CrossMark

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

Leonardo Ferreira[1] ⊙, Christopher J. Conselice[1,2,3] ⊙, Ulrike Kuchner[1] ⊙, and Clár-Bríd Tohill[1] ⊙
[1] Centre for Astronomy and Particle Physics, University of Nottingham, Nottingham, UK; leonardo.ferreira@nottingham.ac.uk, conselice@gmail.com
[2] Jodrell Bank Centre for Astrophysics, University of Manchester, Oxford Road, Manchester, UK
[3] Alan Turing Fellow, Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK

Abstract

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

1. Introduction

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

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

To make progress in understanding the evolution of galaxies, it is crucial to identify merging galaxies correctly. In order to separate galaxies into mergers and nonmergers—initially focusing on the nearby universe—quantitative morphology tools were developed that use parameters such as the asymmetry index (A; e.g., Conselice et al. 2000; Conselice 2003). Merging galaxies are often identified through a combination of these morphological measurements such as the concentration, asymmetry, and smoothness (CAS) parameters (e.g., Conselice et al. 2003; Lotz et al. 2004). However, mergers are not uniquely identifiable in this parameter space and some do not fall into the selection criteria at all (e.g., Conselice 2006; Lotz et al. 2008). Therefore, care has to be taken to calibrate their usage.

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

In addition to high merger rates, distant galaxies have much higher star formation rates than today, peaking at $z \sim 2$ (e.g., Madau & Dickinson 2014). We further know that galactic structure is highly dependent on the star formation rate in the sense that intensely star-forming galaxies generally appear more clumpy and irregular than quiescent galaxies at all redshifts (e.g., Windhorst et al. 2002; Guo et al. 2015, 2018; Mager et al. 2018; Sazonova et al. 2021). In fact, these two different types of galaxies—mergers and noninteracting intensely star-forming galaxies—can look very similar by eye, which complicates visual classifications. Even kinematically it can be challenging to distinguish mergers from rotating galaxies with high dispersions (e.g., Simons et al. 2019; Bottrell et al. 2022). In addition, classifications and selections of galaxies after a merger event (post-mergers) are highly contaminated by misclassified isolated galaxies with high specific star formation rates (sSFR). This is because their star-forming regions and dusty interstellar medium can generate asymmetric features reminiscent of (post-)merger features. It is therefore currently unknown if and how we can correctly distinguish whether a galaxy is undergoing intense star formation or some type of merger using galactic structures and morphologies.

One way to approach this question is through novel techniques using machine learning. Recently, tremendous progress has been made in applying supervised deep learning methods to investigate galaxy morphology (e.g., Huertas-Company et al. 2018, 2019, 2020; Cheng et al. 2020; Reiman & Göhre 2019; Martin et al. 2020; Walmsley et al. 2020, 2022). These end-to-end techniques are also very promising for investigating galaxy mergers specifically (Ackermann et al. 2018; Bottrell et al. 2019; Pearson et al. 2019a, 2019b; Ferreira et al. 2020; Wang et al. 2020; Bickley et al. 2021). Additionally, one can also leverage information not only from visual classifications and observations, but also by forward-modeling cosmological simulations to the observational domain (Ćiprijanović et al. 2020, 2021).

We have recently started a machine-learning exercise to determine the merger history of galaxies using cosmological simulation runs from IllustrisTNG (Vogelsberger et al. 2014; Pillepich et al. 2018b; Nelson et al. 2019). In Ferreira et al. (2020), we were able to separate mergers from other types of galaxies in IllustrisTNG to a success rate of 90% up to $z \sim 3$. The present paper is a follow-up to our first paper, in which we now investigate whether it is possible to distinguish merging galaxies from intensely star-forming galaxies. These galaxies have the lowest success rates in classifications from Ferreira et al. (2020). Our task in this paper is to correctly distinguish mergers from star-forming galaxies by only using their morphology and structure.

This paper is organized as follows: in Section 2 we describe the data sets we constructed for this task, from IllustrisTNG (simulations) and CANDELS (observations). A description of the methods we used to train a deep learning model and how we measure the structure of the galaxies in our samples is given in Section 3. We present our results in Section 4 while a discussion on the implications is laid out in Section 5. Finally, we summarize and conclude our findings in Section 6.

2. Data

To test our new deep learning approach, we use simulated galaxies from cosmological simulations post-processed with the SKIRT (Camps & Baes 2015, 2020) dusty radiative transfer code. The simulations are based on IllustrisTNG (Section 2.1), which are used for the construction of the training sample for a convolutional neural network (CNN) that is subsequently applied to observed galaxies from the CANDELS fields (Section 2.3). Our sample definitions for post-mergers and star-forming galaxies are given in Section 2.2. We discuss the pipeline used to generate CANDELIZED mock images from IllustrisTNG in Section 2.4.

2.1. IllustrisTNG

IllustrisTNG is a suite of cosmological, gravo-magneto-hydrodynamical simulation runs with a diverse set of particle resolutions. From highest to lowest resolution, simulations were realized in three comoving simulation boxes of 50, 100, and 300 Mpc h^{-1} length size, aptly named TNG50, TNG100, and TNG300 (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018, 2019; Pillepich et al. 2018a, 2019; Springel et al. 2018).

Our analysis makes use of the TNG100-1 simulation, which has proven to be a good compromise between resolution and volume.[4] TNG100 has already been used extensively in studies that analyze galactic morphologies and structures, including the comparison between simulations and observations (Huertas-Company et al. 2019; Blumenthal et al. 2020), and the use for deep learning (Wang et al. 2020; Bickley et al. 2021; Bottrell et al. 2022). Specifically, Zanisi et al. (2021) showed that TNG100 galaxies reproduce observed objects well, especially disk-dominated sources. While there are some deviations in the small-scale structure of highly concentrated spheroidal systems, this is a minor issue in our analysis since they only make up a small fraction of our sample. In addition, our galaxies are resolution limited at the current redshift of interest, meaning that tiny details of structure are not relevant in this analysis.

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

This approach is a noticeable refinement to our previous treatment in Ferreira et al. (2020), where the research question did not demand the resolution of fine morphological features

---

[4] We tested our selections on TNG50-1, but the resulting samples are too small for deep learning training.

like clumpy regions and tidal features, which are required for the present analysis.

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

$$\frac{M_{\mathrm{DM}}}{M_{\mathrm{total}}} > 0.1, \tag{1}$$

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

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

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

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

2.2. Sample Definitions

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

$$\mu = \frac{M_2}{M_1}, \quad \mu \geqslant 0.25, \tag{3}$$

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

$$\mathrm{sSFR} > 10^{-9.5}\ \mathrm{yr}^{-1} \tag{4}$$

and are not interacting with other galaxies. To isolate noninteracting cases, we exclude any galaxy from the

Table 1
Summary of the Initial IllustrisTNG Sample

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

Note. The numbers in this table represent the sample before each galaxy was post-processed with SKIRT and CANDELIZED mocks pipeline (see the text for details), during which each image was augmented by 20 for four orientations and five different fields.

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

In summary, this selection results in a sample of ~ 6000 post-mergers and $\sim 110,000$ noninteracting star-forming galaxies. While this may be a realistic representation of actual fractions (only $\sim 5\%$ of the sample are post-mergers), training the network requires a balanced data set. We thus use the post-merger sample as the baseline and separate it in bins of redshift, stellar mass, and size, randomly sampling the same number of noninteracting galaxies within each bin. We remove bins without adequate matched numbers of star-forming galaxies. This becomes noticeable in the higher mass bins where post-mergers dominate and very few star-forming galaxies are present.

After matching the samples, we count ~ 4000 galaxies in each class as our final sample. A summary of this sample separated by class and redshifts is available in Table 1. The distribution of redshifts, star-forming rates, stellar masses, and stellar half-mass–radius is shown in Figure 1 for post-mergers in red, and star-forming galaxies in blue. Both classes have very similar physical properties, with a small excess of large, passive and massive post-mergers in comparison to the star-forming galaxies. Additionally, the top-right and bottom-right panels of Figure 1 show the time since the last major-merger event, τ, and the mass ratio, respectively, for post-mergers. The nature of the distribution for τ arises from the average time between snapshots in the simulation of around ~ 0.15 Gyr. This timescale represents one to three snapshots after the coalescence of stellar masses.

2.3. CANDELS Fields

One of the main goals of this work is to predict star-forming and post-merger galaxies in the observed CANDELS imaging data (Grogin et al. 2011; Koekemoer et al. 2011), which comprises high-quality HST observations from COSMOS, UDS, EGS, GOODS-South, and GOODS-North (Grogin et al. 2011; Koekemoer et al. 2011). CANDELS data has been used extensively for galaxy merger studies, with estimated merger rates up to $z \sim 6$ (e.g., Mantha et al. 2018; Duncan et al. 2019;
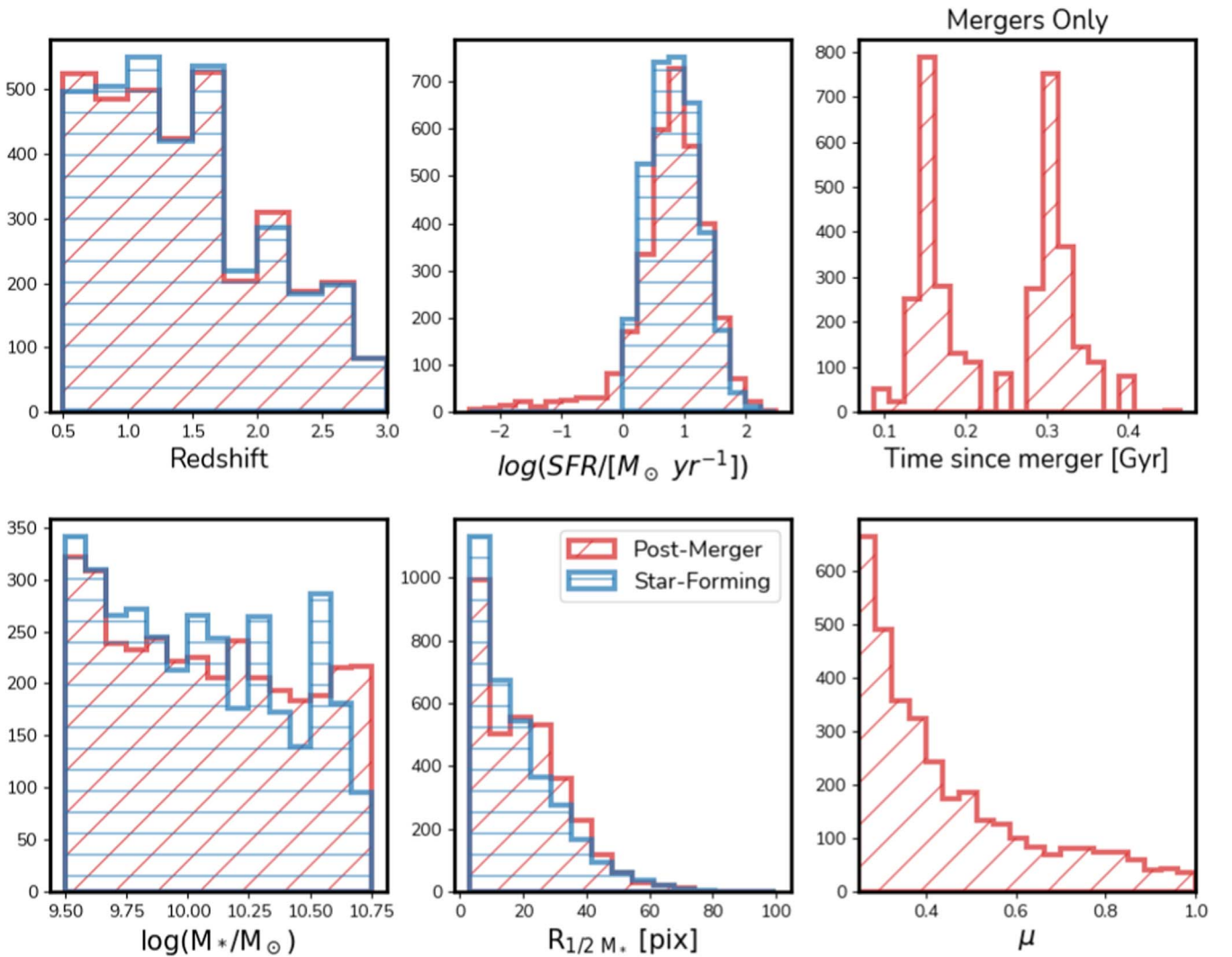
Figure 1. Physical properties of the 8000 IllustrisTNG TNG100-1 selected simulated galaxies. For both types of galaxies, we show distributions for redshifts (top left), star formation rates (top middle), stellar masses (bottom left), and stellar half-mass–radius (bottom middle) in red for post-mergers, and blue for star-forming galaxies. Distributions agree in general, with a small excess of stellar mass and size for the post-mergers. The time since the last major merging event and the mass ratio, μ,—properties unique to the post-mergers—are shown in top-right and bottom-right panels, respectively.

Whitney et al. 2021). Importantly, CANDELS also provides visually classified morphologies (Kartaltepe et al. 2015), as well as photometric redshifts, star formation rates, and stellar mass estimates (Conselice et al. 2007; Duncan et al. 2014, 2018a, 2018b, 2019) from spectral energy distribution (SED) fitting, essential for creating matched samples. Stellar masses are estimated within a 0.2 dex systematic uncertainty, and the outlier fraction of photometric redshifts is smaller than 5%. We do not account for these uncertainties directly in our methods, but we refer the reader to the detailed discussion on the uncertainties associated with these physical measurements in the aforementioned references. Furthermore, the depth of the wide field data is comparable to the detection limit of galaxies in the IllustrisTNG simulations in the stellar mass range studied here.

To select CANDELS galaxies, we first remove all problematic objects according to their quality flags as recorded in the photometric catalog and in the Kartaltepe et al. (2015) catalogs to avoid edges, artifacts, and stars. Following Huertas-Company et al. (2016) and Kartaltepe et al. (2015), we then select galaxies with H-band magnitudes $H < 24.5$ mag.

Because this cut can bias our sample against extended sources, we also include a signal-to-noise ratio (S/N) lower limit of S/N > 50 to exclude any compact source with only a few bright pixels. This magnitude cut removes 1074 sources, while the S/N cut further removes 430 sources. Then we proceed with the same cuts we used to select IllustrisTNG galaxies, using $0.5 < z < 3$ and $M_* > 10^{9.5} M_\odot$. We apply a final cut using the asymmetry ($A > 0.1$; Section 3.3) to remove regular unambiguous galaxies with no apparent disturbed or asymmetric features. This ultimately results in a sample of 23,494 galaxies from all of the CANDELS fields combined.

Finally, we produce cutouts for I_{814}, J_{125}, and H_{160W} bands centering on each selected CANDELS galaxy, each with a field of view of 50 kpc × 50 kpc, using photometric redshifts from Duncan et al. (2019), preserving relative sizes between galaxies. Importantly, this selection does not rely on size measurements that could easily be spurious in interacting or merging galaxies. We do not find any bias in our classifications that could be attributed to small changes of the field of view caused by the photometric redshift uncertainties.
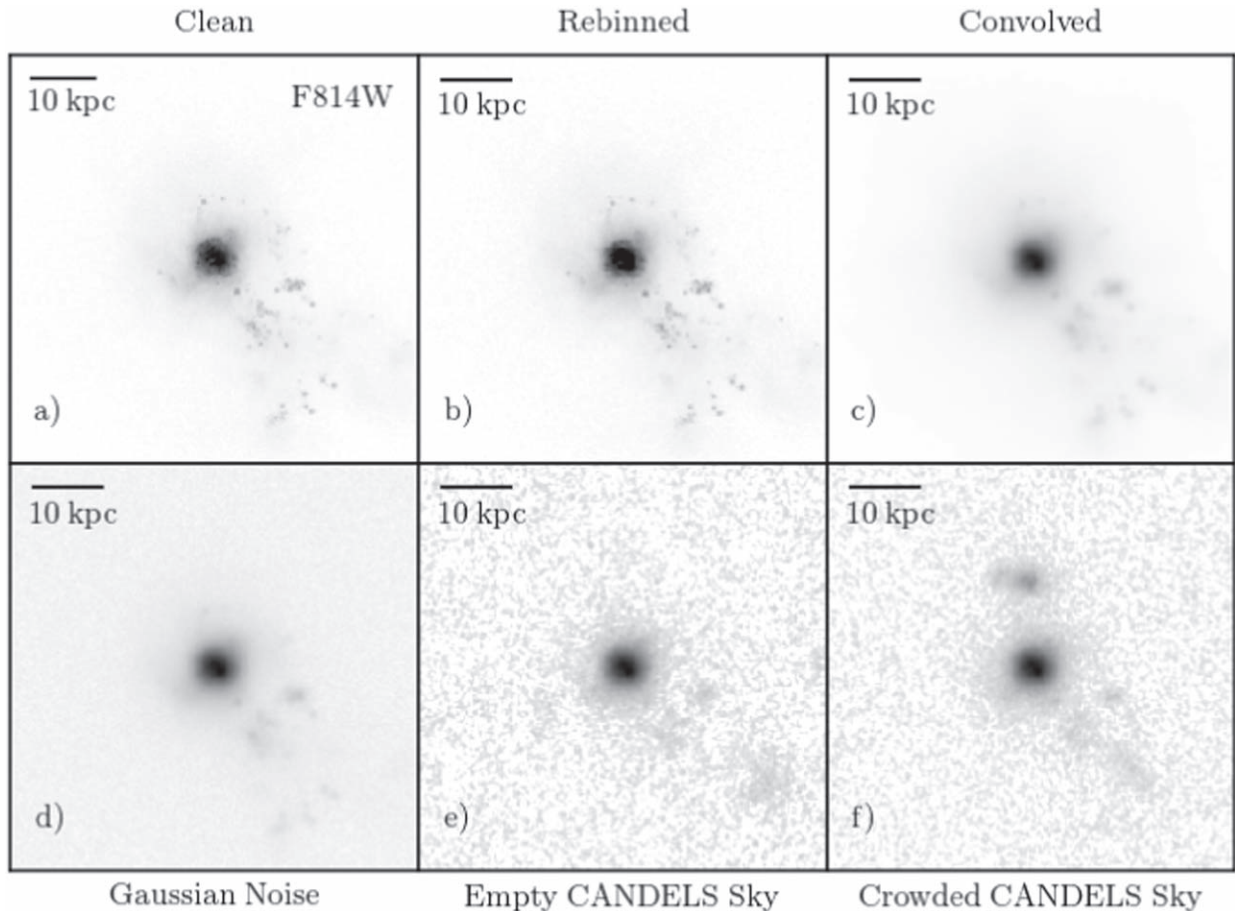
Figure 2. Example of the processing steps of our mock pipeline. (a) Noiseless F814W broadband image generated from the simulated galaxy data cube with $0\rlap{.}''03$ pix^{-1} pixel scale. (b) The same image after rebinning from $z = 0.5$ to $z = 0.6$. (c) Image convolved by the HST F814W PSF. (d) Image with Gaussian noise added. (e) Image added on top of a random patch of the sky within a CANDELS field with no neighboring sources. (f) Image added randomly to a patch of sky with other sources in the field of view. As this patch of the sky is randomly selected, all final images have varying levels of contamination from nearby sources. We quantify this by the total flux in the sky patch before adding the simulated source to it.

2.4. Pipeline to Produce CANDELIZED Mocks

In order to guarantee realistic representations of CANDELS galaxies in the simulated sample, we must include instrumental and cosmological effects to the images of the IllustrisTNG galaxies. An overview of the steps is shown in Figure 2 and is detailed in this section. IllustrisTNG data holds information on the stellar, gas, and dark matter particles for each source. Each particle represents a large physical region that can be described by rich stellar populations, which vary depending on age, mass, and metallicity. The resampling of the star-forming regions is particularly important to avoid problems with the coarse representations (Camps et al. 2016; Trayford et al. 2017).

To create mock broadband images, we thus process each stellar particle with a population synthesis model following the recipes from Trayford et al. (2017) and Vogelsberger et al. (2020). This entails post-processing the simulation data with the Monte Carlo dusty radiative transfer code SKIRT (Camps & Baes 2015, 2020).

Each stellar particle in the simulation is considered as a single stellar population with GALAXEV (Bruzual & Charlot 2003) or MAPPINGSIII (Allen et al. 2008) SEDs based on its stellar mass, absolute metallicity, and age. We choose to adopt these particular templates because, first, they

are implemented in SKIRT and, second, they had been tested previously in similar pipelines to generate mock observations from cosmological simulations (Trayford et al. 2017; Rodriguez-Gomez et al. 2019). Finally, the templates of Bruzual & Charlot (2003) are also those used to derive stellar masses and star formation rates for all of the CANDELS fields in Duncan et al. (2019) that are used in this study.

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

This process produces IFU data cubes over the SKIRT wavelength grid, which we then reduce to broadband images with the same properties as the CANDELS HST images. SKIRT's reference frame used to generate the data cubes is located at a distance of 10 Mpc (initial redshift z_0) of the sources. We must therefore shift the IFU data to each target's
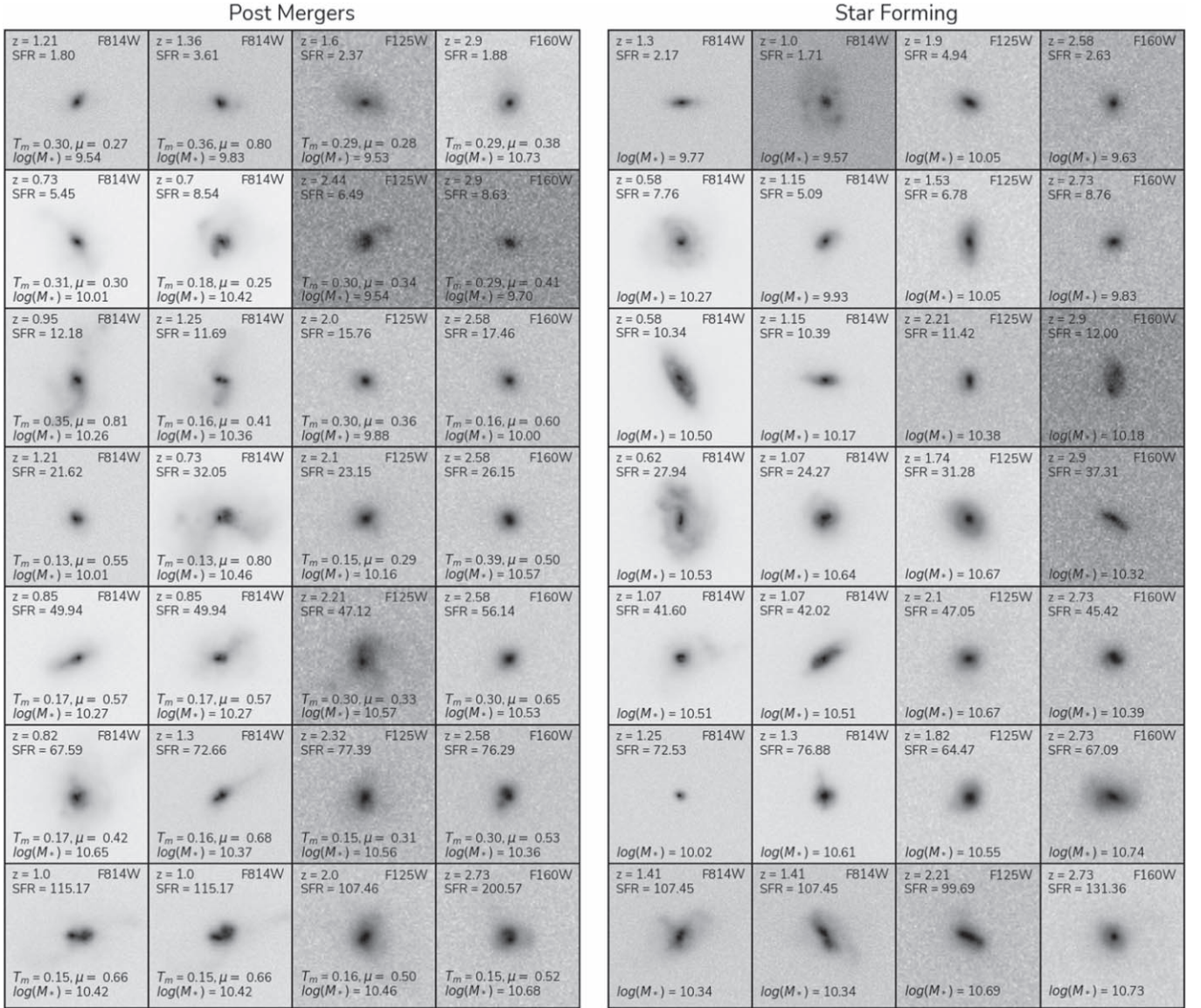
Figure 3. A random selection of IllustrisTNG simulated galaxies in our test sample; shown are post-mergers (left) and star-forming galaxies (right), with their redshifts, SFRs, and stellar masses printed in each stamp. Images are ordered from left to right in redshift, and top to bottom in SFR. For post-mergers, we also display the time since merger, T_m, and the mass ratio μ. All stamps use a square-root normalization.

redshift, z_t, by $(1 + z_t)$ while dimming its flux by

$$\frac{f_t}{f_o} = (1 + z_t)^{-1}\left(\frac{D_{Lo}}{D_{Lt}}\right)^2, \qquad (5)$$

due to cosmological dimming (Hogg 1999; Equation (15)). Next, we convolve the IFU data with the broadband filter's response functions for I_{814}, J_{125}, and H_{160W}. The results are clean, noiseless images from the simulation galaxies at 30 mas pix^{-1} (matching the ACS pixel scale) before adding any PSF effects (Figure 2, images (a)–(c)). We rebin the J and H bands from 30 mas pix^{-1} to the WFC3 images pixel scales of 60 mas pix^{-1}. Examples for stamps where the background was added can be seen in Figure 2, images (e) and (f).

Figure 3 gives randomly selected examples of galaxies in our sample before any contamination from the CANDELS sky is included, separated by their class.

The data-driven paradigm of deep learning methods imposes high requirements on the amount of data necessary to train a model that is capable of generalizing the training data well. In practice, this means that for the majority of models, a successful approach requires tens, hundreds, or even millions of examples. We are far away from these numbers in cosmological simulations. Our initial selection results in a balanced set of ~4000 examples of each class (Section 2.1). Fortunately, in the case of galaxy images, there are ways to increase the initial data set by exploiting aspects of the final image that do not depend directly on the simulated galaxy. In our case, we apply data augmentation to our data set in three ways outlined below. An example of this approach is shown in Figure 4, following the same galaxy in each possible combination of orientation/field.

First, since IllustrisTNG provides the 3D distribution of all particles associated with a galaxy, we generate each galaxy with different line-of-sight projections, treating each new
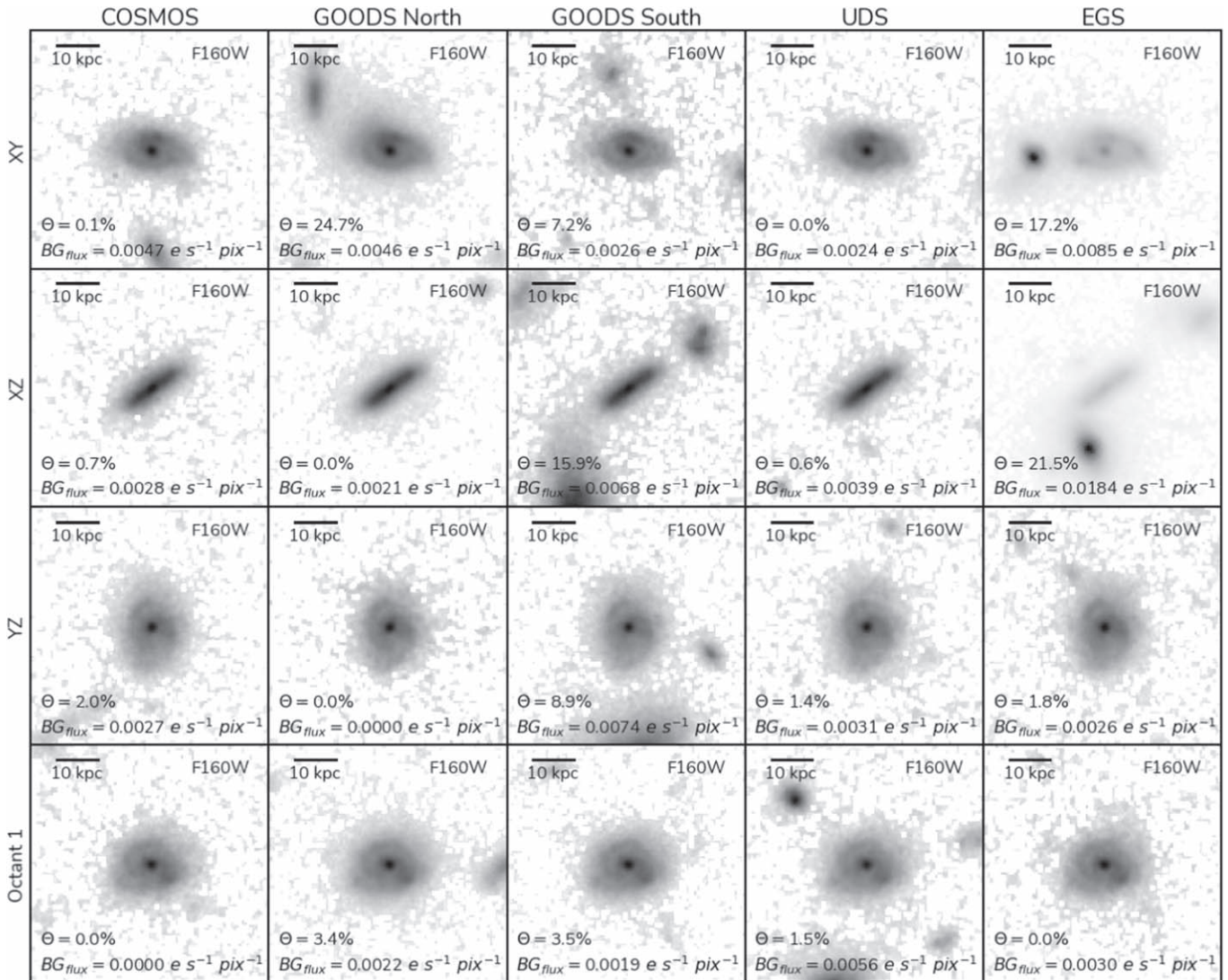
Figure 4. Demonstration of the augmentation pipeline for one random galaxy from TNG100-1 (ID = 192802, $z = 0.55$, at different orientations). We increase our sample by augmenting the data set, reproducing it in four orientations (rows) in each of the CANDELS fields (columns). The simulated galaxy is placed in a random patch of the sky in the CANDELS fields and thus can have other sources in the final cutout. The amount of contamination from neighboring sources varies widely due to the random sampling of the background described in Section 2.4. This contamination is quantified by the overlapping percentage, Θ, and the average flux of the background patch, BG_{flux}.

representation as a new galaxy. We select four different projections, three aligned with the axis of the simulation, *XY*, *XZ*, and *YZ*, respectively, and a fourth line of sight aligned with one octant of the simulation cube.

Second, each CANDELS field has unique observational properties (e.g., different noise levels, and depth). We exploit this aspect and reproduce each of the different orientations from the previous step on top of a random patch of sky of each CANDELS field, taking care to use appropriate noise levels for the simulated galaxy. To find empty patches of sky, we randomly sample the R.A. and decl. within each field, and make a large cutout of the area that is four times larger than the final size of the cutout. Using positions given in the CANDELS catalogs, we then identify all sources within this cutout and reselect a new R.A. and decl. location within the cutout that does not centrally overlap with another source. We allow some degree of overlapping source, but require a unique central position. We do this interactively until a patch of sky that matches all of above

criteria is found. This, combined with all of the orientations, augments our data set 20 times. In addition, this also helps the network to generalize the impact of contamination from neighboring sources, as the same galaxy in one field might be isolated in its cutout, but in a denser environment in another.

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

Overall, our sample increases from ∼8000 examples to ∼160,000. However, having multiples of similar galaxies in our data set can result in overfitting. To reduce this risk, we do not allow different realizations of the same galaxy to fall in both the training sample and the test sample. This ensures that testing and validating are performed on unique data sets.
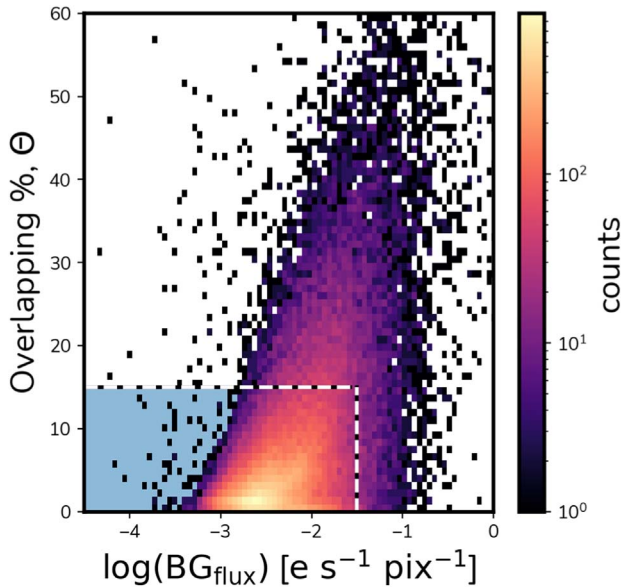
Figure 5. Contamination characterization for 162,000 IllustrisTNG simulated images in our sample. We show the logarithm of the average flux per pixel of the background measured in each cutout, log(BG$_{flux}$) vs. the overlapping percentage, Θ, which indicates how much the central galaxy segmentation map is covered by the segmentation map of the sources in the background. We define a conservative region of low contamination shown by the dashed line and blue area, which contains 90% of the whole sample. Every point represents at least one image.

2.5. Contamination Quantification

Providing realistic levels of contamination and the inclusion of neighboring sources are some of the most important requirements for a good generalization between samples of simulated galaxies and real observations (Bottrell et al. 2019). In an update to what was done in Ferreira et al. (2020), we included realistic contamination in our IllustrisTNG sample, as described in Section 2.4. By comparing clean galaxy realizations to their respective background-added images, we can thus test how our methods behave when faced with a variety of contamination levels, drawing direct conclusions for real world applications. We quantify the degree of contamination in each image using two measurements, which are also listed in Figure 4.

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

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

We use the overlapping percentage, Θ, and the flux of background sources, BG$_{flux}$, to define galaxy images with low contamination. Figure 5 shows the parameter space formed by

these two measurements for the entire sample of $\sim$160,000 simulated and *candelized* images. The blue box framed by the dashed line defines a region of galaxies with low contamination,

$$\Theta < 15\%, \quad BG_{flux} < \times 10^{-1.5} \text{ e s}^{-1} \text{ pix}^{-1}, \quad (6)$$

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

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

3. Methods

We use a deep learning framework with a CNN based on Ferreira et al. (2020) but with significant updates related to the improved and more robust data pipeline that was discussed in Section 2.4. In this section we describe our deep learning analysis (Section 3.1), where we also highlight the improvements to Ferreira et al. (2020). In Section 3.2, we discuss how to avoid overfitting due to the augmentation of the TNG sample, which was part of our sample pipeline. We further wish to compare the resulting classifications to "traditional" classifications. We thus measure nonparametric morphology indices, structural parameters, and Sérsic profiles for both the TNG sample and the CANDELS sample with MORFOME-TRYKA (Ferrari et al. 2015; Albernaz Ferreira & Ferrari 2018; Lucatelli & Ferrari 2019), for which we provide a brief overview in Section 3.3.

3.1. Deep Learning Classifications

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

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

A CNN is an end-to-end method, where the most meaningful spatial features are also learned from the data itself through convolution operations. These features are then combined for a

Table 2
The Best Hyperparameters of Our Architecture Found through Bayesian
Optimization (The GPyOpt 2016)

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

Note. These define the depth, width, and number of trainable parameters of our architecture. This process is done using our set-aside validation samples. The same model is used for all of the CANDELS data sets.

classification task, producing the desired outcome based on the input.

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

1. `number_conv_blocks` define the number of convolutional blocks, each will probe features of different scales;
2. `number_conv_per_block` describe how many convolutions each block will have;
3. `initial_number_filters` define the starting number of filters, which are then doubled after each convolutional block;
4. `initial_kernel_size` is the initial size of the convolutional kernel, which is then reduced by two after each block, down to a minimum of three;
5. `n_fc_layers` and `size_fc_layers` define the number of hidden layers and their respective sizes, respectively;
6. `l2_regularization` and `dropout` are the degrees for each regularization technique used, respectively. l2 regularization is applied to all convolutional layers, where dropout is applied only after the hidden layers.

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

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

Second, we treat the learning rate differently. In Ferreira et al. (2020), we monitored the learning rate decays during training as a hyperparameter. Here, we use cosine annealing, a type of learning rate scheduling (e.g., Loshchilov & Hutter 2016, for an explanation) combined with a regular stochastic gradient descent optimizer (Zhou et al. 2020). This approach probes several different learning rate regimes during training and uses cyclic resets that serve as a way to avoid
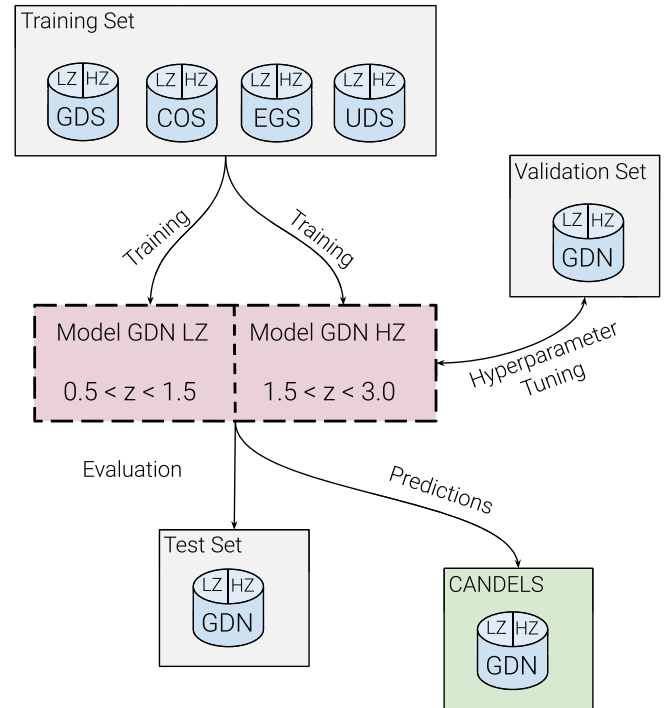

Figure 6. Schematics of the training pipeline leveraging multiple fields for augmentation. Each pair of models, at low redshift (LZ), and high redshift (HZ), is trained only with data that is augmented with the CANDELS fields that are not the target for the model. In this example we show a model designed for predictions on GOODS-North (GDN), trained on data augmented with characteristics of all of the remaining four fields (GDS, COS, EGS, and UDS). This model is also tuned and evaluated in validation and test sets that have only of target CANDELS field augmentations, ensuring that no overfitting of neighboring sources is part of the predictive process.

unstable local minima, improving generalization of the solutions.

All hyperparameters are determined by a Bayesian optimization process (The GPyOpt 2016), and the values for the best model used here are summarized in Table 2. These values can be directly used in conjunction with our public KERAS implementation.

3.2. Augmentations and Overfitting Avoidance

To avoid overfitting pitfalls from using our CANDELS background augmentation pipeline (Section 2.4), we train a suite of models, one for each CANDELS field. Because we have included areas of all of the CANDELS fields as background in our training set, the network could potentially memorize these and use them for predictions, impairing the results. To ensure this is not the case, each CANDELS field has two models—one at low redshift, $0.5 < z < 1.5$, and one at high redshift, $1.5 < z < 3.0$—trained only with images augmented with regions of the other four fields. All data sets (training, validation, and test) are restricted in this way, guaranteeing that any overfitting of the CANDELS background will have no impact on the final application of our models.

An example of this process is outlined in Figure 6, for the models that will be used for predictions in the GOODS-North (GDN) field. The training set contains galaxies augmented with the COSMOS (COS), GOODS-South (GDS), Extended Groth Strip (EGS), and The UltraDeep Survey (UDS) fields while the validation and test sets only contain galaxies from GDN.
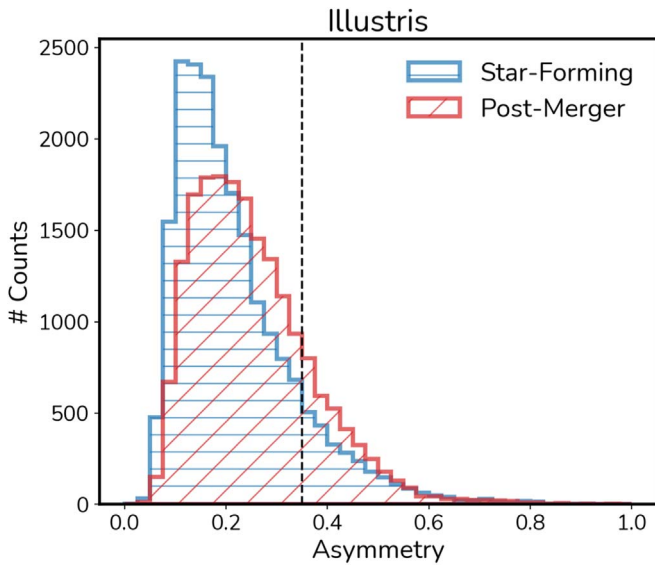
Figure 7. Distribution of asymmetries A measured with MORFOMETRYKA for our TNG100-1 sample of galaxies. Star-forming nonmergers and post-mergers are shown in blue and red, respectively. The dashed vertical line illustrates the typical threshold ($A > 0.35$) used to classify galaxies as mergers.

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

3.3. Galaxy Structure and Morphology

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

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

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

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

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

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

$$(A > 0.1).$$

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

In Figure 7 we show the distribution of asymmetries A measured with MORFOMETRYKA for star-forming galaxies (in blue) and post-mergers (in red) for the simulated galaxies. The distributions largely overlap, though asymmetries for post-mergers are generally slightly higher. The difference between both distributions is small enough that using solely the asymmetry ($A > 0.35$) will produce samples with low completeness and purity, and given that the fraction of merging galaxies is lower than regular star-forming galaxies, it is likely that this approach produces very contaminated samples.

4. Results

Here we discuss what our trained models reveal, first from the test data set of IllustrisTNG selected galaxies (Section 4.1), and then applied to the CANDELS fields (Section 4.2).

4.1. Predictions within IllustrisTNG

We measure the performance of our trained models in our prepared test sets. This is done by training the network with two realizations of the test data sets, one with full HST-matched properties including a CANDELS background patch of the sky (Sections 2.4, 3.2, which we call `realistic mocks`) and one with clean mocks with no sky noise and contamination included (which we call `pristine mocks`). For simplicity, in cases where we only mention the `realistic mocks` without specifying which CANDELS fields it was augmented with, we consider the average of all 20 models described in Section 3.2.

To compare between models and realizations of these data sets, we use traditional performance metrics common for evaluating machine-learning model performance. These consist of receiver operating characteristic (ROC) curves and precision-completeness diagrams (Powers 2011), as well as confusion matrices and their individual indices. Here, we are dealing with a single binary classification task, such that the probabilities of both classes respect the condition $P(\mathrm{NMSF}) + P(\mathrm{PM}) = 1$. Figure 8 displays the overall performance for each network.

The left panel shows four different realizations of the network for comparison purposes. The network is trained twice to generate two different types of models: one labeled `base` that consists of a network trained with the `realistic mocks`, and a second labeled `clean`, which is trained with the `pristine mocks`. Then, each model is applied to both data sets. We do this to measure the best-case scenario within the simulations, in the absence of any contamination or impact from observational effects. Models trained with the `realistic mocks` data set are plotted as solid lines, while models trained with the `pristine mocks` data set are shown by
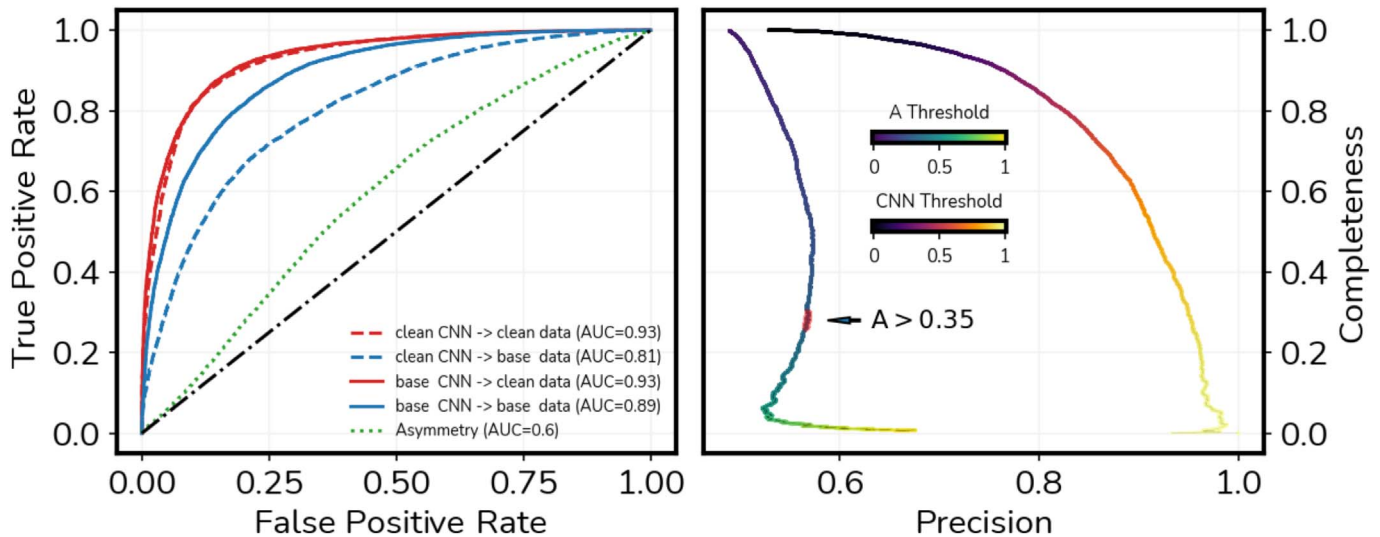
Figure 8. Performance metrics for our four trained models and comparison with the classical asymmetry index *A* for the simulated images. Left: ROC curves for both the network trained with the `pristine mocks` data set (dashed lines) and with the `realistic mocks` data set (solid lines) applied to both data sets, color-coded in red (`pristine`) and blue (`realistic`). The green dotted line indicates the ROC curve for a classifier using only the asymmetry *A*. The area under each curve can be read in the label. Right: precision-completeness diagrams for the baseline network trained with the CANDELS matched mocks with asymmetry *A*, color-coded by classification threshold levels for CNN (`inferno`) and asymmetry (`viridis`). A small region in red is printed over the asymmetry curve to point out the region where the classification threshold is *A* > 0.35.

dashed lines. Furthermore, the color conveys the data set in which the model was applied to, red and blue for `pristine mocks` and `realistic mocks`, respectively. In addition to these, a single parameter classifier based on the asymmetry (*A*) is also evaluated and displayed as the green dotted line. The area under the curve for each case can be found in the legend of the left panel.

The different realizations of our network (`base` and `clean`) cross-correlated with the `realistic mocks` and `pristine mocks` data sets confirm the importance of realistic observational modeling of the mocks (discussed in detail by Bottrell et al. 2019). This is especially important when crossing domains from cosmological simulations to real observations. Figure 8 shows that the `base` network performs just as well as the `clean` network when applied to the `pristine mocks`, resulting in similar performance metrics, as can be seen by the overlapping red curves. However, the `base` network outperforms by ∼10% the `clean` network when applied to the `realistic mocks` data set, as displayed by the difference between the blue curves in Figure 8. This demonstrates that correctly modeled observational features increase the generalization capabilities of the resulting models. A network that is only trained on pristine images will perform poorly in the real observations domain.

Importantly, all cases outperform the asymmetry by 20%–30%. To some extent, this is expected because asymmetries of post-mergers are lower than asymmetries of galaxies that are just in the beginning of their merging event, including cases of closely interacting galaxies. Evidently, the asymmetry function is a much more general morphological descriptor while the network is very specialized for the particular task of dividing post-mergers from star-forming galaxies.

We compare the performance of asymmetry (*A*) and CNN predictions further and show completeness-purity diagrams in the right panel of Figure 8. It displays outcomes for our ensemble of CNN models in the `inferno` color map, and for the classic asymmetry parameter in the `viridis` color map. The commonly used asymmetry value to classify galaxy

mergers is generally higher than (*A* > 0.35), which is shown in the figure by the red patch over the curve. However, here we compare an asymmetry classifier with our neural network to exemplify how one can use the classification threshold of the network as a way to control the trade-off between precision and completeness. This is a useful feature when dealing with unbalanced data sets, like the case for galaxy mergers.

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

Our network is able to correctly identify post-mergers and star-forming galaxies from the IllustrisTNG simulation in ∼80% of the cases. Figure 9 shows the confusion matrix for the `realistic mocks` data set identified within each individual CANDELS field, as well as for the `pristine mocks` sample, where accuracy reaches ∼90%. All classifications are done with the model trained with the `realistic mocks`. We show true positives (TPs) and true negatives (TNs) in blue, and false positives (FPs) and false negatives (FNs) in pink. The CLEAN case represents the best-case scenario, where our current method and data set achieves an even higher performance of ∼91% TPs. A histogram of the redshift distribution for each cell helps to visualize any possible biases in redshift for the misclassification cases. This demonstrates that the models are more likely to correctly classify low-redshift galaxies, as they represent the majority of the samples.

4.1.1. Impact of Redshift

With the goal of applying our models to a wide range of redshifts, we explore how our performance metrics are impacted by increasing redshifts. Following the angular size–distance-relation, galaxies at increasingly larger distances from
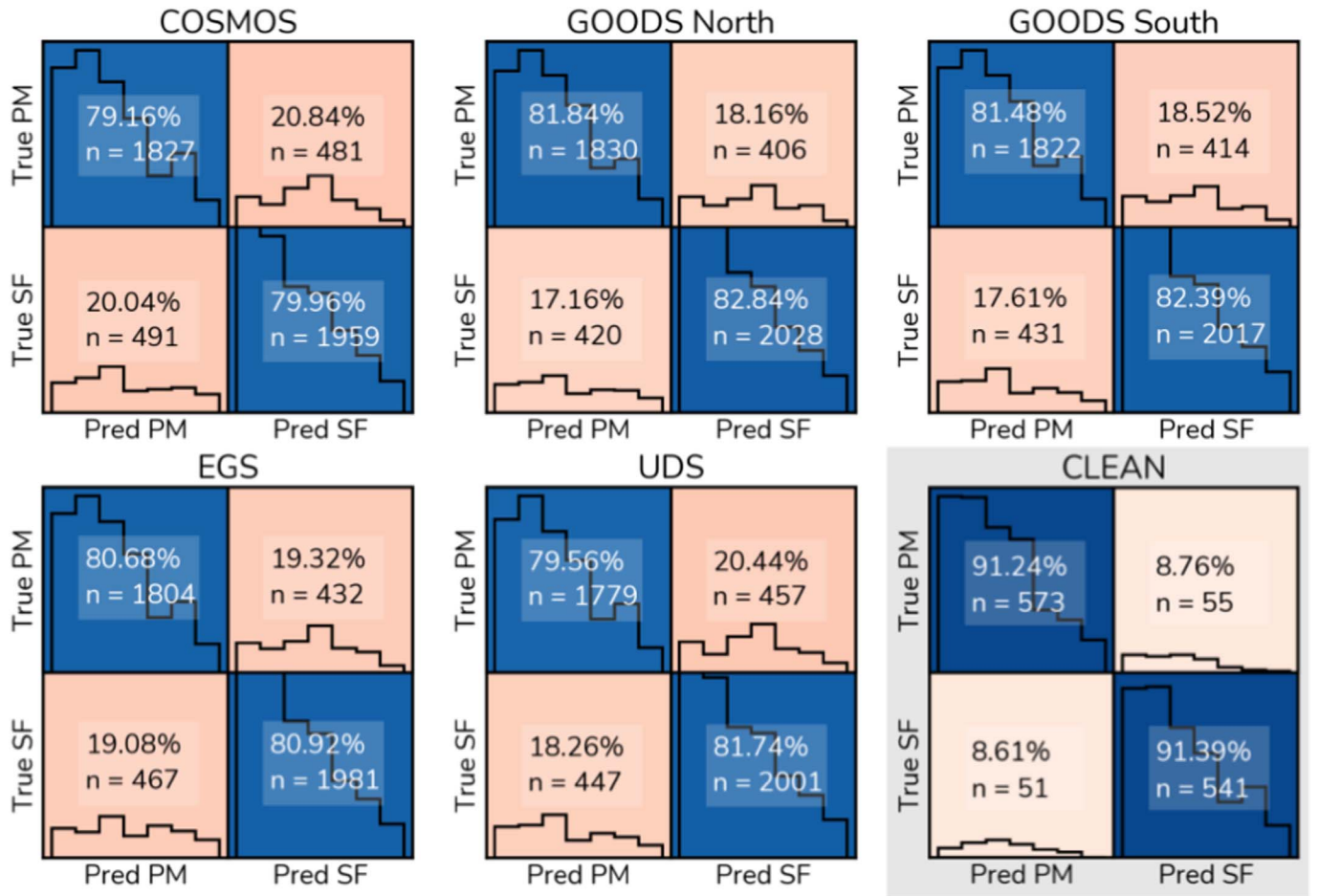
Figure 9. Confusion matrix for all of the samples matched to CANDELS fields as well as the `pristine` sample (highlighted by gray shading in the bottom right). These confusion matrices were evaluated with the ensemble of models trained with the CANDELS matched mocks. We show true negatives (TNs) and true positives (TPs) highlighted in blue while the false negatives (FNs) and false positives (FPs) are shown in pink. The colors are based on the rate percentage, which is also printed in each cell. All of the CANDELS fields have TP and FN rates of around ∼80%. For the `pristine` case, performance can reach as high as ∼90%, marking the intrinsic limit of our method based on the data available. The histograms show the redshift distribution for the galaxies in each category, which demonstrate that it is easier to recover correct classifications at lower redshifts.

low to intermediate redshifts will be greatly impacted by decreasing resolution, which means that morphological features are less well sampled. The right panel of Figure 10 shows this effect on the performance of our models, where the scores of the metrics gradually decrease with increasing redshift, going from 85% accuracy at $z = 0.5$ to around 80% at $z = 2$. The error bars—sampled from bootstrapping our testing samples— follow accordingly.

4.1.2. Contamination Impact on Classification

We use the contamination estimates measured in Section 2.5 to find the contamination failure threshold of our classifier, comparing performance metrics for subsets of the test set selected in bins of both the overlapping percentage, Θ, and the average background flux per pixel, BG_{flux}, as shown in Figure 10. The horizontal black dashed line at 0.8 shows the accuracy of the model when evaluated in the complete test set (80%). The metrics outperform this baseline in subsamples of images with low contamination, decreasing as we increase each of the contamination factors.

As described in Section 2.5, we select the point where the average mean values for each metric fall below the dashed line,

which is our contamination cutoff, i.e.,

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

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

4.2. Classifications on CANDELS

We use our network to carry out predictions in all real CANDELS galaxies at $0.5 < z < 3$, $M_* \geqslant 10^{9.5} M_\odot$, $S/N > 50$, and $H_{MAG} < 24.5$. We filter out regular galaxies using a conservative asymmetry cut of $A > 0.1$ as we are interested only in asymmetric, irregular/peculiar systems. This selection results in a sample of 23,494 galaxies, for which 14,410 have visual classifications from Kartaltepe et al. (2015). Based on the classifications from our networks, we separate these galaxies in post-mergers and noninteracting star-forming galaxies using a threshold probability of 60%. Galaxies with probabilities $50\% < P(PM) \wedge P(SF) < 60\%$ are not considered in any class. These represent 2125 galaxies ($\approx 15\%$) of the sample with visual classifications. Figure 11 showcases some examples of galaxies in the CANDELS fields separated by the
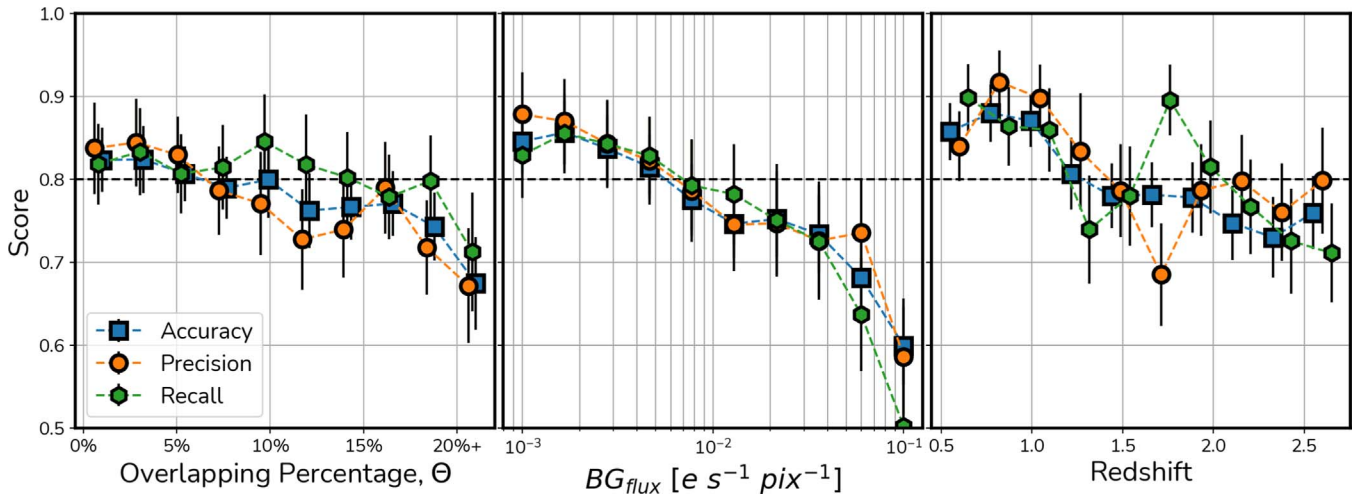
Figure 10. Impact of contamination and redshifts on the performance of our models. The left and central panels show how accuracy, precision, and recall (blue squares, orange circles, and green hexagons, respectively) behave for increasing percentages of overlap (Θ) and for increased background flux (BG_{flux}). In the right panel, we show how the accuracy, precision, and recall of our methods change in bins of $\Delta z = 0.25$ redshift. Error bars are sampled from bootstrapping the test sample. The performance gradually decreases with z, decreasing below 80% beyond $z = 2$. There is a slight uptick at $z = 2.5$, but with large error bars. The cutoff at $z > 2.5$ is the result of a combination of small sample size and redshift effects. The black dashed line at $score = 0.8$ indicates the overall accuracy of the model in the complete test set.

classification of our models. Post-mergers are shown in the left panel, and star-forming galaxies are shown in the right panel.

To investigate how the relative number of post-mergers and star-forming galaxies changes over cosmic time, we divide the CANDELS sample in bins of $\Delta z = 0.25$. Figure 12 shows the change of class fractions change with redshift. We do this analysis in two mass regimes: low-mass galaxies with $9.5 < \log(M_*/M_\odot) < 10.0$ (left panel) and high-mass systems with $\log(M_*/M_\odot) > 10.0$ (right panel). Noninteracting star-forming galaxies are shown by the blue circles while post-mergers are shown in by the red squares. The upper dashed line displays the fraction of galaxies that are not mergers, including the star-forming galaxies and other low-probability cases not included in any class.

For the lower-mass post-mergers, we see an upward trend from ∼15% at $z = 0.5$ to ∼35% at $z = 2$, then a slight decrease beyond $z = 2$. This is still consistent with a ∼35% fraction within the error bars. The star-forming galaxies behave in the opposite way, decreasing from ∼70% at $z = 0.5$ to around ∼55% at $z = 2$. This suggests that among asymmetric galaxies of this mass range, there is an exchange between the classes as we go to higher redshifts up to $z = 2$. This once again emphasizes that classifications of local galaxies that are purely based on the asymmetry (A) are highly contaminated with noninteracting star-forming galaxies. However, this is mitigated at higher redshifts where we find more post-mergers. Nevertheless, samples selected based on A are still dominated by star-forming galaxies, albeit to a lesser extent.

Trends for higher-mass galaxies are substantially different (right panel of Figure 12). While the post-mergers exhibit a similar but steeper upward trend from ∼20% at $z = 0.5$ fraction to ∼50% at $z = 2$, the relative fraction of star-forming galaxies shows a constant value of ∼50% at $0.5 < z < 3$, while the fraction of the rest of the sample (dashed line) goes from ∼75% to ∼50%. This supports the idea that for more massive systems, post-mergers at higher redshifts will eventually become massive passive galaxies with no significant asymmetric features.

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

We know that mergers are more common in the past (e.g., Mundy et al. 2017; Duncan et al. 2019; Ferreira et al. 2020; Whitney et al. 2021), and here we find further evidence that this is also the case for peculiar galaxies, indicating that the nature behind these disturbed morphologies at earlier times can be attributed to merging. To further investigate this, we select all galaxies from Kartaltepe et al. (2015) that are classified as an irregular/peculiar with f_Irr > 0.75, i.e., cases where more than 50% of the visual classifiers agree on the classification, and check how our networks perform on this subset. We observe similar trends with redshift, with the fraction of post-mergers increasing by ∼20% from $z = 0.5$ (∼30%) to $z = 2$ (∼50%), which agrees with the results for the complete sample. Furthermore, our methods classify ∼50% of the galaxies visually classified as potential mergers (f_merger > 0.75) in Kartaltepe et al. (2015) as post-mergers. This is higher than random, but does show the difficulty of obtaining exact matches between mergers determined visually compared with a quantitative process.

4.2.1. Visual Representation of the Classification

As a way to visualize how our networks organize the features extracted from the images to produce the final classification, we generate a 2D representation of the final dense layer of the network corresponding to 128 neurons (128 dimensions) using a UMAP (Figure 13). The color code of the points expresses their respective labels, red for post-mergers and blue for star-forming galaxies. Then, we overplot the positions assigned by the network for unlabeled CANDELS galaxies. We also include some examples of images of CANDELS galaxies close to their original position in this manifold as a way to visualize how the morphologies change
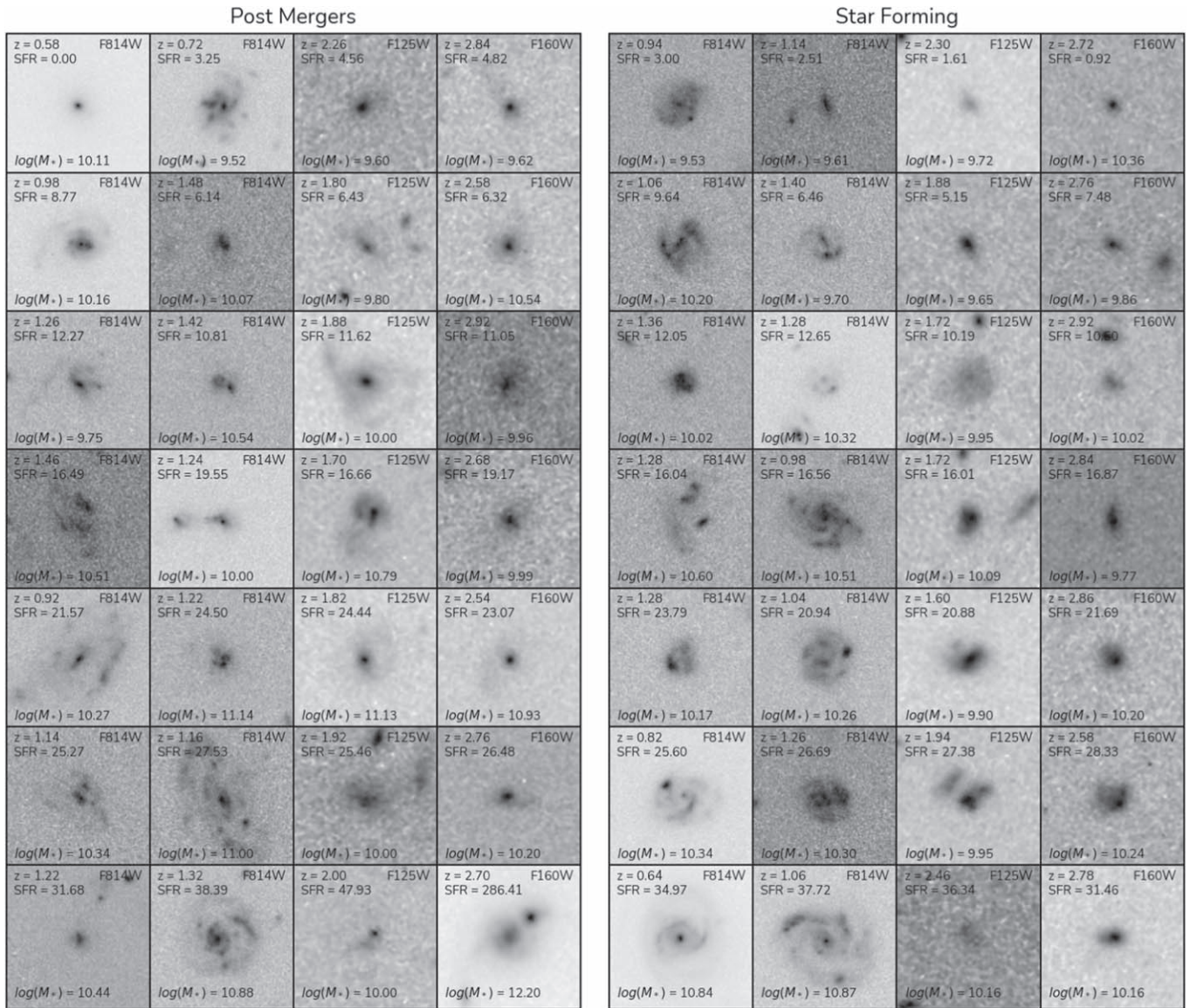
Figure 11. Examples of CANDELS galaxies with $A > 0.1$ classified by our models into post-mergers (left) and star-forming galaxies (right), with their redshifts, SFRs, and stellar masses. Images are ranked from left to right with increasing redshift and top to bottom with increasing SFR. All stamps use a square-root normalization.

with its position. Each region in the parameter space of this diagram is directly related to a probability. The maximum probability is found in the extreme regions farther away from the center, which represents how different these objects are for the network. Images of galaxies that the network struggles to identify are mixed in the bottom middle, representing the region where both probabilities are similar $P(\mathrm{PM}) \sim P(\mathrm{SF})$.

5. Implications

Making use of the classifications from our deep learning models, we first explore the impact of major mergers on classifications above the star-forming main sequence (SFMS) as parameterized by Schreiber et al. (2015; see Section 5.1). We then discuss the structure of the two galaxy classes using Sérsic profile measurements (Section 5.2). In Section 5.3, we update classifications from Ferreira et al. (2020) with our new specialized model, thus increasing certainty for previously undefined classifications. We then add to the discussion

proposed by Bickley et al. (2021) regarding the Bayesian limitations of classifying post-mergers by considering an evolving merger fraction. We finish with Section 5.5, in which we compare extracted features from real CANDELS galaxies to features extracted from IllustrisTNG galaxies, as a way to address the challenges of transferring the model from simulations to real observations.

5.1. Classifications above the Star-forming Main Sequence

The influence of merging on the structure of peculiar/irregular galaxies at intermediate redshifts ($0.5 < z < 3.0$) is directly related to the question of whether merging galaxies can induce more starbursting episodes than galaxies evolving secularly. Enhanced star formation can then lead to more clumpy and asymmetric structures, and thus can impact the morphological appearance of galaxies greatly. By examining the SFMS of galaxies, one can investigate the nature of
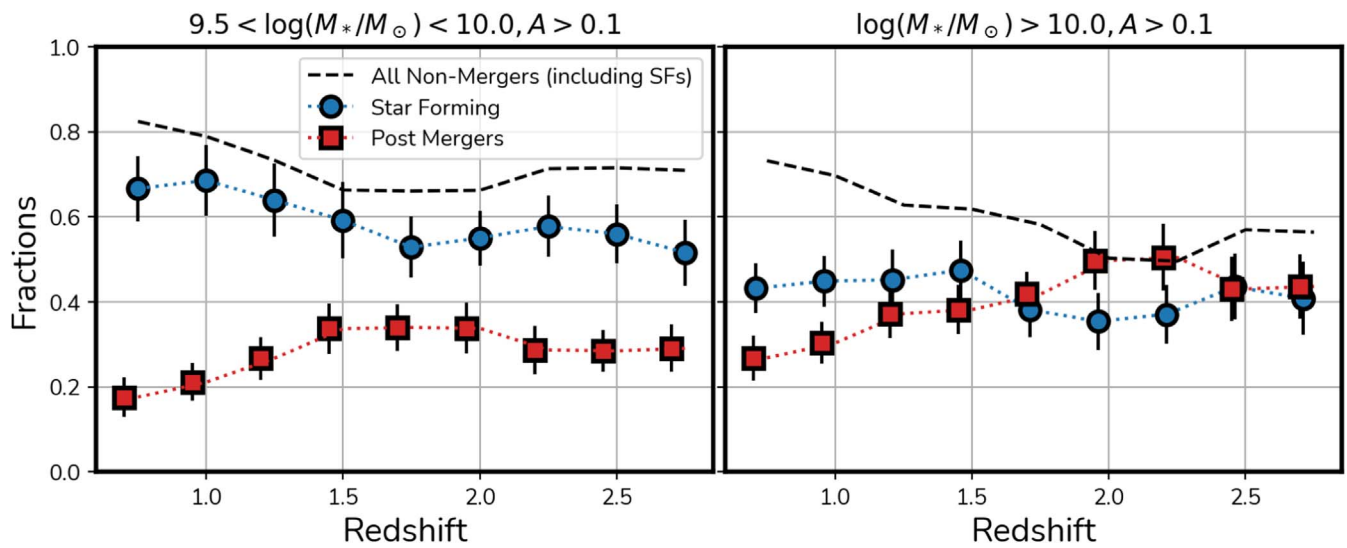
Figure 12. Relative class fractions for post-mergers and star-forming galaxies vs. redshift for real galaxies in the CANDELS fields. The fraction of post-mergers increases from 30% at $z \sim 0.75$ to 50% by $z \sim 2$. Error bars are drawn from bootstrapping the samples and applying the underlying uncertainty associated with the performance of our models, which decreases with redshift.
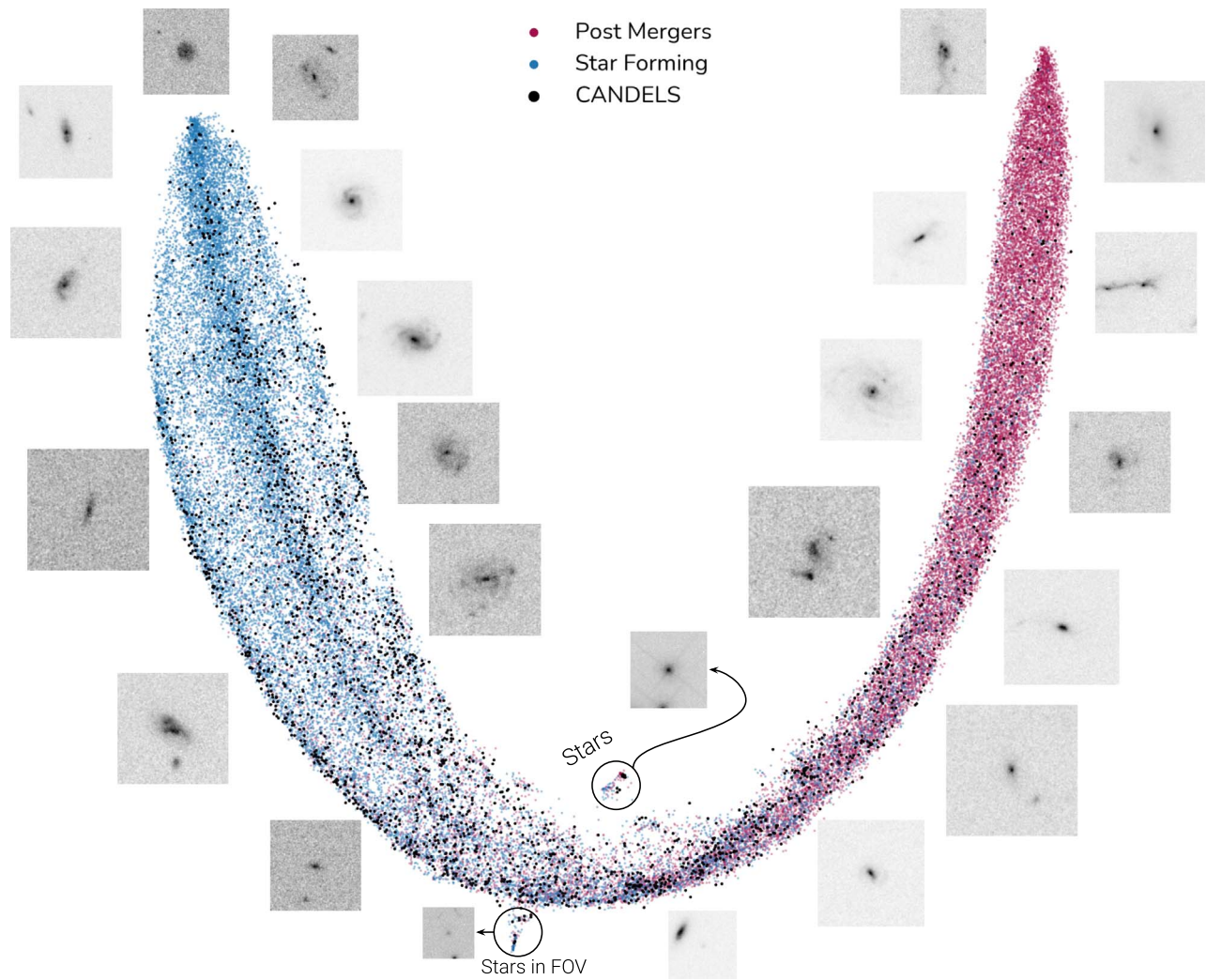

Figure 13. UMAP representation of the output from the last dense layer of the network. This representation shows the parameter space used for the network to generate the final probability. Probabilities are highest in the extremes at the top, and uncertainty increases due to increased contamination as we go along this structure toward the middle. The same random examples of CANDELS galaxies are placed close to their points in this manifold. Small regions identified by circles show the clustering of nongalactic detections in this parameter space, located close to the region of uncertain classifications at the bottom of the UMAP. "Stars" are stars in the center, and "Stars in FOV" correspond to stars at the edge of the stamps.
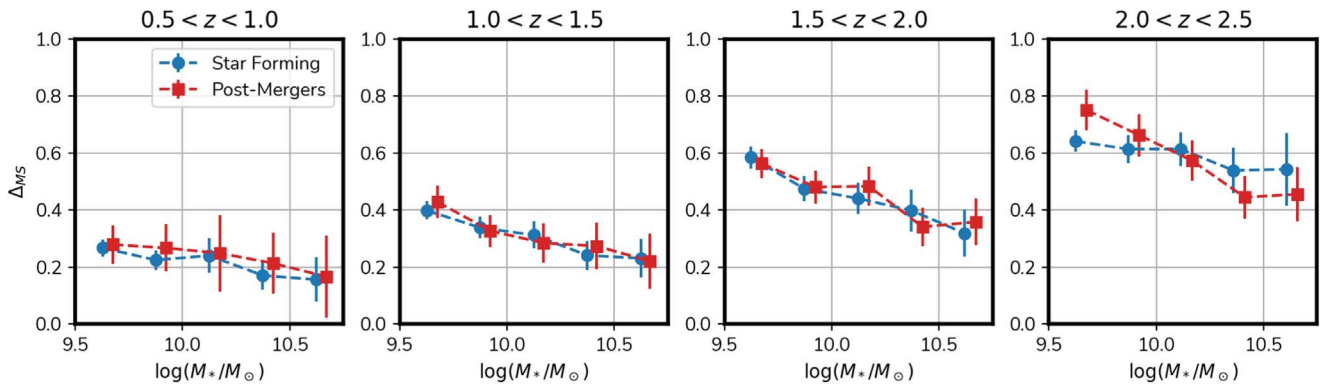
Figure 14. Mean distance to the SFMS (Δ_{MS}) vs. log stellar mass in bins of redshift for CANDELS galaxies above the SFMS. Post-mergers are plotted as red squares, star-forming galaxies as blue circles. Error bars are estimated using bootstrapping and show $\pm 1\sigma$. The classes are indistinguishable between $0.5 < z < 2$, both increase similarly in Δ_{MS} as we increase redshift. This represents the increase in scatter above the main sequence. All galaxies included in this diagram lie above the SFMS, as we are only interested in exploring the scattering above the SFMS. For the last redshift bin ($2 < z < 2.5$), there is a significant difference between the two classes both at the low-mass end and at the high-mass end. At low masses ($\log(M_*/M_\odot) < 10.0$), post-mergers scatter higher than star-forming galaxies with a difference of $\Delta_{MS} \sim 0.1$ dex. At high masses ($\log(M_*/M_\odot) > 10.0$), the trend reverses and star-forming galaxies scatter higher with a Δ_{MS} difference of ~ 0.1 dex. However, care needs to be taken in the interpretation of this trend as it could be spurious or insignificant given the error bars and the performance metrics of our models at high redshift (Figure 10).

galaxies with unusually high SFRs and the formation path that resulted in this physical effect.

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

$$\Delta_{MS} = \log(SFR) - \log(SFR_{MS}), \qquad (7)$$

where $\log(SFR)$ is the log star formation rate of a particular galaxy, and $\log(SFR_{MS})$ is the parameterization from Schreiber et al. (2015). The SFRs and stellar masses used here for CANDELS galaxies were compiled by Duncan et al. (2019) through SED fitting. We refer the reader to this publication for further details. In Figure 14 we show the mean value of each stellar mass bin, for four redshift ranges (one in each panel), separated into star-forming and post-merger galaxies by our classifications. For the $0.5 < z < 2.5$ redshift range (panels A, B, and C), we do not find any impactful difference between the classes and Δ_{MS}, with all offsets well within the error bars. However, for redshifts $2.5 < z < 3.0$, post-mergers with $\log(M_*/M_\odot) < 10.0$ are on average ~ 0.1 dex higher than star-forming galaxies of the same mass. The opposite is found for $\log(M_*/M_\odot) > 10.0$; however, uncertainty is higher here. Additionally, Δ_{MS} increases with redshift in all cases, which describes a larger scatter above the SFMS. However, given the performance metrics of our models at high redshift (Figure 10), we can not claim that this is a real effect. We stress that in Figure 14 we only select galaxies above the SFMS, which is why the distance is always positive.

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

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

5.2. Structure and Light Profiles

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

Figure 15 shows the distribution of Sérsic indices for post-mergers in red, and star-forming galaxies in blue. In general, each class presents very distinct distributions: the post-mergers have a mean Sérsic index $n \sim 1.8^{+0.7}_{-0.6}$ roughly representative of a transition from disks to spheroids; star-forming galaxies have systematically lower Sérsic indexes with $n \sim 1.1^{+0.5}_{-0.5}$, which is more consistent with disk-dominated galaxies. This offset of ~ 1 dex increases for classification thresholds at higher values. The average Sérsic profile (n) of post-mergers increases while the distribution for star-forming galaxies continues with a similar shape. This is quantitative evidence that (1) post-mergers with higher light concentrations are more easily separable from noninteracting star-forming galaxies, and (2) these types of galaxies are intrinsically different from each other.

5.3. Merger Fractions and Rates

By using the new classifications from this work, we can update classifications from Ferreira et al. (2020) for cases where the previous method had ambiguous probabilities for some major mergers and nonmergers.
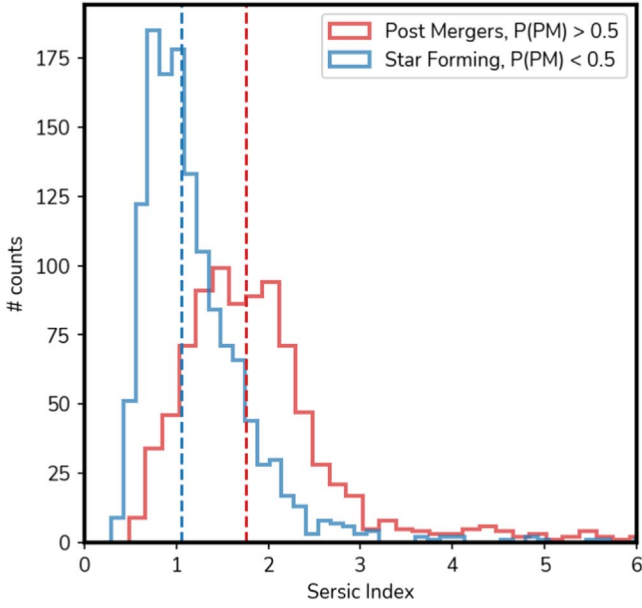
Figure 15. Sérsic index distribution for post-mergers and star-forming galaxies, in red and blue, respectively. Post-mergers display more concentrated light distributions with $n \sim 1.8^{+0.7}_{-0.6}$ while the star-forming galaxies have $n \sim 1.1^{+0.5}_{-0.5}$ consistent with disk-dominated galaxies.

Our new data set accounts for the effects of dust; it is not limited by orientation and probes the rest-frame optical. Thus we can check if any major-merger classifications in the previous work can be attributed to noninteracting star-forming galaxies or if any nonmergers can be reclassified as post-mergers. This is done by comparing the probabilities for major mergers and nonmergers, $P(MM)$ and $P(NM)$, respectively, from Ferreira et al. (2020) to the new probabilities $P(PM)$ and $P(SF)$. We update a nonmerger classification to post-merger if

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

and update the major-merger classifications to nonmerger if

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

In other words, we reclassify galaxies from the previous sample where our new method is more certain about its classification than the previous one. This leads to ~5% of major mergers reclassified as star-forming nonmergers, which lowers the overall merger fractions at lower redshifts and keeps it similar at higher redshifts. In Figure 16 we compare the new merger fraction measurements, in green, to the results from Ferreira et al. (2020), in gray.

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

$$f_m(z) = 0.011 \pm 0.002 \times (1 + z)^{2.71 \pm 0.31}, \qquad (8)$$

with errors estimated with bootstrapping, agrees with the previous measurement in Ferreira et al. (2020) within errors. To measure the galaxy major-merger rate ($\mathcal{R}$), we combine the timescale ($\tau_m = 0.5$ Gyr) used in our selection (Section 2.1) with this merger fraction through
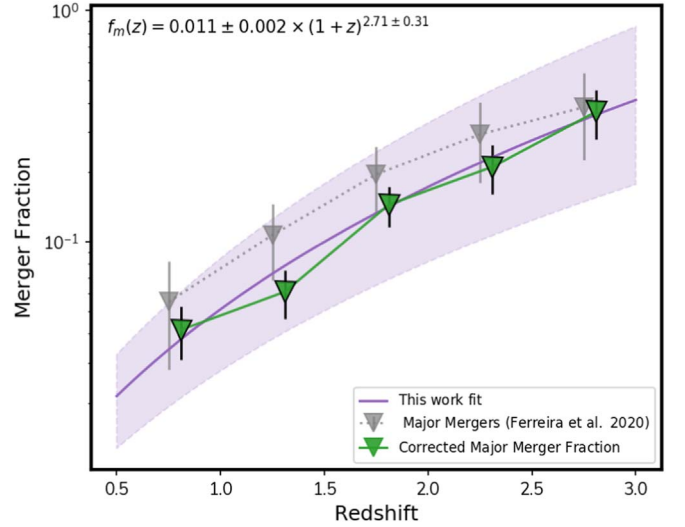
$$\mathcal{R} = \frac{f_m}{\tau_m}. \qquad (9)$$

Figure 16. Major-merger fractions as a function of redshift. We show corrected merger fractions from Ferreira et al. (2020) by reclassifying galaxies with our new method in mergers and nonmergers, shown in green. The original estimates are shown in gray.

The updated galaxy major-merger rate is

$$\mathcal{R} = 0.022 \pm 0.006 \times (1 + z)^{2.71 \pm 0.31}. \qquad (10)$$

We emphasize that this correction is a minor adjustment to the galaxy major-merger rates presented in Ferreira et al. (2020), which remain broadly consistent with each other.

5.4. Bayesian Analysis of Mergers

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

$$P(M|S) = P(S|M) \times \frac{P(M)}{P(S)}, \qquad (11)$$

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

$$P(S) = P(S|M) \times P(M) + P(S|NM) \times P(NM) \qquad (12)$$

where NM standard for nonmergers, where $P(S|NM)$ is the probability of identifying correctly a nonmerger, and the value of $P(NM)$ is the probability that the galaxy is not a merger. We can simplify this if we know, a priori, what the merger fraction is based on previous work. If we denote the merger fraction as f_m, and the machine-learning probability of finding a merger/nonmerger as p_m and p_{mn}, respectively, then we can rewrite

Equation (11), as

$$P(M|S) = \frac{p_{\mathrm{m}} \times f_{\mathrm{m}}}{(p_{\mathrm{m}} f_{\mathrm{m}} + (1 - p_{\mathrm{nm}})(1 - f_m))}. \quad (13)$$

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

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

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

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

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

From this we can conclude that a significant fraction of individual galaxies within the CANDELS imaging may be incorrectly identified as either mergers or nonmergers. From our results here, our method effectiveness for correctly classifying mergers increases from $\sim 40\%$ at $z \sim 0.5$ to $\sim 70\%$ at $z \sim 3$. This is likely what can account for some of our misidentified galaxies as discussed in Section 4 when discussing the success of our method of separating star-forming systems from those that are undergoing mergers.

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

5.5. On Domain Adaptation Issues

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

In our case, the source domain comprises the IllustrisTNG galaxies and the target domain comprises the CANDELS observations. To check if we need to apply domain adaptation techniques to this particular problem, we used uniform manifold approximation and projections (UMAPs; as described in Section 4.2.1; McInnes et al. 2018) to reduce the high dimensional space generated by the features extracted by our network to a 2D-space that is easy to visualize.[5] Then, for each of our trained models, we compare whether the features extracted by the network show similar distributions for Illustris and CANDELS galaxies. In Figure 17 we show UMAPs for each of the CANDELS field models, for low redshift (left) and high redshift (right), color-coded by their class in the case of Illustris and in black for real CANDELS galaxy images. As can be seen, these distributions of simulated galaxies and real observations are clustered together, with very few outliers not following the main cluster. Additionally, we can see that each class forms its own cluster, with overlapping regions, showing that features between classes are distinct and in general not domain specific.[6]

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

6. Summary

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

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

We produced a data set of $\sim 160,000$ images of simulated IllustrisTNG galaxies with realistic visual properties that mimic CANDELS observations in the redshift range $0.5 < z < 3.0$. The images are used to train deep CNNs to distinguish between formation histories of post-mergers and star-forming galaxies. The main conclusions drawn from this work are summarized as follows:

1. The classifier network combined with our new data set produces classification models with a balanced performance of $\sim 80\%$ accuracy, precision, and completeness when applied to a single-band imaging data set, outperforming the asymmetry (A) by at least 25% within the simulated data. Additionally, for pristine images without any contamination and observational effects, the theoretical limit of our model is $\sim 91\%$ accuracy. This is evidence that using the asymmetry (A) alone for ambiguous morphological cases might generate highly contaminated samples.

2. We define two new contamination indicators, the overlapping percentage, Θ, and the average flux of the

---

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

[6] We also tested generating random noise images to check their position in this parameter space. As expected, they cluster away from the image regions, forming their own outlier region, which is far from the main locus where galaxies are found.
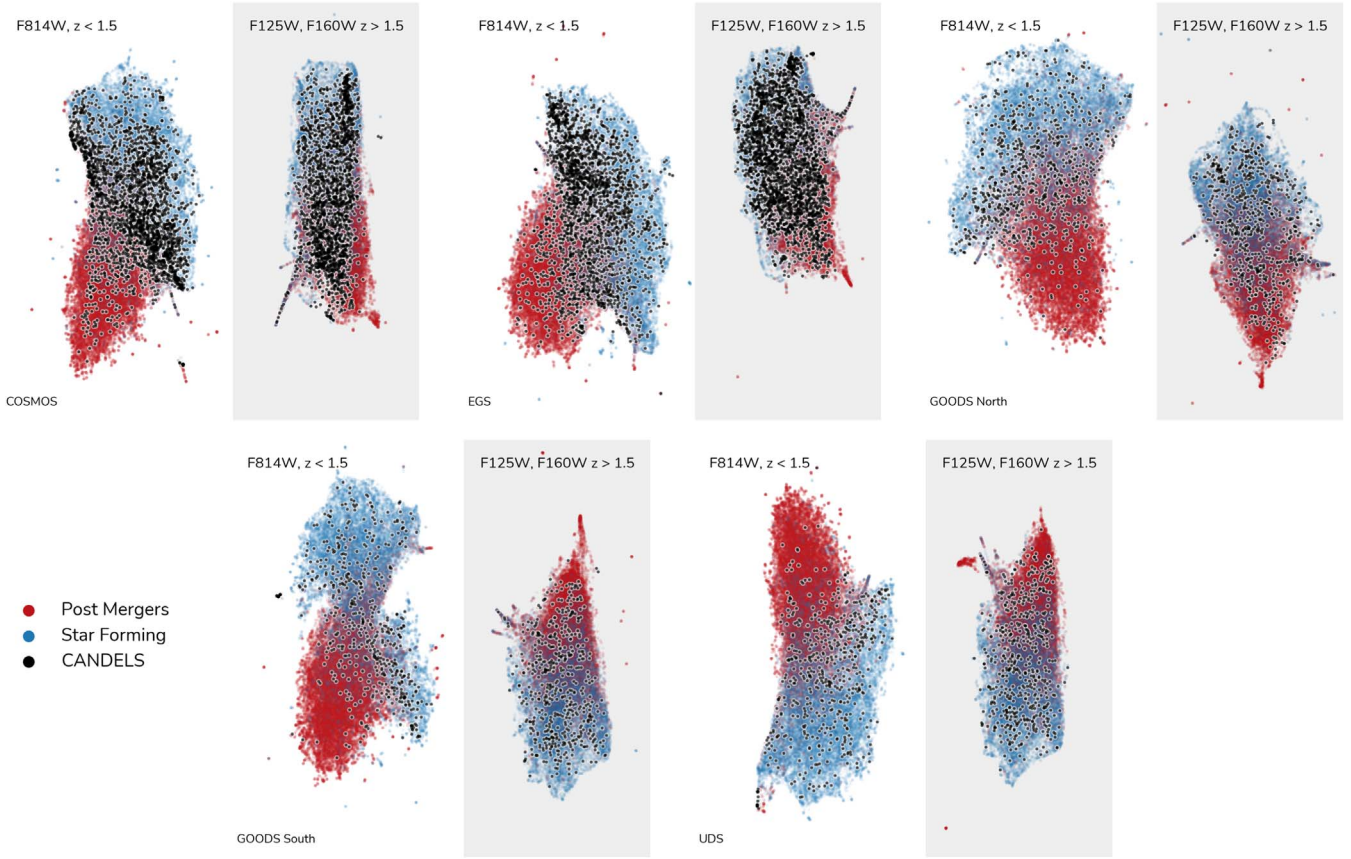
Figure 17. Extracted features by our networks in a UMAP 2D representation. For each model in our ensemble, we generate a UMAP from the extracted features from the last convolutional layer of the trained networks, both when applied to the Illustris galaxies, color-coded by the class, and to the unlabeled CANDELS galaxies shown in black dots. Both Illustris and CANDELS extracted features populate the same region of this representation, showing that the features used by the network to then perform the classification task are in general domain invariant. Additionally, both classes—post-mergers and star-forming galaxies—form separated clusters with some overlapping. Classification could be done in this representation alone, but it is then better organized by the fully connected layers that combine these features to produce the final output probability.

background sources, BG_{flux}, by leveraging how simulated galaxies are combined with true CANDELS background sky patches. Θ controls how sources overlap and are projected in the same stamp, while the BG_{flux} value probes the effect of the brightness of external sources on the classification of the central object. These allow us to explore in detail how deep learning classifications are impacted by contamination. We show that both crowded environments and projections and the relative brightness of external sources to the central galaxy negatively impacted our results. Based on this, we define quality control limits to our approach within the CANDELS fields as $\Theta \sim 10\%$ and $BG_{flux} < 10^{-3}$ e s^{-1} pix^{-1}. Although not universal, these limits provide guidelines for sample selection when applying our models to data.

3. By applying our model to real CANDELS observations of galaxies with high asymmetries, we show that the relative fraction of post-mergers to star-forming galaxies increases with higher redshift for two mass regimes. For low-mass sources ($9.5 < \log(M_*/M_\odot) < 10.0$), the post-merger fraction increases by $\sim20\%$ within $0.5 < z < 2.0$, while the fraction of star-forming galaxies decreases by $\sim15\%$ in the same redshift range. In the high-mass case ($\log(M_*/M_\odot) > 10.0$), the post-merger fraction increases by $\sim25\%$ at $0.5 < z < 2.0$, while the fraction of star-forming galaxies stays broadly constant.

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

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

6. By using our updated data pipeline and models specifically tailored to distinguish between post-mergers and star-forming galaxies, we revisit the merger fractions and merger rates from Ferreira et al. (2020) by correcting ambiguous cases. This leads to updated galaxy merger rates that are slightly lower, but consistent with previously reported rates: $\mathcal{R} = 0.022 \pm 0.006 \times (1 + z)^{2.71 \pm 0.31}$.

7. We show that our models use similar features to classify IllustrisTNG and real CANDELS galaxies, with no clear discrepancy between the two domains. Using the features extracted by the convolutional layers of our network, we generate UMAPs, which visualize the complex parameter space in two dimensions. Features of IllustrisTNG galaxies and CANDELS galaxies overlap for all of the CANDELS fields. Although the CANDELS galaxies do not span the entire feature space of the IllustrisTNG galaxies used here, they are contained within that feature space.

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

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

Data Availability

Ready-to-use models are publicly available for anyone to download.[7] The post-processed IllustrisTNG data used for training is stored in large `TFRecords` binary files and are available upon request.

Appendix

Contamination Network

Based on the contamination measurements described in Section 2.5, we devise a new neural network with the goal of predicting the overlapping percentage (Θ) and the background flux (BG_{flux}) measurements from real observations. In this way, contamination thresholds can be applied to real observational samples in a similar way to what is done in the simulations.

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

We use all of the contamination information from our data pipeline (Section 2.4) to train a neural network to predict these values from the final image, without separating source and background. We use the same network architecture described in this work, but we replace the final sigmoid layer with a linear activation function, and change the loss function as well. The result is a model that can be directly applied to real observations, where the image is the input and the outputs are values for Θ and BG_{flux}.

Figure 18 displays the performance of these predictions based on the original measurements, together with Pearson and Spearman correlation indices. In general, the performance of the model is in good agreement with the original measurements, with rms errors on the order of $\sim 10^{-3}$ for BG_{flux} and $\sim 5\%$ for Θ. These limits are well within the region of the parameter space formed by these indices that we defined as a low-contamination region. Apart from the small bias making the predictions undervalue the truth values, the performance is good enough to separate high-contamination cases from the rest of the sample, which is ultimately our goal. In Figure 19 we show examples of different combinations of Θ and BG_{flux}.

Even though this network is designed to be used within the context of this work as a way to reproduce contamination quantification in the same manner as what was done with the simulated images, we recognize that this can be useful for a wider application. For example, this can be used as a fast selection tool that can remove catastrophically bad cases from big samples in just a couple of seconds; thus, it can be a powerful tool for quick exploration. In this regard, we release this model independent of the classification models presented in Section 3.1.
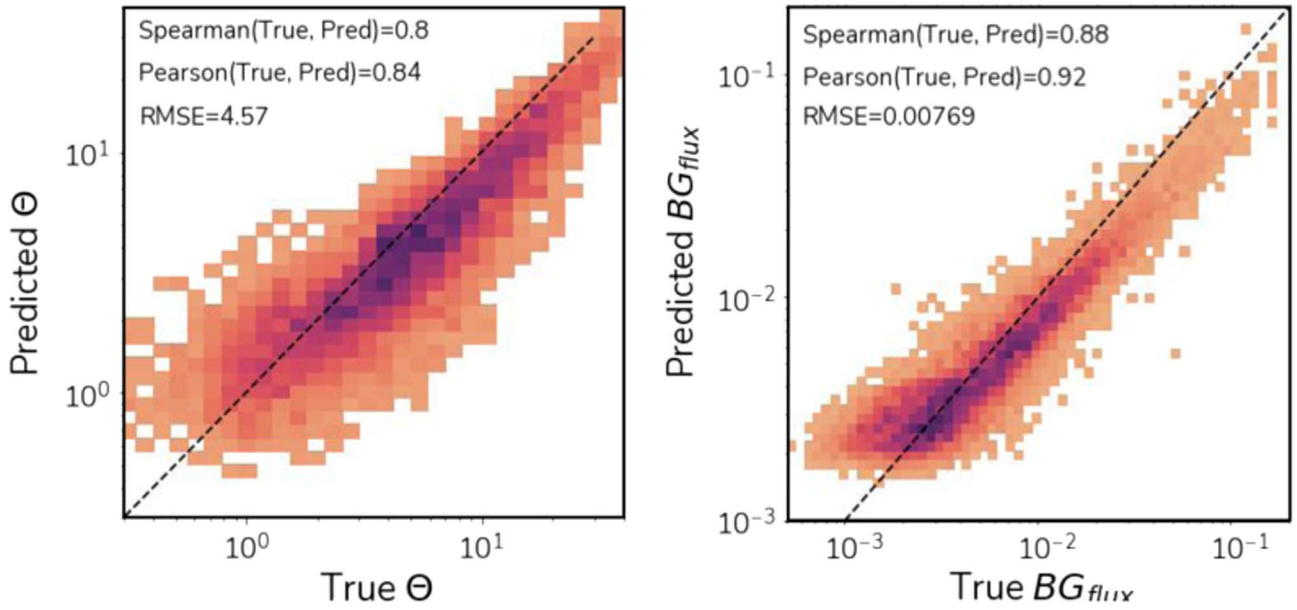
---

[7] https://github.com/astroferreira/FERREIRA2022

Figure 18. Performance of the contamination quantification network. (Top) The relationship between true and predicted values for BG_{flux}, and (bottom) the relationship between truth and predicted values for Θ. Pearson and Spearman correlation indices are displayed for each case, as well as the rms error.
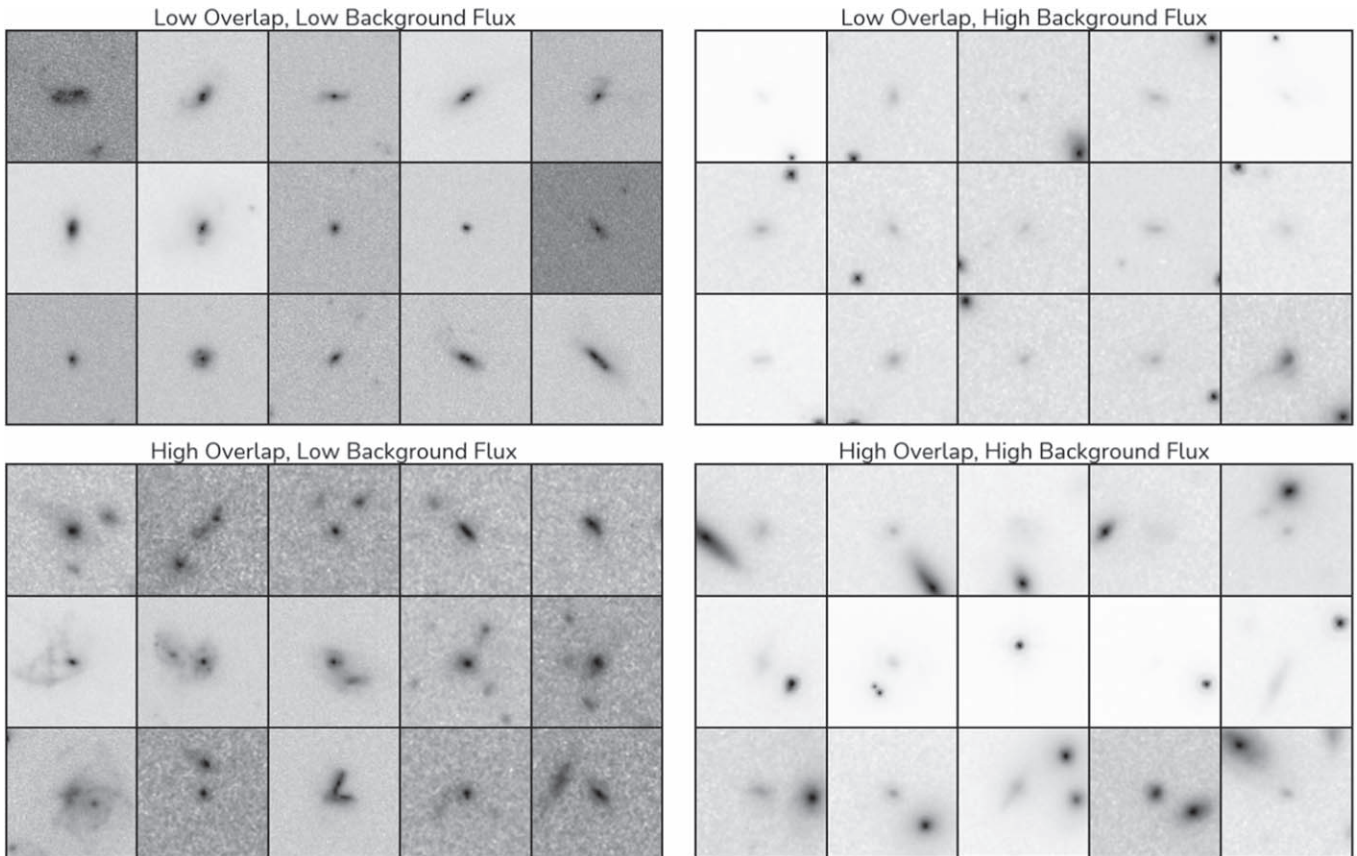

Figure 19. Panel with cutouts of IllustrisTNG galaxies demonstrating four different selections on the O_f and $\langle f_{BG} \rangle$ parameter space. Isolated galaxies with almost no noticeable contamination have low overlap and low background flux (top left). Low overlap and high background flux show cases where the central galaxy is overshadowed by a bright companion, but with no overlap (top right). High overlap and low background flux show galaxies overlapping with similar brightness, cases where projection effects can be misinterpreted as a major merger (bottom left). High overlap and high background flux show central galaxies with very large and bright companions that extend over its segmentation map (bottom right). This illustrates how useful these two measurements can be for proper selections.

ORCID iDs

Leonardo Ferreira ⓘ https://orcid.org/0000-0002-8919-079X
Christopher J. Conselice ⓘ https://orcid.org/0000-0003-1949-7638
Ulrike Kuchner ⓘ https://orcid.org/0000-0002-0035-5202
Clár-Bríd Tohill ⓘ https://orcid.org/0000-0003-2527-0819

References

Ackermann, S., Schawinski, K., Zhang, C., Weigel, A. K., & Dennis Turp, M. 2018, MNRAS, 479, 415
Albernaz Ferreira, L., & Ferrari, F. 2018, MNRAS, 473, 2701
Allen, M. G., Groves, B. A., Dopita, M. A., Sutherland, R. S., & Kewley, L. J. 2008, ApJS, 178, 20
Bertone, S., & Conselice, C. J. 2009, MNRAS, 396, 2345
Bickley, R. W., Bottrell, C., Hani, M. H., et al. 2021, MNRAS, 504, 372
Blumenthal, K. A., Moreno, J., Barnes, J. E., et al. 2020, MNRAS, 492, 2075
Bottrell, C., Hani, M. H., Teimoorinia, H., et al. 2019, MNRAS, 490, 5390
Bottrell, C., Hani, M. H., Teimoorinia, H., Patton, D. R., & Ellison, S. L. 2022, MNRAS, 511, 100
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Camps, P., & Baes, M. 2015, A&C, 9, 20
Camps, P., & Baes, M. 2020, A&C, 31, 100381
Camps, P., Trayford, J. W., Baes, M., et al. 2016, MNRAS, 462, 1057
Cheng, T.-y., Conselice, C. J., Ara, A., & Li, N. 2020, MNRAS, 493, 4209
Ćiprijanović, A., Kafkes, A., Downey, D., et al. 2021, MNRAS, 506, 677
Ćiprijanović, A., Snyder, G. F., Nord, B., & Peek, J. E. G. 2020, A&C, 32, 100390
Conselice, C. J. 2003, ApJS, 147, 1
Conselice, C. J. 2006, MNRAS, 373, 1389
Conselice, C. J. 2014, ARA&A, 52, 291
Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, AJ, 126, 1183
Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, ApJ, 529, 886
Conselice, C. J., Blackburne, J. A., & Papovich, C. 2005, ApJ, 620, 564
Conselice, C. J., Bundy, K., Trujillo, I., et al. 2007, MNRAS, 381, 962
Conselice, C. J., Rajgor, S., & Myers, R. 2008, MNRAS, 386, 909
Dickinson, M., Hanley, C., Elston, R., et al. 2000, ApJ, 531, 624
Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, MNRAS, 444, 2960
Duncan, K., Conselice, C. J., Mundy, C., et al. 2019, ApJ, 876, 110
Duncan, K. J., Brown, M. J. I., Williams, W. L., et al. 2018a, MNRAS, 473, 2655
Duncan, K. J., Jarvis, M. J., Brown, M. J. I., & Röttgering, H. J. A. 2018b, MNRAS, 477, 5177
Ferrari, F., de Carvalho, R. R., & Trevisan, M. 2015, ApJ, 814, 55
Ferreira, L., Conselice, C. J., Duncan, K., et al. 2020, ApJ, 895, 115
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, ApJ, 800, 39
Guo, Y., Rafelski, M., Bell, E. F., et al. 2018, ApJ, 853, 108
Hani, M. H., Gosain, H., Ellison, S. L., Patton, D. R., & Torrey, P. 2020, MNRAS, 493, 3716
Hogg, D. W. 1999, arXiv:astro-ph/9905116
Huertas-Company, M., Bernardi, M., Pérez-González, P. G., et al. 2016, MNRAS, 462, 4495
Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, ApJS, 221, 8
Huertas-Company, M., Primack, J. R., Dekel, A., et al. 2018, ApJ, 858, 114
Huertas-Company, M., Rodriguez-Gomez, V., Nelson, D., et al. 2019, MNRAS, 489, 1859
Huertas-Company, M., Guo, Y., Ginzburg, O., et al. 2020, MNRAS, 499, 814
Jogee, S., Miller, S. H., Penner, K., et al. 2009, ApJ, 697, 1971
Kartaltepe, J. S., Mozena, M., Kocevski, D., et al. 2015, ApJS, 221, 11
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Loshchilov, I., & Hutter, F. 2016, arXiv:1608.03983
Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2008, MNRAS, 391, 1137
Lotz, J. M., Primack, J., & Madau, P. 2004, AJ, 128, 163
Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. 2017, arXiv:1709.02540
Lucatelli, G., & Ferrari, F. 2019, MNRAS, 489, 1161
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Mager, V. A., Conselice, C. J., Seibert, M., et al. 2018, ApJ, 864, 123
Man, A. W. S., Zirm, A. W., & Toft, S. 2016, ApJ, 830, 89
Mantha, K. B., McIntosh, D. H., Brennan, R., et al. 2018, MNRAS, 475, 1549
Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, MNRAS, 480, 5113
Martin, G., Kaviraj, S., Hocking, A., Read, S. C., & Geach, J. E. 2020, MNRAS, 491, 1408
McInnes, L., Healy, J., & Melville, J. 2018, arXiv:1802.03426
Mortlock, A., Conselice, C. J., Hartley, W. G., et al. 2013, MNRAS, 433, 1185
Mundy, C. J., Conselice, C. J., Duncan, K. J., et al. 2017, MNRAS, 470, 3507
Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, MNRAS, 477, 1206
Nelson, D., Pillepich, A., Springel, V., et al. 2018, MNRAS, 475, 624
Nelson, D., Springel, V., Pillepich, A., et al. 2019, ComAC, 6, 2
Papovich, C., Dickinson, M., Giavalisco, M., Conselice, C. J., & Ferguson, H. C. 2005, ApJ, 631, 101
Pearson, W. J., Wang, L., Alpaslan, M., et al. 2019a, A&A, 631, A51
Pearson, W. J., Wang, L., Trayford, J. W., Petrillo, C. E., & van der Tak, F. F. S. 2019b, A&A, 626, A49
Pillepich, A., Nelson, D., Hernquist, L., et al. 2018a, MNRAS, 475, 648
Pillepich, A., Nelson, D., Springel, V., et al. 2019, MNRAS, 490, 3196
Pillepich, A., Nelson, D., Springel, V., Nelson, D., et al. 2018b, MNRAS, 473, 4077
Powers, D. M. W. 2011, J. Mach. Learn. Technol., 2, 37
Reiman, D. M., & Göhre, B. E. 2019, MNRAS, 485, 2617
Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, MNRAS, 449, 49
Rodriguez-Gomez, V., Snyder, G. F., Lotz, J. M., et al. 2019, MNRAS, 483, 4140
Sazonova, E., Alatalo, K., Rowlands, K., et al. 2021, ApJ, 919, 134
Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, A&A, 575, A74
Simons, R. C., Kassin, S. A., Snyder, G. F., et al. 2019, ApJ, 874, 59
Snyder, G. F., Lotz, J. M., Rodriguez-Gomez, V., et al. 2017, MNRAS, 468, 207
Springel, V., Pakmor, R., Pillepich, A., et al. 2018, MNRAS, 475, 676
Tan, M., & Le, Q. V. 2019, arXiv:1905.11946
The GPyOpt, A. 2016, GPyOpt: A Bayesian Optimization Framework in Python, http://github.com/SheffieldML/GPyOpt
Tohill, C., Ferreira, L., Conselice, C. J., Bamford, S. P., & Ferrari, F. 2021, ApJ, 916, 4
Trayford, J. W., Camps, P., Theuns, T., et al. 2017, MNRAS, 470, 771
Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518
Vogelsberger, M., Nelson, D., Pillepich, A., et al. 2020, MNRAS, 492, 5167
Walmsley, M., Lintott, C., Géron, T., et al. 2022, MNRAS, 509, 3966
Walmsley, M., Smith, L., Lintott, C., et al. 2020, MNRAS, 491, 1554
Wang, L., Pearson, W. J., & Rodriguez-Gomez, V. 2020, A&A, 644, A87
Whitney, A., Ferreira, L., Conselice, C. J., & Duncan, K. 2021, ApJ, 919, 139
Williams, R. E., Blacker, B., Dickinson, M., et al. 1996, AJ, 112, 1335
Windhorst, R. A., Taylor, V. A., Jansen, R. A., et al. 2002, ApJS, 143, 113
Zanisi, L., Huertas-Company, M., Lanusse, F., et al. 2021, MNRAS, 501, 4359
Zhou, P., Feng, J., Ma, C., et al. 2020, arXiv:2010.05627