare our machine-extracted clusters to previous works that also utilize unsupervised ML to extract morphological clusters. For example, 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 of 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) include galaxies up to $z \sim 4$ and also include low-redshift galaxies, which 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.

While our machine-found clusters appear to be different galaxy types by eye, we want to investigate their structural and physical properties to determine if they also correlate with their apparent morphologies.

7.3. Comparison to Structural and Physical Properties

We use 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 nonparametric morphological parameters of each galaxy using MORFOMETRYKA (Ferrari et al. 2015) as

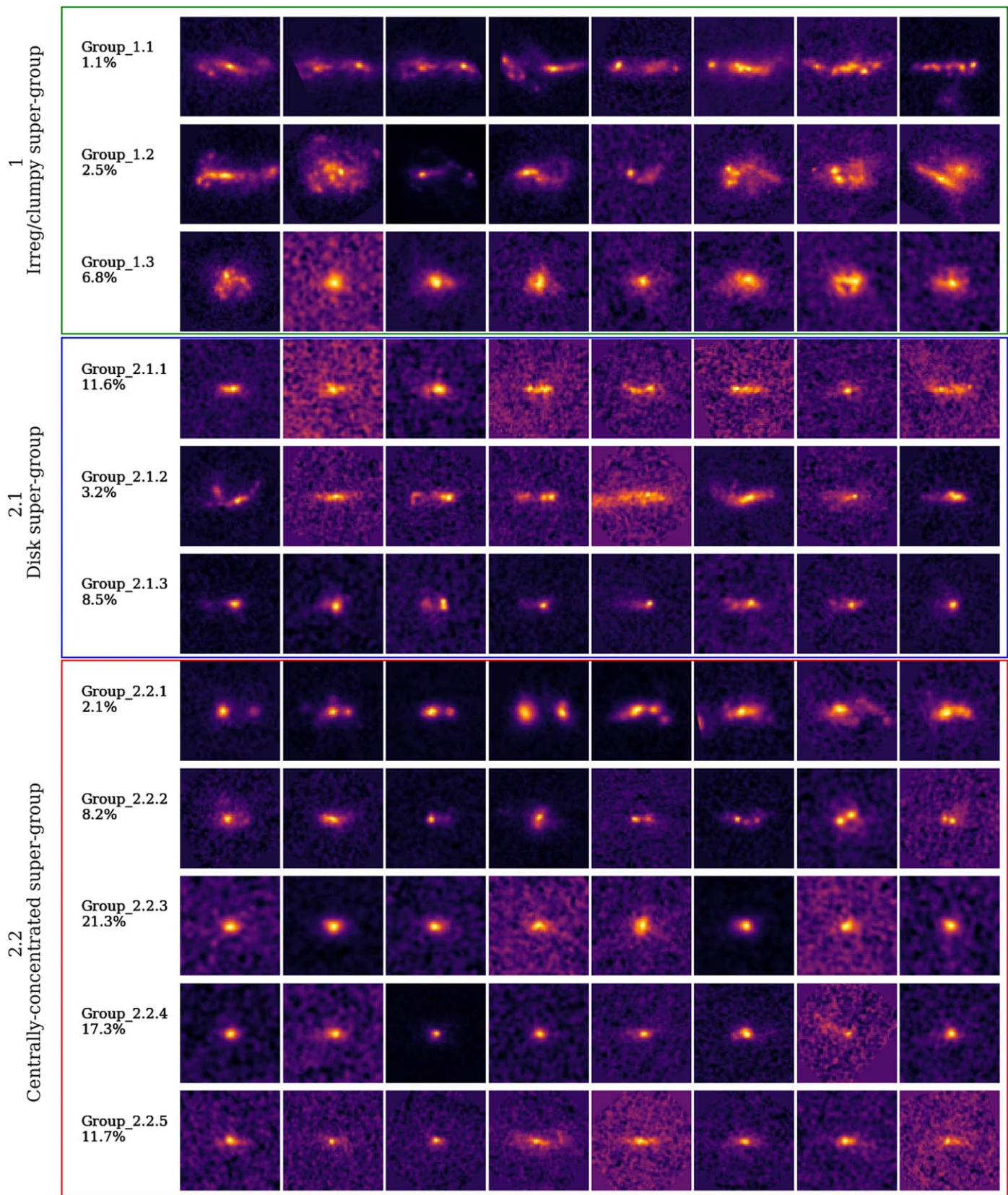


Figure 10. Examples of galaxies from each group found by the HC algorithm. It can be seen that each group is visually distinct from each other. Group names indicate which parent branch they belong to from the HC tree, and parent branches are labeled as supergroups along with the dominant morphology along that branch.

these will benefit from the higher resolution offered by JWST, allowing more accurate measurements compared to those measured with the HST images. MORFOMETRYKA uses

nonstandardised images, along with the associated point-spread function (PSF) images, to calculate these parameters. Investigating the distributions of these structural measurements will give better

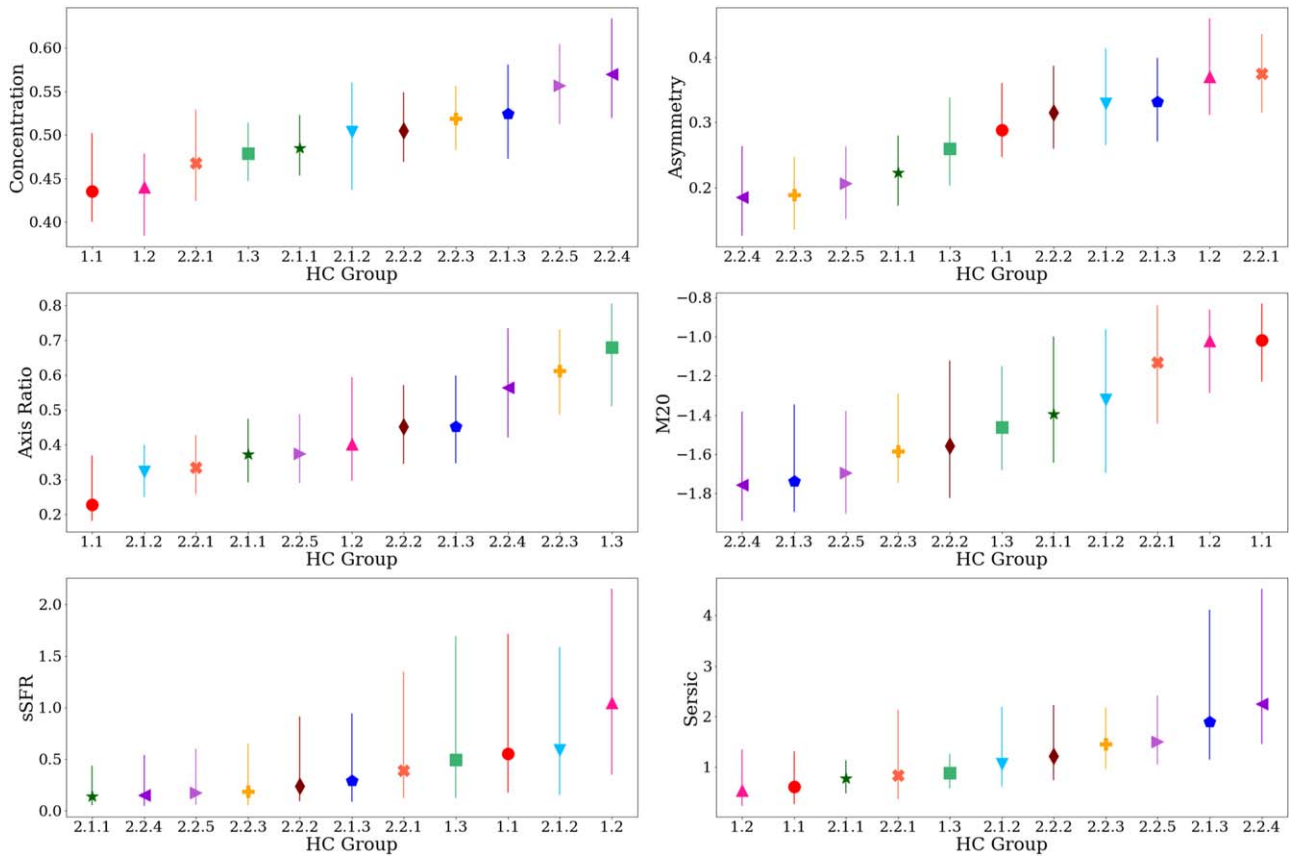


Figure 11. Distributions of the properties of the galaxies within each group determined by the HC algorithm. Each group is distinct from each other in both parametric and nonparametric morphology measurements and also physical properties such as the sSFR. Groups are ordered in increasing value within in measurement. Colored points and symbols are the median, which match across all figures. The error bars show the 25th and 75th percentiles of the distribution of values within each cluster. Note that the error on the mean is ~ 20 times smaller, hence the trends we see are of high significance.

insight into the physical properties of the galaxies within each group, as seen in Figure 11.

Figure 11 shows the median values of the different physical and structural parameters for each of our machine-defined clusters. The error bars show the 25th and 75th percentiles of the distribution of values within each group. It can be seen that each of our morphological groups possesses 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 are still clear transitions between the properties of the galaxies as you move between different morphological classes. Note that the error on the mean is ~ 20 times smaller than the interquartile range. The groups have been ordered in increasing value for each parameter to show this smooth transition better. 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 sSFRs, 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 groups 1.1 and 1.2, which possess almost the opposite properties, with asymmetric light distributions, low central concentrations, and the highest sSFRs of the sample, agreeing with our knowledge that star-forming clumps 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 standardization process has removed the majority of observational effects from the latent space, allowing for a more robust classification scheme. The redshift distributions of our groups are shown in Appendix Figure A4.

Figure 12 shows where each cluster lies in the C–A plane as well as the distribution in Sérsic index with asymmetry. Each colored 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 emphasizes 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 Figure 12 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 4.

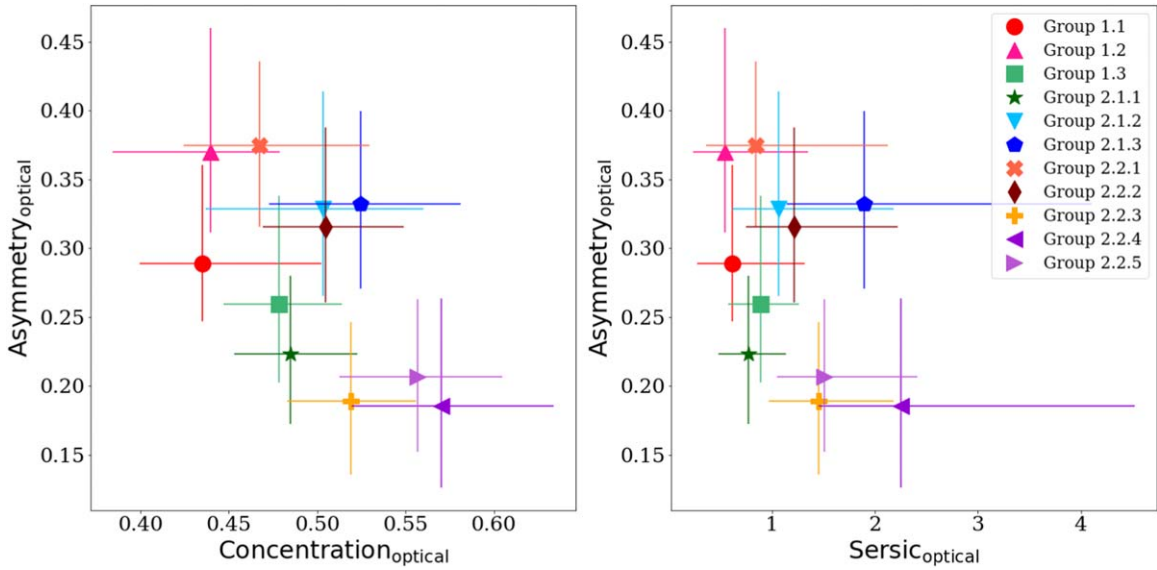


Figure 12. Distributions of the structural properties of the galaxies within each group determined by the HC algorithm. Each point represents one of our groups from our unsupervised ML clustering. Points show the median values and the error bars show the 25th and 75th percentiles of the values within each cluster. The symbols and colors are the same as in Figure 11.

Table 4

Brief Description of Each Group’s Morphology, Number of Samples in Each Cluster, and Median Values of the Structural Measurements

Group	No. of samples	Morphological description	$C_{1/2}$	$A_{1/2}$	$M_{20, 1/2}$	$sSFR_{1/2}$	$Sérsic_{1/2}$
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 semimajor 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

Note. The horizontal lines show the supergroup splits.

Below we give a brief description of the main morphological features of each cluster and compare these to previous studies of high-redshift morphology.

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 the predictions by Ceverino et al. (2015) and Tomassetti et al. (2016).

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 views 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 ratios and asymmetries. 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).

7.3.3. Group 1.3: Clumpy Disks

These galaxies, while 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 high sSFRs, which lead to some clumpy features.

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 an intermediate concentration indicating no clear central region. They possess the lowest sSFRs of our sample with low Sérsic indices. This lack of ongoing star formation could indicate that these galaxies are more evolved with an outside-in formation, so that they will perhaps form their bulge later in their evolution via means other than star formation.

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 M_{20} values due to tails and plumes caused by some interactions. As there are no companions visible, it is possible that these galaxies have been caught after a recent merger, which could account for their disturbed morphologies.

7.3.6. Group 2.1.3: Tadpole Galaxies

Closely resembling a “tadpole-type” galaxy (van den Bergh 1998), with an offset nucleus and a tail resembling the shape of a tadpole. These galaxies are very asymmetric along their semimajor axis and have a range of axis ratios. The origin of this structure is not fully understood; in some cases they could be caused by an offset burst of star formation, tidal interactions, ram pressure stripping, or accretion of cosmic gas (Elmegreen & Elmegreen 2010).

7.3.7. Group 2.2.1: Close Pairs/Double Clumps

Visually dominated by two close objects or two main clumps, these are the most asymmetric galaxies in our sample. They are similar in morphology to those found in 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).

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 are to the left of the main object as opposed to the right with the previous group. This is due to our image standardization 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.

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 Section 7.4.

7.3.10. Group 2.2.4: Compact Spheroids

With 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 discussed in Section 7.4 in more detail. They make up a total of 17.3% of our sample.

7.3.11. Group 2.2.5: Bulge and Disk Components

The final class of galaxy found by our clustering technique is galaxies with edge-on disks and 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. The fact we see these types of galaxies at high redshift is an indicator that bulge and disk formation is already in place very early on.

7.4. Comparison to Visual Classifications

While the aim of this work is to provide a robust, nonbiased approach to classifying galaxy morphology, we want to compare how the network’s 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 ~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 Figure 13 (the same plot but for the supergroups is shown in Appendix Figure A6). 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 into which each galaxy had to be classified. These are the classic spheroid, disk, peculiar, and PSF labels that have been utilized 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.

While we can see some correlations with the visual classifications, such as 70% of the spheroid classifications residing in clusters 2.2.3 and 2.3.4, which agrees with our assessment of these clusters and their structural parameters; and groups 2.1.1 and 2.2.5, which are disk dominated, agreeing

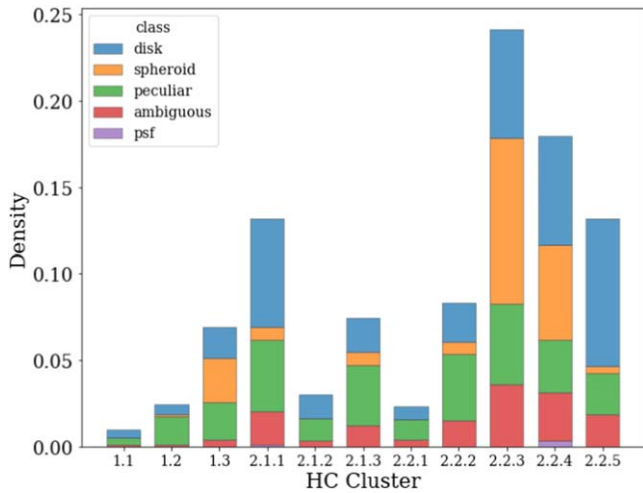


Figure 13. Bar plot of visual classifications from Ferreira et al. (2023) within each of our HC-found clusters. We have a visual classification for $\sim 50\%$ of our sample.

with our expectations there are few other strong correlations. The aim of this work is to show that these traditional “morphological types” are not representative of the wide variety of galaxy morphologies and structures we see in the high- z Universe and that we need to understand better 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 the misclassification of face-on disks with spheroids, and so suggest a combination of visual classifications and structural parameters would help to resolve this issue. We are combining such an approach in our feature extraction process here, as the network has all the information about the pixel-by-pixel light distribution of each galaxy. In recent work, Vega-Ferrero et al. (2024) 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 nonoblate structure.” Looking at our clusters we find that group 2.2.3, while having $\sim 45\%$ of the spheroid classifications, also has a large proportion of disk-like classifications. This could be due, in part, to the reasons stated by Vega-Ferrero et al. (2024) above. The properties of this group more closely resemble spheroid-like galaxies with high concentrations of light, low sSFRs, low asymmetries, and larger Sérsic indices.

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 subsample of images with detailed measurements and then applied to a much larger, unlabeled data set, ideal for the future of galaxy surveys and galaxy evolution studies.

7.5. Impact of Noise

One common issue with visual classifications and both parametric and nonparametric 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 utilizing unsupervised ML to extract different morphological features, and why we optimize our latent space to focus on encoding the galaxy light and not the background noise (see Section 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 S/N level of our images. To test this we select a sample of 15 high-S/N ($14 < S/N_p < 20$) galaxies from each machine-defined cluster and simulate each galaxy at various S/N_p levels from 12 down to 2 (for details of how these images are simulated see Tohill et al. 2021). An example of the simulated noisy images can be seen in Figure 14. We assume for this test that these high-S/N_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 Figure 15. The black points show the fraction of galaxies where their classifications changed and the error bars show 1σ of the binomial distribution. The first point is plotted at an S/N_p of 14, and it is located at zero in the plot as we are assuming these galaxies to be in the correct cluster. The median S/N_p of our entire galaxy sample is shown by the red vertical line. It can be seen that at the median S/N_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 S/N_p level. As the S/N_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 fractions of galaxies changing 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 nonparametric measurement system. A good example of this can be seen in Figure 14, where the fainter structures in the galaxy become indistinguishable from the background noise around S/N_p = 5. From this we can argue that the different morphological clusters found by our method are robust to various levels of noise however, for S/N_p < 6 there is less certainty of the morphological classification from this method. For reference, 82% of our sample has S/N_p > 6, so we can be confident in the majority of our morphological classifications.

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 of the fraction of galaxies in each cluster split into four bins: $2 < z < 3$, $3 < z < 4$, $4 < z < 5$, and $z > 5$. In order to ensure any

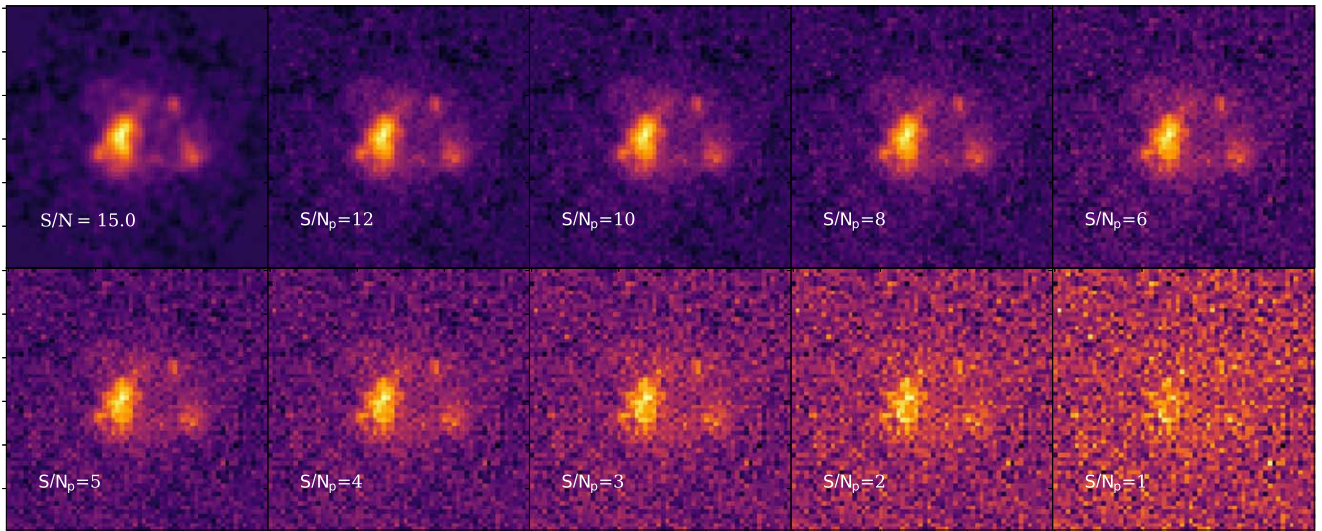


Figure 14. An example of our simulated noisy galaxy images. In the top-left panel we show the original JWST cutout followed by the same galaxy at different simulated S/N_p levels.

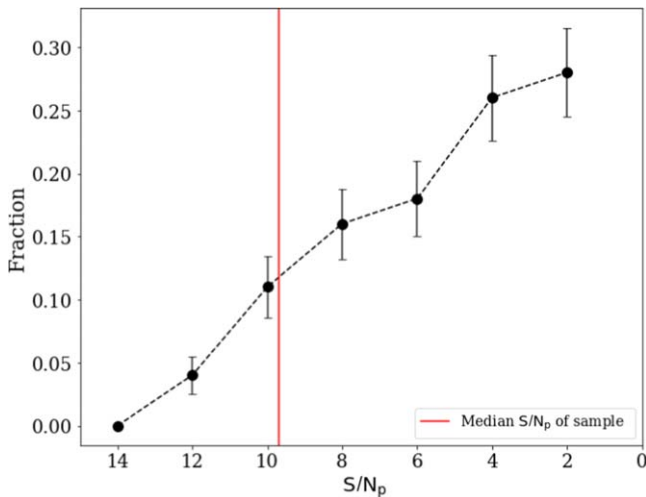


Figure 15. Fraction of galaxies that change classification in our machine-defined clusters vs. the S/N_p of the galaxy. The errors are given as 1σ of the binomial distribution. The red vertical line shows the median S/N_p of our full sample.

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 Figure 16. The groups are split by their HC supergroup for ease of visualization and so it is easier to compare the 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 we will need larger samples to study this in depth, which will be possible in the near future with more data from JWST. While 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. Huertas-Company et al. (2023) 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 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), $> 10^9 M_\odot$, and investigate our two spheroid-dominated classes (2.2.3 and 2.2.4), we find that the amount of galaxies decrease 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 be dominant or already in place at high redshift (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. However, it is a good sign that even with the variety of morphologies that are found by exploring the network-extracted features, similar results are found in other works, who used very different methods. We will study the evolution of these galaxies with redshift in more depth in a follow-up paper.

7.7. Morphology of High- $sSFR$ Galaxies

We analyze the morphology of high star-forming galaxies within our machine-found clusters by making a cut at $\log(sSFR) > 0.5$, as this is representative of our galaxy population in the high- z regime. This cut can be seen in Figure 17. While we are most likely missing a lot of the fainter and lower-mass galaxies at high redshift, we expect to be more complete for the

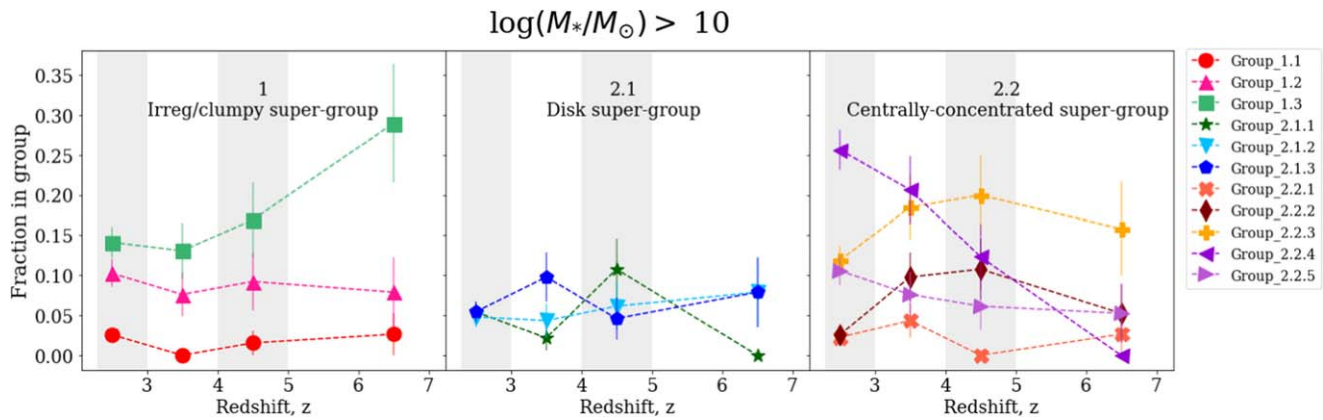


Figure 16. Evolution of the most massive galaxies within each morphological cluster with redshift. The errors are given as 1σ of the binomial distribution. The colors and symbols are the same as in Figures 11 and 12. The groups are divided by their HC supergroup for ease of visualization and comparison. The shaded regions show the bin widths for each point. The last bin includes all galaxies above $z > 5$. See text for details.

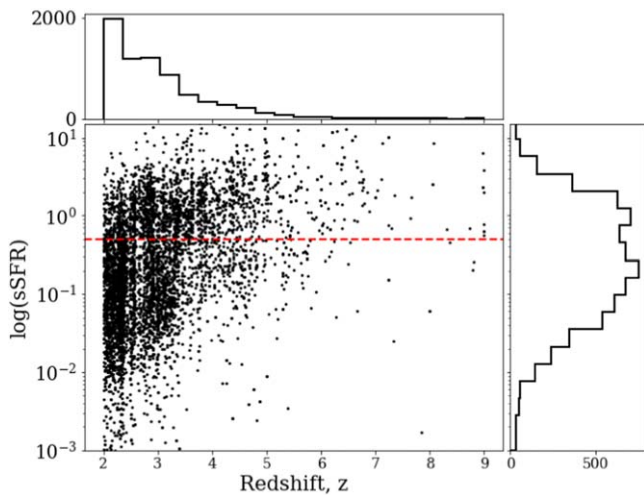


Figure 17. Distribution of the sSFRs of galaxies within our matched sample vs. redshift. The red line shows the sSFR cut of $\log(\text{sSFR}) > 0.5$.

high-sSFR galaxies as their high SFRs are what allows us to detect even the most distant galaxies.

We find that four 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 four 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 as structural parameters as these groups also possess the lowest concentrations of light and Sérsic indices, and possess high asymmetry values. On top of these, they are also the two highest- M_{20} groups, showing that these systems could be picked up easily using traditional methods. The clump disk galaxy type in our classification scheme appears visually to be slightly messy, containing some clumpy morphologies, which could indicate they are undergoing intense star formation. However, looking at the nonparametric measurements alone we find this group to have intermediate CAS and M_{20} values and so could be missed if selection cuts were to be made using

these structural measurement systems. They are, however, found by our classification system, showing again that our network is able to pick up on features that are missed by current measurements. The final group we look at is the disturbed disks, which, like our clump disk group, are asymmetric but possess intermediate concentrations 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 neighbors, it could be argued these galaxies have recently undergone a merger, which could account for both their disturbed morphologies 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 constrain their formation and evolution better. We acknowledge that the classifications presented in this work do not allow us to confirm the formation histories 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.

8. Summary

We present our work utilizing unsupervised ML to perform feature extraction on high-redshift galaxies imaged with JWST. We apply an HC 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 NIRCcam on JWST we find 11 separate morphological clusters that possess different morphological features, physical properties, and structural measurements, e.g., sSFR, CAS- M_{20} 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:

1. We remove the observational biases imposed on the images by standardizing 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.
2. 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.
3. 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 nonparametric parameters.
4. We explore the highest-redshift sample to date utilizing unsupervised ML. 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 summarized below.

1. We find that there is a wide variety of galaxy morphologies already in place at high redshift. In total we find 11 distinct morphological types for our sample.
2. We confirm that our unsupervised machine-defined clusters support work to construct a visual classification scheme suitable for high redshift, while sidestepping the issue of applying predefined categories to new observational regimes.
3. We find a decrease in concentrated spheroidal-type galaxies with redshift as found by others, and find that disk-like galaxies dominate at high redshift, though these are typically clumpy and/or disturbed in morphology.
4. 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 provide a more detailed and better suited classification system.

As mentioned in the paper, 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 analyses, all of which will help improve galaxy evolution studies such as the work carried out in this paper. Such approaches to galaxy morphology classification are also required to handle the amount of data expected with the future of JWST and similar surveys.

Acknowledgments

The authors would like to thank the University of Nottingham for providing the computational infrastructure needed to produce the networks used in this paper. We also thank Phil Parry for timely help with this infrastructure that made this work possible. C.T. and T.H. acknowledge funding from the Science and Technology Facilities Council (STFC). We acknowledge support from the ERC Advanced Investigator Grant EPOCHS (788113). We thank the CEERS team for making their data open to the public, which we have made use of within this work. We thank the anonymous referee for their thorough review, which significantly helped to improve the presentation of this work. This work is based on observations made with the NASA/ESA Hubble Space Telescope (HST) and NASA/ESA/CSA James Webb Space Telescope (JWST) obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute (STScI), which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST and NAS 526555 for HST. The specific observations analyzed can be accessed via doi:[10.17909/z7p0-8481](https://doi.org/10.17909/z7p0-8481).

Appendix A Dimensionality of the Latent Space

In Figure A1 we show the variations in the reconstruction of both the training and test sets versus the dimensionality of the latent space. The solid points show the average MSE for the last 50 epochs (our patience value, see text for details) and the error bars show the 1σ spread of the variations. As with the validation set (see Figure 5) the reconstruction loss will continue to improve with more latent features, however the aperture loss plateaus off. This indicates that the network is using the extra dimensions to encode some background noise in the images, about which we do not care.

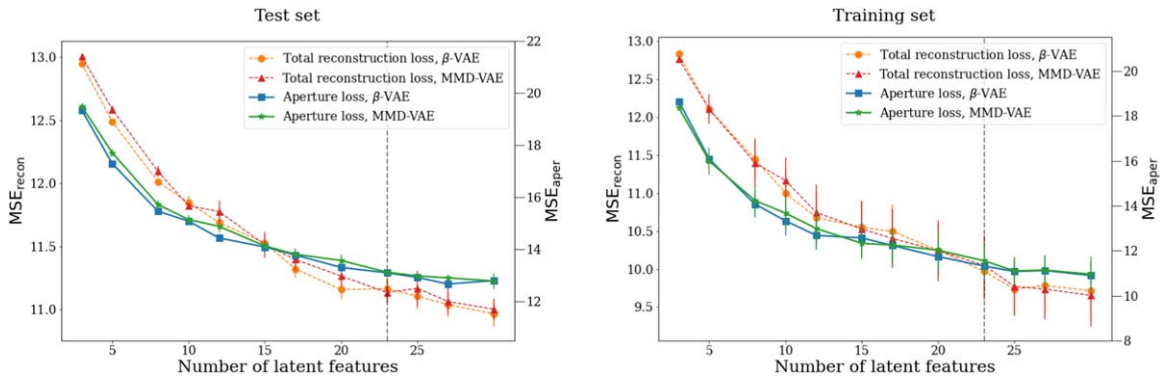


Figure A1. Variation in the total reconstruction loss and the reconstruction loss within an aperture (see Section 5.1 for details) for the test and training sets (same as in Figure 5). The solid points show the average MSE for the last 50 epochs and the error bars show the 1σ spread of the variations. The dashed line shows the chosen dimensionality of the latent space for our analysis.

Appendix B Visualizing the Latent Space

In Figure A2 we show the full Spearman’s correlation matrix for each of our 23 latent dimensions and the measured galaxy properties. In Table 3 we list the highest correlated latent variable for each of the galaxy properties. 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 M_{20} . 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 the 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 for the galaxies.

In Figure A3 we show generated images from our trained latent space. Plotted are the latent features that had the highest Spearman’s correlation coefficients with measured physical and structural properties. These are not real galaxy images but images generated from the encoded space learned by our network. We generate these images by calculating the mean and range of each latent feature and then interpolating along each feature, generating a new image with the network each time, while the other latent features are kept constant. This shows that our network has learned a smooth feature space (which is expected with a VAE), as the morphology transitions smoothly along, and between, each latent feature. We can also see correlations with known morphologies. 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.

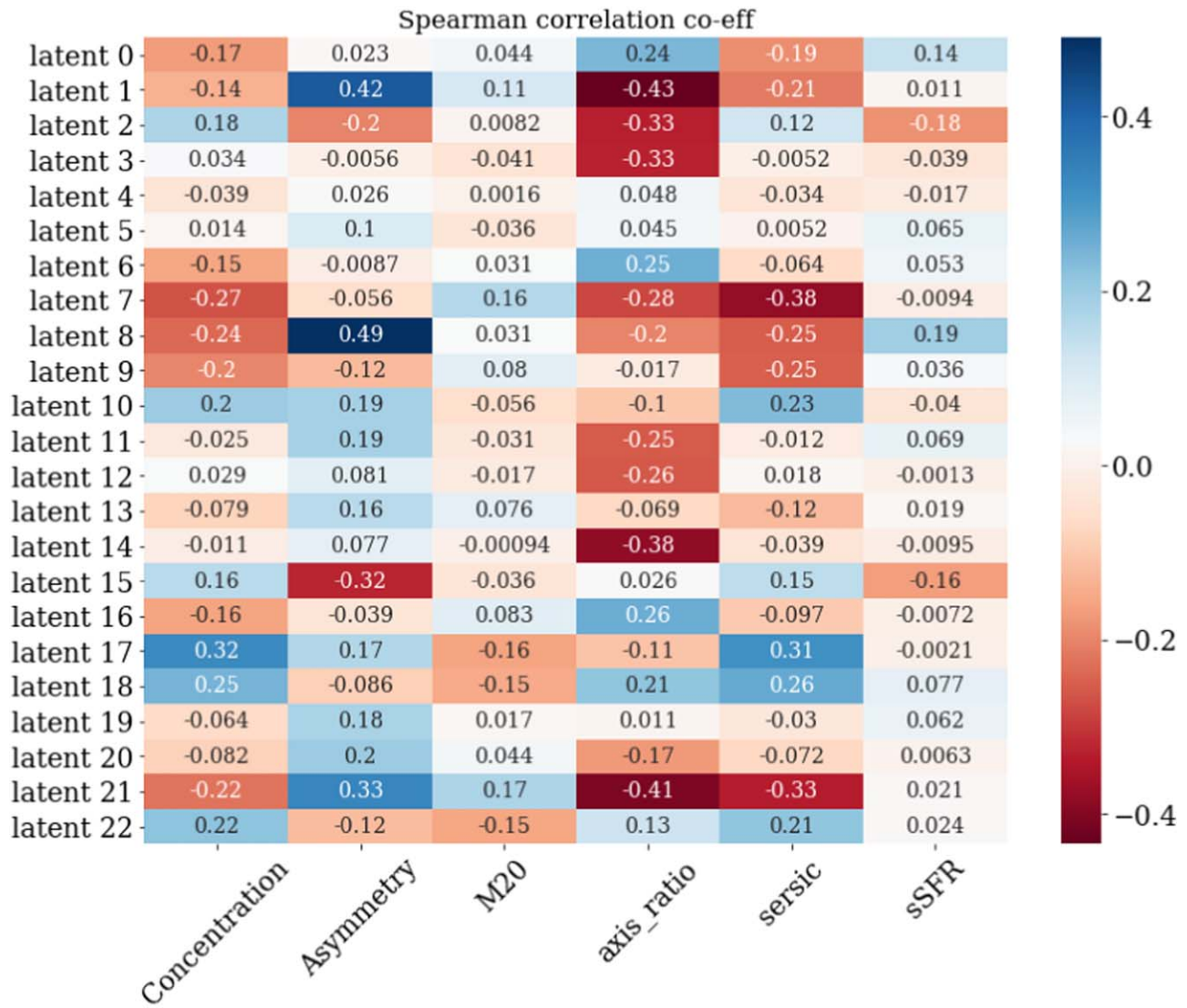


Figure A2. Distribution of the Spearman’s rank correlation coefficients for each latent dimension in our trained MMD-VAE and each measured galaxy property.

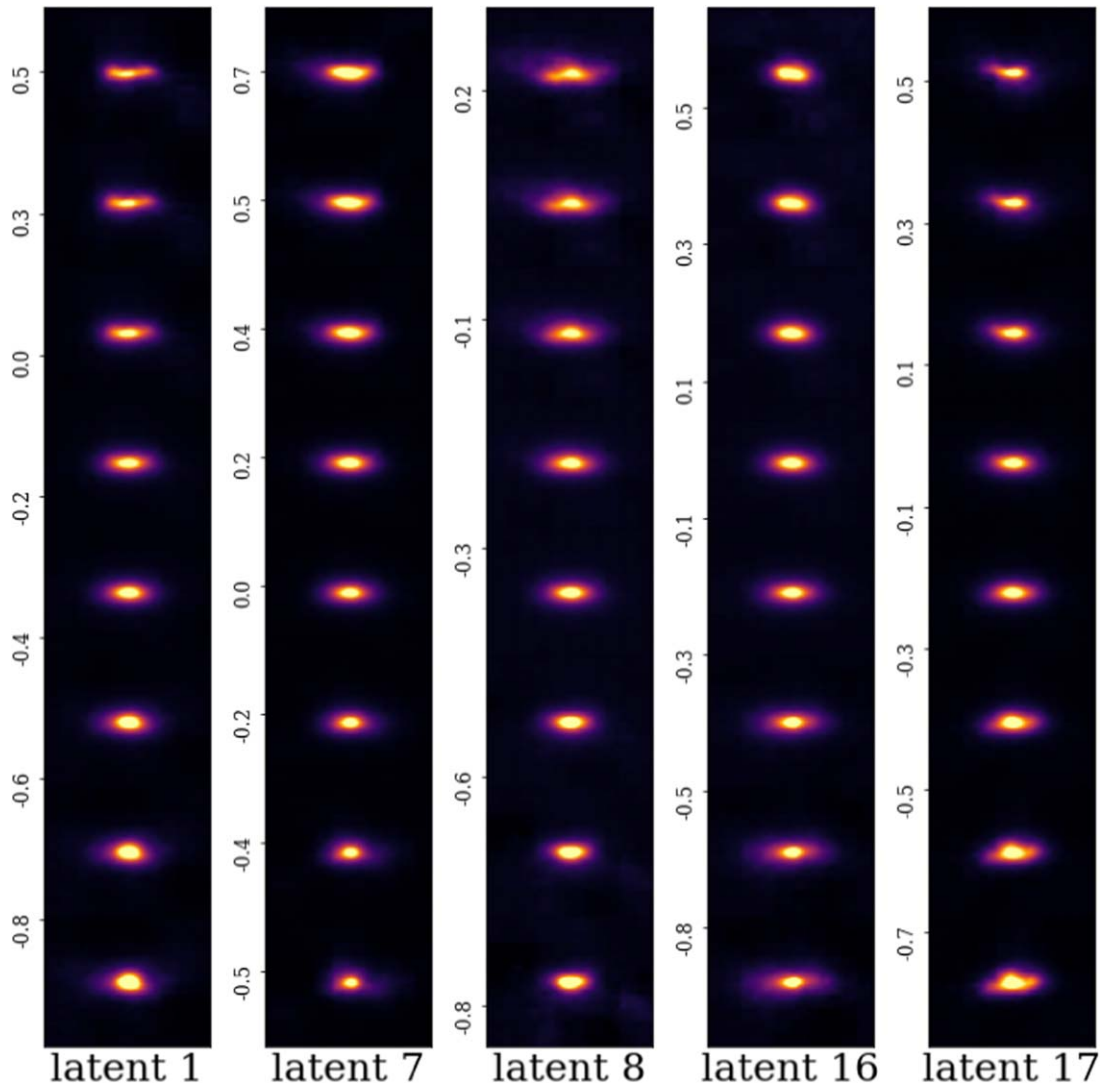


Figure A3. Generated images from the trained latent space. Here we show examples of generated galaxy images along different dimensions in the latent space to show the smooth transition between morphologies. The latent dimensions shown are those with the highest correlation to measured properties (see Figure 8 for more details). We calculate the mean of each feature and linearly interpolate between $\pm 2\sigma$ while keeping all other latent features constant to show the morphology encoded by each dimension.

Appendix C Correlation of Latent Space with Nonphysical Properties

In Figure A4 we show the distribution of the redshifts within each morphological class. It can be seen that there is no significant difference between any of the groups and that they all span similar ranges, indicating that our network is not splitting based on redshift effects or different observational bands. This further supports the power of our standardization process at removing observational effects from our images,

such as apparent size and brightness, as otherwise the network would have used these features to separate the groups. It can be seen that the groups in the first supergroup have slightly larger errors than the other groups, but this is due to the small number of galaxies in each of these groups as they are the most irregular of our sample. The same can be seen for the S/N_p of each group in Figure A5. We see that the network is not using the noise in an image to split the groups. This is important as we have taken care to try to prevent the network from learning noise in an image as a feature (see Section 5.1).

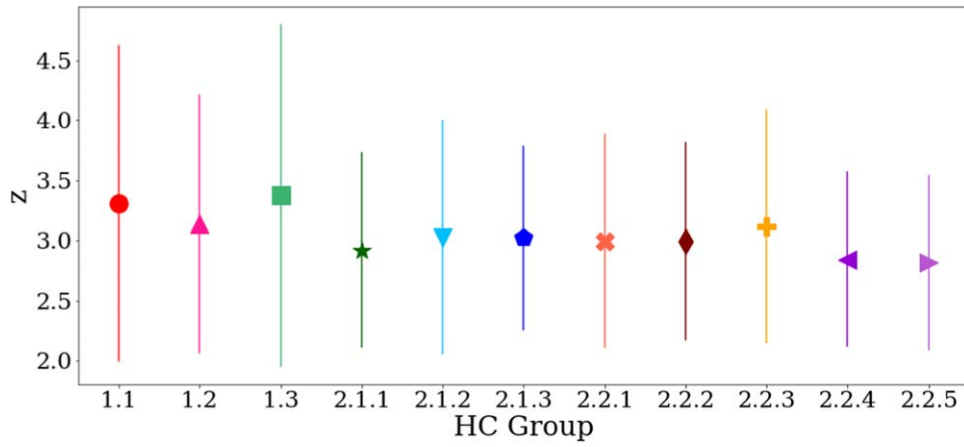


Figure A4. Distributions of the redshifts of galaxies within each morphological cluster. It can be seen that the network does not split groups based on their redshift, which further supports the claim that our standardization approach successfully removes many observational biases such as apparent size and apparent brightness. This also allows us to study further the evolution of galaxies within each morphological class with cosmic time. The colored points and symbols are the same as the other figures in this work. The solid points show the means and the error bars show the standard deviation within each group.

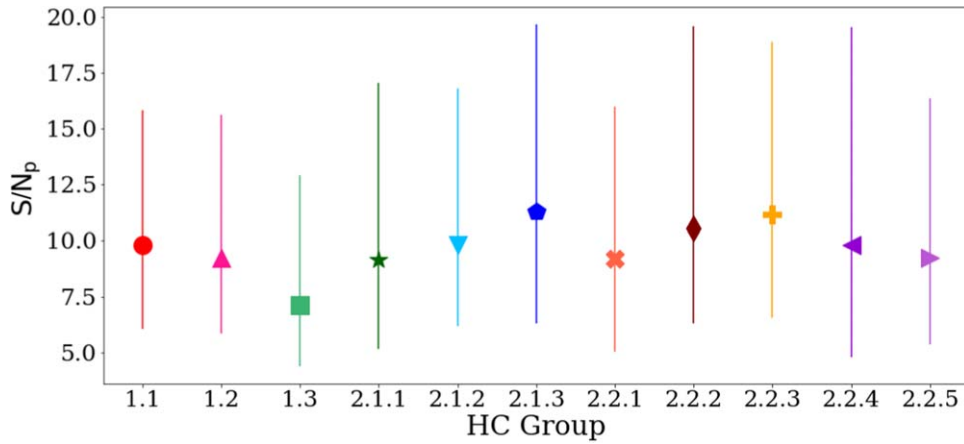


Figure A5. Distributions of the S/N_p of galaxies within each morphological cluster. It can be seen that the network does not split groups based on their S/N_p . The colored points and symbols are the same as the other figures in this work. The solid points show the means and the error bars show the standard deviation within each group.

Appendix D Comparison to Visual Classifications

In Figure A6 we show the comparison between our HC supergroups and the visual classifications from Ferreira et al. (2023). As

it can be seen there are no strong correlations except for the centrally concentrated supergroup (2.2) possessing most of the spheroidal classifications. This further supports the argument that the visual classification systems used at low redshift are not suitable for the high-redshift regime.

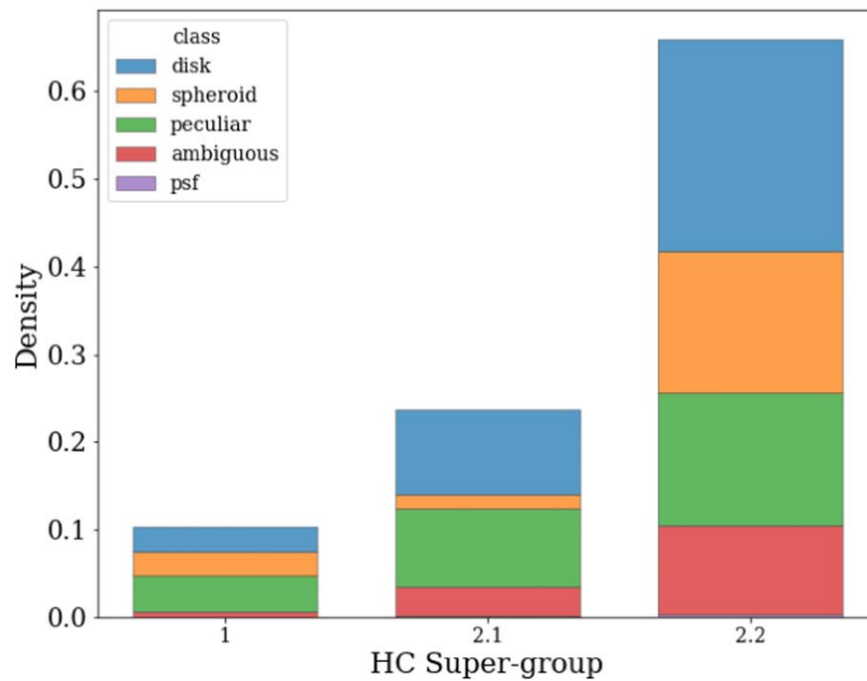


Figure A6. Bar plot of visual classifications from Ferreira et al. (2023) within each of our HC-found supergroups. We have a visual classification for $\sim 50\%$ of our sample.

ORCID iDs

C. Tohill <https://orcid.org/0000-0003-2527-0819>
 S. P. Bamford <https://orcid.org/0000-0001-7821-7195>
 C. J. Conselice <https://orcid.org/0000-0003-1949-7638>
 L. Ferreira <https://orcid.org/0000-0002-8919-079X>
 T. Harvey <https://orcid.org/0000-0002-4130-636X>
 D. Austin <https://orcid.org/0000-0003-0519-9445>

References

- Adams, N. J., Conselice, C. J., Ferreira, L., et al. 2023, *MNRAS*, 518, 4755
 Bagley, M. B., Finkelstein, S. L., Koekemoer, A. M., et al. 2023, *ApJL*, 946, L12
 Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, *MNRAS*, 393, 1324
 Baron, D., & Poznanski, D. 2017, *MNRAS*, 465, 4530
 Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. 2011, in Annual Conf. on Neural Information Processing Systems 25 (NIPS 2011), Advances in Neural Information Processing Systems, ed. R. Zemel et al. (Granada: Neural Information Processing Systems Foundation), M46
 Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, *AJ*, 119, 2645
 Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, *MNRAS*, 399, 1191
 Ceverino, D., Primack, J., & Dekel, A. 2015, *MNRAS*, 453, 408
 Chen, X., Duan, Y., Houthoofd, R., et al. 2016, arXiv:1606.03657
 Cheng, T., Huertas-Company, M., Conselice, C. J., et al. 2021, AAS Meeting, 53, 103.05
 Cheng, T.-Y., Conselice, C. J., Aragón-Salamanca, A., et al. 2020b, *MNRAS*, 493, 4209
 Cheng, T.-Y., Li, N., Conselice, C. J., et al. 2020a, *MNRAS*, 494, 3750
 Conselice, C. J. 2003, *ApJS*, 147, 1
 Conselice, C. J., Blackburne, J. A., & Papovich, C. 2005, *ApJ*, 620, 564
 Conselice, C. J., Mortlock, A., Bluck, A. F. L., Grützbauch, R., & Duncan, K. 2013, *MNRAS*, 430, 1051
 Cowie, L. L., Hu, E. M., & Songaila, A. 1995, *AJ*, 110, 1576
 de Albernaz Ferreira, L., & Ferrari, F. 2018, *MNRAS*, 473, 2701
 Dieleman, S., Willett, K. W., & Dambre, J. 2015, *MNRAS*, 450, 1441
 Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, *MNRAS*, 476, 3661
 Dressler, A. 1980, *ApJ*, 236, 351
 Duncan, K., Conselice, C. J., Mundy, C., et al. 2019, *ApJ*, 876, 110
 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
 Elmegreen, D. M., Elmegreen, B. G., & Hirst, A. C. 2004, *ApJL*, 604, L21
 Elmegreen, D. M., Elmegreen, B. G., Rubin, D. S., & Schaffer, M. A. 2005, *ApJ*, 631, 85
 Ferrari, F., de Carvalho, R. R., & Trevisan, M. 2015, *ApJ*, 814, 55
 Ferreira, L., Adams, N., Conselice, C. J., et al. 2022, *ApJL*, 938, L2
 Ferreira, L., Conselice, C. J., Duncan, K., et al. 2020, *ApJ*, 895, 115
 Ferreira, L., Conselice, C. J., Sazonova, E., et al. 2023, *ApJ*, 955, 94
 Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, *ApJL*, 946, L13
 Gretton, A., Borgwardt, K., Rasch, M. J., Scholkopf, B., & Smola, A. J. 2008, arXiv:0805.2368
 Grogan, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35
 Gullberg, B., Smail, I., Swinbank, A. M., et al. 2019, *MNRAS*, 490, 4956
 Guo, Y., Jooe, S., Finkelstein, S. L., et al. 2023, *ApJL*, 945, L10
 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
 Holmberg, E. 1958, *MeLuS*, 136, 1
 Huertas-Company, M., Iyer, K. G., Angeloudi, E., et al. 2023, arXiv:2305.02478
 Hutter, F., Hoos, H. H., & Leyton-Brown, K. 2011, in Learning and Intelligent Optimization, ed. C. A. C. Coello (Berlin: Springer), 507
 Jacobs, C., Glazebrook, K., Calabrò, A., et al. 2023, *ApJL*, 948, L13
 Johnson, S. C. 1967, *Psychometrika*, 32, 241
 Kartaltepe, J. S., Rose, C., Vanderhoof, B. N., et al. 2023, *ApJL*, 946, L15
 Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, *MNRAS*, 341, 54
 Kingma, D. P., & Welling, M. 2013, arXiv:1312.6114
 Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36
 Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
 Li, R., Napolitano, N. R., Roy, N., et al. 2022, *ApJ*, 929, 152
 Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, *MNRAS*, 389, 1179
 Locatello, F., Bauer, S., Lucic, M., et al. 2018, arXiv:1811.12359
 Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, *ApJ*, 672, 177
 Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, 128, 163
 Madau, P., & Dickinson, M. 2014, *ARA&A*, 52, 415
 Margalef-Bentabol, B., Huertas-Company, M., Charnock, T., et al. 2020, *MNRAS*, 496, 2346
 Martin, G., Kaviraj, S., Hocking, A., Read, S. C., & Geach, J. E. 2020, *MNRAS*, 491, 1408
 McInnes, L., Healy, J., & Melville, J. 2018, arXiv:1802.03426

- Mortlock, A., Conselice, C. J., Hartley, W. G., et al. 2013, *MNRAS*, **433**, 1185
- Patton, D. R., Pritchet, C. J., Carlberg, R. G., et al. 2002, *ApJ*, **565**, 208
- Abraham, R. G., van den Bergh, S., Glazebrook, K., et al. 1996, *ApJS*, **107**, 1
- Robertson, B. E., Tacchella, S., Johnson, B. D., et al. 2023, *ApJL*, **942**, L42
- Sandage, A. 1986, *A&A*, **161**, 89
- Schade, D., Lilly, S. J., Crampton, D., et al. 1995, *ApJL*, **451**, L1
- Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, *MNRAS*, **440**, 889
- Snoek, J., Rippel, O., Swersky, K., et al. 2015, arXiv:1502.05700
- Spindler, A., Geach, J. E., & Smith, M. J. 2021, *MNRAS*, **502**, 985
- Thorne, B., Knox, L., & Prabhu, K. 2021, *MNRAS*, **504**, 2603
- Tohill, C., Ferreira, L., Conselice, C. J., Bamford, S. P., & Ferrari, F. 2021, *ApJ*, **916**, 4
- Tomassetti, M., Dekel, A., Mandelker, N., et al. 2016, *MNRAS*, **458**, 4477
- Toomre, A., & Toomre, J. 1972, *ApJ*, **178**, 623
- Tortorelli, L., & Mercurio, A. 2023, *FrASS*, **10**, 51
- Tuccillo, D., Huertas-Company, M., Decencière, E., et al. 2018, *MNRAS*, **475**, 894
- van den Bergh, S. 1998, *Galaxy Morphology and Classification* (Cambridge: Cambridge Univ. Press)
- van den Bergh, S., Abraham, R. G., Ellis, R. S., et al. 1996, *AJ*, **112**, 359
- van den Bergh, S., Abraham, R. G., Whyte, L. F., et al. 2002, *AJ*, **123**, 2913
- van der Wel, A., Chang, Y.-Y., Bell, E. F., et al. 2014, *ApJL*, **792**, L6
- Vega-Ferrero, J., Huertas-Company, M., Costantin, L., et al. 2024, *ApJ*, **961**, 51
- Walmsley, M., Lintott, C., Géron, T., et al. 2022a, *MNRAS*, **509**, 3966
- Walmsley, M., Scaife, A. M. M., Lintott, C., et al. 2022b, *MNRAS*, **513**, 1581
- Whitney, A., Ferreira, L., Conselice, C. J., & Duncan, K. 2021, *ApJ*, **919**, 139
- Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, *MNRAS*, **435**, 2835
- Wilman, D. J., Fossati, M., Mendel, J. T., et al. 2020, *ApJ*, **892**, 1
- Xu, Q., Shen, S., de Souza, R. S., et al. 2023, *MNRAS*, **526**, 6391
- Zhang, H., Primack, J. R., Faber, S. M., et al. 2019, *MNRAS*, **484**, 5170
- Zhao, S., Song, J., & Ermon, S. 2017, arXiv:1706.02262