Euclid preparation

XIII. Forecasts for galaxy morphology with the Euclid Survey using deep generative models


ABSTRACT

We present a machine learning framework to simulate realistic galaxies for the Euclid Survey, producing more complex and realistic galaxies than the analytical simulations currently used in Euclid. The proposed method combines a control on galaxy shape parameters offered by analytic models with realistic surface brightness distributions learned from real Hubble Space Telescope observations by deep generative models. We simulate a galaxy field of 0.4 deg^2 as it will be seen by the Euclid visible imager VIS, and we show that galaxy structural parameters are recovered to an accuracy similar to that for pure analytic Sérsic profiles. Based on these simulations, we estimate that the Euclid Wide Survey (EWS) will be able to resolve the internal morphological structure of galaxies down to a surface brightness of 22.5 mag arcsec^{-2} and the Euclid Deep Survey (EDS) down to 24.9 mag arcsec^{-2}. This corresponds to approximately 250 million galaxies at the end of the mission and a 50% complete sample for stellar masses above 10^{10} M_⊙ (resp. 10^{9} M_⊙) at a redshift z ~ 0.5 for the EWS (resp. EDS). The approach presented in this work can contribute to improving the preparation of future high-precision cosmological imaging surveys by allowing simulations to incorporate more realistic galaxies.

Key words. techniques: image processing – surveys – galaxies: structure – galaxies: evolution – cosmology: observations

1. Introduction

The Euclid Survey (Laureijs et al. 2011) will observe 15 000 deg^2 (35% of the visible sky) over six years, both in the near-infrared and in the optical at a spatial resolution approaching that of the Hubble Space Telescope (HST). With a field of view of 0.53 deg^2, compared to that of the HST (0.003 deg^2), it will probe the sky at a rate around 175 times faster. It will therefore only take around five hours to observe an area equivalent to the COSMOS field (Scoville et al. 2007), which is still the largest contiguous area...
ever observed by HST and needed around 40 days of observations. In addition to the EWS at an expected nominal depth of 24.5 mag at 10μr for extended sources in the visible (Cropper et al. 2016), Euclid will also observe 40 deg\(^2\) about two magnitudes deeper (EDS). The limiting surface brightness for the EWS in the visible will be 29.8 mag arcsec\(^{-2}\). We refer the reader to Scaramella et al. (in prep.) for precise information about the Euclid surveys and their depths.

Euclid will produce an unprecedented amount of high spatial resolution images that will have a lasting legacy value in a variety of scientific areas, including cosmology and galaxy formation. In order to ensure that the scientific objectives are met, realistic simulations are needed for testing and calibrating algorithms. A standard approach to simulating galaxy images is through analytic Sérsic models (Sérsic 1963). It is well known that galaxies can be modelled, to a first approximation, with two Sérsic functions, one for the bulge component and the other for the disk. Sérsic models have the advantage of being fully described by three parameters, the Sérsic index, which controls the steepness of the profile; the effective radius, which measures a characteristic size for the galaxy; and the axis ratio, which reflects the overall shape of the galaxy. Many previous investigations have shown that Sérsic models reproduce fairly well the average surface brightness distribution of galaxies (e.g., Peng et al. 2002). However, because of their simplicity, they are not well suited to describe complex galactic structure such as spiral arms, bars, clumps, or more generally asymmetric features. This is important for the Euclid mission, however, since the spatial resolution of the visible detector will permit a significant number of galaxies to be resolved. Complex galaxy morphologies can have an impact in the core science of the mission since they can affect the measurement of shear for weak lensing analysis. They are also central to a variety of scientific cases in the field of galaxy formation. The Euclid data will be particularly important to constrain the processes that shape the structures of galaxies and quench star formation, and will allow us to study the relations between detailed morphology, environment, active galactic nuclei activity, and stellar mass, among others (e.g., Lotz et al. 2008; van der Wel et al. 2014; Huertas-Company et al. 2013; Chen et al. 2020; Kocevski et al. 2012; Ferrera et al. 2020; Conselice 2014). Therefore, in order to quantify the possible effects of resolved structures on the image processing pipeline algorithms and to best prepare the scientific analysis of the data, it is important to produce simulations that include realistic galaxy morphologies beyond Sérsic models.

In this work we investigate a novel approach based on generative models to simulate galaxies for the Euclid Survey. We first show that our method can generate realistic Euclid field galaxies with a level of control of the global shapes that is similar to that of analytic profiles, but with the addition of complex morphologies. We then use the generated images to forecast the number of galaxies for which Euclid will resolve the internal structure.

The paper proceeds as follows. In Sect. 2 we introduce the data sets used to analyse Euclid morphological capacities and for training our models. In Sect. 3 we describe the deep generative model used in this work and its training procedure. In Sect. 4 we present our results for the generation of realistic galaxies. In Sect. 5 we use the simulated galaxies to forecast the Euclid morphological limits. We discuss the results of the paper in Sect. 6, and conclude in Sect. 7.

2. Data

We use two data sets for this work: the Euclid Flagship galaxy catalogue (Castander et al., in prep.), hereafter the Euclid/Flagship catalogue, and the COSMOS survey (Scoville et al. 2007). We use the first to simulate best the expected Euclid data as the goal of the paper is to forecast Euclid capacities. The second is used to train our deep learning model so that we learn how to simulate realistic galaxies.

2.1. Target set: Euclid Flagship catalogue

To quantify the performance of our model in Euclid-like conditions and establish morphological forecasts for the mission, we used the Euclid Flagship catalogue. We accessed the catalogue through CosmoHub, a platform that allows the management and exploration of very large catalogues, best described in Tallada et al. (2020) and Carretero et al. (2017).

The Flagship catalogue was built using a semi-empirical halo occupation distribution (HOD) model and was intended to reproduce the global photometric and morphological properties of galaxies as well as the clustering. We refer the reader to Merson et al. (2013) for more details. In order to produce a catalogue close to the real Universe, the morphological parameters, which is what we mainly use in this work, are calibrated on the CANDELS survey (Dimauro et al. 2018) and 3D model fitting on the GOODS fields (Giavalisco et al. 2004) by Welikala et al. (in prep.). Details about the catalogue production will be presented in Castander et al. (in prep.). Each simulated galaxy in the catalogue is made of two components, a bulge and a disk. The bulge component is modelled as a Sérsic profile with an index varying from \( n = 0.3 \) to \( n = 6 \). The disk component is rendered using an exponential profile \( (n = 1) \). The version of the Euclid Flagship catalogue used in this work contains 710 million galaxies distributed over 1200 deg\(^2\), from which we took a random subsample of 44 million galaxies. The distributions of the main morphological parameters used in this work are presented in Fig. 1: the half-light radius \( r_\text{e} \), the axis ratio \( q \), and the Sérsic index \( n \). We also show the apparent magnitudes of the galaxies as measured by VIS, which is the visible imager of Euclid (Cropper et al., in prep.), as well as the redshift and the stellar mass distributions, which we use in Sect. 5 to perform our forecasts. Finally, we show the bulge-to-disk component flux fraction (hereafter bulge fraction).

We note here that the Euclid Flagship catalogue is a pure tabular catalogue. The procedure currently used within the Euclid Consortium to generate the galaxies is described in Sect. 4.1.1, when we compare our galaxies to the current analytic ones. Our work in this study is to use this catalogue of double Sérsic profile parameters to generate the 2D images of the internally structured galaxies.

2.2. Training set: COSMOS

The training set is based on the COSMOS survey. COSMOS is a survey of a 2 deg\(^2\) area with the Hubble Space Telescope Advanced Camera for Surveys (ACS) Wide Field Channel using the F814W filter. The final drizzle pixel scale is of 0.03 pixel\(^{-1}\) and the limiting point source depth at 5σ is 27.2 mag. The central wavelength of the F814W filter roughly corresponds to that of the VIS filter (550–900 nm) and the spatial resolution and depth are better than those expected from the Euclid Survey. Therefore, the data set is well suited and is expected to be close enough to the Euclid data, allowing us to generate mock Euclid fields without being affected by the dependence of morphology on wavelength and without introducing undesired effects owing to extrapolations.

Our selected sample is based on the catalogue by Mendelbaum et al. (2012), which has a magnitude limit of 25.2
and contains 87,630 objects. The catalogue provides, for each galaxy, the best-fit parameters of a one-component and a two-component Sérsic fit by Leauthaud et al. (2007), updated in 2009. In this work, we use only the one-component fitting information. In Fig. 1 we show the distribution of the COSMOS morphological parameters of galaxies compared to those in the Euclid Flagship catalogue. Although the distributions are similar, there are some noticeable differences which might cause a problem. The most obvious one is the magnitude. Since COSMOS is magnitude limited, the sample does not contain as many faint galaxies as the simulation. The half-light radii of the Euclid Flagship catalogue bulge component also extend to smaller values than those in the observations. They are also generally rounder than the observed ones, but the values of axis-ratios span a similar range. The Sérsic index distributions are also different because, as explained previously, the Euclid Flagship disk component always has a Sérsic index of 1. In addition, in the COSMOS catalogue the Sérsic indices of the bulge component are clipped at \( n = 6 \) to be compatible with GaLSim, which creates a noticeable spike at the edge of the distribution. The mass fraction and redshift is derived by Laigle et al. (2016). As we show in the following sections, these differences, although present, do not have a significant effect on our methodology. The most important desirable property is that simulated galaxies cover a similar range to observations. That way, the neural network used in our model is not compelled to extrapolate. This is essentially the case in the distributions shown in Fig. 1, except for very small bulge components and for very faint galaxies, both of which are not expected to present significant features. We address these points in the following sections.

In addition to the catalogue, the authors also provide 128 × 128 pixel stamps centred on each galaxy where neighbouring galaxies have been removed. This is important for training our model on a unique galaxy per stamp. Therefore, the impact of galaxy blending in the morphology forecasts will not be studied in this work. In addition, the size of the stamps inherently limits the size of galaxies that we will be able to generate. The radius of the stamp being 64 pixels, every galaxy with a half-light radius larger than \( \sim 2'' \) will be cut by the limits of the stamp. For this reason, in this work we are limited to, and thus only consider, galaxies smaller than \( 2'' \). Nevertheless, galaxies with a radius bigger than \( 2'' \) represent only 0.6% of the Euclid Flagship catalogue, and thus have no major impact on our results.

The COSMOS images are pre-processed before they are used for training, as illustrated in Fig. 2. We first degrade the spatial sampling from 0.03 pixel\(^{-1}\) to 0.1 pixel\(^{-1}\), which corresponds to the pixel scale of VIMOS, and then pad the image with the appropriate noise. We use the GaLSim (Rowe et al. 2015) method described in Sect. 5 of Mandelbaum et al. (2012). Since the pixel scale increases, the final stamp needs to be padded with noise to keep the size of 128 × 128 pixel. The method does this automatically by adding a noise realisation with the same characteristics as in the original stamps, which also takes into account the different correlations in the original noise. Doing so, the resulting images are still at the size of the COSMOS stamps. Since the pixel size is increased, we can crop up to a factor of three without losing spatial covering. However, because our model is more efficient with images that have a number of pixels which is a power of two (for parity reasons between the compression and decompression steps of our deep learning network), we
crop our image by only a factor of two, resulting in images of 64 × 64 pixel. The purpose of this cropping is to accelerate the training. We finally rotate the stamps so that the galaxy semi-major axis is aligned with the x-axis of the image. With this configuration we ensure that our model will learn to produce only ‘horizontal’ galaxies and therefore position angles can be manually added in post-processing. This has the additional advantage of reducing the complexity and hence allowing the neural network to focus the attention on the more important physical properties of the object. Figure 2 illustrates these pre-processing steps used for the training of our model, and the final galaxy as it would be seen by VIS. Because galaxies produced by our model will be noise-free and not convolved by the PSF, we do not need to change the noise level and the PSF for the training. Thus, the inputs of our model have the noise characteristics and the PSF of the HST images. These two transformations, to go from HST to Euclid data will be added a posteriori. More information about those transformations are described Sects. 4.1.1 and 4.2.2.

We use the COSMOS catalogue and images only for the training of our model. To test the performance of our model, we only use the Flagship catalogue described in the previous section.

3. Euclid emulator with generative models

In this section we describe the methodology for emulating Euclid galaxies using the COSMOS sample described in the previous section.

The generation of synthetic data (images, language, videos) has significantly improved in recent years thanks to new deep learning-based generative models. Generative models are a type of unsupervised machine learning algorithms that are trained to generate unseen data. There are several architectures; variational autoencoders (VAE: Kingma & Welling 2013), generative adversarial networks (GANs: Goodfellow et al. 2014; Arjovsky et al. 2017), and autoregressive models (van den Oord et al. 2016) are the main ones. They all learn a probability distribution function of the pixel distribution, which can be sampled to generate new data. Generative models have already been used in astrophysics for a variety of different purposes. For example, with VAEs radio galaxies can be simulated (Bastien et al. 2021) or images of overlapping galaxies can be reconstructed separately (LSST Dark Energy Science Collaboration 2021). Using GANs, Yi et al. (2020) have simulated missing data from the cosmological microwave background, while Villaescusa-Navarro et al. (2021) have simulated gas density maps. Storey-Fisher et al. (2020) and Margalef-Bentabol et al. (2020) have used GANs to detect outliers in imaging surveys. Autoregressive flows can be used to compare simulations and observations (e.g., Zanisi et al. 2021).

In this work we use a VAE. Variational autoencoders estimate an explicit latent space, which is an important advantage for simulating galaxies with known parameters. The compression-decompression architecture inherent to the VAEs along with the Kullback–Leibler term in the loss (see Sect. 3.1.1 and Eq. (3)) force the latent representation to be meaningful and regular. In addition, VAEs are known to be more stable during training, and less subject to mode collapse (lack of diversity in the generation) than GANs.

3.1. Model

Our model for generating galaxies is based on the work by Lanusse et al. (2021, hereafter L2021) who describe in detail the architecture and specifics of the training procedure. We also illustrate the architecture of the two components of our model in Figs. B.1 and B.2.

The goal of our work is to simulate and test galaxies with more realistic shapes than the classical analytic profiles while keeping a control on the shape parameters, such as axis ratios, effective radii, and fluxes. To this end, our model is made of two distinct parts: a variational autoencoder (Kingma & Welling 2013), which learns how to simulate real galaxies from observations, and a normalising flow (Jimenez Rezende & Mohamed 2015) in charge of mapping catalogue parameters to the VAE latent space. Both parts are merged together after training, resulting in an architecture called a flow variational autoencoder (FVAE). We describe in the following the global properties of these two models.

3.1.1. Galaxy generation with a variational autoencoder

A VAE is a deep generative model which is trained to generate new data (galaxies) by learning a probability distribution from the training data. To this end, the VAE first compresses the input image x into a low-dimensional space, also called latent space, which contains a compact and meaningful representation of the input data. Similar objects are compressed into neighbouring vectors. This is achieved with a convolutional neural network called the encoder, which can be represented as a non-linear function $E_\theta$, $\theta$ being its trainable parameters. While a classical autoencoder compresses the input image only into a vector z, a VAE replaces that low-dimensional vector with a probability distribution function (PDF) $p_\theta(z|x)$. In our case, $p_\theta(z|x)$ is set to be a multivariate Gaussian distribution. This is equivalent to choosing the prior for the distribution of points in the latent space to be Gaussian. Similar galaxies will be encoded into similar regions of the distribution. Having a distribution instead of a point estimate makes the latent space continuous, allowing one to sample new regions from it and to produce new galaxies arising from the same probability density function as the data.

A sample z is then drawn from the distribution $p_\theta$. This constitutes the input of a second convolutional neural network called the decoder $D_\theta$, which typically has an architecture symmetric to that of the encoder. The decoder decompresses the latent representation z using transposed convolutions to produce a new image $\hat{x}$, $D_\theta(z) = \hat{x}$. The output of the decoder can be seen as the probability that the input data x effectively come from the latent space vector z (i.e. $D_\theta(z) = p_\theta(x|z)$). During training, the goal is to reconstruct x with the best possible accuracy (i.e. $\hat{x} = x$) ensuring that the distribution encoded within the latent space is a good representation of the data. The amount of information loss in the compression–decompression is the first term of the neural network loss function $L$, which is used to adapt $\theta$ and $\Theta'$ through a gradient descent minimisation. From a statistical point of view, this accuracy is defined as the negative log-likelihood of x given z, which can be written using the expectation value:

$$L = -E_{z \sim D_\theta(z)} \left[ \log p_\theta(x|z) \right].$$  

(1)

In practice, we can simply see the reconstruction accuracy as the mean square error between the reconstructed image and the input:

$$L = ||x - \hat{x}||^2.$$  

(2)

In addition, in order to regularise $p_\theta$, a second term is added to penalise the encoder when it produces distributions too far from
a normal Gaussian distribution \( \mathcal{N}(0, 1) \). This difference between \( p_\theta(z \mid x) \) and \( \mathcal{N}(0, 1) \) is estimated using the Kullback–Leibler divergence (Kullback & Leibler 1951):

\[
\mathbb{KL} = \mathbb{E} \log p_\theta(z \mid x) - \log \mathcal{N}(0, 1).
\]

The final loss function for the VAE reads

\[
\mathcal{L} = - \mathbb{E}_{z \sim q_\phi(z \mid x)} \left[ \log p_\theta(x \mid z) \right] + \beta \mathbb{E} \left[ \log p_\theta(z \mid x) - \log \mathcal{N}(0, 1) \right],
\]

where \( \beta \) allows us to vary the importance of the terms during training.

Lanusse et al. also introduce two additional features in order to produce images deconvolved by the PSF and without noise. To learn noise-free galaxies, a different version of the log-likelihood for the reconstruction term of the loss function is used. Instead of applying it directly to the pixels, it is done in Fourier space in order to weight the reconstruction error less on the high frequencies (noisy regions). The Fourier transform of the input and of the output is computed, and divided by the power spectrum of the noise. By dividing the Fourier transform of the image by the power spectrum of the noise, a smaller weight is given to the pixels with a high frequency. It ensures that the decoder learns that producing images without noise is not an error. In order to produce deconvolved images, the last convolutional layer of the decoder is not trainable and is set to be equal to the PSF. That way, the model produces an image that looks like the input image before being convolved by the PSF in the second last layer.

3.1.2. Sample of the shape parameters with the regressive flow

The VAE described in the previous subsection can generate realistic galaxies by sampling from the encoded latent space. However, it cannot do so for a given size or ellipticity because it lacks the information about the mapping between the structural parameter space of the galaxy and the latent space.

To learn that mapping, L2021 propose a conditional normalising flow, based on autoregressive algorithms (MAF: Papamakarios et al. 2017, MADE: Germain et al. 2015). A normalising flow is a bijection \( g_\phi \), which transforms a distribution \( q \) into another distribution \( p \) with an invertible transformation \( g \). We use it here to learn the mapping between a latent space with a fixed distribution \( q \), referred to as the flow latent space, and the distribution \( p \) inside the VAE latent space. This mapping can be made conditional to some input parameters \( y \) such as galaxy size or ellipticity. In other words, \( g_\phi \) is a function of both the latent space vector \( z \) and the physical parameters of the galaxy \( y \).

If the mapping is well learnt, when we sample a vector \( z_{\text{flow}} \) from the flow latent space distribution \( q \) and pass it through \( g_\phi \), then it will output a vector \( \hat{z} \) in the VAE latent space

\[
\hat{z} = g_\phi(z_{\text{flow}}, y).
\]

such that \( \hat{z} \) is very similar to the vector \( z \), which would have been encoded by the VAE’s encoder from a galaxy \( x \) with physical parameters \( y \)

\[
\hat{z} \approx E_\phi(x_\gamma) .
\]

With this mapping, we now know where to sample into the VAE latent space in order to decode a galaxy with precise physical parameters: to simulate a galaxy, we need to map \( z_{\text{flow}} \) and \( y \) to the VAE latent space, and then decode the vector with the decoder to produce an image of a galaxy that has the physical properties given by \( y \).

In practice, the training procedure is done the other way around: we learn how to map a vector \( z = E_\phi(x) \) into a vector \( z_{\text{flow}} \) of the flow latent space. Because \( g_\phi \) is a bijection, learning the mapping from the flow latent space to the VAE latent space or the other way around is the same task, but doing it in this direction is much easier because of the loss. The loss we use is the negative log likelihood of \( z \) under the distribution of the flow latent space \( q \)

\[
\mathcal{L}_{\text{flow}} = \mathbb{E}_{z \sim p} \left[ - \log p(z) \right]
\]

\[
= \mathbb{E}_{z \sim p} \left[ - \log q(g^{-1}(z)) + \log \det J_g(z) \right],
\]

where \( J_g \) is the determinant Jacobian of \( g \), which is used to compute the loss and optimise the weights of \( g \).

To implement the flow, we use the probabilistic library of TensorFlow, TensorFlow probability. With this library it becomes straightforward to implement the bijector \( g \), with a chain of masked autoregressive layers, described in Germain et al. (2015). The transformations of the distribution made by the successive layers (shifts of the mean and stretch of the dispersion) are conditioned to the physical parameters of the flow’s input. Then, thanks to the Distribution object of the library, with only one command it is possible to sample the transformed distribution (e.g. to get \( z_{\text{flow}} \)), but also to take the log likelihood for the computation of the loss.

3.1.3. Final model

The final model (schematic representation in Fig. 3) combines the decoder part of the VAE with the regressive flow described in the previous subsection. Therefore, the input of the final model is a galaxy catalogue. The flow samples a Gaussian noise vector, which is concatenated with the catalogue parameters to produce a vector in the latent space. The vector is then decoded by the generator of the VAE, producing the image of a new galaxy with the corresponding input parameters from the catalogue. The use of a continuous distribution enables the generation of new galaxies that resemble real ones, but have never been observed before.

3.2. Training procedure

The main goal of this work is to produce Euclid-like realistic galaxies. We use pre-processed COSMOS galaxies (described in Sect. 2) to train the VAE. We train it for 250 000 steps, which means 3900 epochs (one epoch is when the whole training set has been seen by the network) with a batch size of 64 (the batch size is the number of images with which we perform each gradient descent). The latent space has a dimensionality of 32.
The Euclid Consortium currently creates analytic galaxies with 

4.1.1. Simulations with pure Sérsic profiles

The Euclid Consortium currently creates analytic galaxies 

4.1. Simulation of composite galaxies

4.2. Qualitative inspection

4.2.1. Individual noise-free galaxy simulation

We first qualitatively evaluate our simulations. Figure 4 shows 

4. Emulation of VIS images

In this section we analyse the properties of simulated galaxies 

4.1. Simulation of composite galaxies

4.1.1. Simulations with pure Sérsic profiles

The Euclid Consortium currently creates analytic galaxies 

4.2. Qualitative inspection

4.2.1. Individual noise-free galaxy simulation

We first qualitatively evaluate our simulations. Figure 4 shows 

4. Emulation of VIS images

In this section we analyse the properties of simulated galaxies 

4.1. Simulation of composite galaxies

4.1.1. Simulations with pure Sérsic profiles

The Euclid Consortium currently creates analytic galaxies 

4.2. Qualitative inspection

4.2.1. Individual noise-free galaxy simulation

We first qualitatively evaluate our simulations. Figure 4 shows 

with a constant magnitude set to 24, a fixed Sérsic index of 1.5 and a varying axis-ratio $q$ and half-light radius $r_e$. Figure 6 shows a grid of galaxies with fixed $r_e$ and magnitude but varying axis-ratio and Sérsic index. We can clearly observe the expected trends. Galaxies become rounder as we move from left to right, and bigger from top to bottom in Fig. 5. In Fig. 6 galaxies become more concentrated as the Sérsic index increases from left to right. The images also show several examples presenting non-trivial symmetric shapes. An important limitation to note is that our model is fixed to produce images of size $64 \times 64$ pixel. Very large galaxies might therefore be truncated.

### 4.2.2. Large field simulation

In addition to individual stamps, we also generate two large fields of 0.4 deg$^2$ at the depths of the EWS and the EDS (see a portion of those fields in Fig. 7). We take a subsample of the Euclid Flagship catalogue and generate every galaxy without noise and deconvolved by the PSF. We then convolve the stamp by a unique VIS PSF (no PSF variations are modelled). All the stamps are then placed in the large field into their corresponding positions according to the catalogue. We finally add the expected noise level of the EWS and the EDS in two different realisations of the same field. The background noise (coming mostly from background sources and from the zodiacal light) is simulated by Gaussian noise with the expected standard deviation for the VIS camera (Cropper et al. in prep.; Scaramella et al. in prep.; priv. comm.). The photon noise is simulated with a Poisson distribution added to every pixel, considering the cumulative exposure times presented by Laureijs et al. (2011).

More information will be given about the noise realisations in Merlin et al. (in prep.). We do not simulate any instrumental effects such as cosmic rays, ghosts, charge transfer inefficiency, or read-out noise, considering thus an ideal case of a VIS image processing pipeline. In Fig. 7 we show a random region of the large fields, and highlight some interesting galaxies.

### 4.3. Quantification of structural properties

This visual assessment of the previous subsection confirms that our model behaves as expected both in generating complex shapes and controlling structural parameters. However, in order for the simulation to be useful to test algorithms, it is required that the control on the structural parameters is comparable to what is achieved with analytic profiles.

#### 4.3.1. Surface brightness profiles

We compare the radial profiles of generated galaxies with the profiles of analytic galaxies with the same global properties. Figure 8 compares and shows the radial profile for three bulge components, disk components, and the combination of the two components, simulated with our model and with GalSim. All the images are convolved by the VIS PSF but are without noise. We show both the profile along the major axis and the azimuthally averaged profile. The former is useful to identify deviations from a smooth profile, and thus highlights where the irregularities take place. The latter, computed by averaging the luminosity at a given radius $r$ from the galaxy centre in all directions, allows us to check if the average profile behaves as expected compared to the Sérsic model. Overall, the figure shows the expected behaviour. Some profiles deviate significantly from a Sérsic profile along the major axis. An example for this is the disk component shown in the bottom row of Fig. 8, where we can see a spiral arm feature that creates variation in the radial light profile.
Fig. 7. Illustration of a large field simulation produced by our FVAE. Top and bottom panels: same field simulated at the depths of the EDS and the EWS, respectively. The stamps show zoomed-in regions where some galaxies present morphological diversity. In the large field images, we use the IRAF 'zscale' that stretches and clips the low and high values to better highlight the differences between the EWS and the EDS. The stamps are in linear scale, which better emphasises the structures. With the stretching induced by the zscale, all the structures disappear and only the global shape is still recognisable. Finally, the apparent different level of background between the stamps and the global image is also due to the different brightness scale (different maximum and minimum values in each of them).
Euclid Collaboration: Euclid preparation. XIII.

BULGES

DISKS

COMPOSITE

Fig. 8. Examples of three radial profiles of galaxies generated with GalSim and our model. Each group of two columns represents the different components of the galaxy: bulge, disk, and composite (bulge plus disk) from left to right. Within each group the top row shows the images by our model (left) and by the Sersic model (right). The bottom line represents the light radial profiles, along the major axis (left) and the average profile (right). The orange lines correspond to our model, and blue to the Sersic profile. The dashed grey line represents the EWS noise level. Our simulations show more diverse profiles, but the average closely matches the analytic expectations. The irregularities at very low S/N on the FV AE profiles are a sign that the model does not produce perfectly noise-free galaxies.

However, the average profiles tend to follow the analytic expectations since irregularities are averaged out. Therefore, the generated galaxies seem to present the desired behaviour (i.e. complex surface brightness distributions), which on average match a Sersic profile. An additional interesting result seen in Fig. 8 is that the combination of the two components also behaves very similarly when compared to a double-component analytic galaxy (see the composite galaxy column).

4.3.2. Surface brightness fitting

We now fit Sersic models to quantify how accurately the shape parameters are recovered in a statistical sense. For this purpose we use the Galapagos package (Barden et al. 2012; Haübler et al. 2013). Galapagos is a high-level wrapper for SExtractor (Bertin & Arnouts 1996) and Galfit (Peng et al. 2002) to automatically fit large samples of galaxies. Because two-component Sersic fits are generally less stable than one-component fits (e.g. Simard et al. 2011, Bernardi et al. 2014, Dimauro et al. 2018) we decide to produce the two components separately in two distinct realisation of the field. Thus, we have two different fields, one with only the bulge component and one with only the disk component. We then fit each field with the one-component Sersic model. This allows us to test the reliability of the fits while reducing the degeneracies. Since our objective is to compare our simulation to an analytic one, a single Sersic fit is enough for our purpose.

Using the Euclid Flagship catalogue, we generate with our model a galaxy field of 0.4 deg² (i.e. around 2500 galaxies with magnitude lower than 25), following the same procedure as in Sect. 4.2.2. We then use the same procedure and subsample to produce the same field with the pure Sersic profiles. The two fields are therefore identical in terms of number of galaxies and positions, and contain galaxies with the same structural properties.
Fig. 9. Results of 2D Sérsic fits to the surface brightness distributions of bulge components. In every panel, the orange points and histograms represent the results for the FV AE galaxies and the blue for the analytic galaxies. Each panel represents a different parameter, as labelled. For each parameter the true value of the parameter is plotted as a function of the inferred one from the best-fit model. A perfect fit corresponds to the diagonal. In addition, above and to the right of each plot are the projected distributions of the scatter plot. Finally, the inset plot shows the distribution of the error (fitted value minus true value). To make the scatter plot less crowded, only half the galaxies are plotted, but the error histograms and the projected distributions are computed on the entire field (for more details of the error distributions, see Table 1).

Figures 9 and 10 show the fitting results concerning bulge and disk components, respectively, for five parameters: half-light radius $r_e$, axis ratio $q$, Sérsic index $n$, centroid position $X$, and total magnitude. We recall that the goal of this comparison is not to quantify the absolute accuracy of the fits, but to compare the relative behaviour of our simulations with a baseline. A future publication in preparation will quantify in detail the accuracy of structural parameters in both the EWS and the EDS. Overall, the structural parameters are recovered with similar dispersion for the FV AE and the analytic simulation. This is a first quantitative confirmation of the visual inspection of the previous sections. Our model is able to produce realistic galaxy images while preserving information on the parametric structure. The global distributions of the predicted parameters are very similar, which confirms that our model has correctly learnt the entire distribution of the training set, and is thus able to span the entire parameter space of the Euclid Flagship catalogue.

Looking in more detail, the FVAE results present a slightly larger dispersion in all recovered parameters. This is expected since the analytic simulations represent a perfect match for the model that is fitted. This is not the case for the FVAE simulations, which present more complex profiles. We give the statistical details of the fitting distribution errors (median, first, and third quartile) in Table 1, corresponding to the distributions in the insets in each panel of Figs. 9 and 10.

The systematic offsets might be more problematic. The figure shows that the systematic shifts for the bulge components are very similar for the analytic and the FV AE fields which means that using a FV AE does not introduce any noticeable systematic effects. The only parameter that presents a small bias towards larger values is the axis ratio $q$. This might be because of a lack of very elongated bulges in the training data set. The disk components present a slightly higher systematic bias though, as shown in Fig. 10. Indeed, FV AE galaxies tend to be systematically larger and rounder than their analytical counterparts and show an almost constant offset of 0.15 on the Sérsic index. It is not obvious whether these offsets are a consequence of the simulation or whether it is related to the fitting procedure itself. A possible explanation for the larger offsets is that disk components are generally more extended and with flatter profiles than bulge components, thus they also present more complexity and structure. Alternatively, it can also be related to the simulation itself. Our training set is based on a single-component fit with a
continuous distribution of the Sérsic index. However, the Sérsic index of the disk component in the Euclid Flagship catalogue is fixed to $n = 1$. This means that there is only a small number of examples in the training set with exactly $n = 1$, which can affect the quality of the generation. Finally, we can see that for the magnitude, the fit of our galaxies also differs very little from the Sérsic fits, even if the flux is not something that is parametrised magnitude, the fit of our galaxies also differs very little from the Sérsic fits, even if the flux is not something that is parametrised.

Table 1. Accuracy of fitting results.

<table>
<thead>
<tr>
<th></th>
<th>FVAE</th>
<th>Bulges</th>
<th>Disks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$q_1$</td>
<td>$\mu_{1/2}$</td>
<td>$q_3$</td>
</tr>
<tr>
<td>$X$</td>
<td>$-0.36$</td>
<td>$0.01$</td>
<td>$0.40$</td>
</tr>
<tr>
<td>mag</td>
<td>$-0.33$</td>
<td>$-0.06$</td>
<td>$0.03$</td>
</tr>
<tr>
<td>$r_e$</td>
<td>$-0.23$</td>
<td>$0.25$</td>
<td>$1.77$</td>
</tr>
<tr>
<td>$q$</td>
<td>$-0.10$</td>
<td>$0.00$</td>
<td>$0.07$</td>
</tr>
<tr>
<td>$n$</td>
<td>$-0.64$</td>
<td>$0.23$</td>
<td>$0.52$</td>
</tr>
</tbody>
</table>

Notes. For each parameter shown in Figs. 9 (bulges) and 10 (disks), we present the first quartile ($q_1$), the median ($\mu_{1/2}$), and the third quartile ($q_3$) of the fitting error distributions.

5. Forecasts for galaxy morphology with Euclid

The previous sections have shown that our proposed framework successfully generates galaxies with realistic and resolved structure. Our simulations can therefore be used to establish some forecasts on the number of galaxies for which Euclid will be able to resolve the internal structure beyond a Sérsic profile.

5.1. Identifying galaxies with resolved structure

Our goal is to quantify the fraction of galaxies that present significant structures that deviate from a pure analytic profile. For that purpose we have designed a method to distinguish galaxies with internal structure from smooth objects. We assume that any type of complexity in the galaxy surface brightness distribution, hereafter called structure, will result in a deviation from an analytical profile. This is particularly clear in the disk component shown in Fig. 8. We therefore establish a criterion to characterise the smoothness of a galaxy by computing the derivatives of the semi-major axis profile. For illustration purposes, we show in Fig. 11 three toy profiles. A pure analytical profile, a profile presenting a strong structure, and a slightly perturbed one. When the profile is smooth the first derivative is also smooth, changing its sign only at the centre of the galaxy. If we consider only a one-sided profile, the derivative never goes to zero (i.e. it has no roots). Its second derivative is also smooth, and has only one root that we call a ‘natural zero’. When the galaxy is strongly perturbed, the profile will significantly differ from a pure analytical profile. For a Sérsic profile the light curve decreases from the centre to the edge of the galaxy; instead, for example in a galaxy presenting a spiral arm, the major axis profile increases in the location of the arm. This increase (change of slope) will cause a sign change in the first derivative, and thus two changes in sign in the second derivative, as can be seen in the second column of Fig. 11. However, the roots of the first derivative are not always enough to detect a variation from a smooth profile, as illustrated in the third column of the figure; the profile can be slightly perturbed, with a change of slope in the profile, but this does not make the profile rise as in the second column of the figure, but significantly changes the rate of decrease. Thus, the first derivative will not change in sign (the profile does not increase), but the second derivative will (the rate of the decrease changes).

Therefore, we conclude that the presence of a zero on the second derivative of the light profile (without counting the natural zero) is a reasonable indicator of a galaxy with complex structures. We note that there will be additional zeros at the edge of the profile when it becomes flat. However, these roots will be all consecutive, giving us a way to distinguish zeros coming from a structure from ones coming from the end of the profile. Thus, we can consider a galaxy being structured if its second derivative has two roots (without considering the first natural one), which are far enough from each other. This also prevents the high-frequency perturbations in the profile that we do not want to consider as a structure. We find that, at the VIS resolution, a minimum distance of 1 pixel (approximately one PSF FWHM) between roots is a reasonable choice. To make sure that we do not miss structures that are not along the semi-major axis, we also search for structures with the same method along the semi-minor axis of the galaxy.

We show in Appendix A two random selections of galaxies which have been classified with and without structure. Our method successfully isolates galaxies with perturbed or asymmetric profiles.
Three toy profiles that illustrate our structure detection method. Left panel: smooth galaxy without structure, middle panel: strongly perturbed galaxy, and right panel: slightly perturbed object. For each profile, its luminosity is plotted as a function of the distance to the galaxy centre in arbitrary units (blue solid lines). Their corresponding first and second derivatives are also plotted (orange and green solid lines, respectively). We can see that the number of roots in the second derivative is a good indicator of perturbed galaxies.

Forecast of the number of galaxies with internal structures for the EWS and the EDS, regarding surface brightness. Left panel: fraction of galaxies with resolved structure as a function of surface brightness. Right panel: total number of galaxies with resolved structure as a function of surface brightness. The red squares are for structures discernible for the EWS at $2\sigma$ around the noise level. The blue stars represent the same information, but for the EDS.

5.2. Resolved complex morphologies in Euclid

We use this technique to infer the fraction of galaxies for which *Euclid* will be able to resolve internal morphological structure beyond Sérsic profiles. We simulate galaxies without noise and compute the semi-major axis profile and consider only pixels $2\sigma$ above the noise level. We then plot in the left plot of Fig. 12 the fraction of galaxies presenting structures as a function of the surface brightness $S_b$ of the galaxy, defined as

$$S_b = m + 2.5 \log_{10}(\pi q_{\text{tot}} r_{\text{tot}}^2),$$

(10)

where $r_{\text{tot}}$ (in arcsecond) and $q_{\text{tot}}$ are the global (disk and bulge components) half-light radius and axis ratio of the galaxy. Thus, $\pi q_{\text{tot}} r_{\text{tot}}^2$ represents the area of the galaxy.

We can see that the fraction of galaxies with resolved structures decreases with increasing surface brightness, as expected. The behaviour of the EWS and the EDS is self-similar, but the EDS is shifted towards fainter surface brightness. The difference is of the order of 2 magnitudes: less than 10% of galaxies present detailed structures above $2\sigma$, beyond a surface brightness of 22.5 mag arcsec$^{-2}$ for the EWS and 24.9 mag arcsec$^{-2}$ for the EDS. The statistical fluctuations on the curve are similar because we compute our structure indicator on the same realisations of galaxies with only the S/N changing.

We also provide the total number of galaxies per bin in the right panel of Fig. 12. We simply multiply the fraction of objects with structure by the total number of galaxies in the 15 000 deg$^2$ of the EWS and in the 40 deg$^2$ of the EDS. We conclude that *Euclid* will observe around 250 million galaxies that are significantly more complex than the analytical profiles during the six years of the mission.

Figure 13 shows a 2D representation of the fraction of galaxies with resolved structures above $1\sigma$ and $2\sigma$ of the noise as a function of magnitude and half-light radius. The first line represents *Euclid* capacities for the EWS, and the second for the EDS. First column is the percentage of galaxies presenting structures above $1\sigma$ of the noise level, and the second column above $2\sigma$. The colour-coding is the same as in Figs. 12 and 14. The blue number in each column (row) indicate the mean percentage of the corresponding column (row).
same behaviour, namely that the EDS goes around two magnitudes deeper to probe morphologies. The fraction of galaxies decreases in the limit of the distributions when we increase the level of acceptance from $1\sigma$ to $2\sigma$. The figure summarises the following expected behaviour: (1) the brighter the galaxy, the larger the number of resolved structures (top to bottom gradient) and (2) the fraction becomes smaller for extremes (very small and very large galaxies) at constant magnitude. The decrease at small sizes is a consequence of resolution. At large sizes it is related to S/N. We recall that we did not plot galaxies bigger than $2''$ because of the built-in size limitation of our model, but we expect the decreasing trend to continue at larger radii.

Finally, in Fig. 14 we forecast the fraction and the total number of galaxies with resolved structures as a function of physical properties of galaxies, namely stellar mass and redshift. We conclude that the EWS will be able to reach a 50% completeness regarding the detection of internal structures of galaxies down to $10^{10.6} M_\odot$ at $z \sim 0.5$. The EDS reaches down to a stellar mass of $10^{9.6} M_\odot$ up to $z \sim 0.5$.

We note here that we are probing the internal structures of the galaxies, and not assessing whether the galaxy is resolved or not. We thus consider, in our forecasts, that intrinsically smooth galaxies such as spheroids have no structures, even if they are resolved by Euclid. Since our model is trained on real data, it is reasonable to assume that the fraction of different morphologies is well reproduced. The numbers we provide are therefore an estimate of the fraction of galaxies with complex internal structure, beyond a Sérsic model.

6. Discussion: A framework to simulate future surveys

This work presents a novel framework to generate galaxies with realistic morphologies, while keeping control on the global structural properties. It can be used to calibrate algorithms for future experiments such as Euclid in which the impact of complex galaxy shapes might become significant. This is the case for example for galaxy deblending or even shear estimations. We discuss in this section possible limitations of a large-scale use of generative models for galaxy generation.

One possible bottleneck is execution time. We therefore quantify the execution speed of our framework compared to that of a classical analytic generation. We use two different environments: with and without GPUs. We used a 16 CPU Intel Xeon Bronze 3106, and an NVIDIA Tesla P40 GPU. We then tested our method with increasing batch sizes, going from one galaxy at a time to 64. The results of the different experiments are summarised in Fig. 15. Each measurement refers to the execution time of a standard analytic simulation. The training time is not discussed here as it has to be done only once, and does not enter execution time discussions.

The figure confirms that a GPU is around four times faster than a CPU environment in all configurations. We also see that the batch size has a dramatic impact on the execution speed. For a batch size of one, our method is more than a 100 times slower than a traditional approach. However, the difference is reduced to a factor of around 5 if larger batches are used. It is interesting to note, however, that the execution time does not depend linearly on the batch size. This is a well-known behaviour (Wilper et al. 2020). We note that for this figure, as in all this work, we simulate galaxies by a sum of two components. As explained before, we did that to match the current Euclid simulation strategy. Nevertheless, we highlight here that we are capable of creating complex galaxies with only one component, and therefore all the times of the figure could possibly be divided by two if we were simulating only one component.

Another possible limitation of our proposed framework is the fact that it is trained on observations which therefore contain biases that can propagate the simulation. In particular, we used
here the COSMOS GalFit t fitting as a ground truth to condition the autoregressive flow. The impact of this could be assessed by using different independent fitting codes on the same data sets and comparing the results. This is an ongoing effort as part of the Euclid Morphology Challenge, which will be presented in a forthcoming publication. The diversity of generated galaxies is also limited by the quality of the observations used for training, in this case HST observations. This restricts the range of parameters that our model can probe without extrapolation. This could be mitigated with additional data sets, but we have not explored that in this work. For this reason, we do not recommend using our framework to simulate images without noise as features below the HST noise level are not constrained. We also note that our model is limited regarding the size of the galaxies it can generate because of the fixed size of the training stumps. Simulating galaxies with half-light radii bigger than $2''$ is not recommended. Some galaxies larger than $1.5''$ and with a small Sérsic index (flux above the half-light radius not negligible) can also produce some artefacts if the size of the galaxies is recovered from the training distribution, and because the faint end of the galaxy will be cut. We used this large limit of $2''$ to do our morphological forecasts because those artefacts do not cause problems in our structure detection algorithm. In addition, to produce galaxies fainter than the limiting magnitude of the training set (25.2 mag), we assumed that the galaxy morphology is not correlated with the magnitude, which is of course an approximation. Finally, to establish morphology forecasts, we assume that the amount of structures produced by our model is the same as in real galaxies. Our model may tend to produce galaxies that are smoother than in real galaxies. Therefore our forecast may have underestimated the number of objects with complex morphologies. On the contrary, our choice to use fields without any instrumental effects but the PSF could decrease the effective number of detected galaxies. Finally, the number of low-magnitude galaxies in the Euclid Flagship catalogue could be underestimated, for example compared to the catalogue (Connolly et al. 2014) used for the Legacy Survey of Space and Time (LSST) at the Vera C. Rubin Observatory (Ivezić et al. 2019). This lack of faint galaxies could increase the numbers presented here, especially for the EDS.

7. Summary and conclusion

We have presented a data-driven method for simulating deconvolved and noise-free galaxies with morphologies more realistic and complex than purely analytic Sérsic profiles. The proposed approach is based on a combination of deep generative neural networks trained on observations, which allows one to generate realistic galaxies while preserving a control of the global shape of the surface brightness profiles. We have shown that the structural parameters of the generated galaxies are recovered with similar accuracy compared to that derived for analytic profiles. Our proposed approach, although around five times slower than an analytic simulation can be used to generate realistic simulations for future missions and experiments, and therefore calibrate algorithms under more realistic conditions.

We have used this new framework to establish the first forecasts on the number of galaxies for which Euclid will be able to provide resolved morphological structure beyond Sérsic profiles. We find that Euclid will resolve the internal structure of around 250 million galaxies. This corresponds to a 50% stellar mass complete sample above 10^{10.6} (10^{9.6}) at a redshift $z = 0.5$ for the EWS (EDS). This is a first estimation of the capabilities of Euclid for estimating galaxy morphologies, which are a key ingredient for a variety of galaxy evolution-related science cases.

Looking ahead, there is an ongoing effort of the authors to adapt the VAE to work in a multi-band mode, which will enable the generation of galaxies in the two infrared bands of the Euclid near-infrared imager. We also plan to train a flow on different sets of parameters. Our method can, for example, be conditioned on the orientation and the environment of galaxies to take into account gravitational shear effects. We could also condition our flow on the redshift and initial mass function in order to find their impact on the evolution of structures.

Acknowledgements. We thank the IAC where the first author was in long term visit during the production of this paper, with a special thanks to the TRACES team for their support. We would also like to thank the Direction Information de l’Observatoire (DIO) of the Paris Meudon Observatory for the management and support of the GPU we used to train our deep learning models. We also thank the Centre National d’Études Spatiales (CNES) and the Centre National de la Recherche Scientifique (CNRS) for the financial support of the PhD in which this study took place. This work has made use of CosmoHub. CosmoHub has been developed by the Port d’Informació Científica (PIC), maintained through a collaboration of the Institut de Física d’Altes Energies (IFAE) and the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) and the Institute of Space Sciences (CSIC and IEEC), and was partially funded by the “Plan Estatal de Investigación Científica y Técnica y de Innovación” program of the Spanish government. The Euclid Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of Euclid, in particular the Academy of Finland, the Agenzia Spaziale Italiana, the Belgian Science Policy, the Canadian Euclid Consortium, the Centre National d’Études Spatiales, the Deutsches Zentrum für Luft- und Raumfahrt, the Danish Space Research Institute, the Fundação para a Ciência e a Tecnologia, the Ministerio de Economía y Competitividad, the National Aeronautics and Space Administration, the Netherlands Onderzoekscollege Voor Astronomie, the Norwegian Space Agency, the Romanian Space Agency, the State Secretariat for Education, Research and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. A complete and detailed list is available on the Euclid web site (http://www.euclid-ec.org). Softwares: Astropy (Astropy Collaboration 2013, 2018), GalSim (Rowe et al. 2015), IPython (Perez & Granger 2007), Jupyter (Kluyver et al. 2016), Matplotlib (Hunter 2007), NumPy (Hernández et al. 2020), TensorFlow (Abadi et al. 2016), TensorFlow Probability (Dillon et al. 2017).

References

Arjovsky, M., Chintala, S., & Bottou, L. 2017, Wasserstein GAN
A90, page 16 of 21
Appendix A: More illustrations of the capabilities of our model.

Fig. A.1. Simulation by our model of bulge components with radius smaller than 0\′′1 (i.e. smaller than one pixel). These bulge components are at the end of the Euclid Flagship radii distribution (Fig. 1) outside the COSMOS training domain. As can be seen, our model is able to extrapolate from those cases. Because the object is not resolved, it is almost purely the VIS PSF.

Fig. A.2. Double-Sérsic component galaxies currently used in the Euclid Consortium.
Fig. A.3. Random examples of galaxies considered as having structure. The stamps are cut at twice the effective radius.
Fig. A.4. Random examples of galaxies considered as having no structure. The stamps are cut at twice the effective radius.
Appendix B: Detailed architectures of the FVAE

Fig. B.1. Schematic architecture of the VAE. Each convolution is followed by a ReLU activation and a dropout.
Fig. B.2. Schematic architecture of the regressive flow. Each dense layer is followed by a ReLU activation.