

## Euclid preparation

### LXVII. Deep learning true galaxy morphologies for weak lensing shear bias calibration

Euclid Collaboration: B. Csizi<sup>1,\*</sup>, T. Schrabback<sup>1</sup>, S. Grandis<sup>1</sup>, H. Hoekstra<sup>2</sup>, H. Jansen<sup>1</sup>, L. Linke<sup>1</sup>, G. Congedo<sup>3</sup>, A. N. Taylor<sup>3</sup>, A. Amara<sup>4</sup>, S. Andreon<sup>5</sup>, C. Baccigalupi<sup>6,7,8,9</sup>, M. Baldi<sup>10,11,12</sup>, S. Bardelli<sup>11</sup>, P. Battaglia<sup>11</sup>, R. Bender<sup>13,14</sup>, C. Bodendorf<sup>13</sup>, D. Bonino<sup>15</sup>, E. Branchini<sup>16,17,5</sup>, M. Brescia<sup>18,19,20</sup>, J. Brinchmann<sup>21,22</sup>, S. Camera<sup>23,24,15</sup>, V. Capobianco<sup>15</sup>, C. Carbone<sup>25</sup>, J. Carretero<sup>26,27</sup>, S. Casas<sup>28</sup>, F. J. Castander<sup>29,30</sup>, M. Castellano<sup>31</sup>, G. Castignani<sup>11</sup>, S. Cavuoti<sup>19,20</sup>, A. Cimatti<sup>32</sup>, C. Colodro-Conde<sup>33</sup>, C. J. Conselice<sup>34</sup>, L. Conversi<sup>35,36</sup>, Y. Copin<sup>37</sup>, F. Courbin<sup>38,39,40</sup>, H. M. Courtois<sup>41</sup>, M. Cropper<sup>42</sup>, A. Da Silva<sup>43,44</sup>, H. Degaudenzi<sup>45</sup>, G. De Lucia<sup>7</sup>, J. Dinis<sup>43,44</sup>, M. Douspis<sup>46</sup>, F. Dubath<sup>45</sup>, X. Dupac<sup>36</sup>, S. Dusini<sup>47</sup>, M. Farina<sup>48</sup>, S. Farrens<sup>49</sup>, F. Faustini<sup>50,31</sup>, S. Ferriol<sup>37</sup>, S. Fotopoulou<sup>51</sup>, M. Frailis<sup>7</sup>, E. Franceschi<sup>11</sup>, S. Galeotta<sup>7</sup>, B. Gillis<sup>3</sup>, C. Giocoli<sup>11,52</sup>, A. Grazian<sup>53</sup>, F. Grupp<sup>13,14</sup>, L. Guzzo<sup>54,5</sup>, S. V. H. Haugan<sup>55</sup>, W. Holmes<sup>56</sup>, I. Hook<sup>57</sup>, F. Hormuth<sup>58</sup>, A. Hornstrup<sup>59,60</sup>, P. Hudelot<sup>61</sup>, S. Ilić<sup>62,63</sup>, K. Jahnke<sup>64</sup>, M. Jhabvala<sup>65</sup>, B. Joachimi<sup>66</sup>, E. Keihänen<sup>67</sup>, S. Kermiche<sup>68</sup>, A. Kiessling<sup>56</sup>, M. Kilbinger<sup>49</sup>, B. Kubik<sup>37</sup>, K. Kuijken<sup>2</sup>, M. Kümmel<sup>14</sup>, M. Kunz<sup>69</sup>, H. Kurki-Suonio<sup>70,71</sup>, S. Ligori<sup>15</sup>, P. B. Lilje<sup>55</sup>, V. Lindholm<sup>70,71</sup>, I. Lloro<sup>72</sup>, D. Maino<sup>54,25,73</sup>, E. Maiorano<sup>11</sup>, O. Mansutti<sup>7</sup>, S. Marcin<sup>74</sup>, O. Marggraf<sup>75</sup>, K. Markovic<sup>76</sup>, M. Martinelli<sup>31,76</sup>, N. Martinet<sup>77</sup>, F. Marulli<sup>78,11,12</sup>, R. Massey<sup>79</sup>, E. Medinaceli<sup>11</sup>, S. Mei<sup>80</sup>, M. Melchior<sup>74</sup>, Y. Mellier<sup>81,61</sup>, M. Meneghetti<sup>11,12</sup>, G. Meylan<sup>38</sup>, M. Moresco<sup>78,11</sup>, L. Moscardini<sup>78,11,12</sup>, S.-M. Niemi<sup>82</sup>, C. Padilla<sup>83</sup>, S. Paltani<sup>45</sup>, F. Pasian<sup>7</sup>, K. Pedersen<sup>84</sup>, V. Pettorino<sup>82</sup>, S. Pires<sup>49</sup>, G. Polenta<sup>50</sup>, M. Poncet<sup>85</sup>, L. A. Popa<sup>86</sup>, F. Raison<sup>13</sup>, A. Renzi<sup>87,47</sup>, J. Rhodes<sup>56</sup>, G. Riccio<sup>19</sup>, E. Romelli<sup>7</sup>, M. Roncarelli<sup>11</sup>, E. Rossetti<sup>10</sup>, R. Saglia<sup>14,13</sup>, Z. Sakr<sup>88,63,89</sup>, A. G. Sánchez<sup>13</sup>, B. Sartoris<sup>14,7</sup>, P. Schneider<sup>75</sup>, A. Secroun<sup>68</sup>, G. Seidel<sup>64</sup>, S. Serrano<sup>30,90,29</sup>, C. Sirignano<sup>87,47</sup>, G. Sirri<sup>12</sup>, L. Stanco<sup>47</sup>, J. Steinwagner<sup>13</sup>, P. Tallada-Crespi<sup>26,27</sup>, D. Tavagnacco<sup>7</sup>, H. I. Teplitz<sup>91</sup>, I. Tereno<sup>43,92</sup>, R. Toledo-Moreo<sup>93</sup>, F. Torradeflot<sup>27,26</sup>, I. Tutusaus<sup>63</sup>, E. A. Valentijn<sup>94</sup>, L. Valenziano<sup>11,95</sup>, T. Vassallo<sup>14,7</sup>, G. Verdoes Kleijn<sup>94</sup>, A. Veropalumbo<sup>5,17,96</sup>, Y. Wang<sup>91</sup>, J. Weller<sup>14,13</sup>, G. Zamorani<sup>11</sup>, E. Zucca<sup>11</sup>, A. Biviano<sup>7,6</sup>, M. Bolzonella<sup>11</sup>, E. Bozzo<sup>45</sup>, C. Burigana<sup>97,95</sup>, M. Calabrese<sup>98,25</sup>, D. Di Ferdinando<sup>12</sup>, J. A. Escartin Vigo<sup>13</sup>, R. Farinelli<sup>11</sup>, J. Gracia-Carpio<sup>13</sup>, S. Matthew<sup>3</sup>, N. Mauri<sup>32,12</sup>, A. Pezzotta<sup>13</sup>, M. Pöntinen<sup>70</sup>, V. Scottez<sup>81,99</sup>, M. Tenti<sup>12</sup>, M. Viel<sup>6,7,9,8,100</sup>, M. Wiesmann<sup>55</sup>, Y. Akrami<sup>101,102</sup>, V. Allevalo<sup>19</sup>, S. Anselmi<sup>47,87,103</sup>, M. Archidiacono<sup>54,73</sup>, F. Atrio-Barandela<sup>104</sup>, M. Ballardini<sup>105,11,106</sup>, A. Blanchard<sup>63</sup>, L. Blot<sup>107,103</sup>, S. Borgani<sup>108,6,7,8</sup>, S. Bruton<sup>109</sup>, R. Cabanac<sup>63</sup>, A. Calabro<sup>31</sup>, G. Cañas-Herrera<sup>82,110</sup>, A. Cappi<sup>11,111</sup>, F. Caro<sup>31</sup>, C. S. Carvalho<sup>92</sup>, T. Castro<sup>7,8,6,100</sup>, K. C. Chambers<sup>112</sup>, S. Contarini<sup>113</sup>, A. R. Cooray<sup>113</sup>, G. Desprez<sup>114</sup>, A. Díaz-Sánchez<sup>115</sup>, J. J. Diaz<sup>116</sup>, S. Di Domizio<sup>16,17</sup>, H. Dole<sup>46</sup>, S. Escoffier<sup>68</sup>, A. G. Ferrari<sup>32,12</sup>, P. G. Ferreira<sup>117</sup>, I. Ferrero<sup>55</sup>, A. Finoguenov<sup>70</sup>, A. Fontana<sup>31</sup>, F. Fornari<sup>95</sup>, L. Gabarra<sup>117</sup>, K. Ganga<sup>80</sup>, J. García-Bellido<sup>101</sup>, T. Gasparetto<sup>7</sup>, E. Gaztanaga<sup>29,30,118</sup>, F. Giacomini<sup>12</sup>, F. Gianotti<sup>11</sup>, G. Gozalias<sup>119,70</sup>, C. M. Gutierrez<sup>120</sup>, A. Hall<sup>3</sup>, H. Hildebrandt<sup>121</sup>, J. Hjorth<sup>84</sup>, A. Jimenez Muñoz<sup>122</sup>, S. Joudaki<sup>118</sup>, J. J. E. Kajava<sup>123,124</sup>, V. Kansal<sup>125,126</sup>, D. Karagiannis<sup>127,128</sup>, C. C. Kirkpatrick<sup>67</sup>, A. M. C. Le Brun<sup>103</sup>, J. Le Graet<sup>68</sup>, L. Legrand<sup>129</sup>, J. Lesgourgues<sup>28</sup>, T. I. Liaudat<sup>130</sup>, A. Loureiro<sup>131,132</sup>, J. Macias-Perez<sup>122</sup>, G. Maggio<sup>7</sup>, M. Magliocchetti<sup>48</sup>, C. Mancini<sup>25</sup>, F. Mannucci<sup>133</sup>, R. Maoli<sup>134,31</sup>, J. Martín-Fleitas<sup>135</sup>, C. J. A. P. Martins<sup>136,21</sup>, L. Maurin<sup>46</sup>, R. B. Metcalf<sup>78,11</sup>, M. Miluzio<sup>36,137</sup>, P. Monaco<sup>108,7,8,6</sup>, A. Montoro<sup>29,30</sup>, A. Mora<sup>135</sup>, C. Moretti<sup>9,100,7,6,8</sup>, G. Morgante<sup>11</sup>, Nicholas A. Walton<sup>138</sup>, L. Pagano<sup>105,106</sup>, L. Patrizii<sup>12</sup>, V. Popa<sup>86</sup>, D. Potter<sup>139</sup>, I. Risso<sup>140</sup>, P.-F. Rocci<sup>46</sup>, M. Sahlén<sup>141</sup>, E. Sarpa<sup>9,100,8</sup>, A. Schneider<sup>139</sup>, M. Sereno<sup>11,12</sup>, P. Simon<sup>75</sup>, A. Spurio Mancini<sup>142,42</sup>, J. Stadel<sup>139</sup>, K. Tanidis<sup>117</sup>, C. Tao<sup>68</sup>, N. Tessore<sup>66</sup>, G. Testera<sup>17</sup>, R. Teyssier<sup>143</sup>, S. Toft<sup>60,144</sup>, S. Tosi<sup>16,17</sup>, A. Troja<sup>87,47</sup>, M. Tucci<sup>45</sup>, C. Valieri<sup>12</sup>, J. Valiviita<sup>70,71</sup>, D. Vergani<sup>11</sup>, G. Verza<sup>146,147</sup>, and P. Vielzeuf<sup>68</sup>

(Affiliations can be found after the references)

Received 5 September 2024 / Accepted 21 January 2025

\* Corresponding author; benjamin.csizi@uibk.ac.at

## ABSTRACT

To date, galaxy image simulations for weak lensing surveys usually approximate the light profiles of all galaxies as a single or double Sérsic profile, neglecting the influence of galaxy substructures and morphologies deviating from such a simplified parametric characterisation. While this approximation may be sufficient for previous data sets, the stringent cosmic shear calibration requirements and the high quality of the data in the upcoming *Euclid* survey demand a consideration of the effects that realistic galaxy substructures and irregular shapes have on shear measurement biases. Here we present a novel deep learning-based method to create such simulated galaxies directly from *Hubble* Space Telescope (HST) data. We first build and validate a convolutional neural network based on the wavelet scattering transform to learn noise-free representations independent of the point-spread function (PSF) of HST galaxy images. These can be injected into simulations of images from *Euclid*'s optical instrument VIS without introducing noise correlations during PSF convolution or shearing. Then, we demonstrate the generation of new galaxy images by sampling from the model randomly as well as conditionally. In the latter case, we fine-tune the interpolation between latent space vectors of sample galaxies to directly obtain new realistic objects following a specific Sérsic index and half-light radius distribution. Furthermore, we show that the distribution of galaxy structural and morphological parameters of our generative model matches the distribution of the input HST training data, proving the capability of the model to produce realistic shapes. Next, we quantify the cosmic shear bias from complex galaxy shapes in *Euclid*-like simulations by comparing the shear measurement biases between a sample of model objects and their best-fit double-Sérsic counterparts, thereby creating two separate branches that only differ in the complexity of their shapes. Using the Kaiser, Squires, and Broadhurst shape measurement algorithm, we find a multiplicative bias difference between these branches with realistic morphologies and parametric profiles on the order of  $(6.9 \pm 0.6) \times 10^{-3}$  for a realistic magnitude-Sérsic index distribution. Moreover, we find clear detection bias differences between full image scenes simulated with parametric and realistic galaxies, leading to a bias difference of  $(4.0 \pm 0.9) \times 10^{-3}$  independent of the shape measurement method. This makes complex morphology relevant for stage IV weak lensing surveys, exceeding the full error budget of the Euclid Wide Survey ( $\Delta\mu_{1,2} < 2 \times 10^3$ ).

**Key words.** gravitational lensing: weak – methods: data analysis – techniques: image processing – galaxies: fundamental parameters

## 1. Introduction

Identifying the origin of the accelerated expansion of the Universe by constraining the dark energy equation of state parameter  $w$  is one of the most challenging and pressing open questions in cosmology. To tackle the task of unravelling the characteristics of dark energy, several next-generation surveys such as *Euclid* (Laureijs et al. 2011; Euclid Collaboration: Mellier et al. 2025c), the *Nancy Grace Roman* Telescope (Spergel et al. 2015), and the Legacy Survey of Space and Time at the *Vera C. Rubin* Observatory (Ivezic et al. 2019) will need to measure weak lensing (WL) image distortions at extremely high accuracy. These distortions have been imprinted on the observed shapes of distant galaxies by the gravitational fields of the foreground cosmic large-scale structure. Such measurements call for precise calibrations to meet the tight requirements. Detailed and realistic image simulations are hence a key ingredient for the latest generation of weak lensing surveys, as they allow for the cosmic shear measurement methods that will be applied to be calibrated, and thus the full predictive power of the data for the inference of cosmological parameters can be leveraged.

There have been many efforts to quantify the effect of the properties of image simulations on shape measurements (see for example Hoekstra et al. 2017; Hernández-Martín et al. 2020) and to subsequently improve the simulation quality to more closely match the real observations concerning, for example, galaxy number densities, morphological properties, redshifts and magnitudes, and instrumental or atmospheric effects (Mandelbaum et al. 2018; Kannawadi et al. 2019; MacCrann et al. 2022; Li et al. 2022, 2023). Until now, however, the galaxy morphologies included in these simulations have lacked the complexity and irregularity of real galaxies, or they have relied on injecting real high-resolution *Hubble* Space Telescope (HST) observations into image simulations by adjusting the imaging properties to match the target survey's instrument and observing properties (Li et al. 2022). The latter method is applicable to ground-based stage III experiments but has been shown to cause strong issues in the shear estimation

due to noise correlations (Euclid Collaboration: Scognamiglio et al. 39 2025). For the former simulation framework, the light distribution of an object is simulated as a single analytic Sérsic profile or as a two-component model consisting of the sum of a Sérsic bulge and an either Sérsic or exponential disc. While stage III surveys, such as the Dark Energy Survey (Dark Energy Survey Collaboration 2016), the Hyper-Suprime-Cam Survey (Aihara et al. 2018), and the Kilo-Degree Survey (de Jong et al. 2013), were able to rely on this simplification due to the lower shape measurement bias requirements, novel stage IV projects such as *Euclid* will have to account for the influence of galaxy substructures on the cosmic shear analysis. In weak lensing, cosmic shear is measured from spatial correlations in galaxy ellipticities using large source samples, thus requiring simulations that accurately reproduce the shapes of real objects to calibrate the measurement. Previous attempts at creating more realistic simulations include the emulation of lensing data from HST images (Mandelbaum et al. 2012a, 2018), the generation of galaxies via shapelet functions (Massey et al. 2004), and the sampling of different deep learning models trained on real data (Lanusse et al. 2021). Results from the GREAT3 challenge (Mandelbaum et al. 2015) showed a percent-level bias difference with respect to parametric galaxy morphologies for most shape measurement methods using HST emulation, which was more prominent for simulated space-based data due to the high resolution and small pixel scales, while the Shear Testing Programme 2 (Massey et al. 2007a) revealed a sub-percent bias difference using the shapelet galaxies, albeit at ground-based pixel scales and point-spread functions (PSFs). The analysis of the full Euclid Wide Survey requires a shear calibration accuracy to better than 0.2% (Cropper et al. 2013), making it imperative to account for the impact of galaxy morphologies. The effect stems from the fact that second-order moments of galaxy light profiles, which are commonly used by shape measurement algorithms, are coupled with higher-order moments by the shear (Massey et al. 2007b; Zhang & Komatsu 2011; Bernstein 2010). These higher-order moments are, however, dominated by morphologies deviating from simple parametrisations. Previous estimates determined that the Euclid Wide

Survey will be able to resolve substructures down to a surface brightness of  $22.5 \text{ mag arcsec}^{-2}$  and down to  $24.9 \text{ mag arcsec}^{-2}$  in the Deep Fields (Euclid Collaboration: Bretonnière et al. 2022a). This results in approximately 250 million galaxies with resolved morphologies over the entire mission lifetime. As there does not exist a well-established parametric model of galaxy morphologies that is more realistic than a two-component description, aside from simulating a substructure using parametric shapes that are disturbed with knots in the light profile (Sheldon & Huff 2017), a different path is needed to simulate galaxy images.

With deep learning techniques on the rise for tasks of computer vision, such as image generation or classification, capitalising on this growing research field for the aforementioned goal is a promising approach. Spindler et al. (2021), Lanusse et al. (2021), Smith et al. (2022), and Holzschuh et al. (2022) have shown how variational autoencoders (VAEs; Kingma & Welling 2013) and generative adversarial networks (GANs; Goodfellow et al. 2014) can be applied to generate galaxy images with high-resolution training data, a method that has since also been applied for forecasts on galaxy morphologies with *Euclid* (Euclid Collaboration: Bretonnière et al. 2022a). Aside from VAEs and GANs, diffusion models have gained popularity as a powerful image generation method (Ho et al. 2020), but they are accompanied by increased complexity. Aside from image generation, latent space machine learning models have also found other applications, for example, galaxy classification (Cheng et al. 2021), discovery of strong lenses (Cheng et al. 2020), modelling of galactic dust extinction (Thorne et al. 2021), and dimensionality reduction of galaxy spectra (Portillo et al. 2020).

In this work, we propose a new generative model that allows for noise-free and PSF-independent reconstruction of real galaxy images and is able to generate a distribution of new objects following an input distribution of morphological parameters. While it is mostly interesting to generate new images due to the necessity of including  $10^7$ – $10^9$  galaxies in order to reach the precision of the *Euclid* shear calibration requirements (Euclid Collaboration: Martinet et al. 2019), the noise-free reconstruction of inputs with minimal residuals can also be relevant for the *Euclidisation* procedure proposed by Euclid Collaboration: Scognamiglio et al. (2025a), where the authors presented a method to convert HST images of isolated galaxies (hereafter, postage stamps) into *Euclid* observations. Such a pipeline requires PSF (de)convolution, shearing, and down-sampling steps on a noisy image, which gives rise to noise correlations that impact the shape measurement (Gurvich & Mandelbaum 2016). Moreover, our model can be applied to future high signal-to-noise ratio observations in the Euclid Deep Fields in order to obtain a larger sample of observed but noise-free galaxies that can be injected into simulations without introducing correlated noise or relying on the output of a black-box machine learning model.

In this paper, we perform a proof-of-concept study on the calibration of biases from a complex galaxy morphology for the *Euclid* mission. First, we present the architecture and training data of our deep learning model based on the wavelet scattering transform after summarising the weak lensing formalism and the theoretical background on galaxy morphological statistics in Sects. 2 and 3. We then compare our reconstructed images with their inputs in terms of their structural parameters and a set of morphological statistics in Sect. 4. Afterwards, we describe our method for generating new galaxies according to either an input distribution of Sérsic indices or their overall visual characteristics. Finally, in Sect. 6, we quantify the shear bias introduced

by galaxy substructures through the comparison of simulations of *Euclid* VIS-like postage stamps with samples from our model and their respective best-fit parametric models before concluding in Sect. 7.

## 2. Weak lensing formalism

### 2.1. Definitions

Weak lensing describes the coherent, statistical distortion of objects in the Universe due to gravitational deflection by mass density fluctuations along the line of sight, see Massey et al. (2010), Kilbinger (2015), and Mandelbaum (2018) for reviews. The local differential mapping between lensed and unlensed coordinates can be described via the Jacobian matrix, whose elements are

$$\mathcal{A}_{ij} = \delta_{ij} - \partial_i \partial_j \Psi, \quad (1)$$

where  $\Psi$  denotes the lensing potential,  $\delta_{ij}$  is the Kronecker delta, and  $\partial_{i,j}$  are the derivatives along the respective coordinate of the lens plane. This matrix can be parametrised by the introduction of a convergence  $\kappa$  and a two-component shear  $\gamma = (\gamma_1, \gamma_2)$ , to simplify the Jacobian to

$$\mathcal{A}_{ij} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}. \quad (2)$$

Usually, the shear is expressed as a complex number  $\gamma = \gamma_1 + i\gamma_2$  and in terms of the reduced shear

$$g = g_1 + ig_2 = \frac{\gamma}{1 - \kappa}, \quad (3)$$

which is advantageous due to this quantity's connection to the ellipticity  $\epsilon$ , making it directly measurable. In the absence of preferential intrinsic galaxy orientations, the expectation value of the intrinsic ellipticity  $\langle \epsilon_s \rangle$  vanishes, leading to

$$g = \langle \epsilon_o \rangle, \quad (4)$$

where  $\langle \epsilon_o \rangle$  is the expectation value of the observed ellipticity. This simple relation shows that a measurement of the ellipticity of galaxies allows measurement of the shear when averaging over a sufficiently large number of sources.

### 2.2. Shear measurement

There exist several methods for measuring the shear from a galaxy image. It can be done directly from galaxy brightness moments, for instance with the Kaiser, Squires, and Broadhurst (KSB) shape measurement algorithm (Kaiser et al. 1995; Hoekstra et al. 1998), or using forward-modelling of a light distribution and then fitting it to the data by maximising a likelihood function, such as IM3SHAPE (Zuntz et al. 2013), *lensfit* (Miller et al. 2007), or LENSIMC (Euclid Collaboration: Congedo et al. 2024b). Moreover, METACALIBRATION (Sheldon & Huff 2017) has been applied with both of these types of methods to get, in principle, model bias-corrected estimates of the shear signal. We measure galaxy shapes in this work using the KSB implementation from the HSM module of the GalSim package (Hirata & Seljak 2003; Rowe et al. 2015). Therein, second-order brightness moments

$$Q_{ij} = \frac{\int I(\theta) W(\theta) \theta_i \theta_j d^2\theta}{\int I(\theta) W(\theta) d^2\theta} \quad (5)$$

are used to infer the complex ellipticity

$$\epsilon_0 = \epsilon_1 + i\epsilon_2 = \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22}}. \quad (6)$$

Here,  $\theta = (\theta_1, \theta_2)^T$  is the position vector relative to the object centre, which is defined such that the weighted first-order brightness moment vanishes,  $I(\theta)$  is the image light distribution, and  $W(\theta)$  is an arbitrary weight function, which is usually a Gaussian. Within the KSB formalism the ellipticity is then corrected for the impact of the point-spread function using further brightness moments and measurements of stars.

Model fitting algorithms such as LENSMC on the other hand employ a Bayesian approach to forward-model the pixel data and then estimating the ellipticity as the mean of a posterior distribution

$$\hat{\epsilon} = \int \epsilon p(\epsilon|\mathbf{D}) d\epsilon \quad (7)$$

by sampling the galaxy model parameter space, for instance with Markov-Chain Monte-Carlo (MCMC), and marginalising over nuisance parameters. Here,  $p(\epsilon|\mathbf{D})$  is the ellipticity marginal posterior on the pixel data  $\mathbf{D} = I(\theta)$  given by

$$p(\epsilon|\mathbf{D}) = \frac{1}{p(\mathbf{D})} \int p(\mathbf{D}|\epsilon, \xi, \phi) p(\epsilon, \xi, \phi) d\xi d\phi, \quad (8)$$

with  $p(\mathbf{D})$  being the marginal likelihood and  $\xi$  and  $\phi$  as the intrinsic and linear nuisance parameters (Euclid Collaboration: Congedo et al. 2024b).

### 2.3. Bias calibration

Shear estimators are affected by a range of different bias sources, for example from selection biases or PSF correction errors (Bernstein & Jarvis 2002; Hirata & Seljak 2003; Fenech Conti et al. 2017). In linear approximation, the shear bias, defined as the difference between an observed reduced shear  $g_i^{\text{obs}}$  and the true reduced shear  $g_i^{\text{true}}$  with  $i = 1, 2$  (assuming no mixing between the two components), can be written as

$$g_i^{\text{obs}} - g_i^{\text{true}} = \mu_i g_i^{\text{true}} + c_i + n_i, \quad (9)$$

with  $\mu_i$  and  $c_i$  being the multiplicative and additive shear biases, respectively, and  $n_i$  as a statistical noise component. The indices  $i = 1, 2$  denote the shear component along the Cartesian axes and along the  $\pi/4$  diagonals, respectively. Alternatively, the bias can be defined via a spin-2 equation with spin-0 and spin-4 multiplicative biases, which then facilitates an inclusion of non-linear shear bias terms, if required (Kitching & Deshpande 2022). The magnitude of the multiplicative bias is generally a function of galaxy morphological parameters (see, e.g., Hernández-Martín et al. 2020), galaxy signal-to-noise ratio (Schrabback et al. 2010; Hoekstra et al. 2015), redshift (Kannawadi et al. 2019), and blending (MacCrann et al. 2022), therefore requiring a redshift tomography-dependent calibration.

To mitigate the effect of biased galaxy shape estimation on shear analysis (and in consequence cosmological inference), several techniques can be exploited. On one side, methods such as shape and pixel noise cancellation (Massey et al. 2007b; Jansen et al. 2024) have been shown to be able to efficiently scale down the necessary simulation volume for shear calibration, which in return is used to correct the survey measurements for the determined bias. Other methods such as METACALIBRATION

have been employed successfully to remove noise biases, model biases and selection effects from the shape measurement directly during the measurement process (Sheldon & Huff 2017; Sheldon et al. 2020). Nevertheless, detailed simulated images are needed due to blending, detection (Hoekstra et al. 2021), and redshift blending (MacCrann et al. 2022; Li et al. 2023). In this work, we will focus on the influence of complex galaxy morphologies on the shear measurement bias. This is a model bias that enters the shape measurement due to inaccuracies of the underlying model (either the fitted profiles or the estimation of ellipticities via moments), and thus it depends on the applied method. While this specific model bias has been previously ignored for ground-based surveys with lower resolution imaging, where substructure is not well resolved, this assumption does not necessarily hold for a space-based mission such as *Euclid*, given its unprecedented precision and survey area.

## 3. Galaxy morphologies

### 3.1. Sérsic profiles

The overall complex structure of galaxies is a product of their complicated evolution history, see Conselice (2014) for a review. Nevertheless, the observations of a majority of galaxies at currently feasible resolutions for large ground-based WL surveys can be well approximated using a simple analytic prescription. The most common parametric form of a galaxy-like light distribution is the Sérsic profile

$$I(r) \propto \exp \left[ -b_n \left( \frac{r}{r_e} \right)^{1/n} \right], \quad (10)$$

where  $n$  is the Sérsic index,  $r_e$  is the half-light radius, and  $b_n$  is a scaling factor that depends on  $n$  (Sérsic 1963). With this model, galaxy light profiles can be simulated as either a single Sérsic or a double Sérsic profile consisting of separate bulge and disc components. Simulating galaxies in this way is advantageous due to the simplicity and existing measurements of the model parameters across previous survey areas, which enables a simple yet realistic generation of simulated footprints.

### 3.2. Morphological statistics

*Euclid*'s VIS instrument (Cropper et al. 2016; Euclid Collaboration: Cropper et al. 2024) will observe billions of galaxies over  $14\,000 \text{ deg}^2$ , with many of them covering only a few pixels given the  $0''.1$  pixel scale. A large subset will have shapes that can be well approximated by a Sérsic parametrisation, but a non-negligible portion of the sample will also display structural features such as irregularities, clumps, spiral arms, or tidal streams. Moreover, the fraction of such peculiar galaxies changes gradually with redshift, with high- $z$  objects ( $z > 1.2$ ) showing irregularities more commonly, especially since optical observations show rest-frame UV structures for high- $z$  galaxies, which are dominated by star-forming regions (Abraham et al. 1996; Conselice et al. 2005; Bundy et al. 2005). Even though these objects will have mostly low signal-to-noise ratios, parametric fits can still result in large residuals for a substantial number of objects. Given that the *Euclid* WL sample will include galaxies with redshifts up to  $z \approx 2$  (Euclid Collaboration: Ilbert et al. 2021), accurate galaxy shapes will accordingly be even more relevant.

Statistical proxies for disturbed morphologies can be evaluated on input HST data and the deep learning model output

to validate its capability to capture modes that deviate from smooth structures. [Hackstein et al. \(2023\)](#) summarised a set of such proxies to estimate the power of galaxy image generators. For instance, the MID statistics (multi-mode, intensity, deviation) by [Freeman et al. \(2013\)](#) provide an estimate of peculiar features of a galaxy morphology by tracing the existence and intensity ratio of multiple nuclei as well as the deviation from simple elliptical representations. Similarly, the Concentration, Asymmetry & Smoothness (CAS) morphology indicators ([Conselice et al. 2000](#); [Conselice 2003](#)) trace irregular shapes by defining the following set of estimators:

$$C = 5 \log_{10} \left( \frac{r_{80}}{r_{20}} \right); \quad (11)$$

$$A = \sum_{i,j} \frac{|I_0(\theta_i, \theta_j) - I_\pi(\theta_i, \theta_j)|}{|I_0(\theta_i, \theta_j)|} - |B_\pi|; \quad (12)$$

$$S = \sum_{i,j} \frac{|I_0(\theta_i, \theta_j) - I_S(\theta_i, \theta_j)|}{|I_0(\theta_i, \theta_j)|} - |B_S|. \quad (13)$$

The concentration  $C$  measures the bulge concentration by relating the radii  $r_{80}, r_{20}$  of apertures within which 80% and 20% of the total flux are located. Additionally, the asymmetry parameter  $A$  and the smoothness  $S$  quantify the rotational symmetry with respect to the flux, and the magnitude of small-scale structures, respectively. Here,  $I_0(\theta_i, \theta_j)$  is the galaxy image intensity at pixels  $\theta = (\theta_i, \theta_j)^T$ ,  $I_\pi(\theta_i, \theta_j)$  is the same image rotated by  $\pi$  around the image centre,  $|B_\pi|$  is the average asymmetry of the rotated image background,  $I_S(\theta_i, \theta_j)$  is the image smoothed by a boxcar filter, and  $|B_S|$  is the average smoothness of the background.

Furthermore, another such statistic, the Gini coefficient, is sensitive to the intensity concentration in a compact component of the light profile and can be calculated with

$$G = \frac{1}{k(k-1)\bar{I}_0} \sum_i^k (2i-k-1) I_0(\theta_i). \quad (14)$$

The value  $I_0(\theta_i)$  is the  $i$ -th pixel value of the individual galaxy, here with pixels sorted by increasing intensity, and  $\bar{I}_0$  is the mean over all  $k$  pixels ([Abraham et al. 2003](#)). We also calculated the  $M_{20}$  coefficient

$$M_{20} = \log_{10} \left( \frac{\sum_i Q_i}{Q_{\text{tot}}} \right), \quad (15)$$

where  $Q_{\text{tot}}$  is the second-order moment of the total galaxy light distribution (sum over all pixel fluxes multiplied by their squared distance to the image centre) and  $\sum_i Q_i$  are the second-order moments summed over only the brightest 20% of pixels, so  $\sum_i I_0(\theta_i) < 0.2 I_0(\theta)$ . This parameter is thereby able to indicate merger signatures and clumpiness ([Lotz et al. 2004](#)).

## 4. Reconstruction of HST data

### 4.1. The wavelet scattering transform

One main attribute of deep learning models is a process of dimensionality reduction. To learn the distribution of training images, these models usually compress the 2D array of image pixels into a latent vector  $\mathbf{z}$ , which is then later expanded to image size using convolutional layers. While these latent representations within VAEs and GANs effectively constitute a black-box, where usually no physical meaning can be attributed to the

latent variables, we employ an image compression method that is based on the wavelet scattering transform (WST; [Mallat 2012](#)) as an encoder for the network. Such convolutional neural networks (CNNs) have previously been proposed by [Bruna & Mallat \(2013\)](#) and applied to common deep learning test data sets by [Angles & Mallat \(2018\)](#). This operation is useful, due to the transform's ability to capture morphological information. With this mathematically motivated latent space, we can later on also sample galaxies by clustering them according to their wavelet scattering coefficients. Thus we avoid learning a latent variable model or encountering the typical limitations of other generative models, such as mismatches between aggregate posterior and prior in VAEs ([Tomczak & Welling 2018](#)) or mode collapse for GANs ([Salimans et al. 2016](#)).

The wavelet scattering transform is an operation that applies a set of convolutions by dilated and rotated wavelet filters  $\psi_\lambda$  to a 2D array. A family  $\{\psi_\lambda\}_{j,l}$  of such filters is specified by a dyadic sequence of scales  $2^j$  with  $j \in \mathbb{Z}, J \geq j > 0$ , a number of rotations with angles  $l, l \in \mathbb{Z}, L \geq l > 0$ , and a rotation operation  $r_l$  on the data  $\mathbf{x}$ :

$$\{\psi_\lambda\}_{j,l} = \frac{1}{2^j} \psi_\lambda \left( \frac{r_l^{-1} \mathbf{x}}{2^j} \right). \quad (16)$$

Given an input image  $I_0(\theta)$ , the zeroth-order scattering transform is defined simply as the mean of the input. The first-order coefficients are calculated by convolving the image with the family of wavelet filters  $\{\psi_\lambda\}_{j_1, l_1}$  and then by taking the mean of the modulus of the obtained scattering fields. Similarly, the second-order scattering coefficients are given by the convolution of the first-order fields by another set of wavelets  $\{\psi_\lambda\}_{j_2, l_2}$ :

$$s_0 = \langle I_0(\theta) \rangle; \quad (17)$$

$$s_1^{j_1, l_1} = \langle |I_0(\theta) \star \{\psi_\lambda\}_{j_1, l_1}| \rangle; \quad (18)$$

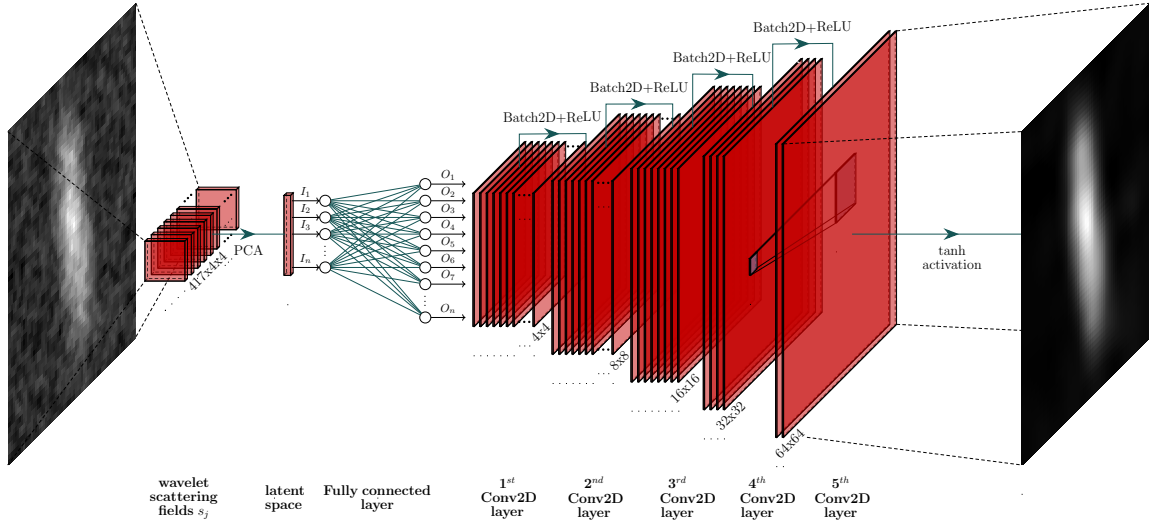
$$s_2^{j_1, l_1, j_2, l_2} = \langle | |I_0(\theta) \star \{\psi_\lambda\}_{j_1, l_1} \star \{\psi_\lambda\}_{j_2, l_2} | | \rangle. \quad (19)$$

Here,  $\star$  designates a convolution. This results in  $J^n L^n + 1$  scattering coefficients, where  $n$  is the maximum order of the scattering transform, which exceeds the target latent space dimension for typical values of  $J = 4, L = 8$ . By averaging over all orientations  $(l_1, l_2)$ , one can reduce this number to  $J + J^2 + 1$  ([Greig et al. 2023](#)). This, however, does not preserve the angular dependence of the wavelet filtering and thus only probes the size scales of the image, which thus reduces the morphological information. We only average over the orientations  $l_1$ , which preserves shape information but discards information on the orientation. Additionally, one can further limit the dimensionality by a factor of  $2^{n-1}$ , by discarding coefficients with  $j_1 \geq j_2$ , as they only contain high-frequency information ([Cheng & Ménard 2021](#)). This is almost exclusively noisy pixel-level information, which will not be highly relevant for *Euclid* with respect to HST, given the resolution difference. For the reduced scattering coefficients of  $I_0(\theta)$ , one thus ultimately obtains

$$s_1^{j_1} = \langle \langle |I_0(\theta) \star \{\psi_\lambda\}_{j_1, l_1}| \rangle \rangle_{l_1}; \quad (20)$$

$$s_2^{j_1, j_2, l_2} = \langle \langle | |I_0(\theta) \star \{\psi_\lambda\}_{j_1, l_1} \star \{\psi_\lambda\}_{j_2, l_2} | | \rangle \rangle_{l_1}, \quad j_2 > j_1. \quad (21)$$

Here,  $\langle \dots \rangle_{l_1}$  denotes an average over all  $l_1$  indices. The choice of  $J$  and  $L$  depends on the size of the images and the scales and angles that need to be probed. In principle, the wavelet scattering transform can be extended to higher orders; this is, however, computationally expensive. [Bruna & Mallat \(2013\)](#) have furthermore shown that the information content is extremely



**Fig. 1.** Architecture of the CNN. Noisy input images are embedded into a latent space vector by performing a PCA on the wavelet scattering fields  $s_2^{j_1, j_2, j_3}$  with  $J = 4, L = 8$ , which is then propagated through one fully connected layer and five convolutional layers, each with a  $5 \times 5$  kernel, batch normalisation and ReLU. The final output of the generative model is produced by a tanh activation function.

small beyond the second order. We compute wavelet scattering coefficients with the `kymatio` package with `pytorch` backend (Andreux et al. 2020; Paszke et al. 2019).

#### 4.2. Training data

Our galaxy image training data set consists of HST images observed in the F814W filter as part of the COSMOS program (Scoville et al. 2007). The `GalSim` package supplies a catalogue of postage stamps of deblended, PSF deconvolved galaxies with magnitudes down to  $m_{AB}^{F814W} \leq 23.5$  from this survey (Leauthaud et al. 2007; Mandelbaum et al. 2012b, 2014). For the sample selection of this data set, some additional cuts were imposed on the COSMOS data to reject objects with contamination from stars or image defects, as well as objects lying in masked regions of the ground-based *BVIz* imaging, to ensure good photometric redshift estimates. The exact cuts can be found in the appendix of Mandelbaum et al. (2014).

We draw these 56 062 galaxies on  $64 \times 64$  images at a pixel scale of  $0''.05$ , that is half of the nominal pixel size of the *Euclid* VIS instrument, and convolve them by a simple Gaussian PSF with  $\sigma = 0''.07$  for training. We can later easily deconvolve the noise-free reconstructions by the same PSF without introducing noise correlations. Next, we discard galaxies with a low signal-to-noise ratio ( $S/N \leq 10$ ), as well as large galaxies that exceed the image size of the postage stamps to avoid truncation. This is done by creating a  $3\sigma$  binary segmentation map and removing objects whose edges do not lie within the stamp. Alternatively, we could increase our postage stamp size, although only to the detriment of much longer training times. Moreover, galaxies that exceed 64 pixels at  $0''.05$  are not relevant for the cosmic shear analysis due to their angular size of  $\geq 3'$ . Such objects do not carry a significant shear signal and can thus be excluded from the analysis. These cuts leave us with 46 720 galaxies, which we divide into 43 520 training images and 3200 test images.

#### 4.3. Network architecture and training

Using the aforementioned WST, we first calculate the scattering fields of our training images up to second order with  $J =$

$4, L = 8$ , resulting in a 3D vector of size  $(1 + LJ + L^2J(J - 1)/2, w/2^J, h/2^J)$  for an image with size  $(w, h)$ . To further reduce the dimensionality we perform a principal component analysis (PCA, see Shlens 2014) to compress the data into a latent space embedding  $\{z_s(I_{0,i})\}$  of input images  $I_{0,i}$  with 64 components, which has been tested before by Angles & Mallat (2018) for deep learning generative models. This proved to be more robust on noisy inputs than training the network directly on the scattering coefficients, as the scattering coefficients first capture morphological information at fixed scales, which mostly do not contain the high-frequency noise, which is then further de-noised by the PCA. This two-step process facilitates the selection of morphological modes only, and produces a latent space that is not arbitrary, but carries significant, correlated information. Later on, however, we also compute the reduced scattering coefficients of the reconstructed noise-free HST data, to classify the objects and facilitate conditional sampling of galaxies.

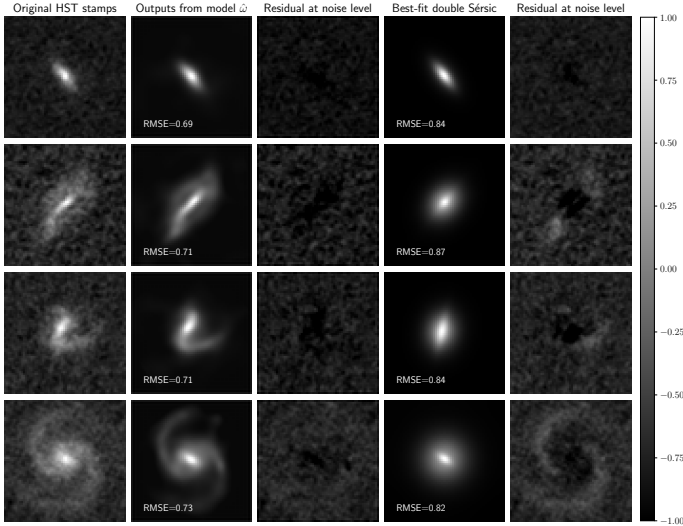
Overall, the model architecture resembles an autoencoder, with a CNN decoder, but a manual encoder to create latent vectors. The CNN itself consists of a linear fully connected layer followed by 5 transposed convolutional layers with batch normalisation and ReLU activation functions (Agarap 2018). These iteratively expand the compressed latent space data via convolutions with  $5 \times 5$  kernels until a final tanh activation to get the generator output. Figure 1 depicts an overview of the CNN architecture. We then also calculate the second-order reduced scattering coefficients  $s_2^{j_1, j_2, j_3}$  (hereafter,  $s_2$ ) of the generated reconstructions, which have previously been shown to be able to trace morphologies in galaxy images (Cheng & Ménard 2021).

We trained our model  $\hat{\omega}$  on the embeddings  $\{z_s(I_{0,i})\}$  by minimising a  $L^1$  loss function such that

$$\hat{\omega}(\{z_s(I_{0,i})\}) = \operatorname{argmin}_{\omega \in \mathcal{G}} \sum_{i=1}^n |I_{0,i} - \omega(\{z_s(I_{0,i})\})|, \quad (22)$$

where  $\mathcal{G}$  represents the class of CNNs with the specified architecture. The generator is trained in batches of size 128 and its hyperparameters are iteratively improved by an ADAM optimiser (Kingma & Ba 2014).

To compare the outputs of our model with the original images, we first perform a qualitative inspection by plotting



**Fig. 2.** Qualitative check of the HST reconstruction from the generator  $\hat{\omega}$ . The left column shows original input galaxies, while the second column displays the output of the generator after training. In the middle we plot the residual at noise level between the first two columns, followed by the best-fit double Sérsic from `pysersic` and the residuals between the original HST stamps and the Sérsic fits. We also display the root mean squared error (RMSE) values between the original image and the reconstruction, for both the parametric and realistic columns.

the samples from the generative model on a few test images next to the corresponding HST galaxy. Additionally, we estimate the best-fit double-Sérsic profile of each galaxy to visualise the gain of our model with respect to common parametric methods. To obtain this fit, we use the `pysersic` package, which employs Bayesian inference methods for this task (Pasha & Miller 2023). After estimating a prior on the fit parameters from the input image, the code finds a posterior distribution by either full MCMC or using stochastic variational inference (SVI, Hoffman et al. 2013). We use the latter (mainly due to its speed), which initially finds the maximum a posteriori (MAP) parameters with SVI and then samples from a narrow Gaussian distribution around these values to obtain the best-fit model.

In Fig. 2, we show this comparison between the input HST data and the realisations from  $\hat{\omega}$  for a selection of galaxies. As one can observe, the CNN is able to easily capture details that deviate from the parametric representation, resulting in an overall reduced residual and a detection of features in the surface brightness that consistently surpasses the capabilities of the employed parametric models. This gain is naturally not as pronounced for galaxies that closely match the regular disc-bulge or elliptical morphology, as for example visible in the first row of the plot. Nevertheless, our model can introduce additional sub-structure, which is relevant for shear bias calibration. We also display the pixel-wise root mean square error (RMSE) values between the original HST image and either the generative model reconstruction or the parametric fit. As all images are normalised to the same interval and the reconstructions do not contain noise, the values are not qualitative comparisons, but rather a way to show that the best-fit Sérsic profile is consistently less accurate.

While there are of course residuals besides pure noise, extremely small-scale deviations from the original images are not concerning for the shear calibration simulations. As the *Euclid* VIS instrument operates at half the pixel resolution of our output, the generated sample has to be processed by a *Euclidisation* pipeline which adds a correct PSF and re-samples the image

with roughly  $2 \times 2$  binning, thus reducing the overall resolution. Such differences might affect the possibility of training the CNN directly on *Euclid* Deep Field data, which is an option that could be investigated in the future.

The proposed model is not the first generative neural network for galaxy morphologies developed recently, as previously another architecture has been used within the context of *Euclid* (Euclid Collaboration: Bretonnière et al. 2022a, 2023). There are, however, some key differences. The presented architecture facilitates extremely fast training and sampling, the latter of which is necessary for the large simulation volume required for a successful *Euclid* shear calibration campaign. Moreover, it allows for an easy extension to multi-channel inputs and outputs in order to learn and simulate morphologies in multiple filter bands jointly. This improvement will be part of future work and is an important step in obtaining true morphologies across both *Euclid* instruments and for calibration of colour gradient biases (Semboloni et al. 2013).

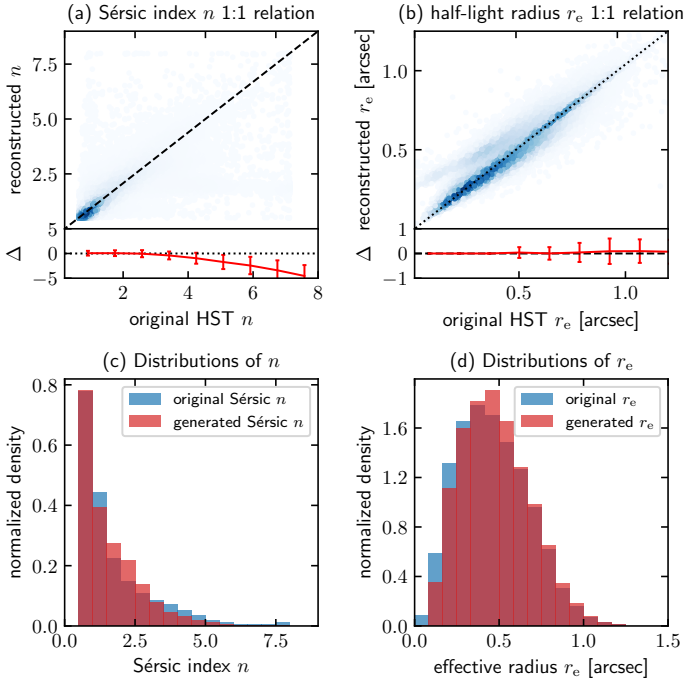
#### 4.4. Recovery of galaxy structural parameters

As an additional step for the validation of the reconstructions, we look at the distribution of the common galaxy structural parameters in the test data set, namely the Sérsic index  $n$  and the half-light radius  $r_e$ . This allows us to check if the generative model generalises well upon application to data outside of the training set, which is paramount for subsequent sampling stability and reliability. Moreover, as shape measurement biases depend on the  $n$  and  $r_e$  distributions, an accurate recovery is necessary for the next steps of the main goal of the work.

We again estimate these Sérsic model properties by performing fits of single Sérsic profiles on the original HST image, as well as on the generated output. We mimic the inputs with our new sample by matching the flux and noise properties between the input and the generated image. Additionally, we restrict our model fitting by fixing the priors on both data sets to assure identical centroid positions and fluxes for each galaxy pair.

Recovering the general input distributions is an additional indicator for success of the generative model and important for sampling of new galaxy images, as these parameters are necessary for realistic generations, due to the existing knowledge on spatial distribution and number densities of these structural parameters for true galactic populations (Shen et al. 2003). Moreover, a correlation between shear measurement bias and Sérsic index has previously been shown by Pujol et al. (2020) and Hernández-Martín et al. (2020). Therefore, samples with accurate Sérsic index distributions in the tomographic redshift bins are required for calibration of the shape measurement.

Figure 3 shows the 1:1 relations and histograms between the best-fit parameters calculated on the original HST image and the corresponding reconstruction output by the generator. The latter is thereby matched by flux to its original counterpart and then given a noise level that resembles the one on the original HST input image. We see that our generative model  $\hat{\omega}$  generalises well on the test data, as the measured structural parameters are recovered well for the majority of galaxies. The mean difference of Sérsic indices is  $\langle \Delta n \rangle = 0.13$  ( $-0.005$ ) with a standard deviation of  $\Delta \sigma_n = 0.68$  (0.46), where the numbers in parentheses are the values when only considering objects up to  $n \leq 4.0$ . For the half-light radii, the mean scatter and its standard deviation are similarly small, with  $\langle \Delta r_e \rangle = -0.02$  ( $-0.01$ ) and  $\Delta \sigma_{r_e} = 0.08$  (0.07). Towards high Sérsic indices  $n \geq 4$ , the fit accuracy breaks down, although it should be noted that the sample size is small at these values. Still, the overall distribution remains precise, with some

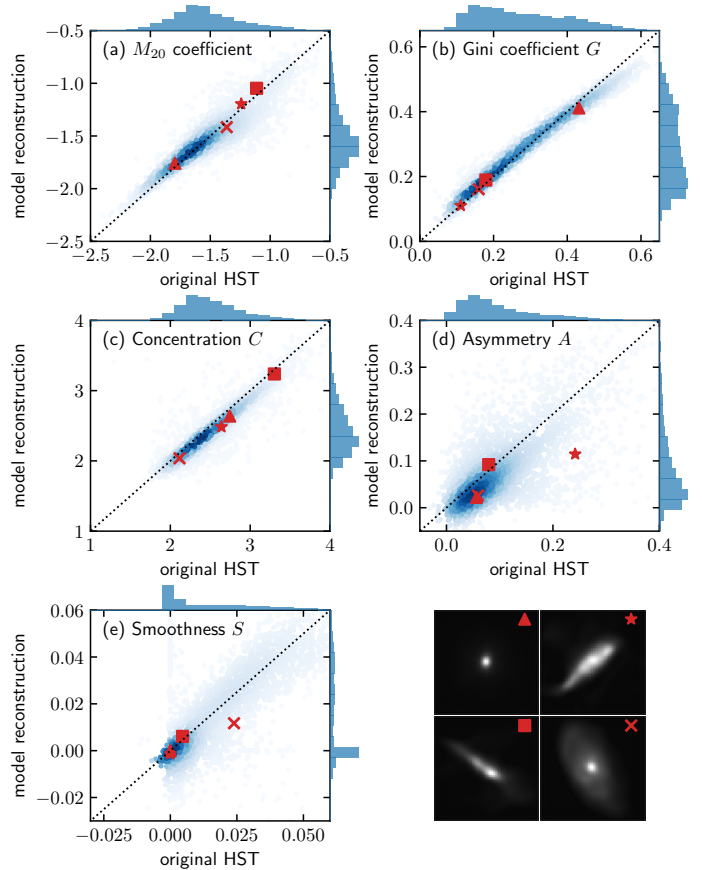


**Fig. 3.** Comparison of Sérsic index  $n$  and half-light radius  $r_e$  recovery on original and reconstructed images. Panel a shows the 1:1 relation for the Sérsic index, and panel b shows the 1:1 relation for the half-light radius. Panels c and d display the overall distributions of the two parameters in both samples. The residual  $\Delta$ -plots show the means and standard deviations of the relative difference between the original and reconstructed subsets over equi-spaced bins.

excess at intermediate Sérsic indices in the generated sample. The scatter towards higher values of  $r_e$  on the reconstructed sample can be mostly attributed to the fitting procedure, where the models for strongly peaked galaxies with high Sérsic indices are not easily distinguishable using the SVI posterior estimation and thus often produce offset half-light radii. To check the recovery of these galaxies, we show in Appendix A a sub-sample of such galaxies with the strongest fit offsets between the HST image and the generator output. Looking at the actual images of the galaxies with the strongest offset from the 1:1 relation, we see that the CNN recovered their overall shape similarly well, meaning that the large difference is a product of the degrading fit accuracy. This mostly happens for very concentrated galaxies (high- $n$  objects) with high S/N.

#### 4.5. Recovery of morphological statistics

Next, we check the statistics on disturbed morphologies for both galaxy image samples. Again, an accurate reconstruction with subsequent noise and flux matching should be able to recover the values of the input for the various morphology parameters introduced in Sect. 3. In Fig. 4, we compare a set of morphological proxies, namely  $M_{20}$ , the Gini coefficient  $G$ , and the CAS tracers. Concentration  $C$ ,  $M_{20}$  and  $G$  are recovered well for the majority of galaxies. The parameters  $A$  and  $S$  follow the 1:1 relation as well, although with more scatter compared to the other parameters, especially towards the higher end of their respective ranges. We also depict four example galaxies with varying shapes to indicate where the main galaxy populations reside within the plots. The estimation of these statistical parameters is again dependent on the noise matching, which is presu-



**Fig. 4.** Comparison of morphological proxies between original HST images ( $x$ -axis) and reconstructions by the generative model ( $y$ -axis). Panel a shows the  $M_{20}$  index, and panel b shows the Gini coefficient. Panels c–e display the CAS statistics, and panel f depicts four example images from the test data set with their values for the respective parameters shown according to the coloured markers.

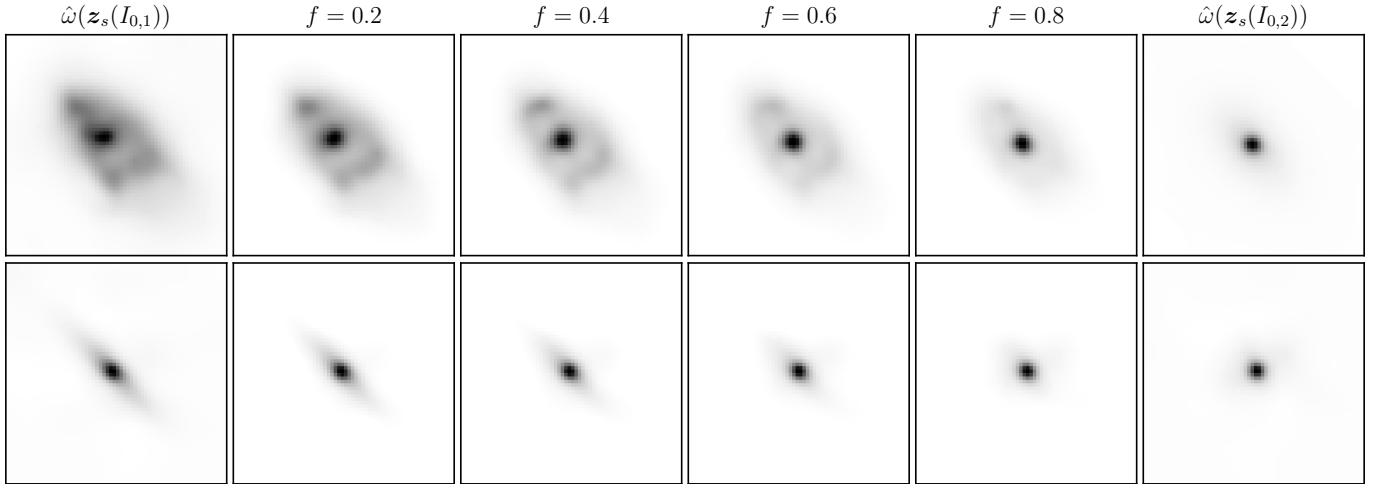
ably the strongest source of scatter, as shown by Conselice et al. (2000), Conselice (2003), and Lotz et al. (2004). Furthermore, offsets can be introduced by the segmentation algorithm applied to separate the galaxy from the background, where redshift-dependent biases may arise due to surface brightness dimming (Freeman et al. 2013). We implement the segmentation method from the photutils package (Bradley et al. 2023) for our analysis. Overall, we find a good and robust recovery of most of the morphological proxies, which proves the capabilities of the reconstructive power of our generative model.

In general, there is only limited knowledge so far on the distribution of these properties in observed samples of galaxies across redshift bins, as not a large amount of data exists, which has a high enough resolution to reliably determine such proxies. We do however expect a dependency of the shear bias from complex morphologies on some of the parameters, as shapes with more disturbance from smooth profiles should lead to larger overall deviations for an ellipticity estimator. In Sect. 6, we will check this dependency for *Euclid*-like simulations.

## 5. Generation of new galaxies

### 5.1. Galaxy-galaxy interpolation

To generate new galaxies from our trained model  $\hat{\omega}$ , several options are at hand. The common approach for VAEs or GANs



**Fig. 5.** Interpolation between two input galaxies with embeddings  $\{z_s(I_{0,1})\}$  and  $\{z_s(I_{0,2})\}$ . The columns in the middle display the intermittent results by feeding interpolated embeddings in 0.2 increments into the generative model  $\hat{\omega}$ . The images are displayed inverted to shown faint details.

is to sample directly from the distribution of latent space variables. This, however, is accompanied by the risk of also generating images whose latent values originate from the multivariate distribution spanned by the training set, but do not possess shapes that fit into the pool of observed galaxies. This can occur for example when the latent space is not compact, resulting in a possibility of arbitrary output shapes. Additionally, this makes conditional sampling difficult without training a secondary latent space model (see, e.g., Lanusse et al. 2021) or using a larger set of input galaxies that can be binned without restricting the generalisation power of the generative model.

Another possible path for the conditional generation of galaxies is linear interpolation, by leveraging the linearity of the scattering fields (Angles & Mallat 2018), which extends onto our PCA components, as the PCA is itself a linear operation. Such latent space linear arithmetic calculations have also been shown to allow for informative sampling in GANs (Bojanowski et al. 2017) and prescribe a common test for generative model performance. Assuming that the GalSim COSMOS data set is representative of the general plethora of possible galaxy shapes, every additional galaxy can be interpreted as an intermediate shape and ergo as an interpolation between two different galaxies from the whole sample. The only galaxy population that presumably does not fit into this space are highly irregular galaxies, for them, the shapes are not conformable with any known physical model anyway, as their name suggests. Thus, a potential irregular object created from a generator cannot be definitively confirmed or refuted as a realistic representation. Still, caution is required also for common shapes, due to the fact that the latent space between two objects is not necessarily fully covered, so interpolating between any arbitrary galaxy pair might not lead to realistic shapes. Interpolation between two edge-on galaxies that are rotated by  $90^\circ$  with respect to each other could ensue intermediate realisations of, for example, cross-like shapes and therefore produce overall more irregular morphologies, which we found when just randomly interpolating between galaxies from the training set. Hence, a fine-tuning of the operation is necessary to ensure realistic shape distributions. Another caveat of training data set size is the small area of the COSMOS field, with potentially large cosmic variance, meaning that the assumption that the full multidimensional parameter space is covered can be wrong. In the future, the training set needs to be expanded towards a more diverse galaxy sample.

Given the embeddings  $z_s(I_{0,1})$  and  $z_s(I_{0,2})$  of two original HST galaxies calculated with the formalism described in Sect. 4, we can obtain a linearly interpolated latent space vector  $z_s(I_{0,f})$  with

$$z_s(I_{0,f}) = (1 - f) z_s(I_{0,1}) + f z_s(I_{0,2}), \quad (23)$$

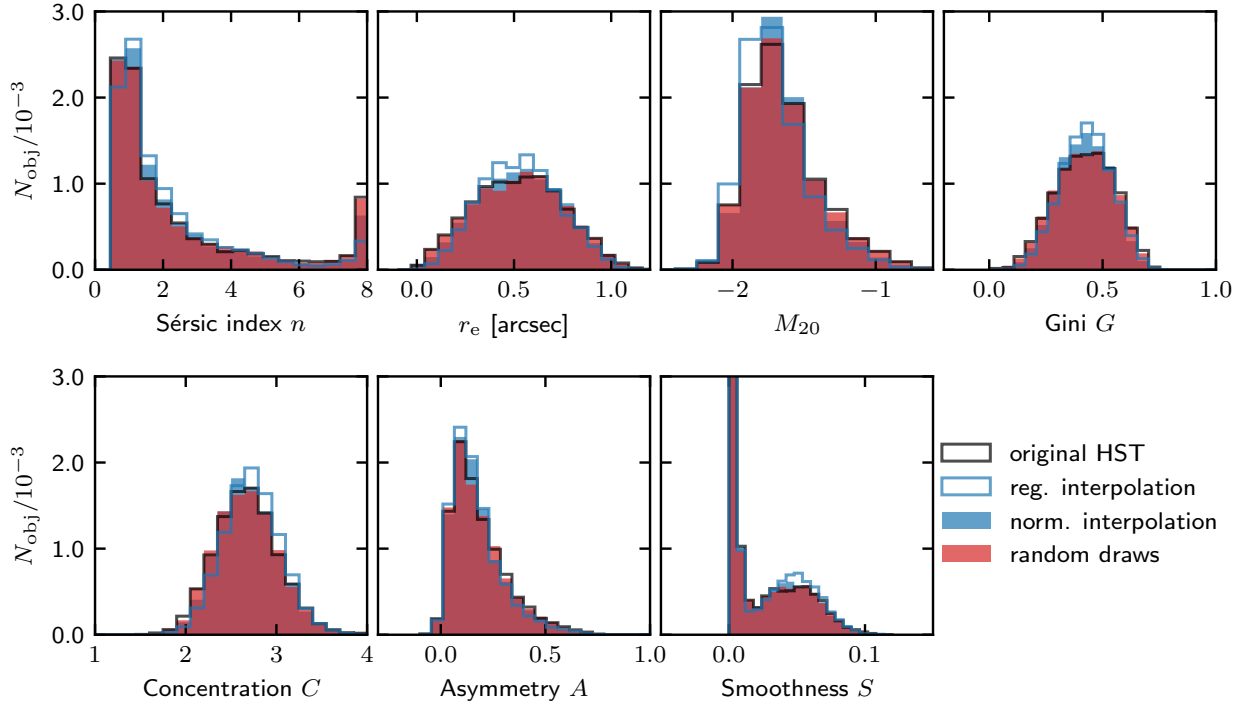
where  $f \in [0, 1]$  is the fraction of interpolation between the individual latent space components of both initial galaxies. Feeding such new realisations into the trained generative model allows for a transformation from a linear operation in the latent space into a non-linear interpolation in image space. In principle, this may be extended to extrapolations, with  $f$  values that lie outside of the mentioned interval. Alternatively, one could calculate the interpolation directly on the scattering fields, as the WST is, however, not invertible, and a gradient descent method is thus needed, which requires the training of a secondary CNN for a regularised inversion.

Latent space interpolation has previously been shown to be prone to distribution mismatch, where the latent priors are narrowed by sampling in this manner, leading to possibly incomplete coverage of the posterior distribution of images (Kilcher et al. 2017). To alleviate this issue, several methods can be incorporated, as for example normalised interpolation (Agustsson et al. 2017), where the intermediate embedding vectors are given by

$$z_s(I_{0,f}) = \frac{(1 - f) z_s(I_{0,1}) + f z_s(I_{0,2})}{\sqrt{(1 - f)^2 + f^2}}. \quad (24)$$

Randomly loading pairs of embeddings from the joint training+test data set and interpolating between them with an arbitrary value for  $f$  thus constitutes another simple way, aside from random draws of the latent space distribution, to generate new objects. The number of possible novel galaxy instances hereby by far exceeds the requirements for the *Euclid* shear calibration, as a 0.1 spacing for  $f$  already produces visibly varying morphologies and allows for combinations of the order of  $O(10^{10})$  with  $\sim 50\,000$  training objects, which can be further increased in the future by incorporating a larger postage stamp sample.

Figure 5 shows how this procedure translates into the image space. Depicted are original galaxies on the left- and right-most subplots, with four intermediate realisation obtained via 0.2 increments for  $f$  in Eq. (23). It is apparent how the overall shape



**Fig. 6.** Comparison of structural parameter distributions between original HST images reconstructed by the generator (black), samples composed by interpolating between galaxies (blue line), normalised interpolation (blue), and samples formed by random draws from the latent distribution (red). The term  $N_{\text{obj}}$  is the number of objects found per bin.

is shifted in a continuous way amidst the two sample objects, demonstrating the capabilities of the CNN and the interpolation procedure.

To preserve the input shape distribution and avoid unrealistic morphologies due to non-compactness of the embeddings space or orientation-related sampling issues, a more restrictive interpolation may thus need to be incorporated. While the standard random interpolation might produce realistic shapes for ellipticals, this is not generally the case for all pairs of galaxies, thus requiring a fine-tuning of the interpolation.

## 5.2. Interpolation fine-tuning

To circumvent such concerns, several solutions can be realised. For once, we can limit the interpolation fraction. This, however, reduces the amount of clearly discernible galaxies from our data set and could result in too many similar looking galaxies that do not necessarily cover the range of realistic shapes that will be observed by *Euclid*. Another option is to disturb existing galaxy images not in a specific direction of the latent space, but by diffusion or random walks in the neighbourhood of their embedding vectors. This though could again give rise to an overall unrealistic distribution of numerous too similar objects or risk the generation of non-physical objects, as we do not have knowledge of the latent space topology mapped by the optimised generative model.

Therefore, we test a different method that should allow for more variable generation and simplify the realisation of conditional sampling by Sérsic index or  $r_e$  distributions. To be able to use the full range of interpolation fractions  $f$  and diminish the likelihood of unrealistic light profiles, we first rotate all original HST galaxies to match along their major axis, which we set as the axis of the best-fit Sérsic model. In consequence, we avoid the possibility of generating objects with for example cross-like

shapes, as the interpolation then always takes place along the same axis. Galaxies for which a clear symmetry cannot be reasonably assigned, that is irregulars, do not pose a threat to this framework, as they have intrinsically peculiar shapes and will therefore naturally lead to generations of new irregular instances, irrespective of their rotation. Then, we recreate the embeddings for these new images and re-train the generator  $\hat{\omega}$  on this data set. We choose the  $45^\circ$  diagonal with respect to the  $x$ -axis as the designated direction, as we can thereby steer clear of rotating large galaxies previously residing along this direction out of the image bounds. It should be noted that rotation in the image space can result in information loss due to interpolation onto a new pixel grid with GalSim. This should, however, be irrelevant towards the application for *Euclid*, as the pixel scales differ by a factor of two and fine details will be smeared out by the re-binning and noise application. With this new data set of generated galaxies, we can obtain new instances over the full range of  $f$  by drawing pairs of objects and interpolating between them along the diagonal axis.

## 5.3. Random draws versus interpolations

Next, we check whether one of both interpolation methods, regular and normalised, provides an advantage with respect to the common random multivariate sampling approach for galaxy generation if no conditionality is needed. For this, we employ all three techniques to create  $10^4$  random galaxies, respectively, and also randomly choose  $10^4$  galaxies from the reconstructed HST training data. Afterwards, we compare the distributions of structural parameters and  $CAS + GM_{20}$  statistics between all sets of objects and selection from the input training set.

Figure 6 displays the histograms of the measured properties for all subsets. Clearly, random latent space drawing can more consistently trace the parameters distributions of the original

HST data set sample, with almost identical histograms. The regular interpolation method on the other hand is similarly reliable on the Sérsic index recovery, but fails to capture the distribution tails for the effective radius and the  $CAS+GM_{20}$  morphology proxies, even though the means are captured correctly.

This is a logical consequence of the method, due to the aforementioned distribution mismatch. The probability of drawing a galaxy from one of the tails is low to begin with, and the likelihood of generating such an object will be decreased upon interpolation with a sample that most likely does not reside in the same regime. This leads to a narrowing of the latent distribution, which translates to the image space and hence to the measured properties. As can be seen in the plot, normalised interpolation reduces this effect, but is still not capable of achieving the quality of the results from random draws.

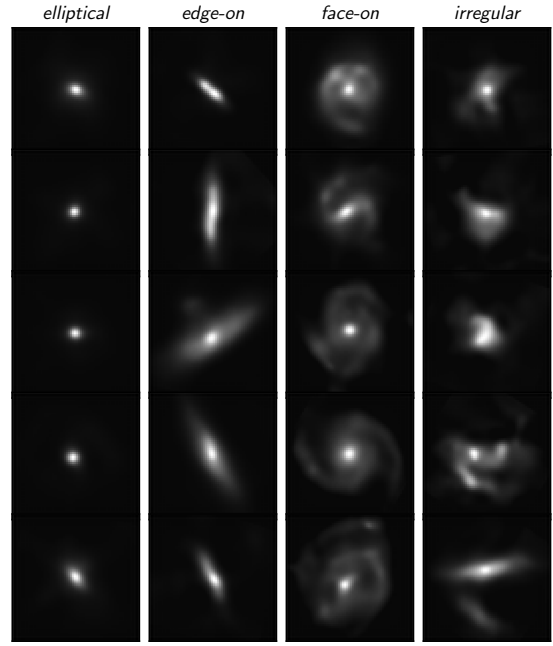
Still, we note that linear interpolating along a predefined axis did not produce objects with non-physical structural parameters or morphology statistics outside the input distribution. This proves the capability of the technique for the task at hand and indicates a tightly packed latent space. While the random draw method delineates a powerful tool for galaxy generation if only the input distribution shall be recovered, we will further explore the interpolation approach for conditional sampling of galaxy properties.

We note here that while the  $CAS+GM_{20}$  parameters provide a well-established set of morphological estimators, there is no clear prior information on which values or distributions would describe non-physical shapes. Moreover, irregular galaxies can most likely reside anywhere in this parameter space, as their origins lie in turbulent process that can create a wide range of complex structures. Thus, a distribution match between samples and observation does not necessarily prove the realism of the generations, but is still a valuable indicator of the model performance.

#### 5.4. Galaxy classification

One step towards conditional sampling can be to group the galaxy data set roughly into main categories of for example ellipticals, spirals, and irregular galaxies. Overall, there exist a multitude of methods for galaxy classification along the Hubble sequence. This can be achieved using intrinsic galaxy properties such as star-formation rate (SFR) or colour (Kennicutt 1998), joint analysis of morphological tracers (e.g. Gini- $M_{20}$ ), or even citizen science projects such as Galaxy Zoo (Darg et al. 2010). While classification with Galaxy Zoo is able to make more distinct classifications, it cannot handle the amount of data in stage IV surveys such as *Euclid*. Machine learning techniques trained on Galaxy Zoo results, however, have recently been shown to be able to classify *Euclid* morphologies directly from the images (Euclid Collaboration: Aussel et al. 2024a). This, as well as using morphological proxies, requires pixel data and relies on rather arbitrary thresholds for classification. Leveraging galaxy population properties, which on the one hand only uses photometry, is usually only able to robustly separate the bimodal distribution of early-type ellipticals and late-type spiral galaxies (Baldry et al. 2004). For deep generative models, a classification can also be obtained via the learned embeddings. Here, studies for instance on GANs have previously shown the manifold clustering capabilities of their low-dimensional latent space distributions (Mukherjee et al. 2018).

We here employ a different method by performing a clustering analysis on our galaxies via the wavelet scattering transform. Given the scattering coefficients of each image, we can determine if and how the multi-dimensional parameter set corre-



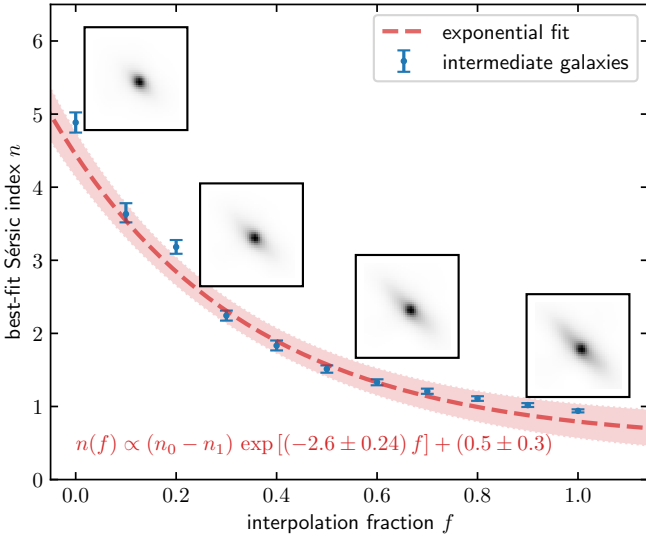
**Fig. 7.** Example galaxies for the four classes *elliptical*, *edge-on*, *face-on*, and *irregular*, classified by fitting a bGMM to their second-order reduced scattering coefficients  $s_2$  and assigning labels to each component.

lates with each galaxy’s intrinsic light profile and accordingly its overall class affiliation. For this, we calculate the second-order reduced scattering coefficients  $s_2$  with  $L = 4, J = 6$  for each galaxy, as they have been shown to correlate with galaxy morphology (Cheng & Ménard 2021). Next, we fit a Bayesian Gaussian mixture model (bGMM) to their distribution. In general, a Gaussian mixture model is a probabilistic model that describes a weighted sum of  $k$  multivariate normal distributions  $\mathcal{N}$  given by

$$p(\mathbf{x}) = \sum_n^k \pi_n \mathcal{N}(\mathbf{x} | \mu_n, \Sigma_n), \quad \sum_n^k \pi_n = 1, \quad (25)$$

where  $\pi_n$  is the weight of the  $n$ -th component and  $\mu_n, \Sigma_n$  are the respective means and covariance matrices,  $p(\mathbf{x})$  is the distribution of input vectors, in our case with  $\mathbf{x} = s_2$ . In a bGMM, the parameters of the model are not found with an expectation maximisation algorithm, as is the case for common GMMs, but by variational inference of an approximate posterior using a Dirichlet prior on the parameters and then maximising the log-likelihood  $\ln \mathcal{L}(p(\mathbf{x}))$ . We use the bGMM implementation from the `scikit-learn` package (Pedregosa et al. 2011). Afterwards, we attribute a keyword to each component, based on the general visual appearance of the items within the respective cluster. These keywords are *elliptical*, *edge-on*, *face-on*, *irregular*.

Figure 7 visualises a random sample of galaxies for each class. It should be acknowledged that the class *edge-on* hereby does not necessarily only consist of disc plus bulge galaxies with an inclination angle of  $90^\circ$  relative to the line of sight, but only groups objects which appear elongated and could thus also include, for example, lenticular galaxies with a large axis ratio. Moreover, the differentiation between face-on spirals and a subset of the irregular class is not trivial, resulting in possible contamination of the peculiar sample by regular spiral galaxies. Besides, the exact differentiation between face-on and edge-on is overall rather arbitrary, as naturally galaxies exist over the full



**Fig. 8.** Dependence of the Sérsic index on the interpolation fraction  $f$ . The blue points and error bars are the respective best-fit Sérsic index and the estimated error from `pysersic`. The red dashed curve shows our best-fit parametric model, with the dotted region designating the  $1\sigma$  confidence interval. The image insets are inverted stamps at intermediate interpolation points.

range of inclination angles. While this method is by no means a classifier that can currently compete with machine learning models such as `Zoobot` (Walmsley et al. 2022), it still showcases the morphological information carried by the wavelet scattering transform, as also shown by Cheng & Ménard (2021). Due to the fact that we do not necessarily need to sample from overall morphological galaxy populations for the quantification of the shear bias, a more complex model is not needed. Still, these general morphology flags provide a first method towards conditional sampling of the COSMOS data set by drawing objects from the latent distribution of each cluster.

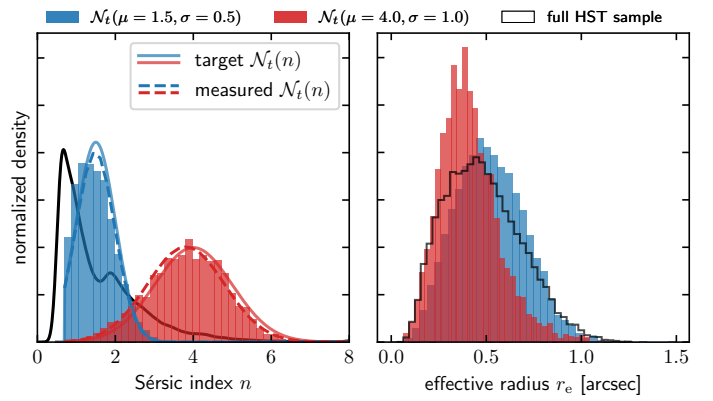
### 5.5. Conditional sampling

Image simulations for shear calibration are needed not only for random fields, but may need specific structural property distributions such as for cluster fields, therefore requiring conditional sampling methods. To sample new galaxies by their Sérsic indices, we can leverage the latent space interpolation technique. To draw an object at a specific Sérsic index  $n_i$ , we draw a galaxy with  $n_0 > n_i$  and another object with  $n_1 < n_i$  from the joint reconstructed training/test data set. Then, conditional sampling only requires finding a robust functional correlation between the interpolation fraction  $f$  and the Sérsic index. To obtain such an empirical description of this dependence, we draw random pairs of galaxies with  $n_0 > n_1$  and calculate their intermediate realisations by propagating the interpolated latent representations in 0.1 steps for  $f$  through the generator, then measure their structural parameters and finally parametrise the functional form of the Sérsic index development.

We find that the correlation of  $f$  and  $n$  can be approximated well by an exponential law following

$$n(f) \propto (n_0 - n_1) e^{-\beta f} + \eta, \quad (26)$$

where the parameters  $\beta = 2.6 \pm 0.24$ ,  $\eta = 0.5 \pm 0.30$  were determined empirically as the mean values from an exponential fit to the  $n(f)$  relation for multiple galaxy pair draws. Figure 8



**Fig. 9.** Results of conditional sampling of galaxies with a normal Sérsic index distribution. The left plot shows the obtained Sérsic index distributions, on the right we show the corresponding effective radii. The black curve depicts the overall distributions from the HST training data.

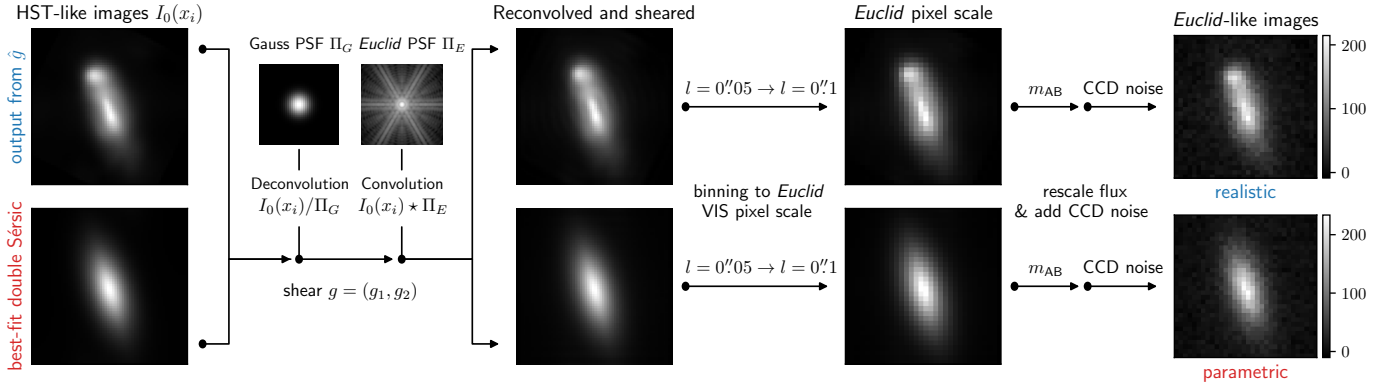
shows the functional form of  $n(f)$  with the  $1\sigma$  confidence interval and example images of a galaxy pair for the interpolation along the exponential law. A similar parametrisation can also be constructed for the half-light radius, to sample galaxies according to this property.

The main drawback of this approach is the fit accuracy at high Sérsic indices. As precise fits of Sérsic profiles are difficult for strongly peaked galaxies without time-intensive sampling, the correct interpolation might suffer under galaxy draws that are mismatched from the target index, thus leading also to differing intermediate representations. To alleviate contamination of the target sample, we reject every galaxy that deviates by more than  $\Delta n = 0.5$  from its target value. We observe that the rejection fraction increases with  $n$ , which confirms the hypothesis of mismatches due to high-Sérsic fitting difficulties.

Given this correlation, we can sample either specific single objects or distributions of Sérsic indices from the generator  $\hat{\omega}$ . To achieve this, we draw values for  $n$  randomly from the target prior. Then, we choose two galaxies with  $n_0 > n > n_1$  from the train/test images and calculate the required interpolation fraction according to Eq. (26) and generate the corresponding image from the interpolated embedding vector. Afterwards, we measure the Sérsic index of the obtained object to validate the result.

To present the capability of this sampling method, we apply this routine to generate two arbitrarily chosen samples of  $10^4$  objects following a truncated normal distribution  $\mathcal{N}_t(\mu, \sigma, [\text{min}, \text{max}])$  for the Sérsic index over the interval  $[0.5, 8.0]$ , peaked at means of  $\mu = 2$  and  $\mu = 4$  and with respective standard deviations of  $\sigma = 0.5$  and  $\sigma = 1.0$ . Figure 9 shows a comparison of the target Sérsic index distributions and the obtained samples, as well as the distributions of the corresponding half-light radii. We find that we are able to accurately construct galaxies given a target Sérsic index distribution. Moreover, the technique also allows for sampling of individual objects at specific  $n$  values over the range of the training data distribution. Additionally, the effective radius of the generated sample does not strongly depart from the range of the original HST distribution, which confirms that the interpolation approach does not create galaxies outside of the input range that might thus exhibit unrealistic shapes.

While the rejection fraction of generated galaxies with offsets larger than 0.5 from their target value increases strongly



**Fig. 10.** Procedure for creating *Euclid* VIS-simulated galaxies from the generator output and their parametric double-Sérsic counterparts. First, the generated images are deconvolved by their Gaussian training PSF, and then a shear is applied, after which they are convolved with the *Euclid* PSF from Appendix C. Next, the galaxies are re-binned to the correct pixel scale of  $l = 0''.1$ . Finally, the total image flux is rescaled according to an apparent magnitude following Eq. (28) and CCD noise is added using the GalSim.CCDNoise tool.

with  $n$  (up to  $\sim 0.8$  for  $7 < n < 8$ ), the majority of galaxies, especially at low Sérsic indices, where the observed number density is higher anyway, are created accurately and with morphological proxies that trace the input distribution, as shown in Fig. 6. Furthermore, even high rejection fractions are not concerning given the speed of the generation, as  $\sim 10^4$  objects can be propagated through the generator in approximately one minute using only one CPU core. This can easily be accelerated even more with GPUs, given that the model architecture is written entirely with pytorch and already optimised for usage with cuda devices.

This conditional method allows for the direct and fast simulation of a large number of galaxies for insertion into full GalSim scenes and the *Euclid* VIS pipeline. Not only can one sample using a specific value or distributions for Sérsic indices, but also by galaxy class (*elliptical, face-on, edge-on, irregular*) via interpolation between objects from the respective label. Additionally, the two methods can be united, to for instance create a population of edge-on galaxies at Sérsic index  $n = 2$ . Still, the diversity of such generated data is strongly susceptible on the size of the training data. This can be improved in the future by larger training data sets from different HST fields or Euclid Deep Field observations. Right now, this is not needed for the main goal of this work, which is the calibration of the shear measurement bias due to galaxy substructures.

## 6. Quantification of shear bias from complex galaxy morphologies

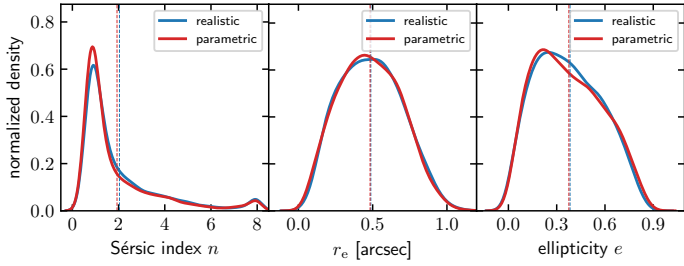
### 6.1. Measurement setup

To reach the scientific requirements of *Euclid*, the bias needs to be calibrated to an accuracy of  $|\delta\mu| < 2 \times 10^{-3}$ ,  $|\delta c| < 1 \times 10^{-4}$  (Cropper et al. 2013). However, most of this error budget is needed for PSF modelling errors. Hence, even sub-percent level biases have to be accounted for, as they exceed the target precision by more than one order of magnitude. Now that we have presented our method for generating galaxies with realistic, noise-free morphologies from the GalSim COSMOS sample, we will apply the technique for the estimation of the shape measurement bias by complex galaxy structures. The difference in shear biases between realistic and parametric objects can be leveraged for this task and determine if this difference is relevant for Euclid Wide Survey cosmic shear measurements. We perform the estimates here using KSB due to its speed, future work will apply

more modern shape measurement methods such as LENSIMC and METACALIBRATION.

First, we need a sample of generated galaxies and their parametric counterparts which are both rendered identically (aside from the complexity of their shapes) and possess the same noise and PSF properties. For this, we sample randomly from the latent space of the trained CNN and create  $5 \times 10^5$  new images. We mainly want to quantify the bias with respect to the whole range of possible shapes and as a function of structural parameters and morphological proxies, a conditional sampling approach is therefore not needed. Afterwards, we perform double-Sérsic fits for each generated object using pysersic. We note that we here do not account for redshift–magnitude correlations yet, which can change the distribution of shapes due to increased irregularity at higher redshifts. For tomographic bias difference estimates this has to be accounted for. This can for example be done by adding the redshift as an additional latent dimension and learning the joint distribution of redshifts and shapes.

As we try to determine the bias for *Euclid* VIS, the generated galaxies have to be modified towards a *Euclid*-like emulation. Such an *Euclidisation* procedure has previously been developed by Euclid Collaboration: Scognamiglio et al. (2025a), although we can here abstain from noise whitening and symmetrisation. Instead, we just add CCD noise (consisting of Poisson shot noise from source and background, as well as Gaussian read-out noise) to the image using the corresponding GalSim function. Accounting for both of these noise sources is important for *Euclid*, given its lower sky background in comparison to previous ground-based weak lensing surveys. First, however, we deconvolve the generated images by their original Gaussian PSF, apply a shear and convolve them with a fixed *Euclid* PSF, as presented in Tewes et al. (2019) and Jansen et al. (2024). Next, we draw the galaxies at the correct pixel scale of  $0''.1$ , producing  $32 \times 32$  pixel postage stamps. Then the galaxies need correct fluxes and noise levels. For this, we assigned to each galaxy a magnitude  $m_{\text{gal}}$  following the distribution from the test+training data set but allowed magnitudes of up to  $m_{\text{AB}} \leq 24.5$ , which approximately corresponds to the VIS target-limiting magnitude. Finally, the noise was added as described above. For this emulation, we used the assumptions on the *Euclid* VIS detector and observing conditions from Tewes et al. (2019), namely the nominal VIS exposure time  $t_{\text{exp}}$ , gain  $G_{\text{VIS}}$ , read-out noise  $R$ , zero point  $Z_p$ , and sky brightness  $m_{\text{sky}}$  (Refregier et al. 2010; Laureijs et al. 2011; Niemi et al. 2015; Cropper et al. 2016). The sky level and galaxy



**Fig. 11.** Sérsic index  $n$ , half-light-radius  $r_e$  and ellipticity  $e$  distributions measured on the realistic (blue), and parametric (double-Sérsic, red) sample galaxies. The data are shown as a kernel density estimate on the histograms. The vertical dashed lines indicate the mean of the respective distribution.

flux at given object and sky magnitudes were thereby calculated following [Tewes et al. \(2019\)](#) with

$$F_{\text{sky}} = \frac{t_{\text{exp}} l^2}{G_{\text{VIS}}} 10^{-0.4(m_{\text{sky}} - Z_p)}, \quad (27)$$

$$F_{\text{gal}} = \frac{t_{\text{exp}}}{G_{\text{VIS}}} 10^{-0.4(m_{\text{gal}} - Z_p)}, \quad (28)$$

where  $l$  is the pixel scale in arcsec. These parameters will have to be adjusted in the future once accurate measurements have been performed following the performance verification of the *Euclid* mission. Figure 10 exhibits an overview of this process to obtain pairs of realistic and parametric galaxies that emulate *Euclid* VIS images. To perform shape noise cancellation, we additionally create duplicates of the original HST-like images rotated by 90 degrees for both sets of simulated galaxies, but with an identical noise field (the noise is not rotated), which increases the effective number of galaxies by a factor of two to  $10^6$ .

We separately save an oversampled image of the applied *Euclid* PSF, which is needed by the KSB method for a shear estimate. This PSF image can be found in Appendix C. On top of that, we assign each object to one of the galaxy classes introduced in Sect. 5 by calculating their second-order wavelet scattering coefficients and also measure the S/N following the definition by [Tewes et al. \(2019\)](#) and the  $CAS + GM_{20}$  statistics of the galaxies to later quantify the bias as a function of these properties.

## 6.2. Realistic morphologies versus double-Sérsic

As a first measurement, we take the sample of  $10^6$  noise-less galaxies from the generator and their parametric double-Sérsic fits and emulate *Euclid* VIS images following the described prescription, creating two measurement branches. Initially, we do this at fixed magnitudes over the range  $m_{\text{AB}} \in [20.0, 24.0]$ . With this, we intend to quantify the overall effect of the bias as a function of magnitude up to approximately the limit of the Euclid Wide Survey. As the fine details of galaxy substructures get increasingly washed out with a decreasing S/N, the galaxy morphology bias may exhibit a dependence on the S/N. While there is an additional dependency of galaxy complexity on redshift and magnitude correlations, we do not yet account for this here. Still, in the future this has to be addressed, as two independent effects can affect the bias difference: The evolution of the galaxy population with more irregular objects at higher redshift and thus lower S/N, as well as the improved resolution of substructures at higher S/N, with a possibly stronger impact on the shape measurement. As an initial estimate of this effect, we will later on

show the dependence of the bias on the  $CAS + GM_{20}$  statistics at a fixed magnitude. As aforementioned, this can in the future be implemented by adding latent space dimensions consisting of the relevant conditional parameters, which are for example magnitude and redshift.

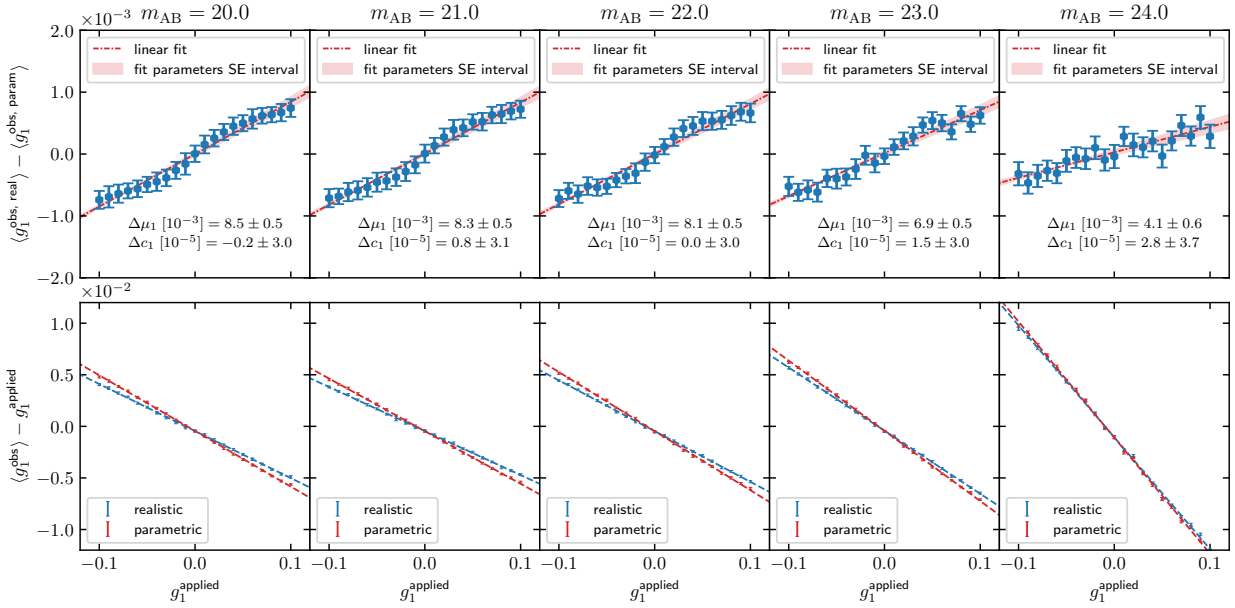
To ensure that both branches have identical Sérsic indices, half-light radii, and ellipticity distributions, we measured these properties with `pysersic` on the HST-like generator sample and the fitted double-Sérsic profiles. This is an important validation in order to warrant that the measured bias difference does not originate from deviations in said distributions, which has been shown to have a direct influence on the bias ([Hoekstra et al. 2015](#); [Hernández-Martín et al. 2020](#); [Pujol et al. 2020](#)). In Fig. 11 we show the distributions of these properties in both branches for the full set of  $10^6$  galaxies. As one can see, the distributions match over the full data set, which therefore guarantees that no significant influence from these parameters can be expected on the shear bias difference from complex morphology. Slight differences can also be attributed to the fast fitting procedure that might not always find the true best-fit Sérsic model on either of the two branches. Moreover, no real Sérsic fit can be obtained for complex objects with multiple brightness peaks for instance, naturally leading to differences between branches. One can argue that such differences can lead to a bias itself. This, however, would also be there in relation to real galaxies, where similar issues arise from the simplicity of a parametric light distribution. Thus, such discrepancies are a part of the morphology bias that should not be removed by using only objects with identical structural parameter distributions.

During the emulation, we apply shears over a range of values for  $g_1$ , while we fix  $g_2 = 0$ . We choose a shear interval of  $g_1^{\text{applied}} \in [-0.1, 0.1]$  and sample it in increments of 0.01. Then, we measure the shear components with KSB for each generated pair of galaxies. A key difference to the usual parametric image simulations is the fact that our training galaxies, and thus also the CNN output, already contain a non-zero cosmic shear, as they originate from real HST data. This can neither be removed nor accounted for, as we do not know the shear of the generated galaxies. However, that only adds an additional source of shape noise and also affects parametric models where usually an observed ellipticity distribution is used as the input.

After measurement of the shear on all galaxies over the specified shear interval, we calculate the mean  $\langle g_1^{\text{obs}} \rangle$  and standard error  $\sigma / \sqrt{N}$  of the observed shear over the entire sample, where  $\sigma$  is the standard deviation of the measured shear distribution over the full sample and  $N$  designates the number of galaxies in the sample, that is  $N = 10^6$ . Thereby, we employ the shape noise cancellation, where the mean and errors are calculated over the pre-computed means of each galaxy pair. If KSB fails for one of the two objects, or both, we simply reject the full galaxy pair to avoid selection bias effects for now. This way, we obtain the multiplicative and additive biases for both measurement branches, namely the realistic and the parametric samples, using Eq. (9). We determine the errors on the linear fits of the bias via bootstrapping of random samples from the shear measurement and obtaining mean and standard deviations of the fit parameters from the bootstrap sample. In the absence of other bias sources that differ between the branches, the difference

$$\langle g_i^{\text{obs, real}} \rangle - \langle g_i^{\text{obs, param}} \rangle = \Delta\mu_i g_i^{\text{applied}} + \Delta c_i \quad (29)$$

can be used as a metric to determine the bias introduced on the shape measurement from realistic galaxy morphologies as opposed to simple Sérsic profiles. While there exist other sources



**Fig. 12.** Results of the cosmic shear measurement with KSB on  $1 \times 10^6$  simulated galaxies at a fixed magnitude. The simulation was performed once with realistic galaxies from the deep learning model and once with their double-Sérsic counterparts. The bottom row depicts the individual shear biases measured on both branches, the upper row shows the difference of the mean shear at each increment with a linear fit and the standard error confidence interval for the fitted parameters  $\Delta\mu_1$ ,  $\Delta c_1$ .

of bias, as for example from pixel noise or a cut-off of the light profile at the edges of the postage stamp (which should be neglectable as our model was only trained on images that do not exceed the specific stamp size), these effects should be apparent in both data sets. Accordingly, these effects should be irrelevant for the main aim of this work, as we do not expect these bias sources to vary between both measurement branches due to their same distributions in S/N,  $n$ ,  $r_e$ , and ellipticity.

While the aforementioned magnitude limit designates the target for the inclusion into the cosmic shear analysis for *Euclid*, low surface brightness galaxies with high half-light radii might exhibit extremely low S/N despite lying in the overall magnitude range. Therefore, we applied a cut at S/N = 10 in the parametric branch and did not include galaxies from both branches into the bias estimation that fall within this sample.

In Fig. 12, we show the results of measuring the shear bias following Eq. (29), where each column shows the measurements at a specific fixed magnitude that was applied to all sample objects. This allows one to trace how the bias changes with decreasing magnitude. While in general morphology and magnitude are tightly correlated, we neglected this correlation on purpose in this case to be able to trace the development of the bias over matching shape distributions. Overall, it is apparent that there is a significant offset between the linear biases of the two branches over the entire magnitude range. This offset does not appear to strongly change with increased magnitude and thus decreased S/N up to  $m_{AB} = 22.0$ , with values at a  $\sim 0.8\%$  level for  $\Delta\mu_1$  and a  $\sim 10^{-5}$  level for  $\Delta c_1$  for each respective measurement with the KSB algorithm. At a fainter magnitude of  $m_{AB} = 24.0$  though, the bias is significantly reduced below the 0.5% level for  $\mu_1$ . We note here also the increased scatter at lower S/N, due to the degradation of the KSB shear estimator for faint galaxies (Hirata & Seljak 2003). Likewise the absolute value of the shape measurement bias increases at this magnitude (see the steep slopes in the lower panel plots). Schrabback et al. (2010) proposed corrections for noise-related multiplicative bias components for this algorithm, though we deem these here not nec-

essary, as KSB only serves as an initial fast test. Noise-related differences occur in both branches, cancelling out in the relative comparison of  $\mu$ . We also notice and confirm the non-linearity in the GalSim KSB shear estimator previously detected by Jansen et al. (2024) for high shears. Ignoring higher-orders of the shear bias also biases the linear term. We show antisymmetric quadratic fits to the shear bias difference in Appendix D and list the corresponding parameters for comparison.

Overall, the results show that realistic galaxies with complex morphologies and substructures can bias the shape measurement at a relevant level given *Euclid*'s tight requirements for the cosmic shear analysis. Our results confirm the order of magnitude of this effect as found in the space-based PSF branch of the GREAT3 challenge (Mandelbaum et al. 2015). While the effect is reduced at lower S/N, where the majority of the galaxy population in the Wide Survey lies, the bias can still not be neglected there and thus has to be accounted for. Still, our sample for now assumes the presence of galaxies of all types within the COSMOS sample at each magnitude, which is not the case in reality. In true observations, disturbed and irregular structures will preferentially occupy the high-redshift and faint-magnitude regime, which could thus reduce or increase the amount of bias in specific tomographic redshift bins. In the future, we will extend this analysis to provide a shear bias calibration that incorporates the distinct target redshift tomography of the Euclid Wide Survey that will be applied for the cosmological weak lensing analysis. The presented work is mostly a proof-of-concept study, which shows that complex morphologies should be added to the *Euclid* simulation pipeline for accurate shear calibration. However, even if KSB and the missing morphology-redshift correlation were to overestimate the bias by a full order of magnitude, it would still be relevant for the Survey. For instance, looking at the  $m_{AB} = 24.0$  estimates, it is likely that the bias difference is underestimated. The sample includes more elliptical galaxies than the typical faint and high- $z$  sample, which would mean even tomographic bins with mostly faint objects would be biased significantly due to increased complexity.

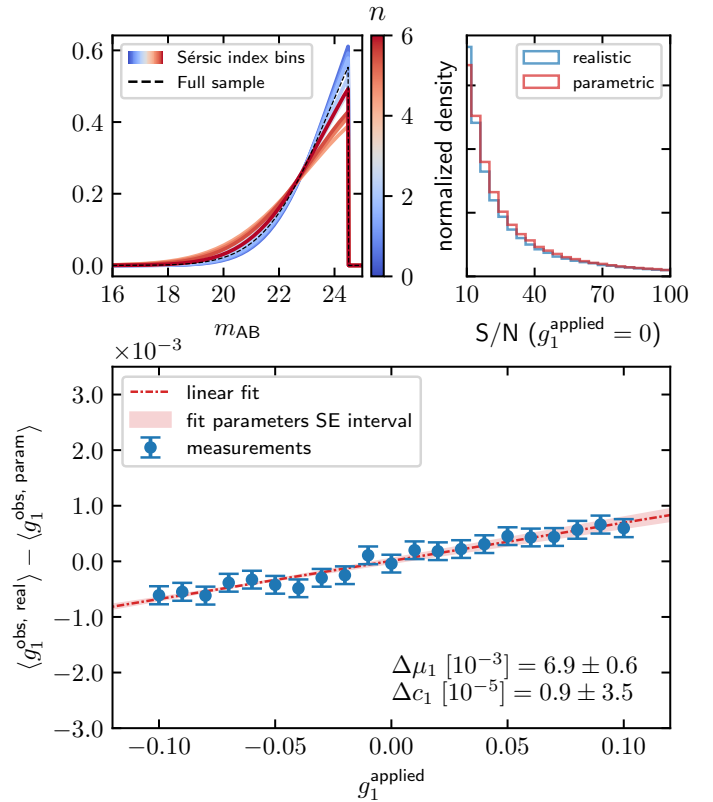
Moreover, different shape measurement methods need to be tested as well, especially LENSMC (Euclid Collaboration: Congedo et al. 2024b), as it is the designated code for the *Euclid* data release 1 (DR1). While KSB is fast and thus advantageous for initial testing, it will not be used for stage IV surveys and hence does not provide an estimate that is fully useful for future cosmological analyses. Nevertheless, these initial findings show the importance of the accurate calibration of this effect, because multiplicative biases at the  $\sim 0.5\%$  level exceed the necessary calibration precision by a factor of 2.5 (Cropper et al. 2013). As the applied PSF is not elliptical though, the additive bias should be consistent with zero, which is the case within the errors. More galaxies and a more realistic PSF are required in the future to check if PSF anisotropy has a significant impact on the bias difference from complex morphologies.

These initial estimates provide a rough quantification of the order of magnitude of the influence by complex galaxy morphologies on the shape measurement, where the magnitude and the S/N of the image correlate with the bias level. This can be expected, as the details of galaxy substructures are washed out with decreasing S/N, which gives rise to the expectation of a reduced bias level at faint magnitudes and hence a low S/N. Our results support this assumption. As mentioned, the amount and degree of complexity of galaxy shapes is a function of redshift, and it might thus still affect the bias for true distributions. If a population of galaxies at low  $z$  and bright magnitude has a higher percentage of elliptical objects than galaxies at higher redshifts, the bias can be reduced in comparison to a sample with increased peculiarity. Hence, it is a necessary step to extend the model towards realistic correlations between morphological properties and redshift; this, however, is beyond the scope of this work.

### 6.3. Bias for a realistic magnitude distribution

Next we also estimate the mean morphology related bias for a galaxy population that follows the expected magnitude distribution of the Euclid Wide Survey (Euclid Collaboration: Scaramella et al. 2022). Given that we use the COSMOS data as a training sample, we can use its magnitude and Sérsic index distributions to model the corresponding properties of our emulated *Euclid* data set. To ensure closely matching characteristics, we separate the magnitude distribution into Sérsic index bins and then fit a truncated Gaussian with an upper limit set at  $m_{AB} = 24.5$  on each component. Then, we sample magnitudes for each generated object according to the magnitude distribution in the respective bin. Thereby, we intend to mimic the true observations as closely as possible with a basic magnitude-morphology relation. However, this does not account for possible correlations between magnitude and tracers of more complex morphologies. This will be achieved in future work as we expand the framework to account for such dependencies and create fully realistic galaxy populations within each tomographic redshift bin. Finally, we again convert the generated sample into VIS-like postage stamps with additional shears over the same interval and then measure the S/N and the estimated  $g_1$  from KSB.

Figure 13 shows the initial magnitude and measured S/N distributions, as well as the bias estimate on  $10^6$  galaxies in the lower subplot. We again observe an overall shear bias difference, that is  $\Delta\mu_1 = (6.9 \pm 0.6) \times 10^{-3}$  for the multiplicative bias and  $\Delta c_1 = (0.9 \pm 3.4) \times 10^{-5}$  as the additive component. The  $c$ -bias is again consistent with zero within its errors, as expected for a circular PSF. Overall though, it is apparent that

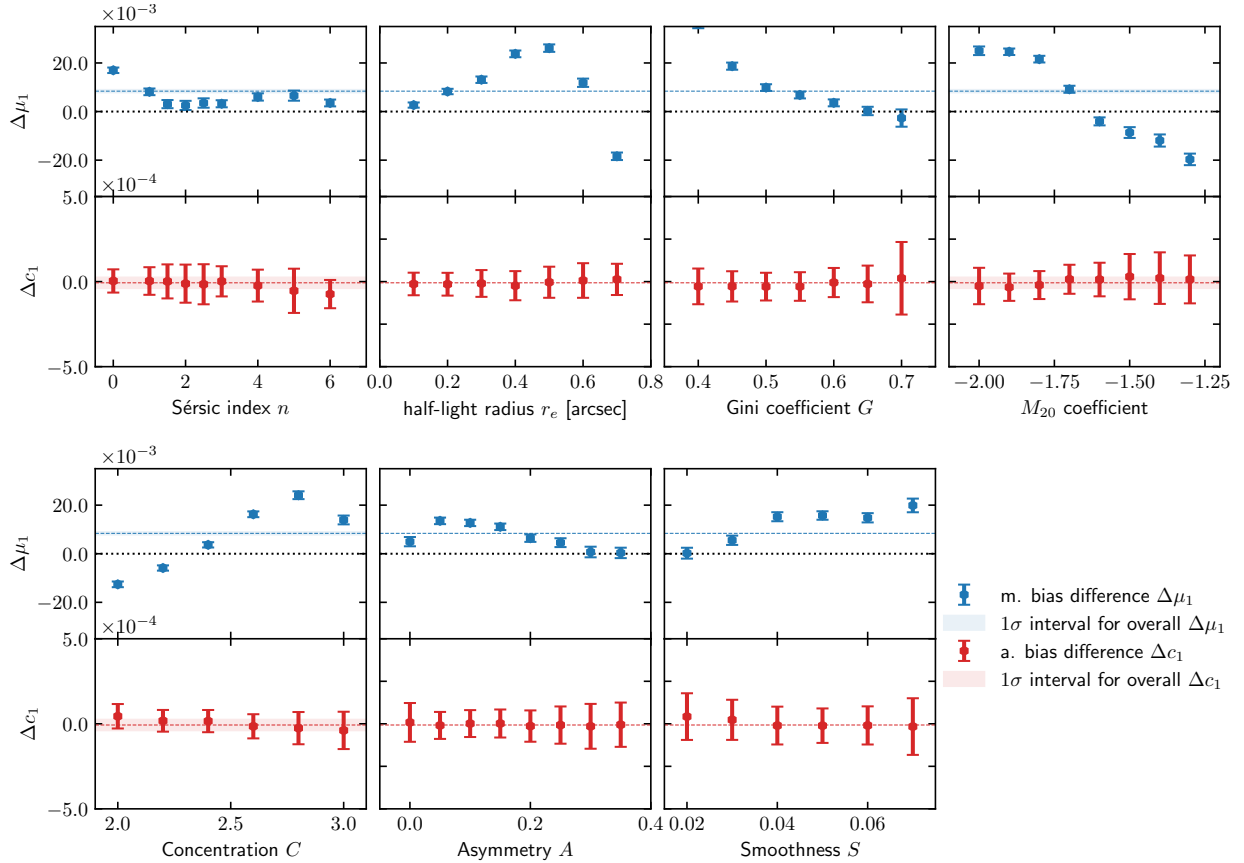


**Fig. 13.** Shear bias difference estimate on a realistic magnitude distribution. The first subplot shows the magnitude distribution modelled with truncated Gaussians over a set of Sérsic index bins and the second subplot shows the measured S/N distribution after the *Euclid*-like conversion of the generated images. The large panel shows the estimate on the shear bias between the realistic and parametric branches calculated according to Eq. (29).

either the *Euclid* cosmic shear analysis needs to be aware of the value of the  $\mu$  and  $c$  biases across the tomographic bins or the simulation pipeline has to be extended to include galaxies with realistic, complex morphologies. While the exact amount of bias still has to be determined using LENSMC and, optionally, other shape measurement codes, a bias at the  $|\delta\mu| \gtrsim 7.0 \times 10^{-3}$  level cannot be neglected, as it lies above the target calibration accuracy determined for the *Euclid* science goals. While it is of course clear that the bias might be reduced for LENSMC, we note that the order of magnitude found using KSB lies in the range of previous estimates from Mandelbaum et al. (2015). Therein, also other shape measurement methods relying on for example model fitting were tested, with similar bias differences of at least  $10^{-3}$  found. This suggests that we may expect a similar order of magnitude for LENSMC and assume relevance for the *Euclid* analysis. The aforementioned *Euclid* requirements were, however, determined for the full Wide Survey footprint and will hence be lower for DR1.

### 6.4. Bias dependence on morphological parameters

The contributions to the bias originate from peculiar shapes themselves, such as spiral arms, merger remnants, tidal streams, or other possible sources of complexity. Moreover, substructures also exist to some degree in galaxies whose shape can be extremely well modelled by a single or double Sérsic profile. Such features correlate with the morphological proxies and



**Fig. 14.** Results of shear bias difference measurements as a function of structural parameters and morphological proxies. The points with error bars depict the multiplicative and additive biases in the respective bin, the dotted lines with the blue or red areas show the means and errors for  $\Delta\mu_1$ ,  $\Delta c_1$  when measuring over the whole sample of  $1 \times 10^6$  galaxies. The data points are placed at the lower bound of the respective bin, the right-most points include all objects above this threshold.

the Sérsic profile parameters presented in Sect. 3. Using the  $CAS+GM_{20}$  statistics as well as Sérsic index  $n$ , and half-light-radius  $r_e$ , we can determine the bias as a function of these properties.

To do so, we take the measurement at fixed magnitude  $m_{AB} = 20.0$  (to ignore S/N effects on the shear bias difference), bin the measured distributions of  $g_1, g_2 = 0$  for each parameter and then average over the shears of all objects in the respective bin. As the overall number of sample galaxies is not exceptionally high, the errors in bins with low number counts can thus again be potentially large. Nevertheless, this gives a rough estimate on how the occurrence of specific disturbed morphologies can effect the shape measurement.

In Fig. 14, we show the results of this analysis. Looking at the additive bias difference  $\Delta c_1$ , we show that it is consistent with zero for all bins. This is expected due to the isotropy of the PSF. In contrast, the multiplicative bias difference displays clear dependencies on all the metrics. First of all, it decreases slightly with Sérsic index, which can be anticipated, as higher Sérsic indices indicate highly peaked galaxies with usually smooth profiles at the given pixel scale. Looking at the half-light radius, we observe an initial increase of the bias difference, followed by a sharp decline after  $r_e \gtrsim 0.5$ . Within the Gini- $M_{20}$  space, both parameters induce a decrease in the multiplicative bias with increased value. High  $G$  values indicate strongly peaked galaxies, where the bias difference here drops to zero as both branches are almost identical. Furthermore,  $M_{20}$  is related to merger signatures such as for instance multiple nuclei as it traces the over-

all moments of the brightest image pixels and does not rely on circular apertures such as  $C$  (Lotz et al. 2004).

Additionally, we detect a correlation between the concentration  $C$  and  $\Delta\mu$ . While initially higher  $C$  relates to an increased bias, very concentrated objects again lead to a reduction of the bias difference. The concentration parameter  $C$  has furthermore been shown to have a strong positive connection to intrinsic galaxy properties such as stellar mass and velocity dispersion (Graham et al. 2001). Given that the more massive galaxies, as well as high Sérsic index galaxies, are almost exclusively elliptical and thus do not possess spiral arms or large active star-forming regions, these result can be anticipated, as images of such objects resemble their Sérsic-fits more closely and thereby do not alter the estimated shear to the same degree as early-type galaxies. For Sérsic indices  $n > 4$  this results in a remaining bias of  $\Delta\mu_1 \sim 0.3\%$ . One could argue that for such objects the bias should reduce even lower, given the absence of any substructure. However, the sub-sample of such galaxies also contains cases where the best-fit Sérsic profile might be highly peaked, but the underlying true shape can exhibit weak tails or even multiple nuclei that cannot be modelled by a Sérsic fit. Thus, the set of galaxies which closely match in both branches is contaminated, resulting in a small residual bias.

The trends for the asymmetry and the smoothness or clumpiness also indicate interesting behaviour. Asymmetries of galaxies are mainly caused by merger signatures, spiral structures, and star-forming regions (Conselice 2003; Hambleton et al. 2011). The low- $A$  population includes ellipticals and high-Sérsic index galax-

ies, where the previous morphological tracers showed a clearly reduced  $\Delta\mu$ . Similarly, the symmetry, which traces the clumpiness, shows an increase of the bias with  $S$ . This means that high- $S$  objects, which often possess multiple nuclei that are not captured by Sérsic models, lead to strong shape differences in the branches and thus increase the relative bias difference. For the asymmetry on the other hand, the bias decreases with  $A$ , although with an initially low value for the lowest- $A$  bin. The correlation is therefore not purely linear, similar to the concentration.

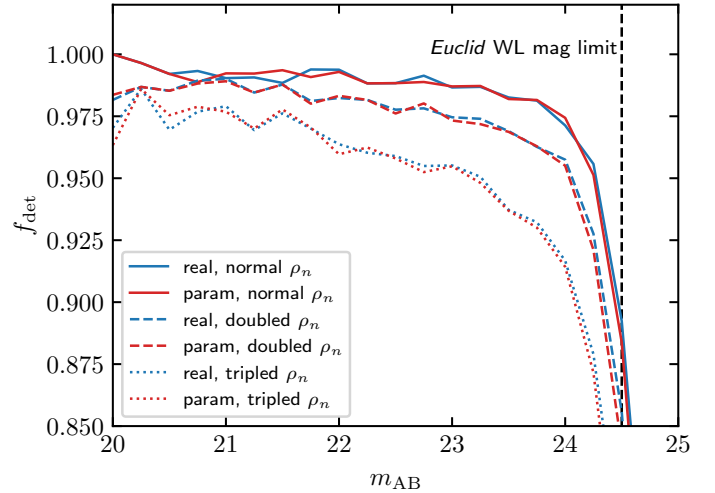
Aside from the intrinsic increase of asymmetry and clumpiness with increasing redshift, an additional effect amplifies this effect. Observing all galaxies in the same filter, in this case the *Euclid* VIS band, means that higher redshift objects will be seen at restframe wavelengths further in the UV regime, which then leads to even more complex and irregular observed morphologies. Especially for *Euclid*, where very faint objects will represent the majority of the sample, this could increase the shear bias from complex morphology even more at low  $S/N$ , which is counteracted by the washing out of structures at faint magnitudes. Thus, simply using the F814W band to mimic *Euclid* VIS images can also influence the bias difference, as the two bands do not fully coincide, with VIS being much broader. Future improvements of the generative model will alleviate this issue by training on multi-band data which can be used to then combine to the full VIS baseline.

Overall, these results show that the shear bias measurement does not depend on only structural parameters such as Sérsic index or effective radius, but can also be traced by statistics for complex morphologies. As neglecting substructure of galaxies in shear calibration naturally introduces an estimator-dependent bias that is relevant for stage IV surveys, the bias is clearly dependent on the analysed galaxy population and, the degree of peculiarity in their shapes. Especially spirals, merger signatures and multiple components increase the bias substantially. On the other hand, its magnitude is irrelevant for *Euclid*'s accuracy requirements for ellipticals and in general symmetric or strongly peaked galaxies.

### 6.5. Detection and blending

Cosmic shear results are not only dependent on the intrinsic biases of the shape measurement methods themselves, but also on selection effects and the detection process (Hirata & Seljak 2003; Hoekstra et al. 2017; Fenech Conti et al. 2017; Euclid Collaboration: Martinet et al. 2019; Kannawadi et al. 2019; Hoekstra et al. 2021). This can enter the results in two ways: Either by actual detection differences in calibration simulations and real observations or by failure of the shape measurement method to determine the ellipticities of individual galaxies. The detection significance of a faint object is usually dependent on its alignment relative to the shear or the PSF (Hirata & Seljak 2003; Bernstein & Jarvis 2002). The detection process happens with a circular kernel, meaning that the images of galaxies with shear perpendicular to their major axis will appear rounder because of the shear, leading to an increased detection probability. This can lead to a multiplicative bias. Moreover, the detection bias estimate is correlated with the shape noise, as objects might be only detected in one of the rotated versions if shape noise cancellation is applied. In simulations, detection bias can thus be removed by requiring the detection of both rotated counterparts, which is, however, not feasible for real observations (Pujol et al. 2020; Hoekstra et al. 2021). We note that objects with failed shear estimates for one of the pair galaxies were previously omitted for the analysis of  $\Delta\mu_1$ , meaning that the results presented in the previous subsections do not include selection bias from KSB.

Given our two measurement branches, it is a necessary step to check if there is a difference in the detection with



**Fig. 15.** SECTRATOR detection fraction in the realistic (blue) and parametric (red) branch. The solid, dashed, and dotted lines show the results for a regular, doubled, and tripled source number density  $\rho_n$ .

SECTRATOR (Bertin & Arnouts 1996) if we do not look at isolated galaxies, but full image scenes with potential blending effects. This is relevant for METACALIBRATION too (Sheldon & Huff 2017), which will also be applied for *Euclid* WL. This method is generally unbiased for isolated galaxies, and can account for selection effects if the selection is performed on a measured quantity of the artificially sheared images (Sheldon & Huff 2017), but will suffer from detection biases or other possible selection effects (Sheldon et al. 2020).

To first of all check if there is a relevant difference from detection itself, we create image scenes per measurement branch, realistic and parametric, with zero applied shear. We use galaxies from our model drawn with the aforementioned magnitude-morphology correlation and paste them at random positions in scenes of size  $4000 \times 4000$  pixels, roughly matching the size of one *Euclid* VIS CCD. We furthermore include additional galaxies up to  $m_{AB} = 29.0$ , as Hoekstra et al. (2017), Euclid Collaboration: Martinet et al. (2019) previously showed the significance of including faint galaxies beyond the *Euclid* WL magnitude limit for the shear measurement bias. Simultaneously, matching scenes with the  $90^\circ$ -rotated galaxies are created, to perform shape noise cancellation and increase the sample size. Afterwards, Gaussian read noise and Poisson shot noise are added to the image, with matching noise fields in rotated scene pairs. While our simulations for now ignore blending with large and bright objects, as well as stars and diffraction spikes, we expect this to influence both branches in the same manner. In the current analysis we only investigate the impact of realistic and parametric light distributions as such. Moreover, galaxies are not randomly positioned in reality, leading to increasing blending fractions due to clustering, especially by faint objects, which we here ignore for now. We assume an average galaxy number density of  $\rho_n \sim 250 \text{ arcmin}^{-2}$  up to  $m_{AB} = 29.0$ , but also repeat the process with doubled and tripled number densities to increase the amount of blended objects.

After rendering of the simulations, we run SECTRATOR on the images scenes and first compare the blending and detection fractions of both measurement branches. Looking at the blending fraction  $f_{\text{blend}}$ , we find that SECTRATOR flags more objects as blended in the realistic branch, and also detects overall more galaxies therein. The overall blending fraction is 2.8%, 4.2%, or 9.2% for realistic objects with normal, doubled, or tripled number density, respectively, or 2.1%, 4.9%, and 8.5% with double-

Sérsic profiles (calculated from the number of detected galaxies and the number of flagged de-blends). We note that these numbers strongly depend on the SEXTRACTOR parameters, which we list in Appendix E.

The increased blending fraction arises from the complex structure and most likely irregular galaxies, where objects with multiple nuclei might be de-blended into separate objects. While the overall difference is not large, with around 2% more detections in the realistic branch over the full magnitude range (only 0.2% below  $m_{AB} = 24.5$ ), this still has to be kept in mind. SEXTRACTOR sometimes struggles to distinguish irregular objects as single galaxies and thus detects some of them inaccurately. Overall, some discrepancy between the detections is expected, especially at fainter magnitudes, where most of the detection differences can be observed. If the best-fit double-Sérsic has a lower surface brightness, for example for asymmetric objects with faint tails that are only fitted with a very broad profile with large  $r_e$ , a portion will fall below the SEXTRACTOR detection limit, while the respective corresponding realistic object might still be selected. We show a cutout of a simulated frame with examples of detection differences in Appendix E.

To show how the detection differences depend on magnitude, we depict in Fig. 15 the detection fraction  $f_{\text{det}}$  for all measurements. The sample is almost complete up to a magnitude of around 24.0 with the standard galaxy number density, with some loss of objects for increased  $\rho_n$ , which can be expected due to blending. The detection fraction for both types of simulations is almost identical, although with slight differences in the number of detected galaxies, that is  $\sim 0.2\%$  more in the realistic branch at standard  $\rho_n$  for  $m_{AB} \leq 24.0$ . Which exact galaxies are detected varies slightly over the magnitude bins, as seen by the solid lines. Furthermore, there is more scatter for bright objects at higher blending fractions, which can be relevant for simulations where sources are not randomly positioned. Therefore, estimates have to be made on how such slight selection effects apparent even at bright magnitudes for both measurement branches are relevant and how selections effects between real observations and parametric simulations in general affect *Euclid*.

## 6.6. Detection bias

In order to determine the actual shear biases from detection-related selection, we create the simulations with standard  $\rho_n$  for each value of the constant shears  $g_1 \in [-0.1, 0.1]$  used for the previous estimates on the morphology bias and run SEXTRACTOR on the full frames. We also produce identical images with the  $90^\circ$  rotated versions to determine detection differences due to orientation with respect to the shear and analyse them in the same manner. Then we select only the ellipticities of galaxies that were actually detected by searching the nearest-neighbour galaxies for each input in the output catalogue. If the centroid offset is larger than three pixels (Hoekstra et al. 2021), we count this as a mismatch and thus a non-detection by SEXTRACTOR, as it is then most likely not a detection of an input shear catalogue object but from the additionally added faint galaxies with  $24.5 < m_{AB} \leq 29$ . Afterwards, we determine the multiplicative biases in both branches for three different cases:

- i Removing objects that are non-detections by SEXTRACTOR, separately in each branch:  $\mu^D$
- ii Removing non-detections and blended objects as identified by SEXTRACTOR with  $0 < \text{FLAGS} \leq 3$ , separately in each branch:  $\mu^{\text{DB}}$
- iii Removing object with failed KSB measurements, without removing rotated counterparts if they are measured by KSB, ignoring detection differences, separately in each branch:  $\mu^{\text{KSB}}$ .

We determine the detection and selection biases separated by the contributions from SEXTRACTOR and KSB, including from blending. To mimic the detection bias results from Hoekstra et al. (2021) for a *Euclid*-like setup, we here do not perform a S/N cut, but simply select galaxies with  $20 \leq m_{AB} \leq 24.5$ .

To furthermore avoid model-dependent influences on the detection bias and isolate the SEXTRACTOR contribution, the bias is hereby not calculated with the KSB estimates, but with the true input ellipticities after shearing of the galaxies, following Fenech Conti et al. (2017), Kannawadi et al. (2019), and Hoekstra et al. (2021). Here, however, we do not have intrinsic ellipticities of the objects, at least not in the realistic branch, as they are not parametric profiles, but just sampled shapes from the generative model. As the parametric galaxies are by construction modelled on the realistic branch and we have shown in Fig. 11 that their ellipticity distributions match, we use the input ellipticities from the Sérsic objects for both branches. The complex observed ellipticities, consisting of the intrinsic shape  $\epsilon_s$  and the added reduced shear  $g = g_1 + ig_2$  with  $g_2 = 0$  and  $|g| < 1$  are given by

$$\epsilon_o = \frac{\epsilon_s + g}{1 + g^* \epsilon_s}, \quad (30)$$

where  $*$  denotes a complex conjugate (Seitz & Schneider 1997; Hoekstra et al. 2021). The complex intrinsic ellipticities are calculated from the absolute ellipticity values of the Sérsic fits and their rotation angles. The bias is then again determined via a linear fit of the means of the ellipticities  $\epsilon_o$  with the applied cuts i–iii as a function of  $g_1$ .

We first determine the biases from all three cases separately for realistic and parametric objects, and then compute the bias differences between the branches  $\Delta\mu^D, \Delta\mu^{\text{DB}}, \Delta\mu^{\text{KSB}}$  to quantify how the selection process affects the simulation sets differently. There, the difference is given by  $\Delta\mu^x = \mu_{\text{real}}^x - \mu_{\text{param}}^x$ , with  $x \in \{D, \text{DB}, \text{KSB}\}$ . We summarise the results in Table 1 for all three cases.

A strong source of selection bias comes from KSB itself, where keeping objects where one from each rotated galaxy pair is not measured produces a strong bias that differs between the branches, leading to a residual of  $\Delta\mu^{\text{KSB}} = -(3.8 \pm 3.1) \times 10^{-3}$ . This is higher than the effect from KSB selection determined by Hoekstra et al. (2021), but also very noisy due to the large size of the error. We note that there exist differences in the respective simulation and measurement setups that can account for such variations. Different KSB implementations and morphologies were applied, which naturally leads to variations. The GalSim KSB implementation fails for a substantial number of objects ( $\sim 4\%$ ), with more failures in the realistic branch. This can be expected due to the existence of irregular shapes that make an ellipticity estimate more difficult.

More importantly, the SEXTRACTOR results significantly differ in both branches due to detection and blending. We find that the detection and detection-plus-blending bias estimates lie close to the values from Hoekstra et al. (2021) for parametric objects. The slight differences can be accounted for by varying PSF and morphology models (Sérsic vs. double-Sérsic), as well as the choices of detector, background, and noise characteristics. The biases for realistic galaxies, however, strongly deviate from these results. Overall, this leads to a bias difference  $\Delta\mu^{\text{DB}} = (6.3 \pm 0.9) \times 10^{-3}$  associated with bias from the detection or removal of objects flagged as a blend. This increase compared to the only detection-related bias difference  $\Delta\mu^D = (2.2 \pm 0.8) \times 10^{-3}$  is especially interesting, as it suggests that blending leads to a positive bias in the realistic branch, but to a negative bias for

**Table 1.** Detection and selection biases.

	$\mu^D [10^{-3}]$	$\mu^{DB} [10^{-3}]$	$\mu^{KSB} [10^{-3}]$
without S/N cut:			
realistic	$-7.0 \pm 0.6$	$-5.6 \pm 0.7$	$-28.3 \pm 2.2$
parametric	$-9.2 \pm 0.6$	$-11.9 \pm 0.6$	$-25.3 \pm 2.4$
$\Delta\mu^x$	<b><math>2.2 \pm 0.8</math></b>	<b><math>6.3 \pm 0.9</math></b>	<b><math>-3.3 \pm 3.1</math></b>
with S/N cut:			
realistic	$-3.5 \pm 0.7$	$-2.8 \pm 0.7$	$-15.1 \pm 1.0$
parametric	$-4.6 \pm 0.6$	$-6.8 \pm 0.7$	$-11.3 \pm 0.9$
$\Delta\mu^x$	<b><math>1.1 \pm 0.8</math></b>	<b><math>4.0 \pm 0.9</math></b>	<b><math>-3.8 \pm 1.2</math></b>

**Notes.** Summary of detection/selection bias contributions from SEXTRACTOR and KSB for both branches, and the bias differences  $\Delta\mu^x$  between both branches, with  $x \in \{D, DB, KSB\}$ , for the influences from detection, detection plus blending, and KSB measurement selection.

parametric shapes. This can be explained by the aforementioned blending differences for irregular shapes, where objects modelled with realistic morphologies are more likely to be flagged. Thus, not only are galaxies that are actually blended omitted but also isolated objects that look similar to blends. We find that the realistic objects which are additionally flagged as blended in comparison to the parametric simulations have almost exclusively large effective radii. Such galaxies often show for example spiral arms or other complex substructure in the realistic branch, which can sometimes lead to SEXTRACTOR flagging despite the absences of actual blending. Flagging due to the shape itself will, however, most likely appear in a galaxy pair for both versions, leading to no selection effects. While it then may seem counter-intuitive that the bias decreases compared to the double-Sérsic simulations, it can be explained by looking at the galaxies that are flagged in both branches: Some of them may be irregular and thus also have a flagged rotated counterpart. This effectively reduces the number of objects with differently flagged rotations, which in return reduces the number of objects that contribute to the blending bias on the ellipticity estimate.

We note that using the true input ellipticities of the profiles in both branches can be an additional source of systematic error in the estimate of the bias in the realistic branch. This can though not be corrected for, as no true ellipticities exist for non-parametric shapes. An alternative option could be to determine the true ellipticities with a shear response on the noise-free shapes, similar to the METACALIBRATION method. This though exceeds the scope of this work. Using the measured KSB  $g_1$  values on the other hand would also potentially lead to difference due to the morphology bias, although this might be not significant, as mostly faint galaxies induce a detection bias, where the difference due to morphology is low anyway. Overall, we expect that using the identical ellipticities in both branches is the most robust method to avoid model bias influences.

Another effect that will affect both the shear measurement and detection biases are correlated ellipticities. For example in dense regions, such as galaxy clusters, the intrinsic shapes may be correlated and more heavily blended, while also exhibiting overall higher shears outside the 10% interval where the shears are currently measured. This will be addressed in future work by simulating scenes with realistic positions and correlated shapes.

As we applied an additional cut of  $S/N > 10$  during the morphology bias estimate, the detection and blending bias that is relevant for the results from Sect. 6.3 will be lower, as there will be fewer objects without a detection. With this additional S/N cut, the bias differences are reduced to  $\Delta\mu^D = (1.1 \pm 0.8) \times 10^{-3}$ ,

$\Delta\mu^{DB} = (4.0 \pm 0.9) \times 10^{-3}$ , and  $\Delta\mu^{KSB} = -(3.8 \pm 1.2) \times 10^{-3}$ , respectively. While the absolute values in the respective branches increase when low signal-to-noise galaxies are included, the relative difference between the branches stays stable within the error bars.

The results show that realistic morphologies not only bias the shape estimator itself, but can account for detection and selection differences between parametric image simulations and observed data. Their different blending characteristics can induce strong biases in the ensemble shape measurement, independently of the algorithm applied to measure the shear signal. Even when neglecting these additional blending effects, the effect from the detection alone leads to a bias that is significant for the *Euclid* DR1 and independent of the applied shape measurement method. This underlines the need to derive accurate bias corrections from simulations that include complex morphologies for the shape measurement methods that are used for the *Euclid* science analysis.

## 7. Conclusions

In this work, we have performed a proof-of-concept study about the calibration of model biases from complex galaxy morphologies for the *Euclid* mission. We have presented a new deep generative CNN for high-resolution, noise-free galaxy postage stamps that uses a physically motivated latent space via the wavelet scattering transform. In addition to fast training and a latent space model that encodes morphological information, which can also be independently applied in unsupervised learning for galaxy classification, the model generalises well over the plethora of galaxy shapes. It is able to recover structural parameters and morphological proxies such as the  $CAS + GM_{20}$  statistics between input and output distributions, which are also correlated with each other. We showed that various sampling techniques can be leveraged to obtain new objects from the generative model, for example, also via conditional sampling by the Sérsic index  $n$ . In subsequent work, we will extend the model to incorporate correlated sampling via redshift, magnitude, and structural parameter distributions, as for now we have only assumed a connection between the magnitude and the Sérsic index.

Next, we estimated the bias difference introduced by realistic galaxy morphologies in shape measurements. For this, we simulated galaxies from the model together with their parametric counterparts, applying additional shear. With the KSB method, we found a multiplicative bias difference at the  $\Delta\mu_1 \sim 0.7\%$  level for a realistic magnitude and S/N distribution with an empirical dependency on the Sérsic index. This bias therefore lies above the target shear accuracy for *Euclid* by a factor of three. While this method will not be applied to *Euclid* data, it confirms the findings from Mandelbaum et al. (2015) for a *Euclid*-like setup and suggests that the bias will have to be calibrated for *Euclid* given the results with other shape measurement methods from this previous work. We detected a correlation between  $\Delta\mu_1$  and the S/N, the Sérsic index  $n$ , the half-light radius, and the  $CAS + GM_{20}$  statistics of the galaxies. Furthermore, we showed that the bias reduces heavily to below a  $\sim 0.3\%$  level for objects whose shapes closely resemble each other, for example smooth profiles at high S/N or low-S/N objects. This proves that the origin of the measured model biases really lies in the complex structures and is not an artefact of the image rendering process. Still, this first measurement assumes a random distribution of galaxy complexity across magnitudes. While we implemented a relation between  $m_{AB}$  and  $n$  based on observations, there is still not much quantitative information about the dependency on other morphological proxies. In reality, where the

fraction of irregular galaxies is highest at high redshifts, the bias might be reduced, as the complex substructures that arise from various mechanisms and thereby emerge in specific regions of the  $CAS+GM_{20}$  space could be washed out at a low S/N. This would lead to an overall lower effect on the shear estimation. However, if that is not the case, high- $z$  samples could be more strongly biased due to their increased irregularity. Either way, the bias difference from galaxy morphology is most likely a function of the source redshift. Therefore, ignoring the effect will influence the tomographic shear calibration and thus also bias the cosmological results.

We also showed that the detection by SExtractor is not completely identical between both measurement branches when creating matching images scenes. The detection process recovers almost the same galaxies, but more galaxies are detected below the  $m_{AB} = 24.5$  magnitude limit of the Euclid Wide Survey in the realistic image scenes, which we attribute mostly to the fact that in some cases SExtractor de-blends complex shapes with multiple brightness peaks into separate galaxies. Thus, selection effects caused by detection and blending discrepancies lead to a bias difference of  $\Delta\mu \sim 4.0 \times 10^{-3}$  between realistic and parametric objects over the full sample with an applied S/N > 10 cut. This implies that realistic morphologies must be included in weak lensing image simulations in order to reach the *Euclid* requirements for the estimation of detection biases alone, disregarding model biases, which are independent of the employed shape measurement method.

Furthermore, KSB is not the designated shape measurement algorithm for *Euclid*; hence future calibrations will need to be performed with LENSMC or METACALIBRATION, where the shape measurement bias could be decreased due to the forward modelling plus fitting nature of LENSMC and the generally unbiased approach of the shear response measurement by METACALIBRATION, where detection and selection effects are still relevant. Fitting codes, for example, have previously been shown to exhibit a lower morphology bias in comparison to KSB (Massey et al. 2007a; Mandelbaum et al. 2015; Hoekstra et al. 2021), though they were only tested with simpler galaxy models via shapelets or emulation from HST data and still at a level relevant for *Euclid*. While this does not necessarily mean that a measurement with LENSMC will be less biased due to complex morphology, the procedure of the code indicates that the bias will be reduced, as LENSMC fits more realistic double-Sérsic models to the data. Thus, a wide variety of shapes can be modelled with such model fitting codes, likely leading to a reduction in the relative biases between realistic and parametric galaxies. This will be determined quantitatively in future work. Once the redshift dependency has been embedded into the model framework, the cosmic shear morphology bias of *Euclid* can be calibrated as a function of tomographic redshift bins for both  $g_1$  and  $g_2$  using LENSMC and METACALIBRATION. We note that this project is a proof-of-concept study whose aim is to obtain a general estimate of the relevance of the bias and to present the deep learning model that can be applied for the full calibration of this bias within the *Euclid* simulation pipeline. Nevertheless, even without the caveats of an analysis solely with KSB, the detection bias results show that there will be residual biases in the shear estimate due to detection differences in simulations and observations if complex morphologies are not accounted for. This alone justifies the usage of realistic shapes for shear calibration.

Eventually, the CNN can be even further extended towards multiple colour channels using HST training data from multi-band observations of, for example, CANDELS data in the GOODS, COSMOS, UDS, and AEGIS fields (Koekemoer et al. 2011; Grogin et al. 2011; Stefanon et al. 2017). This will enable cali-

bration of the *Euclid* cosmic shear analysis for the colour gradient bias, which is caused by the wavelength dependence of the *Euclid* PSF and the spatial colour gradients within galaxies, allowing future calibrations of the colour gradient bias to go beyond the simplified bulge and disc analyses employed in Semboloni et al. (2013), Er et al. (2018). This bias cannot be directly calibrated by METACALIBRATION or METADETECTION (Sheldon et al. 2020), which underlines the necessity of this analysis for next-generation cosmological surveys. With realistic multi-band morphologies from the generator, we intend to determine a colour gradient bias that depends on the actual local SED distribution of each galaxy and thus permits more accurate calibration.

**Acknowledgements.** The Innsbruck authors acknowledge support provided by the Austrian Research Promotion Agency (FFG) and the Federal Ministry of the Republic of Austria for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK) via the Austrian Space Applications Programme with grant numbers 899537, 900565, and 911971, as well as support provided by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant 415537506. HHO acknowledges support by the European Union (ERC-AdG OCULIS, project number 101053992). LL is supported by the Austrian Science Fund (FWF) [ESP 357-N]. GC acknowledges support provided by the United Kingdom Space Agency with grant numbers ST/X00189X/1, ST/W002655/1, ST/V001701/1, and ST/N001761/1. ANT acknowledges the UK Space Agency for its support. 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 Agenzia Spaziale Italiana, the Austrian Forschungsförderungsgesellschaft funded through BMK, the Belgian Science Policy, the Canadian Euclid Consortium, the Deutsches Zentrum für Luft- und Raumfahrt, the DTU Space and the Niels Bohr Institute in Denmark, the French Centre National d'Etudes Spatiales, the Fundação para a Ciência e a Tecnologia, the Hungarian Academy of Sciences, the Ministerio de Ciencia, Innovación y Universidades, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Nederlandse Onderzoeksschool Voor Astronomie, the Norwegian Space Agency, the Research Council of Finland, 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 ([www.euclid-ec.org](http://www.euclid-ec.org)).

## References

- Abraham, R. G., van den Bergh, S., Glazebrook, K., et al. 1996, *ApJS*, 107, 1
- Abraham, R. G., van den Bergh, S., & Nair, P. 2003, *ApJ*, 588, 218
- Agarap, A. F. 2018, arXiv e-prints [arXiv:1803.08375]
- Agustsson, E., Sage, A., Timofte, R., & Van Gool, L. 2017, arXiv e-prints [arXiv:1711.01970]
- Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, *PASJ*, 70, S8
- Andreux, M., Angles, T., Exarchakis, G., et al. 2020, *J. Mach. Learn. Res.*, 21, 1
- Angles, T., & Mallat, S. 2018, arXiv e-prints [arXiv:1805.06621]
- Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, *ApJ*, 600, 681
- Bernstein, G. M. 2010, *MNRAS*, 406, 2793
- Bernstein, G. M., & Jarvis, M. 2002, *AJ*, 123, 583
- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
- Bojanowski, P., Joulín, A., Lopez-Paz, D., & Szlam, A. 2017, Proc. of the 35th Int. Conf. on Machine Learning, PMLR80, [arXiv:1707.05776]
- Bradley, L., Sipócz, B., Robitaille, T., et al. 2023, <https://doi.org/10.5281/zenodo.8248020>
- Bruna, J., & Mallat, S. 2013, *IEEE Trans. Pattern Anal. Mach. Intell.*, 35, 1872
- Bundy, K., Ellis, R. S., & Conselice, C. J. 2005, *ApJ*, 625, 621
- Cheng, S., & Ménard, B. 2021, arXiv e-prints [arXiv:2112.01288]
- Cheng, T.-Y., Li, N., Conselice, C. J., et al. 2020, *MNRAS*, 494, 3750
- Cheng, T.-Y., Huertas-Company, M., Conselice, C. J., et al. 2021, *MNRAS*, 503, 4446
- Conselice, C. J. 2003, *ApJS*, 147, 1
- Conselice, C. J. 2014, *ARA&A*, 52, 291
- Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, *ApJ*, 529, 886
- Conselice, C. J., Blackburne, J. A., & Papovich, C. 2005, *ApJ*, 620, 564
- Cropper, M., Hoekstra, H., Kitching, T., et al. 2013, *MNRAS*, 431, 3103
- Cropper, M., Pottinger, S., Niemi, S., et al. 2016, in Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave, eds. H. A. MacEwen, G. G. Fazio, M. Lystrup, et al., *SPIE Conf. Ser.*, 9904, 99040Q
- Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010, *MNRAS*, 401, 1043
- Dark Energy Survey Collaboration (Abbott, T., et al.) 2016, *MNRAS*, 460, 1270

- de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, *Exp. Astron.*, **35**, 25
- Er, X., Hoekstra, H., Schrabback, T., et al. 2018, *MNRAS*, **476**, 5645
- Euclid Collaboration (Martinet, N., et al.) 2019, *A&A*, **627**, A59
- Euclid Collaboration (Ilbert, O., et al.) 2021, *A&A*, **647**, A117
- Euclid Collaboration (Brettonniere, H., et al.) 2022a, *A&A*, **657**, A90
- Euclid Collaboration (Scaramella, R., et al.) 2022b, *A&A*, **662**, A112
- Euclid Collaboration (Brettonniere, H., et al.) 2023, *A&A*, **671**, A102
- Euclid Collaboration (Aussel, B., et al.) 2024a, *A&A*, **689**, A274
- Euclid Collaboration (Congedo, G., et al.) 2024b, *A&A*, **691**, A319
- Euclid Collaboration (Scognamiglio, D., et al.) 2025a, *A&A*, **694**, A262
- Euclid Collaboration (Cropper, M., et al.) 2025b, *A&A*, in press, <http://dx.doi.org/10.1051/0004-6361/202450996>
- Euclid Collaboration (Mellier, Y., et al.) 2025c, *A&A*, in press, <http://dx.doi.org/10.1051/0004-6361/202450810>
- Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, *MNRAS*, **467**, 1627
- Freeman, P. E., Izbicki, R., Lee, A. B., et al. 2013, *MNRAS*, **434**, 282
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2014, in *Advances in Neural Information Processing Systems*, eds. Z. Ghahramani, M. Welling, C. Cortes, et al. (Red Hook, NY, USA: Curran Associates Inc.), 27
- Graham, A. W., Trujillo, I., & Caon, N. 2001, *AJ*, **122**, 1707
- Greig, B., Ting, Y.-S., & Kaurav, A. A. 2023, *MNRAS*, **519**, 5288
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, **197**, 35
- Gurvich, A., & Mandelbaum, R. 2016, *MNRAS*, **457**, 3522
- Hackstein, S., Kinakh, V., Bailer, C., & Melchior, M. 2023, *Astron. Comput.*, **42**, 100685
- Hambleton, K. M., Gibson, B. K., Brook, C. B., et al. 2011, *MNRAS*, **418**, 801
- Hernández-Martín, B., Schrabback, T., Hoekstra, H., et al. 2020, *A&A*, **640**, A117
- Hirata, C., & Seljak, U. 2003, *MNRAS*, **343**, 459
- Ho, J., Jain, A., & Abbeel, P. 2020, in *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20* (Red Hook, NY, USA: Curran Associates Inc.)
- Hoekstra, H., Franx, M., Kuijken, K., & Squires, G. 1998, *ApJ*, **504**, 636
- Hoekstra, H., Herbonnet, R., Muzzin, A., et al. 2015, *MNRAS*, **449**, 685
- Hoekstra, H., Viola, M., & Herbonnet, R. 2017, *MNRAS*, **468**, 3295
- Hoekstra, H., Kannawadi, A., & Kitching, T. D. 2021, *A&A*, **646**, A124
- Hoffman, M., Blei, D. M., Wang, C., & Paisley, J. 2013, *J. Mach. Learn. Res.*, **14**, 1303
- Holzschuh, B. J., O'Riordan, C. M., Vegetti, S., Rodriguez-Gomez, V., & Thuerey, N. 2022, *MNRAS*, **515**, 652
- Ivezic, Z., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, **873**, 111
- Jansen, H., Tewes, M., Schrabback, T., et al. 2024, *A&A*, **683**, A240
- Kaiser, N., Squires, G., & Broadhurst, T. 1995, *ApJ*, **449**, 460
- Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, *A&A*, **624**, A92
- Kennicutt, R. C. Jr 1998, *ARA&A*, **36**, 189
- Kilbinger, M. 2015, *Rep. Prog. Phys.*, **78**, 086901
- Kilcher, Y., Lucchi, A., & Hofmann, T. 2017, arXiv e-prints [arXiv:1710.11381]
- Kingma, D. P., & Ba, J. 2014, arXiv e-prints [arXiv:1412.6980]
- Kingma, D. P., & Welling, M. 2013, arXiv e-prints [arXiv:1312.6114]
- Kitching, T. D., & Deshpande, A. C. 2022, *Open J. Astrophys.*, **5**, 6
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, **197**, 36
- Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., et al. 2021, *MNRAS*, **504**, 5543
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints [arXiv:1110.3193]
- Leauthaud, A., Massey, R., Kneib, J.-P., et al. 2007, *ApJS*, **172**, 219
- Li, X., Miyatake, H., Luo, W., et al. 2022, *PASJ*, **74**, 421
- Li, S. S., Kuijken, K., Hoekstra, H., et al. 2023, *A&A*, **670**, A100
- Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, **128**, 163
- MacCrann, N., Becker, M. R., McCullough, J., et al. 2022, *MNRAS*, **509**, 3371
- Mallat, S. 2012, *Commun. Pure Appl. Math.*, **65**, 1331
- Mandelbaum, R. 2018, *ARA&A*, **56**, 393
- Mandelbaum, R., Hirata, C. M., Leauthaud, A., Massey, R. J., & Rhodes, J. 2012a, *MNRAS*, **420**, 1518
- Mandelbaum, R., Lackner, C., Leauthaud, A., & Rowe, B. 2012b, <https://doi.org/10.5281/zenodo.3242143>
- Mandelbaum, R., Rowe, B., Bosch, J., et al. 2014, *ApJS*, **212**, 5
- Mandelbaum, R., Rowe, B., Armstrong, R., et al. 2015, *MNRAS*, **450**, 2963
- Mandelbaum, R., Lanusse, F., Leauthaud, A., et al. 2018, *MNRAS*, **481**, 3170
- Massey, R., Refregier, A., Conselice, C. J., David, J., & Bacon, J. 2004, *MNRAS*, **348**, 214
- Massey, R., Heymans, C., Bergé, J., et al. 2007a, *MNRAS*, **376**, 13
- Massey, R., Rowe, B., Refregier, A., Bacon, D. J., & Bergé, J. 2007b, *MNRAS*, **380**, 229
- Massey, R., Kitching, T., & Richard, J. 2010, *Rep. Prog. Phys.*, **73**, 086901
- Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, *MNRAS*, **382**, 315
- Mukherjee, S., Asnani, H., Lin, E., & Kannan, S. 2018, *Proc. of 33rd AAAI Conf. on Artificial Intelligence*
- Niemi, S.-M., Cropper, M., Szafraniec, M., & Kitching, T. 2015, *Exp. Astron.*, **39**, 207
- Pasha, I., & Miller, T. B. 2023, *J. Open Source Software*, **8**, 5703
- Paszke, A., Gross, S., Massa, F., et al. 2019, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, eds. H. Wallach, H. Larochelle, A. Beygelzimer, et al. (Red Hook, NY, USA: Curran Associates Inc.)
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Portillo, S. K. N., Parejko, J. K., Vergara, J. R., & Connolly, A. J. 2020, *AJ*, **160**, 45
- Pujol, A., Sureau, F., Bobin, J., et al. 2020, *A&A*, **641**, A164
- Refregier, A., Amara, A., Kitching, T. D., et al. 2010, arXiv e-prints [arXiv:1001.0061]
- Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, *Astron. Comput.*, **10**, 121
- Salimans, T., Goodfellow, I., Zaremba, W., et al. 2016, in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, eds. D. Lee, M. Sugiyama, U. Luxburg, et al. (Red Hook, NY, USA: Curran Associates Inc.)
- Schrabback, T., Hartlap, J., Joachimi, B., et al. 2010, *A&A*, **516**, A63
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
- Seitz, C., & Schneider, P. 1997, *A&A*, **318**, 687
- Semboloni, E., Hoekstra, H., Huang, Z., et al. 2013, *MNRAS*, **432**, 2385
- Sérsic, J. L. 1963, *Boletín de la Asociación Argentina de Astronomía La Plata Argentina*, **6**, 41
- Sheldon, E. S., & Huff, E. M. 2017, *ApJ*, **841**, 24
- Sheldon, E. S., Becker, M. R., MacCrann, N., & Jarvis, M. 2020, *ApJ*, **902**, 138
- Shen, S., Mo, H. J., White, S. D. M., et al. 2003, *MNRAS*, **343**, 978
- Shlens, J. 2014, arXiv e-prints [arXiv:1404.1100]
- Smith, M. J., Geach, J. E., Jackson, R. A., et al. 2022, *MNRAS*, **511**, 1808
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints [arXiv:1503.03757]
- Spindler, A., Geach, J. E., & Smith, M. J. 2021, *MNRAS*, **502**, 985
- Stefanon, M., Yan, H., Mobasher, B., et al. 2017, *ApJS*, **229**, 32
- Tewes, M., Kuntzer, T., Nakajima, R., et al. 2019, *A&A*, **621**, A36
- Thorne, B., Knox, L., & Prabhu, K. 2021, *MNRAS*, **504**, 2603
- Tomczak, J. M., & Welling, M. 2018, in *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, eds. A. Storkey, & F. Perez-Cruz (PMLR), *Proc. Mach. Learn. Res.*, **84**, 1214
- Walmsley, M., Lintott, C., Géron, T., et al. 2022, *MNRAS*, **509**, 3966
- Zhang, J., & Komatsu, E. 2011, *MNRAS*, **414**, 1047
- Zuntz, J., Kacprzak, T., Voigt, L., et al. 2013, *MNRAS*, **434**, 1604

- 1 Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria
- 2 Leiden Observatory, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands
- 3 Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
- 4 School of Mathematics and Physics, University of Surrey, Guildford Surrey GU2 7XH, UK
- 5 INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
- 6 IFPU, Institute for Fundamental Physics of the Universe, via Beirut 2, 34151 Trieste, Italy
- 7 INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
- 8 INFN, Sezione di Trieste, Via Valerio 2, 34127 Trieste, TS, Italy
- 9 SISSA, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, TS, Italy
- 10 Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
- 11 INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
- 12 INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- 13 Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
- 14 Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany

- <sup>15</sup> INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy
- <sup>16</sup> Dipartimento di Fisica, Università di Genova, Via Dodecaneso 33, 16146 Genova, Italy
- <sup>17</sup> INFN-Sezione di Genova, Via Dodecaneso 33, 16146 Genova, Italy
- <sup>18</sup> Department of Physics “E. Pancini”, University Federico II, Via Cinthia 6, 80126 Napoli, Italy
- <sup>19</sup> INAF-Osservatorio Astronomico di Capodimonte, Via Moiariello 16, 80131 Napoli, Italy
- <sup>20</sup> INFN section of Naples, Via Cinthia 6, 80126 Napoli, Italy
- <sup>21</sup> Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal
- <sup>22</sup> Faculdade de Ciências da Universidade do Porto, Rua do Campo de Alegre, 4150-007 Porto, Portugal
- <sup>23</sup> Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
- <sup>24</sup> INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
- <sup>25</sup> INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
- <sup>26</sup> Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
- <sup>27</sup> Port d’Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain
- <sup>28</sup> Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, 52056 Aachen, Germany
- <sup>29</sup> Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
- <sup>30</sup> Institut d’Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860, Castelldefels Barcelona, Spain
- <sup>31</sup> INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
- <sup>32</sup> Dipartimento di Fisica e Astronomia “Augusto Righi” – Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- <sup>33</sup> Instituto de Astrofísica de Canarias, Vía Láctea, 38205 La Laguna, Tenerife, Spain
- <sup>34</sup> Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
- <sup>35</sup> European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044, Frascati Roma, Italy
- <sup>36</sup> ESAC/ESA, Camino Bajo del Castillo, s/n, Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
- <sup>37</sup> Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, F-69100 Villeurbanne, France
- <sup>38</sup> Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
- <sup>39</sup> Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
- <sup>40</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Luíís Companys 23, 08010 Barcelona, Spain
- <sup>41</sup> UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, 4 rue Enrico Fermi, 69622 Villeurbanne, France
- <sup>42</sup> Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking Surrey RH5 6NT, UK
- <sup>43</sup> Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisboa, Portugal
- <sup>44</sup> Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
- <sup>45</sup> Department of Astronomy, University of Geneva, ch. d’Ecogia 16, 1290 Versoix, Switzerland
- <sup>46</sup> Université Paris-Saclay, CNRS, Institut d’astrophysique spatiale, 91405 Orsay, France
- <sup>47</sup> INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
- <sup>48</sup> INAF-Istituto di Astrofisica e Planetologia Spaziali, via del Fosso del Cavaliere, 100, 00100 Roma, Italy
- <sup>49</sup> Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191 Gif-sur-Yvette, France
- <sup>50</sup> Space Science Data Center, Italian Space Agency, via del Politecnico snc, 00133 Roma, Italy
- <sup>51</sup> School of Physics, HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK
- <sup>52</sup> Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Via Irnerio 46, 40126 Bologna, Italy
- <sup>53</sup> INAF-Osservatorio Astronomico di Padova, Via dell’Osservatorio 5, 35122 Padova, Italy
- <sup>54</sup> Dipartimento di Fisica “Aldo Pontremoli”, Università degli Studi di Milano, Via Celoria 16, 20133 Milano, Italy
- <sup>55</sup> Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029, Blindern 0315, Oslo, Norway
- <sup>56</sup> Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
- <sup>57</sup> Department of Physics, Lancaster University, Lancaster LA1 4YB, UK
- <sup>58</sup> Felix Hormuth Engineering, Goethestr. 17, 69181 Leimen, Germany
- <sup>59</sup> Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
- <sup>60</sup> Cosmic Dawn Center (DAWN), Copenhagen, Denmark
- <sup>61</sup> Institut d’Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France
- <sup>62</sup> Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France
- <sup>63</sup> Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
- <sup>64</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
- <sup>65</sup> NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
- <sup>66</sup> Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
- <sup>67</sup> Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, 00014 University of Helsinki, Finland
- <sup>68</sup> Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
- <sup>69</sup> Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
- <sup>70</sup> Department of Physics, P.O. Box 64 00014 University of Helsinki, Finland
- <sup>71</sup> Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
- <sup>72</sup> NOVA optical infrared instrumentation group at ASTRON, Oude Hoogeveensedijk 4, 7991PD Dwingeloo, The Netherlands
- <sup>73</sup> INFN-Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
- <sup>74</sup> University of Applied Sciences and Arts of Northwestern Switzerland, School of Engineering, 5210 Windisch, Switzerland
- <sup>75</sup> Universität Bonn, Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
- <sup>76</sup> INFN-Sezione di Roma, Piazzale Aldo Moro, 2 - c/o Dipartimento di Fisica, Edificio G. Marconi, 00185 Roma, Italy
- <sup>77</sup> Aix-Marseille Université, CNRS, CNES, LAM, Marseille, France
- <sup>78</sup> Dipartimento di Fisica e Astronomia “Augusto Righi” – Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, 40129 Bologna, Italy
- <sup>79</sup> Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK
- <sup>80</sup> Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
- <sup>81</sup> Institut d’Astrophysique de Paris, 98bis Boulevard Arago, 75014 Paris, France
- <sup>82</sup> European Space Agency/ESTEC, Keplerlaan 1, 2201, AZ Noordwijk, The Netherlands
- <sup>83</sup> Institut de Física d’Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
- <sup>84</sup> DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 155, 2200 Copenhagen, Denmark

- <sup>85</sup> Centre National d'Etudes Spatiales – Centre spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
- <sup>86</sup> Institute of Space Science, Str. Atomistilor, nr. 409 Măgurele, Ilfov 077125, Romania
- <sup>87</sup> Dipartimento di Fisica e Astronomia “G. Galilei”, Università di Padova, Via Marzolo 8, 35131 Padova, Italy
- <sup>88</sup> Institut für Theoretische Physik, University of Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany
- <sup>89</sup> Université St Joseph; Faculty of Sciences, Beirut, Lebanon
- <sup>90</sup> Satlantis, University Science Park, Sede Bld, 48940 Leioa-Bilbao, Spain
- <sup>91</sup> Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
- <sup>92</sup> Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
- <sup>93</sup> Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, Plaza del Hospital 1, 30202 Cartagena, Spain
- <sup>94</sup> Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- <sup>95</sup> INFN-Bologna, Via Irnerio 46, 40126 Bologna, Italy
- <sup>96</sup> Dipartimento di Fisica, Università degli studi di Genova, and INFN-Sezione di Genova, via Dodecaneso 33, 16146 Genova, Italy
- <sup>97</sup> INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy
- <sup>98</sup> Astronomical Observatory of the Autonomous Region of the Aosta Valley (OAVdA), Loc. Lignan 39, I-11020 Nus (Aosta Valley), Italy
- <sup>99</sup> ICL, Junia, Université Catholique de Lille, LITL, 59000 Lille, France
- <sup>100</sup> ICSC - Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing, Via Magnanelli 2, Bologna, Italy
- <sup>101</sup> Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain
- <sup>102</sup> CERCA/ISO, Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
- <sup>103</sup> Laboratoire Univers et Théorie, Observatoire de Paris, Université PSL, Université Paris Cité, CNRS, 92190 Meudon, France
- <sup>104</sup> Departamento de Física Fundamental. Universidad de Salamanca. Plaza de la Merced s/n., 37008 Salamanca, Spain
- <sup>105</sup> Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
- <sup>106</sup> Istituto Nazionale di Fisica Nucleare, Sezione di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
- <sup>107</sup> Center for Data-Driven Discovery, Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
- <sup>108</sup> Dipartimento di Fisica - Sezione di Astronomia, Università di Trieste, Via Tiepolo 11, 34131 Trieste, Italy
- <sup>109</sup> Minnesota Institute for Astrophysics, University of Minnesota, 116 Church St SE, Minneapolis, MN 55455, USA
- <sup>110</sup> Institute Lorentz, Leiden University, Niels Bohrweg 2, 2333, CA Leiden, The Netherlands
- <sup>111</sup> Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France
- <sup>112</sup> Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
- <sup>113</sup> Department of Physics & Astronomy, University of California Irvine, Irvine, CA 92697, USA
- <sup>114</sup> Department of Astronomy & Physics and Institute for Computational Astrophysics, Saint Mary's University, 923 Robie Street, Halifax, Nova Scotia B3H 3C3, Canada
- <sup>115</sup> Departamento Física Aplicada, Universidad Politécnica de Cartagena, Campus Muralla del Mar, 30202 Cartagena, Murcia, Spain
- <sup>116</sup> Instituto de Astrofísica de Canarias (IAC); Departamento de Astrofísica, Universidad de La Laguna (ULL), 38200 La Laguna, Tenerife, Spain
- <sup>117</sup> Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK
- <sup>118</sup> Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
- <sup>119</sup> Department of Computer Science, Aalto University, PO Box 15400, Espoo FI-00 076, Finland
- <sup>120</sup> Instituto de Astrofísica de Canarias, c/ Via Lactea s/n, La Laguna 38200, Spain. Departamento de Astrofísica de la Universidad de La Laguna, Avda. Francisco Sanchez, La Laguna 38200, Spain
- <sup>121</sup> Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany
- <sup>122</sup> Univ. Grenoble Alpes, CNRS, Grenoble INP, LPSC-IN2P3, 53, Avenue des Martyrs, 38000 Grenoble, France
- <sup>123</sup> Department of Physics and Astronomy, Vesilinnantie 5, 20014 University of Turku, Finland
- <sup>124</sup> Serco for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada 28692, Madrid, Spain
- <sup>125</sup> ARC Centre of Excellence for Dark Matter Particle Physics, Melbourne, Australia
- <sup>126</sup> Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia
- <sup>127</sup> School of Physics and Astronomy, Queen Mary University of London, Mile End Road, London E1 4NS, UK
- <sup>128</sup> Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town 7535, South Africa
- <sup>129</sup> ICTP South American Institute for Fundamental Research, Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil
- <sup>130</sup> IRFU, CEA, Université Paris-Saclay, 91191 Gif-sur-Yvette Cedex, France
- <sup>131</sup> Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Stockholm SE-106 91, Sweden
- <sup>132</sup> Astrophysics Group, Blackett Laboratory, Imperial College London, London SW7 2AZ, UK
- <sup>133</sup> INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy
- <sup>134</sup> Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy
- <sup>135</sup> Aurora Technology for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada 28692 Madrid, Spain
- <sup>136</sup> Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal
- <sup>137</sup> HE Space for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada 28692 Madrid, Spain
- <sup>138</sup> Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- <sup>139</sup> Department of Astrophysics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
- <sup>140</sup> INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
- <sup>141</sup> INFN-Sezione di Genova, Via Dodecaneso 33, 16146 Genova, Italy
- <sup>142</sup> Theoretical astrophysics, Department of Physics and Astronomy, Uppsala University, Box 515, 751 20 Uppsala, Sweden
- <sup>143</sup> Department of Physics, Royal Holloway, University of London, TW20 0EX, UK
- <sup>144</sup> Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA
- <sup>145</sup> Cosmic Dawn Center (DAWN), Copenhagen, Denmark
- <sup>146</sup> Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA
- <sup>147</sup> Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010 New York, NY, USA

## Appendix A: Sérsic-fit difference examples

Figure A.1 shows a sample of galaxies with fit results that strongly deviate between the original image and the estimate of the reconstruction with matched noise, as described in Sect. 4.

It is clear that exclusively strongly peaked, elliptical galaxies with high S/N are affected by low-quality fits obtained using `pysersic` concerning the Sérsic index, while the galaxies with inaccurate half-light radius estimates also include objects with faint features or clear bulge plus disc visibility. For such cases, the SVI method finds inaccurate posterior estimates for at least one of both images. Either full MCMC sampling or improved noise matching (which is not very relevant at high S/N) could be able to reduce this mismatch, albeit only with much longer fitting times ( $\sim 10$  s with Laplacian SVI vs.  $> 60$  s with MCMC). Given the objective of this work, this is not necessary, as we are for now mainly interested in the overall reconstructive power of the generative model for the estimation of the shear bias. In the future, however, accurate Sérsic fits are needed if realistic galaxies of a distinct distribution of structural parameters need to be injected into full-size *Euclid* VIS simulations. Different ways of generating structural parameter fits for billions of galaxies have been investigated within the *Euclid* Morphology Challenge (Euclid Collaboration: Bretonnière et al. 2023).

In principle, the monotonic decrease of the reconstruction accuracy seen in Fig. 3 could also be related to the architecture of the model. The generator tries to minimise the loss during training, for which especially the most abundant populations of morphologies have to be reconstructed accurately. Thus, the overall sparsely populated region of  $n \geq 4.0$  could be neglected. However, looking at the reconstructed images, we do not think that this is the case here.

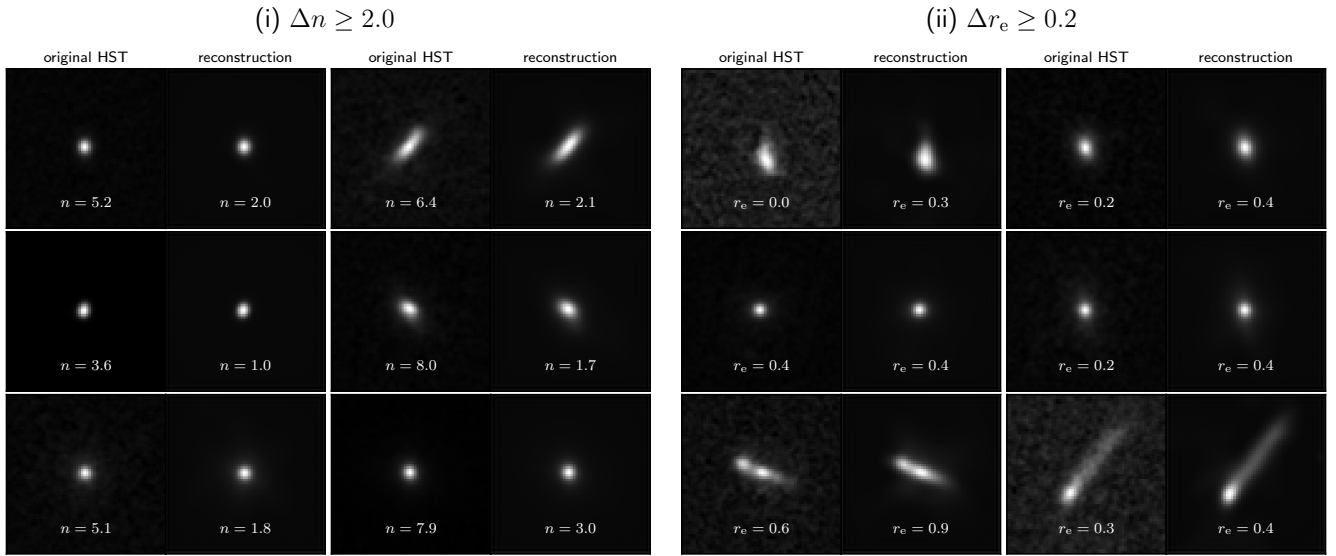
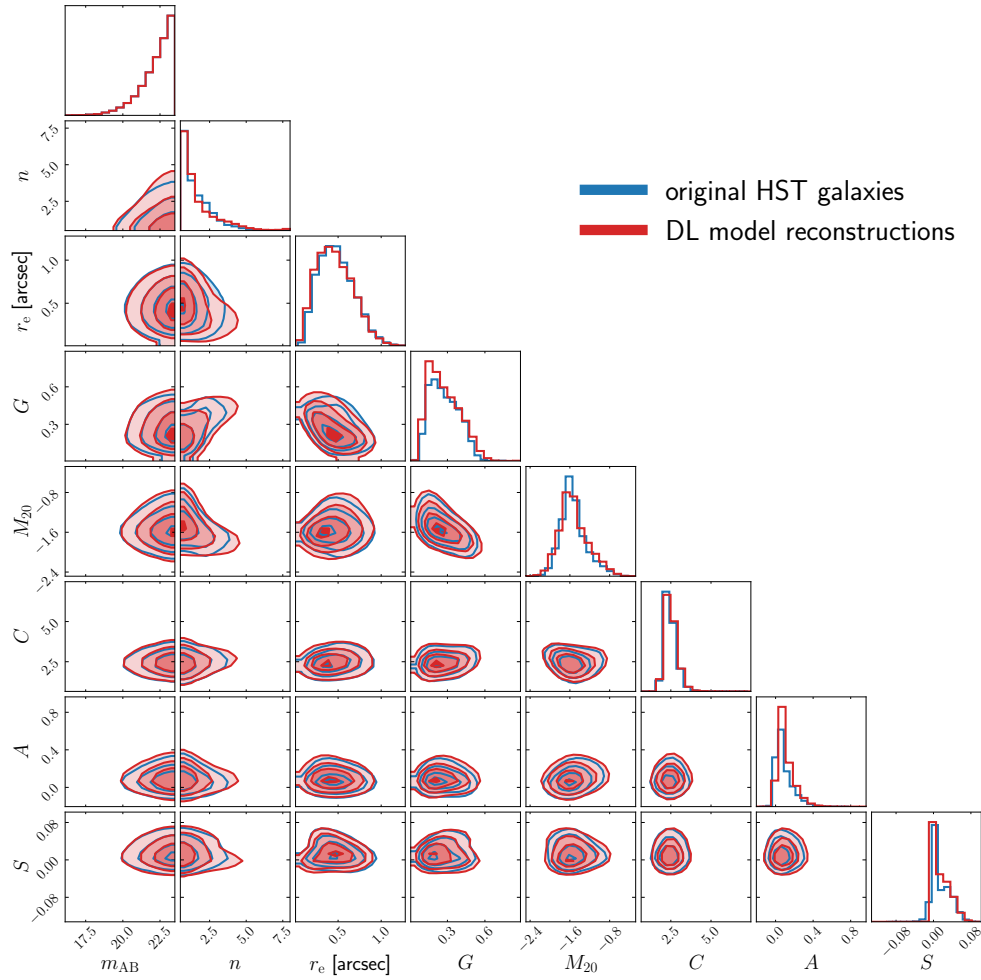


Fig. A.1. Example of galaxies where the difference between the original and generated  $n$  or  $r_e$  exceeds  $\Delta n = 2$  or  $\Delta r_e = 0.2$ .

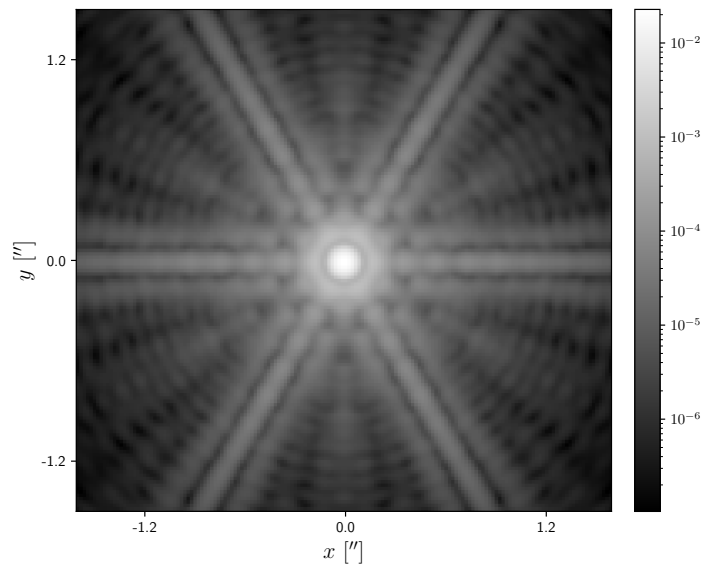
## Appendix B: Two-dimensional morphological parameter distributions

In Fig. B.1 we show a comparison between the 2D correlated distributions of the training input sample and the measurements on the reconstructed outputs from the deep learning model. We can see that the histograms and 2D contours match overall well, with also correlated parameter distributions recovered well by the model.



**Fig. B.1.** Two-dimensional parameter distributions of magnitude, structural parameters, and morphological proxies. The blue contours show the measurements done directly on the HST images, while the red contours depict the results of measuring on the reconstructed deep learning model outputs.

**Appendix C: *Euclid*-like point-spread function**



**Fig. C.1.** Oversampled version of the *Euclid*-like PSF on a logarithmic grey scale. Axes are scaled in arcseconds.

Figure C.1 shows the *Euclid*-like PSF used within this work and mentioned in Sect. 6. The generated galaxies are deconvolved with their training PSF and then convolved with this PSF afterwards to create *Euclid*-like simulations, as illustrated in Fig. 10. For more accurate PSF correction, the PSF model is usually oversampled. Figure C.1 displays a version of the PSF that is oversampled by a factor of 5 on a logarithmic grey scale.

The PSF model is created from a stack of monochromatic PSFs over the wavelength range of the VIS bandpass from 500 nm to 900 nm, with weighting of each component by a stellar spectrum, that is, a Vega spectrum. Additionally, it assumes the *Euclid* telescope's optical characteristics such as mirror size and obscuration. A more detailed description on how this PSF was modelled can be found in Tewes et al. (2019) and Jansen et al. (2024).

## Appendix D: Non-linearity of the shear bias

The results of the shear bias measurement at bright magnitudes shown in Fig. 12 exhibit a strong non-linear component, which has previously also been found by Jansen et al. (2024). To show how second-order effects in the shear bias affect the linear term, we also fit an antisymmetric quadratic function to the shear bias difference. This function is defined as

$$g_i^{\text{obs}} - g_i^{\text{true}} = \alpha |g_i^{\text{true}}| g_i^{\text{true}} + \mu_i g_i^{\text{true}} + c_i + n_i, \quad (\text{D.1})$$

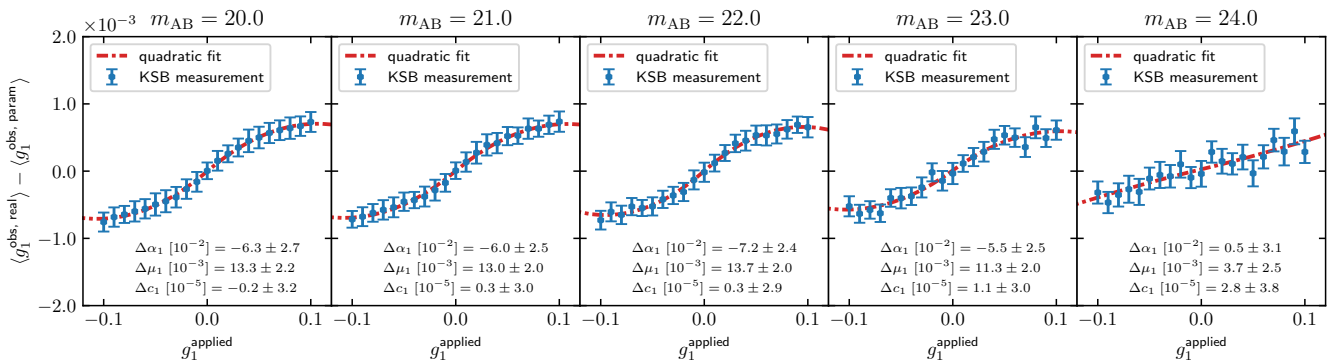
where  $\alpha_i$  is an additional fit parameter. Following Kitching & Deshpande (2022), the linear shear bias can also be written as a spin-2 equation

$$g^{\text{obs}} = (1 + \mu_{i,0}) g^{\text{true}} + \mu_{i,4} (g^{\text{true}})^* + c + n. \quad (\text{D.2})$$

There,  $\mu_{i,0}$  and  $\mu_{i,4}$  are spin-0 and spin-4 quantities, respectively, and \* denotes a complex conjugate. This simplifies the extension to higher-order terms of the shear bias with integer and half-integer power of the shear bias. We refer to Kitching & Deshpande (2022) for a detailed derivation and description of this process. Here, we simply fit the quadratic multiplicative biases with (D.1).

We show in Fig. D.1 the same KSB measurements of the shear bias difference as in Fig. 12. However, in Fig. D.1, the measurements are shown with the antisymmetric quadratic fit from Eq. (D.1). It is apparent that this function fits the data better for bright samples, but no clear higher-order contribution can be seen in the faint galaxy sample, resulting in an overall low term  $\Delta\alpha_1$  there. Compared to the fully linear estimates, the linear term  $\Delta\mu$  is strongly increased at low magnitudes, but fits the estimate within the error bars for the measurement at  $m_{\text{AB}} = 23.0$ .

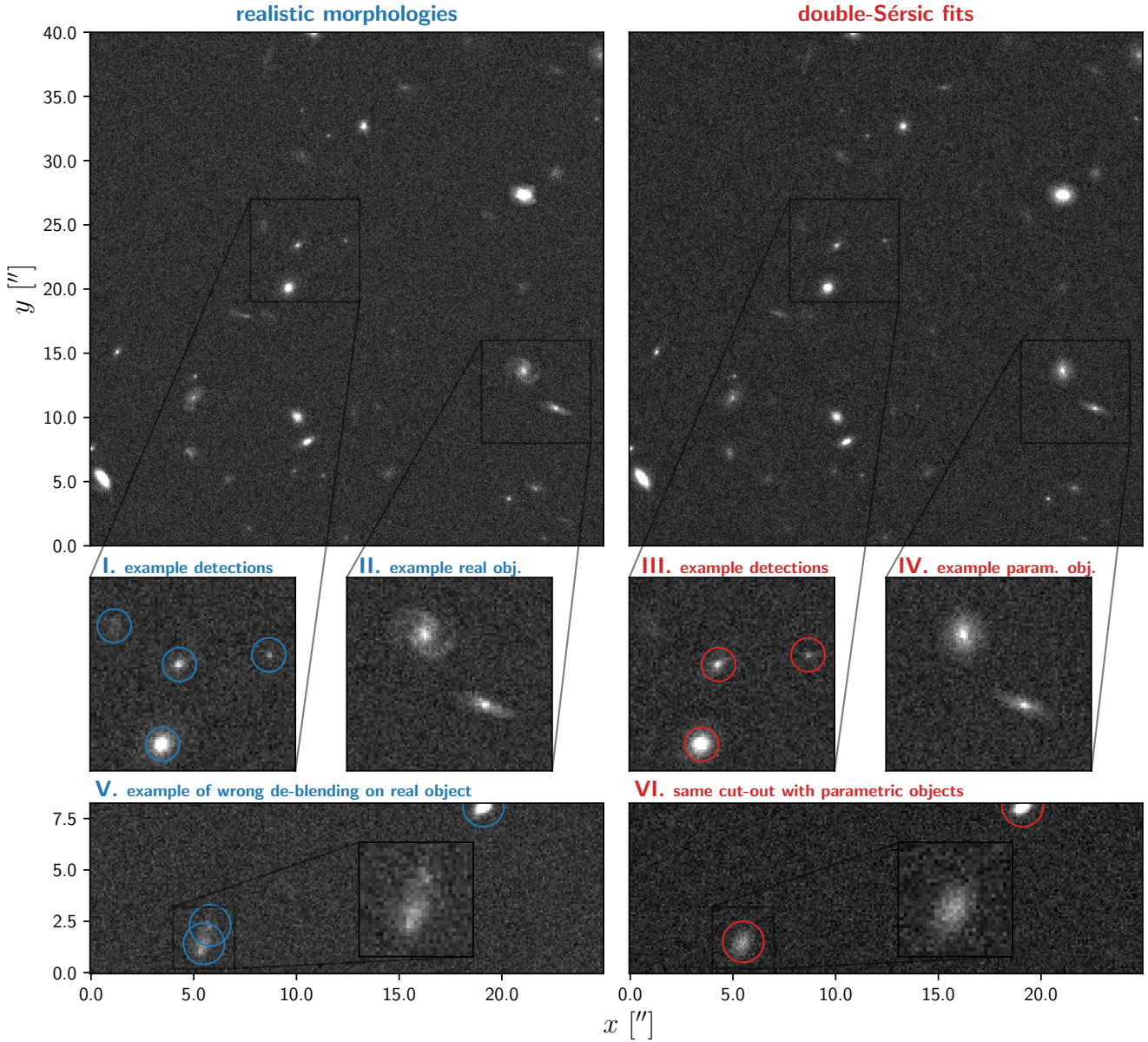
While the non-linearity seems to be relevant for bright objects, we do not observe strong higher-order terms for faint samples or for a realistic magnitude and S/N distribution as shown Fig. 13, which is why we do not explore similar comparisons for this measurement here.



**Fig. D.1.** Results of the cosmic shear measurement with KSB on  $10^6$  simulated galaxies at fixed magnitude, once with realistic galaxies from the deep learning model, and once with their double-Sérsic counterparts. The plots show the relative difference between separate bias estimates on both branches, realistic and parametric galaxies. The red curve shows the best-fit antisymmetric quadratic model according to Eq. (D.1). Inside the plot we summarise the difference  $\Delta$  of the fit parameters  $\alpha_1$ ,  $\mu_1$  and  $c_1$ .

## Appendix E: Scene example for detection comparison

We here show in Fig. E.1 a comparison of image scenes generated from the two branches to illustrate differences in the SExtractor detection, as described in Sect. 6. We show identical cutouts from simulations with a doubled number density  $\rho_n = 500 \text{ arcmin}^{-2}$  to showcase the potential differences in the detections. We clearly see in the zoom-ins II. and IV. that there are objects with for example spirals in the realistic simulations that lack adequate modelling of their morphologies, leading to potential additional deblending into multiple components also at bright magnitudes from the faint structures that are omitted when using Sérsic profiles. Moreover, stamps I. and III. show how detection differences might arise for faint objects, that is, when the surface brightness is not fully identical in the branches (the faint object in the upper-left corner is not above the detection threshold in the parametric branch). The lower panel zoom-ins depict an example where the de-blending produces more objects and a higher blending fraction in the realistic branch. This is of course dependent on the detection threshold, and de-blending settings in the SExtractor configuration file, which for our setup we chose as `DETECT_THRESH = 1.5`, `DETECT_MINAREA = 6`, `DEBLEND_NTHRESH = 32`, `DEBLEND_MINCONT = 0.005`, and `CLEAN_PARAM = 1.0`.



**Fig. E.1.** Example from the full image simulations in both branches used for the SExtractor detection comparison. The top panels show  $400 \times 400$  pixel cutouts from an image scene, with doubled number density  $\rho_n$ , with pixel intensities clipped between the 10<sup>th</sup> and 90<sup>th</sup> percentiles. Below are zoom-ins into interesting regions: Examples I. and III. illustrate a detection difference, where a realistic object lies just above the threshold, while its best-fit double-Sérsic counterpart is not detected. The zoom-ins II. and IV. show an example of a clearly more complex spiral galaxy and how this is represented in the parametric branch, highlighting the ability of the model to produce more realistic image simulations. The lower panels V. and VI. show another example cutout from the same full scene where a single object is de-blended into two components by SExtractor.