

The PAU Survey: background light estimation with deep learning techniques

L. Cabayol-Garcia,^{1★} M. Eriksen,^{1★†} A. Alarcón[ⓑ],^{2,3} A. Amara,⁴ J. Carretero,^{1†}
R. Casas,^{2,3} F. J. Castander,^{2,3} E. Fernández,¹ J. García-Bellido,⁵ E. Gaztanaga,^{2,3}
H. Hoekstra[ⓑ],⁶ R. Miquel,^{1,7} C. Neissner,^{1†} C. Padilla,¹ E. Sánchez,⁸ S. Serrano,²
I. Sevilla-Noarbe,² M. Siudek[ⓑ],¹ P. Tallada^{8†} and L. Tortorelli⁹

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

²*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*

³*Institut d'Estudis Espacials de Catalunya (IEEC), E-08193 Barcelona, Spain*

⁴*Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Burnaby Road, Portsmouth PO1 3FX, UK*

⁵*Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, E-28049 Madrid, Spain*

⁶*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, The Netherlands*

⁷*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*

⁸*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), E-28040 Madrid, Spain*

⁹*Institute for Particle Physics and Astrophysics, ETH Zürich, Wolfgang-Pauli-Str 27, CH-8093 Zürich, Switzerland*

Accepted 2019 November 20. Received 2019 November 20; in original form 2019 October 4

ABSTRACT

In any imaging survey, measuring accurately the astronomical background light is crucial to obtain good photometry. This paper introduces BKGNET, a deep neural network to predict the background and its associated error. BKGNET has been developed for data from the Physics of the Accelerating Universe Survey (PAUS), an imaging survey using a 40 narrow-band filter camera (PAUCam). The images obtained with PAUCam are affected by scattered light: an optical effect consisting of light multiply reflected that deposits energy in specific detector regions affecting the science measurements. Fortunately, scattered light is not a random effect, but it can be predicted and corrected for. We have found that BKGNET background predictions are very robust to distorting effects, while still being statistically accurate. On average, the use of BKGnet improves the photometric flux measurements by 7 per cent and up to 20 per cent at the bright end. BKGNET also removes a systematic trend in the background error estimation with magnitude in the *i* band that is present with the current PAU data management method. With BKGNET, we reduce the photometric redshift outlier rate by 35 per cent for the best 20 per cent galaxies selected with a photometric quality parameter.

Key words: instrumentation: photometers – light pollution – techniques: photometric.

1 INTRODUCTION

The positions, fluxes, and other properties of galaxies and stars can be determined by analysing images of the sky. Modern imaging surveys can cover large areas of sky efficiently, resulting in measurements for large numbers of galaxies to faint magnitudes (e.g. the Dark Energy Survey (DES); Abbott et al. 2018). Improving the accuracy of the measurements is crucial for future weak lensing surveys, e.g. LSST and *Euclid* (Laureijs et al. 2011; Ivezić et al.

2019), to ensure that the results are not dominated by systematic errors.

For imaging surveys, accurate flux measurements are essential: they are used to select samples of galaxies, or to infer their physical properties. A key step towards a reliable flux estimate is the determination of the background, which needs to be subtracted. The main source of background is the brightness of the night sky, which may vary due to a range of effects, such as illumination by the Moon, airglow, and light pollution. Instrumental effects can contribute as well, and in this paper we focus on scattered light, which is the result of light deflecting from the instrument optical path appearing at a different region of the detector (Romanishin 2014).

Different approaches have been used to estimate the sky background (Bijaoui 1980; Newell 1983), and example implementations include DAOPHOT (Stetson 1987) and SEXTRACTOR (Bertin &

* E-mail: lcabayol@ifae.es (LC-G); eriksen@pic.es (ME)

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

Arnouts 1996). DAOPHOT measures the background as the mode of the uniformly scattered pixels at a certain full width at half-maximum (FWHM) of the given target source. On the other hand, SEXTRACTOR meshes the background and reconstructs a ‘background map’ with the background estimated at each particular mesh location. Other methods aim to be more robust in the presence of nearby sources. In Teeninga et al. (2015), the background is estimated at a location without nearby sources, while Popowicz & Smolka (2015) is based on the removal of small objects and an interpolation of missing pixels. In this paper, we propose a new approach based on a deep neural network (NN) to predict the background and its associated error.

Over the last few years, deep learning algorithms have resulted in revolutionary advances in machine learning and computer vision (Voulodimos et al. 2018). Theoretical breakthroughs in training deep NNs (Werbos 1982) or convolutional neural networks (CNNs; LeCun et al. 1989; Lecun et al. 1998; Zeiler & Fergus 2013), together with powerful and efficient parallel computing provided by graphics processing units (GPUs; Krizhevsky, Sutskever & Hinton 2012) have led to groundbreaking improvements across a variety of applications. The number of deep learning projects in cosmology is quickly increasing. This includes astronomical object classification (Carrasco-Davis et al. 2018; Cabayol et al. 2019), gravitational wave detection (George & Huerta 2018), and directly constraining cosmological parameters from mass maps (Fluri et al. 2018; Herbel et al. 2018).

Extracting the source photometry requires a significant amount of data engineering and parameter tweaking. This can be particularly challenging for noisy sources. Deep learning has already been successfully applied to different steps in source photometry extraction. Examples include point source detection (Vafaei Sadr et al. 2019), cosmic-ray detection (Zhang & Bloom 2019), or point spread function (PSF) modelling (Herbel et al. 2018). Deep learning has also been used to directly estimate photometric redshifts from images (D’Isanto & Polsterer 2018; Pasquet et al. 2019). These algorithms implicitly include steps for the background subtraction. Understanding these image processing steps can optimize the performance of e.g. redshift estimation and galaxy classification.

Our goal is to develop and test a deep learning background subtraction method using data from the Physics of the Accelerating Universe Survey (PAUS). PAUS is an imaging survey that measures high-precision photo- z s to faint magnitudes ($i_{AB} < 22.5$), while covering a large area of sky (Martí et al. 2014). This is possible due to the PAUCam instrument (Castander et al. 2012; Padilla et al. 2016, 2019), an optical camera equipped with 40 narrow bands (NBs) covering a wavelength range from 450 to 850 nm (Casas et al. 2016). PAUS reaches a photo- z precision $\sigma(z)/(1+z) \sim 0.0035$ for the best 50 per cent of the sample, compared to typical precision of 0.05 for broad-band measurements (Eriksen et al. 2019). The scientific goals of PAUS include the measurement of intrinsic alignments of galaxies out to $z \sim 0.75$, the study of their spectral energy distributions (SEDs), detailed studies of intermediate-scale cosmic structure (Stoher et al. 2018), and improvements of image simulations (Tortorelli et al. 2018).

PAUS imaged the COSMOS field as a calibration area given the availability of spectroscopic redshifts. The PAUS photo- z catalogue for the full COSMOS sample with $i_{AB} < 22.5$ contains outliers when compared to the spectroscopic redshifts. Some of these outliers simply arise from noisy photometry, but others are due to a strongly varying continuum produced by scattered light. The excess scattered light decreases the signal-to-noise ratio (SNR) of the photometric

measurement and also alters the statistics of the values of the pixels from which the continuum is estimated. These effects can potentially bias the flux measurements.

In this paper, we present BKGNET, a CNN capable of learning the underlying behaviour of scattered light and other distorting effects present in the PAUCam images. BKGNET predicts the background at the location of the target sources and the error associated with the background prediction. Although it is built to improve the PAUS photometry, BKGNET can be also applied to other future imaging surveys such as LSST (Ivezić et al. 2019) and *Euclid* (Laureijs et al. 2011). The code is available at <https://gitlab.pic.es/pau/bkgnet>.

The structure of this paper is as follows. In Section 2, we describe the PAU Survey and the PAUCam camera and present the modelling of scattered light using scattered-light templates. In Section 3, we introduce the specific network we have developed, as well as defining the training and testing process. Sections 4 and 5 contain the results obtained for simulated and real PAUCam images, respectively. In Section 6, we validate the network predictions on real target locations and we conclude and summarize in Section 7.

2 MODELLING SCATTERED LIGHT

PAUCam images contain substantial amounts of scattered light, which mostly affect the edge regions of some CCDs. Scattered light increases the amount of background in the affected regions and distorts the expected statistics of the pixel values used to estimate the background. Therefore, the scattered light present in PAUCam images can lead to an incorrect estimate of the background if not properly modelled, thus biasing the photometry. Moreover, the elevated background lowers the SNR of the measurements. In 2016 the PAU camera was modified in order to mitigate the effect of scattered light by introducing baffles on all the edges of the NB filters of each filter tray. Although this reduced the amount of scattered light, residuals still remain. In the latest COSMOS data reduction, around 8 per cent of exposures taken before the camera intervention are flagged as affected by scattered light, and therefore dismissed. After the intervention, this number reduced to 5 per cent of the exposures, such that on average 7 per cent of data in the COSMOS field are lost due to scattered light. In this section, we present the PAUCam scattered light model we are using throughout the paper.

2.1 The PAUS observations

PAUS has been observing since the 2015B semester and as of 2019A, PAUS has taken data for 160 nights. The current data cover 10 deg² of the CFHTLS fields¹ W1, W2; 20 deg² in W3 and 2 deg² of the COSMOS field.² The PAUS data are stored at the Port d’Informació Científica (PIC), where the data are processed and distributed (Tonello et al. 2019).

In this paper, we focus only on the data from the COSMOS field, which were taken in the semesters 2015B, 2016A, 2016B, and 2017B (the low efficiency was caused by bad weather). The COSMOS field observations comprise a total of 9749 images, 343 images for each NB. From these images, 4928 were taken before the camera intervention and 4821 after. The basic exposure times

¹http://www.cfht.hawaii.edu/Science/CFHTLS_Y_WIRCam/cfhtlsdeepwifields.html

²<http://cosmos.astro.caltech.edu/>

in the COSMOS field are 70, 80, 90, 110, and 130 s from the bluest to the reddest filter trays.

2.2 PAUS images processing

The PAUCam instrument (Castander et al. 2012; Padilla et al. 2016, 2019) is an optical camera equipped with 40 NBs, covering a wavelength range from 450 to 850nm (Casas et al. 2016). The NB filters have 13nm FWHM and a separation between consecutive bands of 10nm. The camera is also equipped with *ugrizY* broadband filters that so far have been mainly used by external observers. The camera has 18 red-sensitive fully depleted Hamamatsu CCD detectors (Casas et al. 2012), although only the eight central CCDs are used for NB imaging. Each CCD has 4096×2048 pixels with a pixel scale of $0.26 \text{ arcsec pixel}^{-1}$. The NB filter set effectively measures a high-resolution photometric spectrum ($R \approx 50$).

In this project, we use images that have already been corrected for various instrumental effects (Serrano et al., in preparation) in the PAUS NIGHTLY pipeline. This pipeline performs basic instrumental de-trending processing. The electronic effects are corrected using a master bias, which is an observation with the shutter closed and zero exposure time. To correct pixel-to-pixel variations we use dome flats, which are obtained by imaging a uniformly illuminated screen. The astrometry of the narrow-band images is calibrated comparing to the positions of *Gaia* DR2 stars (Gaia Collaboration 2018). The photometry calibration is done relative to SDSS stars (Castander et al., in preparation).

Once the images have been corrected, we perform forced photometry to extract the galaxy flux. The current PAUDM pipeline, similar to DAOPHOT, predicts the background noise as the median of the pixels within a ring placed around the target source. However, this algorithm requires a (fairly) flat background for an accurate estimate. This assumption breaks down when either the annulus or source extraction regions are affected by scattered light. In addition, other sources of errors in the background estimation are undetected sources, cosmic rays, or cross-talk. In order to minimize the effect of any of these artefacts, the pixels inside the annulus are 3σ clipped before computing the median.

The default PAUDM radii for the annulus region are $r_{\text{in}} = 30$ and $r_{\text{out}} = 45$ pixels (Serrano et al., in preparation). The annulus is selected to be sufficiently far away from the galaxy to avoid light leaking inside the ring and not too far so that the background is representative. Throughout this paper, we use the default configuration to compare this commonly used approach to our deep learning algorithm. However, in Appendix B, we study the effect of a variable annulus.

2.3 Scattered-light templates

Fig. 1 shows four PAUCam images in the NB filter NB685 before the camera intervention (first and second images on the left) and after the camera intervention (third and fourth images). They show scattered light near the edges of the CCD, displaying a spatially varying amount of scattered light. The scattered light patterns change from before the intervention (two images on the left) to after (two images on the right). The scattered light pattern is also filter dependent. The images taken in each filter show their own distinctive patterns, meaning that the pattern depends on the filter used.

One way to quantify and model the scattered light is to create background pixel maps per NB. This is done with the following steps:

- i. *Select images*: Select a group of NB images from the same bands since they have the same scattered light pattern.
- ii. *Compute median*: For each of the images, compute the median background level in the central regions, μ_{BKG} , which are unaffected by scattered light.
- iii. *Estimate ratios*: Divide every image by its median to obtain a pixel ratio map.
- iv. *Mask sources*: Mask the images sources by masking all pixels above a given pixel ratio threshold.
- v. *Combine images*: Combine all individual pixels maps with a median to get a single scattered-light template (SLT) for all the selected images.

If the background were flat and followed Poisson statistics, all pixels in the ratio map should fluctuate along unity. However, if the image is affected by scattered light, the scattered-light templates in affected regions will have a value above unity. We can understand this ratio as approximately the percentage of extra light (scattered light) compared to the flat background. Notice that this model takes into account that scattered light depends on the amount of light falling on the CCD.

The procedure in step (v) can be written as

$$\text{SLT}(x, y) = \text{median}_j \left[\frac{I_j(x, y)}{\mu_{\text{BKG}}} \right], \quad (1)$$

where I_j is image j and the median is over the selected images (step i). To determine the amount of scattered light we can follow the previous procedure to step [iii]. This way we obtain normalized background images that should fluctuate around unity if they contain a flat background, but would have values above one if they are affected by scattered light.

The top panel of Fig. 2 shows some of these normalized images for the NB685 filter. It shows the background pixel value from side to side of the image before (black dashed line) and one after (orange solid line) the camera intervention. The plot shows an increasing background on the edges of the CCD before the camera intervention. After the intervention, the amount of scattered light is considerably reduced. Unfortunately, it is still present and thus needs to be accounted for.

We can use all the normalized background images in a given NB to create a general scattered-light template for that band (also splitting before/after the intervention). The bottom panel in Fig. 2 shows the resulting mean of each scattered-light templates (one per band) as a function of NB. The mean of the scattered-light templates gives information about the amount of scattered light in a given band. We can clearly see the effect of the intervention on the amount of scattered light, which is reduced.

2.4 Scattered-light templates as scattered light correcting method

If the scattered-light templates modelling is sufficiently accurate, it can be used to correct the scattered light on PAUCam images. Assuming that all the images from a given NB follow the same scattered light pattern scaled by the CCD sky background, a way of correcting scattered light would be

$$\tilde{I}(x, y) = I(x, y) - (\text{SLT}(x, y) - 1)\mu_{\text{BKG}}, \quad (2)$$

where we subtract from a given target image ($I(x, y)$) the scattered-light templates scaled by the mean background of such image (μ_{BKG}). Notice that instead of subtracting the scattered-light templates, we subtract the scattered-light templates without the flat sky

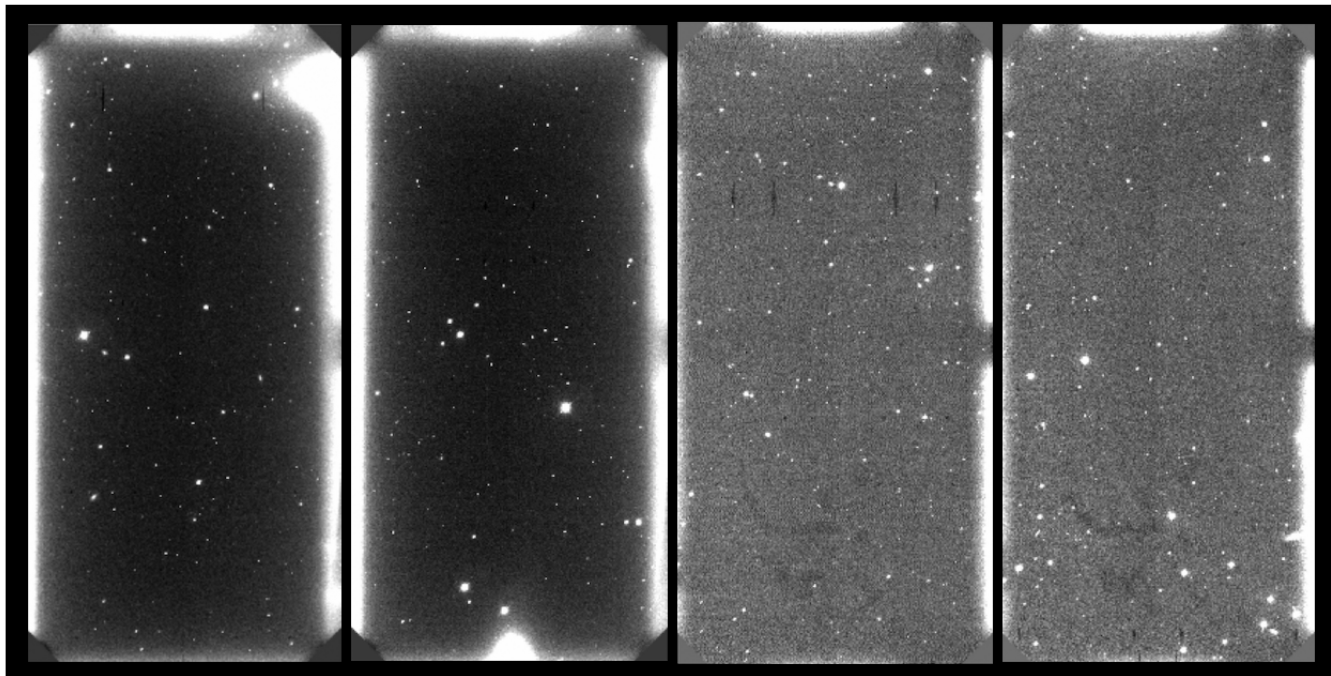


Figure 1. Images taken with the PAUCam, corresponding to the NB685 filter. *Left:* The first two images correspond to PAUCam images before the camera intervention. Notice that both exhibit the same scattered light pattern. *Right:* The two images on the right correspond to PAUCam images after the intervention. Again, both present the same scattered light pattern, but different to the first two images on the left. This shows the changes in scattered light patterns with the intervention.

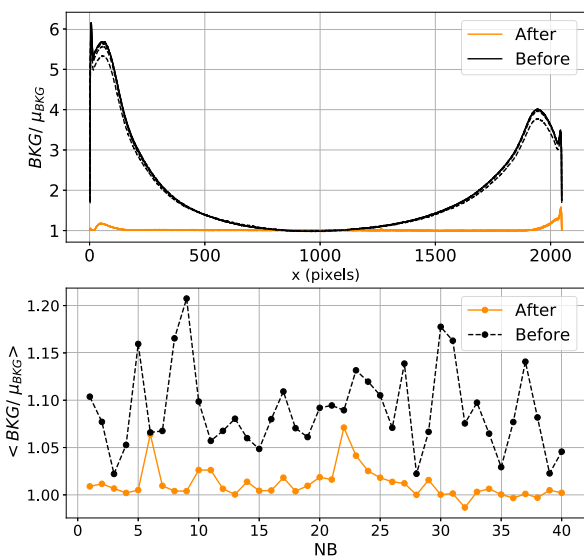


Figure 2. *Top:* Normalized background light content in each pixel as a function of the pixel position in the image for different images before (black dashed line) and after (orange solid line) the camera intervention. Each pixel value is divided by the mean background in the image. Regions without scattered light should fluctuate around unity. Regions affected by scattered light should be above unity. *Bottom:* Mean value of the normalized background curves considering all the images taken in that band, for the 40 narrow photometric bands.

background. This way, the regions without scattered light are barely affected.

Fig. 3 shows the original CCD image (left), after correcting with the scattered-light templates (middle) and the scattered-light

templates used for correction (right). Visually, the scattered light pattern in the original image (left) disappears after applying the scattered-light templates correction (middle). However, although the correction seems visually almost perfect, this method has a drawback. Even though scattered light follows approximately a pattern given a band, there might be fluctuations due to other external conditions. For example, the weather, Moon illumination, and other observing conditions may induce variations between different observations in an NB. To be more precise estimating the correction, one should create a template per band and per night, such that the observing conditions are similar. However, for creating a scattered-light templates per night, there might be an insufficient number of images to obtain an accurate modelling of the correction pattern. Bright stars also contribute to scattered light and this cannot be corrected with the scattered-light templates.

Fig. 4 shows the background level for a specific image in the NB685 filter before and after the correction with the scattered-light templates. In this case, the image is corrected without considering any split on night to generate the scattered-light templates. This means that all images, despite being observed on different night and with different observing conditions, are used to build the scattered-light templates. The image without correction displays large peaks at both edges and those are clearly corrected by the scattered-light templates. However, both sides of the CCD still show bumps that are caused by scattered light residuals.

3 BKGNET: A DEEP LEARNING BASED METHOD TO PREDICT THE BACKGROUND

In this section, we start by describing the BKGNET architecture. We then describe our training and test samples and describe the training process. As a reference, in Appendix A, we introduce the

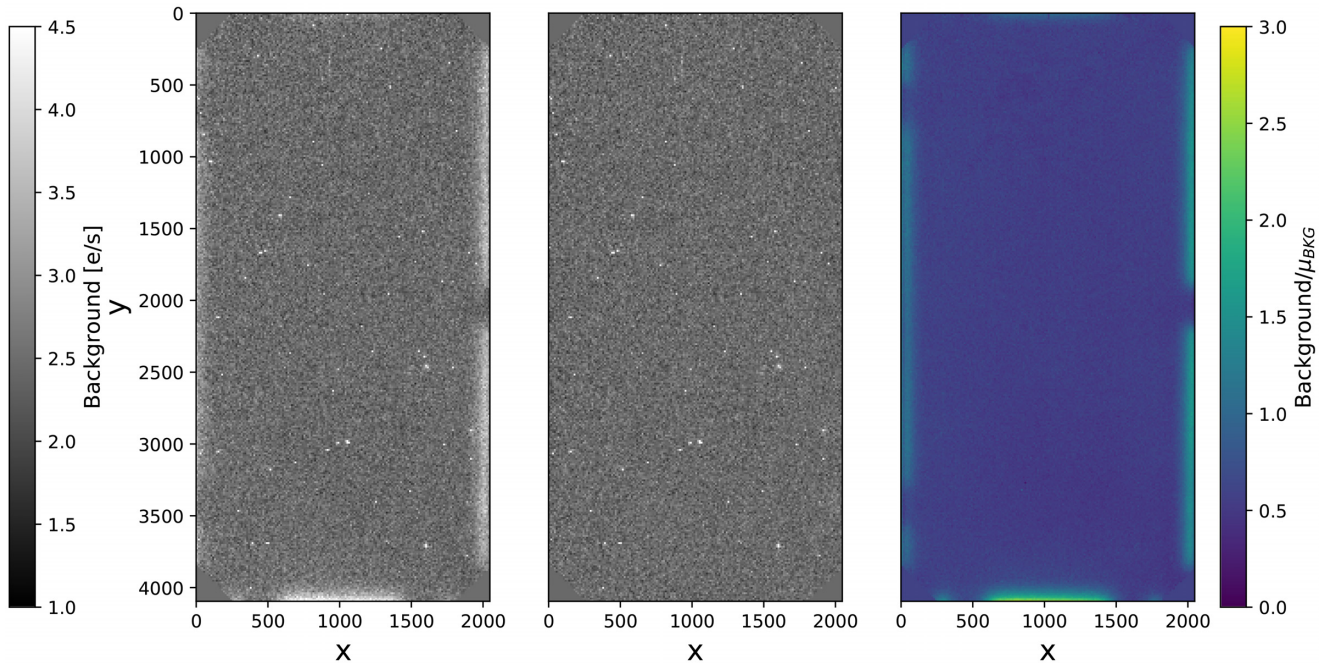


Figure 3. *Left:* Image taken in the NB685 filter showing a scattered light pattern on the edges. *Middle:* Previous image corrected with the scattered-light template. *Right:* The scattered-light template generated with equation (1) considering all images taken the same observation night as the original image.

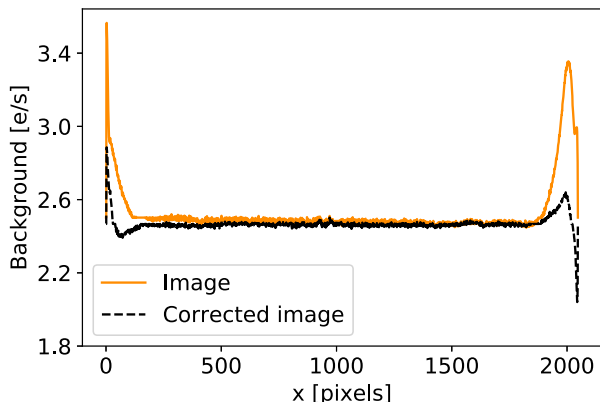


Figure 4. Background pixel values across the image. The original image (orange solid line) displays high peaks on the edges caused by scattered light. After correcting with the scattered-light templates (dashed black line) the peaks are reduced, but some residuals remain. The images are in e/s.

basics of deep learning and CNNs, together with some terminology definitions.

3.1 Neural network architecture

BKGNET³ is built using the PYTORCH library (Paszke et al. 2017). It has two main blocks: a CNN and a linear NN.

Fig. 5 shows the BKGNET architecture. The CNN block handles the information coming from the image itself, as the background we want to recover is encoded in the pixel values. The inputs are 120×120 pixels stamps containing the target galaxy in the centre. This

³<https://gitlab.pic.es/pau/bkgnet>

choice for the stamp size is a compromise between having enough pixels while keeping the computing requirements (memory, GPU) within reasonable limits.

As Fig. 5 shows, the CNN contains five blocks of convolutional layer (red layer), pooling layer (yellow layer), and batch normalization layer (blue layer). In each convolutional layer, the network learns to gradually capture different features in the image. The first layers learn low-level features, like edge detection, while having more layers leads the network to learn high-level features (Zeiler & Fergus 2013).

The scattered light model depends on parameters that are not encoded in the stamps. These are the position of the stamp in the original image, the NB used to observe the galaxy and a before/after intervention flag informing the network when the galaxy was observed. We also include the target galaxy magnitude from a reference catalogue, as it contains information about the number of pixels that are affected by the galaxy. To help the network learning the scattered light patterns, the previously mentioned parameters are provided to the linear NN, together with the CNN's output. The magnitude of the galaxy and the coordinates of the image are added as fixed parameters per stamp.

The NB filter and the intervention flag are discrete variables with 40 possible values for the band (1–40) and two for the intervention flag (0/1). The combination of the two is, however, directly related to the scattered light pattern, and we can effectively convert these two discrete variables into a single one with values from one to 80. We add the band and intervention information using an embedding. The embedding replaces each combination with 10 trainable parameters. Before embedding the band and intervention information, BKGNET learns to encode each scattered light pattern using the 10 numbers that should best characterize the pattern. Therefore, the linear network receives the output from the convolution layers, two galaxy coordinates, the magnitude, and 10 numbers from the embedding.

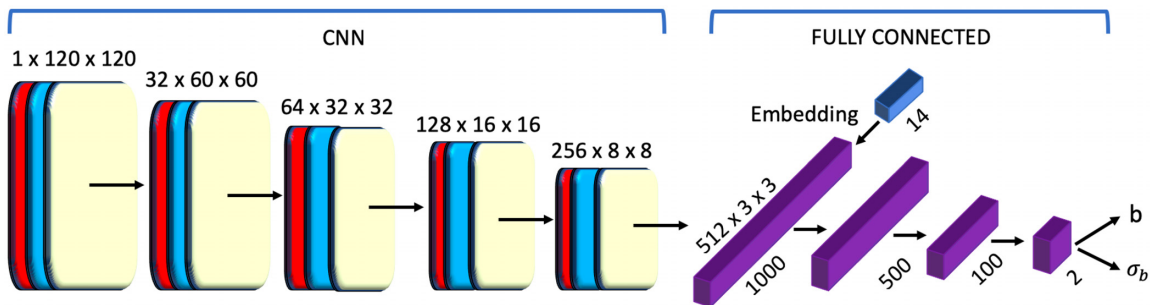


Figure 5. BKGnet scheme: The first set of layers corresponds to a CNN to which one inputs the images. The CNN output, together with extra information are input to a linear NN. The numbers on each of the convolutional layers represent the layer’s dimension. The first number corresponds to the number of channels. The second and third numbers are the dimension of the stamp in that layer.

3.2 Data: training and test samples

BKGNET’s inputs are stamps with the target galaxy in the centre. However, to train the network we use empty CCD positions, meaning regions where there are not target sources. This way, we can estimate the ground truth background value at the central CCD region (where there is supposed to be a target galaxy) and train the network to recover this value. The estimation of the true background values used as training sample labels is done by computing the mean background inside a circular aperture of a given fixed radius in the central region of the stamp. Therefore, these measurements have an associated uncertainty that directly depends on the aperture radius. Assuming that the background is purely Poissonian, then

$$\sigma_{\text{label}}^2 = \frac{N_a b}{t_{\text{exp}}}, \quad (3)$$

where t_{exp} is the exposure time, b is the background estimated as the mean of the pixels inside the aperture, i.e. the background label, and N_a is the number of pixels inside the circular aperture, directly related with the choice of aperture radius. We have fixed it to eight pixels as a balance between the error of the ground truth measurement and having a precise background measurement in the exact galaxy location. To select empty stamps for the training sample we identify sources by cross-correlating the sky coordinates of a given image location with the sky coordinates of the sources in the COSMOS catalogue (Laigle et al. 2016).

In any deep learning algorithm, the training and test samples should be as similar as possible. In our case, our training sample does not contain target galaxies whereas the test sample does. We therefore add simulated galaxies at the centre of the empty training stamps. The simulated galaxies are constructed with parameters based on PAUS data: the Sersic profile parameters, r_{50} , I_{50} , and the magnitude in the i band. The Sersic profile describes the surface brightness profile (I) of a galaxy. The radius r that contains 50 per cent of the light intensity (I_{50}) is r_{50} . These simulated galaxies may differ from the real ones. For this reason, we mask the central 16×16 pixels in both the training and test samples. Although the simulated galaxy is now masked, it is still important to include it, as for some profiles the galaxy light extends outside the masked region. Without the simulated galaxy, BKGNET fails on testing bright sources. As the label is estimated in an eight-pixel-radius aperture, 16×16 pixels is the minimum area such that the network does not see the pixels used for the estimation of the ground truth background.

We normalize the stamp before feeding the network. There are different ways of doing this. We apply a normalization stamp by stamp, where we use the mean and the standard deviation of each

stamp to normalize it. We have chosen this normalization method as it performs better on our data set.

We use all the PAUCam images in the COSMOS field to train and validate the network. We have 4928 PAUCam images before the intervention and 4821 after (see Section 3.2 for details). For each of them, we sample around 40 stamps per CCD image, giving a total of around 400 000 stamps. We use 90 per cent of them for training and the remaining 10 per cent for validation.

3.3 Training process and loss function

Supervised deep learning algorithms are trained comparing the true value with the algorithm’s prediction. The agreement between the prediction and the true value is evaluated with a loss function. The choice of loss function depends on the kind of problem one is facing, (e.g. classification, regression). A typical loss function for classification problems is the cross-entropy loss, whereas in regression problems the mean squared error is commonly used. With BKGnet we want the network to associate an uncertainty to each prediction. In supervised deep learning, there are some methods based on Bayesian statistics that deal with uncertainties associated with the predictions (e.g. Kendall & Gal 2017; Kendall, Gal & Cipolla 2017).

The method we use assumes that the distribution $p(\mathbf{y}|f^w(\mathbf{x}))$ is Gaussian, where \mathbf{y} are the background label values, \mathbf{x} are the inputs, and $f^w(\mathbf{x})$ are the network background predictions. Therefore, the loss function is defined

$$\text{Loss} = -\log p(f^w(\mathbf{x})) = \frac{(f^w(\mathbf{x}) - y)^2}{\sigma^2} + 2 \log \sigma. \quad (4)$$

In this way, we train the network to provide both, the background prediction $f^w(\mathbf{x})$ and its associated error σ . Notice that the second term on the right-hand side prevents the network from predicting a large error that minimizes the first term.

With the loss function in equation (4), the network provides an error on the quantity $f^w(\mathbf{x}) - y$, which has an associated uncertainty $\sigma_{\text{pred}}^2 + \sigma_{\text{label}}^2$. The error on the prediction is therefore

$$\sigma_{\text{pred}} = \sqrt{\sigma_{\text{bkgnet}}^2 - \sigma_{\text{label}}^2}, \quad (5)$$

where σ_{bkgnet}^2 is the error provided by the network and σ_{label}^2 is the error of the background label. The error of the background label is defined in equation (3).

BKGNET is trained in 60 epochs with a batch size of 100 stamps using the ADAM optimizer (Kingma & Ba 2014) and a learning rate of 10^{-5} (see Appendix A for terminology). The training takes about two hours using an NVIDIA TITAN V GPU.

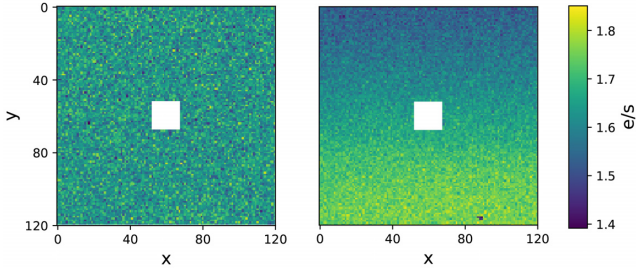


Figure 6. Simulated stamps. On the left, a stamp with a flat background. On the right a stamp with a background with a gradient.

4 TESTING BKGNET ON SIMULATIONS

We test the performance of BKGNET with simulated data. We study how well we can predict the background with the network and explore what data are needed and how these data need to be treated before feeding the network. Throughout the rest of the paper, we compare the BKGNET predictions to those obtained by calculating the background inside an annulus around the target source before and after correcting the image with the scattered-light templates.

4.1 Simulated PAUCam background images

The final simulated image $I_{\text{sim}}(x, y)$ can be expressed as

$$I_{\text{sim}}(x, y) = A \times \frac{t_{\text{exp}} \times \text{SLT}(x, y) + P(t_{\text{exp}} \times \text{SLT}(x, y))}{t_{\text{exp}}}, \quad (6)$$

where SLT is the scattered-light template used to generate the image. In this way, the simulated image shows the same scattered light pattern as the PAUCam images. To generate the Poisson noise, we first multiply with the exposure time (t_{exp}) to convert the template from e/s to electrons. Additionally, the template is scaled with a factor A to simulate a wide range of background levels. Finally, we generate a realization of Poisson sky noise $P(\cdot)$ that we add to the image and convert it back to e/s.

Fig. 6 illustrates the effects of scattered light in our simulated stamps. On the left, we show a simulated stamp image with a flat background pattern; and, on the right, with a gradient on the background. This gradient is caused by scattered light (see Figs 1 and 3). Both stamps show a central 8×8 pixel masked region, blocking the light of the galaxy.

4.2 BKGNET predictions on simulations

Throughout this section, we train and test on stamps without target galaxies (empty positions). This allows us to test whether it is possible to predict the background with this network's assembly. We also fix the band we are testing to the NB685 filter after the camera intervention. This choice is a compromise between having a considerable amount of scattered light without being completely dominated by it. Before the intervention, the amount of scattered light in some of the CCD images is very large and might not be an adequate choice to test the network. On the other hand, after the intervention, some of the CCDs barely contain scattered light, and those would not be a good choice either. The NB685 filter contains a considerable amount of scattered light and therefore it is a representative example. We do not need to simulate all bands, as here we only want to test the viability of the scattered light prediction with BKGNET and to have a better understanding of the network's behaviour. To quantify the background prediction

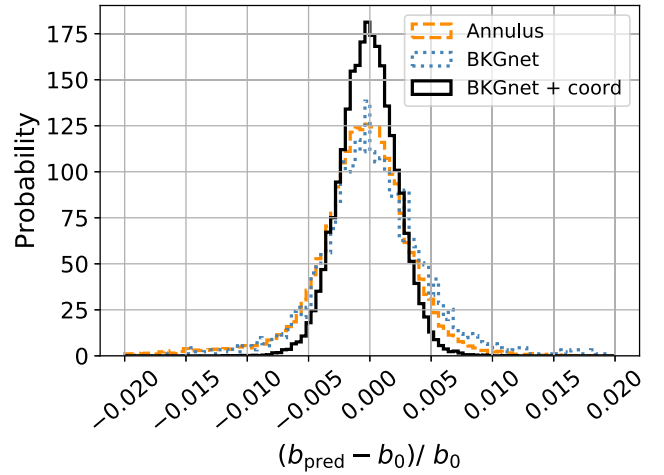


Figure 7. Relative error distributions for the BKGNET (green without coordinate information and orange with coordinate information) and the annulus predictions. b_0 is the background label and b_{pred} is the background prediction, either for the annulus or for BKGNET.

accuracy, we use

$$\sigma_{68} \equiv 0.5 \left(b_{\text{quant}}^{84.1} - b_{\text{quant}}^{15.9} \right), \quad (7)$$

with quantiles set to 84.1 and 15.9 percentage values. This quantity, σ_{68} , is equivalent to a 1σ error for a normal distribution, but it is less affected by outliers.

Fig. 7 compares the accuracy with which BKGNET predicts the background to the PAUS default approach. As described before, PAUS estimates the background computing the median inside an annulus centred on the galaxy after the pixel values have been σ -clipped. The plot shows the relative error distribution of the predictions for both methods. We have tested BKGNET with and without embedding the image coordinates of the galaxy. The BKGNET performance improves significantly with the coordinate information. This is not surprising because the amount of scattered light depends on the CCD position (see Section 2). Although scattered light is encoded in the image, the CCD position also includes essential information for the prediction. The network might need it to create something similar to the scattered-light template. BKGNET achieves a $\sigma_{68} = 0.0038$ with information coming only with the stamps. Including the coordinate information, this improves to $\sigma_{68} = 0.0022$. Therefore, the network improves by 70 per cent with the coordinates embedding. The default background estimate shows tails on both sides of the distribution, and yields $\sigma_{68} = 0.0033$, which means BKGNET improves the estimate by 42 per cent.

Fig. 8 shows the spatial background map (left) and the relative error on the prediction of this map with the annulus background predictions (third panel) and the BKGNET background predictions (second panel). The precision is lower at the edges of the CCD for the annulus-based method, where scattered light is present. This indicates that the tails in Fig. 7 are caused by scattered light. On the other hand, one can see that BKGNET is able to account for the presence of scattered light.

The right-hand panel of Fig. 8 shows the background reconstruction using a k-Nearest Neighbors (kNN) algorithm. In addition to the CNN and annulus methods, we have also tested the kNN (Cover & Hart 2006), Support Vector Regression (SVR) (Drucker et al. 1996), Random Forest (RF) (Breiman 2001), and NN techniques. We used the SCIKIT-LEARN implementations (Pedregosa et al. 2011)

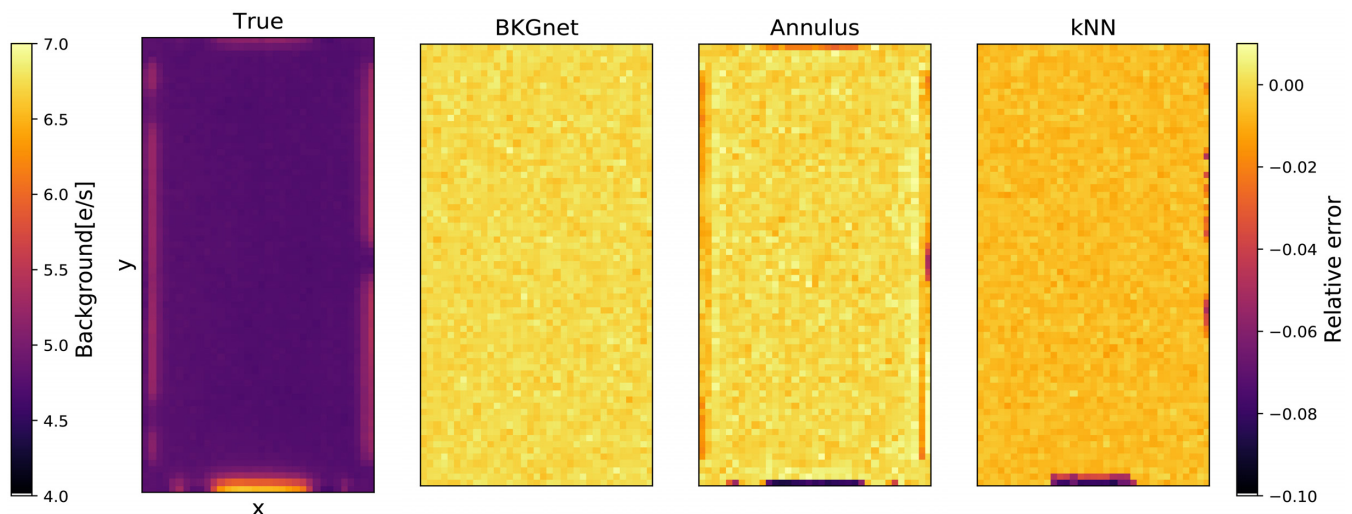


Figure 8. *Left:* CCD reconstruction with the true background values used to train the network. We sample these background values consecutively and we reconstruct the original image by placing each value in the position it was sampled from. *Second:* Accuracy on the background prediction with BKGNET in the different image positions. We can see there are not spatial patterns. *Third:* Accuracy on the background prediction with the annulus in the different image positions. We can see there are not spatial patterns. *Right:* Accuracy on the background prediction with a kNN in the different image positions.

to run the kNN, RF, and SVR algorithms. For these tests, unlike the CNN, which can handle images, we input the embedded information and the median pixel value. As shown in Fig. 8, the kNN measures the background with less accuracy than BKGNET and the annulus method in the flat background regions. The background measurements are biased by about 3 percent, which is six times larger than the relative errors (0.5 percent). The prediction also shows patterns on the edges with an error six times higher than those on flat regions. In contrast, for BKGNET the precision only degrades by a factor of 1.2 when we compare the centre with the border positions. Concerning the other methods, the NN provides better predictions than the kNN, although it increases σ_{68} by a factor of 2.5 with respect to BKGNET. It also shows patterns on the edges with four times higher errors than in flat background regions. Further, the errors of the RF and SVR algorithms are a factor of six and four, respectively, higher than those of the BKGNET method, rendering these methods too imprecise.

In the following sections, we test the background estimation method on real PAUCAM images. Based on the performance on simulations, we will only consider the BKGNET and the annulus methods.

5 BKGNET ON PAUCAM IMAGES

In the previous section, we have shown that BKGNET is able to accurately predict strongly scattered light backgrounds on simulated blank images including only the background. However, in real PAUCAM data other complications such as cosmic rays, electronic cross-talk, read-out noise, and dark current may affect the performance on the estimation of the background. Moreover, correlations between pixels might be introduced during the data reduction process. To examine the impact of these real-life effects we use actual PAUCAM images. To assess the accuracy of our measurements, we test network on empty stamps, i.e. without target galaxies.

We will use all the images available in COSMOS, splitting the data into those obtained before the camera intervention (in 2016A) and after. This split yields 4928 images and 4821 images for

before and after the intervention, respectively. As these numbers are similar, we can easily balance the number of stamps before and after the intervention in our training sample. Although the training sample does not contain target galaxies in the centre, sources might be placed in other stamp’s positions. To avoid outliers in the training set, e.g. a stamp with a bright star covering most of the background or a bright object too close to the centre, we filter the training stamps based on the maximum pixel value. All stamps with a pixel containing more than 100 000 counts are excluded from the training sample.

We also exclude 40 images from each subsample before training the network. These 80 images are not used to train the network, but are kept to test it. This is important, as we need to test the network on images it has never seen before. To generate the test set, instead of sampling randomly from the CCDs, we sample stamps consecutively in intervals of 60 pixels. This ensures that we test all CCD regions, including regions affected by scattered light.

Fig. 9 shows the results when we use BKGNET to predict the background on PAUCAM images in empty regions. We also show results when the background is estimated using an annulus, and when we first correct the background variations using a scattered-light template (‘annulus + sky’). Fig. 9 shows the value of σ_{68} (equation 7) of the relative error distribution on the prediction for the 40 different bands. Because we are using the relative error, the comparison between the results before and after the intervention is not representative, as the background levels are different. For instance, in the first filter tray (NB455–NB515), the background before the intervention is between three and five times higher than after.

We focus first on the results before the camera intervention (left-hand panel in Fig. 9). Images before the camera intervention contain more scattered light than those after (see Fig. 2). This makes the scattered-light template modelling more unstable than the modelling of images after the intervention. We find that correcting with the scattered-light template does not improve the annulus method result in every band. In the bluest NBs, i.e. those with the highest amount of scattered light, the scattered-light template seems to decrease

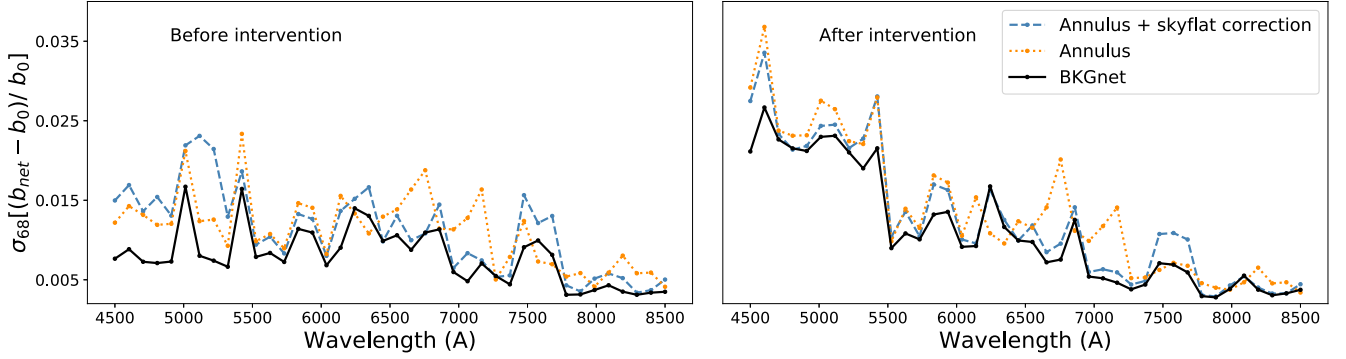


Figure 9. σ_{68} of the relative error in the background prediction for the 40 NBs. *Left:* Before the intervention. *Right:* After the camera intervention. In almost all cases BKGNET performs better than the default approach that employs an annulus to estimate the background.

Table 1. Average σ_{68} of the relative error in the background prediction across all the bands for BKGNET trained before and after the camera intervention. We list the results for the data sets without filtering out stamps affected by sources (‘sources’), and if we remove these (‘filtered’).

	Before		After	
	Filtered	Sources	Filtered	Sources
Annulus	0.011	0.011	0.014	0.014
+ SLT	0.011	0.011	0.011	0.013
BKGnet	0.008	0.008	0.011	0.011

the accuracy of the background prediction. On the other hand, BKGNET improves the accuracy compared to the other two methods, especially on the bluest filter tray. On average considering all bands, the network reduces the σ_{68} by 37 per cent compared to the scattered-light template and up to 50 per cent if we only consider the eight bluest NBs.

If we consider the results after the camera intervention (right-hand panel of Fig. 9), we see that the scattered-light template improves the annulus method prediction in all the bands. This is expected from the top panel in Fig. 2, which shows that scattered light trends are stable after the camera intervention. Before the intervention the scattered-light template fails in the bluer bands, which no longer happens after the camera intervention. Nevertheless, BKGNET performs even better: on average, after the intervention it achieves an 18 per cent improvement compared to the scattered-light template correction.

Table 1 lists the average value of σ_{68} of the relative error in the background prediction for the three methods: annulus, annulus + scattered-light template, and BKGNET. When training and testing, we first exclude the stamps with a maximum pixel value above 100 000. By doing this, we avoid stamps with very bright nearby sources that might bias the prediction. To examine the impact of this step, we also list the results when contaminated stamps are included (‘sources’) in Table 1. These results show that the filtering does not make a difference before the intervention, but the performance improves somewhat for the correction that uses the scattered-light templates. The small difference suggests that scattered light is the main source of bias. For the images without bright sources taken after the intervention, BKGNET and the scattered-light template give the same σ_{68} . Therefore, it is possible that BKGNET learns the underlying behaviour of scattered light in a similar way as the scattered-light template. However, as the network also sees the stamp, the correction it infers is more flexible than applying a

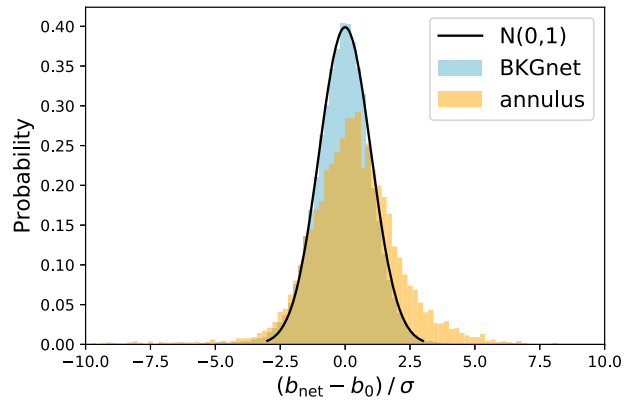


Figure 10. The distribution of $(b_{\text{net}} - b_0)/\sigma$, where b_{net} is the background prediction and b_0 the true background. σ is the uncertainty in the prediction. We expect the distribution to be a Gaussian centred on zero with unit variance. We show the distribution for the annulus (orange) and BKGNET (blue) predictions.

scattered-light template. This indicates that BKGNET is able to learn how to estimate the background in the presence of other artefacts (e.g. sources or cosmic rays).

BKGNET also provides an estimate for the uncertainty associated with the background prediction. To test the accuracy of this estimate, we use the empty stamps and study the distribution of $(b_{\text{net}} - b_{\text{true}})/\sigma$, where b_{net} and σ are the network predictions. If the errors are correct, this distribution should be a Gaussian with zero mean and unit variance.

Fig. 10 shows the theoretical Gaussian we should recover and the measured distributions for the annulus and BKGNET predictions. The BKGNET results fit the theoretical Gaussian, which means that our errors are robust. In contrast, the annulus predictions underestimate the uncertainties by 47 per cent. Therefore, BKGNET provides a more reliable estimate of the uncertainty in the background determination.

6 BKGNET VALIDATION

The results presented in Section 5 show that, compared to the annulus-based methods, BKGNET yields better background estimates (see Fig. 9 and Table 1). It also provides accurate estimates for the associated uncertainty. However, these tests were done on stamps without galaxies. Here, we increase the realism of

the problem and quantify the performance of BKGNET at galaxy positions.

6.1 Generating the PAUS catalogue with BKGNET predictions

We use BKGNET to estimate the background for galaxies in the COSMOS field. We compare the results to those from the PAUDM catalogue, which uses an annulus to determine the background. These catalogues contain around 12 million flux measurements, approximately half of them done on images taken before the intervention and the other half on images taken after the intervention. The galaxy fluxes are obtained subtracting the background from the PAUS raw signal measurements,

$$F = S - N_a b, \quad (8)$$

where F is the net galaxy flux, S is the total signal measured inside the aperture, N_a is the number of pixels inside the aperture, and b is the predicted background per pixel. When the background is estimated with an annulus, the error on the net flux is

$$\sigma^2 = (S - b) + N_a \sigma_b^2 + N_a^2 \left(\frac{\pi}{2}\right) \frac{\sigma_b^2}{N_b}, \quad (9)$$

where b and σ_b are the background and the background error in that region and N_b is the number of pixels inside the annulus. The $\pi/2$ factor arises from that fact that we use the median of the pixels inside the annulus instead of the mean.⁴

For BKGNET, the error on the galaxy flux is

$$\sigma^2 = (S - b) + N_a (b + RN^2) + N_a^2 \sigma_b^2, \quad (10)$$

where RN is the read-out noise.

Equations (9) and (10) reflect the differences in the flux uncertainty when the background is measured with an annulus or with BKGNET. In general, there are three main contributors to the flux uncertainty: the uncertainty in the net galaxy flux, the uncertainty in the background estimate, and the uncertainty introduced by the background subtraction. For both background estimation methods, we assume that the uncertainty in the net galaxy flux is captured by shot noise. For BKGNET, the background uncertainty is also described by shot noise (equation 10), but we add a read-out noise contribution to the background error. For the PAUDM measurements, the background uncertainty is given by the mean variance per pixel (equation 9). Therefore, for PAUDM, this term should also account for other error contributions besides shot noise. The third terms in equations (9) and (10) are the contributions from background subtraction uncertainties. In PAUDM, this is determined by the subtraction of a background measured in the annulus. In contrast, in equation (10), we use the uncertainty provided by the network within the aperture where the flux is estimated.

6.2 Validating the catalogues

The estimation of the background using an annulus is a viable method when the background is flat. In PAUS data, scattered light only affects objects near the edges of the images. Hence, for most of the galaxies in PAUS data the background should be (approximately) flat and we should not expect large differences between the BKGNET and the PAUDM catalogues. Comparing the fluxes estimated with equation (8), we find a 2 per cent difference between the two approaches. On the other hand, the uncertainties

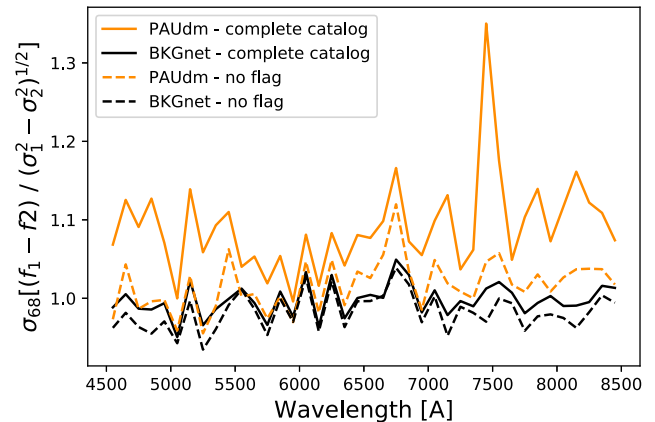


Figure 11. BKGNET validation with the duplicates distribution test. We plot the width of the distribution defined in equation (11) as a function of wavelength for the catalogue generated with BKGNET (black line) and the current PAUDM catalogue (orange line). The dashed line corresponds to the results excluding all objects flagged in PAUDM. The solid line includes all objects.

estimated with BKGNET (equation 10) are 4 per cent lower than for PAUDM.

We need to determine which catalogue provides better photometry estimates. To do so, we use the fact that PAUCam takes multiple observations of the same object in all NB filters. We can compare different exposures of the same object, which should be comparable once the background noise is subtracted. This is formulated as

$$D \equiv \frac{(e_1 - e_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)}}, \quad (11)$$

where e_i are different exposures of the same object and σ_i the associated uncertainties. The distribution of D should be a Gaussian with unit variance if the photometry is robust and the errors are properly accounted for. We call this the duplicates test.

Fig. 11 shows the results of the duplicates test as a function of wavelength. We estimate $\sigma_{68}[D]$ (equation 11) for each NB with the BKGNET (black line) and PAUDM (orange line) catalogues. It is possible to flag photometric outliers based on an ellipticity parameter to detect strongly varying backgrounds. The dashed lines in Fig. 11 show the results when we exclude such flagged objects. Dropping flagged objects does not significantly change the measurements for BKGNET, but we see a clear improvement for the PAUDM measurements. The improvement is particularly prominent for the NB755 filter (at 7500Å), which is affected by telluric absorption of O₂ in the atmosphere. Interestingly, BKGNET seems to know how to deal with these objects, showing that BKGNET is more robust towards various sources of bias, not only scattered light. When we consider all NBs, we find $\langle \sigma_{68}[D] \rangle = 1.00$ for BKGNET, which is what we would expect for correct photometry. On the other hand, the current PAUDM catalogue yields $\langle \sigma_{68}[D] \rangle = 1.10$, i.e. it overestimates the uncertainties.

The measurement uncertainties should depend on the brightness of the source. To explore this we show $\sigma_{68}[D]$ as a function of the Subaru i_{Auto} magnitude in Fig. 12. In the PAUDM catalogue there is a strong trend with magnitude. At the bright end, the fluxes differ by more than 20 per cent compared to the expectation. This trend disappears when we predict the background and uncertainties with BKGNET. To explore the origin of the trend further we used the background prediction from BKGNET but the errors from the

⁴<http://wise2.ipac.caltech.edu/staff/fmasci/ApPhotUncert.pdf>

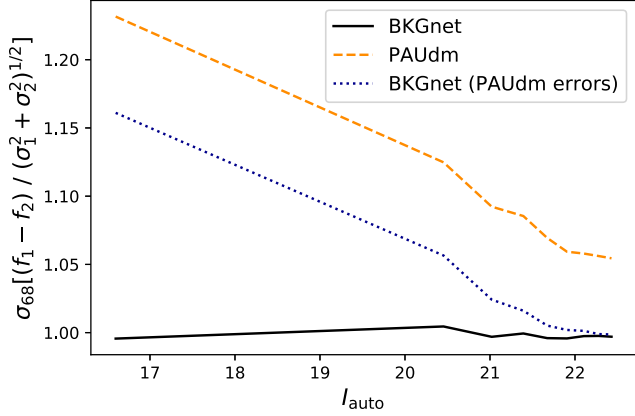


Figure 12. BKGNET validation with the duplicates distribution test. We plot the width of the distribution defined in equation (11) as a function of i_{auto} in the Subaru i band for the catalogue generated with BKGNET (black solid line), the current PAUDM catalogue (orange dashed line) and a mixed catalogue with the predictions from BKGNET and the errors from PAUDM (blue dotted line).

annulus method. As the blue dotted line in Fig. 12 shows, we find the same trend with magnitude. This implies that it is caused by the estimated uncertainties for the annulus method. Moreover, the blue dotted line lies below the PAUDM line. The only difference between these two curves is the background value prediction (not the error). Therefore, the predictions with PAUDM are more accurate than those with the annulus method.

To further validate the BKGNET catalogue, we run BCNz2 (Eriksen et al. 2019) using the fluxes determined using BKGNET. For this test, we exclude the objects flagged in the PAUDM catalogue, in order to use exactly the same objects as in (Eriksen et al. 2019). However, as shown in Figs 11 and 12, we do not need to exclude these objects. The photo- z s are compared to secure spectroscopic estimates from zCOSMOS DR3 (Lilly et al. 2007) with $i_{\text{AB}} < 22.5$. We split the sample based on a quality parameter defined as:

$$Q_z \equiv \frac{\chi^2}{N_f - 3} \left(\frac{z_{\text{quant}}^{99} - z_{\text{quant}}^1}{\text{ODDS}(\Delta z = 0.01)} \right), \quad (12)$$

where $\chi^2/(n_f - 1)$ is the reduced chi-squared from the template fit and the z_{quant} are the percentiles of $(z_{\text{photo}} - z_{\text{spec}})/(1 + z_{\text{spec}})$. The ODDS is defined as

$$\text{ODDS} \equiv \int_{z_b - \Delta z}^{z_b + \Delta z} dz p(z), \quad (13)$$

where z_b is the mode of the $p(z)$ and Δz defines a redshift interval around the peak. In PAUS, a galaxy is considered an outlier if

$$|z_{\text{photo}} - z_{\text{spec}}| / (1 + z_{\text{spec}}) > 0.02. \quad (14)$$

Notice that this outlier definition is very strict. In broad-band photometry, a common outlier definition is $|z_{\text{photo}} - z_{\text{spec}}| > 0.15(1 + z_{\text{spec}})$, e.g. Ilbert et al. (2006) and Bilicki et al. (2018).

Table 2 lists the outlier rate and the photometric redshift precision obtained with BCNz2 for the two catalogues. To quantify the redshift precision we use σ_{68} (equation 7). The photometric redshift precision does not improve significantly between the two catalogues, but we find a reduction in the outlier rate. If we consider the complete sample (100 per cent) this improvement is small. This might be because in the full sample the outliers are dominated by photo- z outliers, rather than outliers on the photometry itself.

Table 2. Photo- z outlier rate and accuracy obtained with BCNz2 for the BKGNET and the PAUDM catalogues. The percentages correspond to the samples selected by the photo- z quality parameters Q_z .

Percentage	Outlier percentage		$10^3 \sigma_{68}$	
	BKGNET	PAUDM	BKGNET	PAUDM
20	3.5	5.4	2.0	2.1
50	3.8	5.1	3.6	3.7
80	10.4	11.3	5.8	6.0
100	16.7	17.5	8.4	8.6

However, if we cut using the Q_z parameter to get the best 20 per cent and 50 per cent of the sample, we notice that the outlier rate reduces significantly. These outliers should be dominated by photometric outliers. For the best 50 per cent of the sample we reduce the number of outliers by 25 per cent, whereas for the best 20 per cent of objects this improvement rises to 35 per cent. This shows once more that BKGNET is a statistically accurate method that is also robust.

7 CONCLUSIONS

Imaging surveys need accurate background subtraction methods to obtain precise source photometry. We have developed a deep learning method to predict the background for astronomical images that are affected by scattered light. The algorithm has been developed to predict the background on images taken with PAUCam. The edges of PAUCam images are affected by scattered light (see Fig. 1), especially in the bluer bands. In 2016, the camera was modified to reduce the amount of scattered light. While the amount of scattered light decreased drastically, PAUCam images still contain a significant amount of scattered light (see Fig. 2).

For each band, the scattered light follows the same spatial pattern within the CCD and scales approximately linear with the background level. We have constructed scattered-light templates and background pixel maps by combining images taken with the same NB and normalized by their background level. These scattered-light templates show the scattered light variation across the CCD and can be used to correct for scattered light (see Fig. 3). Nevertheless, background fluctuations due to external conditions (e.g. Moon, seeing, airmass) can trigger differences on scattered light from night to night. To accurately correct scattered light with scattered-light templates, we would need to generate a scattered-light templates per NB and night. However, even then, fluctuations during the night or a small number of available images in a given band can lead to inaccurate corrections.

We therefore developed BKGNET, a deep-learning-based algorithm that predicts the background and its associated uncertainty behind target sources accounting for scattered light and other distorting effect. BKGNET consists of a CNN followed by a linear NN (see Fig. 5). In the training set we use empty stamps, i.e. without a target galaxy, so that we can estimate the true background and use it for training. We need to simulate target galaxies in the training sample before masking the central region, otherwise the network fails when applied to bright and large sources.

We first tested the predictions on PAUCam empty stamps, i.e. without target galaxies. For data taken before the intervention, BKGNET improves over the scattered-light templates + annulus prediction by 37 per cent. The scattered-light templates correction fails in many of the bands, specially on the bluer filter tray, which is affected the most by scattered light (left-hand panel of Fig. 9). For data taken after the intervention, BKGNET improves over the

scattered-light templates + annulus prediction by 17 per cent (right-hand panel of Fig. 9).

BKGNET also predicts the uncertainty associated with the background prediction. For that, we use the log likelihood of a Gaussian centred at the background true value as loss function (equation 4). To validate BKGNET, we test on empty positions and estimate the difference between the prediction and the background level, divided by the estimated uncertainty. For the annulus method, we find that the errors are underestimated by 47 per cent (Fig. 10). On the other hand, with BKGNET this quantity is normally distributed around zero with unit variance, showing that the uncertainties are correctly estimated.

We generated a PAUS catalogue for the COSMOS field using BKGNET to predict the background. To validate the catalogue, we took advantage of having multiple measurements of the same object. The resulting distribution of differences in flux measurements should be a Gaussian of unit variance (equation 11). The results demonstrate that BKGNET improves the photometry with respect to the current background subtraction algorithm. We test the performance for the full catalogue and a catalogue where we exclude all objects flagged in the current catalogue version. When excluding flagged objects, we find very similar results with the BKGNET catalogue and the current catalogue. However, when testing the full catalogue, we find a large improvement for BKGNET. It specially improves the results in a region with high atmospheric absorption, demonstrating that it is more robust against sources of bias while still being statistically accurate. It also removes a strong systematic trend with *i*-band magnitude, which disappears when the uncertainties are estimated with the network.

Finally, as the aim of PAUS is to provide accurate redshifts for large samples of galaxies, we have run the BCNZ2 code using the BKGNET catalogue. BKGNET reduces the outlier rate by a 25 per cent and 35 per cent, respectively, for the best 50 per cent and 20 per cent photo-*z* samples, while the accuracy is not affected.

With BKGNET we have optimized the background subtraction task, one of the image processing steps in photometric surveys that can improve the redshift estimation and classification of galaxies. Deep learning algorithms that predict these quantities directly from images have to subtract the background internally. Therefore, the understanding from BKGNET will also help to optimize such deep learning algorithms. Although the network has been tested with PAUCam images, the concept should also be applicable to future imaging surveys as *Euclid* and LSST.

ACKNOWLEDGEMENTS

Funding for PAUS has 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) and University College London. The PAUS participants from Spanish institutions are partially supported by MINECO under grants CSD2007-00060, AYA2015-71825, ESP2015-88861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IECC and IFAE are partially funded by the CERCA program of the Generalitat de Catalunya. The PAU data center 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. CosmoHub has been developed by PIC and was partially funded by the 'Plan Estatal de Investigación Científica y Técnica y de Innovación' program of the Spanish government.

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776247. AA is supported by a Royal Society Wolfson Fellowship. MS has been supported by the National Science Centre (grant UMO-2016/23/N/ST9/02963).

REFERENCES

- Abbott T. M. C. et al., 2018, *ApJS*, 239, 18
 Alexander S., Gleyzer S., McDonough E., Toomey M. W., Usai E., 2019, preprint (arXiv:1909.07346)
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
 Bijaoui A., 1980, *A&A*, 84, 81
 Bilicki M. et al., 2018, *A&A*, 616, A69
 Breiman L., 2001, *Mach. Learn.*, 45, 5
 Cabayol L. et al., 2019, *MNRAS*, 483, 529
 Carrasco-Davis R. et al., 2019, *PASP*, 131, 108006
 Casas R. et al., 2012, in Holland A. D., Beletic J. W., eds, Proc. SPIE Conf. Ser., Vol. 8453, High Energy, Optical, and Infrared Detectors for Astronomy V. SPIE, Bellingham, p. 845326
 Casas R. et al., 2016, in Evans C. J., Simard L., Takami H., eds, Proc. SPIE Conf. Ser., Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, p. 99084K
 Castander F. J. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser., Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84466D
 Cover T., Hart P., 2006, *IEEE Trans. Inf. Theor.*, 13, 21
 D'Isanto A., Polsterer K. L., 2018, *A&A*, 609, A111
 Drucker H., Burges C. J. C., Kaufman L., Smola A., Vapnik V., 1996, Support Vector Regression Machines, Proceedings of the Neural Information Processing Systems Conference, Denver, Colorado
 Eriksen M. et al., 2019, *MNRAS*, 484, 4200
 Fluri J., Kacprzak T., Refregier A., Amara A., Lucchi A., Hofmann T., 2018, *Phys. Rev. D*, 98, 123518
 Gaia Collaboration, 2018, *A&A*, 616, A1
 George D., Huerta E. A., 2018, *Phys. Lett. B*, 778, 64
 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
 Ivezić, Ž. et al., 2019, *ApJ*, 873, 111
 Kendall A., Gal Y., 2017, preprint (arXiv:1703.04977)
 Kendall A., Gal Y., Cipolla R., 2017, preprint (arXiv:1705.07115)
 Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
 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., USA, p. 1097
 Laigle C. et al., 2016, *ApJS*, 224, 24
 Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
 LeCun Y., Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W., Jackel L. D., 1989, *Neural Comput.*, 1, 541
 Lecun Y., Bottou L., Bengio Y., Haffner P., 1998, *Proc. IEEE*, 86, 2278
 Lilly S. J. et al., 2007, *ApJS*, 172, 70
 Martí P., Miquel R., Castander F. J., Gaztañaga E., Eriksen M., Sánchez C., 2014, *MNRAS*, 442, 92
 Newell E. B., 1983, in Proceedings of the Workshop on Astronomical Measuring Machines. Edinburgh, p. 15
 Padilla C. et al., 2016, in Evans C. J., Simard L., Takami H., eds, Proc. SPIE Conf. Ser., Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, p. 99080Z
 Padilla C. et al., 2019, *AJ*, 157, 246
 Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, *A&A*, 621, A26
 Paszke A. et al., 2017, 31st Conference on Neural Information Processing Systems, Long Beach (CA)
 Pedregosa F. et al., 2011, *J. Mach. Learn. Res.*, 12, 2825

- Popowicz A., Smolka B., 2015, *MNRAS*, 452, 809
- Romanishin W., 2014, An Introduction to Astronomical Photometry Using Ccds. Createspace Independent Pub, University of Oklahoma, Norman
- Stetson P. B., 1987, *PASP*, 99, 191
- Stohtert L. et al., 2018, *MNRAS*, 481, 4221
- Teeninga P., Moschini U., Trager S. C., Wilkinson M. H. F., 2015, in IEEE International Conference on Image Processing (ICIP), Canada, p. 1046
- Tonello N. et al., 2019, *Astron. Comput.*, 27, 171
- Tortorelli L. et al., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 35
- Vafaei Sadr A., Vos E. E., Bassett B. A., Hosenie Z., Oozeer N., Lochner M., 2019, *MNRAS*, 484, 2793
- Voulodimos A., Doulamis N., Doulamis A., Protopapadakis E., 2018, *Comput. Intell. Neurosci.*, 2018, 1
- Werbos P. J., 1982, in Drenick R. F., Kozin F., eds, System Modeling and Optimization. Springer, Berlin, Heidelberg, p. 762
- Xu B., Wang N., Chen T., Li M., 2015, preprint (arXiv:1505.00853)
- Zeiler M. D., Fergus R., 2013, preprint (arXiv:1311.2901)
- Zhang K., Bloom J., 2019, *J. Open Source Soft.*, 4, 1651

APPENDIX A: CONVOLUTIONAL NEURAL NETWORKS

Machine learning methods are data analysis techniques where the algorithm learns from the data. In particular, one of the most popular class of algorithms are NNs (Werbos 1982), which are designed to recognize patterns, usually learned from training data (*supervised method*). They are mainly used for regression and classification problems (Alexander et al. 2019). Deep learning is a subset of machine learning that refers to a development of NN technology, involving a large number of layers.

Deep learning methods, and in general any supervised machine learning method, model a problem by optimizing a set of trainable weights 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 propagate through all the network layers and then, the network gives a prediction for each of the input samples. After that, by comparing with the known true value, which is technically called *label*, the network estimates a prediction error with a given loss function. After that, *back propagation* takes place. *Back propagation* consists of computing the contribution of each weight on the prediction error. Such contributions are calculated with the partial derivative of the loss with respect to each of the weights. The *weight optimization* is the weights correction based on the quantities calculated in the *back propagation* to reduce the error in the next iteration.

In this work, we use a CNN (Lecun et al. 1998; Zeiler & Fergus 2013). Our network contains four differentiated types of layers:

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, one can convolve the input with as many kernels as desired. 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.

Pooling layers: This layer reduces the dimensionality of the set of convolved images. It applies some function (e.g. sum, mean, maximum) to a group of spatially connected pixels and reduces the dimensions of such group. For example, it takes two consecutive pixels and converts them to the mean of both. 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).

Fully connected layer: This layer is usually the last layer of the network. Its input is the linearized outcome of the previous ones (in our network, convolutions + poolings). It applies a linear transformation from the input to the output. The slope and bias of the linear transformation are the learning parameters.

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 an NN and avoids overfitting problems.

After each convolution and fully connected layer there is an activation function that transforms the outcome. Activation functions are non-linear functions that map the outcome of a layer to the input of the following one. An example of an activation function is the Rectified Linear Unit (ReLU) (Krizhevsky et al. 2012), although we use a variation of this function called LeakyReLU (Xu et al. 2015), with which we find better results.

Other terms that one needs to be familiar with are *epoch* and *batch*. An *epoch* is an iteration over the complete training data set. It is common practice to avoid feeding the network with all the training sample together. Instead, the training data are divided in groups of a certain size and each of these groups is called *batch*. Feeding the network in batches helps it learn faster as in every iteration over a batch, it back-propagates updating all the weights. Then, instead of updating once per epoch, it updates as many times as there are batches. The amount of variation allowed per iteration is regulated by the *learning rate*.

APPENDIX B: VARIABLE ANNULUS

Currently, the PAUS photometry pipeline uses an annulus region to estimate the astronomical background behind a target source. The inner and outer radii are fixed at 30 and 45 pixels from the source, respectively. However, for each galaxy, one could adjust the annulus parameters to minimize the effect of the target galaxy flux falling inside the ring and background variations between the target and the annulus location.

We can quantify the amount of extra light falling inside the annulus coming from the galaxy (Δ_F) and scattered light (Δ_B). For a flat background, Δ_B should be independent of the annulus location. However, this term does depend on the annulus location when the background varies. On the other hand, Δ_F is minimized by an annulus further away from the source and depends on the galaxy size and the PSF. We define

$$\Gamma \equiv \frac{|\Delta_F + \Delta_B|}{\sigma_b}, \quad (\text{B1})$$

where σ_b is the error on the background subtraction. Γ measures the relative error on the background prediction due to scattered light and the source contribution. We use Γ to study the effect of a variable annulus by minimizing the quantity as a function of annulus radius.

We have tested this on simulations. The background images are simulated as in Section 3.2. The images also contain simulated galaxies with the same size (r_{50}) and PSF distribution as observed PAUS galaxies. Using simulations allows us to evaluate Δ_B on the background simulations and Δ_F on the galaxy simulations. This method cannot directly be applied to observed images since it requires distinguishing between the galaxy flux and the background. However, it shows under which conditions the annulus approach

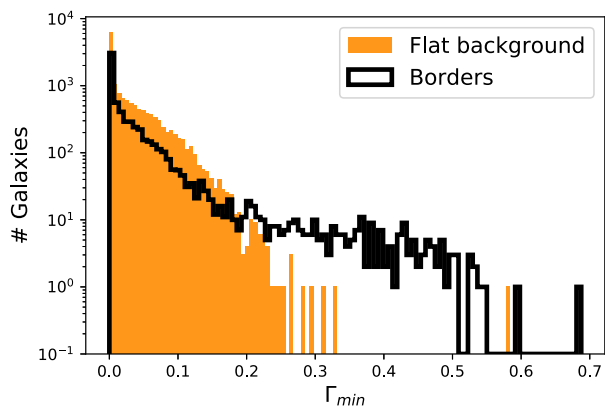


Figure B1. The Γ_{\min} distribution for galaxies in the border (scattered light affected, black) and in the centre (flat background, filled orange) of the image.

degrades. Our simulations only contain Poisson noise and scattered light. We know that real PAUCam images have a more complicated noise pattern and therefore we would also expect the distributions of Γ_{opt} to shift towards higher values.

Fig. B1 shows the histogram of minimum Γ measurements (equation B1) for a set of simulated galaxies. Γ is evaluated for different annulus radii, moving the inner and outer radii between

one and 40 pixels from the source with fixed $r_{\text{out}} - r_{\text{in}} = 15$ pixels. We split the results in galaxies in the centre of the image (flat background, orange) and galaxies in the borders (scattered light, black). Both histograms show a large fraction of galaxies for which the annulus can predict the background very accurately (low gamma values). For galaxies in the centre, we expect a flat background and therefore good results with the annulus method.

There are also many galaxies (Fig. B1) on the border with accurate measurements. First of all, not all positions at the border are affected by scattered light. For those positions with scattered light, the annulus can be placed very close to the target galaxy if it is small ($r_{50} \approx 1$ or 2 pixels). Then, the background variation from the target source to the annulus position would be small. Nevertheless, the distribution of border galaxies also shows a tail corresponding to galaxies for which the annulus estimation significantly degrades. In these cases, the optimal annulus is either too close to the source or capturing a strong background variation. If the background variation is very strong, the annulus will tend to get closer to the target source in order to minimize Δ_{B} . However, this is not possible for bright and large galaxies since getting closer to the target increases Δ_{F} . The annulus method cannot be used to measure the background in large galaxies in varying background regions since it strongly degrades.

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