

# The PAU survey: estimating galaxy photometry with deep learning

L. Cabayol,<sup>1</sup>★ M. Eriksen,<sup>1</sup>† A. Amara,<sup>2</sup> J. Carretero<sup>1b</sup>,<sup>1</sup>† R. Casas,<sup>3,4</sup> F. J. Castander,<sup>3,4</sup> J. De. Vicente<sup>1b</sup>,<sup>5</sup>  
E. Fernández,<sup>1</sup> J. García-Bellido,<sup>6</sup> E. Gaztanaga<sup>1b</sup>,<sup>3,4</sup> H. Hildebrandt,<sup>7</sup> R. Miquel,<sup>1,8</sup> C. Padilla,<sup>1</sup>  
E. Sánchez,<sup>5</sup> S. Serrano,<sup>3</sup> I. Sevilla-Noarbe<sup>3</sup> and P. Tallada-Crespí<sup>5</sup>†

<sup>1</sup>*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, E-08193 Bellaterra, Spain*

<sup>2</sup>*Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciana Building, Burnaby Road, Portsmouth PO1 3FX, UK*

<sup>3</sup>*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain*

<sup>4</sup>*Institut d'Estudis Espacials de Catalunya (IEEC), 08193 Barcelona, Spain*

<sup>5</sup>*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), 28040 Madrid, Spain*

<sup>6</sup>*Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*

<sup>7</sup>*Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, D-44780 Bochum, Germany*

<sup>8</sup>*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*

Accepted 2021 June 30. Received 2021 June 30; in original form 2021 April 13

## ABSTRACT

With the dramatic rise in high-quality galaxy data expected from *Euclid* and Vera C. Rubin Observatory, there will be increasing demand for fast high-precision methods for measuring galaxy fluxes. These will be essential for inferring the redshifts of the galaxies. In this paper, we introduce LUMOS, a deep learning method to measure photometry from galaxy images. LUMOS builds on BKGNET, an algorithm to predict the background and its associated error, and predicts the background-subtracted flux probability density function. We have developed LUMOS for data from the Physics of the Accelerating Universe Survey (PAUS), an imaging survey using a 40 narrow-band filter camera (PAUCam). PAUCam images are affected by scattered light, displaying a background noise pattern that can be predicted and corrected for. On average, LUMOS increases the SNR of the observations by a factor of 2 compared to an aperture photometry algorithm. It also incorporates other advantages like robustness towards distorting artefacts, e.g. cosmic rays or scattered light, the ability of deblending and less sensitivity to uncertainties in the galaxy profile parameters used to infer the photometry. Indeed, the number of flagged photometry outlier observations is reduced from 10 to 2 per cent, comparing to aperture photometry. Furthermore, with LUMOS photometry, the photo-z scatter is reduced by  $\approx 10$  per cent with the Deepz machine-learning photo-z code and the photo-z outlier rate by 20 per cent. The photo-z improvement is lower than expected from the SNR increment, however, currently the photometric calibration and outliers in the photometry seem to be its limiting factor.

**Key words:** techniques: image processing – techniques: photometric – galaxies: photometry – cosmology: observations.

## 1 INTRODUCTION

Wide-field galaxy surveys are a powerful tool for cosmology. Galaxy redshifts are the most fundamental property of any cosmological or galaxy evolution study. Spectroscopic surveys, e.g. the Sloan Digital Sky Survey (SDSS, Ahumada et al. 2020), measure very high precision redshifts, however, these are only possible for on the order of a million objects (e.g. BOSS, Dawson et al. 2013). In contrast, imaging surveys are  $\approx 2$  orders of magnitude ahead in terms of number of objects but these are observed with a lower spectral resolution, which makes the redshift measurements less precise. Current and past imaging surveys, e.g. The Kilo-Degree Survey (KiDS, de Jong et al. 2013), The Hyper Supreme-Cam Subaru (HSC, Aihara et al. 2018) or The Dark Energy Survey (DES, DES

Collaboration et al. 2021) have detected hundreds of millions of galaxies and oncoming surveys like *Euclid* (Laureijs et al. 2011) or Vera C. Rubin Observatory (LSST, Ivezić et al. 2019) will increase this number to billions. Consequently, fast and precise methods to analyse and extract galaxy properties (e.g. flux, size, shape) are needed.

There are many different algorithms to estimate galaxy photometry. One very widely used example is SEXTRACTOR (Bertin & Arnouts 1996), which applies aperture photometry (Ni et al. 2019) inspired by the Kron (1980) first moments algorithm. This technique measures the flux of the targeted galaxies by placing an aperture around the source and measuring the light captured inside such aperture. Another technique is model fitting (Hesley 1999), which consists of fitting the galaxy image to a theoretical model and extract its photometry. This includes the GAAP (Kuijken 2008) algorithm, which estimates the total flux by fitting the pixelated galaxy images to polar shapelets (Refregier 2003; Massey & Refregier 2005), separating the galaxy image into components with explicit rotational symmetries.

\* E-mail: [lcabayol@ifae.es](mailto:lcabayol@ifae.es)

† Also at Port d'Informació Científica (PIC), Campus UAB, C. Albareda s/n, 08193 Bellaterra (Cerdanyola del Vallès), Spain.

There are many other examples as e.g. PROFOUND (Robotham et al. 2018), T-PHOT (Merlin et al. 2015) and TRACTOR (Lang, Hogg & Schlegel 2016), and each of them is adjusted to outperform the others on a particular data set. For instance, a photometry algorithm can be optimized to work very well on images with many blended galaxies (e.g. Boucaud et al. 2020) while another can be intended to improve the photometry of very noisy galaxies. Therefore, depending on the type of data and the science goals, different methodologies are applied to improve the photometry estimation.

Although all these algorithms have proven to work well, they also have their shortcomings. Aperture photometry works very well on clean images, but it is not robust towards distorting effects such as e.g. blended galaxies, variable background light and cosmic rays. On the other hand, model fitting is sensitive to model parametrization. In contrast, this is not the case for machine-learning techniques, which learn a model adapted to the data. Furthermore, deep learning has proven to be very powerful on image recognition and computer vision (e.g. Girshick 2015; Zhao et al. 2019), which makes it a robust tool to work on images with artefacts and variant effects. Also, the evaluation of a trained machine-learning algorithm is very fast, which is relevant when dealing with very large amounts of data. For instance, Haigh et al. (2021) compares several source-extraction codes and conclude that currently there is no tool sufficiently fast and accurate to be well suited to large-scale automated segmentation.

Deep learning has already been applied to different steps of astronomical imaging photometry, e.g. photometry of blended galaxies (Boucaud et al. 2020), PSF simulation (Herbel et al. 2018), cosmic-ray rejection (Zhang & Bloom 2020) or source detection (Hausen & Robertson 2020). The power of deep learning techniques on object detection or image recognition tasks makes these steps of the data reduction, among others, very suitable candidates to apply machine learning. While addressing them with classical methods can be difficult and computationally expensive, deep learning is an effective tool to tackle the problem.

In this paper, we present LUMOS,<sup>1</sup> a deep learning-based algorithm to extract the photometry from astronomical images. It consists of a CNN that works on input galaxy images and a mixture density network (MDN) that outputs the probability distribution of galaxy fluxes. LUMOS builds on BKGNET (Cabayol-Garcia et al. 2020), a deep learning algorithm to predict and correct strongly varying astronomical backgrounds. LUMOS estimates the probability distribution of the background-subtracted galaxy flux, which requires the implicit estimation and subtraction of the background noise.

We have developed and tested LUMOS using images from the Physics of the Accelerating Universe Survey (PAUS). PAUS is an imaging survey that measures high precision photo- $z$  to faint magnitudes  $i_{AB} < 22.5$ , while covering a large area of sky (Martí et al. 2014). This is possible thanks to the PAUCam instrument (Castander et al. 2012; Padilla et al. 2016, 2019), a camera equipped with 40 narrow-band filters covering the optical spectrum (Casas et al. 2016). With BCNZ2 (Eriksen et al. 2019), a template-fitting algorithm, PAUS reaches a photo- $z$  precision  $\sigma(z)/(1+z) \approx 0.0035$  for the best 50 per cent of the sample, compared to typical precision of 0.05 for broad-band measurements. Similar results are obtained with DELIGHT, a hybrid template-machine-learning photometric redshift algorithm that uses Gaussian processes (Soo et al. 2021). Both methods are further improved with DEEPZ (Eriksen et al. 2020), a deep learning algorithm to measure photo- $z$ s that reduces the  $\sigma_{68}$

scatter by 50 per cent compared with the template-fitting method, having the largest improvement at fainter galaxies. Furthermore, Alarcon et al. (2021) present a high precision photo- $z$  catalogue in the COSMOS field computed using a combination of PAUS NB and 26 broad, intermediate, and NBs covering the UV, visible and near-infrared spectrum. Although LUMOS has been developed for PAUS, it can be adapted to any imaging survey, like e.g. *Euclid* or Rubin.

The structure of this paper is as follows. In Section 2, we present the PAUS data used in this paper and the simulations we use for training. Section 3 presents different flux estimation alternatives that we have used to compare to LUMOS performance. In Section 4, we introduce LUMOS, its architecture and the training procedure. Section 5 presents LUMOS results on simulations, including validation of the flux probability distributions, a comparison with alternative flux estimation methods and deblending tests. Finally, Section 6 shows LUMOS results on the PAUS data, including single exposure photometry, co-added fluxes and photometric redshifts obtained with LUMOS photometry. Conclusions and discussion can be found in Section 7.

## 2 DATA

In this section, we present PAUS data (Section 2.1) and TEAHUPOO simulations (Section 2.2), the simulated galaxy images used throughout the paper.

### 2.1 PAUS data

PAUS data are taken with the William Herschel Telescope (WHT), at the Observatorio del Roque de los Muchachos in La Palma at the Canary Islands (Spain). Images are obtained with the PAUCam instrument (Castander et al. 2012; Padilla et al. 2019), an optical camera equipped with 40 narrow band (NB) filters, covering a wavelength range from 4500 to 8500 Å (Casas et al. 2016). The NB filters have 130 Å full width half-maximum (FWHM) and a separation between consecutive bands of 100 Å. The camera has 18 red-sensitive fully depleted Hamamatsu CCD detectors (Casas et al. 2012), although only the eight central CCDs are currently used for NB imaging. Each CCD has  $4096 \times 2048$  pixels with a pixel scale of  $0.263 \text{ arcsec pix}^{-1}$ . The NB filters are mounted in five trays with eight filters per tray that can be exchanged and placed in front of the central CCDs. The NB filter set effectively measures a low-resolution spectrum ( $R \approx 50$ ).

PAUS has been observing since the 2015B semester and as of 2021A, PAUS has taken data for 160 nights. It partially covers the CFHTLS fields<sup>2</sup> W1, W2, and W3 and the COSMOS field.<sup>3</sup> Currently, PAUS data have a 40 NB coverage of  $10 \text{ deg}^2$  in W1 and W2,  $20 \text{ deg}^2$  in W3, and the  $2 \text{ deg}^2$  of the COSMOS field. The PAUS data are stored at the Port d'Informació Científica (PIC), where the data are processed and distributed (Tonello et al. 2019).

#### 2.1.1 COSMOS sample

This paper focuses on data from the COSMOS field (Lilly et al. 2009), which were taken in the semesters 2015B, 2016A, 2016B, and 2017B. These are the data used in the BKGNET paper and also in all the photo- $z$  studies published so far. The COSMOS field observations

<sup>1</sup>The code is available at <https://github.com/PAU-survey/lumos> under a GPL-3 license.

<sup>2</sup>[http://www.cfht.hawaii.edu/Science/CFHTLS\\_Y\\_WIRCam/cfhtlsdeepwidefields.html](http://www.cfht.hawaii.edu/Science/CFHTLS_Y_WIRCam/cfhtlsdeepwidefields.html)

<sup>3</sup><http://cosmos.astro.caltech.edu/>

comprise a total of 9749 images, 243 images in each NB. While observing COSMOS, the PAU camera was modified to mitigate the effect of scattered light by introducing baffles on the edges of the NB filters of each filter tray. This camera intervention changed the noise patterns of PAUS images. Half of the images in COSMOS were taken before the camera modification and the other half, after. The baseline exposure times in the COSMOS field are 70, 80, 90, 110, and 130 s from the bluest to the reddest filter trays. The complete photometry catalogue comprises 64 476 galaxies to  $i_{AB} < 23$  in 40 NB filters. This corresponds to  $\approx 12.5$  million galaxy observations ( $\approx 5$  observations per galaxy and NB filter).

### 2.1.2 Standard data reduction pipeline

PAUS data reduction process consists of two pipelines: the NIGHTLY and the MEMBA pipelines. The NIGHTLY pipeline (Serrano et al. in preparation) is the first step and performs an instrumental de-trending processing, e.g. electronic or illumination biases. Electronic biases are corrected with an overscan subtraction, an observation with the shutter closed and zero exposure time. Vignetting of the telescope corrector, pixel-to-pixel variations or dead pixels are detected and corrected with dome flats, 10-s exposures of a uniformly and homogeneously illuminated screen. The point spread function (PSF) is modelled with PSFEX (Bertin 2011) using star observations from the COSMOS Advanced Camera for Surveys (ACS, Leauthaud et al. 2007). Cosmic rays are detected and masked from the image using a Laplacian edge detector (van Dokkum 2001). The astrometry of the NB images is calibrated using SCAMP (Bertin 2006) with the *Gaia* DR2 stars (Gaia Collaboration et al. 2018) as a reference catalogue.

The MEMBA pipeline (Serrano et al. in preparation) is the second step in the data reduction. It applies forced aperture photometry to targets selected from an external detection catalogue (see Section 3.2 for more detail). In the COSMOS field, the detection parent catalogue is provided by Ilbert et al. (2008) and the photometry calibration is relative to SDSS stars (Castander et al. in preparation). A brief description of the photometry calibration can be found in Eriksen et al. (2019).

The machine-learning algorithm developed in this paper presents an alternative to MEMBA. It aims to be more robust in the presence of distorting effects as blending or scattered light. It also aims to be provide better photometry measurements of galaxies with errors in the parent catalogue galaxy profile parameters.

## 2.2 TEAHUPOO simulations

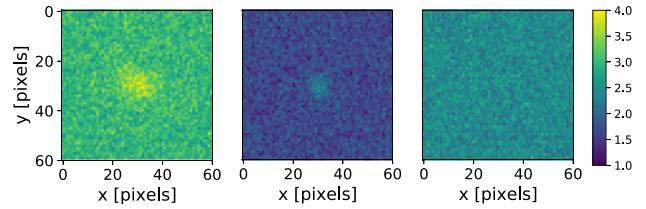
For this work, we have constructed the TEAHUPOO<sup>4</sup> simulations, a set of PAUS-like galaxy image simulations. Three examples of TEAHUPOO galaxies with  $i$ -band magnitudes  $i_{AB} = 18.5, 20.3,$  and  $22.3$  are shown in Fig. 1. Note that already at  $i_{AB} \approx 20$  it is hard to distinguish the galaxy from the background noise and with  $i_{AB} > 22$ , the galaxy signal is visually masked by background fluctuations.

TEAHUPOO light profiles are modelled with a single Sérsic profile. Each galaxy is generated using

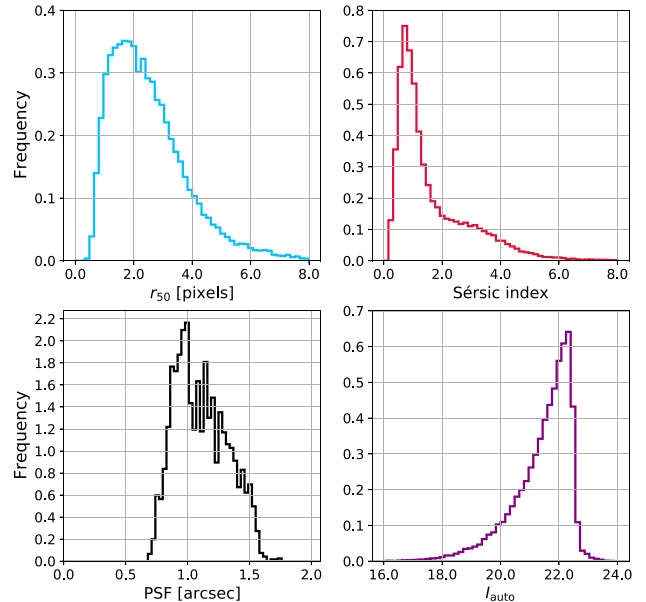
$$I(R) = I_e \exp \left\{ -(2n_s - 1/3) \left[ \left( \frac{R}{R_{50}} \right)^{1/n_s} - 1 \right] \right\}, \quad (1)$$

where  $I_e$ ,  $r_{50}$ , and  $n_s$  are the surface brightness, the half-light radius and the Sérsic index, respectively. As the shape and the size of the

<sup>4</sup>Named after the favourite sandwich of the first author.



**Figure 1.** From the left- to right-hand panel, TEAHUPOO galaxy images with  $i$ -band magnitudes 18.5, 20.3, and 22.3, in PAUS ‘NB685’. These are simulated with a exposure time of 90 s, the baseline PAUS exposure time in the ‘NB685’ filter.

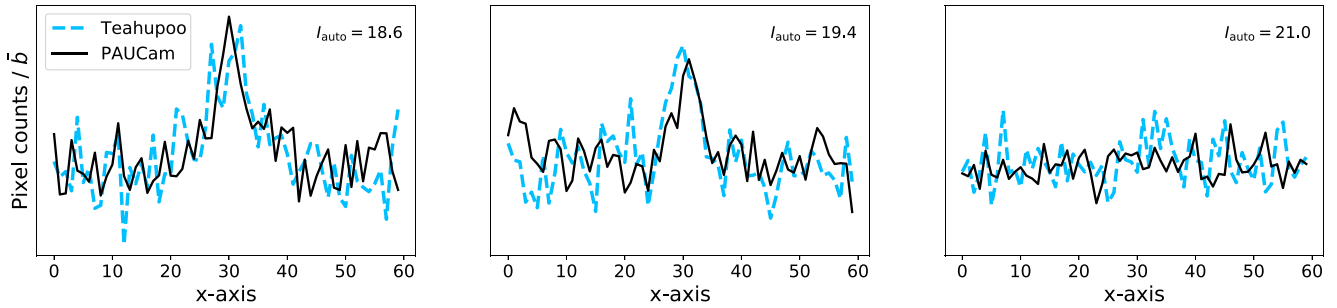


**Figure 2.** Distributions of the half-light radius ( $r_{50}$ ) (top left-hand panel), Sérsic index ( $n_s$ ) (top right-hand panel), the PSF FWHM (bottom left-hand panel) and the  $I_{\text{auto}}$  magnitude of PAUS galaxies in the COSMOS field (bottom right-hand panel).

galaxy are correlated, to ensure galaxies are realistic, we jointly sample the half-light radius, the Sérsic index and the ellipticity from their distributions in the COSMOS field. Elliptical galaxies are simulated by elongating the half-light radius according to the  $b/a$  distribution in the COSMOS field. Fig. 2 shows the distributions of  $r_{50}$  (top left-hand panel) and the Sérsic index (top right-hand panel) of PAUS galaxies in the COSMOS field, which are provided by Ilbert et al. (2009).

TEAHUPOO image simulations are  $60 \times 60$  pixels image generated with ASTROPY (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018). ASTROPY methods evaluate the galaxy profiles (equation 1) at the centre of each pixel instead of integrating along the pixel. This is problematic for small and steep galaxy profiles (high  $n_s$ ), where the flux changes significantly along the pixel. To correct for this effect, galaxies are initially drawn in a  $600 \times 600$  grid with a later size reduction. Furthermore, drawing on a larger grid allows shifting the galaxy at a sub-pixel level from the centre. Including sub-pixels shifts in TEAHUPOO galaxy images has also proven important to reduce the number of photometry outliers on real PAUCam galaxies.

TEAHUPOO images use background cutouts from real PAUCam images. These cutouts can contain artefacts such as other galaxies, cosmic rays or crosstalk. This has proven very important for our



**Figure 3.** Comparison between single pixel row light profile for pairs of PAUCam galaxies (solid black line) and TEAHUPOO simulated galaxies (dashed light blue). TEAHUPOO galaxies are constructed to exactly mimic its real PAUCam pair. The plot shows the pixel value along the central row of pixels (crossing the source) divided by the mean image background noise. Therefore, the central peak at  $x = 30$  corresponds to the galaxy light contribution. From the left-to-right-hand panel, the galaxy magnitudes are  $i_{AB} = 18.6, 19.4,$  and  $21.0$ .

network, as it learns how to make predictions when they are present (see e.g. Sections 5.3 and 6.1). The background light noise patterns across the CCD are NB dependent and changed when the PAUCam camera was modified. For this reason, the background stamps are taken from a PAUCam image observed with the same NB filter we are simulating and we track if it was observed before or after the camera intervention (see Sections 4.1). The galaxy signal is also wavelength-dependent. Consequently, we independently sample the galaxy flux from forty flux distributions, one per NB filter.

The simulated galaxy is convolved with the PSF as detected in the source image of the background cutout (Bertin 2011). The distribution of PSFs in the COSMOS field is also displayed in Fig. 2 (bottom left-hand panel). Using the same PSF for the galaxy and the background noise is crucial. Otherwise, the network could artificially learn that the galaxy has a different PSF than the background and use this to estimate the clean flux, which would not work on PAUCam data. Before combining the background cutout and the galaxy, we simulate photon shot noise on the galaxy. Note that other sources of additive noise, e.g. readout or electronic noise, are not required as the background cutout already includes them. This is another benefit of using real PAUCam background stamps, as simulating realistic noise is often hard and could easily lead to differences between simulations and data.

As simulations are constructed from real PAUS flux measurements and PAUCam background cutouts, outlier measurements in any of these two might end up represented in the TEAHUPOO images. Some examples of this could be background images with spurious effects or outlier flux measurements at the catalogue level (i.e. artificially low or high fluxes). To reduce the number of affected TEAHUPOO galaxies, the flux distribution used to create them is clipped at 0 and  $1000 \text{ e}^-/\text{s}$ . This ensures that neither negative fluxes nor artificially bright examples are represented in the image simulations. Furthermore, we have also proceeded with a visual inspection of the PAUCam images. This filters out very poor observations, but cannot deal with local effects in regions of the CCD, e.g. saturated pixels, and therefore a few outliers will still leak into the TEAHUPOO catalogue.

The methodology used to generate TEAHUPOO images is very similar to that of the BALROG simulations (Suchyta et al. 2016; Eckert et al. 2020). The main similarity between BALROG and TEAHUPOO is that both methodologies add the simulated galaxy on real survey images. In contrast, there are also differences as BALROG uses GALSIM (Rowe et al. 2015) to draw the simulated galaxies, while TEAHUPOO galaxies are built with ASTROPY. Also, TEAHUPOO galaxies are constructed in a super-resolution grid, which increases

the resolution of small objects and allows to include sub-pixel shifts from the centre of the stamp.

### 2.3 Comparison between PAUCam and TEAHUPOO galaxies

Supervised machine-learning algorithms require a training sample, i.e. a set of data with a known solution that is used to find the non-linear mapping from the input to the output of the network. Having a good and large training sample is a crucial part of the training and ideally, we would train LUMOS on a sample of PAUCam images with known photometry. However, in absence of that, we are using TEAHUPOO images for training. These simulations need to be representative of the testing data, and small differences between PAUS and TEAHUPOO galaxies can lead to a degradation of the predictions.

To test the similarity between PAUCam and TEAHUPOO galaxies, we have generated a controlled sample of TEAHUPOO-PAUCam galaxy pairs. Given a PAUS galaxy, its simulated pair is constructed with a Sérsic modelling using the same profile parameters ( $r_{50}, n_s$ ) and the same amount of light. The simulated background stamp is selected from a sourceless region in the same image as its real PAUCam pair.

Fig. 3 shows three TEAHUPOO-PAUCam galaxy pairs with  $i$ -band magnitudes of 18.6, 19.4, and 21.0, respectively. The plot shows the pixel values along the central row of the image normalized with the mean background (excluding the source) of the stamp. Therefore, pixels without galaxy light contribution should be fluctuating around unity, while pixels with higher values will be showing the galaxy light profile. In general, PAUS and TEAHUPOO galaxies fit well up to background light and shot noise fluctuations. The first plot on the left-hand side is a clear example of this. In the middle galaxy, the two galaxies also match reasonably well, however, it also exhibits a small shift between the galaxy peaks, possibly due to an astrometry inaccuracy. The right plot shows that the comparison on fainter sources is much harder, as fainter galaxies can barely be distinguished from background fluctuations (see also Fig. 1).

Several effects could bring variations between TEAHUPOO and PAUCam galaxies. For instance, inaccuracies in any step of the data reduction process as e.g. the photometric calibration, the astrometry or the PSF measurement will not be fully represented in the simulations. The PSF is a clear example, as currently a single PSF measurement is taken from each PAUCam image assuming the PSF is constant across the CCD, which could potentially yield to discrepancies between simulations and data. Additionally, if the real galaxy cannot be modelled with a single Sérsic modelling, that would

also imply a difference between the two images. These discrepancies will propagate into larger errors in the flux estimation. None the less, note that inaccuracies in the calibration, the modelling or the PSF would also affect the measurement with other flux estimation methods (e.g. aperture photometry or model fitting). Furthermore, LUMOS uses both the galaxy image and the image of the modelled profile, which allows it to provide flux uncertainties that take into account discrepancies between the galaxy and the model. It also allows detecting inaccuracies in e.g. the astrometry or the PSF, which will also be considered in the uncertainty measurement (see last paragraph in Section 6.1).

### 3 FLUX ESTIMATION METHODS

In this section, we introduce other flux estimation methodologies that we will use to compare to LUMOS performance. Particularly, we consider a profile fitting methodology (Section 3.1), aperture photometry (Section 3.2), and a linear weighted sum of the galaxy pixels (Section 3.3).

#### 3.1 Profile fitting

In a profile fitting approach, the background-subtracted galaxy image is fitted to a theoretical galaxy model to infer the profile amplitude. Assuming that the galaxy can be modelled as  $I(r) = I_e R(r)$ , where  $I_e$  is the profile amplitude and  $R(r_i)$  corresponds to a Sérsic light profile at pixel  $i$ , we can fit the image to the theoretical profile with

$$\chi^2 = \sum_i \frac{(f_i - I(r_i))^2}{\sigma_{F,i}^2}, \quad (2)$$

where  $i$  sums over pixels,  $f$  is the background subtracted flux ( $f \equiv F - B$ , with  $F$  and  $B$  being the total flux and the background noise) and  $I(r_i)$  is the galaxy theoretical model in pixel  $i$ . Assuming Poisson errors, the previous equation becomes

$$\chi^2(I_e) = \sum_i \frac{(f_i - I_e R(r_i))^2}{I_e R(r_i) + B}, \quad (3)$$

where  $B$  is the mean background per pixel. The total flux is measured as the  $I_e$  minimizing equation (3). Note that the parameter  $I_e$  appears twice in the equation, which makes the closed form not feasible. Instead, we have minimized equation (3) with a Neelder–Mead algorithm from SCIPY (Jones et al. 2001).

#### 3.2 Aperture photometry

Aperture photometry (Mighell 1999) is widely used in a large number of surveys e.g. DES (Drlica-Wagner et al. 2018) or Pan-STARRS (Magnier et al. 2020), and also in PAUS. This approach measures all the pixel contributions inside an aperture of radius  $R$  with subtraction of the background light, i.e.

$$f = \iint_R dA (S(r) - \bar{b}), \quad (4)$$

where  $S$  is the signal per pixel,  $\bar{b}$  is the mean background per pixel,  $R$  the aperture radius, and  $A$  is the aperture area.

In PAUS, the apertures are elliptical and their areas are set to target a fixed amount of galaxy light, in our case 62.5 percent of the flux. Therefore, obtaining the total flux requires scaling the measurement (equation 4) by 1/0.625. Given a target percentage of light,  $R$  is estimated using a simulated galaxy profile (Sérsic index, size, ellipticity) convolved with the image PSF. The background light

is measured as the mean of the pixel values within an annulus of  $R_{in} = 30$  pixels,  $R_{out} = 45$  pixels centred at the targeted galaxy. The pixels within the ring are  $\sigma$ -clipped beforehand, to prevent from artefacts biasing the background measurement.

#### 3.3 Weighted pixel sum

In aperture photometry, all pixels within the aperture contribute equally to the total flux, i.e. pixels at the galaxy border with low SNR and pixels at the galaxy centre contribute the same. In terms of total SNR, this is not optimal, especially for small and faint galaxies where all the signal is distributed among very few pixels. Weighting differently each pixel contribution could increase the SNR of the measurements.

There are different choices of weights, with some providing a higher SNR than others. Indeed note that aperture photometry is just a simple case of pixel weighting, where pixels within the aperture have a unity weight and those outside do not contribute.

The weighting we are interested in is that giving the most optimal unbiased linear solution. This means the unbiased estimator providing the maximum SNR, which can be written as

$$\text{SNR} = \frac{\sum_i w_i m_i}{\sqrt{\sum_i w_i^2 (m_i + b_i)}}, \quad (5)$$

where  $b_i$ ,  $m_i$ , and  $w_i$  are the background mean value, the signal mean value and the optimal weight in pixel  $i$ , respectively. Maximising the SNR (equation 5) as a function of the pixel weights  $w_i$  leads to

$$w_x = \frac{\sum_i m_i}{\sum_i m_i^2 / (m_i + b_i)} \frac{1}{1 + b_x / m_x}, \quad (6)$$

where  $w_x$  is the optimal weight of pixel  $x$  (see Appendix A for a more detailed derivation). This is the linear estimator providing the most precise unbiased flux measurements. However, obtaining the optimal weights in equation (6) requires a perfect knowledge of the galaxy light profile and consequently, uncertainties in the pixel signal and background noise degrade the precision of the flux measurements. None the less, this methodology puts a limit on how well linear methods can measure galaxy fluxes, which can be used to benchmark the LUMOS performance.

### 4 LUMOS: MEASURING FLUXES WITH A CNN

In this section, we describe LUMOS, our deep learning algorithm to measure the photometry of astronomical objects. A short introduction to the deep learning concepts needed to understand this section can be found in Appendix B. We will discuss LUMOS input data (Section 4.1), its architecture (Section 4.2) and the training procedure (Sections 4.3 and 4.4). In Sections 4.2–4.4, we describe technical details of the network’s architecture and the training procedure, which are not absolutely required to understand the main results of this work.

#### 4.1 Input data

The LUMOS input consists of two types of data. The most important input includes two images of  $60 \times 60$  pixels. The first one is made of an image cutout centred at the target galaxy and, given that PAUS pixel scale is 0.263, it covers  $16 \times 16$  arcsec of the night sky. Although most PAUS galaxies have a half-light radius between 1 and 3 pixels, which would not require such a large cutout, in Cabayol-Garcia et al. (2020), we already showed that the network needs larger stamps to accurately model the background noise fluctuations and scattered light patterns.

The second image contains the convolved galaxy profile drawn using the parameters from an external catalogue. We have tested other possibilities like, e.g. using the true galaxy profile, the PSF profile or both separately, obtaining the best results with the convolved galaxy profile. Note that in the training phase the input galaxy cutouts are TEAHUPOO image simulations (see Section 2.2), while in the testing phase these are of real galaxy cutouts.

While the network can directly estimate the photometry using only images, we have found that additional information improves the results. This information is the second type of input and it currently includes:

(i) The  $i$ -band magnitude of the target galaxy obtained from the external catalogue. This is not strictly needed as the network works without this information. However, it helps providing better photometry uncertainties and so far LUMOS requires other information (galaxy profile, coordinates) from an external catalogue anyway.

(ii) The CCD coordinates: PAUCam images contain scattered light with a band dependent spatial pattern across the CCD. This makes the CCD position and the band relevant information for the network (see figs 1 and 2 in Cabayol-Garcia et al. 2020).

(iii) The NB filter identification. The galaxy flux distribution is different for each NB filter. Furthermore, the scattered light pattern also depends on the NB filter (see fig. 2 in Cabayol-Garcia et al. 2020). Therefore, the band provides valuable extra knowledge of the expected flux and background noise pattern.

(iv) A camera intervention flag: The camera was modified while observing the COSMOS field (see Section 2). Therefore, the network also benefits from knowing if an image was taken before or after the camera intervention. This information is combined with that of the NB filter and given as an  $80 \times 10$  trainable matrix (see Section 4.2 for more details). In practice, the network effectively works for all intents and purposes as having 80 different filters instead of 40.

How these inputs are combined and given to the network is described in the next subsection.

## 4.2 LUMOS architecture

The LUMOS architecture (see Fig. 4) has two differentiated parts, a CNN and MDN. The CNN works directly on the input images and it is built up with five blocks of convolution-pooling-batch normalization (see Appendix B for a more detailed explanation). These layers are represented in Fig. 4 as orange, blue and purple stacked blocks, respectively. The CNN's output is transformed into a 1D array, including the galaxy image information and then combined with external information regarding its  $i$ -band magnitude, its position in the CCD, the NB filter it was observed with and the camera intervention flag (see Section 4.1). For the band information, the network uses a  $80 \times 10$  matrix, where each combination of band  $\times$  camera intervention flag (before/after) is represented by 10 features to be trained.

The CNN's output array combined with the galaxy information is the input of the MDN. MDNs (Bishop 1994) are a variant of neural networks that combine a feed-forward network and a mixture model. The MDN returns the probability distribution of the total flux prediction as a linear combination of  $k$  distributions. These distributions could be any sort of basis function, e.g. Gaussians or multiquadratic functions. However, in this work, we have only considered Gaussian distributions, in such a way that

$$p(f|G) = \sum_i^k \alpha_i N_i(\mu_i, \sigma_i), \quad (7)$$

where  $f$  is the galaxy flux,  $G$  is the galaxy data,  $N_i(\mu_i, \sigma_i)$  is the  $i$ th Gaussian outcome distribution, and the parameters  $\alpha_i$  are the so-called mixing coefficients. These parameters give the relative contribution of each Gaussian component to the total probability distribution. Note that the set of mixing coefficients must sum to unity and therefore they can be understood as the prior probability of each Gaussian component.

The MDN in LUMOS consists in four dense layers with parameters 5000:1000:100:15. It outputs five mixing coefficients ( $\alpha$ ) together with five pairs of  $(\mu, \sigma)$  parametrizing the Gaussian components ( $N$ ). This kind of network architecture was already used on PAUS data in Eriksen et al. (2020), where it was used to predict redshift probability distributions.

The choice of loss function is a crucial step in the construction of a neural network. LUMOS combines two loss functions, the first as in equation (8) and a modified version of it in equation (10).

A common loss function for a Gaussian MDN is

$$\mathcal{L}_{\text{MDN}} = \sum_{i=1}^k \left[ \log(\alpha_i) - \frac{(f_i - \bar{f})^2}{\sigma_i^2} - 2 \log(\sigma_i) \right], \quad (8)$$

which corresponds to the negative log likelihood of a Gaussian distribution with mean  $\bar{f}$  and variance  $\sigma_i^2$ .

In LUMOS, we have also considered a variation of the previous loss motivated by the fact that physical galaxies have positive flux values. This alternative loss also corresponds to the Gaussian negative log-likelihood, but integrated from 0 to  $\infty$ , which changes the PDF normalization.

The logarithm of a Gaussian PDF ( $G(x)$ ) integrated from 0 to  $\infty$  is

$$\begin{aligned} \log(G(x)) = & \log\left(\frac{2}{\pi}\right) + \log\left(\exp\left(-\frac{\frac{1}{2}(x - \mu)^2}{\sigma^2}\right)\right) \\ & + \log\left(\sigma \text{Erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + \sigma\right), \end{aligned} \quad (9)$$

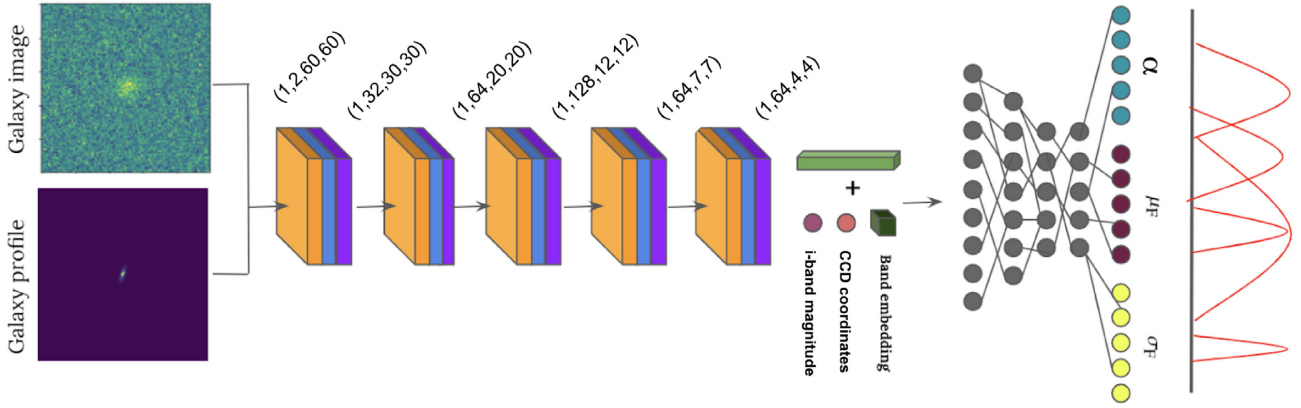
which leads to the following loss function

$$\begin{aligned} \mathcal{L}_{\text{MDN}} = & \sum_{i=1}^k \log(\alpha_i) - \frac{(f_i - \bar{f})^2}{\sigma_i^2} - 2 \log(\sigma_i) \\ & - \log\left(\text{Erf}\left(\frac{\bar{f}}{\sqrt{2}\sigma_i}\right) + 1\right), \end{aligned} \quad (10)$$

where again  $\bar{f}$  is the true flux and  $\sigma_i$  is the flux uncertainty. Therefore, the truncation of the Gaussian distribution effectively corresponds to an additional term in the loss function. LUMOS combines both loss functions and uses equation (8) on objects with  $i_{\text{AB}} < 20.5$  and equation (10) for the rest of the sample. Choosing a single loss function for all magnitudes also works, however, combining the two improves the photometry for the faintest and the brightest galaxies.

## 4.3 Unsupervised transfer learning

Transfer learning (Tan et al. 2018; Zhuang et al. 2019) is a deep learning technique that aims to adapt a model trained to make predictions on a particular task to work on a similar but different problem. One example is a classifier trained to distinguish between cats and dogs adapted to distinguish horses and zebras. Instead of training from scratch, the zebra-horse classifier takes the parameters of the cat-dog classifier as initial parameters, in such a way that the network has already learnt to extract shared features like e.g. detect edges or shapes. Such mutual features are commonly extracted in the shallower layers of the network, while deeper layers pick up more



**Figure 4.** LUMOS architecture: Orange cubes correspond to convolutional layers, blue cubes are pooling layers and the purple ones are batch normalization layers. The vectors between two CNN blocks corresponds to the dimension of the previous block’s output and the input dimension of the following. The CNN’s output is linearized (green stick), combined with external information (coordinates, NB filter, and  $i$ -band magnitude) and input to an MDN. The MDN outputs a probability distribution of the total flux as a linear combination of five Gaussians.

subtle data traits. For this reason, many times transfer learning only requires training deeper layers of the network, while shallower layers can be shared among the different networks.

The same idea can be applied to adapt models trained on simulations to perform well on data (Tercan et al. 2018; Eriksen et al. 2020). To train a supervised network one needs data with a known solution (labelled data). Many times there are labelled data available, but not enough to train a network from scratch. A possible solution in such cases is to train the network on simulations and use a small labelled data set to adapt the model to the data. This requires two consecutive trainings, one initial on simulations and once this finishes, an additional one on data with the network parameters from the training on simulations as a starting point. Domínguez Sánchez et al. (2019) also explores the possibility of using transfer learning to adapt a model trained on data from one survey to another one. Particularly, it adapts a morphology classifier trained on SDSS images to work on DES images.

LUMOS is trained on TEAHUPOO image simulations (see Section 2 for more details) and we cannot apply supervised transfer learning as there are no data with known photometry. Instead, we will use the compatibility of independent observations of the same galaxy in the same NB filter to apply what we call *unsupervised transfer learning*.

For that, we have collected a set of PAUCam observed galaxy pairs, with two independent images of the same galaxy observed with the same NB filter. LUMOS should predict compatible flux PDFs for the two observations of the same object, learning to ignore differences in e.g. the background noise or the PSF. Therefore, after training LUMOS on TEAHUPOO image simulations, we retrain it on the set of PAUCam galaxy pairs, forcing compatibility between the two flux measurements. With this procedure, we make sure that LUMOS has seen real data before evaluating the network on the test sample.

The unsupervised transfer learning loss function is constructed comparing the probability distribution of two observations. Before comparing the PDFs, these need to be calibrated with the image zero-point (see Section 2.1 for more details). The PDFs are parametrized with five Gaussian distributions (as provided by LUMOS, see Section 4.2), in such a way that each Gaussian component in the first observation is compared to all the components in the second observation. The negative log likelihood

of the difference between the two predicted PDFs takes the form of

$$\begin{aligned} \mathcal{L}_{\text{UTL}} = & \sum_i \sum_j \log \alpha_i + \log \beta_j - \frac{1}{2} \frac{(f_i - f_j)^2}{\sigma_i^2 + \sigma_j^2} \\ & - \sqrt{\sigma_{f_i}^2 + \sigma_{f_j}^2} - \log \left[ 1 + \text{Erf} \left( \frac{f_i}{\sqrt{2}\sigma_i} \right) \right] \\ & - \log \left[ 1 + \text{Erf} \left( \frac{f_j}{\sqrt{2}\sigma_j} \right) \right] \\ & + \log \left[ 1 + \text{Erf} \left( \frac{f_i/\sigma_i^2 + f_j/\sigma_j^2}{\sqrt{1/\sigma_i^2 + 1/\sigma_j^2}} \right) \right], \end{aligned} \quad (11)$$

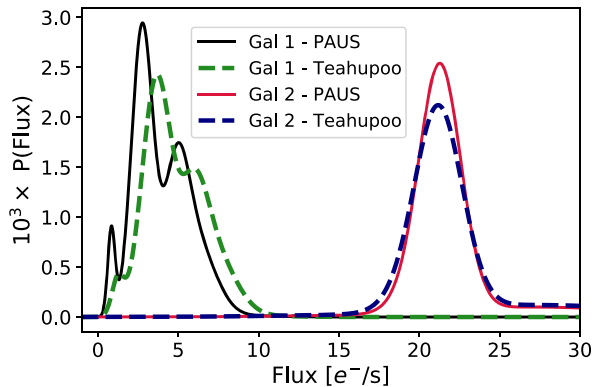
where  $f_i$  ( $f_j$ ) is the  $i$ th ( $j$ th) Gaussian component of the first (second) exposure, and similarly for  $\sigma$ .

#### 4.4 Training procedure

LUMOS is trained with 500 000 simulated images constructed combining simulated galaxies and real PAUCam backgrounds (see Section 2.2). The sample is split into 90 percent for training and 10 percent for validation. We also generate 10 000 independent simulated galaxies for testing. LUMOS is trained for 100 epochs with an initial learning rate of  $10^{-4}$ , which is reduced to a factor of 10 every 40 epochs. We use ADAM (Kingma & Ba 2015) as optimization algorithm. The training takes about 20 h using an NVIDIA TITAN V GPU.

The network is trained with a batch size of 500 galaxies, which is a relatively large batch size. As simulations are constructed from real PAUS flux measurements and PAUCam background cutouts, TEAHUPOO image simulations might contain some outliers. We have applied a filtering to reduce them (see the last paragraph in Section 2.2), however, a few poor examples can still be part of TEAHUPOO images and these are very difficult to detect. Having a large batch size reduces their effect on the overall loss function and consequently in the training.

After the supervised training, we apply the unsupervised transfer learning. This part is trained with 20 000 galaxy pairs. A danger of applying unsupervised transfer learning is that nothing prevents the



**Figure 5.** Flux probability distributions provided by Lumos for two PAUCam galaxies (Gal1, Gal2, solid lines) and their TEAHUPOO imitations (dashed lines).

network from biasing the flux predictions, since equation (11) does not directly constrain the flux prediction but the pairwise consistency. For this reason, we have also included a supervised training with 20 000 simulations. The loss function ends up combining equations (8), (10), and (11) and therefore becomes

$$\mathcal{L} = \mathcal{L}_{\text{MDN}} + \mathcal{L}_{\text{UTL}}. \quad (12)$$

These two losses are weighted equally in the loss function. Nevertheless, this is not strictly necessary and one could consider weighting them differently.

LUMOS has already been trained on reliable simulations, therefore its parameters should be close to optimal before the unsupervised transfer learning training. Consequently, instead of training all the network parameters, the transfer learning only varies the parameters in the last linear layer, while all the others are frozen. The unsupervised transfer learning phase is trained for 100 epochs, with an initial learning rate of  $10^{-5}$ , which is reduced by a factor of 10 every 50 epochs.

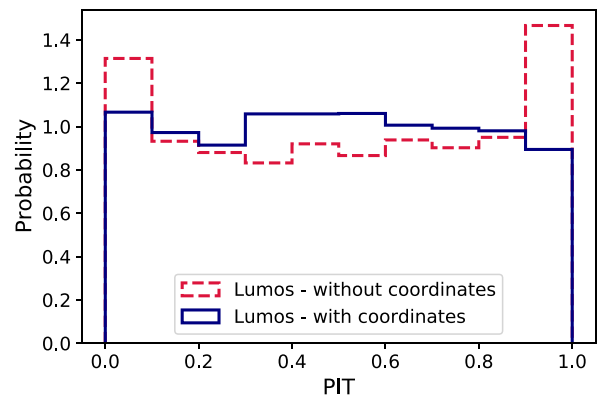
## 5 LUMOS FLUX MEASUREMENTS ON SIMULATIONS

In this section, we will test LUMOS on TEAHUPOO galaxies. We will first validate the flux probability distributions (Section 5.1), followed by a comparison with other flux estimation methods (Section 5.2). Finally, we will test how well LUMOS performs on blended galaxies (Section 5.3).

### 5.1 Flux probability distributions

Most photometry algorithms only provide a flux measurement and its uncertainty. In contrast, LUMOS provides the flux probability distributions as a linear combination of five Gaussians (see Section 4.2). Fig. 5 shows the predicted flux PDFs for two PAUCam galaxies (solid lines) and two TEAHUPOO galaxies mimicking them (dashed lines) (see Section 2.3). The flux probability distributions on data and on the simulations are very similar, providing additional confidence on the reliability of TEAHUPOO galaxies.

Also, notice that the resulting PDFs are not Gaussian. This can especially be seen in the faintest galaxy, which displays secondary peaks on the left and right of the main one. This type of PDF is common in LUMOS predictions, where fainter galaxies exhibit more non-Gaussianities than brighter ones. In general, LUMOS PDFs are



**Figure 6.** PIT distribution of the LUMOS flux PDFs on a set of 10 000 TEAHUPOO galaxies. We have tested LUMOS PDFs with (solid blue line) and without (red dashed line) including the CCD coordinates. If the CCD coordinates are not included, LUMOS provides outliers (peaks at 0 and 1) corresponding to scattered light affected objects.

more Gaussian at redder bands, where galaxies are also brighter. At the blue end, many PAUS galaxies have fluxes very close to 0 for which LUMOS commonly provides very non-Gaussian PDFs (see Section 6.3 for further discussion).

#### 5.1.1 Probability integral transform (PIT) on simulations

PIT (Dawid 1984; Gneiting et al. 2005; Bordoloi, Lilly & Amara 2010) tests the quality of the probability distribution and it is defined by

$$\text{PIT} \equiv \int_{-\infty}^{f^*} df \phi(f), \quad (13)$$

where  $f^*$  is the true flux value and  $\phi(f)$  is the probability distribution. When  $\phi(f)$  faithfully represents the true value, the PIT distribution is the uniform distribution  $U[0,1]$ .

In Fig. 6, we have estimated the PIT value for 10 000 TEAHUPOO galaxies with known true flux. The plot shows two distributions, one including (solid blue line) and another not (dashed red line) including the CCD coordinates of galaxies in the training and test set. When the training does not include the coordinates, the PIT distribution displays two peaks of outliers at the first and last bin of the plot. These outliers correspond to galaxies with strongly varying background light that require accurate knowledge of the background noise patterns in the different NB filters.

In contrast, when the training includes the CCD coordinates, the PIT test displays a flat  $U[0,1]$ , showing that LUMOS provides robust flux probability distributions and that CCD coordinates are essential information for cutouts with varying backgrounds. This is consistent with the results in fig. 7 of Cabayol-Garcia et al. (2020), where the CCD coordinates proved essential to predict accurate backgrounds (solid black line).

#### 5.1.2 Single flux and flux uncertainty measurements

Although the flux PDF provides more information, many applications require a single flux measurement and uncertainty. Flux point-like estimates can be computed with different statistical estimators as the mean, the median or the peak. For (almost) Gaussian PDFs, these estimators provide very similar flux measurements. However,



when the PDFs move away from gaussianity, these estimators can provide significant differences among them. As an example, in LUMOS multiple peaked distributions tend to provide higher flux measurements with the median than with the peak. This is because these PDFs commonly belong to faint objects with the main peak very close to zero and secondary peaks and the tail moving towards higher fluxes (see Fig. 17 for an example).

For the flux uncertainty, the most straightforward estimator is the standard deviation. However, another possibility is

$$\sigma_{68} \equiv \frac{1}{2}[q_{84} - q_{16}], \quad (14)$$

where  $q_{16}$  ( $q_{84}$ ) are the 16th (84th) quantiles. For Gaussian distributions, these two estimators coincide. However, in the case of non-Gaussian PDFs,  $\sigma_{68}$  is more robust towards distributions with tails but provides higher uncertainties in the presence of multiple peaks (see Section 6.3 for further discussion).

Even though the median or  $\sigma_{68}$  are more robust with noisy PDFs, they require the explicit PDF construction from the Gaussian parametrisation provided by LUMOS. This is time-consuming, since we already have more than 10 million galaxy exposures in the small COSMOS field. A fast alternative is to analytically determine the mean and the variance from the Gaussian component parameters. The mean flux ( $f$ ) is estimated as

$$f = \sum_i \alpha_i \Delta \mu_i, \quad (15)$$

where  $\alpha_i$  and  $\mu_i$  are the mixing coefficient and the expected value of the  $i$ th Gaussian component, respectively. The associated variance ( $\sigma_f^2$ ) is then given by

$$\sigma_f^2 = \sum_i \left[ \alpha_i \left( \sigma_i^2 + \|\mu_i - \sum_j \alpha_j \mu_j\|^2 \right) \right], \quad (16)$$

where  $\sigma_i^2$  is the variance of the  $i$ th Gaussian component.

## 5.2 Comparison with different flux estimation methods

While LUMOS has proven to provide reliable flux probability distributions, other methodologies as model fitting or aperture photometry are also able to provide accurate flux estimates (e.g. Lang et al. 2016; Drlica-Wagner et al. 2018; Kuijken et al. 2019). In this section, we will use simulations to compare the performance of LUMOS with a profile fitting method (Section 3.1), aperture photometry (Section 3.2) and a linear weighted sum of pixels (Section 3.3). To quantify the quality of the flux measurements, we will use

$$\text{Bias} : \text{Median} [(f - f_0)/f_0], \quad (17)$$

$$\text{Dispersion} : \sigma_{68} [(f - f_0)/f_0], \quad (18)$$

where  $f_0$  is the ground truth flux.

Fig. 7 compares the bias (left-hand panel) and the dispersion (right-hand panel) in the flux predictions as a function of the  $I_{\text{auto}}$  magnitude for the four methods. The model-fitting method (purple dashed-dotted line) displays a systematic increment of the bias with magnitude, with a 20 per cent bias at the faint end. PAUS galaxies are already hard to distinguish from background fluctuations for magnitudes  $i_{\text{AB}} > 20$  (see Fig. 2), which could be severely complicate the fitting at fainter magnitudes.

The second method is the pixel weighted sum. It is unbiased for objects with  $i < 21$ , but fainter objects are 10 per cent biased. While the optimal weights (equation 6) ensure unbiased flux estimates, these

also require perfect knowledge of the galaxy profile and the background light. On simulations, the galaxy light distribution is known, however, the background light is not. Therefore, at the faint end where the galaxy signal is comparable to the background fluctuations, the weights (equation 6) seem to be very sensitive to inaccuracies in the background noise. In contrast, aperture photometry and LUMOS provide unbiased estimates to a 5 per cent level up to magnitude 22.

In terms of dispersion, LUMOS is the most precise method with  $\sigma_{68} = 0.36$ . This implies a 28 per cent improvement with respect to the linear pixel weighting method, which is the second-best method with  $\sigma_{68} = 0.47$ . As expected, the optimal weighted sum is the most precise unbiased linear method, but it degrades at fainter magnitudes. LUMOS overcomes the linear optimal weighting with a non-linear mapping and provides good flux estimates at all magnitudes.

The previous results combine measurements from all NB filters. Considering measurements from each NB filter independently, bluer bands show a higher dispersion, which is expected since their SNR is lower. Furthermore, the flux measurements are unbiased to a 3 per cent level in all NB filters and these also show unbiased for all galaxy sizes ( $r_{50}$  and PSFs).

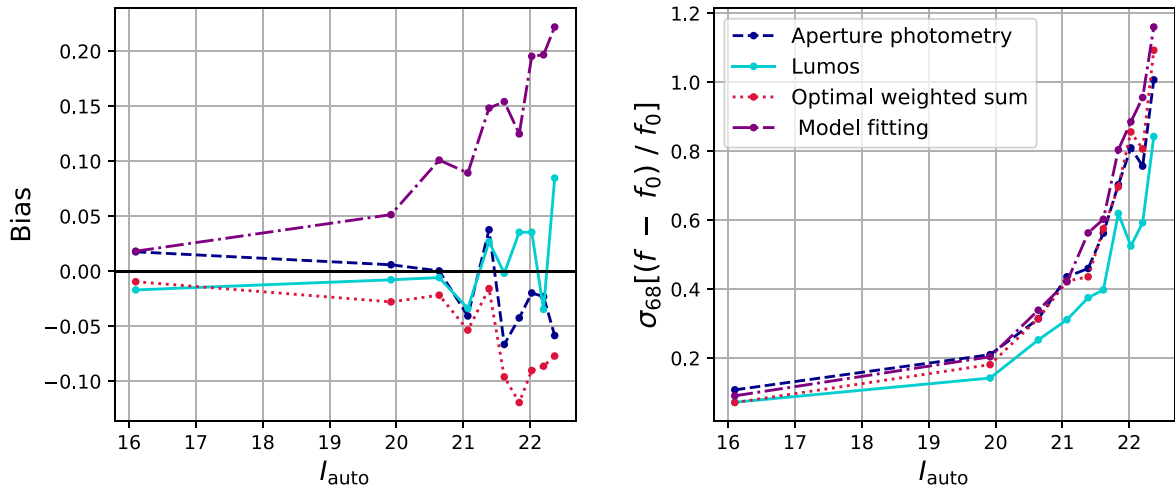
## 5.3 Deblending with LUMOS

Blending is the superposition of galaxies with other astrophysical objects along the line of sight. It affects the photometric and shape measurements contributing to systematics in weak-lensing studies (Arcelin et al. 2021). Deblending will be a challenge for future ground-based photometric surveys such as Rubin (Ivezić et al. 2019) or *Euclid* (Laureijs et al. 2011) and it has recently been approached with deep learning techniques (e.g. Arcelin et al. (e.g. Boucaud et al. 2020; Arcelin et al. 2021)

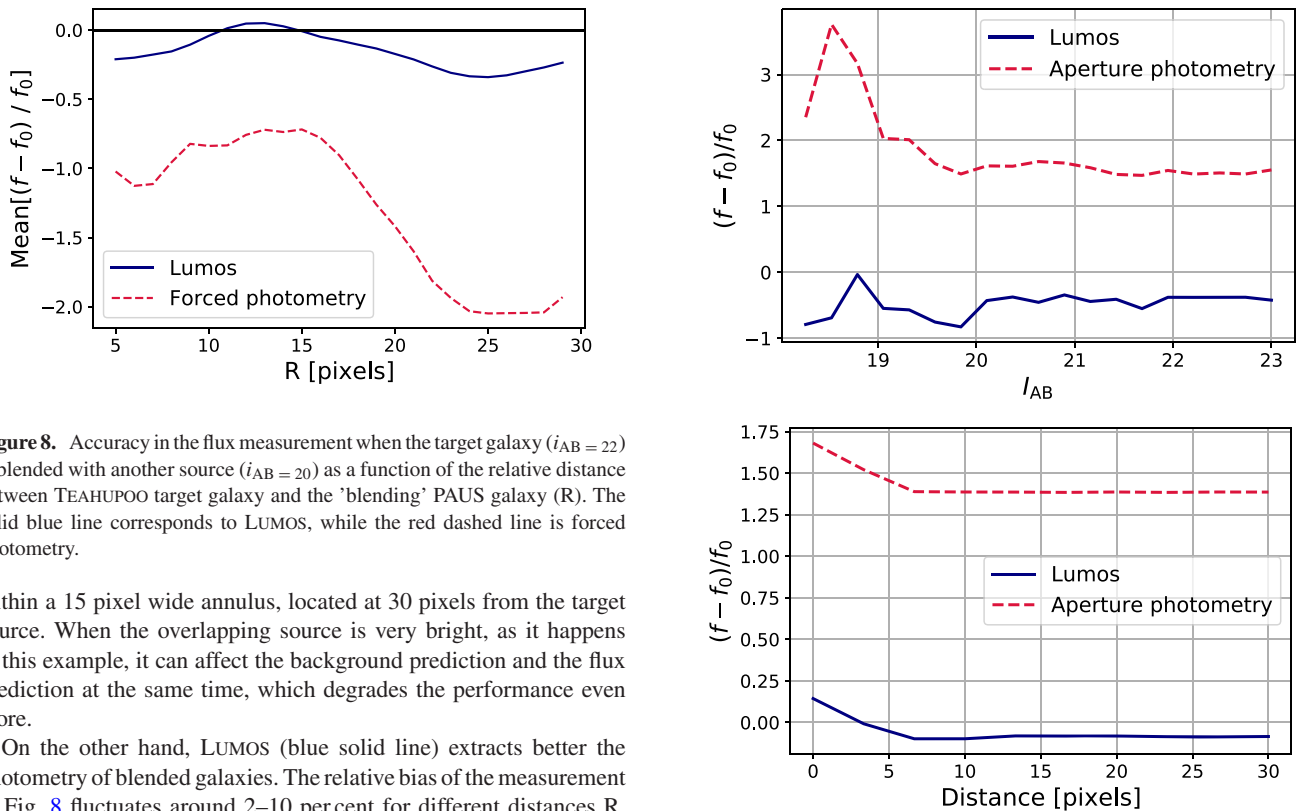
In this section, we test if LUMOS can extract the galaxy photometry when the target galaxy is blended with another object. Even though LUMOS is not explicitly constructed to predict the photometry in the presence of other galaxies, the simulated training sample is built using background cutouts from PAUCam images (see Section 2). These image cutouts are centred at random CCD positions, without minding if there are any other artefacts nearby. Consequently, the training sample contains examples of blended galaxies to learn from. This is one benefit of machine-learning algorithms. These are flexible enough to learn how to extract the photometry of blended sources by only including examples in the training sample, without explicitly constructing the algorithm for this task.

We have generated 3600  $60 \times 60$  pixel realizations of the same target TEAHUPOO galaxy, which is placed at the central pixel of the stamp. Each of these realizations also contains the same PAUS galaxy centred on a different pixel at a time. The PAUS galaxy moves across the stamp in steps of one pixel, in such a way that the realization where it is located at the central pixel corresponds to a total blending with the TEAHUPOO galaxy. At the end of the day, with all the realization the PAUS galaxy covers all the pixels in the stamp.

Fig. 8 shows the accuracy in the flux measurement as a function of distance to the overlapping source. The target galaxy is fainter than the overlapping one, with  $i_{\text{AB}} = 22$  and  $i_{\text{AB}} = 20$ , respectively. With aperture photometry (dashed red line), the flux is considerably biased for all distances  $R$ . This is expected since summing all contributions within the aperture does not differentiate if the light belongs to the target source or the overlapping one. Therefore, for  $R < 15$ , the bias measurement is caused by light from the overlapping source accounted inside the aperture. At larger values of  $R$ , the background noise prediction is also affected by the overlapping source. In the PAUS aperture photometry pipeline, the background is estimated



**Figure 7.** Comparison of the bias (left-hand panel) and the dispersion (right-hand panel) among the flux measurements with LUMOS, aperture photometry, model fitting and optimal pixel weighting. These results include galaxy image simulations in the 40 PAUS NB filters. The I-band magnitude corresponds to the AUTO magnitude as measured by the *HST*–ACS on the COSMOS field.



**Figure 8.** Accuracy in the flux measurement when the target galaxy ( $i_{\text{AB}} = 22$ ) is blended with another source ( $i_{\text{AB}} = 20$ ) as a function of the relative distance between TEAHUPOO target galaxy and the ‘blending’ PAUS galaxy (R). The solid blue line corresponds to LUMOS, while the red dashed line is forced photometry.

within a 15 pixel wide annulus, located at 30 pixels from the target source. When the overlapping source is very bright, as it happens in this example, it can affect the background prediction and the flux prediction at the same time, which degrades the performance even more.

On the other hand, LUMOS (blue solid line) extracts better the photometry of blended galaxies. The relative bias of the measurement in Fig. 8 fluctuates around 2–10 per cent for different distances R. Unlike aperture photometry, LUMOS is able to distinguish between the two galaxies and consider the overlapping one a source of noise.

In Fig. 9, we have explored more the LUMOS deblending capability as a function of the magnitude and distance to the overlapping galaxy. For the top plot, we have simulated 20 image cutouts with an  $i_{\text{AB}} \approx 21$  galaxy at the centre (the same for the twenty realizations). In each of the cutouts, we have included a second source always located at five pixels from the centre, but which varies brightness among realizations. The plot shows the photometry accuracy with LUMOS (blue solid line) and aperture photometry (red dashed line) for different magnitudes of the overlapping sources. For this study, we

**Figure 9.** Accuracy in the flux predictions in the presence of overlapping sources for LUMOS (solid blue line) and aperture photometry (dashed red line) as a function of Top panel: magnitude of the overlapping source. Bottom panel: distance in pixels between the target and the overlapping source.

have also applied a  $\sigma$ -clipping of the annulus so that the background measurement in the aperture photometry is more robust. As indicated in the previous test, LUMOS shows more robust towards overlapping nearby sources, even when these are bright. In all cases, aperture photometry provides biased measurements due to the proximity

of the source, which is always accounted within the aperture pixels.

For the bottom plot, we have proceeded similarly. We have generated ten realizations of the same galaxy with magnitude  $i_{AB} \approx 21$  and in each of the cutouts, we have included a second source with  $i_{AB} \approx 22$ , but located at a different distance from the target source. The plot exhibits the photometry accuracy as a function of the distance between the target and the overlapping source. Again, the plot shows much better accuracy for LUMOS than aperture photometry.

## 6 LUMOS PHOTOMETRY ON PAUS DATA

In this section, we present the photometry extracted from PAUCam images in the COSMOS field with LUMOS. First, we show single observation measurements (Section 6.1) and compare them to SDSS measurements (Section 6.2). We then discuss the co-added flux measurements (Section 6.3) and show the photometric redshift results with LUMOS photometry (Section 6.4).

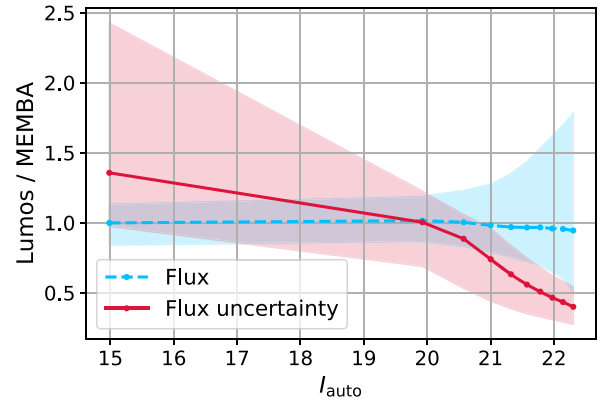
### 6.1 Single exposure measurements

PAUS has taken around 10000 images in the COSMOS field (Section 2.1), which contain ten million galaxy observations. In this section, we will only show the results on the spectroscopic sample with  $i_{AB} < 22.5$ , which are  $\approx 3$  million exposures from 15000 galaxies. For the targeting, we use the Ilbert et al. (2008) catalogue, the same external catalogue that MEMBA (the PAUS aperture photometry algorithm, see Section 2.1) uses. Given a PAUCam image, LUMOS matches it with the detection catalogue and creates the cutouts around the sources. Each source is also matched to the external catalogue, which provides a value for the half-light radius, the ellipticity and the Sérsic index. From these values, LUMOS generates the modelled galaxy profiles (see Section 4.1). Then, it provides the flux probability distribution of the observed galaxies.

#### 6.1.1 Flux and flux error measurements

In this section, we will use flux and flux error point estimates calculated as the mean and the variance of the LUMOS flux PDFs (see Section 5.1), which corresponds to the total galaxy flux. For MEMBA, we will use flux measurements from aperture photometry using the background subtraction from BKGNET, which has proven more accurate than that estimated with an annulus. Fig. 10 shows the flux (dashed blue line) and flux error (solid red line) ratios between LUMOS and MEMBA photometry in all NB filters. The shaded areas correspond to the 16th and 84th quantiles. For the full sample, the flux ratio between the two photometries is 0.99. In magnitude bins, this ratio oscillates between 0.95 and 1.02, with LUMOS measuring  $\approx 4$  per cent less flux in the brightest ( $i_{AB} < 18$ ) and the faintest ( $i_{AB} > 22$ ) bins. At the faintest end, the spread in the flux ratio increases, which is natural since these galaxies are noisier. Studying each NB filter independently, all the ratios but those from the three bluest bands ('NB455', 'NB465', and 'NB475') oscillate between 0.95 and 1.03. The three bluest bands display a  $\approx 0.9$  ratio between MEMBA and LUMOS, which is attributed to very faint galaxies with negative flux measurements in MEMBA, which are not allowed in LUMOS. MEMBA has proven accurate enough to obtain very precise phototzs, therefore measuring similar fluxes with MEMBA and LUMOS is a good first test.

Altogether, LUMOS also provides 40 per cent lower flux uncertainties than MEMBA displaying a lower error for 85 per cent of



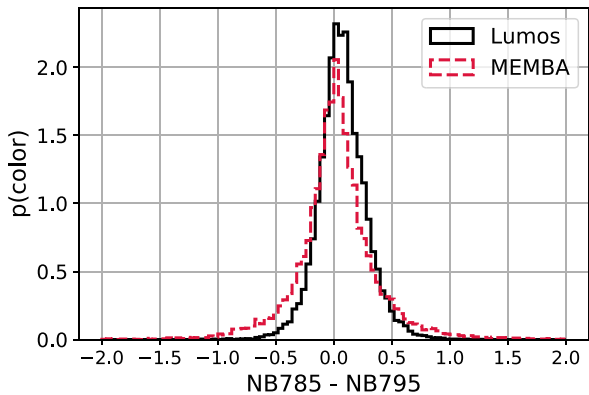
**Figure 10.** Flux and flux uncertainty ratios between LUMOS and MEMBA photometry in equally populated magnitude bins. The shaded areas correspond to the 16th and 84th quantiles.

the measurements. The ratio between LUMOS and MEMBA flux uncertainties (Fig. 10) is not constant with magnitude. LUMOS shows 60 per cent lower errors for objects with  $i_{AB} > 22$ . This number monotonically decreases to e.g. 30 per cent at  $i_{AB} = 21$  or 10 per cent at  $i_{AB} = 20.5$ , while some of the brightest objects display lower errors with MEMBA. At the brightest end, one can note a large scatter of the error ratio. This is attributed to some very bright galaxies with a significantly large error with LUMOS. However, note that aperture photometry provides a purely statistical error, while in LUMOS any additional source of uncertainty, e.g. an artefact in the cutout, inaccuracies in the profile parameters used to infer the photometry ( $n_s$ ,  $r_{50}$ ) or data reduction issues underrepresented in the training simulations will be also accounted in the error estimate. As LUMOS is provided with both the galaxy image and the model, it is able to capture discrepancies or potential sources of inaccuracies and account for them in the flux uncertainty. Following this line of study, inaccuracies in the profile parameters would most likely have a larger impact on the photometry of large, bright and resolved galaxies, where e.g. slightly underestimating  $r_{50}$  could easily lead to a quite biased flux measurement. This could also explain the increment of spread in the uncertainty ratio at the brightest end.

#### 6.1.2 Colour histograms

Assuming that galaxies have an underlying distribution of colours, the width of the colour histograms is an estimation of the uncertainty in the photometry measurements. The intrinsic width of the colour histogram is broadened by photometry uncertainties and consequently, the photometry providing narrower colour histograms is that with lowest uncertainties. Using colour histograms to compare photometries was used in Wright et al. (2016), where they presented and applied LAMBDA to improve the Galaxy and Mass Assembly (GAMA, Driver et al. 2011) photometry.

Fig. 11 shows the NB785-NB795 colour distribution (more colour histograms can be found in Appendix E). By eye it can be already noted that LUMOS provides a narrower colour distribution than MEMBA. We have estimated the width of such colour histograms with  $\sigma_{68}$  (equation 14) and  $\sigma_{95}$  (equivalent to  $\sigma_{68}$  but considering the 2.5 and 97.5 quantiles, i.e. the width accounts for 95 per cent of the data). Concretely, MEMBA provides  $\sigma_{68} = 0.26$ , while LUMOS results in  $\sigma_{68} = 0.19$ , which corresponds to a 30 per cent lower



**Figure 11.** MEMBA (red dashed line) and LUMOS (solid black line) colour histogram for NB colour (NB785-NB795). Note that the photometry with lower uncertainties is that displaying a narrower colour histogram, which in this case is the LUMOS photometry.

effective width. Considering  $\sigma_{95}$ , LUMOS reduces the width a factor of  $\approx 3$ , from 0.74 to 0.41. This also suggests that LUMOS reduced the number of photometry outliers, which are not affecting  $\sigma_{68}$  but enlarge  $\sigma_{95}$ . Such photometric outliers are located asymmetrically on the tails of the distribution, which triggers the skewness of the histograms and therefore a shift in the median of the MEMBA histogram with respect to that of LUMOS. This can be noted in the NB785-NB795 colour histogram, but also in other colour histograms in Fig. E1.

All the other NBs also show narrower colour histograms with LUMOS (see Appendix E for more details). Furthermore, the relative difference in  $\sigma_{95}$  is systematically higher than with  $\sigma_{68}$ . This is likely related with exposures with noisy photometry and outliers, which lay in the tails of the colour histograms (see Appendix E for more details).

### 6.1.3 Validation of the flux uncertainties

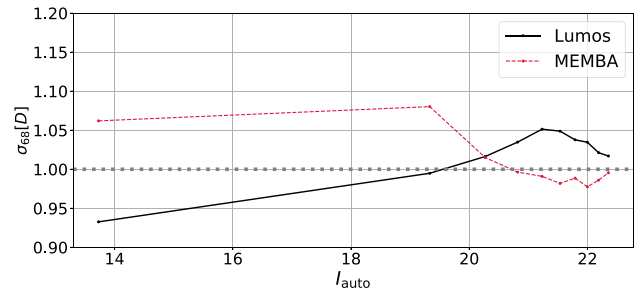
To test LUMOS flux uncertainties and ensure that these are not artificially low, we have made use of PAUS taking multiple observations of the same galaxy in the same NB filter. Given two observations of the same galaxy in the same NB, their flux measurements and uncertainties must be compatible. This is formulated as

$$D \equiv \frac{(f_1 - f_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)}}, \quad (19)$$

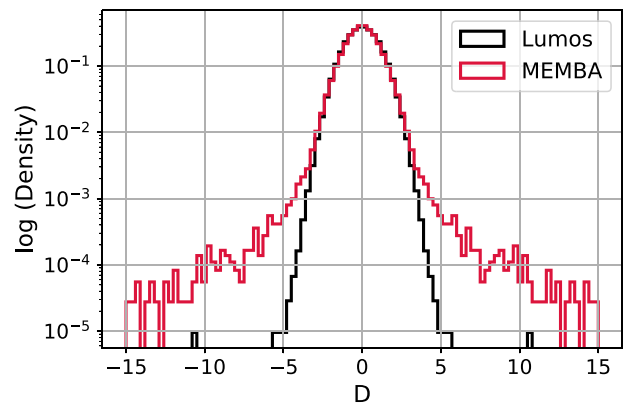
where  $f_1$  and  $f_2$  are the flux estimates of two exposures of the same object and  $\sigma_1$  and  $\sigma_2$  are their associated uncertainties.

Fig. 12 shows the width of the  $D$  distribution in equally populated magnitude bins. If the photometry uncertainties are properly accounted for, the distribution of  $D$  (equation 19) should be a Gaussian with unit standard deviation. To be less affected by outliers, we have estimated the width of  $D$  with  $\sigma_{68}$  (equation 14) and both LUMOS and MEMBA display a quite constant unity  $\sigma_{68}[D]$  along the tested magnitude range (solid black and red dashed lines, respectively). In the case of MEMBA, the background estimation with aperture photometry was providing 20 per cent underestimated errors at the bright end. This trend was fixed using BKGNET (see fig. 11 in Cabayol-Garcia et al. 2020).

Fig. 13 shows the distribution of the quantity defined in equation (19) with LUMOS (solid black) and MEMBA (dashed red)



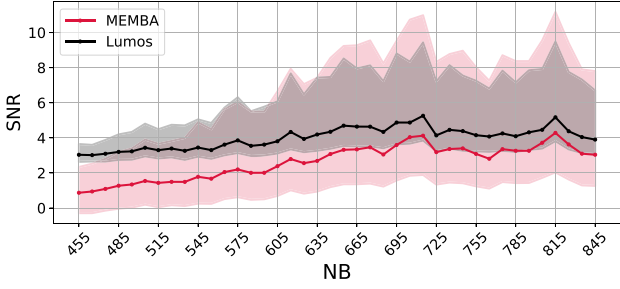
**Figure 12.** The width of the distribution in equation (19) for LUMOS and MEMBA flux predictions in equally populated magnitude bins. Robust uncertainties must provide a unity width (marked by the thick grey dotted line).



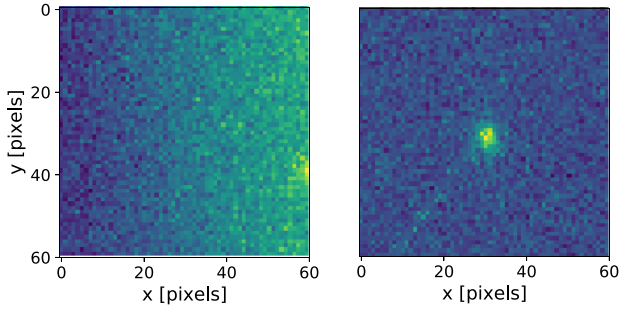
**Figure 13.** Distribution of  $D$  (equation 19) estimated with LUMOS (solid black) and MEMBA (dashed red) photometries in logarithmic scale.

photometries. As expected, both of them fit a Gaussian with zero mean and unit variance, however, we can note a tail of outliers in the MEMBA photometry not present with LUMOS. This can be connected to LUMOS providing not purely statistical uncertainties. While inaccuracies in the profile parameters or contaminating effects at the image level are not considered in MEMBA flux errors, LUMOS is flexible enough to provide an error estimate that already takes into account these effects. As an example, if the parent catalogue provides a 10 per cent underestimated  $r_{50}$  for a particular galaxy, with aperture photometry the aperture size and consequently the flux measurement will be also underestimated. However, the flux uncertainty will only account for the statistical variation in the pixels within the aperture, while the error in the galaxy profile will not be considered. In contrast, LUMOS is provided with the galaxy and the galaxy modelled image and therefore, differences between these two are captured and accounted for in the flux uncertainty.

Fig. 14 compares the median SNR per NB in MEMBA and LUMOS photometries for galaxies with  $i_{AB} < 22.5$ . The shaded areas correspond to the 16th and 84th quantiles of the SNR distribution. For the complete photometry catalogue, on average LUMOS provides a 54 per cent higher SNR. Furthermore, it gives a higher median SNR at all wavelengths, although the increment with respect to MEMBA is higher in bluer bands. For galaxies with  $i_{AB} > 22$ , the SNR is 2.5 times higher in LUMOS. The ratio increases to 3 taking into account only the bluest NB (‘NB455’) and decreases to a factor of 2 for the reddest one (‘NB845’). This is natural considering that LUMOS gives the greatest improvement in terms of SNR for faint



**Figure 14.** Median SNR per NB filter with LUMOS and MEMBA flux measurements. Shaded areas are the 16th and 84th quantiles of the SNR distribution.



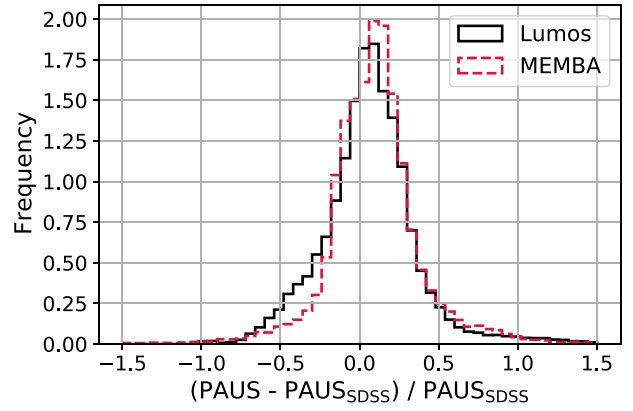
**Figure 15.** Observations affected by scattered light (left-hand panel) and cosmic rays (right-hand panel). While MEMBA provides outlier flux measurements for both observations, LUMOS estimates a flux close to that measured in other exposures of the same object.

objects. Altogether,  $\approx 85$  per cent of the observations have higher SNR with LUMOS photometry.

#### 6.1.4 Observation's flagging

The MEMBA pipeline already provides an outlier flag for its measurements. This is a discrete value that flags objects with problematic image reductions, e.g. saturated pixels, crosstalk, cosmetics, distortion or undesirable artefacts near the target source such as scattered light, cosmic rays and blending. LUMOS uses the reduced PAUCam images, which are affected by all these effects. However, as we already showed in Cabayol-Garcia et al. (2020) and earlier in this paper (Section 5.3), LUMOS deals with recurrent problems as scattered light or blending. Fig. 15 shows examples of scattered light (left-hand panel) and cosmic-ray (right-hand panel) affected observations. For the former, MEMBA provides a flux of  $-43.87 e^-/s$ , while LUMOS measures  $23.01 e^-/s$ . In the cosmic-ray example, MEMBA provides a calibrated flux of  $-211.21 e^-/s$ , while with LUMOS this is  $110.76 e^-/s$ . Other observations of the same galaxy in the same NB filter provide a mean flux  $100.75 e^-/s$ , which suggests that the LUMOS measurement is closer to the correct flux.

Currently, with aperture photometry 10 per cent of the observations are flagged. Within these flagged objects, 72 per cent are observations affected by scattered light, 18 per cent have image distortion effects and the rest is distributed among other minority effects such as e.g. crosstalk, cosmic rays or cosmetics. We have observed that LUMOS predictions are only affected in the presence of image distortions, cosmetics and saturated pixels, which reduces the number of flagged observations from 10 to 2 per cent. This reduction highlights that LUMOS is more robust towards outliers in the photometry, which it



**Figure 16.** Comparison between PAUS flux measurements and SDSS measurements convolved with PAUCam filters (PAUS synthetic fluxes). The solid black line corresponds to LUMOS measurements and the red dashed line to MEMBA.

is particularly interesting since the network is not explicitly trained to deal with artefacts as cosmic rays or crosstalk signals. However, by using real PAUCam background cutouts, we include examples of such effects in the training sample from which LUMOS learns to make robust predictions. As a result, LUMOS increases the size of the galaxy sample that is considered reliable.

## 6.2 Comparison with SDSS spectroscopy

To further validate the flux estimates, we have compared our measurements with synthetic PAUS photometry. These are constructed convolving SDSS galaxy spectra with the PAUCam filter throughput. Unfortunately, the synthetic PAUS data corresponds to a bright sample with a magnitude limit  $i_{AB} < 20.5$ , which only provides validation of bright sources. Comparing PAUS with PAUS synthetic data requires having spectra and PAUS photometry of the same galaxies and matching them by sky position (we have paired galaxies within 0.5 arcsec). It also requires scaling the synthetic PAUS fluxes with a multiplicative zero-point (zp). The zero-point is obtained by minimizing the  $\chi^2$  between PAUS observations and PAUS synthetic fluxes, i.e.

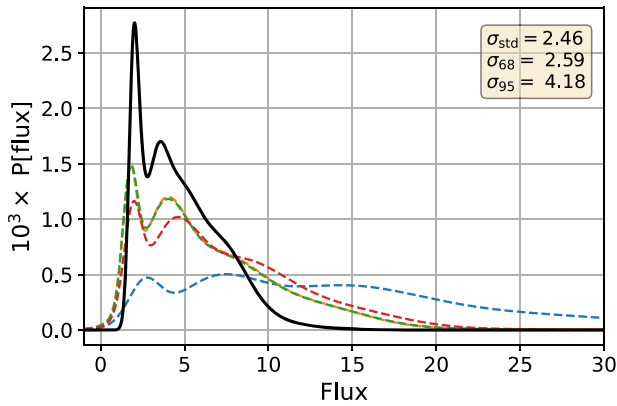
$$\chi^2 = \sum_i \frac{(f_{SDSS,i} - zp \cdot f_{SDSS_{PAUS},i})^2}{\sigma_{SDSS,i}^2 + zp^2 \cdot \sigma_{SDSS_{PAUS},i}^2}, \quad (20)$$

where SDSS is the observed SDSS photometry,  $SDSS_{PAUS}$  is the SDSS-PAUS synthetic flux and the sum (i) is over the *gri* bands. This minimization provides a median of 1.64 with a  $\sigma_{68} = 1.01$ .

Fig. 16 shows that both MEMBA and LUMOS agree well with SDSS-PAUS convolved flux measurements. MEMBA displays a lower spread than LUMOS, with  $\sigma_{68} = 0.22$  and 0.24, respectively. None the less, LUMOS shows a 5 per cent bias while in MEMBA this goes to 10 per cent. Furthermore, the number of observations at more than  $5\sigma$  from the mean of the distribution is reduced by 2 with LUMOS, going from 3 to 1.5 per cent.

## 6.3 Coadded flux measurements

The co-added flux measurements are constructed combining individual observations of the same galaxy in the same NB. Co-adding exposures increases the SNR of the galaxy photometry and it is very helpful to reject wrong observations. Before co-adding individual



**Figure 17.** The co-added flux probability distribution (black solid line) constructed from its individual observations (coloured dashed lines). The upper box displays the  $\sigma_{\text{std}}$ ,  $\sigma_{68}$ , and  $\sigma_{95}$  of the co-added PDF.

observations, a zero-point calibration per image is required, which in our case is done relative to SDSS (see Section 2.1 for more details about PAUS data). One common way of co-adding flux measurements ( $f_{\text{coadd}}$ ), and the currently implemented in PAUS, is a weighted sum of the individual observations

$$f_{\text{coadd}} = \frac{\sum_i f_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2}, \quad (21)$$

where the weights are the inverse variance of the observations and  $f_i$  and  $\sigma_i^2$  are the flux measurement and its variance of the  $i$ th observation, respectively.

LUMOS calibrates the Gaussian components individually with the photometric zero-point in such a way that when these are combined, they already provide a calibrated PDF for the flux observation. With LUMOS, combining point like estimates with equation (21) is still possible. Nevertheless, it can also generate co-added measurements combining the probability distribution of the individual galaxy observations. Fig. 17 shows an example of the co-added flux PDF of a faint galaxy. The dashed coloured lines correspond to the individual observations while the black line is the co-added PDF. This example also shows the benefit of creating co-adds at a PDF level. Combining point-like values would only provide a flux measurement close to the co-added PDF peak. In contrast, combining the PDFs keeps the contributions from the secondary peaks and the tails and therefore, it contains more valuable information about the measurement than a single point-like estimate.

Furthermore, having the PDF allows the calculation of the flux and flux error using different statistical estimators, e.g. the median or the peak. From the co-added flux PDF, we estimate the mean, median, the peak, the variance,  $\sigma_{68}$  and  $\sigma_{95}$ . From the two latest quantities, we construct a measurement of the PDFs Gaussianity,

$$\eta \equiv \sigma_{95} / \sigma_{68} - 1, \quad (22)$$

which can be helpful to decide which are the best flux and flux error estimators. While for a sharp and peaked PDF ( $\eta \gtrsim 1$ ), the peak or the median would provide similar flux estimates, in non-Gaussian PDFs as e.g. the galaxy in Fig. 17 these two estimators would give significantly different measurements. The PDF Gaussianity also affects the flux uncertainty estimators. Broadly, galaxies with  $\eta > 1$  have  $\sigma_{68} < \sigma_{\text{std}}$ , while this is the opposite for galaxies with  $\eta < 1$ . Fig. 17 ( $\eta = 0.56$ ) is also an example of multiple peaked PDF where  $\sigma_{68} > \sigma_{\text{std}}$  (see the top right-hand box in the Figure for more details).

We have tested applying different flux and flux error estimators based on the  $\eta$  parameter. However, at the end of the day we have found that the peak of the flux PDF is the best estimator regardless of the PDF Gaussianity and that using  $\sigma_{68}$ ,  $\sigma_{\text{std}}$  or  $\sigma_{95}/2$  does not lead to a significant difference.

#### 6.4 Photometric redshift estimates

Accurate photo- $z$  estimates are crucial for many science applications. Improving the photometry SNR is expected to improve the photo- $z$  estimates. In this section, we have tested LUMOS photometry with BCNZ2 (Eriksen et al. 2019) and DEEPZ (Eriksen et al. 2020): a template-based method and a deep learning algorithm built specifically for estimating PAUS photo- $z$ s. As there is not a sample of galaxies with known photometry, testing the photo- $z$ s has also been particularly helpful to find and fix some issues in the LUMOS photometry that were missed with other validation tests. One example of this are galaxies which were exhibiting an oscillating photometry. This kind of objects were detected as photo- $z$  outliers and we could trace that these were triggered by galaxy images with sub-pixels shifts with respect to the centre of the stamp.

Fig. 18 shows the photo- $z$  dispersion with BCNZ2 or DEEPZ using LUMOS photometry to  $i_{\text{AB}} < 22.5$ . We have also included the photo- $z$  result with the MEMBA forced aperture photometry as a comparison. With BCNZ2, LUMOS photometry reduces the photo- $z$  scatter by 5–15 per cent for galaxies with  $i_{\text{AB}} > 20.5$ . However, the right-hand panel also shows a small degradation at the faintest galaxies with BCNZ2 on LUMOS photometry (dashed blue line). This degradation is related with the galaxy redshift rather than to its brightness. At high redshift, photo- $z$ s with LUMOS photometry are statistically better, however, there are some high redshift outliers that increase  $\sigma_{68}$ . Rejecting galaxies with spectroscopic redshift ( $z_s$ )  $z_s > 0.8$ , the faintest galaxies ( $i_{\text{AB}} > 22$ ) have a 14 per cent lower photo- $z$  dispersion with LUMOS than with MEMBA photometry.

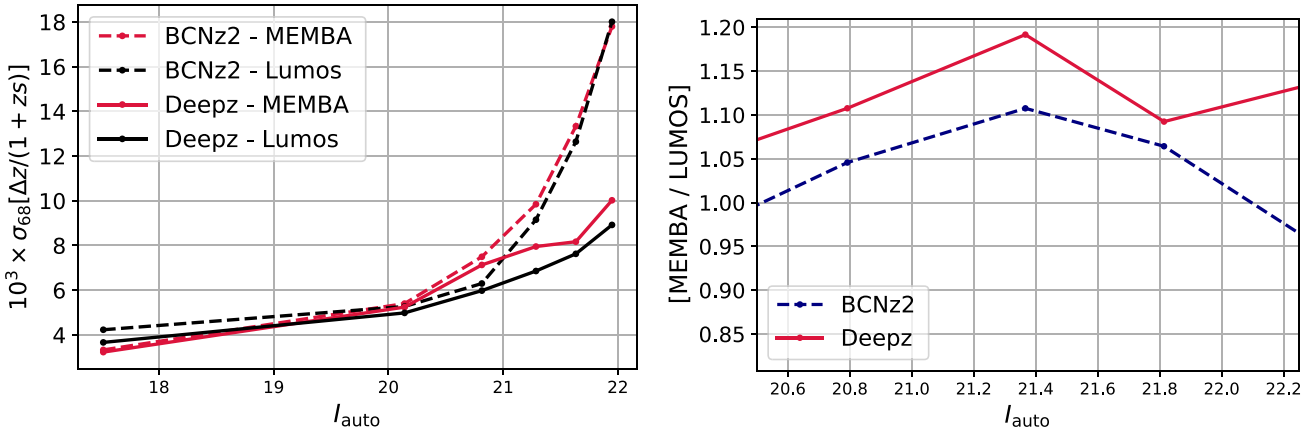
Photo- $z$ s with DEEPZ are not showing this degradation at high redshift. For galaxies with  $i_{\text{AB}} > 20.5$ , the DEEPZ photo- $z$ s are between 10 per cent and 20 per cent more precise with LUMOS photometry. Furthermore, at  $i_{\text{AB}} > 22$  and without any redshift cut, photo- $z$ s are 15 per cent better. This suggests that the minor degradation with BCNZ2 at high redshift is caused by the photo- $z$  code.

With both photo- $z$  codes, the performance is degrading at the brightest end with the LUMOS photometry. This could potentially be triggered by differences between the TEAHUPOO image simulations and the data. Bright galaxies with higher SNR are more resolved. Therefore, discrepancies between the training simulations and the data are more evident and these could have a stronger effect on the network’s performance. Nevertheless, the fraction of objects affected by this effect is small and furthermore, these are the brightest galaxies, which are not those we are more interested in.

In PAUS, a galaxy is considered an outlier if

$$|z_p - z_s| / (1 + z_s) > 0.02, \quad (23)$$

where  $z_p$  is the photo- $z$  and  $z_s$  is the spec- $z$ . This outlier definition is very strict compared to broad-band photometry, where a common outlier definition is  $|z_p - z_s| > 0.15(1 + z_s)$ , e.g. Ilbert et al. (2006) and Bilicki et al. (2018). LUMOS photometry reduces the outlier rate with both BCNZ2 and DEEPZ. With BCNZ2, the outlier rate is reduced by 5 per cent in the complete catalogue. With DEEPZ, the improvement is greater, with 20 per cent less outliers. This number increases to 23 per cent for objects with  $i_{\text{AB}} > 22$ .



**Figure 18.** Left-hand panel: Photo- $z$  precision with BCNz2 and DEEPZ using LUMOS or MEMBA photometry. Right-hand panel: Relative difference between photo- $z$ s with LUMOS or MEMBA photometry.

To the best of our knowledge, we do not know about any photo- $z$  code that could properly deal with flux PDFs and both BCNz2 and DEEPZ require point estimates for the flux and its uncertainty. Here, these quantities are estimated as the peak and  $\sigma_{68}$  of the co-added flux PDF. We have also tested other quantities as e.g. the median and the standard deviation or choosing different estimators based on the  $\eta$  parameters (equation 22). However, the peak and  $\sigma_{68}$  are those providing the best photo- $z$  estimates.

The photo- $z$  improvement obtained with LUMOS photometry is lower than expected considering the increment in the SNR. In Appendix D, we have used PAUS simulated mocks to test BCNz2 performance with SNR and its behaviour with artificially injected issues in the sample photometry. The results suggest that the photo- $z$  improvement should be  $\approx 90$  percent greater than what we actually see on data if the sample had perfect photometry. However, errors in the zero-point calibration and outliers in the flux measurements rapidly degrade the photo- $z$  performance, suggesting that currently these are potential limiting factors of the photo- $z$  performance.

## 7 CONCLUSIONS AND DISCUSSION

Accurate galaxy photometry is a key ingredient for imaging surveys to obtain precise photometric redshifts. We have developed LUMOS, a deep learning method to estimate the galaxy flux for astronomical images. LUMOS is the evolution of BKGNET, a deep learning method that predicts the background light of astronomical images with strongly varying noise patterns. In contrast, LUMOS predicts the background-subtracted galaxy flux, which requires an intrinsic background noise measurement. The algorithm has been developed for PAUCam images.

LUMOS is trained on TEAHUPOO galaxies, image simulations specially built for this work (see Fig. 2). TEAHUPOO galaxy images use PAUCam image cutouts for the background noise. First, astronomical images contain distorting effects, e.g. scattered light or crosstalk, and artefacts like e.g. cosmic rays or blended galaxies, that trigger inaccuracies in the photometry. Including real PAUCam cutouts in our simulations ensures that LUMOS has training examples to learn how to deal with such effects. Without explicitly developing LUMOS to provide photometry in the presence of distorting artefacts, the network provides reliable flux measurements on PAUCam observa-

tions affected by scattered light, cosmic rays or other contaminating effects that require flagging with aperture photometry.

Furthermore, we have tested the LUMOS deblending capability on simulations (see Section 5.3). Without explicitly including blended galaxies in the training sample, LUMOS is able to extract the target galaxy photometry much better than aperture photometry (Fig. 8). While aperture photometry provides a catastrophic flux measurement for blended sources, LUMOS is able to provide a flux with 2–10 percent accuracy, depending on the distance in pixels to the overlapping source. This is particularly interesting since LUMOS has not been written to deblend galaxies, however, this came without additional cost by using deep learning and real PAUCam background noise patterns.

LUMOS consists on a CNN followed by an MDN (Fig. 4). While most photometry algorithms provide a flux value and an associated uncertainty, LUMOS outputs the flux probability distribution as a linear combination of five Gaussian distributions. Even if many science applications require photometry point estimates, having the PDF enables the generation of a co-added flux PDF (Section 6.3). The co-added flux PDF keeps valuable information about the individual flux exposure distributions that would be missed by combining point-like estimates. While using the full PDF would require reworking of the pipelines using the photometry as an input, this can also be part of an end-to-end photometry machine-learning pipeline that goes from images to photo- $z$  estimates. The network could benefit from all the information available in the full PDF to provide more precise photo- $z$  estimates.

On PAUS observations, LUMOS provides fluxes that differ less than 1 percent from the baseline aperture photometry measurements (see Fig. 10). Concerning uncertainties, our photometry errors are 40 percent lower than with aperture photometry. This translates into between 1.5 and 3 times higher SNR in LUMOS than in MEMBA, with the largest improvement at the faint end (Fig. 14). We have run the BCNz2 and DEEPZ codes with LUMOS photometry, resulting in a reduction of the photo- $z$  scatter with both (Fig. 18). The photo- $z$  improvement using LUMOS photometry is greater with DEEPZ rather than with BCNz2, with an overall scatter reduction of 10 percent on the full catalogue and 13 percent for galaxies with  $i_{\text{AB}} > 22$  and an outlier rate reduction of  $\approx 20$  percent. Nevertheless, Appendix D shows that the photo- $z$  improvement is limited by the photometric calibration and outliers in the sample. These outliers can have

different natures as e.g. the LUMOS photometry itself or problems in the reduced PAUCam images.

LUMOS obtains the largest improvement at fainter galaxies, showing less degradation than aperture photometry. Future imaging surveys like *Euclid* or Rubin will observe much deeper galaxies with very low SNR, where LUMOS could be a very helpful tool to improve the photometry. Furthermore, LUMOS has been shown to be robust for blended sources, which could be also beneficial for future deeper surveys where the number of blended galaxies will significantly increase.

Although we have only tested the method on PAUCam images so far, we believe the methodology should readily apply to other imaging surveys. Training the network for other surveys can be addressed by training the network from scratch with simulated galaxies using real background cutouts from the targeted survey. Nevertheless, one potential difficulty of the method applied to deeper surveys is the modelled galaxy profile input requirement. While PAUS galaxies are shallow and therefore, there exists previous observations, deeper surveys like Rubin will observe galaxies for which there is not previous knowledge. Not using the modelled profile in the training barely affects the overall predicted flux measurements, however, this degrades the SNR by 15 per cent.

LUMOS supersedes BKGNET and provides a background subtracted flux measurements, which requires a measure of the background light contribution. Consequently, LUMOS deals with potential correlations between the galaxy flux and the background light that are not easy to address analytically. Moreover, in this work, we have combined two independent networks, LUMOS and DEEPZ, which provides the greatest photo-z obtained. This motivates the construction of an end-to-end pipeline that supersedes LUMOS providing the galaxy photometry and the photometric redshift.

## ACKNOWLEDGEMENTS

The authors thank Malgorzata Siudek, Pablo Renard, Giorgio Manzoni, Jacobo Asorey, Luca Tortorelli and Helena Sánchez-Domínguez for providing feedback on the paper.

The PAU Survey is partially supported by MINECO under grants CSD2007-00060, AYA2015-71825, ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, MDM-2015-0509, PID2019-111317GB-C31 and Juan de la Cierva fellowship and LACEGAL and EWC Marie Skłodowska-Curie grant No 734374 and no.776247 with ERDF funds from the EU Horizon 2020 Programme, some of which include ERDF funds from the European Union. IEEC and IFAE are partially funded by the CERCA and Beatriu de Pinós program of the Generalitat de Catalunya. Funding for PAUS has also been provided by Durham University (via the ERC StG DEGAS-259586), ETH Zurich, Leiden University (via ERC StG ADULT-279396 and Netherlands Organisation for Scientific Research (NWO) Vici grant 639.043.512), Bochum University (via a Heisenberg grant of the Deutsche Forschungsgemeinschaft (Hi 1495/5-1) as well as an ERC Consolidator Grant (No. 770935)), University College London, Portsmouth support through the Royal Society Wolfson fellowship and from the European Union's Horizon 2020 research and innovation programme under the grant agreement No 776247 EWC.

The PAU data centre is hosted by the Port d'Informació Científica (PIC), maintained through a collaboration of CIEMAT and IFAE, with additional support from Universitat Autònoma de Barcelona and ERDF. We acknowledge the PIC services department team for their support and fruitful discussions. We gratefully acknowledge

the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research.

## DATA AVAILABILITY

The PAUS raw data are publicly available through the ING group. A few reduced images are publicly available at <https://www.pausurvey.org>.

## REFERENCES

- Ahumada R. et al., 2020, *Astrophys. J. Suppl.*, 249, 3  
 Aihara H. et al., 2018, *Publ. Astron. Soc. Japan*, 70, S8  
 Alarcon A. et al., 2021, *MNRAS*, 501, 6103  
 Arcelin B., Doux C., Aubourg E., Rouelle C., LSST Dark Energy Science Collaboration, 2021, *MNRAS*, 500, 531  
 Astropy Collaboration et al., 2013, *A&A*, 558, A33  
 Bertin E., 2006, in Gabriel C., Arviset C., Ponz D., Enrique S., eds, ASP Conf. Ser. Vol. 351, *Astronomical Data Analysis Software and Systems XV*. Astron. Soc. Pac., San Francisco, p. 112  
 Bertin E., 2011, in Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds, ASP Conf. Ser. Vol. 442, *Astronomical Data Analysis Software and Systems XX*. Astron. Soc. Pac. San Francisco, p. 435  
 Bertin E., Arnouts S., 1996, *Astron. Astrophys. Suppl. Ser.*, 117, 393  
 Bilicki M. et al., 2018, *A&A*, 616, A69  
 Bishop C. M., 1994, Technical report, Mixture density networks  
 Bordoloi R., Lilly S. J., Amara A., 2010, *MNRAS*, 406, 881  
 Boucaud A. et al., 2020, *MNRAS*, 491, 2481  
 Cabayol-Garcia L. et al., 2020, *MNRAS*, 491, 5392  
 Casas R. et al., 2012, in Holland A. D., Beletic J. W., eds, *High Energy, Optical, and Infrared Detectors for Astronomy V*. SPIE, Amsterdam, Netherlands, p. 845326  
 Casas R. et al., 2016, in Evans C. J., ed., *Ground-based and Airborne Instrumentation for Astronomy VI*. SPIE - International Society for Optics and Photonics, Edinburgh, United Kingdom, p. 99084K  
 Castander F. J. et al., 2012, in McLean I. S., ed., *Ground-based and Airborne Instrumentation for Astronomy IV*. SPIE - International Society for Optics and Photonics, Amsterdam, Netherlands, p. 84466D  
 Dawid A. P., 1984, *J. R. Stat. Soc. A (General)*, 147, 278  
 Dawson K. S. et al., 2013, *AJ*, 145, 10  
 de Jong J. T. A. et al., 2013, *The Messenger*, 154, 44  
 DES Collaboration et al., 2021, preprint ([arXiv:2105.13549](https://arxiv.org/abs/2105.13549))  
 Domínguez Sánchez H. et al., 2019, *MNRAS*, 484, 93  
 Driver S. P. et al., 2011, *MNRAS*, 413, 971  
 Drlica-Wagner A. et al., 2018, *ApJS*, 235, 33  
 Eckert K. et al., 2020, *MNRAS*, 497, 2529  
 Eriksen M. et al., 2019, *MNRAS*, 484, 4200  
 Eriksen M. et al., 2020, *MNRAS*, 497, 4565  
 Fisher R. A., 1922, *Phil. Trans. R. Soc.*, 222, 309  
 Gaia Collaboration et al., 2018, *A&A*, 616, A1  
 Girshick R., 2015, in 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, Santiago, Chile, p. 1440  
 Gneiting T., Raftery A. E., Westveld A. H., Goldman T., 2005, *Mon. Weather Rev.*, 133, 1098  
 Haigh C., Chamba N., Venhola A., Peletier R., Doorenbos L., Watkins M., Wilkinson M. H. F., 2021, *A&A*, 645, A107  
 Hausen R., Robertson B. E., 2020, *ApJS*, 248, 20  
 Heasley J. N., 1999, in Craine E. R., Crawford D. L., Tucker R. A., eds, ASP Conf. Ser. Vol. 189, *Precision CCD Photometry*. Astron. Soc. Pac., San Francisco, p.56  
 Herbel J., Kacprzak T., Amara A., Refregier A., Lucchi A., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 054  
 Ilbert O. et al., 2006, *A&A*, 457, 841  
 Ilbert O. et al., 2008, in Kodama T., Yamada T., Aoki K., eds, ASP Conf. Ser. Vol. 399, *Panoramic Views of Galaxy Formation and Evolution*. Astron. Soc. Pac., San Francisco, p. 169  
 Ilbert O. et al., 2009, *ApJ*, 690, 1236



- Ioffe S., Szegedy C., 2015, in Bach F., Blei D., eds, Proceedings of Machine Learning Research Vol. 37, Proceedings of the 32nd International Conference on Machine Learning. PMLR, Lille, p. 448
- Ivezić Ž. et al., 2019, *ApJ*, 873, 111
- Jones E., Oliphant T., Peterson P. et al., 2001, SciPy: Open source scientific tools for Python. Available at: <http://www.scipy.org/>
- Kelley H. J., 1960, *Ars J.*, 30, 947
- Kennicutt Robert C. J., 1998, *ARA&A*, 36, 189
- Kingma D. P., Ba J., 2015, in Bengio Y., LeCun Y., eds, 3rd International Conference on Learning Representations. ICLR, San Diego, CA
- Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. NIPS'12. Curran Associates Inc., New York, p. 1097
- Kron R. G., 1980, *ApJS*, 43, 305
- Kuijken K. et al., 2019, *A&A*, 625, A2
- Kuijken K., 2008, *A&A*, 482, 1053
- Lang D., Hogg D. W., Schlegel D. J., 2016, *AJ*, 151, 36
- Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
- Leauthaud A. et al., 2007, *ApJS*, 172, 219
- Lecun Y., Bottou L., Bengio Y., Haffner P., 1998, *Proceedings of the IEEE*, 86, 2278
- Lilly S. J. et al., 2009, *ApJS*, 184, 218
- Magnier E. A. et al., 2020, *ApJS*, 251, 6
- Martí P., Miquel R., Castander F. J., Gaztañaga E., Eriksen M., Sánchez C., 2014, *MNRAS*, 442, 92
- Massey R., Refregier A., 2005, *MNRAS*, 363, 197
- Merlin E. et al., 2015, *A&A*, 582, A15
- Mighell K. J., 1999, in Craine E. R., Crawford D. L., Tucker R. A., eds, ASP Conf. Ser. Vol. 189, Precision CCD Photometry. Astron. Soc. Pac., San Francisco, p. 50
- Ni Q., Timlin J., Brandt W. N., Yang G., 2019, *Res. Notes AAS*, 3, 5
- Padilla C. et al., 2016, in Christopher J. E., ed., Ground-based and Airborne Instrumentation for Astronomy VI. SPIE - International Society for Optics and Photonics, Edinburgh, United Kingdom, p. 99080Z
- Padilla C. et al., 2019, *AJ*, 157, 246
- Price-Whelan A. M. et al., 2018, *AJ*, 156, 123
- Refregier A., 2003, *MNRAS*, 338, 35
- Robotham A. S. G., Davies L. J. M., Driver S. P., Koushan S., Taranu D. S., Casura S., Liske J., 2018, *MNRAS*, 476, 3137
- Rowe B. T. P. et al., 2015, *Astron. Comput.*, 10, 121
- Soo J. Y. H. et al., 2021, *MNRAS*, 503, 4118
- Suchyta E. et al., 2016, *MNRAS*, 457, 786
- Tan C., Sun F., Kong T., Zhang W., Yang C., Liu C., 2018, preprint ([arXiv:1808.01974](https://arxiv.org/abs/1808.01974))
- Tercan H., Guajardo A., Heinisch J., Thiele T., Hopmann C., Meisen T., 2018, *Procedia CIRP*, 72, 185
- Tonello N. et al., 2019, *Astron. Comput.*, 27, 171
- van Dokkum P. G., 2001, *Publ. Astron. Soc. Pac.*, 113, 1420
- Wright A. H. et al., 2016, *MNRAS*, 460, 765
- Zhang K., Bloom J. S., 2020, *ApJ*, 889, 24
- Zhao Z., Zheng P., Xu S., Wu X., 2019, *IEEE Transactions on Neural Networks and Learning Systems*, 30, 3212
- Zhuang F., Qi Z., Duan K., Xi D., Zhu Y., Zhu H., Xiong H., He Q., 2019, preprint ([arXiv:1911.02685](https://arxiv.org/abs/1911.02685))

## APPENDIX A: FLUX ESTIMATION METHODS: DERIVATIONS

This appendix derives the linear combination of pixel values giving and unbiased and optimal flux measurement (equation 6). The SNR of the measurement when combining pixels with mean  $m$  and weight  $w$  is

$$\text{SNR} = \frac{\sum_i w_i m_i}{\sqrt{\sum_i w_i^2 (m_i + b_i)}}, \quad (\text{A1})$$

where  $b_i$  is the background mean value. The optimal SNR is found by requiring stationary derivatives for all weights independently, which

results in

$$w_x = \lambda \frac{m_x}{(m_x + b_x)} \quad (\text{A2})$$

where  $\lambda$  is a constant. The flux measurement being unbiased means

$$\sum_i m_i = \sum_i w_i m_i. \quad (\text{A3})$$

Using this requirement, the pixel weights (equation A2) becomes

$$w_x = \frac{\sum_i m_i}{\sum_i m_i^2 / (m_i + b_i)} \frac{1}{1 + b_x / m_x} \quad (\text{A4})$$

Notice that, given a pixel  $x$ , its weight  $w_x$  depends on the true flux ( $m_x$ ) and background ( $b_x$ ) on that concrete pixel.

## APPENDIX B: CONVOLUTIONAL NEURAL NETWORKS

Machine-learning methods are data analysis techniques where the algorithm learns from the data. In particular, one of the currently most popular classes of algorithms are neural networks (Lecun et al. 1998), which are designed to recognize patterns, usually learned from training data (*supervised method*). They are mainly used for regression and classification problems. Deep learning is a subset of machine-learning that refers to the development of neural network technology involving a large number of layers.

Other terms that one needs to be familiar with are *epoch* and *batch*. An *epoch* is an iteration over the complete training data set. However, it is common practice to give the data to the network in *batches*. Feeding the network in batches helps it learn faster as in every iteration over a batch, it updates all the weights. Then, instead of updating once per epoch, it updates as many times as there are batches, which speeds up the process.

Deep learning methods, and in general any supervised machine-learning algorithm, model a problem by optimizing a set of trainable parameters that fit the data. This is done in three stages: *forward propagation*, *back propagation*, and *weight optimization*. The network starts with the *forward propagation*. At this stage, the input data propagates through all the network layers to give a prediction for each of the input samples. After that, by comparing the prediction with the known true value (*label*), the network estimates a prediction error with a loss function. The ultimate goal of the training procedure is to minimize the loss function.

To minimize the loss, one performs a process named *back propagation* after evaluating the loss function. *Back propagation* consists of computing the contribution of each weight to the loss function (Kelley 1960). Such contributions are calculated using the chain rule. The *weight optimization* consists in updating the network parameters using the derivatives estimated. This whole procedure takes place repeatedly, reducing the loss function after each iteration while adapting the parameters to the data. The amount of variation allowed per iteration is regulated by the *learning rate*.

After each layer there is an activation function, which are non-linear functions that map the outcome of a layer to the input of the following one. This is required to produce non-linearities in the model. An example of an activation function is the Rectified Linear Unit (ReLU) (Krizhevsky, Sutskever & Hinton 2012).

In this work, we use a Convolutional Neural Network (CNN; Lecun et al. 1998). Our network contains four differentiated types of layers:

(i) *Convolutional layer*: This layer makes the network powerful in image and pattern recognition tasks. It has a filter, technically

named *kernel* and is usually two-dimensional, which contains a set of trainable weights used to convolve the image. The outcome of this layer is the input image convolved with the kernel. In a given convolutional layer, each of the outputs is a linear combination of the different convolutions. Each of these convolutions will generate a convolved image, which we refer to as *channel*. All of them together are the input of the next layer.

(ii) *Pooling layers*: This layer reduces the dimensionality of the set of convolved images. It applies some function (e.g. maximum, sum, mean) to a group of spatially connected pixels and reduces the dimensions of such group. For example, the Max-Pooling, which we use in this paper, takes a e.g.  $(2 \times 2)$  group of pixels and converts them to their maximum. Although we use it to handle the amount of data generated after the convolutions, it also regularizes the model to avoid learning from non-generalizable noise and details in the training data (also known as overfitting).

(iii) *Batch normalization layer*: In this layer, the network normalizes the output of a previous activation layer. It subtracts the mean and divides by the standard deviation. Batch normalization helps to increase the stability of a neural network and avoids over-fitting problems (Ioffe & Szegedy 2015).

(iv) *Fully connected layer*: These layers are usually the last layers of the network and its input is the linearized outcome of the previous ones (in our network: convolutions, poolings, and batch normalizations). It applies a linear transformation from the input to the output.

### APPENDIX C: FORECASTING THE EFFECT OF ERRORS ON PROFILE PARAMETERS

The algorithms described in Section 3 need information about the galaxy profile properties, making the flux measurement accuracy sensitive to errors on these parameters. Here, we will quantify the effect that errors in the input galaxy parameters have on the flux measurements using a Fisher forecast formalism (Fisher 1922) on TEAHUPOO galaxies. In particular, we will test if neural networks are more robust to errors in the input parameters.

A photometry algorithm ( $\Phi$ ) that measures the flux ( $\tilde{f}$ )

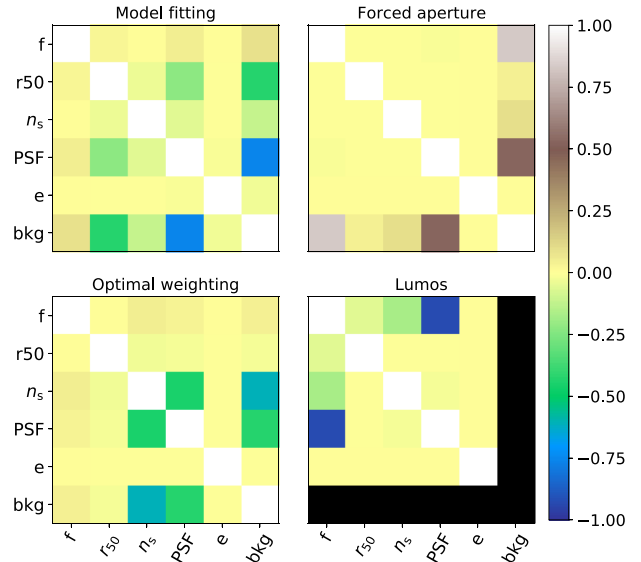
$$\tilde{f} = \Phi(I, f, r_{50}, n_s, PSF, e, b) \quad (C1)$$

is foremost dependent on the galaxy image ( $I$ ), but also parameters such as the total flux ( $f$ ), the half-light radius ( $r_{50}$ ), the Sérsic index ( $n_s$ ), the PSF FWHM, the ellipticity ( $e$ ) and the background light ( $b$ ). This is because e.g. the aperture algorithm uses these quantities to scale the pixels and they are used to construct a profile for the profile fit. We can estimate the error on the flux from these parameters using a Fisher matrix formalism. The Fisher matrix is estimated by

$$FM_{\mu\nu} = \frac{\partial \tilde{f}}{\partial \mu} \left( \sigma_{\tilde{f}}^{-2} \right) \frac{\partial \tilde{f}}{\partial \nu}, \quad (C2)$$

where the indices  $\mu$  and  $\nu$  are the different parameters the total flux depend on (see equation C1).

The covariance matrix of the flux measurements can be estimated as the inverse of the Fisher matrix. Fig. C1 shows the correlation matrices of the parameters  $Flux$ ,  $r_{50}$ ,  $n_s$ ,  $PSF$ ,  $e$ , and  $b$  for the four flux estimation methods described in Section 3. The correlation matrix differs for different galaxy types (e.g. different morphologies or brightness). Here, we have constructed a galaxy with values of  $r_{50}$ ,  $n_s$ , and  $PSF$  corresponding to the mean of their distribution. The model-fitting method (top left-hand panel), together with the optimal weighting (bottom left-hand panel) are those showing more correlation between the flux and the profile parameters. In contrast,



**Figure C1.** The correlation matrix for the parameters  $f$ ,  $r_{50}$ ,  $n_s$ ,  $PSF$ ,  $e$ , and  $bkg$  for: Top left-hand panel: the model-fitting method, Top right-hand panel: Forced aperture photometry, Bottom left-hand panel: Optimal weighted pixel sum and Bottom right-hand panel: LUMOS.

the forced aperture photometry and LUMOS show a lower correlation. The lower correlation makes these methods more robust since the effect of uncertainties in the parameters is also lower.

All methods but LUMOS show a high correlation with the background estimation. The background light is not an input parameter for LUMOS, since it is intrinsically measured inside the method. This makes LUMOS insensitive to external errors on this parameter.

As mentioned, the correlation matrices in Fig. C1 are for a particular common galaxy. To estimate the requirements on these parameters, we have studied how errors in the profile parameters propagate to errors in the flux measurement. For that, we have assumed a 10 per cent prior error in each of the input parameters, such that

$$FM_{\text{comb}} = FM + FM_{\text{priors}}, \quad (C3)$$

where  $FM_{\text{priors}}$  is a diagonal matrix including the inverse prior variance of each parameter. The variance on the flux parameter is then

$$\sigma_f^2 = (FM_{\text{comb}}^{-1})_{ff} \quad (C4)$$

where the matrix subscripts ( $ff$ ) denotes selecting the row/column corresponding to the flux parameter.

Table C1 shows the percentage of error in the parameters that propagates to a 10 per cent error in the flux measurements. While studying a particular parameter, we always assume that the rest are fixed. As the sensitivity to the parameters can vary among galaxy types, the results in Table C1 are averaged out of one hundred independent random galaxies. LUMOS is the method that requires higher errors on the parameters to propagate to a 10 per cent flux error, i.e. it shows as the most robust method. As expected from Fig. C1, the  $PSF$  is the parameter it is more sensitive to, followed by the half-light radius. However, in both cases it is still less sensitive than the other methodologies.

The robustness of LUMOS most likely comes from having the galaxy image. LUMOS is provided with the galaxy and the modelled profile images, which allows the comparison and detection of prob-

**Table C1.** Percentage of error in  $r_{50}$ ,  $n_s$ , PSF, ellipticity ( $e$ ), and background noise ( $b$ ) that propagates to a 10% error in the total flux. Note that when studying one of the parameters, the rest remain fixed. Also note that high errors indicate that the method is more robust, since it requires a large error in the parameter to propagate to a 10% flux error.

	$r_{50}$	$n_s$	PSF	$e$	$b$
Model-fitting	4	4	4	5	1
Forced photometry	21	59	19	1	1
Opt. weighted sum	8	11	4	13	1
LUMOS	28	80	14	83	–

lematic profiles instead of blindly relying on the input parameters. Furthermore, small errors in such parameters are more subtle and difficult to detect when these are encoded in a profile rather than directly inputted into the method.

#### APPENDIX D: PHOTOMETRIC REDSHIFTS WITH BCNZ2 ON PAUS GALAXY MOCKS

In this section, we will run BCNZ2 on the PAUS galaxy mock. The purpose of this test is to have an idea of the photo- $z$  improvement that we should expect on data.

We have generated PAUS photometry with the same pipeline as the Flagship simulations (Castander et al (in prep.)) containing 500K objects over  $25 \text{ deg}^2$  with a redshift limit of 2.25. Initially, galaxies are generated with rest-frame luminosity using abundance matching between the halo mass function and SDSS galaxies. Next, the galaxy redshift is estimated using evolutionary population synthesis models. Then, mock galaxies are matched to the COSMOS galaxies from Ilbert et al. (2008) and extinction and an SED are assigned to each of them. The SED templates also take into account the emission lines.  $H_\alpha$  is computed from the ultra-violet following Kennicutt (1998). The other line fluxes are computed following observed relations. Finally, the SED is convolved with the filter transmission curves to produce the fluxes.

In this work, we have used PAUS NBs, the CFHT  $u$ -band and the Subaru  $BVriz$  broad bands and the noise is included assuming Gaussian uncertainties. We have randomly selected 10 000 noiseless galaxies and included uncertainties to mimic the SNR provided by LUMOS (Fig. D1). For the broad bands, we have generated errors to simulate the SNR in Eriksen et al. (2019).

Fig. D2 shows that the photo- $z$  on the PAUS-mock (black solid line) has on average 90 percent less scatter than on data (blue solid line). This is a large number considering that the SNR on both samples is similar and consequently, a similar performance is expected. Therefore, the photo- $z$ s appear limited by other factors than the photometry SNR.

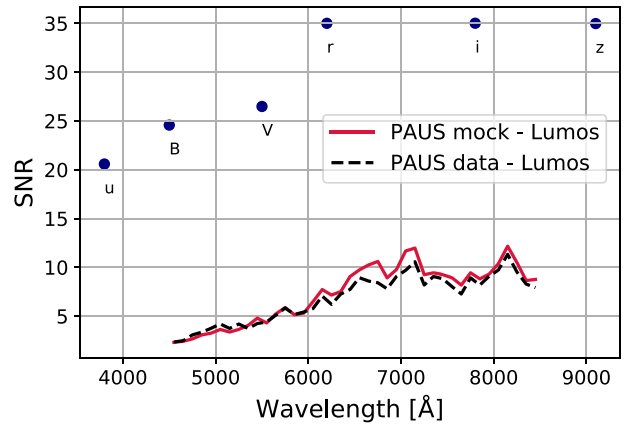
The purple solid line shows the effect of the photometric calibration. For each flux measurements, we have generated five individual observations resembling the number of PAUS observations in the COSMOS field from a Gaussian centred at the co-added flux. Each of these observations is randomly assigned with a photometric zero-point and its uncertainty from the distribution of zero-point in the PAUS COSMOS data. The zero points are scattered with their uncertainty, sampling for every galaxy a new zero-point from a Gaussian. The individual observations are combined back to a single flux error using the scattered zero points.

The median zero-point uncertainty in the PAUS data in COSMOS is  $\approx 4$  per cent. Including this effect in the photo- $z$ s (purple solid line in Fig. D2) degrades the precision by 40 per cent with respect to

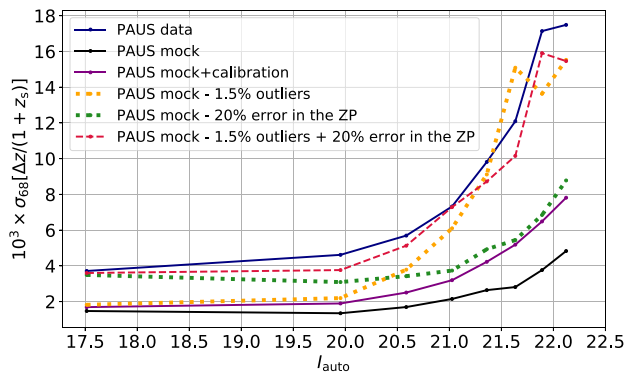
perfect photometry (black solid line), from  $\sigma_{68} = 0.0026$  to  $0.0041$ , with faint galaxies being more affected. Nevertheless, after including the calibration effect, the photo- $z$  scatter is still significantly lower on the PAUS mock compared to the results on data. The coloured dotted lines in Fig. D2 shows the photo- $z$  dispersion with further effects in the photometry that could potentially lead to a photo- $z$  degradation. All the lines are built on the purple one, assuming the calibration effect.

The green dotted line studies the effect of having an addition 20 per cent error in 20 per cent of the photometric zero points. Note that this additional error is not accounted for in the final photometric error and could potentially make a correct observation an outlier. This further reduced the photo- $z$  precision to  $\sigma_{68} = 0.0050$ , significantly affecting bright galaxies. Indeed, this effect is required to understand the photo- $z$  degradation at the bright end on data with respect to simulations.

In the orange dotted line, we have artificially injected 1.5 per cent of outliers in the PAUS mock fluxes. These outliers are directly affecting the co-added flux measurement and therefore, a 1.5 per cent of affected fluxes corresponds to a higher percentage of affected



**Figure D1.** The SNR of the PAUS galaxy simulations used to run BCNZ2. The solid red line corresponds to a PAUS mock with a SNR similar to that provided by LUMOS. As a comparison, the dashed black line corresponds to the observed SNR on PAUS data with LUMOS photometry.



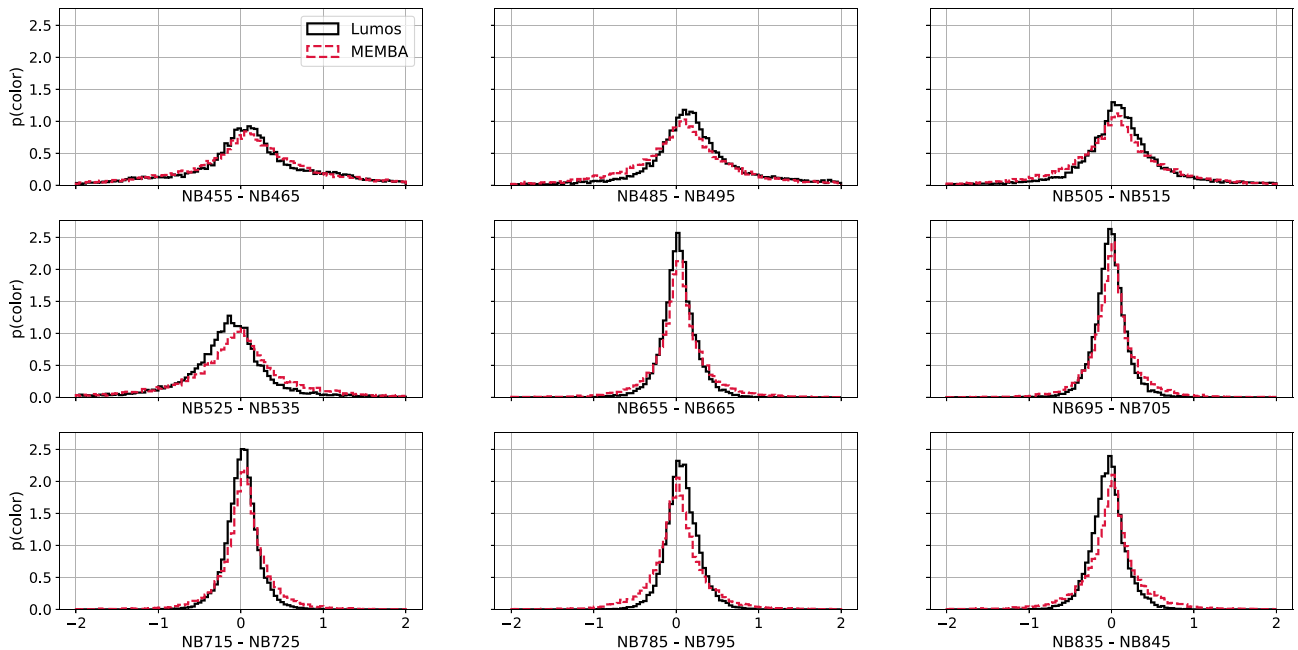
**Figure D2.** Photo- $z$  dispersion as a function of  $i$ -band magnitude using BCNZ2 for a galaxy mock with LUMOS SNR (black solid line) and PAUS data with LUMOS photometry (blue solid line). The purple line includes the photometric calibration on the PAUS mocks. Dotted lines include outliers and calibration errors in the PAUS mock photometry. The red dashed line combines the effect of the two dotted lines.

galaxies. Particularly, 45 per cent of the galaxies in the PAUS mock have at least one affected NB flux measurement and  $\approx 10$  per cent have more than one. Nevertheless, mind that not all the galaxies with affected photometries end up providing worse redshift estimates. Indeed, these outliers barely affect the bright end, where galaxies have high SNR and the photo-z algorithm deals well with an outlier in one of the bands. Contrary, at the faint end outliers increase the photo-z scatter by  $\approx 2$ . We have also tested the effect of other percentages of outliers finding that 1 per cent was too few to explain the degradation on data and 2 per cent, too much. With 1.5 per cent of outliers, the photo-z precision degrades by 80 per cent, providing a  $\sigma_{68} = 0.0074$ . The dashed red line combines the two previous effects and gives a photo-z precision close to that obtained on data.

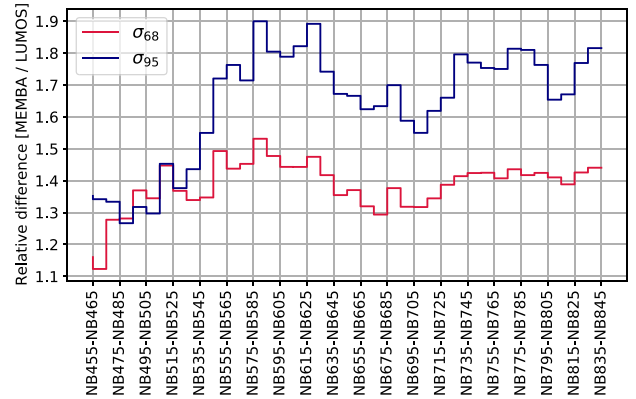
## APPENDIX E: COLOUR HISTOGRAMS IN THE COMPLETE NB SET

Colour histograms can be used to compare different photometries of the same images. Assuming an underlying galactic colour distribution, photometry uncertainties broaden such distribution and therefore the width of the colour histogram provides an idea of the photometry uncertainties. Consequently, the best photometry on a sample of galaxies is that providing narrower colour distributions. In Fig. 11 (Section 6.1), we showed the NB785-NB795 histogram, which displayed a narrower distribution for LUMOS than MEMBA. Here, we show the colour histograms results for the rest of the bands. Fig. E1 shows the colour histogram of nine different NBs with the LUMOS photometry (black solid line) and the MEMBA photometry (red dashed line).

Furthermore, Fig. E2 shows the relative difference in the effective width of the colour histograms with the photometries from LUMOS and MEMBA. As in Section 6.1, the effective widths are estimated with  $\sigma_{68}$  (equation 14) and  $\sigma_{95}$ , which is equivalent to  $\sigma_{68}$  but considering the 2.5 and 97.5 quantiles. LUMOS provides narrower colour histograms in all the NB filters. The relative difference in  $\sigma_{68}$  oscillates between 30 and 40 per cent in all NBs but the bluest,



**Figure E1.** Colour histograms for nine different NB filters using LUMOS (solid black line) or MEMBA (dashed red line) photometries.

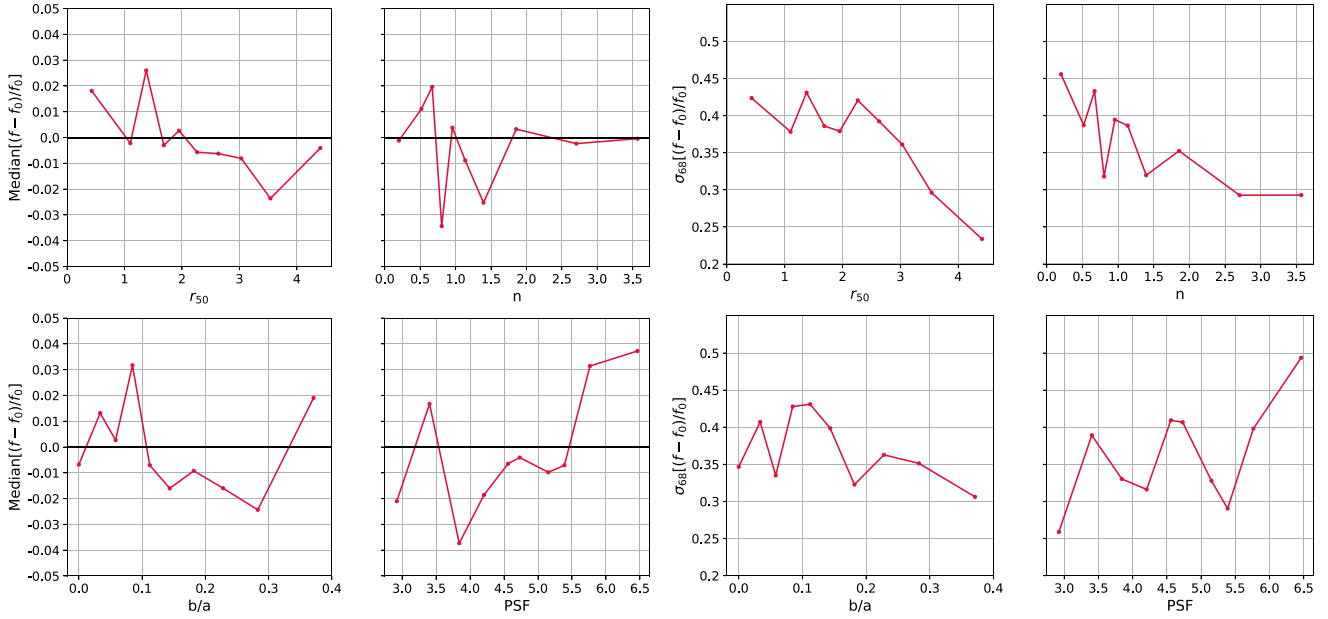


**Figure E2.** Relative difference between the effective width of the colour histograms of the PAUS photometry with LUMOS and MEMBA. The effective width has been estimated with  $\sigma_{68}$  (blue line) and  $\sigma_{95}$  (red line).

where it is  $\approx 15$  per cent. With  $\sigma_{95}$ , the relative effective width is lower at the first eight bluer bands,  $\approx 30$  per cent, and increases to  $\approx 70$  per cent for the rest of the bands. The relative difference in  $\sigma_{95}$  is systematically larger than with  $\sigma_{68}$ , which is likely related with very noisy measurements. Photometry measurements in the tails of the colour histograms will not affect the  $\sigma_{68}$  measurement, however, these will be accounted in  $\sigma_{95}$ . Consequently, Fig. E2 would be showing that the number of outlier observations is lower in LUMOS than in MEMBA.

## APPENDIX F: PHOTOMETRY AND PHOTO-Z CORRELATIONS WITH GALAXY PARAMETERS

Fig. 7 only showed the bias and the precision of the photometry obtained with LUMOS (blue solid line) as a function of magnitude on simulations. In this appendix, we are extending the exploration of the



**Figure F1.** Left-hand panel: Bias in the photometry measurements as a function of the galaxy size ( $r_{50}$ ), galaxy shape (Sérsic index), ellipticity ( $b/a$ ) and PSF. Right-hand panel: Precision in the photometry as a function of the same parameters.

photometry (Section F1) and the photo- $z$  (Section F2) predictions as a function of other galaxy parameters as e.g. the galaxy size or ellipticity.

### F1 LUMOS photometry correlation with galaxy parameters

Fig. F1 shows the bias and the precision of the LUMOS photometry as a function of the galaxy size ( $r_{50}$ ), the galaxy shape (Sérsic index,  $n$ ), the galaxy ellipticity ( $b/a$ ) and the PSF, binned in ten equally populated bins. The photometry does not show a significant bias with any of the galaxy parameters. Note that the galaxy bias increment for higher PSF values is expected since this parameter directly correlates with the quality of the data. The largest galaxies in the data set tend to have slightly underestimated flux predictions ( $\approx 1$ – $2$  per cent). This could be a consequence of the fixed stamp size, which could lead to a small leak of light. Nevertheless, this should have a weaker effect on LUMOS than on other non-trainable algorithms, since the network can learn to predict reasonable fluxes when the galaxy is partially outside the stamp.

The photometry precision is better for larger galaxies. Furthermore, the precision is higher for larger Sérsic indices, which is expected since larger Sérsic indices correspond to bigger and brighter galaxies. In contrast, we do not see any correlation between the photometry precision and the galaxy ellipticity.

### F2 Photo- $z$ correlation with galaxy parameters

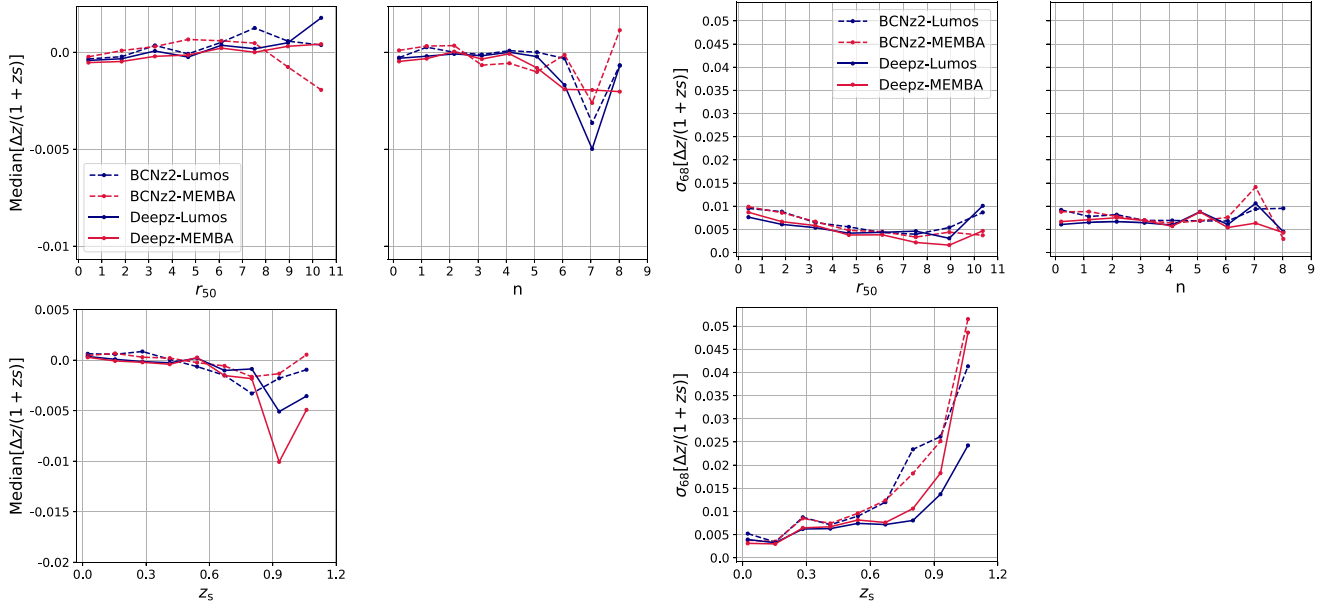
Fig. F2 explores the photo- $z$  performance with the BCNZ2 template fitting and the DEEPZ machine-learning codes as a function of

the galaxy size ( $r_{50}$ ), the galaxy shape (Sérsic index) and the spectroscopic redshift. This is presented for both the LUMOS and the MEMBA photometries. In this case, these quantities are binned in 10 equally spaced bins, so that that we can explore the photo- $z$  performance on the edges of the training set distributions.

Overall, the photo- $z$  bias (first and second columns) is not affected neither by the size nor the shape of the galaxy with any of the codes or photometries. The photo- $z$ s are also unbiased as a function of spectroscopic redshift, only presenting a  $\approx 1$  per cent bias at high redshifts with the MEMBA photometry and the DEEPZ code (blue solid line). Using DEEPZ on the LUMOS photometry also presents a  $\approx 0.5$  per cent bias, while such biases disappear with the BCNZ2 algorithm on both photometries. This suggests that it might be triggered by the photo- $z$  method.

The photo- $z$  precision (third and fourth rows) shows a similar trend with the galaxy size and shape using the MEMBA or LUMOS photometries. Note that LUMOS provides better photo- $z$  precision for small galaxies, while MEMBA gives better photo- $z$ s for larger galaxies. This is potentially related to discrepancies between the training image simulations and the data (see Section 6.4). Such differences affect more large bright galaxies, as these are more resolved. A similar effect can be noted as a function of Sérsic index.

The photo- $z$  precision with spectroscopic redshift presents a similar trend for both photometries, exhibiting better photo- $z$ s for  $z_s > 1$  with both the BCNZ2 and DEEPZ codes on the LUMOS photometry. At high redshifts, the improvement with the LUMOS photometry and the DEEPZ code is remarkable.



**Figure F2.** Left-hand panel: Bias in the photo- $z$  measurements as a function of the galaxy size ( $r_{50}$ ), galaxy shape (Sérsic index) and the spectroscopic redshift ( $z_s$ ). Right-hand panel: Precision in the photo- $z$  as a function of the same parameters.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.