Galaxy Light Profile Convolutional Neural Networks (GaLNets). I. Fast and accurate structural parameters for billion-galaxy samples

Li, R.; Napolitano, N.R.; Roy, N.; Tortora, C.; La Barbera, F.; Sonnenfeld, A.; ... ; Liu, S.

The Astrophysical Journal, 929:152 (20pp), 2022 April 20 https://doi.org/10.3847/1538-4357/ac5ea0

Galaxy Light Profile Convolutional Neural Networks (GaLNets). I. Fast and Accurate Structural Parameters for Billion-galaxy Samples

R. Li[1,2,3], N. R. Napolitano[2,4,5], N. Roy[6], C. Tortora[5], F. La Barbera[5], A. Sonnenfeld[7], C. Qiu[2], and S. Liu[2]
[1] School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
[2] School of Physics and Astronomy, Sun Yat-sen University, Zhuhai Campus, 2 Daxue Road, Xiangzhou District, Zhuhai, People's Republic of China; napolitano@mail.sysu.edu.cn
[3] National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, People's Republic of China
[4] CSST Science Center for Guangdong-Hong Kong-Macau Great Bay Area, Zhuhai, 519082, People's Republic of China
[5] INAF—Osservatorio Astronomico di Capodimonte, Salita Moiariello 16, I-80131—Napoli, Italy
[6] Carmel College, Mala, Thrissur-680732, Kerala, India
[7] Leiden Observatory, Leiden University, P.O. Box 9513, NL-2300 RA Leiden, The Netherlands
Received 2021 November 9; revised 2022 March 13; accepted 2022 March 15; published 2022 April 25

Abstract

Next-generation large sky surveys will observe up to billions of galaxies for which basic structural parameters are needed to study their evolution. This is a challenging task that, for ground-based observations, is complicated by seeing-limited point-spread functions (PSFs). To perform a fast and accurate analysis of galaxy surface brightness, we have developed a family of supervised convolutional neural networks (CNNs) to derive Sérsic profile parameters of galaxies. This work presents the first two Galaxy Light profile CNNs (GaLNets) of this family. The first one is trained using galaxy images only (GaLNet-1), and the second is trained with both galaxy images and the local PSF (GaLNet-2). We have compared the results from GaLNets with structural parameters (total magnitude, effective radius, Sérsic index, etc.) derived from a set of galaxies from the Kilo-Degree Survey by 2DPHOT as a representative of the "standard" PSF-convolved Sérsic fitting tools. The comparison shows that GaLNet-2 can reach an accuracy as high as that of 2DPHOT, while GaLNet-1 performs worse because it misses the information from the local PSF. Both GaLNets are three orders of magnitude faster than standard methods in terms of computational speed. This first application of CNNs to ground-based galaxy surface photometry shows that they are promising tools to perform parametric analyses of very large galaxy samples, like the ones expected from the Vera Rubin/LSST surveys. However, GaLNets can be easily modified for space observations from Euclid and the China Space Station Telescope.

Unified Astronomy Thesaurus concepts: Galaxy structure (622); Galaxy formation (595); Neural networks (1933)

1. Introduction

Galaxy structural parameters, such as the total magnitude mag, the effective radius $R_{\rm eff}$, and the Sérsic n index, are basic measurements for many empirical scaling relations. They are used for the size–mass relation (Shen et al. 2003; Baldry et al. 2012; Lange et al. 2016; Roy et al. 2018), the μ_e–R_e relation (Kormendy 1977; Capaccioli et al. 1992), the fundamental plane (Dressler et al. 1987), the Faber–Jackson relation (Faber & Jackson 1976), etc.

The dependence of these relations on galaxy properties, such as the environment where galaxies reside and their evolution with redshift, is a fundamental test of the galaxy formation and evolution scenario, as they provide evidence of the growth in size and mass of these systems and the processes behind them. For instance, within the hierarchical model in the Λ-cold dark matter cosmology, most present-day massive, spheroidal galaxies are expected to have grown in size and mass via a two-phase assembly: the first phase, at redshift $1 < z < 2$, when galaxies turn to red compact systems (likely the progenitors of the cores of today's ellipticals; e.g., Wellons et al. 2016, 2015) and the second phase (e.g., Oser et al. 2012), characterized by mergers and gas inflows, leading to rejuvenated star formation and dramatic size evolution from $z \sim 1$ to today (Trujillo et al. 2007; Buitrago et al. 2008). On the other hand, less luminous/massive spheroids, i.e., those having magnitude fainter than the knee of the luminosity function, m_*, and a stellar mass smaller than $M_* \lesssim 10^{10.5} M_\odot$, are predicted to have a less dramatic size evolution (e.g., Furlong et al. 2017). Similarly, star-forming galaxies also show a milder evolution than spheroids (see, e.g., van der Wel et al. 2014; Roy et al. 2018).

Overall, hydrodynamical simulations seem to show that the difference in growth between passive spheroids and active disk-dominated systems is mainly due to the fraction of accreted mass, which is higher in the former, especially in the higher-mass bins (e.g., Oser et al. 2010; Furlong et al. 2017).

To fully test these predictions, it is crucial to measure structural parameters for large statistical samples of galaxies over a wider range of redshifts and stellar masses.

Next-generation large sky surveys like Vera Rubin/LSST (Ivezić 2019), the Euclid mission (Laureijs et al. 2011), and the Chinese Space Station (CSST; Zhan 2011), will eventually provide this unique opportunity. For instance, with the Vera Rubin/LSST full depth (5σ point source $r_{\rm AB} \sim 27.5$ coadded), we will be able to identify $m_* + 2$ galaxies in clusters and the field at $z \sim 1.5$ and reach a signal-to-noise ratio (S/N) high enough to study the galaxy structural parameters of m_* galaxies at $z \sim 1.2$ in the *ugrizy* optical bands (Robertson et al. 2017). Similarly, CSST and Euclid will allow accurate surface brightness measurements for $\sim 2 \times 10^8$ galaxies at $1 \leqslant z < 3$

in UV/optical and near-infrared, respectively (see, e.g., Laureijs et al. 2011).

Hence, in the coming years, we will analyze the surface photometry of billions of galaxies in many optical and near-infrared bands, which is a major technological and methodological challenge. If, on the one hand, these measurements need to be performed in a reasonable timescale, on the other hand, the galaxy complexity to account for in the modeling (e.g., multicomponent versus single-component systems) will make the analysis process computationally challenging.

So far, most of the analyses of galaxy structural parameters over large data sets have been mainly based on one-component galaxy models, e.g., using the Sérsic (1968) profile. This has been successfully used to measure the main structural parameters of galaxies in different data sets, e.g., Sloan Digital Sky Survey (SDSS; e.g., Shen et al. 2003; Hyde & Bernardi 2009), GAlaxy Mass Assembly (GAMA; e.g., Baldry et al. 2012; Lange et al. 2015), Kilo-Degree Survey (KiDS; e.g., Roy et al. 2018), Dark Energy Survey (DES; e.g., Tarsitano et al. 2018), Hyper Suprime-Cam Subaru Strategic Program (HSC; e.g., Kawinwanichakij et al. 2021), and the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; e.g., Shibuya et al. 2015)

Indeed, with only three free structural parameters (the total magnitude mag, the effective radius $R_{\rm eff}$, and the Sérsic index n), together with some other shape/position parameters (the axis ratio q, the position angle PA, and the galaxy center coordinates $x_{\rm cen}$, $y_{\rm cen}$), this model has sufficient generality to match most of the light distribution of galaxies of different morphologies (i.e., early- versus late-type galaxies). Also, it has been successfully used to reproduce the surface brightness distribution of galaxies from the inner subarcsecond scales to relatively larger radii, even for ground-based observations (see, e.g., Tortora et al. 2018). Obviously, to fully catch the physics of galaxy evolution, it is mandatory to move toward more complex models considering multicomponent profiles (see, e.g., Dimauro et al. 2018). However, besides the complication of higher degeneracy among the parameters, which is worsened by the low spatial resolution of ground-based observations, these models do add further computational complexity.

Currently, traditional galaxy-fitting codes, like GALFIT (Peng et al. 2002), Gim2d (Simard et al. 2002), and 2DPHOT (La Barbera et al. 2008), are either too slow or need too much manual intervention. Hence, they could satisfy the required speed demand when suitably parallelized and/or using next-generation multi-CPU machines. Although "hybrid codes," like GALAPAGOS (Barden et al. 2012) or PyMorph (Vikram et al. 2010), are trying to overcome these technical difficulties, the computational time per galaxy is still rather high and strongly dependent on the number of model parameters. Furthermore, for most of these codes, the accuracy and success rate are also strongly dependent on the initial conditions (see, e.g., Yoon et al. 2011). Hence, faster, automatic, and more accurate methods are highly needed for next-generation sky surveys.

Machine-learning (ML) tools possibly provide the most viable solution to this need. In the last years, they have been successfully applied to many fields of astronomy to solve typical classification or regression tasks with fast speed and high accuracy. For instance, they have been used to search for strong gravitational lenses either from images (Petrillo et al. 2017, 2019a, 2019b; Jacobs et al. 2019; Cañameras et al. 2020; Li et al. 2020, 2021) or from spectra (Li et al. 2019); for automatic extraction of spectral features (Wang et al. 2017); for unsupervised feature-learning for galaxy spectral energy distributions (SEDs; Frontera-Pons et al. 2017); and for photometric redshifts (Bilicki et al. 2018).

Among the ML tools above, convolutional neural networks (CNNs), in particular, allow us to optimally handle problems related to image processing and feature recognition that are suitable for galaxy morphology and surface brightness analysis. Indeed, galaxy classification has been one of the fields where CNNs have produced early promising results. Dieleman et al. (2015) developed a CNN model to classify the galaxies that have been labeled by human inspection in the Galaxy Zoo project. The CNN reached 99% accuracy while having a much faster speed than human classification. Similarly, Tarsitano et al. (2022) developed a CNN-based approach to classify the galaxies in DES and achieved 86% accuracy for early-type galaxies and 93% for late-type galaxies.

CNNs have also been successfully applied to a variety of galaxy photometry-related studies, such as nonparametric light profile extraction (e.g., Smith et al. 2021; Stone et al. 2021), source deblending (e.g., Boucaud et al. 2020), and galaxy stellar population analysis (e.g., Buck & Wolf 2021).

Tuccillo et al. (2018) for the first time applied CNNs to two-dimensional light profile galaxy fitting on Hubble Space Telescope (HST)/CANDELS data. Their code (DeepLeGATo) manages to fit galaxies with a single Sérsic profile, closely reproducing the results obtained with GALFIT but with much shorter computational time and in a fully automatic way. To achieve the best accuracy, though, DeepLeGATo required some domain adaptation,[8] because the initial simulated training set was not "realistic" enough. Furthermore, being developed for space observations, DeepLeGATo was not designed to account for the seeing-limited point-spread function (PSF).

The impact of the local PSF has been investigated previously in Umayahara et al. (2020), who showed that adding the information of PSF to the galaxy images in deep-learning tools would improve the prediction accuracy for the effective radius in ground-based observations. However, this early attempt did not test the full constraint of a Sérsic model, including all other model and geometrical parameters like the $n$index, the total magnitude, the axis ratio, and the PA. Hence, to our knowledge, there are currently no works that have produced science-ready ML tools to infer galaxy surface photometry, fully accounting for the PSF effect on ground-based observations. For this reason, we have started a project to develop ML tools to analyze the 2D surface brightness of galaxies using primarily ground-based observations. The main aim is to produce a family of fast and accurate GAlaxy Light profile CNNs (GaLNets, in short) to perform single- and multi-Sérsic-profile models of galaxies on different data sets. This will allow us to test the applicability of GaLNets to these data for future ground-based surveys (e.g., Vera Rubin/LSST), but it will eventually be extended to space observations (e.g., Euclid and CSST) with other dedicated GaLNets.

[8] Domain adaptation is a special technique for transferring learning from a given domain encoded in a training sample into another domain characterizing the predictive sample. This is a technique used to solve the discrepancies between the predictions and the true values that can come from a mismatch between the sample features used to train a CNN and the real properties of the predictive sample (see Csurka 2017). For instance, in the specific case of the galaxy structural parameters, the simulated galaxies can have no companion systems, while the real ones do (see Tuccillo et al. 2018).

In this first paper, we develop two GaLNets to fit single Sérsic (1968) profile parameters to galaxies from public ground-based data of the KiDS (de Jong et al. 2015, 2017; Kuijken et al. 2019) as a prototype of high-quality ground-based imaging data sets, but this can be easily generalized to other imaging data sets. To train the two CNNs, we simulated realistic KiDS mock galaxies by fully taking into account the observed seeing to produce PSF-convolved 2D galaxy models that we add to randomly selected "noise cutouts." These mock observations are similar to what we have implemented in the strong-lensing classifiers (see Li et al. 2020, 2021), where we have produced realistic color images of gravitational arcs and multiple images from lensed sources. The two GaLNets differ by the "features" that they use in the training phase: The first one (GaLNet-1) is fed with only galaxy images as input, while the second one (GaLNet-2) is fed with both galaxy images and the "local" PSFs.[9] The reason for introducing GaLNet-1, which is expected to perform generally worse than GaLNet-2 because it lacks PSF information, is to quantify the impact of the PSF on the accuracy of the results and possibly find room for improvement in the PSF model to adopt.

This paper is meant to provide the first analysis of the performance of this novel approach. For this reason, we use only *r*-band observations, as these are the best image-quality data in KiDS (see, e.g., Kuijken et al. 2019). We will also use a limited sample of ∼25,000 galaxies for which we have previously derived independent structural parameters (i.e., total magnitude, effective radii, Sérsic index, axis ratio, etc.; see Roy et al. 2018) using 2DPHOT, a PSF-convolved Sérsic fitting tool based on χ^2 minimization (La Barbera et al. 2008). In order to avoid systematics due to the selection and modeling of the local PSF, we will adopt the PSF models produced by 2DPHOT for the galaxy sample (see Roy et al. 2018 for details) as input for GaLNet-2.

In future papers, we will apply GaLNets to the full multiband KiDS data set to study the structural properties of galaxies as a function of galaxy mass and redshift. In future papers of this series, we will also expand the GaLNets model capabilities to fit multicomponent systems and test their application to future survey facilities like Vera Rubin/LSST, Euclid, and CSST.

This work is organized as follows. In Section 2, we describe how to build the training and testing sample and the CNN architectures. In Section 3, we test our CNNs on simulated and real data. In Sections 4 and 5, we give some discussions and summarize our main conclusions.

2. The CNN Method

CNNs are a special category of neural networks made of layers of artificial neurons called nodes. The nodes contain the convolution part where specific activation functions calculate the weighted sum of the inputs and return an activation map. Convolution is able to save the important information contained in the "features" (e.g., a pattern or a color in an image) by simultaneously reducing the size of the elements carrying such relevant information. This produces massive savings in computational time in high-resolution images, hence making CNNs particularly suitable for both classification and regression tasks related to image analysis.

[9] With "local" PSFs, we refer to the fact that the network is trained with PSF models that take into account the characteristic spatial variation of the PSF across the observed images.

CNNs are trained over specific samples containing the features to be used to make predictions on a series of "targets" (i.e., a pattern, a property, one or more parameters) over a sample of objects that represent the "predictive data." The crucial ingredients in this "training" phase are (1) the perfect correspondence between the sample used for the training (training set) and the sample for which we need to estimate the target parameters (predictive sample) and (2) the size of the training se*t*, which needs to encompass the full target parameter space. These are often not available, as the ground-truth values are not given for real objects one wants to analyze; even if these latter values are provided, they are limited to small statistics from other standard methods. A classic example for this latter case is strong gravitational lensing (see, e.g., Petrillo et al. 2017; Jacobs et al. 2019; Li et al. 2020), for which only a few hundred confirmed cases are available (e.g., SLACS, Bolton et al. 2008; BELLS, Brownstein et al. 2012; BELLSGALLERY, Shu et al. 2016).

In this paper, we tackle a classic regression task where the inputs are images of individual galaxies or pairs of images of galaxies and the corresponding PSFs, and the outputs are seven parameters of the best Sérsic profiles describing the 2D galaxy light distribution. This is defined as

$$I(R) = I_e \exp\{-b_n[\left(\frac{\sqrt{(x-x_0)^2 + (y-y_0)^2/q^2}}{R_{\rm eff}}\right)^{\frac{1}{n}} - 1]\}, \quad (1)$$

where the $R_{\rm eff}$ is the effective radius, I_e is the surface brightness at the effective radius, and (x_0, y_0) is the position of the galaxy center. The position angle PA gives the angle between the (x, y) plane and the sky coordinate frame, which is defined by assuming to be positive from north to east. Finally, q is the axis ratio, and n is the Sérsic index. This latter is known to be a proxy of the galaxy morphology as early-type galaxies generally have $n \gtrsim 2.5$ while late-type galaxies have $n \lesssim 2.5$ (e.g., Trujillo et al. 2007). In Equation (1), for b_n we use the expression provided by Ciotti & Bertin (1999):

$$b_n \approx \begin{cases} 2n - 1/3 + 4/(405n), & n \geqslant 0.36 \\ 0.01945 - 0.8902n + 10.95n^2, & n < 0.36. \end{cases} \quad (2)$$

The total (apparent) magnitude, mag, is defined by

$$\text{mag} = -2.5\log(F_{\rm tot}) + \text{zpt}, \quad (3)$$

where zpt is the zero point of the photometry. $F_{\rm tot}$ is the total flux of the galaxies, defined as

$$F_{\rm tot} = 2\pi R_{\rm eff}^2 I_e e^{b_n} n b_n^{-2n} \Gamma(2n) q, \quad (4)$$

in which Γ indicates the standard Γ function. In summary, the fully 2D galaxy model contains seven free parameters: x_0, y_0, mag, $R_{\rm eff}$, q, PA, and n.

The observed surface brightness profiles of galaxies result from the convolution of the intrinsic (real) galaxy fluxes and the PSFs. Equally, the modeled 2D profile is given by the convolution of the 2D Sérsic profile and the PSF model. This can be computed by the equation

$$M(\text{BG}, \{p_k\}) = \text{BG} + I(\{p_k\}) \circ S, \quad (5)$$

where I is the galaxy surface brightness distribution from Equation (1), $\{p_k\}$ are the model parameters, S is the PSF

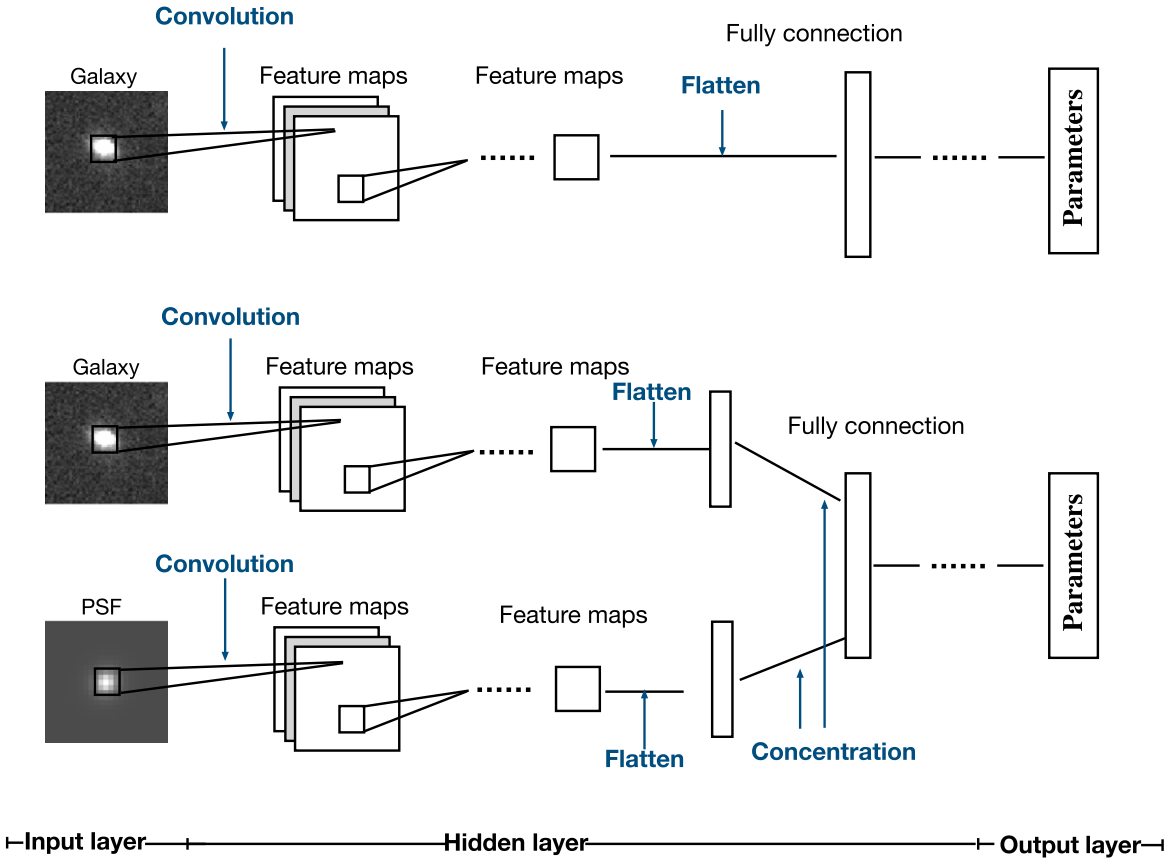

Figure 1. The GaLNet models used in this work. Top: structure of GaLNet-1, with only galaxy images as inputs. Bottom: structure of GaLNet-2, fed by both galaxy images and the corresponding "local" PSFs from 2DPHOT.

model, BG is the value of the local background, and the symbol "o" denotes convolution. In particular, for the PSF we adopt two Moffat profiles (see La Barbera et al. 2008).

2.1. The CNN Architecture

Here, we will describe the CNN method behind GaLNets, including the customization of the CNN architecture, the building of the training data, and the training of the CNN models.

CNN, which mimics the biological perception mechanism, is one of the representative deep-learning algorithms. Depending on whether labels are provided for the training data, this learning process can be supervised or unsupervised. Supervised learning can infer a function from the labeled training data. These data are composed of input objects (usually vectors) and desired output values (labeled targets).

The simple idea behind GaLNets is that one can input galaxy images into the CNNs and require them to output the Sérsic parameters, with the CNN learning how labeled profiles would look like in images with the same noise/background structure and seeing as real observed galaxies. However, especially for ground-based data, besides the intrinsic degeneracies among the Sérsic parameters in Equation (1) (see, e.g., Trujillo et al. 2001b), there are some further degeneracies among these latter and the PSF (see, e.g., Trujillo et al. 2001a). As a very intuitive example, one can think of the intrinsic galaxy axis ratio. The PSF is, by definition, rounder than a flattened galaxy, hence the effect of the atmosphere on the 2D light distribution (i.e., a 2D convolution) produces as a net effect a rounder galaxy. The higher the seeing in arcseconds (which is measured by the FWHM), the rounder the observed galaxy is. Without knowledge of the PSFs, the prediction from the CNN could be biased toward lower axis ratios (i.e., rounder shapes).

In this work, we want to explore (1) how a CNN performs if only galaxy images labeled with the corresponding true intrinsic parameters are used in the training phase, and (2) how a CNN improves if, together with the galaxy image, the corresponding 2D model of the local PSF is also inputted.

To check these two schemes, we build two CNN models, GaLNet-1 and GaLNet-2, respectively. These are schematized in Figure 1 where we use, in particular, a slightly modified VGGNet (Simonyan & Zisserman 2014). The overall structure is common to other CNNs and consists of three parts: the input layer, the hidden layer, and the output layer. The heart of the CNN is the hidden layer, which includes (1) the convolutional layer, (2) the pooling layer, and (3) the fully connected layer. The convolutional layer, containing multiple convolution kernels composed of weights and biases, is used to extract features from the input data. These "feature maps" will be passed to the pooling layer for information filtering to reduce their intrinsic size. The pooling layer contains a preset pooling function used to replace the feature map in a given region with a single point (flattening). The fully connected layer is located at the end of the hidden layer of networks to perform a nonlinear combination of the extracted features. In particular, it combines the low-level learned features into high-level features and passes them to the output layer.

For GaLNet-1, with only galaxy images as input, the hidden layers are made of four convolutional blocks, and in each

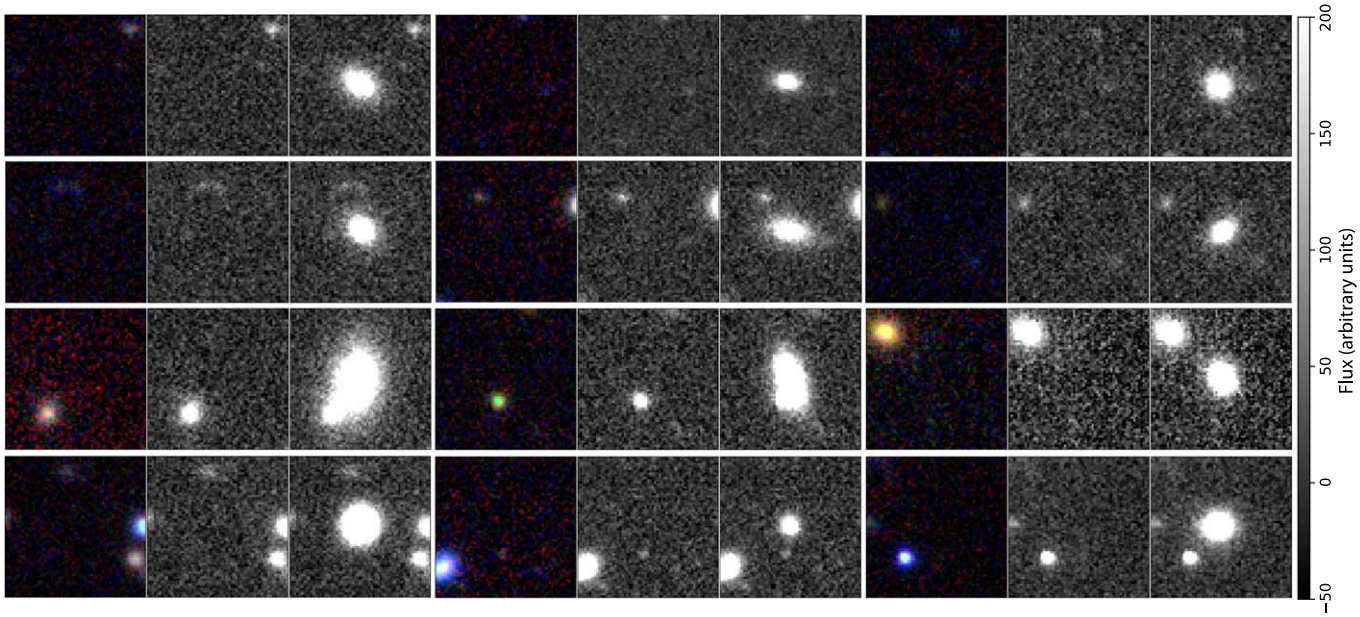

Figure 2. Examples of simulated galaxies for training GaLNets. For each three-panel image, we have the *gri* color combined (left panel), the *r*-band image (central panel) of the background cutouts, and the simulated galaxy added in the cutout center (right panel), according to its labeled Sérsic parameters. We can see the variety of real "contaminants" in the background cutouts, including bluer and redder stars and galaxies of different sizes and colors.

block, there are two weight layers, followed by an averaging 2×2 pooling layer. These convolutional blocks return feature maps, and the feature maps are then fed to three fully connected layers to combine the low-level features with higher-level ones. In this phase, the CNN makes the decision about the seven Sérsic parameters to output through the output layer.

For GaLNet-2, including also the 2D PSF model images as input (see Section 2.2 for details), we designed a two-path architecture: The main path is used to process the images, and the secondary path is used to process the PSFs. The main path is the same as that in GaLNet-1. The secondary path has only one convolutional block composed of two weight layers. After the flattening, the two paths are concentrated and then followed by three fully connected layers. Here the CNNs make the decision about the seven Sérsic parameters after having learned the PSF morphology and output the result through the output layer.

2.2. Training and Testing Data

The training set is built by adding two-dimensional, PSF-convolved, simulated Sérsic profiles to randomly selected *r*-band cutouts from public KiDS data, representing the galaxy "background." These random background images are not empty but can contain other sources to simulate realistic environment situations (see Figure 2 and the description below).

The mock galaxy images constituting the training data are obtained with the following procedure:

1. *Background images.* We randomly obtained 20,000 cutouts with a size of 61×61 pixels from the *r*-band observations in KiDS. We visually make sure that (1) these cutouts are not located in areas with saturated stars or stellar diffraction spikes or reflection halos, and (2) there are no bright galaxies or stars in the central region. These cutouts represent the "background sample" as introduced above.
2. *PSFs.* We collect 25,000 PSFs from previous 2DPHOT runs (see, e.g., Roy et al. 2018), where 2D local PSF models paired to each galaxy are obtained assuming two Moffat profiles (see La Barbera et al. 2008). We obtain 25×25 pixel images of the modeled PSF that make up the "PSF sample." The mean FWHM of the PSFs is $0\farcs70$, and the standard deviation is $0\farcs11$, similar to the one of the full DR4 in Kuijken et al. (2019);
3. *Mock galaxies.* We simulate the Sérsic profiles with a range for the seven parameters as shown in Table 1. The distribution of $x_{\rm cen}$, $y_{\rm cen}$, and $\log R_{\rm eff}$ are Gaussian, mag is exponential, n is F, while q and PA are uniform. These distributions are chosen to mimic the observed parameter distributions in galaxies (see, e.g., Roy et al. 2018).
4. *PSF convolution and final stamps.* We convolve each Sérsic profile with a randomly selected two-dimensional

Table 1
Parameters for Simulating the Training Sample

Parameter	Range	Units	Distribution
$x_{\rm cen}$	−0.4–0.4	arcseconds	normal ($\mu = 0.0$, $\sigma = 0.15$)
$y_{\rm cen}$	−0.4–0.4	arcseconds	normal ($\mu = 0.0$, $\sigma = 0.15$)
mag	17–22	⋯	exponential ($s = 1$)
$R_{\rm eff}$	0.2–4	arcseconds	normal ($\log R_{\rm eff}$) ($\mu = -0.1$, $\sigma = 0.4$)
q	0.2–1.0	⋯	uniform
pa	0.0–180.0	degrees	uniform
n	0.1–8.0	⋯	F ($n = 30$, $d = 5$)

Notes. Range and distribution of parameter values used to simulate the galaxies. μ and σ are the mean value and standard deviation of a normal distribution. s is the scale of the exponential distribution. n is the degrees of freedom in the numerator and d is the degrees of freedom in the denominator.

PSF from the PSF sample, and then a random Poisson noise is added. The resulting 2D profile is finally summed to a random "background" cutout (from step 1) to obtain fairly realistic galaxy images. If the S/N, as computed over a central circular area of radius 1.″5, is larger than 50, this galaxy is kept as training data; otherwise, it is discarded.

Adopting an $S/N > 50$ for the training sample is made for uniformity with typical choices made for real galaxies to ensure reasonable accuracy on the fitted parameters (see, e.g., La Barbera & de Carvalho 2009; Roy et al. 2018). For this first analysis, we do not plan to test the impact of the S/N on GaLNets predictions, which we will leave for future analyses (see also Section 4.2).

We also remark that the use of realistic distributions of parameters as a "prior" for the training sample, in step 3, is based on a series of tests made on different distributions (see the Appendix). To summarize, we find that the unrealistic distributions of the n index will produce some systematic features at $n > 4$, while for other parameters, the use of unrealistic priors will affect the final predictions very little. This implies that although we paid attention to reproducing realistic enough ground-truth parameter distributions for the n index, we cannot exclude that there are still some systematics with respect to the intrinsic parameters. For this reason, we will explore the use of a Bayesian neural network for future developments (see, e.g., Gal & Ghahramani 2015; Wagner-Carena et al. 2021). This will be a natural choice to include error estimates in the individual predicted parameters, which are not available with the current GaLNets. For the time being, as shown in the Appendix, we stress that for variations as large as 30% in the parameter distributions, GaLNets do not suffer significant biases, except for large values of the Sérsic index ($n > 4$).

Finally, following steps 1–4 above, we simulate 220,000 mock galaxies. We split this sample into the training data, to which we assign 200,000 mock galaxies, and the test data, made of the residual 20,000 simulated galaxies. In Figure 2 we show some mock galaxies and their original background images. There are clearly isolated cases with empty backgrounds and cases with close systems of different sizes, shapes, and colors. Here, we remark that according to the procedures above, the PSFs of the simulated galaxies do not match the ones of the background cutouts but are randomly selected from the "PSF sample." However, we do not expect this to affect the final results, as the CNNs focus only on the model of the central galaxy. However, this is a refinement that we will implement in future analyses when GaLNets will implement a self-made PSF model.

On the other hand, a parameter that can impact the final CNN predictions is the size of the cutouts. We have taken 61 pixels ($\sim 12''$) per side because this provides an area large enough to sample most of the light profile of galaxies with $R_{\rm eff} \sim 1''$ (corresponding to $\sim$5 pixels). This latter represents the majority of the galaxies observed in typical ground-based observations, like KiDS (see, e.g., Figure 9 of Roy et al. 2018). However, for larger galaxies, the fraction of the total light enclosed in the cutout can be quite different. For instance, by integrating Equation (1) (without seeing convolution) over the sky plane and considering the galaxy centered in the cutout with $12''$ on a side, the fraction (f) of total projected luminosity within a $12''$ diameter is

1. $f = 99.9\%$ for $R_{\rm eff} = 1''$ and $n = 1$,
2. $f = 91\%$ for $R_{\rm eff} = 1''$ and $n = 4$,
3. $f = 86\%$ for $R_{\rm eff} = 1''$ and $n = 6$,
4. $f = 61\%$ for $R_{\rm eff} = 4''$ and $n = 4$.

Hence, only for galaxies with small $R_{\rm eff}$ ($<1''$) and small n (<6), most of the light is collected into $12''$ wide cutouts. For larger n, even for $R_{\rm eff} \sim 1''$, there is a significant fraction of light emitted outside the cutouts. Finally, for larger $R_{\rm eff}$, only a small fraction of the whole galaxies is included in the cutout. Note that using a relatively small cutout might affect the parameter estimates for standard 2D fitting techniques (e.g., GALFIT Peng et al. 2002; 2DPHOT La Barbera et al. 2008) because both the outer parts of the galaxy and the background are poorly sampled. However, investigating this issue is outside the scope of the present paper.

For this reason, we additionally tested larger cutout sizes, namely 101 and 151 pixels a side, corresponding to $\sim 20''$ and $30''$, respectively, to check how the cutout sizes would affect their results. We have found that GaLNets generally give poorer results for small galaxies, while they do not show significant improvements for larger galaxies (i.e., $R_{\rm eff} > 3''$), despite the background around the galaxies being better sampled. A reason for this result is that the CNN learns how to derive central parameters ($R_{\rm eff}$, n, $I_{\rm e}$; see Equation (1)), which determine the gradient of the light profiles in the central regions. Eventually, the CNNs learn how to infer these parameters from the information they extract from the central steeper light gradient rather than the shallower slopes of the outermost radii. Hence, for the main analysis of this paper, we keep the 61 pixel cutouts as the best compromise for most of the systems we will analyze.

2.3. Training the Network

The input of the CNNs is the 61×61 pixel galaxy images for both GaLNet-1 and GaLNet-2 and the 25×25 pixel PSF images for GaLNet-2 only. The outputs are the seven parameters describing the light distribution of a Sérsic galaxy: $x_{\rm cen}$, $y_{\rm cen}$, mag, $R_{\rm eff}$, n, q, and PA). We train the CNNs by minimizing the "Huber" loss function (Friedman 1999) with an "Adam" optimizer (Kingma & Ba 2014). The "Huber" loss is defined as

$$L_\delta(a) = \begin{cases} \frac{1}{2}(a)^2, & |a| \leqslant \delta \\ \delta \cdot (|a| - \frac{1}{2}\delta), & \text{otherwise} \end{cases}, \quad (6)$$

in which $a = y_{\rm true} - y_{\rm pred}$, where $y_{\rm true}$ is the real value for the simulations and $y_{\rm pred}$ is the value predicted by the CNNs. δ is a parameter that can be preset. Given δ (δ was fixed to 0.001 in this work), the loss will be a square error when the prediction deviation $|a|$ is smaller than δ, otherwise, the loss reduces to a linear function. We preferred the "Huber" loss over standard loss functions like the mean square error (MSE) and the mean absolute error (MAE) because we have found the former to provide better accuracy and robust convergence. The reason to discard the MSE is that this latter gives higher weights to outliers but at the expense of the prediction accuracy of other normal data points. Some of our targets are prone to outliers; see, e.g., the PA that can become rather random when q is close to 1 and degenerate between 0° and 180° when the ground truth

is close to either of these two values. Hence, we want to avoid introducing artificial biases. The MAE is less sensitive to outliers, but it tends to give convergence problems as it does not properly weight the gradient in the loss function with the errors (for instance, it does not allow large gradients to converge in the presence of small errors), and the updated gradient in the training process is always the same. Therefore, even for a small loss value, the gradient is large, affecting the convergence speed of the training process, and sometimes the CNN cannot even find the best-fitting value.

In terms of computational time, using an NVIDIA RTX 2070 graphics processing unit (GPU), the training times for GaLNet-1 and GaLNet-2 have no significant differences, as both need about 2 hr to be trained with the 200,000 galaxy sample. As we checked that the training time scales with the sample size (e.g., a 50,000 sample takes 0.5 hr), then this means that we can train the CNN over a 2 million sample in less than one day with commercial GPUs.

3. Testing the Performances

3.1. Testing on Simulated Data

After training the two GaLNets, we use the test sample (see Section 2.2) to check their performance. In Figure 3, we compare the ground-truth values (x-axis) of each parameter used to simulate the galaxies and the predicted values (y-axis) from GaLNet-1 (left panel) and GaLNet-2 (right panel). To assess how strong the linear relationship is between the two variables in these plots, we adopt three diagnostics: (1) the R squared (R^2), (2) the fraction of outliers, and (3) the normalized median absolute deviation (NMAD). The R^2 is defined as

$$R^2 = 1 - \frac{\sum_i (p_i - t_i)^2}{\sum_i (t_i - \bar{t})^2}, \quad (7)$$

where t_i are the ground-truth values, $_{pi}$ are GaLNets' predicted values, and $\bar{t}$ is the mean value of t_i. According to this definition, R^2 is 0 for no correlation and 1 for the perfect correlation.

The fraction of outliers is defined as the fraction of discrepant estimates larger than 15%, according to the following formula:

$$|\Delta p| = \frac{|p_i - t_i|}{1 + t_i} > 0.15. \quad (8)$$

This definition is usually adopted for outliers in photometric redshift determination (see details in Amaro et al. 2021). Finally, NMAD is defined as

$$\mathrm{NMAD} = 1.4826 \times \mathrm{median}(|\Delta p - \mathrm{median}\ (\Delta p)|). \quad (9)$$

The results obtained for all these parameters are listed in Table 2. From this table and from Figure 3, we see that the predictions from GaLNet-1 and GaLNet-2 are generally good as the R^2 is close to 1. This is particularly true for $x_{\rm cen}$, $y_{\rm cen}$, mag, $R_{\rm eff}$, and q, with GaLNet-2 showing R^2 systematically larger than GaLNet-1 (see below). The R^2 for the position angle PA and Sérsic index n is slightly worse. For PA, this comes from a large scatter around PA $= 0°$ and PA $= 180°$, where the CNNs tend to concentrate the rounder systems and can be easier to swap the 0° with the 180° solutions. For the Sérsic index n, the poorer correlation is caused by the scatter at larger n values. This mainly comes from two factors. First, the central slope of a Sérsic profile with a larger Sérsic index is very steep, which means the light decreases quickly within the central 1–2 pixels, so it is hard for the CNNs to fully "measure" such a steep variation, given the combination of the KiDS pixel size (0.″2) and the PSF (∼0.″6–0.″8 in the KiDS $r$band). Second, the outer part of such a large Sérsic-index profile is flat and is often hidden within the background noise, hence it can be hard for the CNNs to detect a significant light gradient that can help us guess the n values, as learned from the training sample. This effect is even more severe if the CNNs have no information about the PSF. This latter increases the degeneracies both at the low scales (where the seeing suppresses the peaks) and at large scales, where the seeing tends to smooth the profiles and make a shallower slope (see, e.g., Tortora et al. 2018). Encouragingly, the effective radius is a well-constrained parameter from both GaLNets, despite the larger uncertainties on the n index. This can be quantified by the fraction of outliers and the NMAD as a proxy of the scatter, as reported in Table 1. Here we see that the NMAD of $R_{\rm eff}$ is 0.07 for GaLNet-1 and 0.03 for GaLNet-2, comparable to the ones of the well-constrained mag and q. However, the fraction of outliers for $R_{\rm eff}$ (0.028 for GaLNet-1 and 0.017 for GaLNet-2) is larger than the ones of mag (0.0 for both CNNs) and q (0.008 and 0.001, respectively), although these are rather acceptable, especially for GaLNet-2. Finally, the n index shows fractions of outliers of the order of 0.26 and 0.13, which prompt us to find strategies for improvement. This would also help improve the result on $R_{\rm eff}$, discussed above. We remark, though, that the Sérsic index is traditionally a hard parameter to constrain from the ground, regardless of the tools adopted (see, e.g., Trujillo et al. 2001a), hence we consider the absence of systematics seen in Figure 2 as an encouraging result of these first GaLNets.

Going into more details about the performance of the two GaLNets, the fact that GaLNet-2 always performs better than GaLNet-1 shows the importance of the information that the CNN can learn from the "local" PSF, especially for the galaxy shapes. Indeed, the improvement is more evident for the axis ratio q and the Sérsic index n. For the former, as introduced in Section 2.1, GaLNet-2 can correct the prediction depending on how round the 2D PSF is from the input local PSF image, according to what was learned in the training phase. For the latter, as discussed above, GaLNet-2 can account for the effect of the smoothing of the profile, especially in the central pixels. We stress here that for both of these parameters, GaLNet-1 is also trained to account for the PSF, as this is incorporated in the convolution of the model with a random image from the PSF sample (see Section 2.2). However, when making predictions, GaLNet-1 can only guess on the basis of a "self-learned" PSF,[10] as it does not have information on the local PSF (PSF-mismatch, hereafter). This produces a larger scatter on the output parameters because of the well-known degeneracies among the Sérsic parameters (e.g., $Re - n$ and the mag $- n$). On the other hand, GaLNet-2, by avoiding the PSF-mismatch using the local PSF, is able to contain the scatter and overall parameter uncertainties (see Section 3.2).

We can finally visualize the differences between the two CNNs by estimating the residual images obtained by subtracting the 2D Sérsic model from the original galaxy image. This is

[10] The information of the PSF is encrypted in the mapping between the ground-truth labels and the training images, which are by definition convolved with the local PSF. Hence, even without knowing the local PSF, GaLNet-1 learns the "average" effect of the PSF over the training sample.

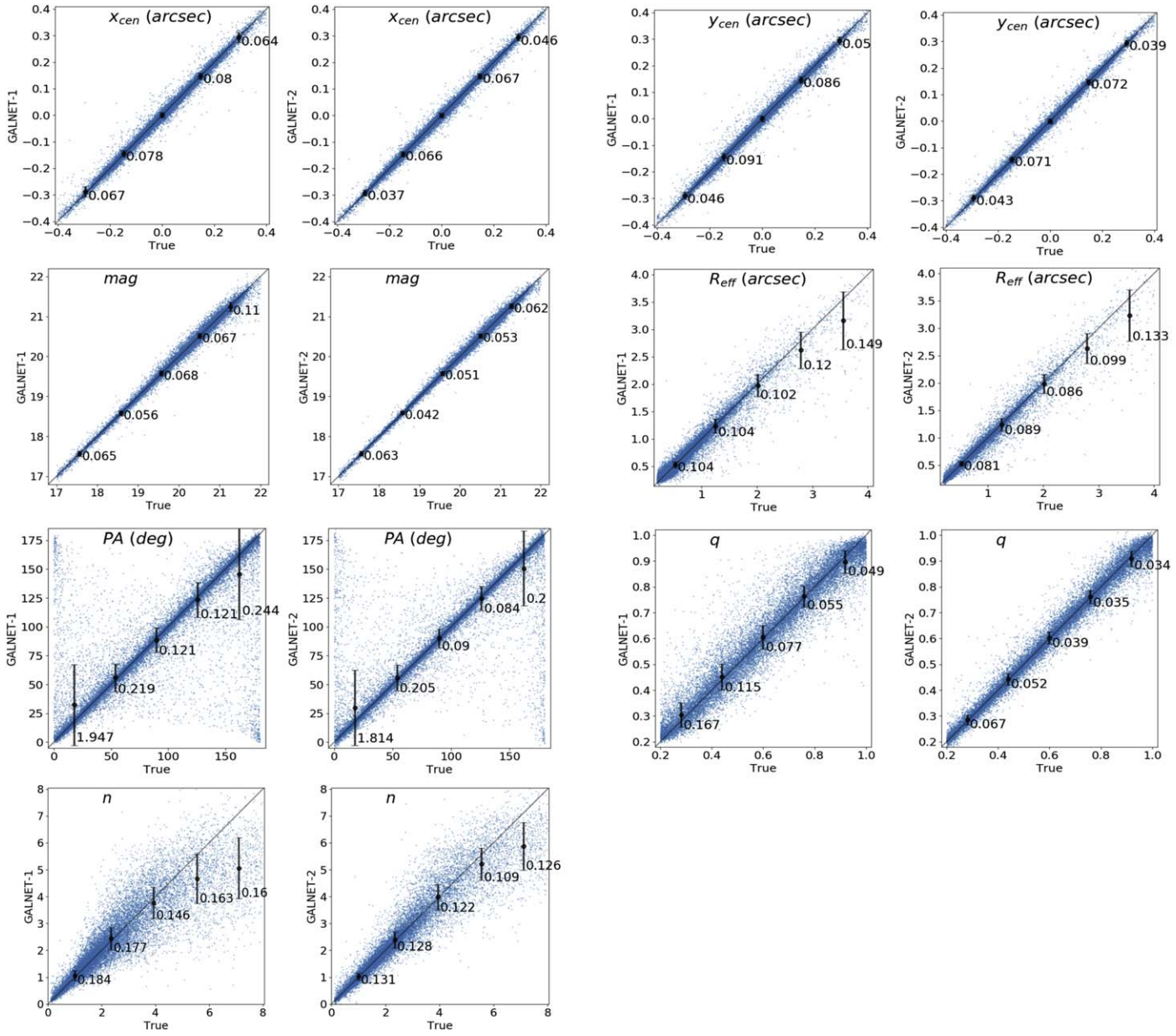

Figure 3. Comparison between the true value and the predicted value from the two GaLNets on simulated data. In each panel, the horizontal axes are the true values and the vertical axes are the predictions from GaLNets. The error bars are the absolute mean errors in each bin, while labels report the absolute errors for mag and relative errors for others.

done as follows. First, we derive a 2D galaxy model with the estimated parameters issued by the CNN then convolve this model with the local PSF (i.e., the one we use as input for GaLNet-2). This PSF-convolved model is subtracted by the galaxy image input in GaLNets: The better the models, the more the residual images look like a (zero) background image.

In Figure 4, we show some examples of the original simulated cutouts and their residual images from the two CNNs. Most of the time, GaLNet-1 and GaLNet-2 provide acceptable residuals (the first and second rows), even in the presence of companions and partially overlapping systems (second rows). However, in some cases, some bad residuals are found from only GaLNet-1 (third row) or both GaLNet-1 and GaLNet-2 (fourth row). For these cases, most of the bad fits are concentrated in the central pixels while the residuals become better outward. We note, though, that the failing cases are generally the ones with bright companion objects that can either affect the closer side of the main galaxy or cannot even be easily deblended. These cases are quite difficult to resolve also with standard methods (see also below).

Finally, we can clearly see that even in the poorer fits, those residuals from GaLNet-2 are generally better than those from GaLNet-1, which is visual proof of the crucial impact of the local PSF on the CNN performance, especially in finding the best fit of the galaxy cores.

3.2. Assessing the Statistical Errors

The simulated galaxies, with the "ground-truth" values of the parameters, allow us to assess the statistical errors (e_s, hereafter). In this paper, these will be quantified as the standard deviation of the scatter between the true values and the

Table 2
Statistical Properties of Simulated Data

CNN Model	x_{cen}	y_{cen}	mag	R_{eff}	PA	q	n
			R^2				
GaLNet-1	0.9885	0.9882	0.9898	0.9206	0.6716	0.9124	0.7280
GaLNet-2	0.9924	0.9915	0.9945	0.9624	0.7692.	0.9781	0.8905
			Fraction of Outliers				
GaLNet-1	0.001	0.001	0.000	0.028	0.210	0.008	0.260
GaLNet-2	0.000	0.001	0.000	0.017	0.165	0.001	0.128
			NMAD				
GaLNet-1	0.0100	0.0099	0.0028	0.0420	0.04840	0.02678	0.1094
GaLNet-2	0.0090	0.0090	0.0022	0.01990	0.0344	0.0139	0.0655

Notes. Statistical properties of the predictions on the simulated testing data. From top to bottom we show the R^2; the fraction of outliers and the NMAD for the central position x_{cen}, y_{cen}; magnitudes mag; effective radius R_{eff}; position angle PA; axis ratio q; and the Sérsic index n. The parameters of GaLNet-2 are always better than those of GaLNet-1.

predictions:

$$e_s^2 = \frac{1}{N-1}\sum_i^N (p_i - t_i)^2, \quad (10)$$

where the $_{pi}$ and the t_i are the predicted and true values in a given interval of $i = 1...N$ galaxies.

Being computed on simulated galaxies, these errors are referred to idealized objects while, as repeatedly specified before, in reality, we will not deal with single Sérsic profiles. However, these errors are the closest estimate that we can have of the contribution of the one-Sérsic "fitting" procedure to the overall error budget. From this point of view, the statistical errors we derive in this section represent a lower limit of real errors.

For each parameter, in Figure 3 we plot the e_s as error bars labeled with the relative error with respect to the mean of the true values in five bins. These latter are defined as $\Delta e = e_s/\overline{p}$, where $\overline{p}$ is the mean of the predicted values in a given bin—except for Δmag, which, being a logarithmic quantity, is defined as the difference between the predicted and the true magnitude, e_s. Generally speaking, the Δmag and the Δe of x_{cen}, y_{cen}, R_{eff}, and q for both GaLNets are smaller than $\sim$0.1, meaning that the two GaLNets are rather accurate to estimate these parameters. However, for PA and n, most of the Δe are between $\sim$0.1 and 0.2, slightly worse than those of other parameters. We also clearly see that in each panel that the Δe of GaLNet-2 is smaller than the corresponding value of GaLNet-1. The differences between the two GaLNets are even larger for PA, q, and n, with an average value of $\sim$0.03–0.05, while for others the differences are generally $<$0.2.

Looking into the variation of the relative errors as a function of the different parameters, we can see that the Δe of magnitudes from both GaLNet-1 and GaLNet-2 are generally smaller than 0.06 and stay almost constant in all the range of magnitudes that we have adopted, i.e., from mag = 17 to 22. This means that, above a given S/N ($\geqslant$50), both GaLNets achieve very high accuracy in estimating the total magnitudes, regardless of how luminous they are. The Δe of R_{eff} is smaller at $R_{eff} < 2''$ ($\sim$0.104 for GaLNet-1 and $\sim$0.085 for GaLNet-2) and become larger at $R_{eff} > 2''$ ($>$0.12 for GaLNet-1 and $>$0.1 for GaLNet-2). This is because most of the galaxies with larger R_{eff} also have large n. These galaxies are more difficult to model either for standard methods (see discussion in La Barbera et al. 2010) and also for GaLNets, as the innermost gradient of the light profiles becomes very steep and difficult to constrain. Some further contributions to the scatter can come from (1) the lack of galaxies with large R_{eff} in the training sample, produced by the log-normal distribution (see Tables 1) and (2) the cutout size (i.e., 12″ on a side), which, in principle, can sample only the central regions of galaxies with large effective radii. Because the effect of the cutout size has been shown to not improve GaLNet's results (see Section 2.2), the effect of the correlated variation of the Sérsic index n and R_{eff} seems to be the most reasonable explanation for the observed errors.

Going to PA, the relative scatters Δe at PA $< 20°$ and PA $> 160°$ are larger than in the other three bins in each panel, because, by symmetry, galaxies with PA $\sim 180°$ are equivalent to the ones with PA $\sim 0°$. This means that the Δe of the bins closer to these values does not represent the real PA statistical errors. In the three other bins, though, the Δe of the PA for GaLNet-1 varies from 0.12 to 0.22, and for GaLNet-2 from 0.084 to 0.20. Considering that a large portion of galaxies has $q > 0.8$, i.e., they are close to round and it is really hard for them to measure the exact PA value, 10%–20% scatter is a satisfactory uncertainty. For q, we find the Δe for both GaLNets becomes smaller for larger q values. However, this does not indicate that GaLNet performs worse at large q. This is because, as the absolute scatter is almost constant for all q, the Δe becomes larger at small q, by definition. It is worth noting the beneficial effect of the PSFs in GaLNet-2, which can improve its performance significantly with respect to GaLNet-1 at small q. Finally, for Sérsic index n, the scatters are generally larger than those of other parameters, $\sim$0.146–0.184 for GaLNet-1 and $\sim$0.109–0.131 for GaLNet-2. In this case, introducing the PSF produces the largest improvement compared to the other parameters.

3.3. Testing on Real Data

The test on mock galaxies is important to check the presence of systematics and to assess the statistical errors. Indeed, because the ground truth is errorless, the relative errors discussed in Section 3.2 just contain the internal errors of the measurement method.

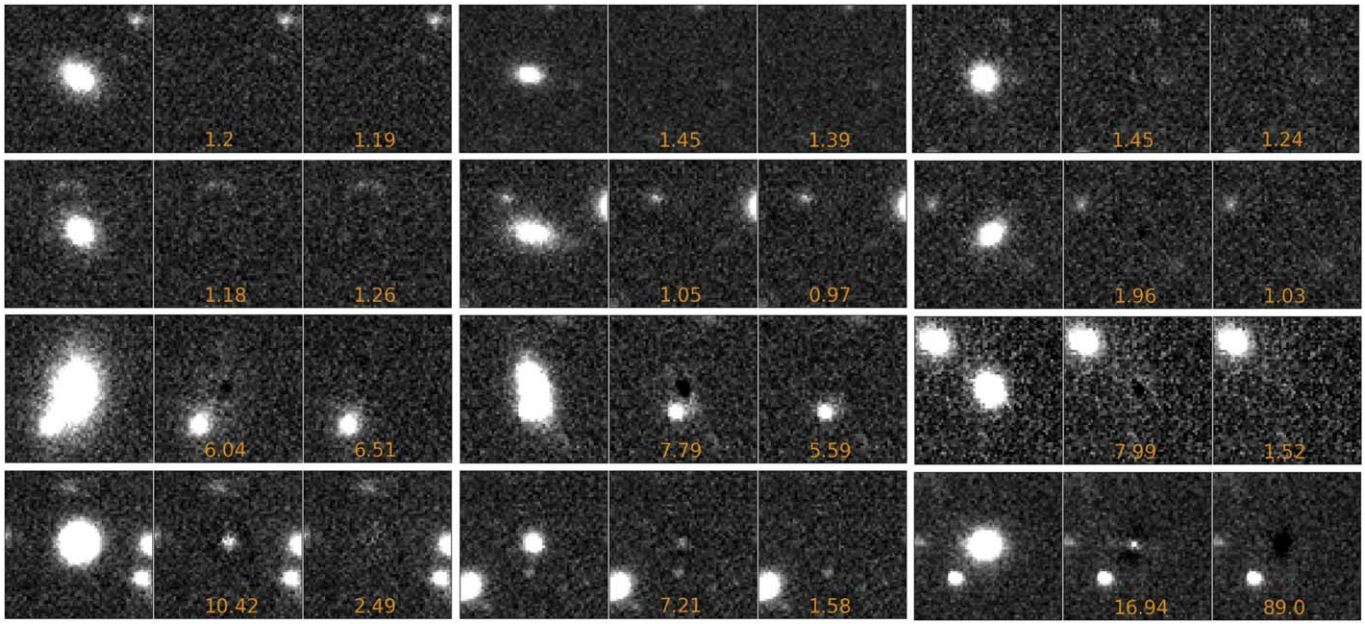

Figure 4. Residual maps, obtained from simulated "Galaxy" images (left panels), after subtracting the reconstructed Séersic models using GaLNets-1 (middle) and GaLNets-2 (right panels). At the bottom of each residual image, we report the a posteriori reduced $\tilde{\chi}^2$ (see Section 3.3 for a definition).

As we finally want to apply the trained GaLNets to real galaxies, we need to check the presence of systematics in more-realistic situations, including the impact of unaccounted substructures on real systems. Here we foresee two main categories of "reality mismatch." First, a "model mismatch": As discussed earlier, for this first GaLNets, we adopt the simplistic choice of a one-Sérsic profile. This is possibly accurate enough for a majority of galaxies, but not for all of them. Hence, we need to bear in mind that, despite how good a method is to fit a simplistic model, the best fit cannot always be optimal, as the real galaxy can have a more complex structure (this is true for both ML and standard tools; see, e.g., Peng et al. 2002, 2011). Second, a "substructure mismatch": This comes from the fact that even real galaxies with a dominant one-Sérsic light distribution can have uneven components (spiral arms, rings, bars, tilted isophotes, etc.). Of course, there might be systems where both mismatch types are present. These systems would be rather difficult to fit with a parametric model by general-purpose tools, so they are not a matter of concern in this context. As mentioned in Section 1, in this first step of GaLNet development, we will mainly focus on the substructure mismatch and ignore the model mismatch, which will be considered in the next steps of the project. In particular, we will consider either multicomponent parametric models or more general nonparametric models (e.g., Lauer 1985; Tody 1986; Ciambur 2015; Bradley et al. 2020; Stone et al. 2021).

For the test on real galaxies, we use 25,000 high-S/N galaxies randomly selected from KiDS for which the Sérsic parameters with the 2DPHOT package (La Barbera et al. 2008) are available (see also Roy et al. 2018). 2DPHOT is an automated software environment used for source detection and deep wide-field images analysis. In particular, 2DPHOT can derive structural parameters of galaxies by fitting the images with two-dimensional PSF-convolved Sérsic models. For each galaxy, 2DPHOT can automatically select two or three nearby sure stars and produce an average two-dimensional PSF by modeling them with two Moffat profiles (see detail in La Barbera et al. 2008). This PSF will be convolved by the modeled Sérsic profile to produce a 2D distribution of the galaxy light. The best-fitting parameters of the Sérsic profile then are obtained by minimizing a given χ^2 function. In the fitting process, the mask images, created by 2DPHOT itself, during the source detection, are also considered.

The real test galaxies are collected according to the following requirements: (i) $S/N > 50$, (ii) $17 < \mathrm{mag} < 22$ (in the r band), (iii) $0\farcs2 < R_{\rm eff} < 4''$, (iv) $0.1 < n < 8$, and (v) $q > 0.2$, which are generally within the parameter intervals of the training data. In order to avoid systematics due to the selection and modeling of the "local" PSF, as input for GaLNet-2 we will adopt the PSF models produced by 2DPHOT for the test galaxy sample (see Roy et al. 2018 for details). This has the advantage of strictly evaluating the relative performance of the two tools, but it has the disadvantage that GaLNet-2 will reproduce the same systematics of 2DPHOT in case of poor PSF model occurrences. However, we expect to produce independent PSF models for extensive future analyses on real data.

Differently from the simulated sample, here we cannot know the true parameter values, but we take the estimates from 2DPHOT as the "ground truth," assuming that these are unbiased estimates of the intrinsic galaxy parameters. Note that this is an idealization. In fact, although 2DPHOT was widely tested with ground-based and HST data, as well as simulated galaxies (La Barbera et al. 2008; Roy et al. 2018), some galaxies, e.g., those not well represented by a single Sérsic profile and/or in case of poor PSF modeling, might still be affected by unaccounted-for systematics.

However, being based on completely different approaches, it is unlikely that both methods suffer the same systematics. Hence, from their comparison, we can learn more about their relative accuracy, in particular whether GaLNets provide parameter estimates consistent with standard fitting approaches.

The comparison between the parameters mag, $R_{\rm eff}$, q, PA, and n, obtained by 2DPHOT and the two GaLNets, is shown in Figure 5. For these one-to-one relations, the R^2, outliers, and

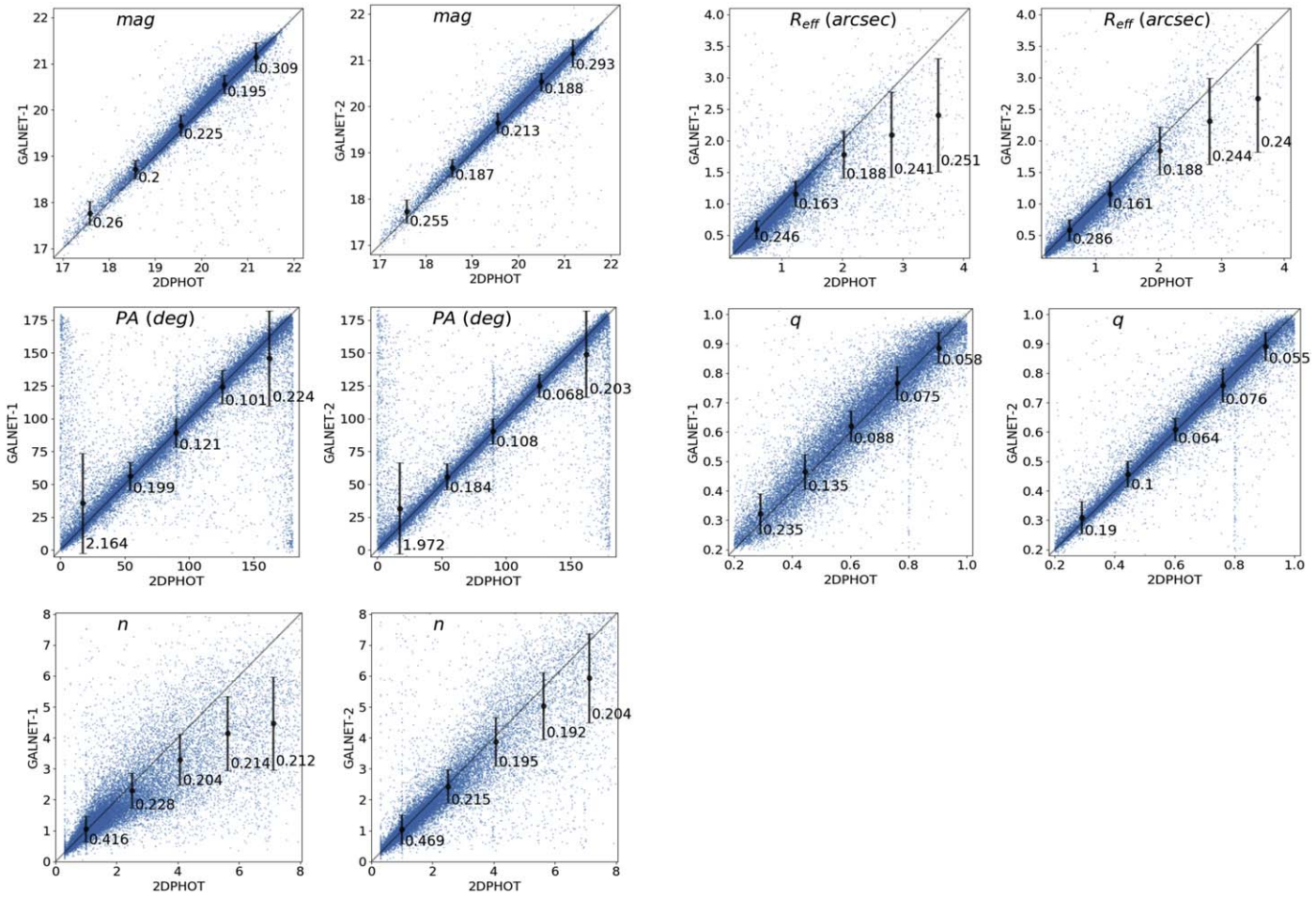

Figure 5. Comparison between the output of GaLNets and 2DPHOT on real galaxies from KiDS. In each panel, the horizontal axes are the values obtained by 2DPHOT and the vertical axes are the predictions from GaLNets. Error bars are the absolute mean errors in each bin, while labels report the absolute errors for mag and relative errors for others.

NMAD are reported in Table 3. For all five parameters, we generally see a fair agreement between the estimates of the two tools, with all data being concentrated around the one-to-one line and the R^2 parameter larger than 0.7. Looking at the comparison of the individual GaLNets, we see that the results agree for both the test sample and real galaxies.

GaLNet-2 shows a tighter consistency with the "ground-truth" estimates than GaLNet-1, especially for q and n. However, the magnitudes and effective radii of galaxies are still reasonably well constrained from GaLNet-1, showing that the PSF is important for these parameters mainly to reduce the uncertainties, but that the mean estimates over large numbers are fairly unbiased.

Compared with simulated data, the R^2 values are poorer overall than those measured in Section 3.1, showing a degradation of the performance with respect to the test sample. The larger scatter seen in Figure 5 is not entirely due to the larger uncertainties of GaLNets, but rather due to the combination of the statistical errors from the two different measurement procedures. Indeed, different from the test sample, where the ground-truth values were errorless, here the 2DPHOT estimates carry their own intrinsic error. If the two methods had similar Gaussian statistics and the uncertainties were uncorrelated, we might expect the scatter in Figure 5 to be $\sim\sqrt{2}$ times larger than the one in Figure 3. For instance, the mean scatters reported in Figure 5 for GaLNet-2 is a factor of $\sim$2 larger than the one in Figure 3 for $R_{\rm eff}$, and a factor of $\sim$1.6 larger for the n index. This confirms that GaLNets do not perform significantly worse on real galaxies, considering that some extra uncertainties on both GaLNets and 2DPHOT come

Table 3
Statistical Properties of Real Data

CNN Model	mag	$R_{\rm eff}$	q	PA	n
		R^2			
GaLNet-1	0.9037	0.6666	0.6528	0.8481	0.4274
GaLNet-2	0.9166	0.7035	0.7411	0.9160	0.7488
		Fraction of Outliers			
GaLNet-1	0.001	0.073	0.239	0.0152	0.329
GaLNet-2	0.001	0.065	0.180	0.0113	0.191
		NMAD			
GaLNet-1	0.0040	0.0497	0.0668	0.0327	0.1242
GaLNet-2	0.0034	0.0259	0.0437	0.0140	0.0786

Notes. Statistical properties of the prediction on real data. From top to bottom we show the R^2, the fraction of outliers, and the NMAD for the magnitudes mag, effective radius $R_{\rm eff}$, axis ratio q, position angle PA, and the Sérsic index n.

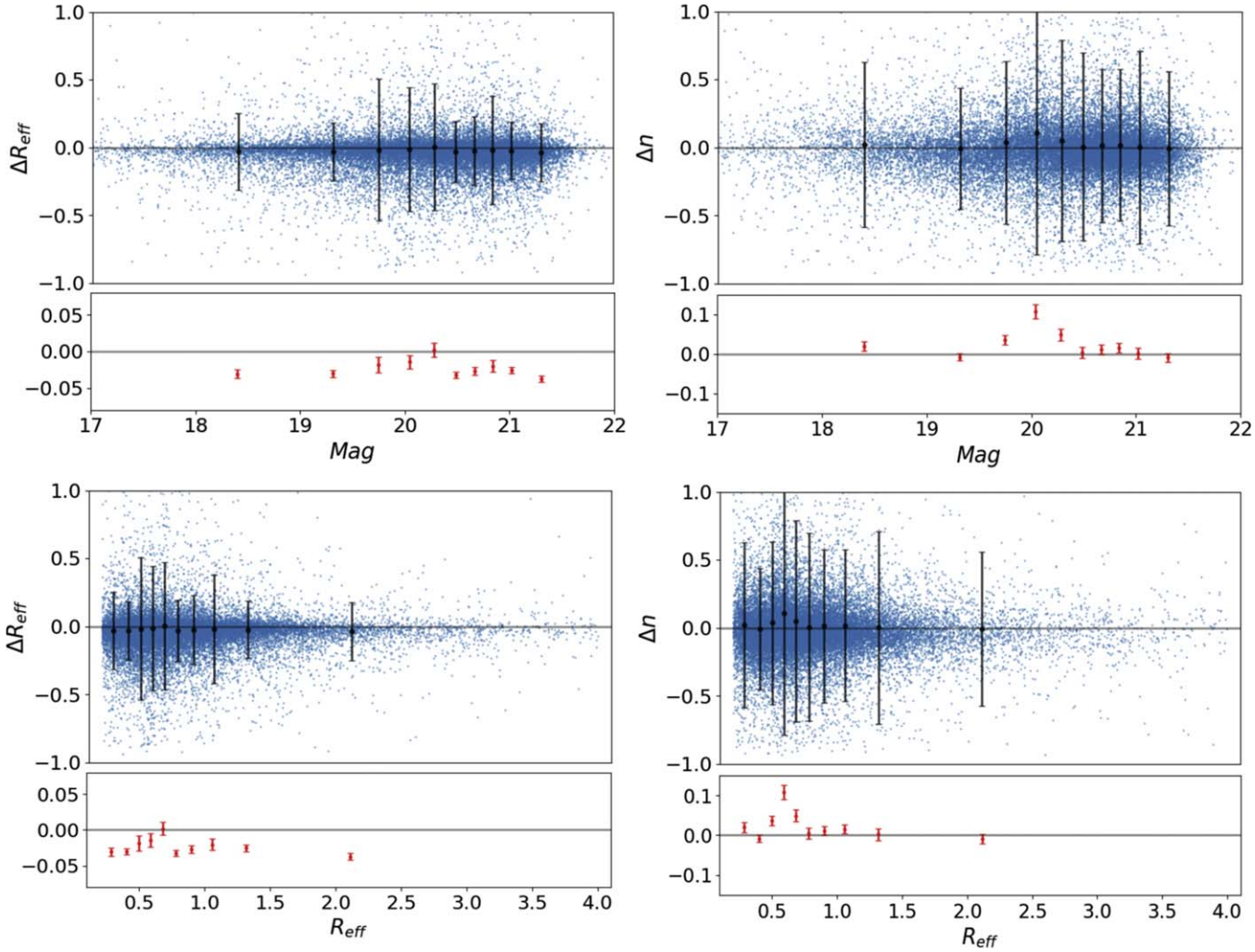

Figure 6. The relative errors of the effective radius ($\Delta R_{\rm eff}$) and Sérsic index (Δn) against the magnitude (mag) and the effective radius ($R_{\rm eff}$). Predictions are from GaLNet-2. In each panel, the blue points are the relative errors, and the dark bars show the standard deviation of the errors, while the red points show the mean of the relative errors and the red bars are the standard error on the mean in each bin.

from the substructure mismatch (i.e., the fact that real galaxies possess extra substructures) or the model mismatch (i.e., the fact that real galaxies might have multiple components, like a bulge and a disk), as discussed in Section 3.3. Moreover, as noted above, both GaLNets and 2DPHOT were used to measure the same galaxies, implying that their uncertainties are not necessarily uncorrelated. Before we move on, we need to point out the presence of some vertical features in Figure 5, suggesting the presence of some systematics in the 2DPHOT estimates. These are found for PA $\sim 90^\circ$ and $q \sim 0.8$ for both comparisons with GaLNet-1 and GaLNet-2. We have visually checked a subsample of these outliers and found that galaxies at PA $\sim 90^\circ$ always have a very round shape. Because for these systems it is hard to define the PA, 2DPHOT tends to give the 90° values, close to the initialization value. As for the second feature at $q \sim 0.8$, we notice that the galaxies always have close pairs, and the light from the pairs has a strong influence on the central galaxies. In some cases, again, 2DPHOT cannot optimize q and finally uses the initial guess as the final output.

In Figure 6, we show the relative errors (top panels) and the standard errors of the mean (bottom panels) of the effective radius and the Sérsic index for GaLNet-2 against the magnitude (mag) and the effective radius ($R_{\rm eff}$), as representative of the behavior seen in Figure 5. These are defined as $\Delta R_{\rm eff} = (R_{\rm eff}^{\rm GN2} - R_{\rm eff}^{\rm 2DPH})/R_{\rm eff}^{\rm 2DPH}$ and $\Delta n = (n^{\rm GN2} - n^{\rm 2DPH})/n^{\rm 2DPH}$, where we have used the apex GN2 and 2DPH for GaLNet-2 and 2DPHOT, respectively. From the two plots in the upper row, we see no clear dependence of the $\Delta R_{\rm eff}$ and Δn on the total magnitude, although the bulk of the scatter seems to increase while the galaxies become fainter. In the left panel of the upper and bottom rows, we see that GaLNet-2 predicts effective radii that are systematically smaller than those from 2DPHOT (as also seen in Figure 5). Statistically, these seem to be marginally significant, as about half of the standard deviation of the mean (red error bars in the bottom panels) generally deviate by more than 2σ from the zero value in the different mag and $R_{\rm eff}$ bins. To investigate in detail the reason for this discrepancy is beyond the purpose of this paper, as this would involve the use of a third tool to cross-check the systematics. This will be explored in more detail in future analyses. The scatter of $\Delta R_{\rm eff}$ seems to also be invariant with $R_{\rm eff}$. Finally, in the right panel of the bottom row, we show Δn against $R_{\rm eff}$. Here we see no clear systematics and fairly similar errors (around 40% and 50%), except for the very small $R_{\rm eff}$. Eventually, for these small galaxies, where the $R_{\rm eff}$ is sampled by $\sim$2 pixels on a side, both

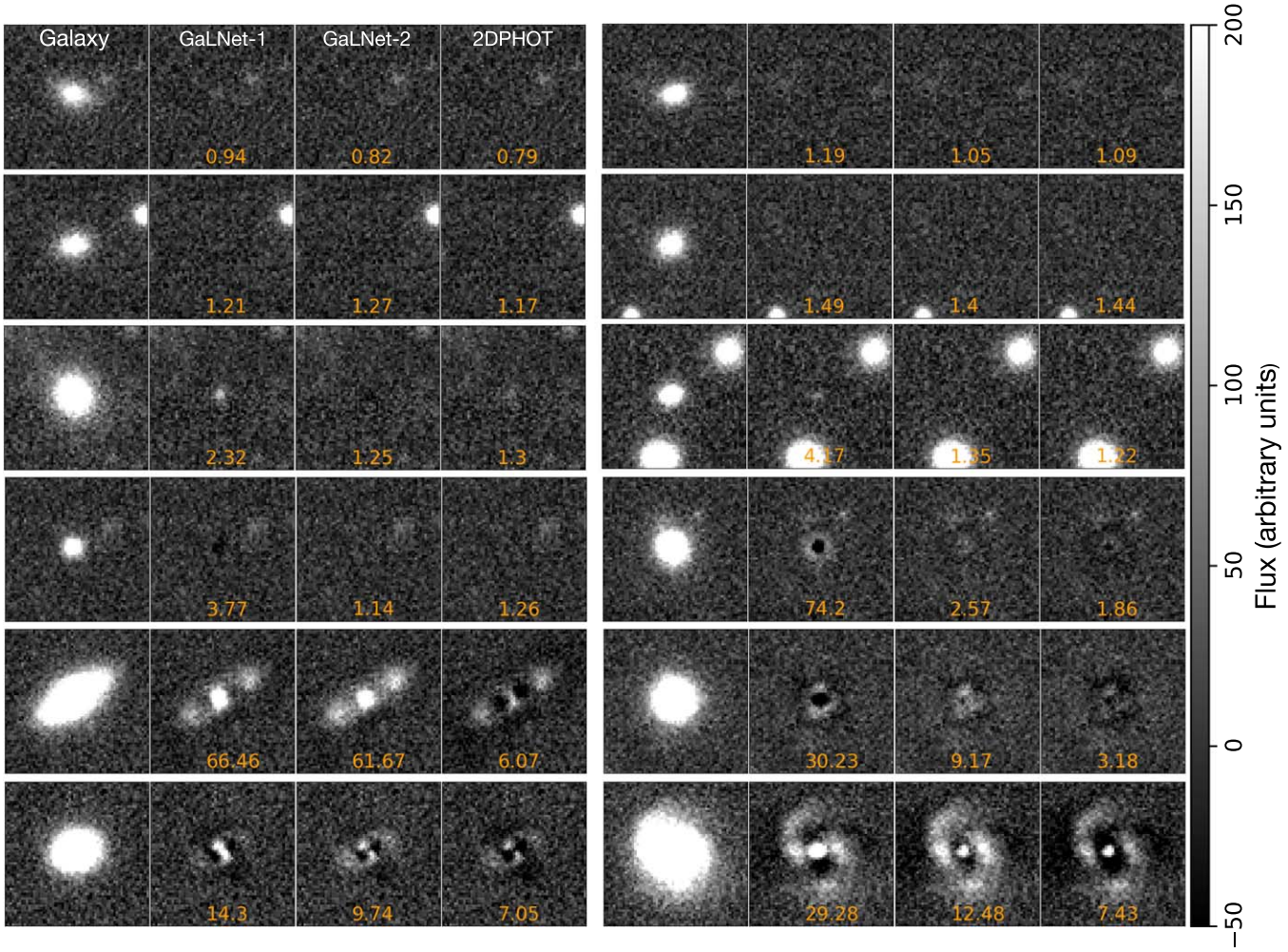

Figure 7. Residuals of real galaxies used for testing the networks. In each panel, the first is the *r*-band image of the real galaxy, the second and third are the residuals obtained from GaLNet-1 and GaLNet-2 predictions, respectively, and the fourth is the residual obtained from 2DPHOT fitting. At the bottom of each residual image, we report the a posteriori reduced $\tilde{\chi}^2$.

2DPHOT and GaLNets can still measure the slope of the light profile with no systematics, but with larger uncertainties. The absence of systematics is particularly encouraging, as this would imply unbiased information on large data sets in this particular region of the parameter space.

In Figure 7 we show some examples of the residuals from the model-subtracted images. These are derived as in Section 3.1. For each target, we show the raw image of the original *r*-band image of the galaxy (first column from the left), and the residuals from GaLNet-1 (second column), GaLNet-2 (third column), and 2DPHOT (fourth column). For most of the galaxies, such as the ones shown in the first two upper rows, the performance of the three tools is fairly good, with almost no residuals.

To quantify this we have computed an a posteriori reduced $\tilde{\chi}^2$, defined as

$$\tilde{\chi}^2 = \sum_i \frac{(p_i - m_i)^2}{\sigma^2_{\rm bkg}}/{\rm dof}, \tag{11}$$

where $_{pi}$ are the observed pixel counts of the galaxy within the effective radius, m_i the model values, $\sigma^2_{\rm bkg}$ the background noise, and dof = (No. of pixels − No. of fit parameters). This definition of χ^2 for both GaLNets and 2DPHOT is different from the one used in Roy et al. (2018) for 2DPHOT and serves here only as an overall estimate of the relative goodness of the final GaLNets' fits and compare these with 2DPHOT. These are reported in Figure 7, where one can see that smaller χ^2 correspond, in general, to smaller local residuals.

In the same Figure 7, we also see cases where GaLNet-1 cannot perform as well as the other two methods, as we can spot darker/brighter dots in the center due to over/under-subtraction, respectively (see, e.g., third and fourth rows). To our surprise, even if the CNNs have been trained on smooth one-component profiles, in the case of bright real substructures like spiral arms, bars, cores, and rings, the two GaLNets tend to predict rather reasonable Sérsic profiles, without catastrophic residuals. This is not what often happens to standard tools based on χ^2 minimization or Markov Chain Monte Carlo (MCMC), which sometimes diverges and returns failing results. As an example, we have checked the tail of the solutions located at the lower-right corner of the total magnitude plot (e.g., 2DPHOT: $20 < {\rm mag} < 22$, GaLNets: $17 < {\rm mag} < 19$, in Figure 5) and the other outliers toward the top-left corner in the same plots. Such large outliers are unexpected for the mag prediction as this is one of the tightest constrained parameters both from 2DPHOT and GaLNets.

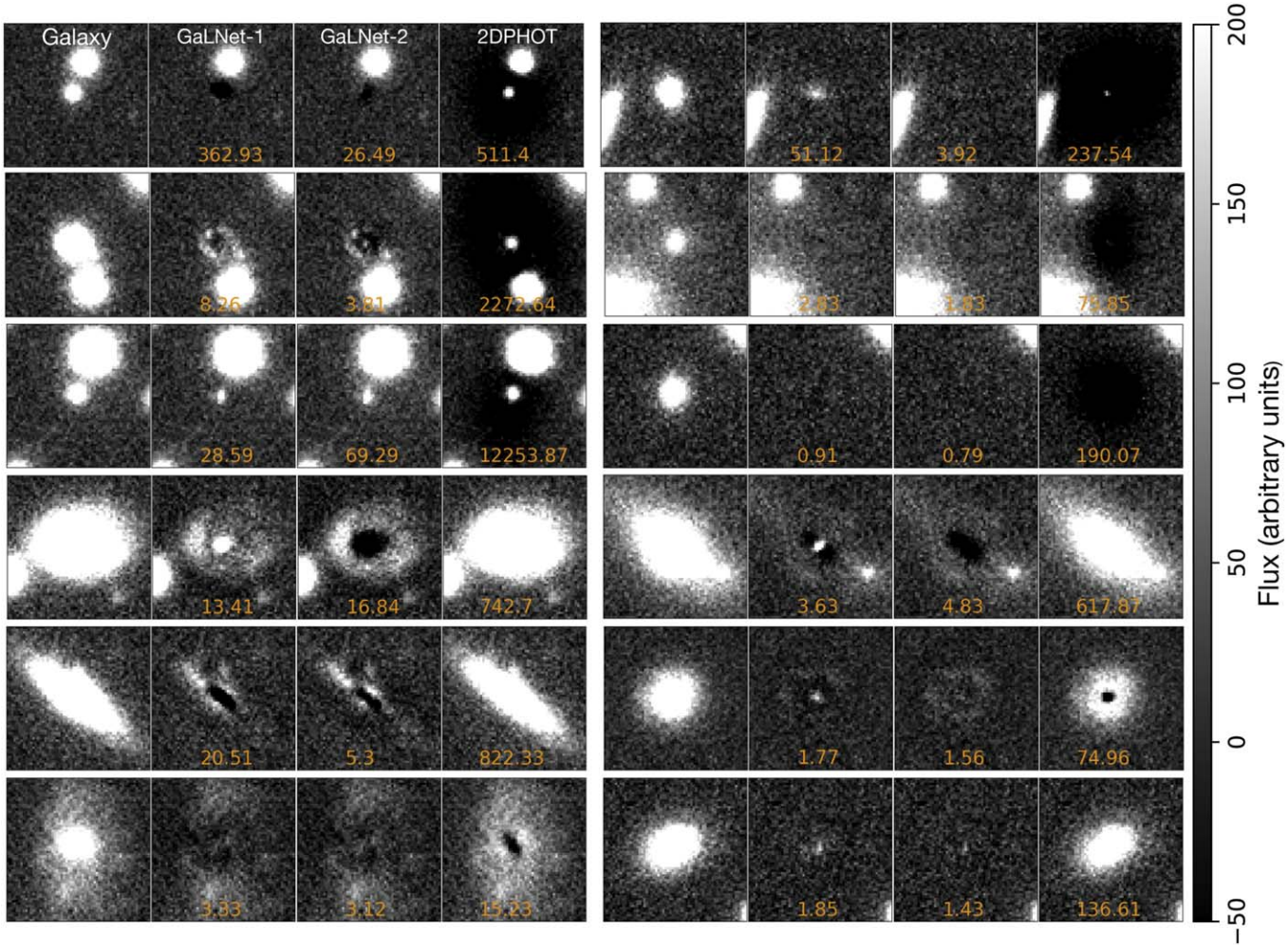

Figure 8. GaLNet vs. 2DPHOT modeling for outliers in the magnitude estimates. These are galaxies expected to have failed to converge to an acceptable model. Panels are arranged as in Figure 7.

In Figure 8 we show the residuals of some of these "outliers." The three upper rows are six outliers from the top-left corner of Figure 5, i.e., systems for which 2DPHOT measures a bright magnitude and the two GaLNets predict they are faint ($r > 21$) objects. It is clear, in all cases, that 2DPHOT overpredicts the magnitude, ending in a wide dark area in the center of the residual image. This is due to a mix of situations: (1) the presence of close objects and (2) the presence of a steep peak in the galaxy center. In this latter case (usually a small bulge, but also an overlapping compact object; see, e.g., top left) the two GaLNets focus on the overall 2D light gradients and ignore the central peak, hence producing a rather satisfactory fit. In all other cases, the presence of close systems could make the segmentation images produced by S-Extractor (Bertin & Arnouts 1996) and used by 2DPHOT to be not very accurate. Therefore, it becomes more difficult for 2DPHOT to find a correct fit, as the contribution of excess light in the outskirts pushes it to converge toward high-n models.

Going to the bottom three rows in Figure 8, i.e., systems for which 2DPHOT measures a faint magnitude and the two GaLNets predict they are bright ($r < 19$) objects, the situation is reversed, but yet it is mainly the standard model failing. Here, the problem seems to be the presence of close companions that 2DPHOT fails to mask, leading to some underfit. The two GaLNets, on the other hand, seem to nicely catch the main smooth component of the galaxies, correctly leaving the peaked components behind.

There are two aspects of this test that are particularly significant. First, the CNN approach does not generally produce catastrophic events, at least in terms of total magnitude, while standard methods sometimes do. Also, CNNs always converge while standard methods sometimes do not. Second, CNNs seem to learn how to correctly "weight" the main component of a galaxy, as they learn from the training sample that this can be mixed to other secondary substructures, either close galaxies or stars, or small embedded systems (see, e.g., Figure 2). This natural "penalization," which is intrinsic in the learning process, turns out to solve very efficiently most of the issues related to standard tools in terms of over/underfitting of the data. These are not negligible capabilities that represent, in our view, a strong advantage of the ML approaches, provided that the structural parameters of the targets fall within the properties of the training sample.

Of course, in all these analyses above, we have left in the background the model mismatch imposed by our initial choice of the single Sérsic profile. Most of the catastrophic outliers, a minor fraction amounting to $\sim$1% of all galaxies analyzed in the comparison shown above, could be resolved by adopting more-realistic two-component models for both the standard and CNN approaches. This will be addressed in forthcoming

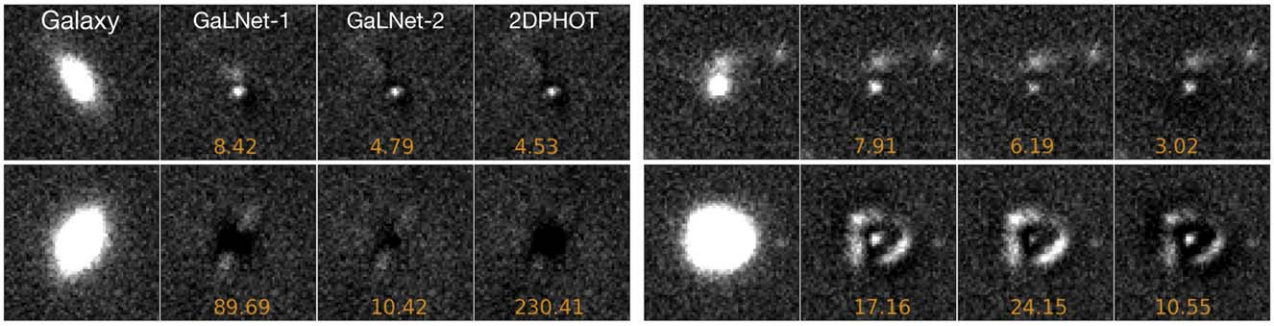

Figure 9. We show some special features, like compact cores (A), interactions (B), residual disks (C), and pseudo-lensing (D), in the residuals of real KiDS galaxies. Panels are arranged as in Figure 7.

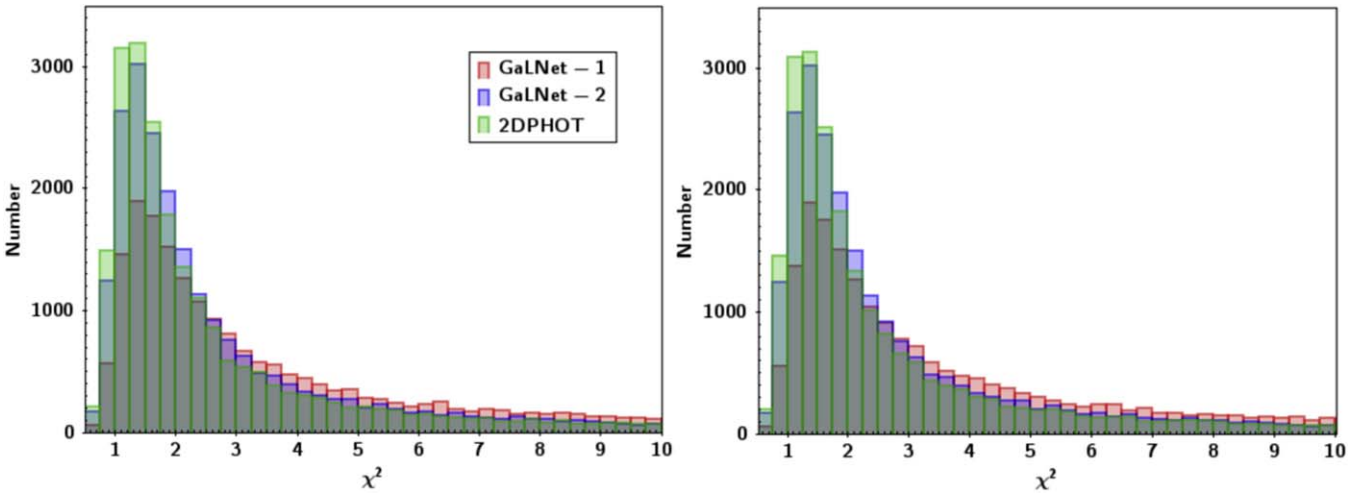

Figure 10. χ^2 distribution of GaLNet-1, GaLNet-2, and 2DPHOT. The ones on the left are obtained using pixels within the individual method effective radius, while the ones on the right are obtained from pixels inside the same aperture defined by the effective radius from GaLNet-2. The *x*range is cut to $\chi^2 = 10$ for the sake of readability, and the distribution is almost flat outside the window.

papers. Here we just stress that GaLNet-2, in particular, is able to capture the main component, when evident, and leave behind other substructures, like clear spiral arms (see Figure 6) as well as more peculiar features, like compact cores, interactions, residual disks, pseudo-lensing configurations, etc. (see Figure 9).

Noticeably, robust one-Sérsic analyses are extremely valuable per se, as they provide crucial parameters for a large sample of galaxies (total luminosity, sizes, and *n* index, as proxies of the morphology) with enormous science implications.

4. Discussion and Perspectives

In this section, we discuss in more detail the possible reasons for the different levels of performance of GaLNet with respect to standard 2D fitting approaches and roughly estimate the gain in terms of computational time. We will also discuss the perspectives in terms of future development of GaLNet tools.

4.1. GaLNets versus Standard Methods

In Section 3 we have compared the results of the two GaLNets against each other, and against 2DPHOT, and discussed the relative scatter as a measure of their internal statistical errors. We have shown that the performance of GaLNet-2 is generally better than that of GaLNet-1 both in reproducing the ground truth of the simulated (test) sample and the estimates from 2DPHOT of the KiDS (real) galaxies.

Here we briefly look at a more quantitative comparison between the different tools and discuss if there are systematic differences and their origins.

In Figure 10, we show the distribution of the $\tilde{\chi}^2$ defined in Equation (11) for the two GaLNets and 2DPHOT. To make a uniform comparison among the different tools, we show both the $\tilde{\chi}^2$ obtained using pixels within the individual method effective radius and the one obtained from pixels inside the same aperture, defined by the effective radius from GaLNet-2, for all of them. As we can see, the two definitions are almost equivalent at all $\tilde{\chi}^2$, as they do not produce appreciable differences in their distributions. The major relative differences are possibly present at larger $\tilde{\chi}^2$, where the differences in the estimated parameters are larger, including strong outliers, as discussed in the previous sections. Generally speaking, we see that the $\tilde{\chi}^2$ of 2DPHOT tend to be better than the one produced by GaLNet-2 because the distribution is peaked slightly toward a smaller $\tilde{\chi}^2$ (closer to 1). GaLNet-2 has a small excess of $\tilde{\chi}^2$ around 2–3, while the two distributions are identical for $\tilde{\chi}^2 > 3$. GaLNet-1, on the other hand, is more broadly distributed with a larger number of systems at $\tilde{\chi}^2 > 3$ than the other two algorithms. As already discussed in Sections 3.1 and 2.2, the main reason for this overall underperformance of GaLNet-1 is the unaccounted-for local PSF.

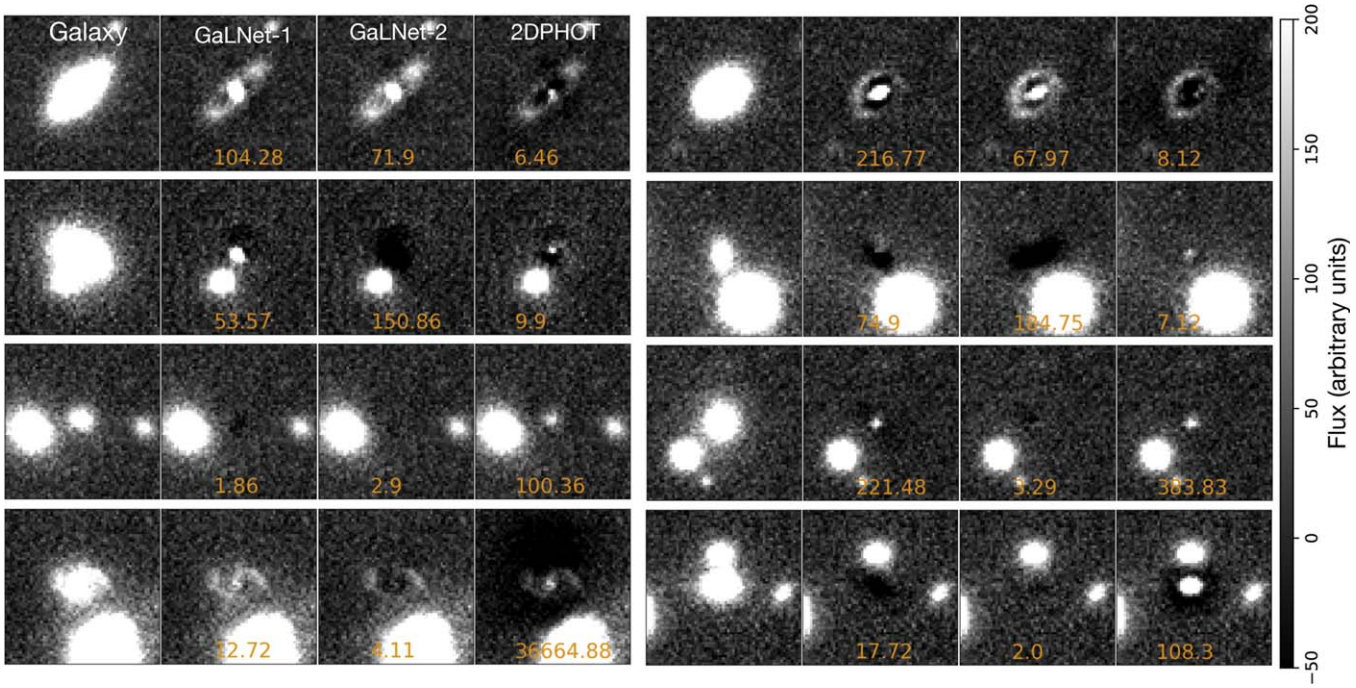

Figure 11. Special cases of the model of real KiDS galaxies. We show residuals with $\tilde{\chi}^2_{\rm GN2} \gg \tilde{\chi}^2_{\rm 2DPH}$ in the top two rows and residuals with $\tilde{\chi}^2_{\rm GN2} \ll \tilde{\chi}^2_{\rm 2DPH}$ in the bottom two rows. Panels are arranged as in Figure 7.

Going to the comparison between GaLNet-2 and 2DPHOT, we have observed that the former performs sensitively worse than the latter, i.e., $\tilde{\chi}^2_{\rm GN2} \gg \tilde{\chi}^2_{\rm 2DPH}$ in two major circumstances. The first is when there are bright substructures in the center (e.g., spiral arms, rings, etc.; see Figure 11, top two rows): In this case, GaLNet-2 tends to predict a flatter profile and underfit the outer regions (hence leaving bright residuals in the galaxy outskirts). Second, when there is a bright companion, GaLNet-2 tends to overfit the model in the center and produce darker regions and a worse $\tilde{\chi}^2$.

Finally, there are cases where 2DPHOT performs much worse than GaLNet-2, i.e., $\tilde{\chi}^2_{\rm 2DPH} \gg \tilde{\chi}^2_{\rm GN2}$ (see Figure 11 last two rows). In around 50% of these cases, we have seen bright sources around the main galaxy, which is generally a compact one. Here 2DPHOT is likely unable to mask out the companion galaxies, and this might produce flatter tails in the measured profile, which is fitted with a larger $R_{\rm eff}$ and/or smaller n, producing an underfitting of the central regions that remain unsubtracted in the residual images. GaLNet-2, on the other hand, being insensitive to the outer regions, can fit better the main galaxy component. For the other 50% of the cases, we see a mix of situations with no clear pattern, including rather crowded regions, bright large galaxies, and strong blended sources, where both 2DPHOT and GaLNet-2 can either perform better or worse, rather unpredictably. Finally, we have a mixed situation with $\tilde{\chi}^2_{\rm 2DPH} < \tilde{\chi}^2_{\rm GN2}$, or vice versa, and both $\tilde{\chi}^2 \gg 1$, where, looking at the residuals, it is hard to really appreciate sensitive differences.

Overall, these detailed comparisons above give useful indications about improvements to implement for the next GaLNets. For instance, by increasing the number of cutouts of the background sample with bright galaxies in them, the CNN will eventually learn how to predict the parameters in the presence of a more severe galaxy–galaxy blending (Section 4.2).

However, according to the current GaLNets versus 2DPHOT results, the broad conclusion from Figure 10 is that GaLNet-2 has a predictive power comparable to traditional methods.

The computational speed, instead, is a major strength of GaLNets. This can depend on many factors, including the performance of the hardware underuse. However, to give a scale of the relative computing times, we have made a statistic for individual galaxy from the 2DPHOT run, which takes a few seconds for the 2D fitting, and a similar time for the preliminary 2D fitting that 2DPHOT performs to determine the initial guess parameters (La Barbera et al. 2008). As a comparison with another standard 2D fitting tool, we have checked the average performance of Galfit on a bunch of KiDS galaxies using the 2DPHOT parameters as a first guess and measured on average $\sim$6 s/galaxy computing time. Even ignoring the time needed for the first guess step, the standard tools need on the order of seconds to converge. On the other hand, we have seen that the CNNs need only $\sim$0.04 s/galaxy running on standard CPUs and about a factor of 10 faster running on commercial GPUs on a laptop. That is, GaLNets can work at a speed that is a factor of $\sim$250 to 1500 faster with respect to standard tools, with hardware that is about 4 times cheaper. If we include the necessary optimization step of the initial guess parameters, which for the CNN is incorporated in the training step and is negligible in time for a billion-galaxy sample, then GaLNets are overall a further factor of 10 faster than the standard tools.

With these exceptional computational capabilities, a billion galaxies in 10 optical + near-infrared bands, like the sample we will collect by combining Rubin/LSST, EUCLID, and CSST, will be doable for GaLNets in a week's time on a midsize GPU server. This makes GaLNets, and other similar ML tools (e.g., DeepLegato) a very promising solution for the next-generation large sky surveys.

4.2. Future Developments

In this section, we first discuss the possibilities for improving the performance of the current GaLNetS. Then we will discuss how to expand their applicability to solve some issues related to the structural parameter measurement, for instance, the deblending (as anticipated in Section 4.1).

The first line in GaLNet improvements is related to the treatment of the PSFs in the training-sample simulations. In this process, randomly selected PSFs were used to convolve the Sérsic profiles. Hence, the PSFs of the central galaxies are different from the local ones of the surrounding galaxies. This can be a problem for GaLNet-2 to learn how to deblend two or more galaxies as this will not reflect real situations where there is a unique PSF. This PSF-mismatch will be solved in the future by first selecting a "PSF sample" and then assembling the "background sample" in the vicinity of the selected PSF, hence building the "PSF-background pair" sample. In this case, when a galaxy is simulated in the center of the background image, the properly paired PSF is convolved with its Sérsic profile. The PSF can be either provided as a Moffat model, as discussed in Section 2.2, or can be possibly provided as the image of the representative star used to estimate the local PSF.

As a second improvement, we will extend the application to fainter galaxies with lower S/N. This paper has only focused on high "average" S/N galaxies. However, the lower-S/N galaxies account for most of the observed targets in galaxy surveys, and they usually have lower masses or are located at higher redshift. These two properties are valuable for studies of the galaxy structure, formation, and evolution and for many scaling relations. Being more sensitive to the central regions of the galaxy, as discussed in Section 2.2, we will check whether GaLNets may perform reasonably well with lower average S/Ns.

As a further improvement, thanks to their computational speed, GaLNets will make it possible to derive the light profiles for multiband images. In particular, we can train a CNN to simultaneously predict all the parameters from all bands. This can help prevent single-band outliers as one can expect a smooth continuity among the structural parameters in different bands (see, e.g., the magnitude outliers in Figure 8). With such a regular multiband fitting, one can expect to reconstruct the 2D model of the galaxy SED for stellar population analysis.

Finally, we will implement the multi-Sérsic profiles in the training sample. Here we expect the degeneracies among the parameters to be increased by the presence of multiple components, but we will possibly take advantage of the simultaneous multiband fitting above to break some of these degeneracies, as the different components might look different in the different bands.

One missing feature of the current GaLNets is that, as with all CNNs, they cannot provide the errors on parameters of individual galaxies (but statistical errors can reasonably be derived from the test samples). Possible solutions here can be Bayesian neural networks (e.g., for the strong-lensing model; Perreault Levasseur et al. 2017) or the adoption of some hybrid algorithms, e.g., using the prediction from CNNs as an initial guess for a standard MCMC code (e.g., Pearson et al. 2021).

To conclude this section, we want to list a few further applications that can be performed by GaLNets, with some simple variations on the training methods. For instance:

1. Star/galaxy separation. Using the information on the shape and internal light profiles, we expect that training GaLNets on simulated PSFs and galaxies will enable the CNNs to work as a classifier to assign the probability of being a star to each modeled object. This is a rather standard task that has been implemented with sophisticated standard or ML techniques (e.g., La Barbera et al. 2008; Khramtsov et al. 2019; Muyskens et al. 2022). However, integrating this step in the analysis will make GaLNets capable of autonomously selecting the stars for the PSF modeling, hence taking a step toward the full automation of the tool.
2. Deblending. As discussed in Section 4.1, we can improve the capability of the CNN to make models of galaxies blended with close companions. This might have some limits (in terms of accuracy) for a general-purpose tool, where one expects the majority of the systems to be almost isolated galaxies. On the other hand, we can specialize a GaLNet only on deblending systems, focusing on a smaller number of relevant parameters (e.g., total luminosity and half-light ratio) but trying to optimize the deblending performance.

5. Conclusions

With the upcoming facilities for wide-field sky surveys, we will enter a new era in which an enormous amount of data will be available to study the life of the universe and the evolution of galaxies therein in much detail. The physical processes driving this evolution are heavily connected to galaxy properties, like mass, size, colors and color gradients, shape, internal kinematics, etc., and are dependent on the epoch the galaxies are observed, i.e., their redshifts.

Measuring the structural properties of millions to billions of galaxies is a prerequisite to performing great science, but it will represent a major challenge for different scientific communities, not to mention other complex and time-consuming tasks like source deblending and star/galaxy separation. In preparation for the upcoming data flows, there is a significant effort from the community to investigate and test new techniques, based on ML, to perform these tasks efficiently. In some specific applications, ML tools are at a rather advanced stage: e.g., photometric redshifts (Carrasco Kind & Brunner 2013; Sadeh et al. 2016) and strong gravitational lens classifiers (Petrillo et al. 2017, 2019b; Li et al. 2020, 2021).

There are, though, other applications that are still at the pioneering stage. This is the case with the galaxy structural parameter measurement, which is a major outcome expected to be provided by these billion-galaxy surveys. To start filling this gap, we have built two CNN tools to fit the surface brightness profiles of galaxies from seeing-limited ground-based observations. One network is fed with only galaxy images (GaLNet-1) while the other uses galaxy images together with the local PSF measured in the galaxy surroundings (GaLNet-2). The two CNNs are trained with 200,000 mock galaxies created by adding PSF-convolved simulated Sérsic profiles on top of randomly selected cutouts from public KiDS images. The outputs of the CNNs include seven parameters used to describe the Sérsic profiles of the galaxies: the coordinates of the center position $x_{\rm cen}$, $y_{\rm cen}$; magnitude mag; effective radius $R_{\rm eff}$; axis ratio q; position angle PA; and Sérsic index n. We have tested

the CNNs on both mock data and real galaxy cutouts collected from KiDS.

We have found that the two CNNs can predict values for the structural parameters that match well with the input values of the simulated galaxies. We have also shown that, when applied to real KiDS galaxies, the predictions are quite consistent with the best-fit values determined from standard PSF-convolved 2D Sérsic profile-fitting procedures (i.e., 2DPHOT; Roy et al. 2018). Specifically, the CNN using the local PSF (GaLNet-2) performs very similarly to 2DPHOT, while the CNN trained with only the galaxy images (GaLNet-1) performs slightly worse, albeit still acceptable. These results demonstrate that GaLNets are a promising class of regression CNNs to perform an accurate analysis of the surface brightness of galaxies. We have also shown that they are fast, up to 1000 times faster than standard tools based on standard statistics like χ^2 minimization, such as 2DPHOT, which is comparable to, if not faster than, other standard tools based on likelihood maximization or MCMC.

Future implementations of GaLNets will add complexity to the ground truth, e.g., using multiple-component models and substructures and obtaining simultaneous predictions in multiple bands. Moreover, we can further generalize them to perform other kinds of difficult tasks, such as source deblending or star/galaxy separation.

For all these reasons, GaLNet will represent a unique and unprecedented tool to perform in a very efficient way a series of tasks on data sets from future ground- and space-based facilities (e.g., HSC DES, Rubin/LSST, Euclid, and CSST).

R.L. acknowledges the science research grants from the China Manned Space Project (Nos. CMS-CSST-2021-B01, CMS-CSST-2021-A01). N.R.N. acknowledges financial support from the "One hundred top talent program of Sun Yat-sen University" grant No. 71000-18841229.

Appendix
Prior Distribution of Simulations

In Section 2.2, we have discussed the impact of the distribution of parameters of the mock galaxies used as training and testing sets on the performance of GaLNets applied to real data. In particular, we have used parameter distributions similar to the ones observed in the galaxy sample we expected to analyze, e.g., normal distribution for $\log R_{\rm eff}$ and F distribution for n. Generally speaking, these prior distributions are unknown, thus we need to test the impact of inaccurate choices on the final predictions from GaLNets. However, in doing this, we should be cautious to adopt unrealistic parameter distributions that might sample a volume in parameter space that is inaccessible to real galaxies. This can, as a minimal effect, degrade the overall training process because of the loss of useful information in realistic parameter combinations. We stress here, though, that this is the first time such experiments were made to evaluate the response of deep learning in galaxy light profiles using all parameters simultaneously. Hence, we can use this test as a pathfinder for future setup, although the overall discussion of the prior distribution is beyond the scope of this paper, although it will be tackled in future analyses.

For this first check, besides the specific choice of the adopted distributions, we also consider how a simple variation of the properties of these distributions can impact the final results. In particular, we want to quantify this effect by testing GaLNets trained on a sample of simulated galaxies for which we have perturbed the prior distributions of the main parameters determining their light profiles, namely$R_{\rm eff}$ and n. This test is meant to reproduce the systematics one might expect if the CNNs are trained over a simulated galaxy sample that deviates from real systems. In particular, for the training data, we adopt the following modifications:

1. We have moved the peak of the $\log R_{\rm eff}$ and$n$ distributions by 30% and 50% (see Figure 12), respectively.
2. We have changed the distributions of $R_{\rm eff}$ and n to flat.

These tests allow us to check GaLNets' response over a wide variety of reality mismatches, from extreme (flat) to reasonably moderate (30%) ones, and estimate their impact on the final parameter predictions. The results are shown in Figure 12 for GaLNet-2. Here, we see the predictions from GaLNet-2 (y-axis) trained on the perturbed sample versus the fitting results obtained by 2DPHOT (x-axis). We stress that the training samples have the two parameters changed at the same time, so this is the cumulative effect of a prior mismatch for both of the parameters with the same amount of deviation from the true distributions (the perturbed ones). In this respect, these all represent some worst-case scenarios of poor knowledge of the intrinsic galaxy parameters, while in reality the effective radius is known more robustly than the $n$index.

In Figure 12, we can see that the impact on $R_{\rm eff}$ and n is different. For the former, moving the peak or changing the distribution to flat almost has no obvious impact on the overall performance. For the n index, at larger values ($n > 4$), there are small systematics for small deviations (30%). However, compared with the original one, the scatters do not increase. For larger mismatches of the priors (50%), the systematics of the n index become rather severe. Going to the flat distribution, surprisingly, the scatters at larger values ($n > 5$) become smaller than that in the original figure and also the mean values seem more aligned with the one-to-one line. However, a peculiar overdensity of estimates systematically above the one-to-one line is evident at $4 \lesssim n \lesssim 6$ (see the red arrow in Figure 12), which is compensated by the long tail at the bottom of the same line, leaving the mean value in those bins well aligned with the one-to-one line. This feature and the fact that the accuracy and scatter of the estimates in the low-n (< 4) bins are worse than the original distribution in the first panel on the left make the flat distribution a reasonable but not compelling option. Finally, we remark that, apart from$R_{\rm eff}$ and n, we also tested the effects of the distribution of other parameters and no obvious changes were found.

To conclude, we believe that a deeper investigation of the prior mismatch is eventually needed, although the first test presented here does not seem to show that there is much room for improvement. It is potentially worth looking more into the impact of the flat distributions on both $R_{\rm eff}$ and n, which possibly mask obvious areas of the parameter space that are away from typical scaling relations of galaxies. Overall, we need to look more into the reasons for the (still) unsatisfactory number of outliers that is possibly driving part of the statistical difference between the predictions from different methods. As discussed in Section 4.2, this might be due more to the deblending issue than the poor performance of GaLNets from biased training sets.

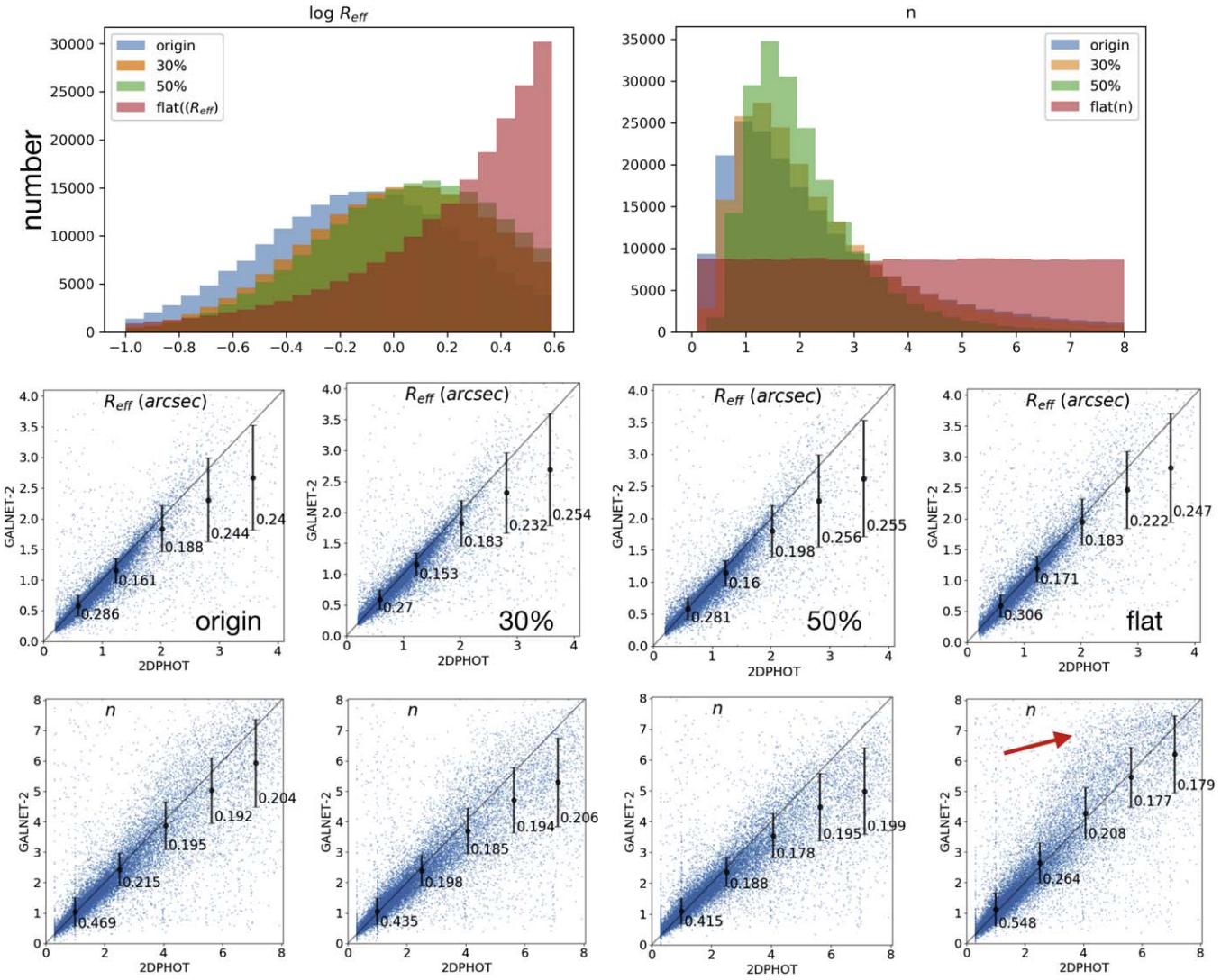

Figure 12. Impact of prior distributions on GaLNet-2 predictions. The two panels in the first row show the distributions used in simulating the training data. Blue is the original one used in the main paper, orange and green are the ones with the 30% and 50% peak moving toward the right, and red is the one with the flat distribution. The second and third rows show the results obtained with the real distribution (origin), 30% and 50% peak offset with respect to the original one, and a flat distribution, respectively. The absolute error bars are also plotted for comparison. The red arrow shows a peculiar feature in the *n*-index predictions (see discussion in the text).

ORCID iDs

R. Li https://orcid.org/0000-0002-3490-4089
N. R. Napolitano https://orcid.org/0000-0003-0911-8884
C. Tortora https://orcid.org/0000-0001-7958-6531

References

Amaro, V., Cavuoti, S., Brescia, M., et al. 2021, in Intelligent Astrophysics, ed. I. Zelinka, M. Brescia, & D. Baron (Cham: Springer), 245
Baldry, I. K., Driver, S. P., Loveday, J., et al. 2012, MNRAS, 421, 621
Barden, M., Häußler, B., Peng, C. Y., McIntosh, D. H., & Guo, Y. 2012, MNRAS, 422, 449
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bilicki, M., Hoekstra, H., Brown, M. J., et al. 2018, A&A, 616, A69
Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964
Boucaud, A., Huertas-Company, M., Heneka, C., et al. 2020, MNRAS, 491, 2481
Bradley, L., Sipőcz, B., Robitaille, T., et al. 2020, astropy/photutils: v1.0.0, Zenodo, doi:10.5281/zenodo.4044744
Brownstein, J. R., Bolton, A., Schlegel, D., et al. 2012, ApJ, 744, 41
Buck, T., & Wolf, S. 2021, arXiv:2111.01154
Buitrago, F., Trujillo, I., Conselice, C. J., et al. 2008, ApJL, 687, L61
Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, A&A, 644, A163
Capaccioli, M., Caon, N., & D'Onofrio, M. 1992, MNRAS, 259, 323
Carrasco Kind, M., & Brunner, R. J. 2013, MNRAS, 432, 1483
Ciambur, B. C. 2015, ApJ, 810, 120
Ciotti, L., & Bertin, G. 1999, A&A, 352, 447
Csurka, G. 2017, arXiv:1702.05374
de Jong, J. T. A., Kleijn, G. A. V., Boxhoorn, D. R., et al. 2015, A&A, 582, A62
de Jong, J. T. A., Kleijn, G. A. V., Erben, T., et al. 2017, A&A, 604, A134
Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441
Dimauro, P., Huertas-Company, M., Daddi, E., et al. 2018, MNRAS, 478, 5410
Dressler, A., Lynden-Bell, D., Burstein, D., et al. 1987, ApJ, 313, 42
Faber, S. M., & Jackson, R. E. 1976, ApJ, 204, 668
Friedman, J. H. 1999, AnSta, 29, 1189
Frontera-Pons, J., Sureau, F., Bobin, J., & Le Floc'h, E. 2017, A&A, 603, A60
Furlong, M., Bower, R. G., Crain, R. A., et al. 2016, MNRAS, 465, 722
Gal, Y., & Ghahramani, Z. 2015, arXiv:1506.02158
Hyde, J. B., & Bernardi, M. 2009, MNRAS, 394, 1978
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, ApJS, 243, 17
Kawinwanichakij, L., Silverman, J. D., Ding, X., et al. 2021, ApJ, 921, 38
Khramtsov, V., Sergeyev, A., Spiniello, C., et al. 2019, A&A, 632, A56
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980

Kormendy, J. 1977, ApJ, 218, 333
Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2
La Barbera, F., & de Carvalho, R. R. 2009, ApJL, 699, L76
La Barbera, F., de Carvalho, R. R., de La Rosa, I. G., et al. 2010, MNRAS, 408, 1313
La Barbera, F., de Carvalho, R. R., Kohl-Moreira, J. L., et al. 2008, PASP, 120, 681
Lange, R., Driver, S. P., Robotham, A. S. G., et al. 2015, MNRAS, 447, 2603
Lange, R., Moffett, A. J., Driver, S. P., et al. 2016, MNRAS, 462, 1470
Lauer, T. R. 1985, ApJS, 57, 473
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
Li, R., Shu, Y., Su, J., et al. 2019, MNRAS, 482, 313
Li, R., Napolitano, N. R., Spiniello, C., et al. 2021, ApJ, 923, 16
Li, R., Napolitano, N. R., Tortora, C., et al. 2020, ApJ, 899, 30
Muyskens, A. L., Goumiri, I. R., Priest, B. W., et al. 2022, AJ, 163, 148
Oser, L., Naab, T., Ostriker, J. P., & Johansson, P. H. 2012, ApJ, 744, 63
Oser, L., Ostriker, J. P., Naab, T., Johansson, P. H., & Burkert, A. 2010, ApJ, 725, 2312
Pearson, J., Maresca, J., Li, N., & Dye, S. 2021, MNRAS, 505, 4362
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2011, GALFIT: Detailed Structural Decomposition of Galaxy Images, Astrophysics Source Code Library, ascl:1104.010
Perreault Levasseur, L., Hezaveh, Y. D., & Wechsler, R. H. 2017, ApJL, 850, L7
Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129
Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2019a, MNRAS, 482, 807
Petrillo, C. E., Tortora, C., Vernados, G., et al. 2019b, MNRAS, 484, 3879
Robertson, B. E., Banerji, M., Cooper, M., et al. 2017, arXiv:1708.01617
Roy, N., Napolitano, N. R., La Barbera, F., et al. 2018, MNRAS, 480, 1057
Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, PASP, 128, 104502
Sérsic, J. L. 1968, Atlas de Galaxias Australes (Córdoba: Observatorio Astronómico de Córdoba)
Shen, S., Mo, H. J., White, S. D. M., et al. 2003, MNRAS, 343, 978
Shibuya, T., Ouchi, M., & Harikane, Y. 2015, ApJS, 219, 15
Shu, Y., Bolton, A. S., Mao, S., et al. 2016, ApJ, 833, 264
Simard, L., Willmer, C. N. A., Vogt, N., et al. 2002, ApJS, 142, 1
Simonyan, K., & Zisserman, A. 2014, arXiv:1409.1556
Smith, M. J., Arora, N., Stone, C., Courteau, S., & Geach, J. E. 2021, MNRAS, 503, 96
Stone, C. J., Arora, N., Courteau, S., & Cuillandre, J.-C. 2021, MNRAS, 508, 1870
Tarsitano, F., Bruderer, C., Schawinski, K., & Hartley, W. G. 2022, MNRAS, 511, 3330
Tarsitano, F., Hartley, W. G., Amara, A., et al. 2018, MNRAS, 481, 2018
Tody, D. 1986, Proc. SPIE, 627, 733
Tortora, C., Napolitano, N. R., Spavone, M., et al. 2018, MNRAS, 481, 4728
Trujillo, I., Aguerri, J. A. L., Cepa, J., & Gutiérrez, C. M. 2001a, MNRAS, 328, 977
Trujillo, I., Conselice, C. J., Bundy, K., et al. 2007, MNRAS, 382, 109
Trujillo, I., Graham, A. W., & Caon, N. 2001b, MNRAS, 326, 869
Tuccillo, D., Huertas-Company, M., Decencière, E., et al. 2018, MNRAS, 475, 894
Umayahara, T., Shibuya, T., Miura, N., et al. 2020, Proc. SPIE, 11452, 1145223
van der Wel, A., Chang, Y.-Y., Bell, E. F., et al. 2014, ApJL, 792, L6
Vikram, V., Wadadekar, Y., Kembhavi, A. K., & Vijayagovindan, G. V. 2010, MNRAS, 409, 1379
Wagner-Carena, S., Park, J. W., Birrer, S., et al. 2021, ApJ, 909, 187
Wang, K., Guo, P., & Luo, A.-L. 2017, MNRAS, 465, 4311
Wellons, S., Torrey, P., Ma, C. P., et al. 2015, MNRAS, 449, 361
Wellons, S., Torrey, P., Ma, C. P., et al. 2016, MNRAS, 456, 1030
Yoon, I., Weinberg, M. D., & Katz, N. 2011, MNRAS, 414, 1625
Zhan, H. 2011, SSPMA, 41, 1441