



HAL
open science

Galaxy Morphology from $z \sim 6$ through the eyes of JWST

M Huertas-Company, K.G Iyer, E Angeloudi, M.B Bagley, S.L Finkelstein, J Kartaltepe, R Sarmiento, J Vega-Ferrero, P. Arrabal Haro, P Behroozi, et al.

► **To cite this version:**

M Huertas-Company, K.G Iyer, E Angeloudi, M.B Bagley, S.L Finkelstein, et al.. Galaxy Morphology from $z \sim 6$ through the eyes of JWST. *Astron.Astrophys.*, 2024, 685, pp.A48. 10.1051/0004-6361/202346800 . hal-04114716

HAL Id: hal-04114716

<https://hal.science/hal-04114716v1>

Submitted on 17 Jan 2025






HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Galaxy morphology from $z \sim 6$ through the lens of JWST[★]

M. Huertas-Company^{1,2,3,4,30}, K. G. Iyer^{5,★★}, E. Angeloudi^{1,4}, M. B. Bagley⁷ , S. L. Finkelstein⁷, J. Kartaltepe¹⁸, E. J. McGrath²⁹ , R. Sarmiento^{1,4}, J. Vega-Ferrero^{1,4,9,10}, P. Arrabal Haro⁶, P. Behroozi^{8,9}, F. Buitrago^{10,11} , Y. Cheng¹², L. Costantin¹³, A. Dekel^{14,15}, M. Dickinson⁶, D. Elbaz¹⁶, N. A. Grogin¹⁹, N. P. Hathi¹⁹ , B. W. Holwerda¹⁷, A. M. Koekemoer¹⁹, R. A. Lucas¹⁹, C. Papovich^{20,21}, P. G. Pérez-González¹³, N. Pirzkal²², L.-M. Seillé²³, A. de la Vega²⁴, S. Wuyts²⁵ , G. Yang^{26,27}, and L. Y. A. Yung²⁸

(Affiliations can be found after the references)

Received 2 May 2023 / Accepted 8 October 2023

ABSTRACT

Context. The *James Webb* Space Telescope's (JWST's) unprecedented combination of sensitivity, spatial resolution, and infrared coverage has enabled a new era of galaxy morphology exploration across most of cosmic history.

Aims. We analyze the near-infrared (NIR $\sim 0.8\text{--}1\ \mu\text{m}$) rest-frame morphologies of galaxies with $\log M_*/M_\odot > 9$ in the redshift range of $0 < z < 6$, compare with previous HST-based results and release the first JWST-based morphological catalog of $\sim 20\,000$ galaxies in the CEERS survey.

Methods. We classified the galaxies in our sample into four main broad classes: spheroid, disk+spheroid, disk, and disturbed, based on imaging with four filters: *F150W*, *F200W*, *F356W*, and *F444W*. We used convolutional neural networks (CNNs) trained on HST/WFC3 labeled images and domain-adapted to JWST/NIRCam.

Results. We find that $\sim 90\%$ and $\sim 75\%$ of galaxies at $z < 3$ have the same early and late and regular and irregular classification, respectively, in JWST and HST imaging when considering similar wavelengths. For small (large) and faint objects, JWST-based classifications tend to systematically present less bulge-dominated systems (peculiar galaxies) than HST-based ones, but the impact on the reported evolution of morphological fractions is less than $\sim 10\%$. Using JWST-based morphologies at the same rest-frame wavelength ($\sim 0.8\text{--}1\ \mu\text{m}$), we confirm an increase in peculiar galaxies and a decrease in bulge-dominated galaxies with redshift, as reported in previous HST-based works, suggesting that the stellar mass distribution, in addition to light distribution, is more disturbed in the early Universe. However, we find that undisturbed disk-like systems already dominate the high-mass end of the late-type galaxy population ($\log M_*/M_\odot > 10.5$) at $z \sim 5$, and bulge-dominated galaxies also exist at these early epochs, confirming a rich and evolved morphological diversity of galaxies ~ 1 Gyr after the Big Bang. Finally, we find that the morphology-quenching relation is already in place for massive galaxies at $z > 3$, with massive quiescent galaxies ($\log M_*/M_\odot > 10.5$) being predominantly bulge-dominated.

Key words. catalogs – galaxies: evolution – galaxies: high-redshift – galaxies: statistics – galaxies: structure

1. Introduction

Understanding how galaxy diversity emerges across cosmic time is one of the major goals of the field of galaxy formation. Questions as to how and when stellar disks form, along with what the main drivers of bulge growth are, as well as how and when galaxy morphology and star formation became connected, remain largely unanswered despite significant progress in recent years (e.g., Conselice 2014; Förster Schreiber & Wuyts 2020). Galaxy morphology remains a key first-order proxy of galaxy diversity since it is a fast and cost-effective way of identifying different types of galaxies in large samples at different cosmic times, solely requiring imaging. Until now, almost all studies involving galaxy structure beyond the local Universe have made use of the *Hubble* Space Telescope (HST), which is the only facility delivering a high enough spatial resolution to study the structure of small high-redshift galaxies (e.g., Abraham et al. 1996; Huertas-Company et al. 2015). As a result, HST imaging has revealed that galaxies at $z > 1$ tend to become more irregular in their light distribution (e.g., Conselice et al. 2000), even if observed in the rest-frame optical more sensitive

to main-sequence stars (e.g., Huertas-Company et al. 2016). It has also revealed that star-forming and quiescent galaxies have different stellar structures at all cosmic epochs probed so far (e.g., Wuyts et al. 2011; van der Wel et al. 2014a; Pearson et al. 2021; Dimauro et al. 2022) and many others. The recent launch of the *James Webb* Space Telescope (JWST) has opened a new window onto galaxy morphology, especially at early epochs, by providing unprecedented spatial resolution and depth combined with infrared coverage that allows for the optical rest-frame emission in the first billion years of the Universe's history to be probed. A few early works analyzing the first set of JWST images started exploring the morphological diversity up to $z \sim 6$ using visual classifications or deep learning models trained on HST images (e.g., Ferreira et al. 2022, 2023; Kartaltepe et al. 2023; Robertson et al. 2023). The general conclusion coming from these initial studies is that there might be more disk-like galaxies at high redshift than previously inferred with HST, although the exact nature of these disk-like galaxies still needs to be confirmed (Vega-Ferrero et al. 2024).

In this work, we take a step forward toward understanding the morphological diversity of galaxies in the early Universe by providing the first and largest publicly available morphological classification of $\sim 20\,000$ galaxies selected in the JWST Cosmic Evolution Early Release Science (CEERS, Finkelstein et al. 2022, 2023) survey and observed in four different wavelengths (*F150W*, *F200W*, *F356W*, and *F444W*). Galaxies are classified

[★] The catalog is available at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/685/A48>

^{★★} Hubble Fellow.

into four broad morphological classes using a convolutional neural network (CNN) trained on HST-based classifications from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS, Koekemoer et al. 2011; Grogin et al. 2011) and domain-adapted to work on JWST images. We then use the new classifications to precisely quantify the changes in galaxy morphology when moving from HST to JWST, study the evolution of morphological fractions from $z \sim 6$ to $z \sim 0$ in the NIR rest-frame ($\sim 0.8\text{--}1\ \mu\text{m}$) by using different filters at different Finally, we revisit the morphology-quenching relation over $\sim 90\%$ of the cosmic history.

The paper proceeds as follows. In Sect. 2, we present the data sets used, namely the CEERS and the CANDELS surveys. Section 3 describes the method used to estimate morphologies on JWST images and Sect. 4 systematically compares HST and JWST based galaxy morphologies. Section 5 studies the evolution of morphological fractions between $z \sim 0$ and $z \sim 6$ as well as the morphology-quenching relation. Section 6 offers a discussion of some of the implications of the results presented. A summary is presented in Sect. 7. Throughout this paper, we use a *Planck* 2013 cosmology (Planck Collaboration XVI 2014).

2. Data

2.1. CEERS

CEERS (Finkelstein et al. 2022, 2023) is an Early Release Science (ERS) program (Proposal ID 1345, PI: Finkelstein) that has been observing the EGS (Extended Groth Strip, Davis et al. 2007) extragalactic deep field (one of the five CANDELS fields, Koekemoer et al. 2011; Grogin et al. 2011) since June 2022, with data made available to the public immediately. For this work, we combined ten NIRC*am* pointings observed in June and December 2022 in four different filters: *F150W*, *F200W*, *F356W*, and *F444W*. For the June pointings, we used the images from the latest CEERS public data release¹. A detailed description of the reduction process can be found in Bagley et al. (2023). For the December data, we used internal data products from the CEERS team reduced with an analogous procedure.

In addition to imaging data, we also used photometric redshifts and physical properties of galaxies derived with Spectral Energy Distribution (SED) fitting. Photometric redshifts were derived using the EAZY code (Brammer et al. 2008), as detailed in Finkelstein et al. (2022). The physical properties of galaxies (i.e., stellar masses and star formation rates for this work) have been determined using the DENSE BASIS method² (Iyer et al. 2019), an SED fitting method that uses a procedure similar to that in Iyer et al. (2019), Olsen et al. (2021), Finkelstein et al. (2022), Mowla et al. (2022), Asada et al. (2023). The code performs a fully Bayesian inference of the Star Formation History (SFH), dust attenuation, and chemical enrichment for each galaxy, using a fully non-parametric Gaussian process-based description for the SFH described in Iyer et al. (2019). The model uses a Chabrier IMF (Chabrier 2003), a Calzetti dust attenuation law (Calzetti et al. 2000), and Madau IGM absorption (Madau 1995). In the fitting process, a redshift prior taken from the confidence interval of the EAZY estimate is used. Setting explicit priors in SFH space using non-parametric SFHs has been shown to be robust against outshining due to younger stellar populations, which can otherwise bias estimates of masses and star formation rates (Iyer & Gawiser 2017; Leja et al. 2019; Lower et al. 2020).

¹ <https://ceers.github.io>

² <https://dense-basis.readthedocs.io/>

We ran GALFIT (Peng et al. 2010) v3.0.5 on the background-subtracted version of the CEERS imaging mosaics, holding the background value fixed at zero. We fit sources with $F356W < 28.5$ mag, using empirical PSFs generated by stacking stars across all CEERS fields. The ERR array, which includes background sky, Poisson, and read noise, was used as the input noise map. Fit regions for each source were determined based on the Kron radius, ultimately encompassing a region roughly 30 times the half-light radius of each source. All sources within the fitting region brighter than 27th magnitude and no more than three magnitudes fainter than the primary source were fit simultaneously. All other sources were masked during fitting. For more details, we refer to McGrath et al. (in prep.). In the analysis that follows, we consider only those sources that encountered no errors during fitting and whose best-fit magnitudes are consistent with the input photometry catalog (GALFIT flag value = 0).

2.2. CANDELS

We used publicly available images and data products from the CANDELS survey (Koekemoer et al. 2011; Grogin et al. 2011) from the five different fields (Galametz et al. 2013; Guo et al. 2013; Nayyeri et al. 2017; Stefanon et al. 2017; Barro et al. 2019). For training the neural networks (see Sect. 3), we used the morphological catalog of Huertas-Company et al. (2015), which provides neural network-based morphological labels trained on the Kartaltepe et al. (2015) visual classifications for all galaxies with $F160W < 24.5$. As the ground truth, we decided to use the labels from the CNN based classification of Huertas-Company et al. (2015) instead of directly using the visual classifications from Kartaltepe et al. (2015) because the former provides an homogeneous and larger sample. Given the excellent agreement shown in Huertas-Company et al. (2015) we do not expect this choice to introduce any significant bias. For this work, we defined four main morphological classes using the criteria also described in Huertas-Company et al. (2015), which have been demonstrated to provide reasonably clean classes. Namely: Class 1: spheroids/pure bulges: $f_{\text{sph}} > 2/3$, $f_{\text{disk}} < 2/3$, $f_{\text{irr}} < 1/10$; Class 2: disks: $f_{\text{sph}} < 2/3$, $f_{\text{disk}} > 2/3$, $f_{\text{irr}} < 1/10$; Class 3: bulge+disk: $f_{\text{sph}} > 2/3$, $f_{\text{disk}} > 2/3$, $f_{\text{irr}} < 1/10$; Class 4: irregulars/disturbed/peculiar: $f_{\text{irr}} > 1/10$; where f_{sph} , f_{disk} and f_{irr} are estimates of the vote fractions of different classifiers for spheroid, disk and irregular features, respectively (see Kartaltepe et al. 2015 and Huertas-Company et al. 2015 for more details). In particular, Huertas-Company et al. (2015) shows that such thresholds result in an agreement with the visual classification of $>95\%$ for the main morphological classes. For the remainder of this work, we consider the labels corresponding to the four classes as a ground truth. We use the terms “spheroids” or “pure bulges” interchangeably to refer to the first class, and “irregulars,” “disturbed,” or “peculiar” to refer to the last class. Additionally, we refer to the combination of classes 1 and 3 as early-type galaxies and the combination of classes 2 and 4 as late-type galaxies. The inclusion of bulge+disks systems into the early-type class is motivated by the results of Huertas-Company et al. (2015), who showed that the bulge+disk class is mainly composed of bulge-dominated systems. We discuss this further in the following sections. Because we removed stars using one of the filters with highest spatial resolution, we do not include a point source classification, which would furthermore decrease the accuracy of our classification because of low statistics.

It is important to note that the names given to the classes are solely used to identify different morphological properties

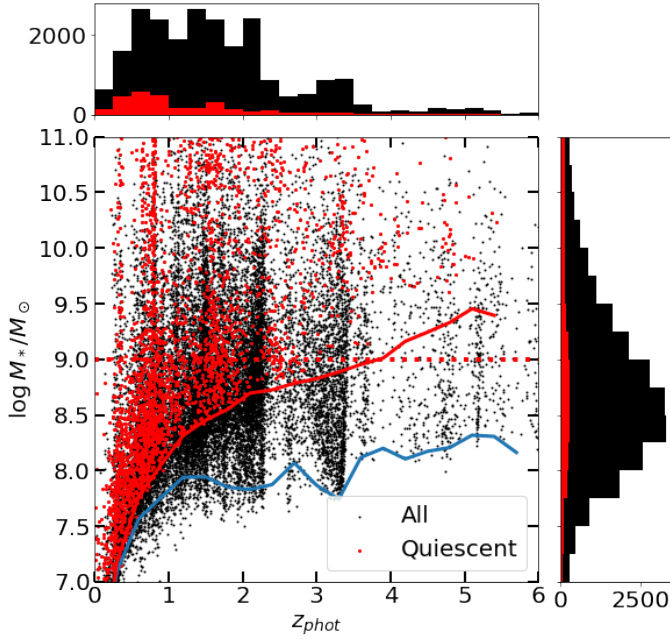


Fig. 1. Distribution of selected galaxies ($F200W < 27$) in the stellar mass-photometric redshift plane. Black points show the complete sample and red point are quiescent galaxies. The dashed red line shows the 10^9 solar mass limit used in this work for scientific analysis. The solid blue and red lines indicate the 90% mass completeness limit for the global sample and for quiescent galaxies respectively (see text for details). Overall our analysis threshold is well above the completeness limit except for the population of high-redshift quiescent galaxies.

based on images and do not necessarily reflect the true physical nature of these objects. In particular, the irregular class can contain a variety of galaxies with different physical properties. The spheroid class generally refers to round and compact galaxies which do not necessarily imply kinematically hot systems, especially at high redshift.

2.3. Sample selection and completeness

We select all galaxies with $F200W[AB] < 27$. The magnitude cut corresponds roughly to a signal-to-noise ratio (S/N) which enables a reliable morphological classification (see Kartaltepe et al. 2023). In addition to the apparent magnitude cut, we also removed obvious stars using a simple procedure described in Appendix A using photometric parameters measured in the $F200W$ filter. The final sample for which we estimate galaxy morphologies is then of 23 674 galaxies. Figure 1 shows the distribution of the selected galaxies in the stellar mass-photometric redshift plane. The lower envelope of the distribution can be used as a proxy for completeness, which is estimated at roughly $10^{8.5} M_{\odot}$ over the considered redshift range. Although the released morphological catalog includes all galaxies, in the remaining of this work, we primarily focus on galaxies more massive than 10^9 solar masses. For a rough estimate of the stellar mass completeness, we follow the method used in Pozzetti et al. (2010) and Ilbert et al. (2013), which is based in computing the lowest stellar mass (M_{lim}^*) which could be observed for each galaxy of magnitude $F200$ given the applied magnitude cut ($F200W[AB] < 27$): $\log(M_{\text{lim}}^*) = \log(M_{\text{lim}}^*) + 0.4(F200W[AB] - 27)$. The completeness is then estimated as the 90th percentile of the distribution of M_{lim}^* , namely, the stellar mass for which 90%

of the galaxies have lower limiting stellar masses. We repeated the measurement for the complete sample and for quiescent galaxies³ and report the results in Fig. 1. We see that the analysis adopted stellar mass threshold is conservatively above the completeness limit for the total sample. It is slightly below the 90% completeness for quiescent galaxies at $z > 4$. It should not significantly affect our main conclusions, as discussed in the following sections.

3. Inferring galaxy morphologies of CEERS galaxies

3.1. Method

Galaxy morphology is commonly estimated with supervised CNNs, which allow for an efficient extraction of image features correlated with galaxy morphology (see the review by Huertas-Company & Lanusse 2023). The main bottleneck of such an approach is the training set, which needs to contain a large enough sample of annotated images, typically performed through visual inspection. However, this is a time-consuming process that potentially undermines the advantage of using machine learning. Therefore, several works have been aimed at trying to find global features (e.g., Walmsley et al. 2022; Ćiprijanović et al. 2023) or transfer a network trained on one data set to another (e.g., Domínguez Sánchez et al. 2019).

Since there are very few available labels on JWST images, we used a standard adversarial domain adaptation (Ganin et al. 2016) in this work to transfer existing labels on HST imaging from the CANDELS survey to JWST-CEERS. An adversarial domain adaptation has been successfully applied to astronomical data in the context of galaxy merger identification (e.g., Ćiprijanović et al. 2020).

The overall idea is that both labeled images from HST and unlabeled images from JWST are shown to the CNN. An adversarial term is added to the loss functions to push the features extracted from the labeled data set to be close to the ones from the unlabeled one. This way, the network is prevented from learning specific features linked to a particular data set. More precisely, in this work, we use the network architecture illustrated in Fig. 2 and the following loss function:

$$E(\theta_f, \theta_y, \theta_d) = \sum_{i=1 \dots N} L_y(G_y(G_f(x_i^s; \theta_f); \theta_y), y_i) - \alpha L_d(G_d(G_f(x_i^t, x_i^s; \theta_f); \theta_d), k_i), \quad (1)$$

where G_f , G_y , and G_d are the feature extractor, classifier, and discriminator networks, respectively, with free parameters θ_f , θ_y , and θ_d . Then, L_y and L_d are the losses of the classifier and discriminator, which in our case are standard cross-entropy losses; x^s and x^t are source (CANDELS) and target (CEERS) images, respectively; y and k are the morphological classes and the data set class respectively; α is a scalar hyper-parameter that adjusts the weight between the two loss terms, thereby acting as a trade-off between classification accuracy and domain invariance. In this work, we set $\alpha = 1$ as in the original implementation. We investigated varying its value without significant impact on the final classification. Because there is a minus sign in Eq. (1), the network is effectively optimized so that the classification loss is minimized (i.e., classification accuracy), while the domain discriminator loss is maximized (i.e., domain invariance). Optimal

³ The procedure to select quiescent galaxies is detailed in Sect. 5.3.1.

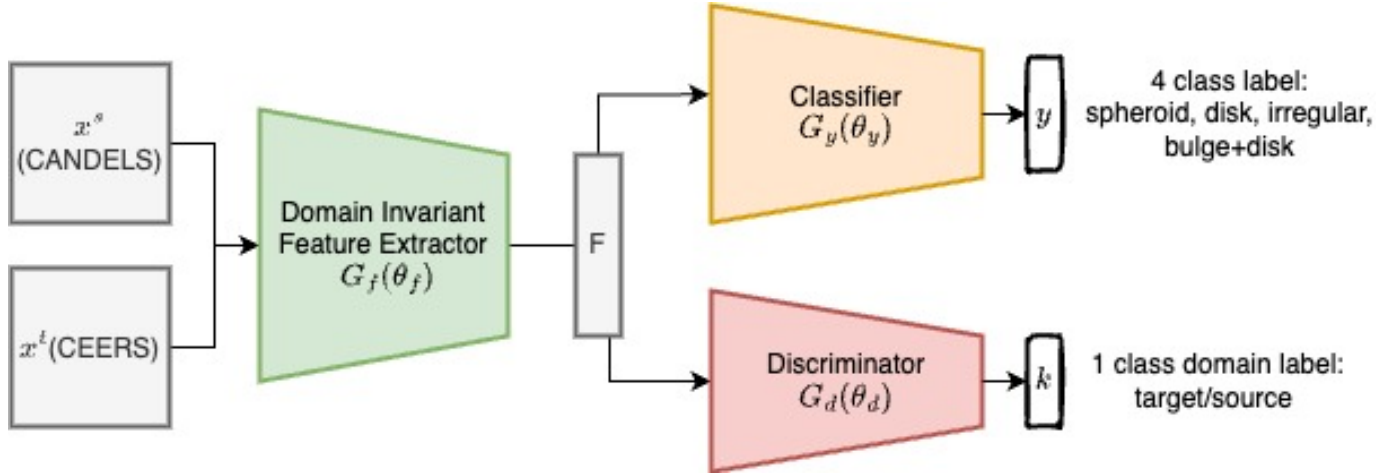


Fig. 2. Schematic representation of the neural network architecture used for classifying CEERS galaxies in this work. A first CNN ($G_f(\theta_f)$) is fed with both labeled and unlabeled CANDELS and CEERS stamps respectively. The computed features are then used as input for two additional CNNs: a discriminator ($G_d(\theta_d)$), which learns to distinguish stamps coming from the two data sets, and a classifier ($G_y(\theta_y)$) which provides a classification in four main morphological classes. More details about the training strategy can be found in the text.

values for the free parameters are thus given by:

$$\begin{aligned} (\hat{\theta}_f, \hat{\theta}_y) &= \arg \min_{\theta_f, \theta_y} E(\theta_f, \theta_y, \hat{\theta}_d), \\ \hat{\theta}_d &= \arg \max_{\theta_d} E(\hat{\theta}_f, \hat{\theta}_y, \theta_d). \end{aligned} \quad (2)$$

We trained four identical networks using as source images, $\sim 50\,000$ $F160W$ stamps from the CANDELS survey distributed in five different fields with labels from (Huertas-Company et al. 2015, see Sect. 2 for more details) and as our target, we used $F150W$, $F200W$, $F356W$ and $F444W$ until convergence (~ 50 epochs). The output of the classifier network is determined by a softmax layer, which provides an array of size four corresponding to a measurement of the probability that a given galaxy image belongs to one of the four classes described in Sect. 2. The stamp size is fixed to 32×32 pixels for both data sets (CEERS and CANDELS) which implies that the effective field of view is also different. We tried interpolating the CANDELS images as a preprocessing step to match the pixel scale of NIRCcam, but it led to worse results. We thus decided to keep the different field of views and let the network learn the corrections.

The classification is ultimately done by applying the trained feature extractor (G_f) and classifier (G_y) to the CEERS target stamps (x^t). To account for uncertainty, we employed an ensemble of ten separate trainings with different initial conditions and slightly different training sets. Specifically, for every run, we randomly initialize the neural network weights and shuffle the training set before splitting into test and training. To estimate the final probability of a given morphological class, we took the average of the outputs from all ten networks. The standard deviation was used to assess the robustness of the classification which we find to be typically below 0.1. Thus, unless otherwise noted, we defined classes as the maximum of the average probability from the ten classifications networks. Excluding objects with uncertain classifications does not significantly impact the main findings.

3.2. Visual inspection

There are very few ground truth labels available for the CEERS images by construction. It is therefore not straightforward to assess the accuracy of the resulting morphological classification

using the domain adaptation procedure described in the previous subsection. We followed two approaches to provide an overview of the quality of the resulting classification. First, we performed visual inspections of randomly selected galaxies belonging to the four classes. Figures 3–5 show some random examples of galaxies ordered by stellar mass and redshift in the $F200W$, $F356W$, and $F444W$ filters, classified in the four different classes⁴. The stamps clearly show distinct morphological features, indicating that the network produces meaningful classifications on JWST images, even without provided labels. The difference between the two main classes (spheroids and disks, shown in the top rows of Figs. 3–5) is obvious, demonstrating (not surprisingly) that the network is clearly able to distinguish between compact and extended sources. The boundary between disks and irregulars is slightly more diffuse, but this is known to be a difficult task. The figure seems to suggest that irregular galaxies might be biased towards images that contain multiple objects in the field of view. Some indeed appear to be perturbed light profiles, but others might simply be foreground or background contamination. It is also worth noticing the different spatial resolution of the different filters. Some compact galaxies appear to be unresolved in the redder wavebands.

3.3. Comparison with visual classifications

In a recent study, Kartaltepe et al. (2023) visually classified a subset of galaxies with $z > 3$ using a similar scheme to the one used for CANDELS. As an independent cross-check of the domain-adapted morphologies, we compare our deep learning-based classification to the visual one in Fig. 6. For this comparison, we use the neural network-based classification in the $F200W$ filter, since it is the primary band also used in Kartaltepe et al. (2023), although classifiers could also inspect other bands. Although this comparison is performed on a reduced data set of ~ 800 galaxies, it is crucial since it compares two independent classifications made on the same images. It is also the set of galaxies for which more discrepancies might

⁴ $F200W$, $F356W$, and $F444W$ are the filters primarily used in the scientific analysis of this work and thus $F150W$ images are not shown to keep a reasonably amount of figures. $F150W$ images are used for comparison with HST as discussed in Sect. 4 and in Appendix B.

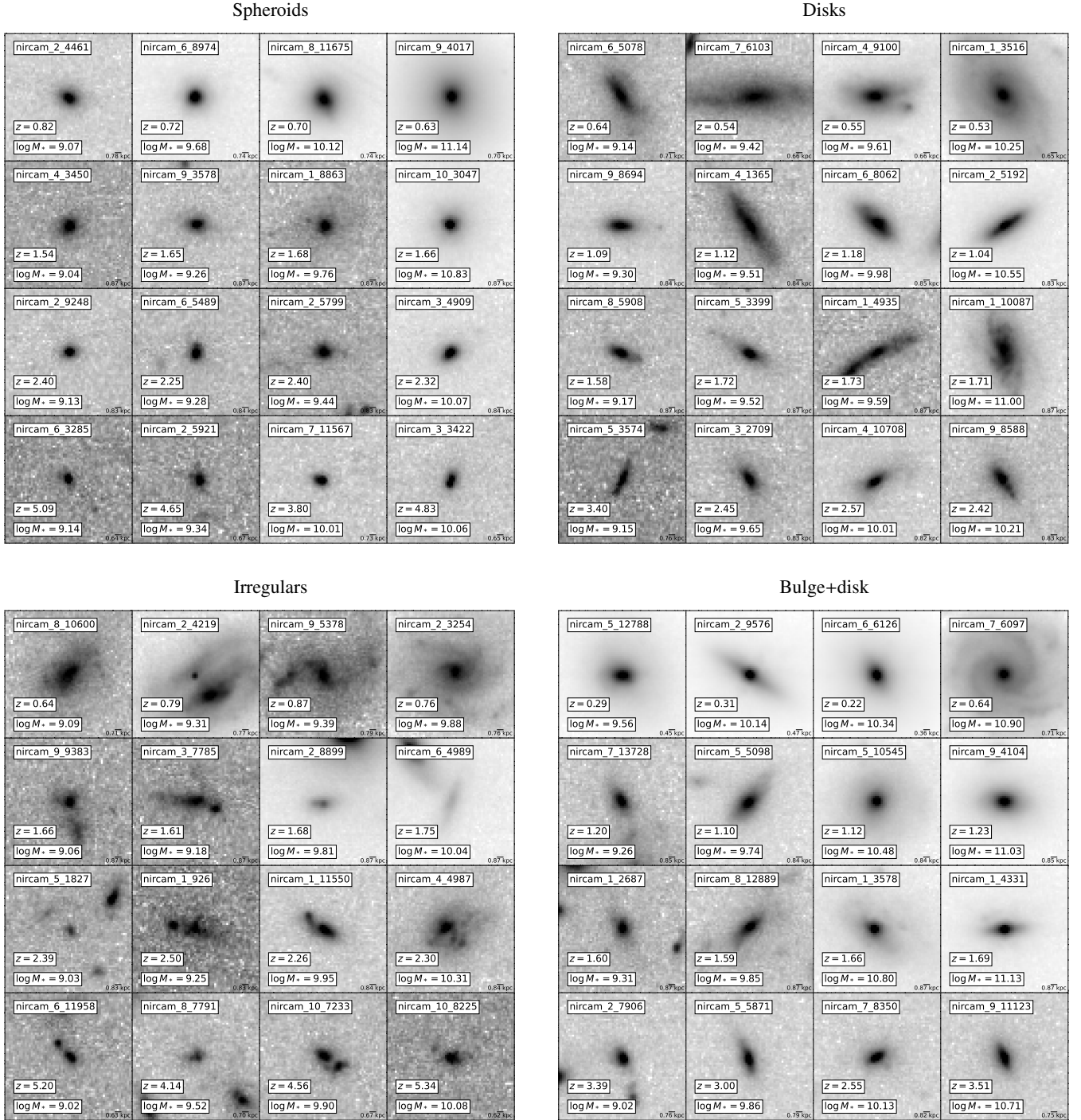


Fig. 3. Example of random stamps of CEERS galaxies observed with the $F200W$ filter, classified into four main morphological classes. Each panel of 16 images illustrates a different class. Top-left: spheroids. Top-right: disks. Bottom-left: irregulars. Bottom-right: composite bulge+disk galaxies. In each group of 16, galaxies are ordered by increasing photometric redshift (top to bottom) and stellar mass (left to right). The physical scale in kpc is shown for every galaxy. A square root scaling has been applied to enhance the outskirts.

be expected, as they represent the faint end of the distribution and cover a redshift range that is not well probed by HST.

Since the classification of Kartaltepe et al. (2023) is more detailed than ours, Fig. 6 shows, for each one of the four deep learning classes, the distribution of visual labels as defined in Kartaltepe et al. (2023). Overall, we observe good agreement between the two classifications. The distributions for the four primary classes are clearly different. More than 90% of galaxies classified as spheroids by the neural network are also

flagged as having a spheroid component visually. All galaxies automatically classified as disks have visually identified disks. Even the class of composite bulge+disk systems shows good agreement with visual classifications, with a clear peak at the Sph+Disk systems. As hinted by the visual inspection of the previous section, the largest discrepancies are between the irregular and disk classes. About 25% of galaxies classified as disks by the neural networks are flagged as irregulars by human classifiers, and the same happens in the other direction. Given the

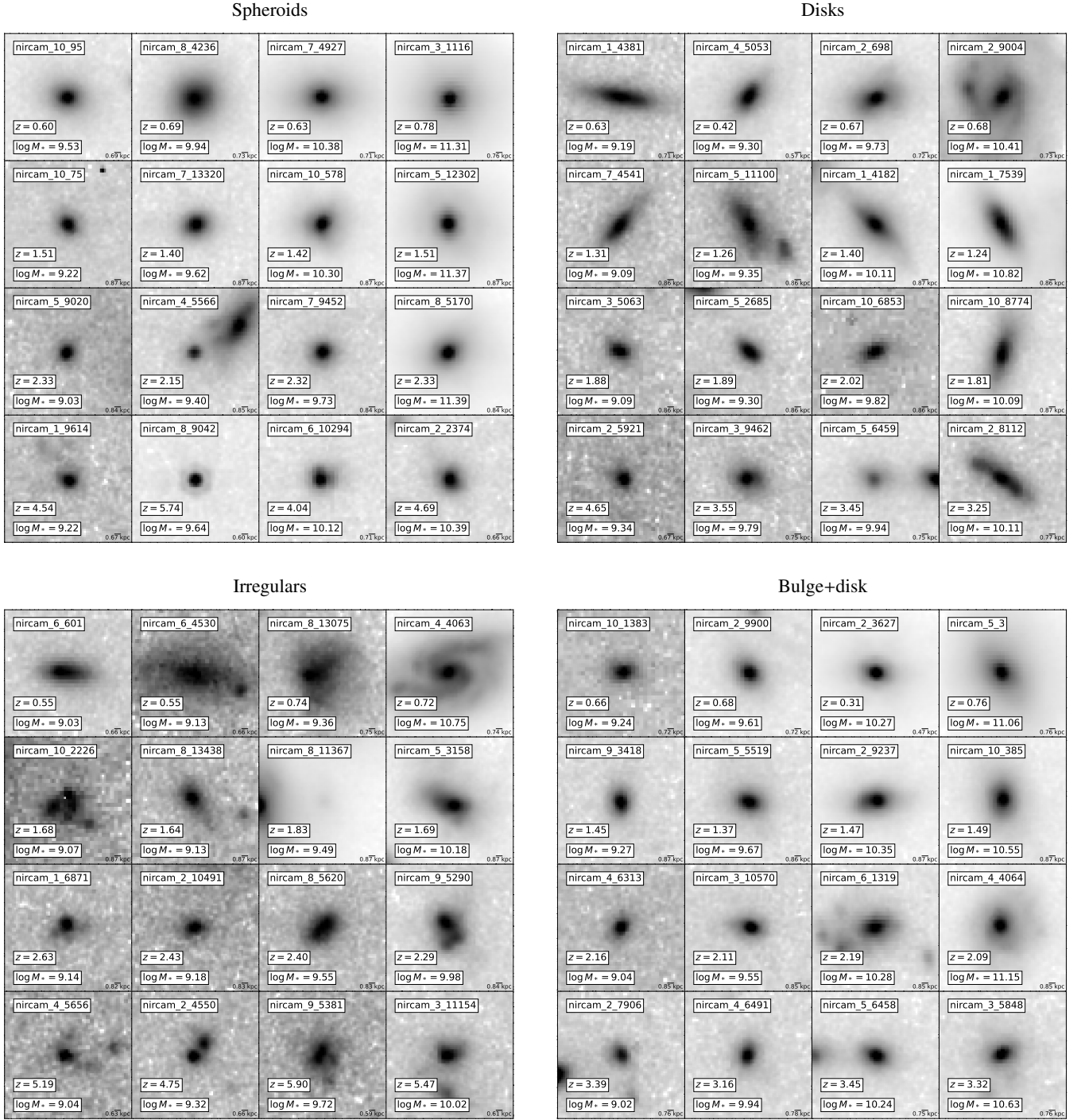


Fig. 4. Example of random stamps of CEERS galaxies observed with the $F356W$ filter, classified in four main morphological classes. Each panel of 16 images illustrates a different class. Top-left: spheroids. Top-right: disks. Bottom-left: irregulars. Bottom-right: composite bulge+disk galaxies. In each group of 16, galaxies are ordered by increasing photometric redshift (top to bottom) and stellar mass (left to right). The physical scale in kpc is shown for every galaxy. A square root scaling has been applied to enhance the outskirts.

difficulty (and somewhat subjective nature) of identifying irregular galaxies and the fact that the two classifications compared here are completely independent, we consider this contamination acceptable.

We emphasize that even if the comparison with the visual classifications from Kartaltepe et al. (2023) concerns only a fraction of our sample, it is the most different from our labeled training set from CANDELS. Therefore, the reported good agreement is a good indication that the proposed domain adap-

tation framework for training has successfully adapted to JWST images.

Nevertheless, in order to provide an additional quantification of the classification accuracy, we investigate in Fig. 7 the correlation between the morphological classes derived in this work and the Sersic index from independent parametric fitting with GALFIT (McGrath et al. in prep.). The figure shows the expected trends. Different morphological classes present different Sersic index distributions, with spheroids more skewed towards larger

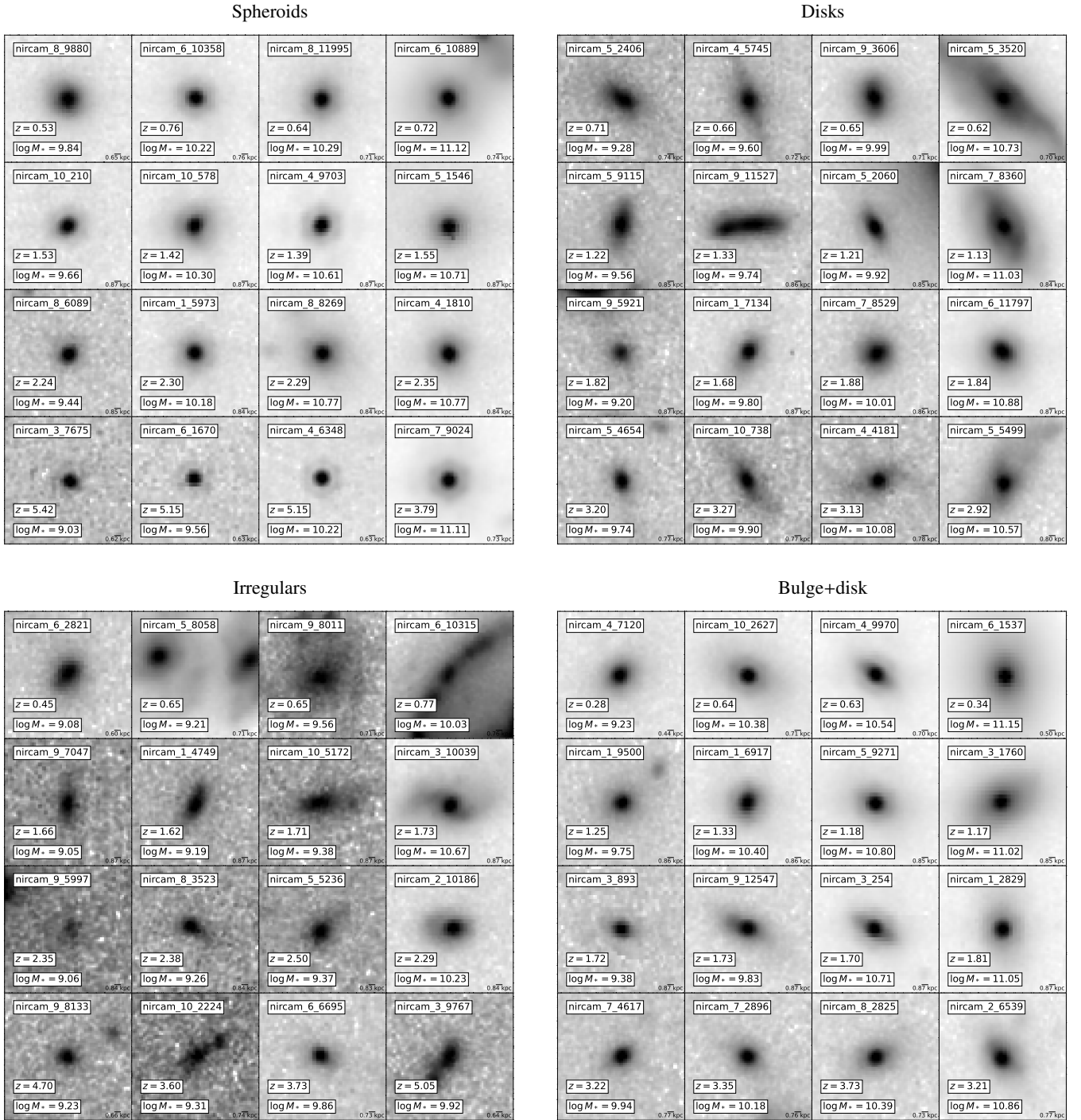


Fig. 5. Example of random stamps of CEERS galaxies observed with the $F444W$ filter, classified into four main morphological classes. Each panel of 16 images illustrates a different class. Top-left: spheroids. Top-right: disks. Bottom-left: irregulars. Bottom-right: composite bulge+disk galaxies. In each group of 16, galaxies are ordered by increasing photometric redshift (top to bottom) and stellar mass (left to right). The physical scale in kpc is shown for every galaxy. A square root scaling has been applied to enhance the outskirts.

values of the Sersic index ($n \sim 3$) and disks peaking close to $n \sim 1$ as expected for exponential profiles. The distribution for bulge+disks lies in between both disks and spheroids confirming that galaxies in this class tend to have a more prominent central concentration but are also more extended than pure spheroids. Irregulars have a very similar distribution to disks but the peak at $n \sim 1$ is less pronounced and contains a population with flatter luminosity profiles. This is also somehow expected for galaxies without a clearly defined center. These trends reinforce the fact

that the proposed morphological classification is robust in a redshift range weakly probed by our training set.

Given that we do not have available visual classifications for all our JWST sample, we cannot perform a direct comparison as for the [Kartaltepe et al. \(2023\)](#) sample for all our galaxies. Visually inspecting a representative sample is time consuming and is somehow at odds with the purpose of this work, which is to provide a large morphological classification with a limited amount of labels. However, we can use the comparison with parametric

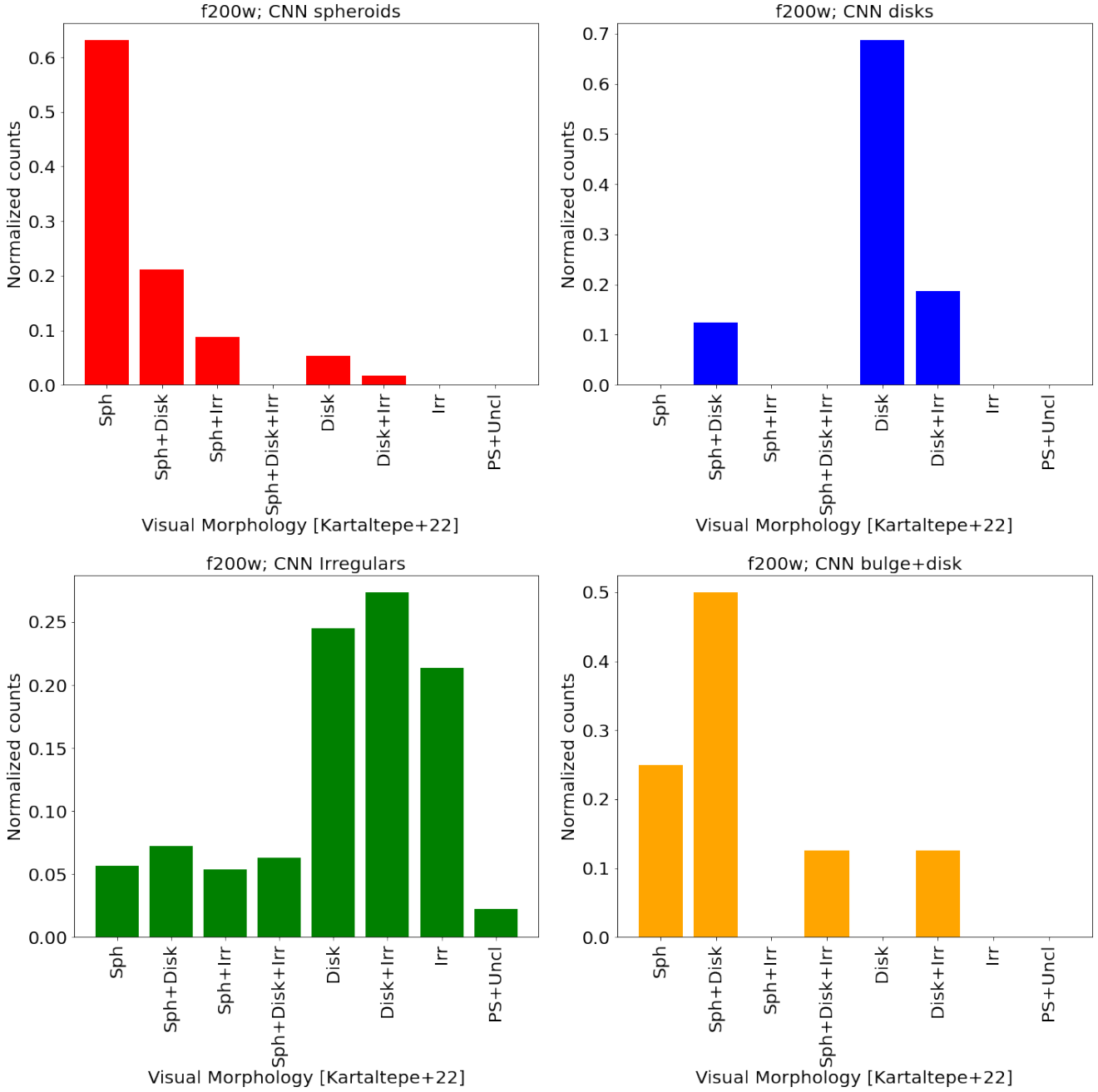


Fig. 6. Comparison between the deep learning based classification in the $F200W$ band presented in this work and the visual classifications of [Kartaltepe et al. \(2023\)](#) for galaxies with $z > 3$. The different panels show the distribution of visual classes for each of the four classes defined in this work. Top-left: spheroids. Top-right: disks. Bottom-left: irregulars. Bottom-right: disks+spheroids.

fits presented in Fig. 7 as a calibration for the robustness of the classification, where no visual labels exist. If the classification is robust we indeed expect a similar correlation with the Sérsic index in different redshifts and stellar mass bins. Figure 8 shows the mean values and normalized absolute deviations of the Sérsic index in different regions of the stellar mass-redshift plane. Overall, we measured similar trends as for the [Kartaltepe et al. \(2023\)](#) subsample. Spheroids and bulge+disks systems consistently present a mean Sérsic index larger than 2 and NMAD values of ~ 0.1 – 0.15 (except for some bins). Disks and irregulars present on average Sérsic indices close to ~ 1 with an homogeneous normalized mean absolute deviation. These measurements are consistent with the values measured in the [Kartaltepe et al.](#)

(2023) sample reported in Table 1, confirming the robustness of the classification over the entire parameter space.

4. From HST to JWST: impact of depth and spatial resolution

The JWST images provide deeper and higher spatial resolution images compared to the HST, which until recently was the primary telescope used to quantify galaxy morphologies. The quality of morphological classification is known to be significantly affected by factors such as resolution and S/N. Hence, it is interesting to compare the classifications of the same galaxies observed with HST and JWST. This is the main focus of this

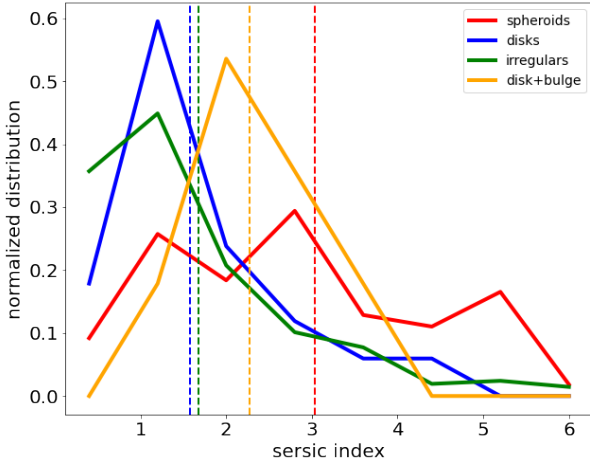


Fig. 7. Sersic index distributions for different morphological classes as labeled. The colored vertical dashed lines indicate the mean values of each distribution.

subsection, and to minimize the wavelength effect, we compared the *F160W* and *F150W* images.

We begin by exploring in Figs. 9 and 10 how basic morphological classifications on the same objects change between JWST and HST. For simplicity and to better understand the differences between the two classifications, we consider two cases: early versus late-type and disturbed versus undisturbed.

In the early-type class, we include both spheroids and bulge+disk galaxies as defined in Sect. 2. Huertas-Company et al. (2015) showed that the bulge+disk class is mainly composed of bulge-dominated systems. The late-type class contains both disks and irregulars. The disturbed class contains irregulars, while the undisturbed class gathers all the remaining three classes (see Sect. 2).

Since the primary differences between the two telescopes are sensitivity and spatial resolution, in addition to a global comparison, we quantify the differences as a function of apparent *F150W* magnitude as a proxy for the S/N and apparent half-light size as a proxy for the resolution.

4.1. Early versus late-type galaxies

The top panel of Fig. 9 shows that there is an overall good agreement between HST- and JWST-based early and late classifications. We find that $\sim 90\%$ of the galaxies between $z = 0-3$ have the same classifications with both instruments. The confusion matrix also shows that the agreement is not completely symmetrical, namely, $\sim 20\%$ of galaxies classified as early-type with HST move to the late-type class with JWST, in agreement with previous results suggesting that disks are more abundant than expected (Ferreira et al. 2022). In Fig. 10 we explore in more detail how the differences depend on apparent magnitude and apparent size. The discrepancy in the relative fraction of early type galaxies between JWST and HST ($\Delta_{\text{early}} = \frac{N_{\text{early}}^{\text{CEERS}} - N_{\text{early}}^{\text{CANDELS}}}{N_{\text{total}}}$) is consistently below $\sim 10\%$ across most of the parameter space. It is only for very small galaxies ($R_e < 0.1''$) that the relative fraction of early-type galaxies measured by JWST drops by about $\sim 30\%$ compared to HST. This increase in discrepancy for small and faint galaxies suggests that it is related to spatial resolution and (possibly) the S/N. We show in Appendix B some example stamps of galaxies with different morphological classifications in CANDELS and CEERS. Interestingly, because of the

distribution of galaxies in the parameter space (shown in the left panel of Fig. 10), these discrepancies only affect a small fraction of objects. Thus, the measured fractions of early and late-type objects as a function of redshift and stellar mass (shown in the top row of Fig. 11) are very similar and fully compatible within the 1σ uncertainties.

4.2. Peculiar galaxies

The differences between the fractions of disturbed and undisturbed galaxies are more significant. The top panel of Fig. 9 shows that the overall agreement is of $\sim 76\%$. This is mostly driven by the fact that $\sim 30\%$ of galaxies classified as irregular with HST are not found to be disturbed with JWST. The rightmost panel of Fig. 10 confirms a clear trend of classifying galaxies as less disturbed in JWST than in HST images. The differences in the relative fractions of disturbed systems ($\Delta_{\text{irr}} = \frac{N_{\text{irr}}^{\text{CEERS}} - N_{\text{irr}}^{\text{CANDELS}}}{N_{\text{total}}}$) range, on average, between $\sim 20\%$ and $\sim 40\%$ and tend to be more pronounced for large objects. This trend is likely due to the fact that the disturbed or irregular class is a poorly defined category of objects that includes all galaxies with some irregularity in their surface brightness profile. Thus, their irregular appearance is highly dependent on the S/N, since fluctuations in the surface brightness caused by noise can be easily interpreted as an irregular light distribution. Moreover, deeper observations will tend to better detect a potential diffuse component around galaxies which can be interpreted as a disk and move the classification from irregular to disk. Given that the CEERS survey's imaging is significantly deeper than the CANDELS data, it is expected that galaxies will appear less disturbed. This trend is confirmed by the rightmost panel of Fig. 10. Extended galaxies have lower S/N per pixel at a fixed magnitude, which may explain why the differences are larger for larger objects. We could argue as well that deeper observation may improve detections of low surface-brightness features around galaxies, which might make them look more irregular and therefore have the opposite effect. However, this is not what Fig. 10 suggests and the reason for this might be that the JWST observations, despite being deep, are not deep enough to detect these low surface brightness features, which are typically seen at depths of $\sim 30 \text{ mag} \cdot \text{arcsec}^{-2}$ or more (e.g., Trujillo & Fliri 2016). Some examples of images with different classifications are shown in Appendix B. Despite these differences, the impact on the measured fraction of disturbed galaxies as a function of stellar mass and redshift reported in Fig. 11 is not significant. The fraction of irregulars tends to be less than $\sim 10\%$ smaller in CEERS than in CANDELS. This can be explained by the fact that the majority of objects lie in a region of the parameter space, where the differences between HST and JWST are less pronounced. We also notice that the leftmost panel of Fig. 11 shows that irregular galaxies completely dominate the galaxy distribution at $z > 3$. This is most likely a consequence of using the *F200W* filter, which probes the UV rest-frame as discussed in the following. It might be also affected by selection biases since we are only representing objects in common in CANDELS and CEERS and that is also why it differs from the results of Kartaltepe et al. (2023).

5. Galaxy morphology since $z = 6$ in the rest-frame NIR

A unique feature of JWST is that it probes longer wavelengths than HST, offering a view of galaxy structure in the mid-infrared. This enables us to probe galaxy structure in the optical

F200W - Mean and NMAD of Sersic Index

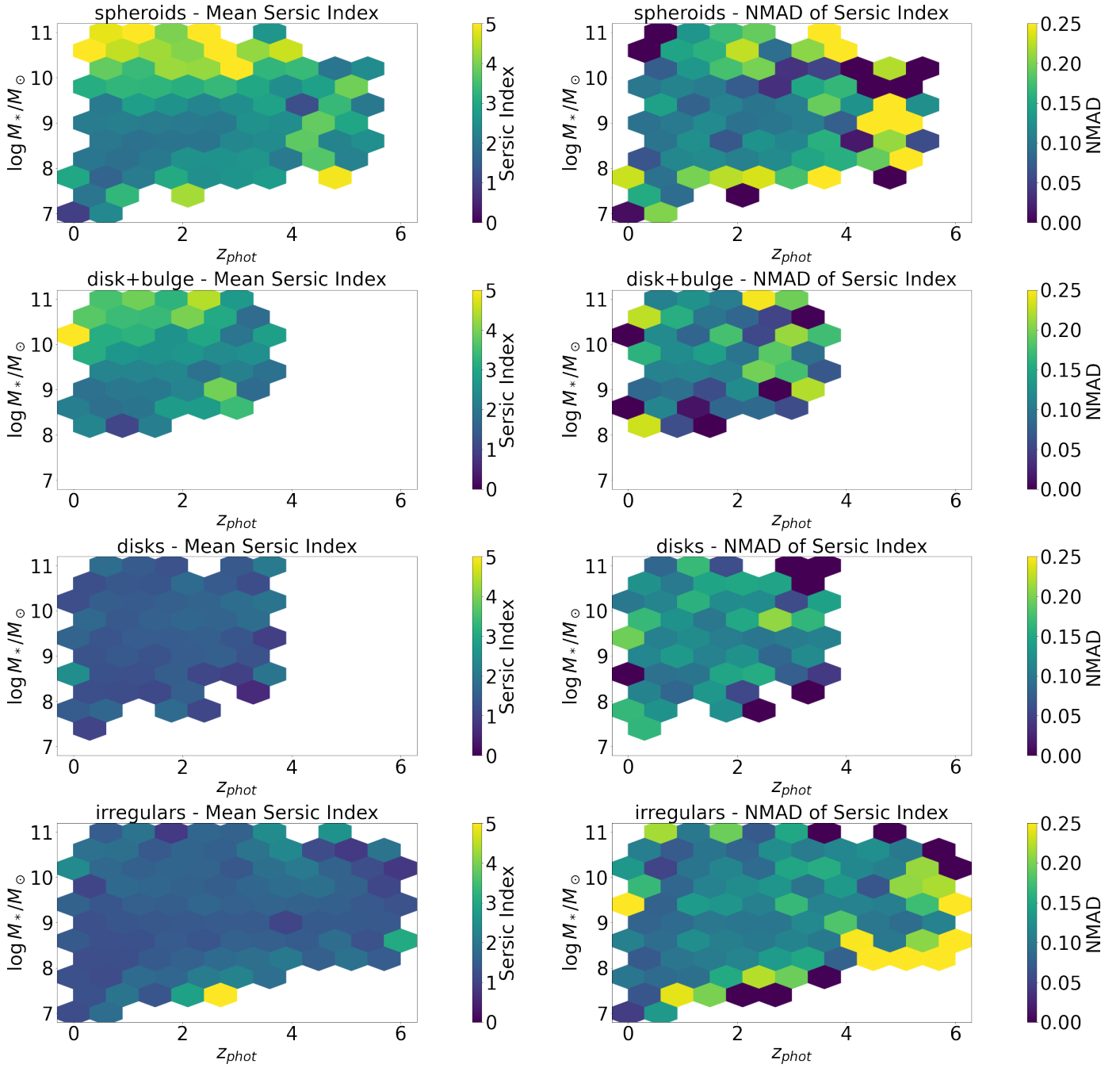


Fig. 8. Correlation between global morphological class and Sersic index. Mean values (left column) and normalized mean absolute deviations (right column) of the Sersic index for different morphological classes in different bins of stellar mass and redshift. From top to bottom: spheroids, disk+bulge, disks and irregulars. We only include objects with converged GALFIT measurements (flag=0) and only cells with more than ten galaxies are reported.

rest-frame up to higher redshifts ($z \sim 6$), but also in the NIR rest-frame with a proxy closer to stellar mass and less biased by stellar age and/or dust light-weighting effects.

5.1. Morphology versus Wavelength

In a similar fashion to the analysis of the previous section, we investigate how galaxy morphology changes with wavelength. Figure 12 compares the fractions of early-type and peculiar

galaxies between the $F200W$ and $F444W$ filters as a function of stellar mass and redshift.

Concerning the abundances of early-type galaxies, the figure shows that the differences between the fractions measured in the two filters become significant for galaxies above $z \sim 3$ and stellar masses larger than $\log M_* \sim 10.5$. In particular, we find that the fraction of early-type galaxies is 30% larger in the longer wavelengths. This result is expected since, at $z > 3$, the $F200W$ filter probes the near-UV and is therefore more sensitive to the

Table 1. Mean and normalized mean absolute deviations (NMAD) of the Sersic index for different morphological classes in the Kartaltepe et al. (2023) sample.

Morphology	Mean Sersic index	NMAD
Spheroids	2.93741	0.191377
Disk+bulge	2.4785	0.117469
Disks	1.59952	0.0926526
Irregulars	1.58476	0.116651

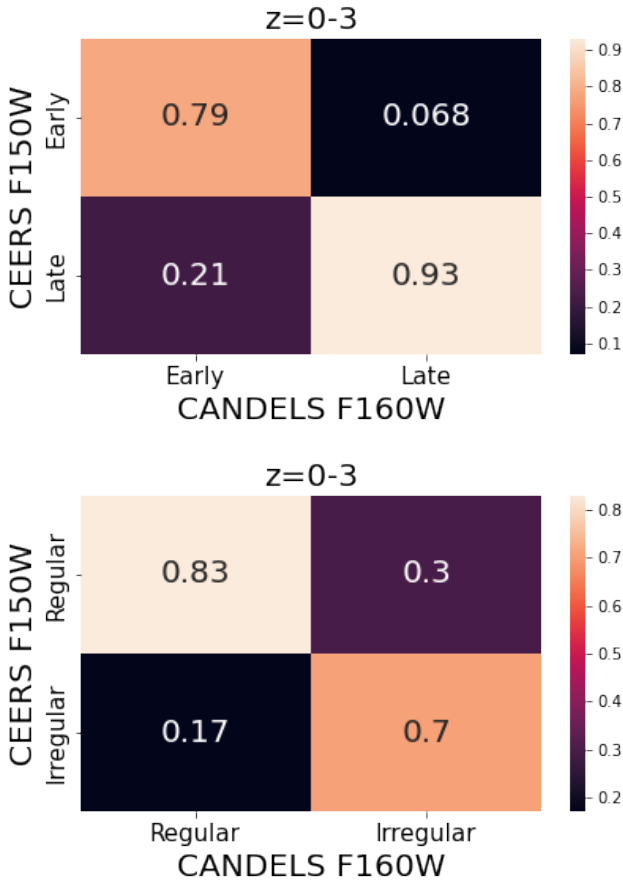


Fig. 9. Confusion matrices showing the overall agreement between early/late (top) and regular and irregular (bottom) classifications with HST-WFC3 $F160W$ imaging and JWST-NIRCam $F150W$ imaging for galaxies in the redshift range of $z = 0-3$.

emission of young stars, which can make the bulge difficult to detect (e.g., Papaderos et al. 2023). The $F444W$ filter, on the other hand, probes a rest-frame wavelength between ~ 1 and $\sim 0.8 \mu\text{m}$ above $z \sim 3$, which is dominated by old stars typically located in the central parts of galaxies (i.e., bulge). For stellar masses lower than 10^{10} solar masses, the differences between the two filters are slightly less pronounced, and the trend is inverted. Specifically, the fraction of early-type galaxies is larger in the bluer filters. This counterintuitive trend may reflect the presence of compact star-forming regions interpreted as bulge components in the $F200W$ filter. It may also be due to a resolution effect, which prevents the disk component to be detected in the longer wavelengths. We discuss this further in Sect. 5. The top row of Fig. 13 explores how these reported differences between the two filters translate into the evolution of early and late-type fractions as a function of stel-

lar mass and redshift. We observe the expected behavior from the trends described in Fig. 12. The fraction of early-type galaxies in the highest redshift bin slightly increases at the high-mass end in the $F444W$ filter, while the opposite is measured at the low-mass end.

The differences between the fractions of disturbed and undisturbed galaxies are more pronounced. At $z > 3$ the fraction of irregular galaxies in the $F444W$ filter can be up to $\sim 50\%$ smaller than in the $F200W$. This might be a signature that the distribution of mass is less irregular than the one of light, as pointed out by previous works who attempted to estimate stellar mass maps from HST imaging (e.g., Wuyts et al. 2012). It is thus interesting to see that this trend is confirmed when probing galaxy morphology at longer wavelengths. It is worth noticing however that a similar trend could be driven by a difference in spatial resolution. The long wavelength imaging has a factor of ~ 2 lower resolution than the short wavelength (see Figs. 3 and 5) so it is less sensitive to substructure, which could make galaxies appear more regular in their light distribution. However, in Sect. 4 we report that the fraction of peculiar galaxies decreases in JWST imaging as compared to HST, even with increased spatial resolution. The trend measured with wavelength hence suggests that it is not purely a resolution effect. The bottom panel of Fig. 13 measures how these differences between the filters is translated into the fraction of peculiar galaxies as a function of stellar mass and redshift. The differences are particularly dramatic at $z > 3$. The fraction of peculiar galaxies completely dominates in the $F200W$ band at all stellar masses. In the $F444W$ filter, however, they only dominate at the low-mass end.

5.2. Morphological evolution

We examine how the fractions of different morphological types depend on stellar mass and redshift using the new JWST-based morphologies derived in this work (see Figs. 14 and 15). To define the morphologies, we use the $F200W$ filter at $z < 1$, $F356W$ at $1 < z < 3$, and $F444W$ at $3 < z < 6$, approximately in the rest-frame $\sim 0.8-1 \mu\text{m}$ (Fig. C.1). This should be mostly sensitive to the emission of old stars and therefore more closely linked to stellar mass. The price to pay (as previously mentioned) is that the spatial resolution is worse in the redder filters potentially biasing the measured evolution of morphologies. Nevertheless, as shown in Fig. C.1, the resolution remains reasonably constant around ~ 0.5 kpc from $z \sim 1$ to $z \sim 6$ and only goes down to ~ 0.2 kpc at $z < 1$. Throughout this section, the uncertainties in the figures reflect both the epistemic uncertainties on the morphological classifications as well as the statistical uncertainties due to limited number counts. For estimating the error caused by classification uncertainties we randomly sample 100 times the classification probabilities for every galaxy using the mean and the standard deviation of the ten ensembles. The statistical uncertainty computed using the 16 and 85 quantiles of the beta distribution, as described in Cameron (2011), is a robust estimator for uncertainties of fractional quantities.

We observed some well-known trends. The fraction of bulge-dominated galaxies (early-type) shows a strong correlation with stellar mass, with the number densities of early-type galaxies steadily increasing above $\sim 10^{10.3}$ solar masses. The behavior is surprisingly similar at all redshifts probed, suggesting similar physical processes for bulge formation at all epochs. Also, the fact that the trends change smoothly with redshift is an indication that there are no significant biases induced by the use of different filters with different spatial resolutions. The main difference is a change in normalization, with early-type galaxies

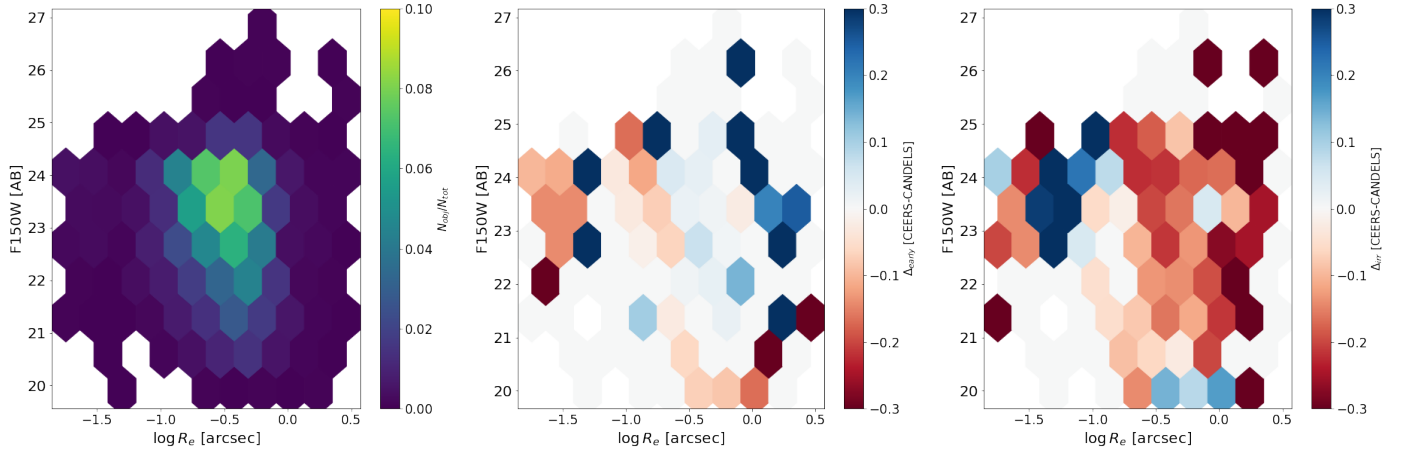


Fig. 10. Differences between CANDELS and CEERS measured morphologies as a function of apparent $F150W$ magnitude and angular half-light radius. Left panel: number density of objects. Middle panel: difference between the fraction of early-type galaxies in CEERS and CANDELS: $\Delta_{\text{early}} = \frac{N_{\text{early}}^{\text{CEERS}} - N_{\text{early}}^{\text{CANDELS}}}{N_{\text{total}}}$. Right panel: difference between the fraction of disturbed galaxies in CEERS and CANDELS: $\Delta_{\text{irr}} = \frac{N_{\text{irr}}^{\text{CEERS}} - N_{\text{irr}}^{\text{CANDELS}}}{N_{\text{total}}}$.

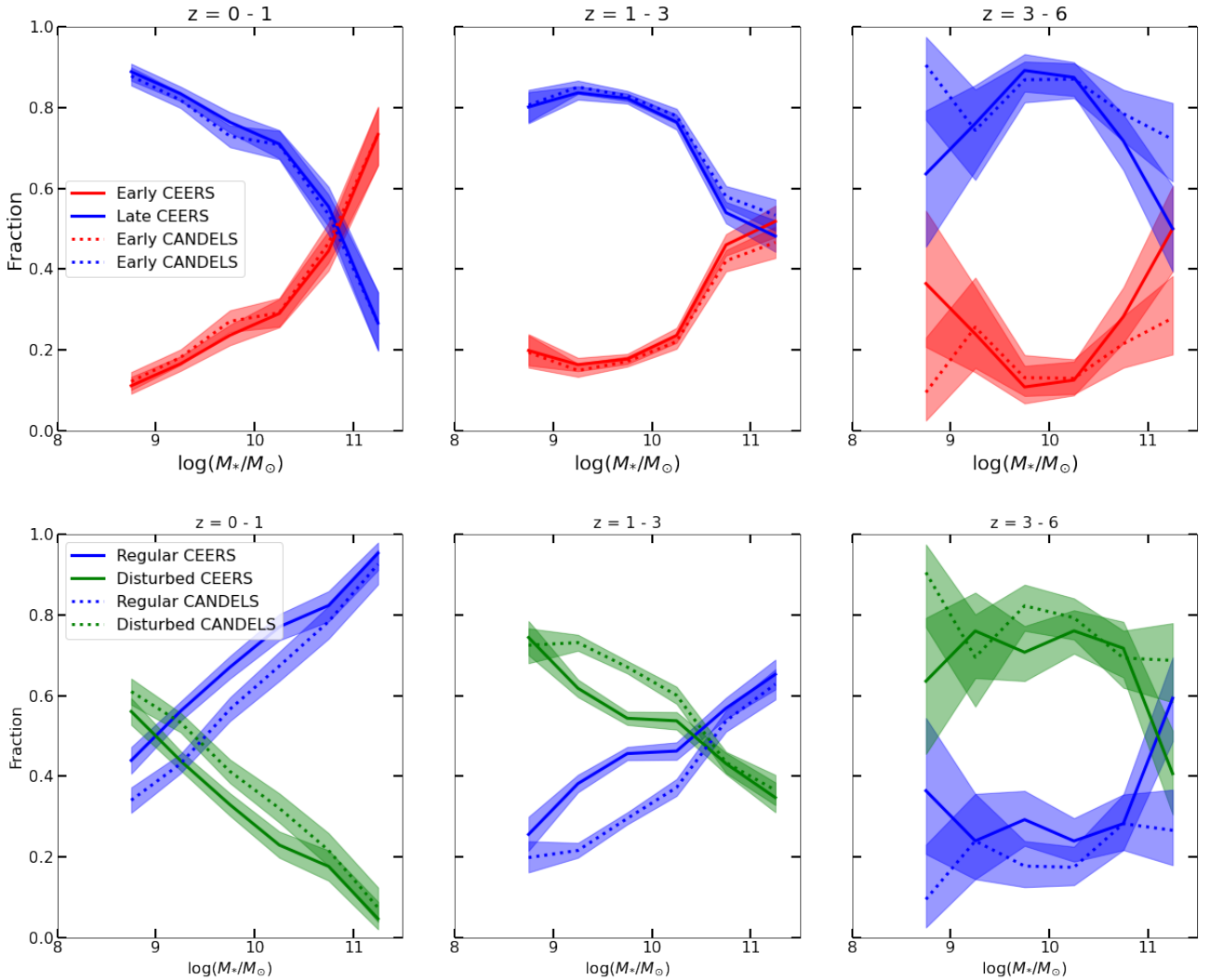


Fig. 11. Comparison between the fractions of early and late-type galaxies (top row) and disturbed and undisturbed galaxies (bottom row) as a function of stellar mass measured in CANDELS (dashed lines) and CEERS (solid lines) for exactly the same galaxies. For CANDELS, the morphologies are inferred in the $F160W$ filter while for CEERS we use $F150W$ in all panels to reduce wavelength induced differences in the morphologies. Each panel indicates a redshift bin as labeled.

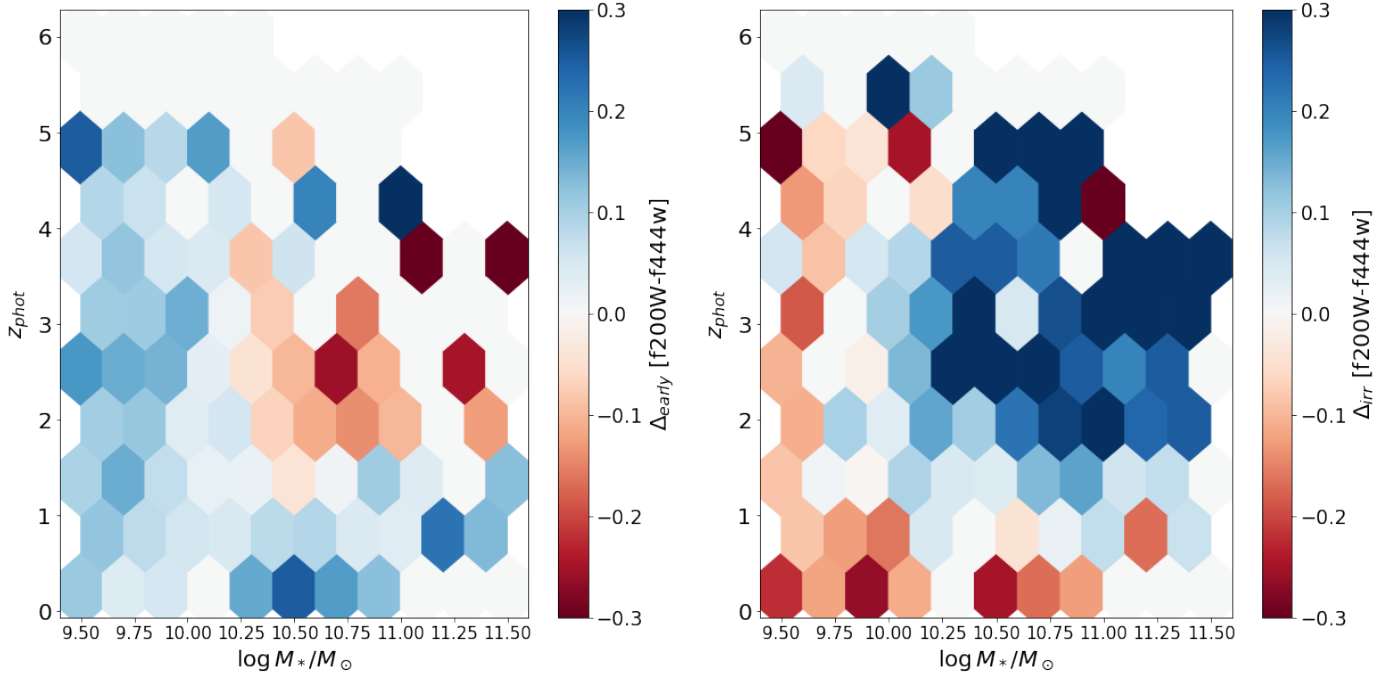


Fig. 12. Differences between morphologies measured in $F200W$ and $F444W$ in CEERS. The left panel shows the difference in the fractions of early-type galaxies ($\Delta_{\text{early}} = \frac{N_{\text{early}}^{F200W} - N_{\text{early}}^{F444W}}{N_{\text{total}}}$) and the right panel differences in the fractions of disturbed galaxies ($\Delta_{\text{irr}} = \frac{N_{\text{irr}}^{F200W} - N_{\text{irr}}^{F444W}}{N_{\text{total}}}$).

becoming more abundant at fixed stellar mass at later times. At $\sim 10^{11}$ solar masses, $\sim 40\text{--}50\%$ of galaxies are bulge-dominated at $z > 3$, while the fraction increases to $\sim 70\%$ at $z < 1$ for the same stellar mass. In fact, Fig. 15 shows that early-type galaxies start dominating the massive end of the galaxy population from $z \sim 3$. These results qualitatively agree with the recent findings by Costantin et al. (2021), who reported a bimodality in bulge formation, with a first wave of bulges ($\sim 30\%$) already in place at $z \sim 6$. Ferreira et al. (2023) tend to find a smaller fraction of bulge dominated galaxies at all redshifts using a similar data set but based on pure visual classifications. It might be because they did not plot the most massive galaxies separately as we do here. The lower mass bins in Fig. 15 do indeed show that late-type galaxies dominate at all redshifts.

Regarding the fraction of peculiar galaxies, we also measure very similar trends at all redshifts. The abundance of irregular galaxies is a strong function of stellar mass at all redshifts, with low-mass galaxies being predominantly peculiar. The stellar mass threshold below which the galaxy population starts to be dominated by irregular galaxies decreases with time. Galaxies less massive than $\sim 10^{10.5}$ solar masses are irregular at $z > 3$, while this is true only for galaxies less massive than $\sim 10^9$ at $z < 1$. It is interesting to see that the abundance of irregular galaxies is still measured to increase with redshift even when probing the rest-frame NIR. This suggests that, at early epochs, the distribution of stellar mass is also perturbed and that it is not only a consequence of the presence of bright star-forming regions emitting in the UV. However, as discussed in Sect. 4, the classification of a galaxy as irregular or peculiar does not only depend on wavelength but is also rather noise-sensitive. The result might therefore be biased because galaxies at fixed stellar mass appear fainter at high redshift. It is also worth noticing, though, that the impact of the S/N on the reported fractions is generally small (see Fig. 11) because it essentially affects large galaxies that represent a small fraction of the galaxy population. Our results qual-

itatively agree with the trends reported by Ferreira et al. (2023) in the sense that “regular” galaxies tend to dominate at lower redshifts, but they did not find such clear trends. Instead, they have found that low-mass galaxies ($\log M_*/M_\odot < 9$) are predominantly disks, while our classification tends to classify the majority of low mass galaxies as irregulars at all redshifts (bottom-left panel of Fig. 15). The differences might be a reflection of the somehow loose definition of the “peculiar class.” A one-to-one comparison between the two classifications would provide more insights on the origins of the differences.

5.3. Morphology-quenching relation

In this section, we investigate the evolution of the relationship between galaxy morphology and star-formation activity, using the new NIR rest-frame morphologies based on JWST, from $z \sim 6$.

5.3.1. Definition of quenched and star-forming samples

To begin, we define quiescent and star-forming galaxies, following the definition of Tacchella et al. (2022), which is based on the timescale for doubling the stellar mass, given the instantaneous star formation rate at the time of observation. Tacchella et al. (2022) measured this timescale by defining the mass-doubling number $D(z) = \text{sSFR}(z) \times t_H(z)$, where $\text{sSFR}(z)$ and $t_H(z)$ represent the specific star-formation rate and the Hubble time at the time of observation, respectively. Therefore, $D(z)$ measures the number of times the stellar mass doubles within the age of the Universe, at a constant sSFR. To account for uncertainties in both the measurements of star formation rates and stellar masses, we sample 100 times from the posterior distributions of both quantities estimated by DENSE BASIS. For simplicity, we assume a Gaussian posterior distribution with standard deviation equal to half the difference between the 84th and 16th quantiles, as

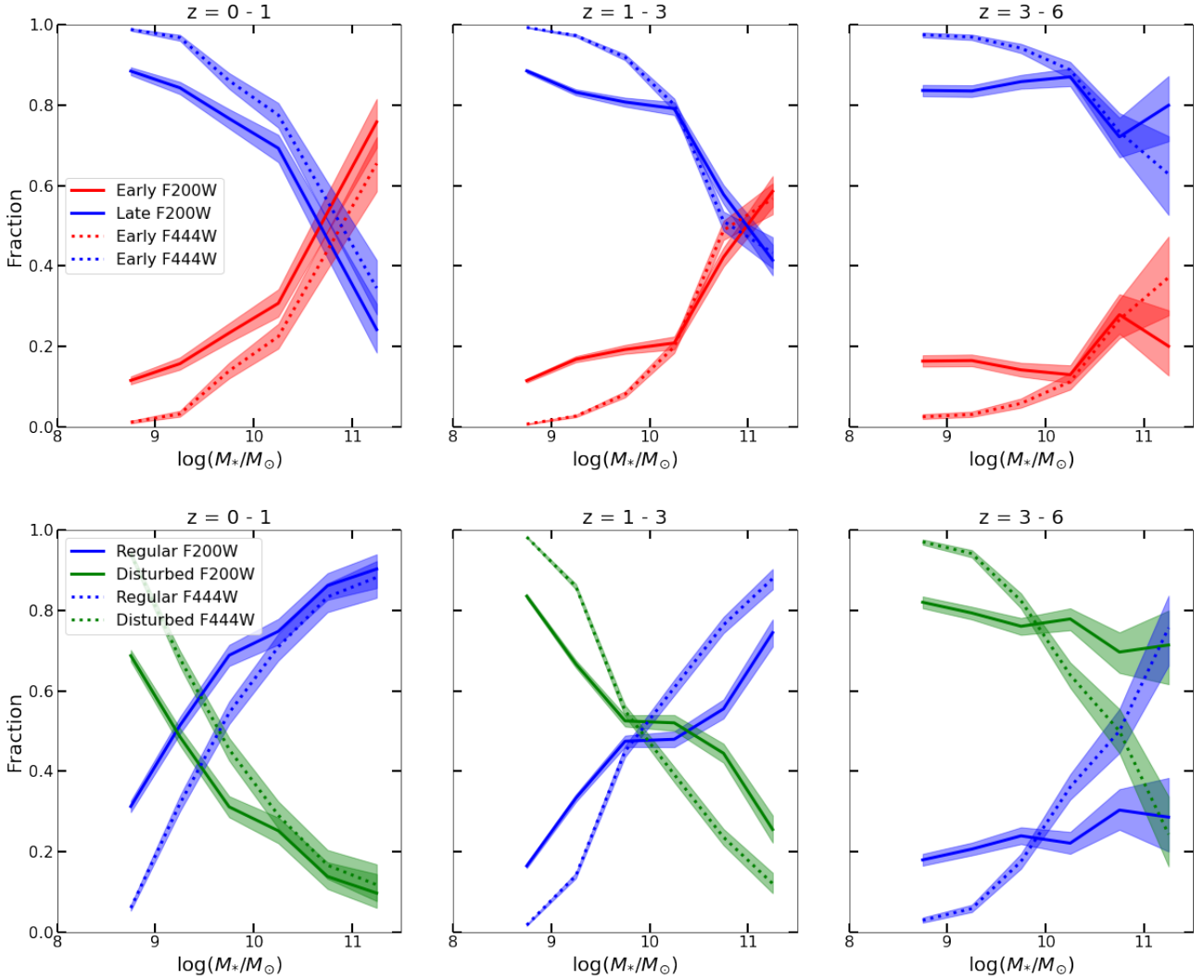


Fig. 13. Comparison of the morphological fractions as function of redshift and stellar mass measured in the *F200W* and *F444W* filters. The top row show the fractions of early (red lines) and late (blue) type galaxies in three redshift bins as labeled in the *F200W* (solid lines) and *F444W* (dashed lines) bands. The bottom row, indicates the fraction of regular (blue) and peculiar (green) galaxies.

estimated by DENSE BASIS (see Sect. 2 for more details). The top row of Fig. 16 displays the contours of the distribution of the 100 samples in the $\log M_* - \log \text{SFR}$ plane, for different redshift bins. The SFR is averaged over a timescale of 100 Myr, where star-forming galaxies are defined as those with $D(z) > 1/3$, transitioning galaxies as those with $1/20 < D(z) < 1/3$, and quiescent galaxies as those with $D(z) < 1/20$, following Tacchella et al. (2022).

Next, we fit the distribution of star-forming galaxies ($D(z) > 1/3$) using a power law with two parameters α and β :

$$\log \text{SFR} = \alpha \times (\log M_*/M_\odot - 10.5) + \beta.$$

We estimated the posterior distributions $p(\alpha, \beta | \{\log M_*, \log \text{SFR}\})$ into three redshift bins of $0 < z < 1$, $1 < z < 3$, and $3 < z < 6$, using an amortized likelihood-free inference approach with a masked autoregressive neural flow. We simulated 100 000 samples, using a flat prior for both α and β , and using a Gaussian distribution for both $\log \text{SFR}$ and $\log M_*$. We then trained a neural flow on the simulations to estimate p . We trained three different flows for the three different redshift bins. To account for systematic uncertainties due to the neural network architecture, we trained five additional random variations

of the masked autoregressive flow in each redshift bin. We find that all models provide consistent results and choose one of them as the primary estimator. The mean and 16th and 84th quantiles of α and β are reported in Table 2.

We observed an increase in the zeropoint (β) as we move to higher redshifts, consistent with previous studies (e.g., Whitaker et al. 2012a). Furthermore, we noted a slight decrease in the slope of the star-forming main sequence in the highest redshift bin, which is also within the ballpark of the range of different values reported in the literature (see Mérida et al. 2023 for a recent compilation). Figure 16 shows that this decrease in slope could be driven by a larger scatter in star formation at fixed stellar mass and the lack of massive objects at these redshifts. Further investigations are necessary to fully understand this trend. However, our primary goal is to obtain a reasonable guess of the location of the main sequence, rather than to analyze the slope and normalization in detail, which is beyond the scope of this work.

The middle and bottom rows of Fig. 16 display the distribution of late-type and early-type galaxies in the $\log M_* - \log \text{SFR}$ plane. In line with the expectations, the vast majority of disk-dominated galaxies are located within 1 dex of the star-forming

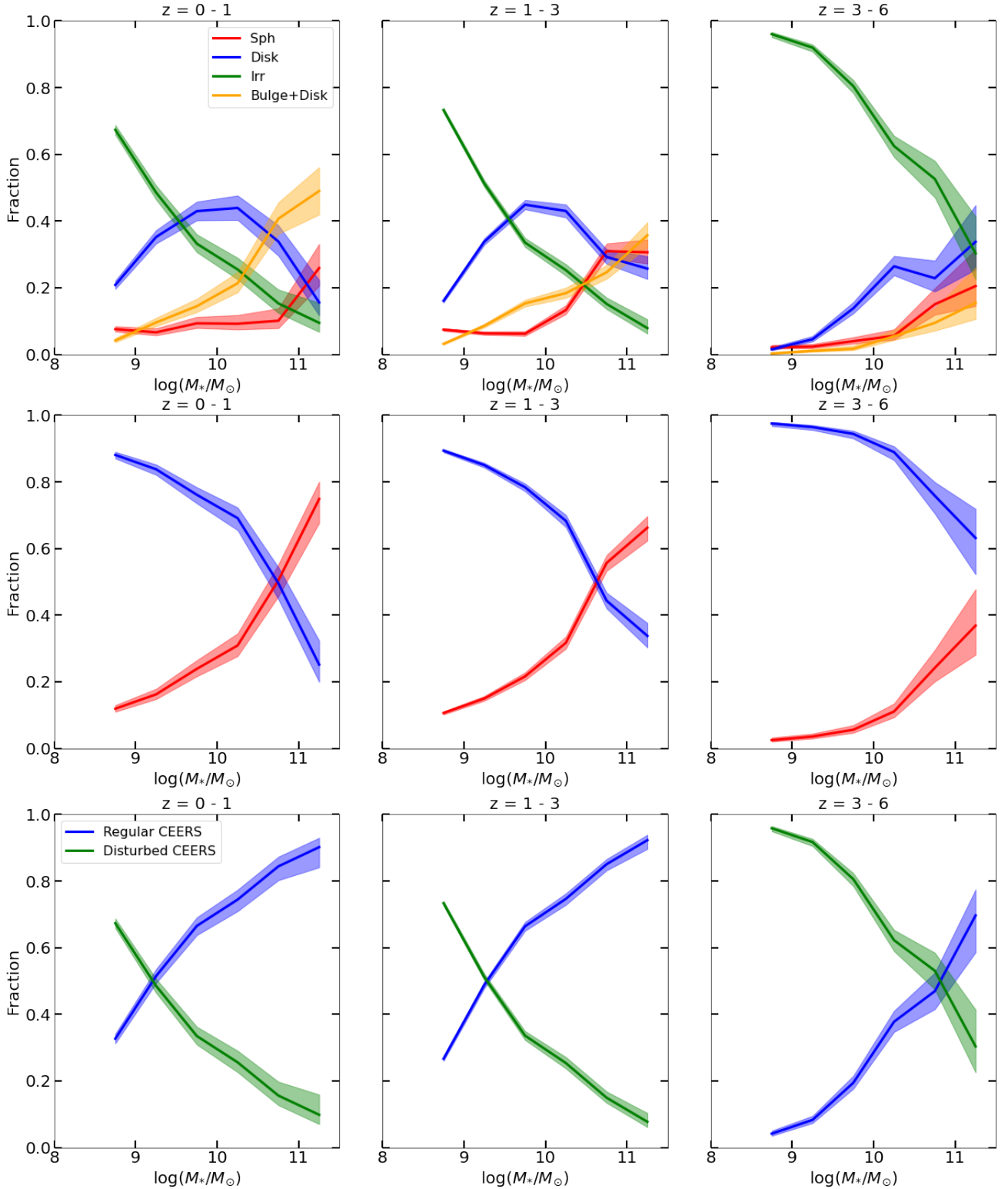


Fig. 14. Evolution of the fractions of different morphological types in rest-frame $\sim 0.8\text{--}1\ \mu\text{m}$ as a function of stellar mass and redshift. Each panel shows a redshift bin as labeled. Filters $F200W$, $F356W$ and $F444W$ are used to infer galaxy morphology in the redshift bins $0 < z < 1$, $1 < z < 3$ and $3 < z < 6$, respectively. Top row: fractions in four morphological classes: spheroids (red), disks (blue), bulge+disk (orange) and peculiar or irregular (green). Middle row: fractions in two broad classes: disk dominated (blue) and bulge dominated (red). Bottom row: fractions in two broad classes: regular (blue) and disturbed (green). The fractions of early and regular galaxies steadily decrease with redshift at all stellar masses. However, we still observe a significant fraction of massive bulge dominated and dusky galaxies up to $z \sim 6$. As in the previous figure, we observe a regular decline of regular and early-type galaxies with increasing redshift with massive galaxies presenting a more evolved morphology.

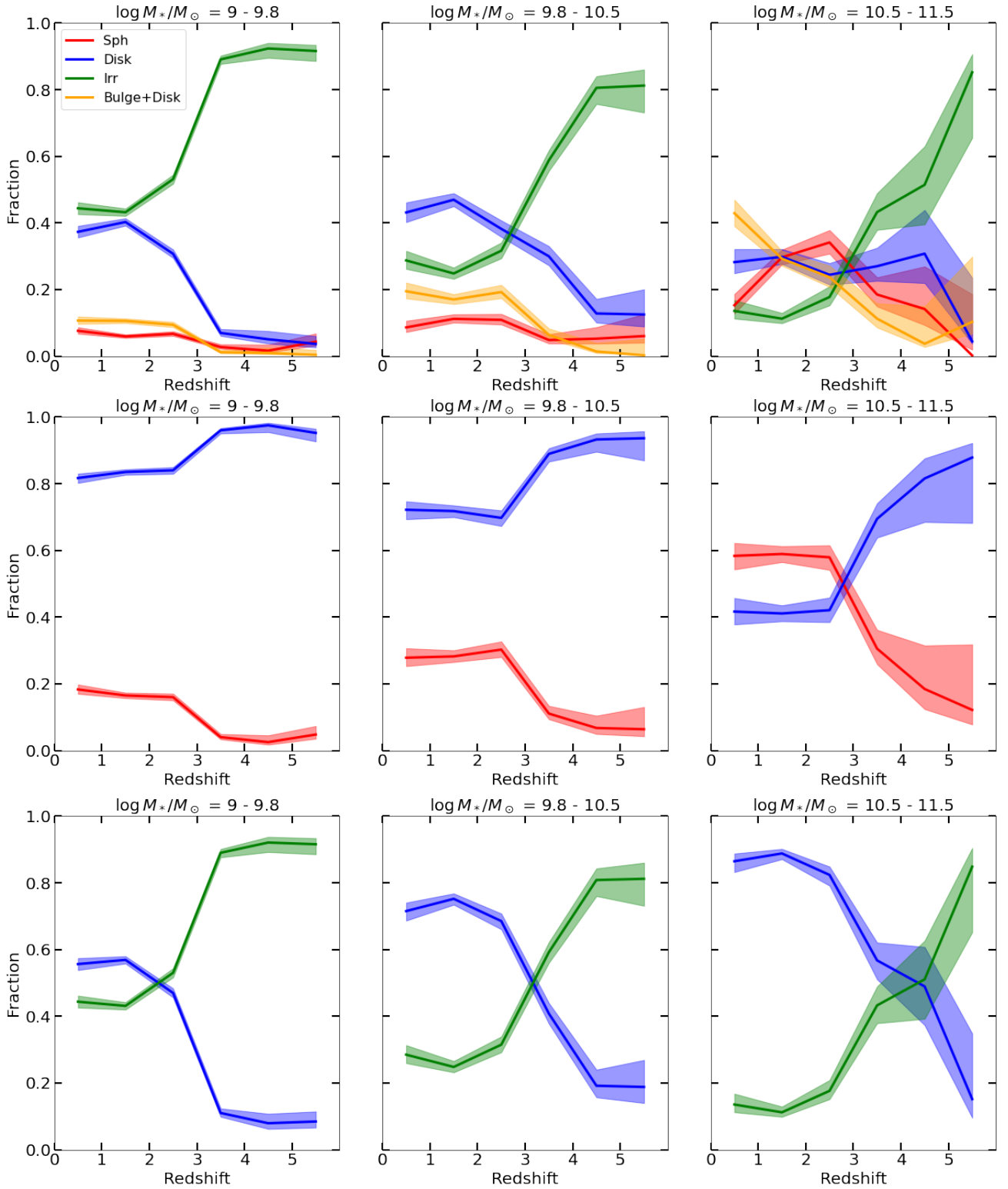


Fig. 15. Evolution of the fractions of different morphological types in rest-frame $\sim 0.8\text{--}1\ \mu\text{m}$ as a function of stellar mass and redshift. Filters $F200W$, $F356W$, and $F444W$ are used to infer galaxy morphology in the redshift bins $0 < z < 1$, $1 < z < 3$ and $3 < z < 6$, respectively. Each panel shows a stellar mass bin as labeled. Top row: fractions in four morphological classes: spheroids (red), disks (blue), bulge+disk (orange), and peculiar or irregular (green). Middle row: fractions in two broad classes: disk dominated (blue) and bulge dominated (red). Bottom row: fractions in two broad classes: regular (blue) and disturbed (green).

main sequence. In contrast, early-type galaxies are located in two distinct regions. At the high-mass end, they tend to lie below the main sequence in the region where quenched galaxies are

found, confirming that massive quenched galaxies tend to be bulge-dominated. Interestingly, this trend seems to hold even in the highest redshift bin. Additionally, the bottom row of Fig. 16

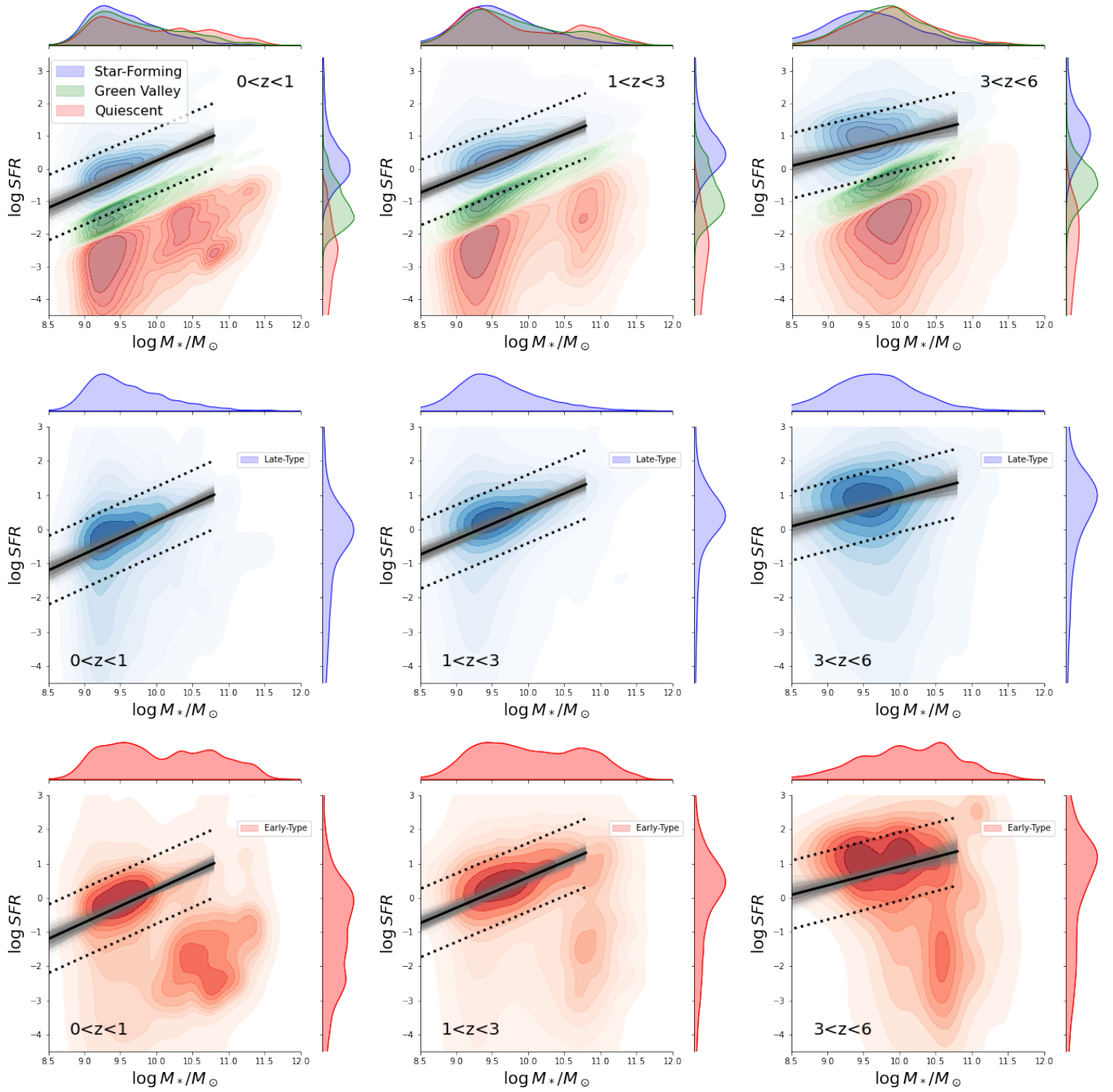


Fig. 16. Stellar mass-and-star formation rate ($M_{\odot} \cdot \text{yr}^{-1}$) plane in different redshift bins as labeled. The top row shows the whole sample analyzed in this work divided in star-forming (blue contours), green valley (green contours), and quiescent galaxies (red contours) based on the mass doubling time (see text for details). The middle row shows the distribution of late-type galaxies and the bottom row the one for early-type galaxies. In all panels, the solid line shows the mean of the posterior power-law fit to star-forming galaxies and the gray shaded regions indicates random samples of the posterior distribution. The dotted lines indicate the location ± 1 dex around the main sequence.

reveals a significant population of low-mass ($\log M_{*}/M_{\odot} < 10$) star-forming bulge-dominated galaxies located at all redshifts, which we investigate further in the following subsections.

5.3.2. Fraction of early-type galaxies

Using the power-law fits discussed in the previous subsection, we computed the fraction of early-type galaxies as a function of the distance to the main sequence, $\Delta \log \text{SFR}$, in different stellar

mass bins, and present the results in Fig. 17. Here, $\Delta \log \text{SFR}$ is calculated as the difference between a galaxy's SFR and the location of the main sequence corresponding to its stellar mass. We find that the fraction of bulge-dominated systems among massive galaxies ($\log M_{*}/M_{\odot} > 10.5$) increases as we move below the main sequence at all redshifts. Despite the large uncertainties due to low statistics, we observe that ~ 70 – 90% of massive galaxies 2–3 dex below the main sequence are early-type. These results suggest that the morphology-quenching relation

Table 2. Slope (α) and normalization (β) of the power-law fit to the star-forming main sequence in different redshift bins.

	$0 < z < 1$	$1 < z < 3$	$3 < z < 6$
α	$0.96^{+0.11}_{-0.12}$	$0.89^{+0.07}_{-0.07}$	$0.55^{+0.16}_{-0.15}$
β	$0.72^{+0.09}_{-0.08}$	$1.04^{+0.09}_{-0.10}$	$1.19^{+0.15}_{-0.15}$

is already in place for the most massive galaxies at $z \sim 5$. At intermediate stellar masses, we also observe a moderate increase in the early-type fraction below the main sequence, at least up to $z \sim 3$. However, at $z > 3$, the fraction remains consistently low ($<10\%$) regardless of star formation activity, indicating that morphological transformations have not yet occurred in significant numbers by $z \sim 5$. We do not observe any dependence of the early-type fraction on star formation activity for low-mass galaxies within the probed redshift range.

5.3.3. Star-forming early-type galaxies

We analyzed the distribution of the star-formation rate for each fixed morphological type. Although the fraction of low-mass early-type galaxies is small at low and intermediate masses, Fig. 16 shows that the majority of those are star-forming, in contrast to what happens at the high-mass end.

To investigate further the nature of these objects and check whether this is a consequence of classification errors, we show in Fig. 18 example stamps of late-type and early-type star-forming galaxies ($D(z) > 0.33$) with similar stellar mass. We only show examples of $z < 1$ galaxies to better appreciate the morphological differences. The figure clearly shows distinct morphologies for both populations. As expected, late-type galaxies are more extended and present more structure than early-type galaxies, whose light profile is smoother and more compact. An inspection of the half-light radii and Sersic indices distributions of both populations also reveals that early-type galaxies are, on average, more compact and have higher Sersic indices. Both results confirm the distinct morphological nature of these star-forming galaxies. We speculate that the population of low-mass early-type galaxies might be experiencing a phase of central star-formation activity (blue nugget) as reported in several previous works (e.g., Lapiner et al. 2023; Huertas-Company et al. 2018; Barro et al. 2016, and references therein). However, an analysis of the non-parametric SFHs inferred by DENSE BASIS does not show any clear differences between the two populations given the large uncertainties at the current S/N. Therefore we cannot firmly conclude that they have clear different formation pathways. Future deep observations targeting these high-redshift galaxies might enable us to break these degeneracies and better constrain their SFHs.

6. Discussion

6.1. The differences between HST and JWST-based galaxy morphologies

Galaxy morphology is a powerful proxy for physical processes in galaxies, but it is also well known to be significantly affected by multiple observational effects such as the S/N, spatial resolution, and cosmological dimming. Most results published in the past decades have been based on HST data, which, despite being a powerful space-based telescope, has been

pushed to its limits for analyzing galaxy structures beyond the local Universe. In particular, many works based on HST data (e.g., Buitrago et al. 2013; Lang et al. 2014; van der Wel et al. 2014a; Huertas-Company et al. 2016) have shown that massive bulge-dominated galaxies exist from at least $z \sim 3$, suggesting that effective dissipative processes efficiently form bulges and destroy disks in the early Universe (e.g., Naab et al. 2007; Tacchella et al. 2016; Lapiner et al. 2023). However, these results may have been partially biased by the depth of observations coupled with cosmological dimming, which makes it difficult to detect a disk component. Our work, based on JWST data with significantly better sensitivity and increased spatial resolution by a factor of ~ 2 , tends to confirm the morphological classifications performed with HST for the vast majority of objects. As expected, we find that JWST classifications tend to present more disks, but the differences affect only a small number of objects that do not change the main trends.

Another key result of the past decades has been that galaxy morphologies become more irregular and disturbed as we move to higher redshifts (e.g., Abraham et al. 1996; Conselice et al. 2000; Guo et al. 2015; Huertas-Company et al. 2016). However, there has been some debate about whether this is due to more disturbed kinematics of stars or if it is biased because HST filters probe bluer light at high redshift. A disturbed appearance of the light distribution can also be caused by poor S/N, which tends to create noise fluctuations that can be interpreted as irregularities in the surface brightness distribution. Kinematic studies of ionized gas at high redshift (e.g., Genzel et al. 2008; Wisnioski et al. 2019) have indeed suggested that the majority of galaxies at $z \sim 1-2$ present clear rotation patterns, suggesting a more regular kinematic structure than what is inferred from the UV rest-frame light, although with higher velocity dispersion than local disks (e.g., Simons et al. 2016). Some works have tried to infer the stellar mass distribution from resolved pixel-based SED fitting (e.g., Wuyts et al. 2012), finding a smoother distribution than the one of light. Other works have also estimated that the contribution of bright UV clumps to the stellar mass is rather modest (Huertas-Company et al. 2020).

The JWST data, with its improved sensitivity, spatial resolution, and longer wavelength coverage allows us to revisit the issue of quantifying the abundance of galaxies with irregular morphologies using the rest-frame NIR, which tracks stellar mass more closely. Our work confirms that resolution and S/N play an important role in defining the class of irregular galaxies. When compared consistently, JWST-based classifications tend to find up to $\sim 30\%$ fewer irregular galaxies than HST-based ones on the same objects. However, despite this decrease, the reported fractions as a function of redshift appear to be very similar between HST and JWST – that is, when they are compared at similar rest-frame wavelengths – because it only affects large objects with a low S/N per pixel, which represent a small fraction of the galaxy population. The fraction of irregular galaxies is thus confirmed to increase and does not seem to be purely driven by an observational bias – at least at first order.

As previously mentioned, JWST does not only enable the quantification of the effect of S/N and resolution in the classification of irregular galaxies, but also allows us to look at morphologies in longer wavelengths than HST and therefore quantify how much of the increase in the fraction of irregular galaxies is driven by the light emitted by young stars. First works using JWST data have suggested that the fraction of regular disks might have been underestimated by HST (e.g., Ferreira et al. 2022; Kartaltepe et al. 2023; Robertson et al. 2023). We confirm

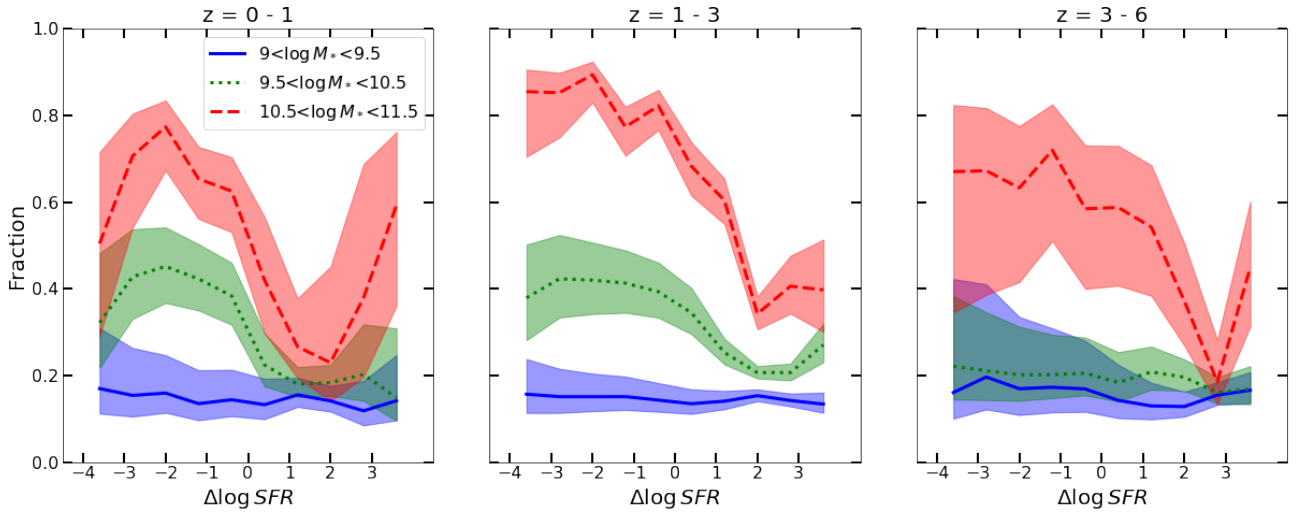


Fig. 17. Fraction of early-type galaxies as a function of $\Delta \log \text{SFR}$. Each panel shows a different redshift bin $-0 < z < 1$, $1 < z < 3$, $3 < z < 6$ – from left to right. The red dashed, green dotted and solid blue lines show galaxies with stellar masses of $10.5 < \log M_*/M_\odot < 11.5$, $9.5 < \log M_*/M_\odot < 10.5$, $9.5 < \log M_*/M_\odot < 10.5$ respectively. The morphology-quenching relation seems to be in place for massive galaxies at $z > 3$.

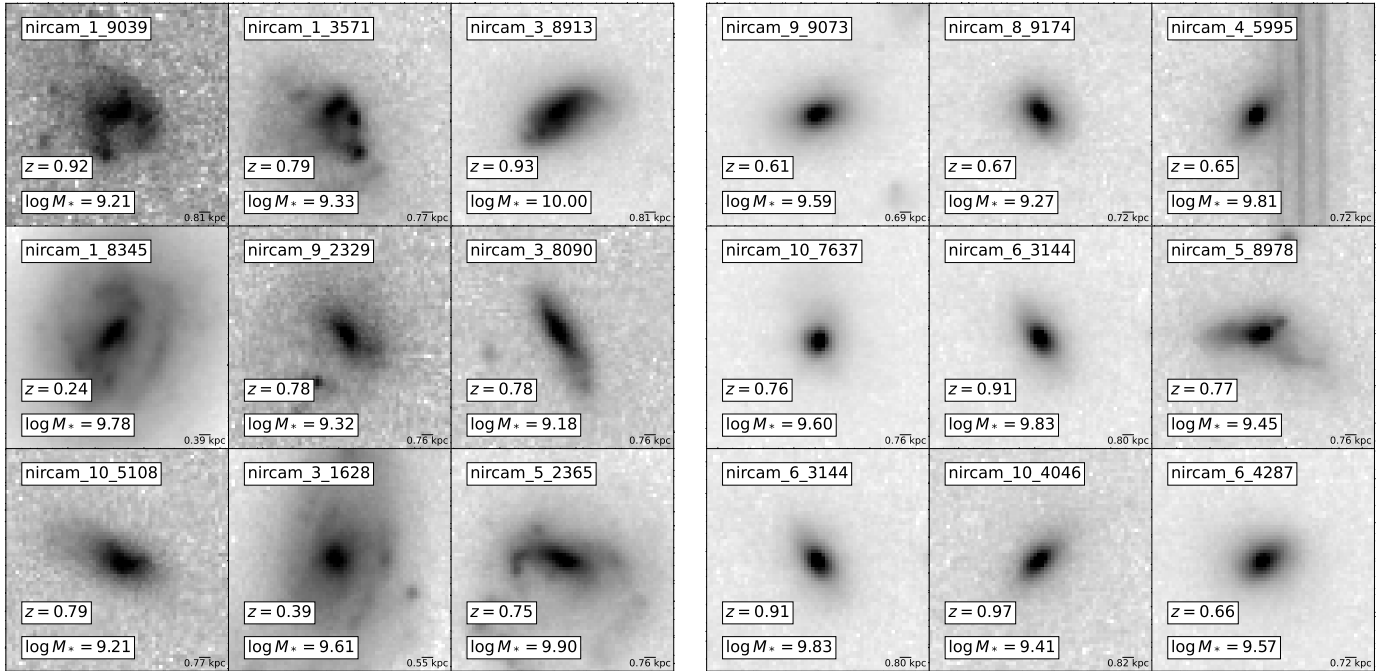


Fig. 18. Random example stamps of low mass ($9 < \log M_*/M_\odot < 10$) star-forming galaxies with late-type (left panel) and early-type morphologies (right panel). Images are in the $F200W$ filter.

in this work that $\sim 30\%$ of the galaxies tend to appear less disturbed when looked in the NIR rest-frame. This would point towards a smoother distribution of mass than previously measured with HST and more in agreement with pixel-based SED fitting results (Wuyts et al. 2012). We find that undisturbed disk-like morphologies are rather common at the high stellar mass end up to $z \sim 5$. However, low mass galaxies are still found to be irregular in their vast majority even in the NIR rest-frame, suggesting more perturbed kinematics than in the local Universe. This is in slight disagreement with the measurements of Ferreira et al. (2023) who find that disks still dominate at the low mass end. The discrepancies might be a consequence of the different definitions of peculiar galaxies. Future spectroscopic

JWST-based observations would allow us to further constrain the physical properties of these galaxies.

6.2. Onset of formation for disks and bulges

Understanding when and how bulges form remains a fundamental question in the field of galaxy formation for which the new JWST is expected to provide new insights. Our work indeed provides a new look into the emergence of morphological diversity from $z \sim 6$ as seen by JWST at longer wavelengths than ever before with HST.

Numerous previous works based on HST have shown that even if the fraction of bulge-dominated galaxies among massive

galaxies decreases as we move to a higher redshift, massive bulge-dominated galaxies are found at least from $z \sim 3$, suggesting that bulge formation starts very early (e.g., Costantin et al. 2021). In this work, we extended the census of bulges up to $z \sim 6$ and we can confirm a decreasing trend in terms of their abundance. At $z > 3$, the majority of massive galaxies are not bulge-dominated. However, we still measure $\sim 20\%$ of bulge-dominated galaxies at $z \sim 6$, which suggests that massive galaxy formation is well evolved already at these early epochs. Our work also shows that the connection between morphology and quenching, which has been reported in a large variety of works up to $z \sim 3$ (e.g., van Dokkum et al. 2008; Whitaker et al. 2012b; van der Wel et al. 2014a; Bluck et al. 2014; Barro et al. 2017 and many others), is already in place at $z \sim 5$. The fraction of bulge-dominated galaxies below the main sequence is close to $\sim 60\%$ also pointing towards a rapid assembly of stellar mass and efficient feedback mechanisms to quench star formation a few hundred million years after the Big Bang. In Fig. 19, we make an attempt to estimate the fraction of the stellar mass in bulges and disk components as a function of stellar mass and redshift. We do so by assuming that for galaxies classified as spheroids, all the mass is in the form of bulges, while for bulge+disk systems, we set it to an arbitrary number of 0.7 of the total mass of the galaxy. We estimated the amount of mass in bulges lies between the lower limit estimate from only spheroids and the upper limit, which comes from the sum of spheroids and composite systems. We understand that this is a fully first-order approximation with strong assumptions, which we still think can be used at least to explore the main trends. We see that up to $z \sim 3$, $\sim 60\%$ of the stellar mass in quiescent galaxies more massive than 10^{10} solar masses is in bulges, which is compatible with previous HST-based measurements (e.g., Lang et al. 2014; Huertas-Company et al. 2016). At $z \sim 5$, the fraction decreases to $\sim 30\%$, which suggests we are approaching the onset of massive bulge formation. However, there is still a significant population of massive bulges in place at these early epochs. We recall that these results might be partially affected by incompleteness as reported in Fig. 1. A careful comparison with the predictions of numerical simulations can help put some additional constraints on feedback models in the early Universe (e.g., Dekel et al. 2023).

Interestingly, for star-forming galaxies, the fraction of stellar mass in bulges remains of the order of $\sim 10\%$ at all epochs, also confirming previous findings (e.g., Dimauro et al. 2022). Although the abundance of these star-forming bulges, sometimes referred as “blue nuggets,” is small, numerous works have suggested they can represent an important phase for bulge build-up (see Lapiner et al. 2023 for a nice review on the topic). Our work confirms the presence of these systems. Although there might still be a disk component that is not detected, the fact that we see this star-forming bulge dominated population at low redshift and with deeper imaging than with HST, suggests that that population exists. Whether the morphological transformations and quenching are causally connected (e.g., Tacchella et al. 2016; Lee et al. 2018; Chen et al. 2020; Dimauro et al. 2022; Costantin et al. 2022; Lapiner et al. 2023) or whether it is just a consequence of progenitor bias (e.g., Lilly & Carollo 2016) is not something that our results can directly address. However, the fact that a galaxy bimodality already exists at these very early epochs, where the impact of progenitor bias should be rather limited given the short amount of elapsed time, suggests that some degree of morphological transformation is taking place before or after quenching.

Another important question is when the first disks are formed. The current cosmological model predicts that mas-

sive haloes are assembled by the merging of smaller ones. Simulations show that low-mass galaxies at high redshift have indeed perturbed kinematics or even prolate shapes (e.g., Tomassetti et al. 2016), which has also been hinted by some observational studies (e.g., van der Wel et al. 2014b; Zhang et al. 2019; Vega-Ferrero et al. 2024). Kassin et al. (2012) and Simons et al. (2017) also find that the abundance of rotationally supported systems increases with stellar mass with a transition mass that is increasing with redshift. A recent theoretical work by Dekel et al. (2020) suggests that gas disks only survive above a characteristic stellar mass of $\sim 10^9$ since, for lower mass systems, the frequent mergers change the spin in less than an orbital time, thus preventing disk formation. Although it is extremely difficult to infer the true disky nature of a system solely based on its apparent morphology (see e.g., Vega-Ferrero et al. 2024 for a discussion on this), our results looking at the NIR rest-frame morphologies suggest that massive “unperturbed” disk objects do exist at $z \sim 5$ (see e.g., Rizzo et al. 2020; Lelli et al. 2021 for similar conclusions based on gas). We also find that the abundance of galaxies morphologically identified as disks strongly depends on stellar mass in qualitative agreement with theoretical predictions and gas kinematic studies up to $z \sim 2$. Following an analogous procedure to what was done for the bulges in Fig. 19, we quantify the fraction of stellar mass in disks in star-forming galaxies as a function of stellar mass and redshift (irregular morphologies are not included in this estimate). Interestingly, we find that the fraction remains relatively constant across cosmic time between $\sim 30\%$ and $\sim 50\%$, suggesting that disk formation above a certain mass happens even at very early epochs. We emphasize that the measurements presented in this work do not allow us to firmly conclude on the true “disky” nature of the galaxies classified as “disks.” We recall that another potential source of bias could be related to the S/N (see Sect. 6.1), which can artificially increase the fraction of irregular galaxies. However, since HST and JWST-based classifications are shown to agree reasonably well, we might deduce that the impact of this bias is not dominant.

7. Summary and conclusions

In this work, we present a morphological classification of around 20 000 galaxies with $F200W[AB] < 27$ using JWST/NIRCam images in four different filters, $F150W$, $F200W$, $F356W$, and $F444W$, obtained from the CEERS survey. We classify galaxies into four classes – spheroid, bulge+disk, disk, and irregular – using a CNN and adversarial domain adaptation. The resulting classification shows excellent agreement with independent visual classifications, demonstrating the successful adaptation of the CNN to the new domain and will be made publicly available as part of the CEERS data products.

We compare our JWST-based classifications with existing HST/WFC3-based classifications. We find that $\sim 90\%$ and $\sim 75\%$ of galaxies with $z < 3$ have the same early-late and regular-irregular class, respectively, in both JWST and HST imaging when considering similar wavelengths. For the smallest and faintest objects, NIRCam based classifications tend to find fewer bulge-dominated and disturbed galaxies, likely due to a combination of SNR and spatial resolution. However, the impact on the measured morphological fractions as a function of cosmic time is minimal.

In the second part of the study, we analyze the rest-frame NIR ($\sim 0.8\text{--}1\ \mu\text{m}$) morphologies of a mass-complete sample ($\log M_*/M_\odot > 9$) of galaxies from $z \sim 6$ to $z \sim 0$. Our main findings are as follows.

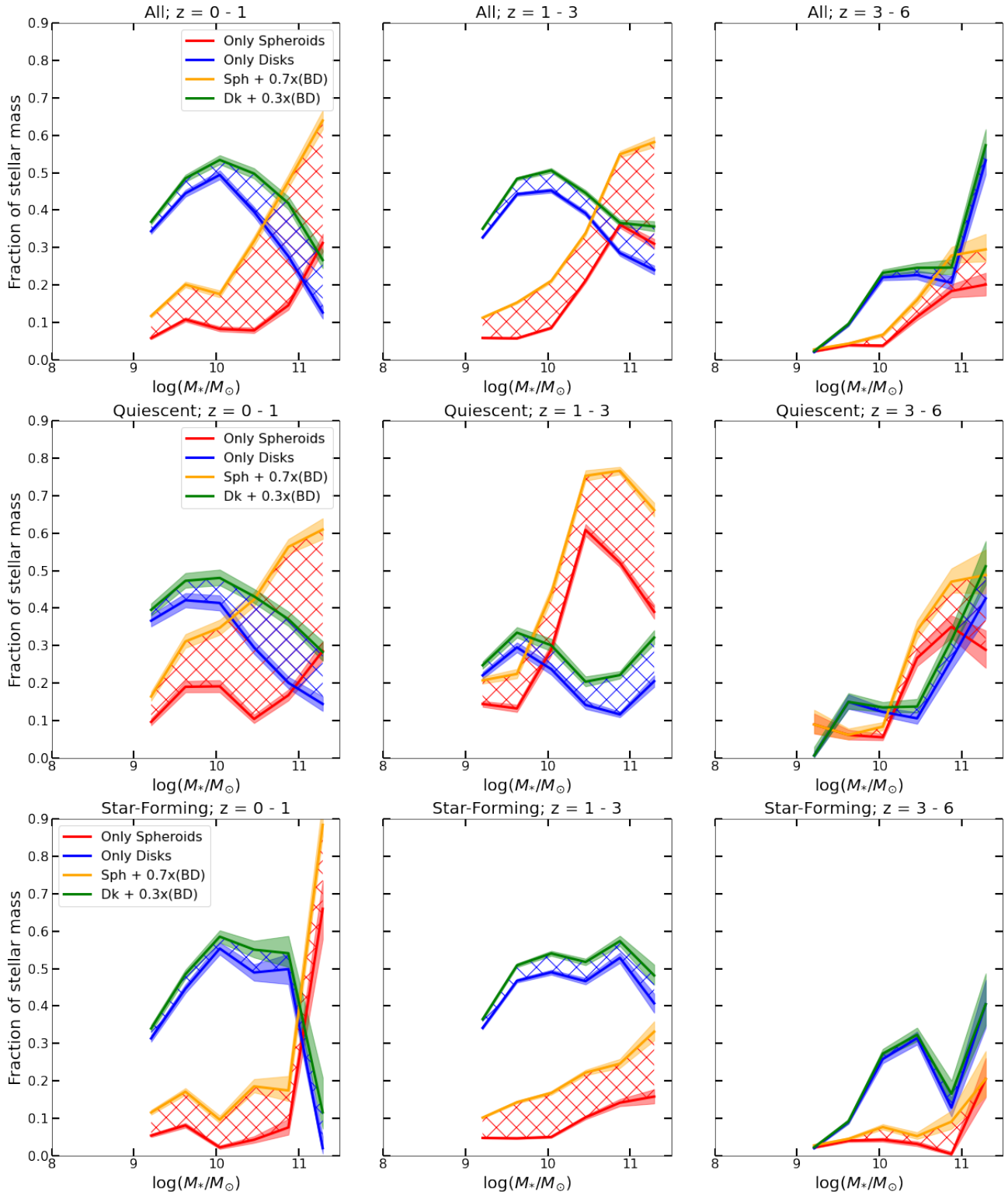


Fig. 19. Fraction of stellar mass in bulges (red hashed region) and disks (blue hashed region) for all (top row), quiescent (middle row) and star-forming (bottom row) galaxies in different redshift bins as labeled.

- The fraction of bulge-dominated galaxies increases at the high-mass end, even at $z \sim 5$, indicating that the processes of bulge formation in massive galaxies are already in place at these early cosmic epochs.
- The fraction of peculiar galaxies also increases with redshift, even in the NIR rest-frame, suggesting that the stellar mass distribution is more disturbed at high redshift, although the S/N may still affect this result.
- The high-mass end of the galaxy distribution ($\log M_*/M_\odot > 10.5$) is dominated by undisturbed disk-like morphologies even at $z \sim 5$, indicating that disk formation may be in place at very early epochs.

- The fraction of early-type galaxies reaches $\sim 70\%$ to $\sim 90\%$ for massive ($\log M_*/M_\odot > 10.5$) quenched galaxies, even at $z \sim 5$, suggesting that the connection between quenching and bulge growth is already established around ~ 1 Gyr after the Big Bang.

Overall, our results indicate a complex morphological diversity already in place ~ 1 Gyr after the Big Bang.

Acknowledgements. M.H.C., E.A., R.S. and J.V.F. acknowledge financial support from the State Research Agency (AEIMCINN) of the Spanish Ministry of Science and Innovation under the grants “Galaxy Evolution with Artificial Intelligence” with reference PGC2018-100852-A-I00 and “BASALT” with reference PID2021-126838NB-I00.

References

- Abraham, R. G., Tanvir, N. R., Santiago, B. X., et al. 1996, *MNRAS*, 279, L47
- Asada, Y., Sawicki, M., Desprez, G., et al. 2023, *MNRAS*, 523, L40
- Bagley, M. B., Finkelstein, S. L., Koekemoer, A. M., et al. 2023, *ApJ*, 946, L12
- Barro, G., Kriek, M., Pérez-González, P. G., et al. 2016, *ApJ*, 827, L32
- Barro, G., Faber, S. M., Koo, D. C., et al. 2017, *ApJ*, 840, 47
- Barro, G., Pérez-González, P. G., Cava, A., et al. 2019, *ApJS*, 243, 22
- Bluck, A. F. L., Mendel, J. T., Ellison, S. L., et al. 2014, *MNRAS*, 441, 599
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, 686, 1503
- Buitrago, F., Trujillo, I., Conselice, C. J., & Häußler, B. 2013, *MNRAS*, 428, 1460
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
- Cameron, E. 2011, *PASA*, 28, 128
- Chabrier, G. 2003, *PASP*, 115, 763
- Chen, Z., Faber, S. M., Koo, D. C., et al. 2020, *ApJ*, 897, 102
- Ćiprijanović, A., Kafkes, D., Jenkins, S., et al. 2020, arXiv e-prints [arXiv:2011.03591]
- Ćiprijanović, A., Lewis, A., Pedro, K., et al. 2023, *Mach. Learn.: Sci. Technol.*, 4, 025013
- Conselice, C. J. 2014, *ARA&A*, 52, 291
- Conselice, C. J., Bershad, M. A., & Jangren, A. 2000, *ApJ*, 529, 886
- Costantin, L., Pérez-González, P. G., Méndez-Abreu, J., et al. 2021, *ApJ*, 913, 125
- Costantin, L., Pérez-González, P. G., Méndez-Abreu, J., et al. 2022, *ApJ*, 929, 121
- Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, *ApJ*, 660, L1
- Dekel, A., Ginzburg, O., Jiang, F., et al. 2020, *MNRAS*, 493, 4126
- Dekel, A., Sarkar, K. C., Birnboim, Y., Mandelker, N., & Li, Z. 2023, *MNRAS*, 523, 3201
- Dimauro, P., Daddi, E., Shankar, F., et al. 2022, *MNRAS*, 513, 256
- Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., et al. 2019, *MNRAS*, 484, 93
- Ferreira, L., Adams, N., Conselice, C. J., et al. 2022, *ApJ*, 938, L2
- Ferreira, L., Conselice, C. J., Sazonova, E., et al. 2023, *ApJ*, 955, 94
- Finkelstein, S. L., Bagley, M. B., Haro, P. A., et al. 2022, *ApJ*, 940, L55
- Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, *ApJ*, 946, L13
- Förster Schreiber, N. M., & Wuyts, S. 2020, *ARA&A*, 58, 661
- Galametz, A., Grazian, A., Fontana, A., et al. 2013, *ApJS*, 206, 10
- Ganin, Y., Ustinova, E., Ajakan, H., et al. 2016, *J. Mach. Learn. Res.*, 17, 1
- Genzel, R., Burkert, A., Bouché, N., et al. 2008, *ApJ*, 687, 59
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35
- Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, *ApJS*, 207, 24
- Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, *ApJ*, 800, 39
- Huertas-Company, M., & Lanusse, F. 2023, *PASA*, 40, e001
- Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, *ApJS*, 221, 8
- Huertas-Company, M., Bernardi, M., Pérez-González, P. G., et al. 2016, *MNRAS*, 462, 4495
- Huertas-Company, M., Primack, J. R., Dekel, A., et al. 2018, *ApJ*, 858, 114
- Huertas-Company, M., Guo, Y., Ginzburg, O., et al. 2020, *MNRAS*, 499, 814
- Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, *A&A*, 556, A55
- Iyer, K., & Gawiser, E. 2017, *ApJ*, 838, 127
- Iyer, K. G., Gawiser, E., Faber, S. M., et al. 2019, *ApJ*, 879, 116
- Kartaltepe, J. S., Mozena, M., Kocevski, D., et al. 2015, *ApJS*, 221, 11
- Kartaltepe, J. S., Rose, C., Vanderhoof, B. N., et al. 2023, *ApJ*, 946, L15
- Kassin, S. A., Weiner, B. J., Faber, S. M., et al. 2012, *ApJ*, 758, 106
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36
- Lang, P., Wuyts, S., Somerville, R. S., et al. 2014, *ApJ*, 788, 11
- Lapiner, S., Dekel, A., Freundlich, J., et al. 2023, *MNRAS*, 522, 4515
- Lee, B., Giavalisco, M., Whitaker, K., et al. 2018, *ApJ*, 853, 131
- Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019, *ApJ*, 876, 3
- Lelli, F., Di Teodoro, E. M., Fraternali, F., et al. 2021, *Science*, 371, 713
- Lilly, S. J., & Carollo, C. M. 2016, *ApJ*, 833, 1
- Lower, S., Narayanan, D., Leja, J., et al. 2020, *ApJ*, 904, 33
- Madau, P. 1995, *ApJ*, 441, 18
- Mérida, R. M., Pérez-González, P. G., Sánchez-Blázquez, P., et al. 2023, *ApJ*, 950, 125
- Mowla, L., Iyer, K. G., Desprez, G., et al. 2022, *ApJ*, 937, L35
- Naab, T., Johansson, P. H., Ostriker, J. P., & Efstathiou, G. 2007, *ApJ*, 658, 710
- Nayyeri, H., Hemmati, S., Mobasher, B., et al. 2017, *ApJS*, 228, 7
- Olsen, C., Gawiser, E., Iyer, K., et al. 2021, *ApJ*, 913, 45
- Papaderos, P., Östlin, G., & Breda, I. 2023, *A&A*, 673, A30
- Pearson, W. J., Wang, L., Brough, S., et al. 2021, *A&A*, 646, A151
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, *AJ*, 139, 2097
- Planck Collaboration XVI. 2014, *A&A*, 571, A16
- Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, *A&A*, 523, A13
- Rizzo, F., Vegetti, S., Powell, D., et al. 2020, *Nature*, 584, 201
- Robertson, B. E., Tacchella, S., Johnson, B. D., et al. 2023, *ApJ*, 942, L42
- Simons, R. C., Kassin, S. A., Trump, J. R., et al. 2016, *ApJ*, 830, 14
- Simons, R. C., Kassin, S. A., Weiner, B. J., et al. 2017, *ApJ*, 843, 46
- Stefanon, M., Yan, H., Mobasher, B., et al. 2017, *ApJS*, 229, 32
- Tacchella, S., Dekel, A., Carollo, C. M., et al. 2016, *MNRAS*, 457, 2790
- Tacchella, S., Conroy, C., Faber, S. M., et al. 2022, *ApJ*, 926, 134
- Tomassetti, M., Dekel, A., Mandelker, N., et al. 2016, *MNRAS*, 458, 4477
- Trujillo, I., & Fliri, J. 2016, *ApJ*, 823, 123
- van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014a, *ApJ*, 788, 28
- van der Wel, A., Chang, Y.-Y., Bell, E. F., et al. 2014b, *ApJ*, 792, L6
- van Dokkum, P. G., Franx, M., Kriek, M., et al. 2008, *ApJ*, 677, L5
- Vega-Ferrero, J., Huertas-Company, M., Costantin, L., et al. 2024, *ApJ*, 961, 51
- Walmsley, M., Slijepcevic, I., Bowles, M. R., & Scaife, A. 2022, in Machine Learning for Astrophysics, proceedings of the Thirty-ninth International Conference on Machine Learning (ICML 2022), online at <https://ml4astro.github.io/icml2022>, 29
- Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012a, *ApJ*, 754, L29
- Whitaker, K. E., Kriek, M., van Dokkum, P. G., et al. 2012b, *ApJ*, 745, 179
- Wisnioski, E., Förster Schreiber, N. M., Fossati, M., et al. 2019, *ApJ*, 886, 124
- Wuyts, S., Förster Schreiber, N. M., van der Wel, A., et al. 2011, *ApJ*, 742, 96
- Wuyts, S., Förster Schreiber, N. M., Genzel, R., et al. 2012, *ApJ*, 753, 114
- Zhang, H., Primack, J. R., Faber, S. M., et al. 2019, *MNRAS*, 484, 5170

¹ Instituto de Astrofísica de Canarias (IAC), La Laguna 38205, Spain e-mail: mhurtas@iac.es

² Observatoire de Paris, LERMA, PSL University, 61 avenue de l’Observatoire, 75014 Paris, France

³ Université Paris-Cité, 5 rue Thomas Mann, 75014 Paris, France

⁴ Universidad de La Laguna, Avda. Astrofísico Fco. Sanchez, La Laguna, Tenerife, Spain

⁵ Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA

⁶ NSF’s National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Ave., Tucson, AZ 85719, USA

⁷ Department of Astronomy, The University of Texas at Austin, Austin, TX, USA

⁸ Department of Astronomy and Steward Observatory, University of Arizona, Tucson, AZ 85721, USA

⁹ Division of Science, National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

¹⁰ Departamento de Física Teórica, Atómica y Óptica, Universidad de Valladolid, 47011 Valladolid, Spain

¹¹ Instituto de Astrofísica e Ciências do Espaço, Universidade de Lisboa, OAL, Tapada da Ajuda, 1349-018 Lisbon, Portugal

¹² University of Massachusetts Amherst, 710 North Pleasant Street, Amherst, MA 01003-9305, USA

¹³ Centro de Astrobiología (CAB), CSIC-INTA, Ctra. de Ajalvir km 4, Torrejón de Ardoz, 28850 Madrid, Spain

¹⁴ Centre for Astrophysics and Planetary Science, Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel

¹⁵ Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, CA 95064, USA

- ¹⁶ Laboratoire AIM-Paris-Saclay, CEA/DRF/Irfu – CNRS – Université Paris Cité, CEA-Saclay, pt courrier 131, 91191 Gif-surYvette, France
- ¹⁷ Physics & Astronomy Department, University of Louisville, Louisville, KY 40292, USA
- ¹⁸ Laboratory for Multiwavelength Astrophysics, School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623, USA
- ¹⁹ Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA
- ²⁰ Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA
- ²¹ George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA
- ²² ESA/AURA Space Telescope Science Institute, Baltimore, MD, USA
- ²³ Aix-Marseille Univ., CNRS, CNES, LAM Marseille, Marseille, France
- ²⁴ Department of Physics and Astronomy, University of California, 900 University Ave, Riverside, CA 92521, USA
- ²⁵ Department of Physics, University of Bath, Claverton Down, Bath BA2 7AY, UK
- ²⁶ Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- ²⁷ SRON Netherlands Institute for Space Research, Postbus 800, 9700 AV Groningen, The Netherlands
- ²⁸ Astrophysics Science Division, NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA
- ²⁹ Department of Physics and Astronomy, Colby College, Waterville, ME 04901, USA
- ³⁰ Center for Computational Astrophysics, Flatiron Institute, New York, USA

Appendix A: Star and galaxy separation

We used a simple procedure to perform a rough star-galaxy separation in the F200W filter using three parameters delivered by SEXTRACTOR: A_{IMAGE} , F_{200} and $CLASS_{STAR}$, which measure the isophotal image major axis, isophotal flux, and a stellarity flag computed with a pre-trained Neural Network, respectively. In Figure A.1 we plot $\log F_{200}$ vs. $\log A_{IMAGE}$ for all galaxies in our sample and for those with a value of $CLASS_{STAR}$ larger than 0.95. We see that the latter tend to be located in a well-defined sequence separated from the bulk of the distribution. We thus identify this sequence as the locus of bright stars and draw an empirical linear relation below which objects are flagged as stars. While this recipe might work reasonably well to identify bright stars, the separation is expected to be much less obvious at faint fluxes, as seen in Figure A.1. However, given that the main analysis of this work focuses on galaxies more massive than 10^9 solar masses, we do not expect the contamination of stars to significantly affect any of the conclusions.

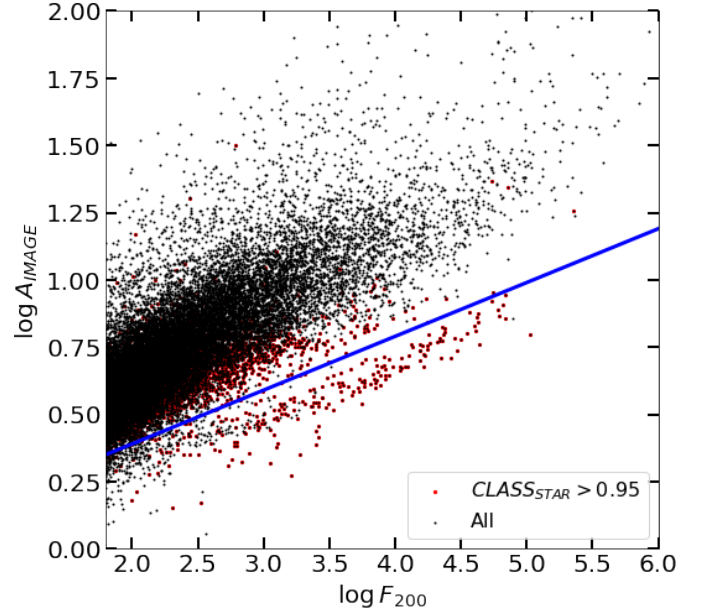


Fig. A.1. $\log F_{200} - \log A_{IMAGE}$ plane used in this work for star-galaxy separation. The black dots show all the sample of selected objects with F200W AB apparent magnitude smaller than 27. The red points are those objects with a value of $CLASS_{STAR} > 0.95$. The blue solid line show the boundary used to separate stars (below the line) from galaxies (above the line).

Appendix B: HST vs. JWST stamps

We show in Figure B.1 some examples of galaxies with different morphological classifications in CEERS and CANDELS observed at a similar wavelength and at the same pixel scale of $0.03''/\text{pixel}$. We can see how resolution and depth can impact the classification. Objects whose classifications change for early-type in CANDELS to late-type in CEERS are typically small (as

shown in Figure 10) and therefore appear rounder in CANDELS. Regarding objects which are classified as irregular in CANDELS but not in CEERS, we see there are more extended and the better sensitivity and resolution of CEERS enables a clearer detection of a diffuse component, which can be interpreted as a disk. Nevertheless, galaxies still tend to present some asymmetries which illustrates the difficulty of defining the irregular class.

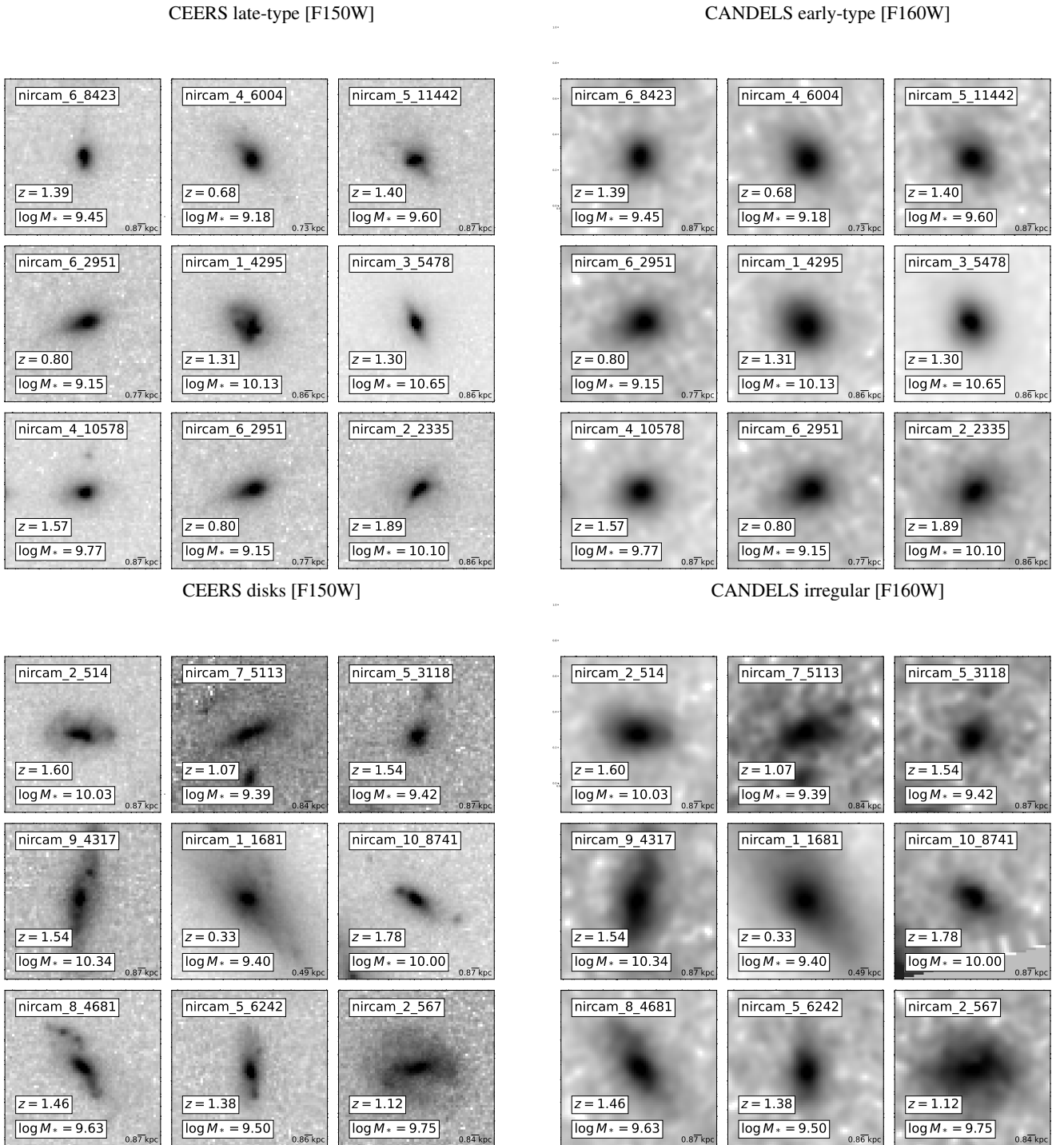


Fig. B.1. Example of random stamps of the same galaxies in CEERS (F150W) and CANDELS (F160W) at the same pixel scale with different morphological classifications. Two top panels show galaxies classified as late-type in CEERS (top-left) and as early-type in CANDELS (top-right). The bottom panels show galaxies classified as disks in CEERS (bottom-left) and irregular in CANDELS (bottom-right). The physical scale in kpc is shown for every galaxy. A square root scaling has been applied to enhance the outskirts.

Appendix C: Spatial resolution and rest-frame wavelength

In this work, we try to quantify the evolution of galaxy morphology in the same rest-frame wavelength by using different filters depending on the redshift bin. This implies that the effective spatial resolution also changes. In Figure C.1 we show the evolution of the rest-frame wavelength and spatial resolution for the three different filters used in this work. The spatial resolution is computed as the full width at half maximum (FWHM) value for every filter reported in the Table 1 of Finkelstein et al. (2023) and divided by 2.35.

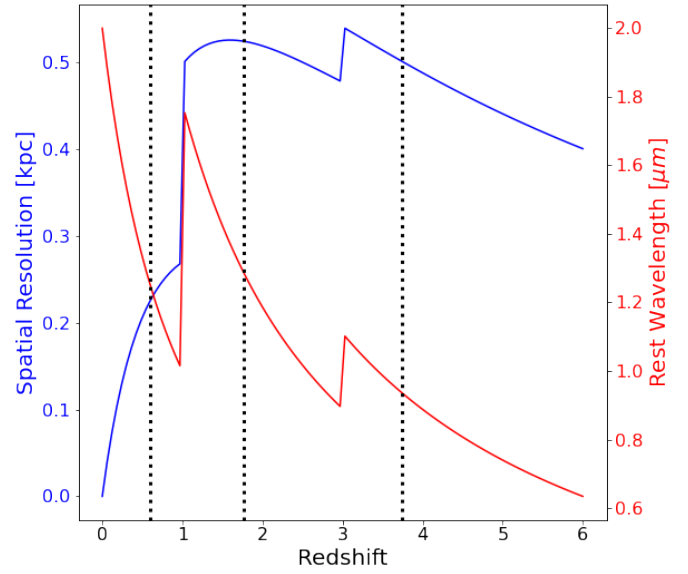


Fig. C.1. Spatial resolution in kpc (blue solid line) and rest wavelength (red solid line) as a function of redshift for the three different filters used in this work: *F200W* at $z < 1$, *F356W* at $1 < z < 3$ and *F444W* at $z > 3$. The dashed vertical lines indicate the mean redshift in each of three redshift bins used throughout this work.

Appendix D: Comparison of deep learning based morphological classifications with Sersic fits

We show in Figures D.1 and D.2 the comparison between parametric morphologies and our deep learning-based morphologies

for the F356W and F444W filters, respectively, as done in the main text for F200W. Similar trends are observed.

F356W - Mean and NMAD of Sersic Index

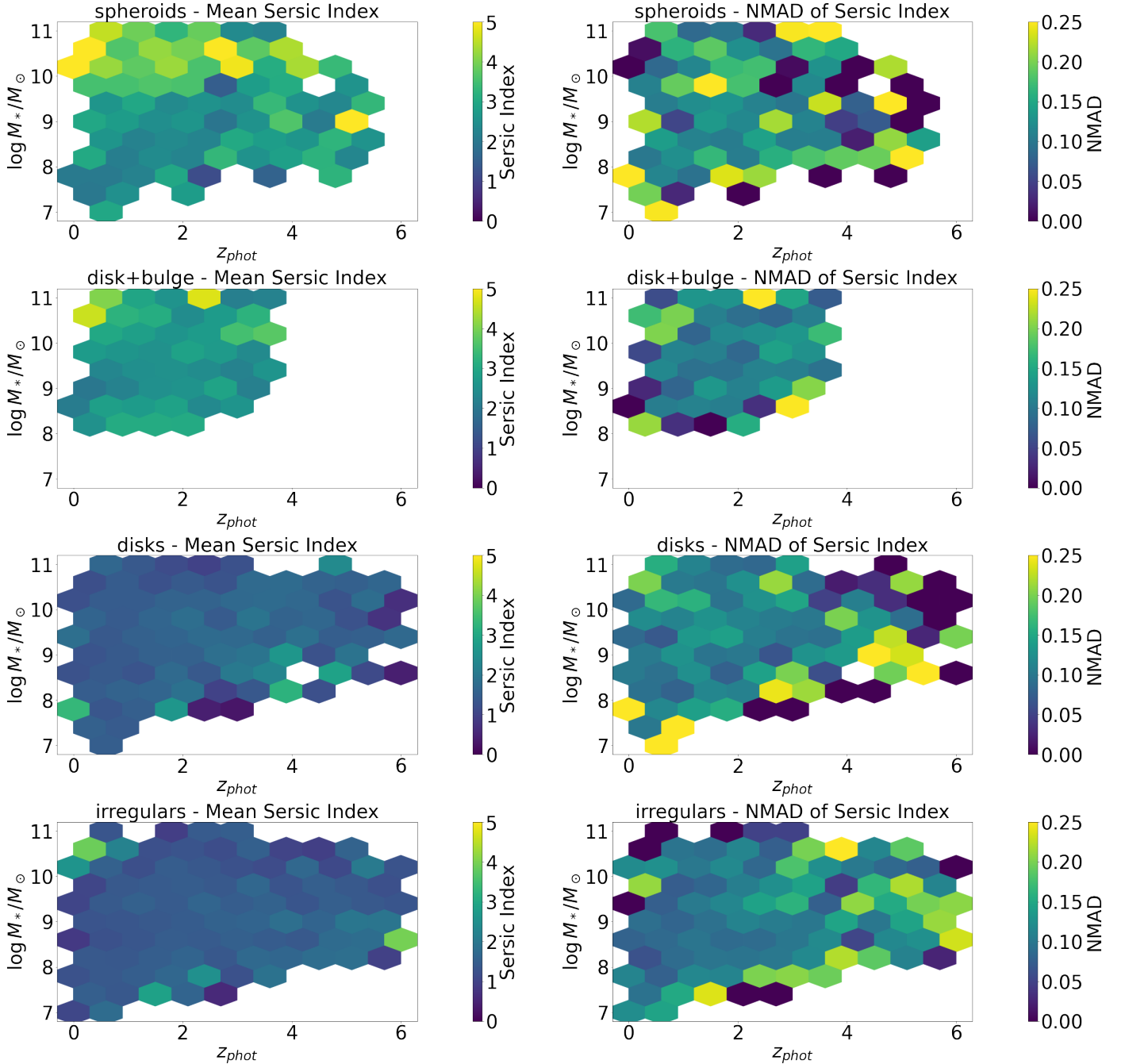


Fig. D.1. Mean values (left column) and normalized mean absolute deviations (right column) of the Sersic index for different morphological classes in the F356W filter in different bins of stellar mass and redshift. From top to bottom: Spheroids, disk+bulge, disks and irregulars. We only include objects with converged galfit measurements (flag=0) and only cells with more than ten galaxies are reported.

F444W - Mean and NMAD of Sersic Index

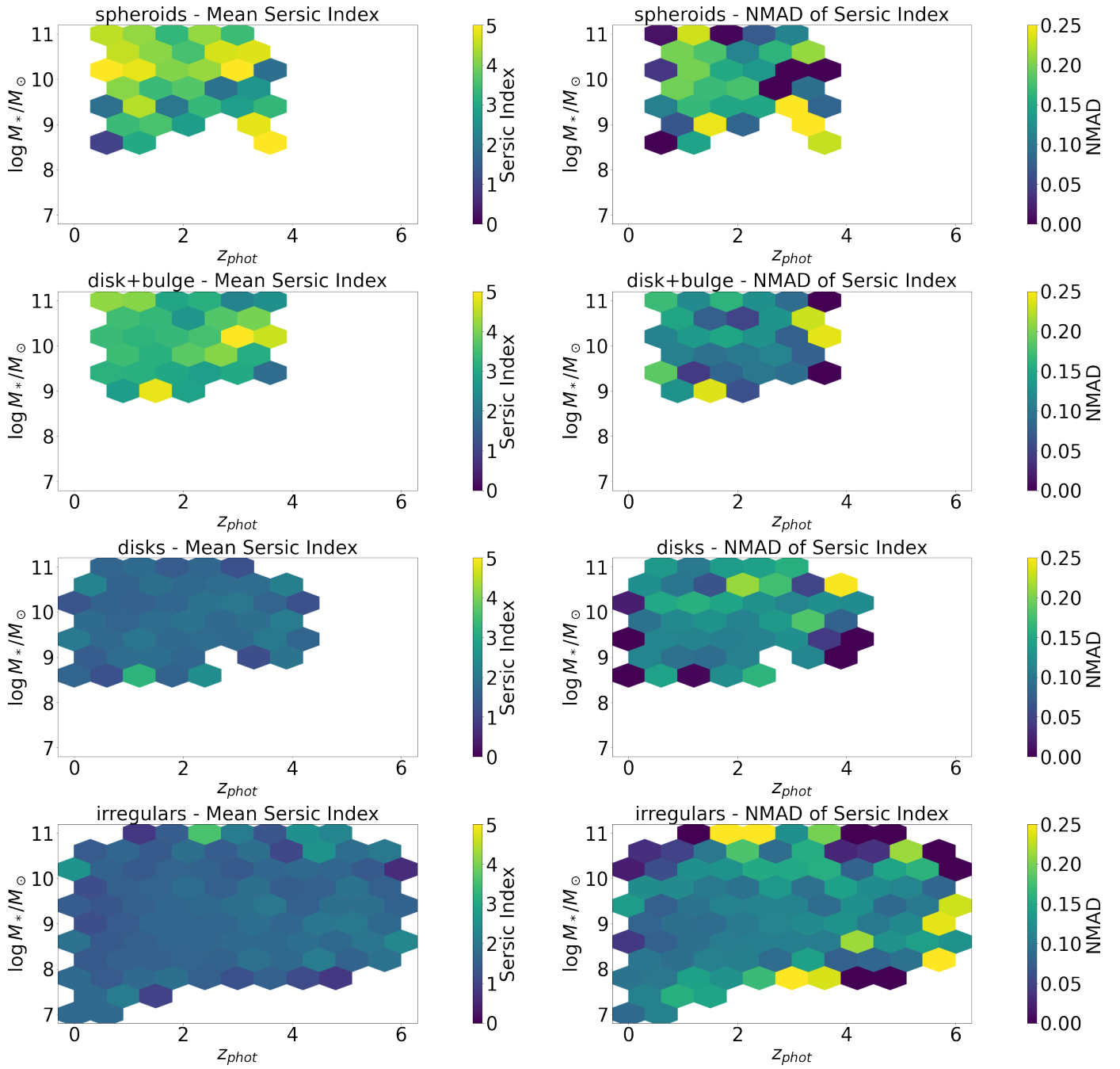


Fig. D.2. Mean values (left column) and normalized mean absolute deviations (right column) of the Sersic index for different morphological classes in the F444W filter in different bins of stellar mass and redshift. From top to bottom > spheroids, disk+bulge, disks, and irregulars. We only include objects with converged galfit measurements (flag=0) and only cells with more than ten galaxies are reported.