Precise photometry and photo-z s with multi narrow-band data and deep learning

Laura Cabayol García

Precise photometry and photo-zs with multi narrow-band data and deep learning

Laura Cabayol García

Programa de Doctorat en Física
Universitat Autònoma de Barcelona
Director: Dr. Martin Børstad Eriksen
Tutor: Dr. Manuel Delfino Reznicek

Intitut de Física de Altes Energies

A thesis submitted for the degree of Philosophae Doctor (PhD)

June 2022

Abstract

In the last decades, galaxy surveys have triggered unprecedented progress in our understanding of the Universe. Better astronomical cameras or more powerful computers have enabled the collection of more and better data. Astronomical images need to be processed to turn into photometric catalogues and ultimately into photometric redshifts. Current galaxy surveys have observed the order of millions of galaxies while upcoming surveys like *Euclid* or LSST will increase these numbers to billions. These data will require fast and precise methods to extract the photometry and the photometric redshift.

In this thesis, we have used data from the Physics of the Accelerating Universe Survey (PAUS) to develop an end-to-end deep-learning algorithm to extract the photometry and predict the photometric redshift from astronomical images. We have built the pipeline in three steps, gradually increasing the complexity of the data-reduction operation. In this step-wise approach, we have optimised each photometry process independently, learning about the data, the network requirements, and its underlying mechanisms.

The first project predicts the background noise in the presence of nuisance artefacts and strongly-varying backgrounds. On average, our deep-learning background measurements improve the photometry by 7% and up to 20% at the bright end. The background measurements also reduce the photometric redshift outlier rate by 35% for the best 20% galaxies.

The second project measures the probability distribution of the photometry in single-exposure images. On average, the deep-learning photometry increases the signal-to-noise of the flux measurements by a factor of two compared to an existing aperture photometry algorithm. This algorithm also incorporates other advantages such as 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. This enables reducing the number of photometry outlier observations from 10% to 2%, compared to aperture photometry.

The thesis also presents a novel methodology to enable better broad-band photometric redshifts using data only available for a fraction of the observations. The method consists of a multi-task neural network that predicts the photometric redshift and the PAUS narrow-band photometry. The photometry estimation is an auxiliary quantity that correlates with the redshift. This forces the network to learn a general solution capable of predicting the photometry and the redshift simultaneously. As the auxiliary data are not used as input to the network, we can evaluate the redshift of any galaxy without such data available. In the COSMOS field, we find that the method predicts photometric redshifts that are 14% more precise down to magnitude $i_{\mathrm{AB}} < 23$ while reducing the outlier rate by 40% with respect to the broad-band photometric redshifts. Furthermore, for simulated data, training on a sample with $i_{\mathrm{AB}} < 23$ the method reduces the photo-z scatter by 15% for all galaxies with

$24 < i_{\rm AB} < 25$.

Finally, the last step expands the single-band photometry measurements to multi-band photometry. Using information from the full galaxy spectral energy distribution, this network predicts the photometry in each of the bands and the photometric redshift. This method duplicates the signal-to-noise ratio of the galaxy photometry with respect to the `Lumos` photometry. Furthermore, colour histograms indicate that multi-band photometry contains less noise that the `Lumos` and the `MEMBA` ones since the colour-histograms width is reduced by 5 and 3, respectively. The photometric redshifts are trained on simulations and adapted to the data using transfer learning. These photo-zs improves `BCNz2` template-based photo-z measurements, particularly at the faint end with 25% more precise photo-z. However, we have still not reached the `Deepz` precision. This project is still work in progress and in the near future we aim to study and improve the photo-z precision at the bright end.

Pròleg

En les darreres dècades, millores tecnològiques com la potència de càlcul dels ordinadors i dels fotodetectors, han provocat un progrés sense precedents en el coneixement de l'Univers. Les exploracions sistemàtiques de l'Univers ens han permès obtenir catàlegs fotomètrics de galàxies, que són necessaris per poder calcular la distància a la qual es troben les galàxies (*redfshift*) i poder així fer mapes de l'Univers. En l'actualitat, s'han observat de l'ordre de milions de galàxies, però en properes exploracions, com per exemple les que faran *Euclid* o LSST, n'observarem de l'ordre de bilions. Totes aquestes dades requeriran mètodes ràpids i acurats per a calcular la fotometria i el *redfshift* de les galàxies.

Aquesta tesi se centra en el desenvolupament d'un algoritme d'aprenentatge profund per mesurar simultàniament la fotometria i el *redfshift* d'una galàxia. L'algoritme s'implementa directament sobre imatges astronòmiques i va d'extrem a extrem, incrementant gradualment la complexitat del procés d'extracció de dades. D'aquesta manera, hem optimitzat cada pas de la reducció d'imatges de manera independent, fet que ens ha permès aprendre els requeriments i mecanismes de les xarxes neuronals emprades. Per desenvolupar el mètode, hem utilitzat dades de l'experiment *Physics of the Accelerating Universe Survey* (PAUS).

La primera part de la tesi s'enfoca en la predicció del soroll de fons de les imatges utilitzant xarxes neuronal convolucionals. De mitjana, l'algoritme millora la fotometria de les galàxies entre un 7 i un 20%. A més a més, les nostres mesures de soroll redueixen un 35% els photo-z atípics presents en la mostra.

La segona part de la tesi, extenem el treball previ i desenvolupem una xarxa neuronal que mesura la distribució de probabilitat de la fotometria en cada banda fotomètrica de manera independent. De mitjana, la nostra fotometria duplica el senyal-soroll de les mesures de flux realitzades amb un codi existent de fotometria d'obertura. El nostre algoritme d'aprenentatge profund també incorpora altres beneficis com robustesa en presència d'elements distorsionats, per exemple raigs còsmics, i menys sensitivitat a inexactituds en els paràmetres que defineixen les galàxies. Això permet reduir el nombre de galàxies amb fotometria atípica d'un 10% a un 2%, en comparació amb la fotometria d'obertura.

La tesi també explora com millorar les mesures del *redfshift* de les galàxies fotografiades amb filtres fotomètrics de banda ampla (baixa resolució de longitud d'ona) utilitzant observacions en bandes estretes. El mètode consisteix en una xarxa neuronal multitasca que prediu el *redfshift* i la fotometria en banda estreta d'una galàxia a partir de la seva fotometria en banda ampla. La fotometria està correlacionada amb el *redfshift*, així la xarxa neuronal pot emprar el coneixement adquirit en la predicció d'una de les quantitats per millorar l'altra. La fotometria en banda estreta no són dades d'entrada a la xarxa neuronal. D'aquesta manera, un cop entrenada la xarxa pot predir el *redfshift* de qualsevol galàxia a partir de la seva fotometria en banda ampla sense requerir fotometria en banda estreta. Al camp "COSMOS´´, el nostre mètode prediu photo-z amb un 14% més de precisió fins a magnituds $i_{\mathrm{AB}} < 23$ i redueix el nombre de photo-z atípics un 40%. A més a més, hem pogut provar en simulacions que la xarxa neuronal multitasca també redueix un 15% la dispersió en el photo-z de galàxies amb magnitud $24 < i_{\mathrm{AB}} < 25$.

L'últim capítol mesura fotometria multibanda i el photo-z de les galàxies a partir de les imatges. Aquesta xarxa utilitza la informació disponible en totes les imatges obtingudes d'una galàxia per fer prediccions en cadascuna de les bandes. La fotometria multibanda duplica el senyal-soroll de la fotmetria banda per banda. Les prediccions del *redshift* són encara treball en procès i tenen encara marge de millora. Estem treballant en entendre una tendència sistemàtica en el photo-z de galàxies brillants.

Contents

Introduction 1

I Concepts 4

1 Deep learning background 5

1.1 Gentle introduction to machine learning . . . 5
1.2 Implementation of deep learning algorithms . . . 7
1.2.1 Convolutional Neural Networks . . . 9
1.2.2 Mixture density networks . . . 12
1.2.3 Multi-task learning . . . 13
1.3 Deep learning in astronomy . . . 13
1.3.1 Object classification . . . 15
1.3.2 Photometric redshift estimation . . . 16
1.3.3 Other applications . . . 16

2 The PAU Survey 17

2.1 Galaxy surveys . . . 17
2.1.1 Imaging surveys . . . 18
2.2 The PAU Survey . . . 20
2.2.1 History . . . 20
2.2.2 PAUCam characteristics and construction . . . 22
2.2.3 PAUCam commissioning . . . 24
2.2.4 Science goals . . . 24

3 Image processing, photometry and photometric redshifts 27

3.1 From raw to science images . . . 27
3.1.1 Bias subtraction . . . 27
3.1.2 Flat fielding . . . 27
3.1.3 Dark current . . . 28
3.1.4 Cosmic rays and other spurious artefacts . . . 28
3.2 Photometry . . . 29
3.2.1 Photometry calibration . . . 31
3.3 Further-processing: co-adding flux measurements and derived properties . . . 32
3.3.1 Co-addition . . . 32

3.3.2 Object classification . . . 33
3.3.3 Photometric redshift estimation . . . 33
3.4 PAUS data management . . . 35
3.4.1 From raw to science images: the `Nightly` pipeline . . . 35
3.4.2 Measuring photometry: the `MEMBA` pipeline . . . 37
3.4.3 PAUS photometric redshifts . . . 39
3.4.4 PAUS data in COSMOS . . . 41

II Galaxy photometry and photo-zs with deep learning 43

4 Background light prediction 44
4.1 Motivation . . . 44
4.2 Modelling scattered-light . . . 45
4.2.1 The PAUS observations . . . 46
4.2.2 Scattered-light templates . . . 46
4.2.3 Scattered-light templates as scattered-light correcting method . . . 49
4.3 `BKGnet`: A Deep Learning based method to predict the background . . . 49
4.3.1 Neural network architecture . . . 49
4.3.2 Data: training and test samples . . . 51
4.3.3 Training process and loss function . . . 52
4.4 Testing `BKGnet` on simulations . . . 53
4.4.1 Simulated PAUCam background images . . . 53
4.4.2 `BKGnet` predictions on simulations . . . 54
4.5 `BKGnet` on PAUCam images . . . 56
4.6 `BKGnet` validation . . . 58
4.6.1 Generating the PAUS catalogue with `BKGnet` predictions . . . 59
4.6.2 Validating the catalogues . . . 60
4.7 Conclusions . . . 63
4.A Variable annulus . . . 65

5 Single band photometry 67
5.1 Motivation . . . 67
5.2 Data . . . 68
5.2.1 PAUS data . . . 69
5.2.2 Teahupoo simulations . . . 69
5.2.3 Comparison between PAUCam and Teahupoo galaxies . . . 71
5.3 Flux estimation methods . . . 73
5.3.1 Profile fitting . . . 73
5.3.2 Aperture photometry . . . 73
5.3.3 Weighted pixel sum . . . 73
5.4 `Lumos`: Measuring fluxes with a CNN . . . 74
5.4.1 Input data . . . 74
5.4.2 `Lumos` architecture . . . 75
5.4.3 Unsupervised transfer learning . . . 77

5.4.4 Training procedure . . . 78
5.5 `Lumos` flux measurements on simulations . . . 79
5.5.1 Flux probability distributions . . . 79
5.5.2 Comparison with different flux estimation methods . . . 82
5.5.3 Deblending with `Lumos` . . . 82
5.6 `Lumos` photometry on PAUS data . . . 84
5.6.1 Single exposure measurements . . . 84
5.6.2 Comparison with SDSS spectroscopy . . . 91
5.6.3 Coadded flux measurements . . . 92
5.6.4 Photometric redshift estimates . . . 93
5.7 Conclusions and discussion . . . 95
5.A Flux estimation methods: derivations . . . 98
5.B Forecasting the effect of errors on profile parameters . . . 98
5.C Photometric redshifts with `BCNz2` on PAUS galaxy mocks . . . 100
5.D Colour histograms in the complete narrow band set . . . 102
5.E Photometry and photo-z correlations with galaxy parameters . . . 103
5.E.1 `Lumos` photometry correlation with galaxy parameters . . . 103
5.E.2 Photo-z correlation with galaxy parameters . . . 106

6 Improving broadband photometric redshifts with multi-task learning 107
6.1 Motivation . . . 107
6.2 Data . . . 109
6.2.1 PAUS data . . . 109
6.2.2 Photometric redshift sample . . . 109
6.2.3 Broadband data . . . 110
6.2.4 Spectroscopic galaxy sample . . . 110
6.2.5 Galaxy mocks . . . 110
6.3 Multi-task neural network to improve broad-band photo-z . . . 110
6.3.1 Multi-task learning . . . 111
6.3.2 Model architecture and training procedures . . . 111
6.4 Photo-z performance in the COSMOS field . . . 114
6.4.1 Photo-z performance metrics . . . 114
6.4.2 Photo-z dispersion . . . 115
6.4.3 Photo-z bias and outlier rate . . . 117
6.4.4 Redshift distributions, N(z) . . . 118
6.5 Photo-z performance on deeper galaxy simulations . . . 120
6.6 Photo-z in colour-space . . . 121
6.6.1 MTL photo-z in colour-space . . . 122
6.6.2 Broad-band degeneracies in colour-space . . . 124
6.6.3 Emission line confusions . . . 126
6.7 Understanding the MTL underlying mechanism . . . 128
6.7.1 Underlying data representation in colour-space with MTL . . . 128
6.7.2 MTL with other galaxy parameters . . . 129
6.7.3 Effect of narrow-band resolution . . . 131
6.8 Discussion and conclusions . . . 131

6.A Self-organising maps . . . 133
6.B Effect of training with photo-z as ground-truth targets . . . 134
6.B.1 Effect of photo-z dispersion in the training redshifts . . . 135
6.B.2 Effect of photo-z outliers in the training redshifts . . . 136

7 Multi-band photometry and photometric redshifts 139
7.1 Motivation . . . 139
7.2 Data . . . 141
7.2.1 PAUS data . . . 141
7.2.2 Spectroscopic sample . . . 141
7.2.3 PAUS image simulations . . . 142
7.3 Network architecture and training procedure . . . 143
7.3.1 Feature extraction with a convolutional neural network . . . 143
7.3.2 Photometry and photo-zs with MDN . . . 145
7.3.3 Training procedure . . . 146
7.4 Multi-band photometry measurements . . . 148
7.4.1 `Aczio` photometry on simulated data . . . 149
7.4.2 `Aczio` fluxes on PAUCam images . . . 151
7.4.3 Flux uncertainties . . . 153
7.4.4 Signal-to-noise on PAUCam data . . . 154
7.4.5 Colour histograms . . . 154
7.4.6 Robustness of the method . . . 156
7.5 Photo-z measurements . . . 157
7.5.1 Photo-zs in COSMOS . . . 157
7.5.2 Photo-z probability distributions . . . 161
7.5.3 Colour space . . . 163
7.6 Conclusions . . . 163
7.A Effect uninformative features in the input of a neural network . . . 165
7.B Simulated broad-band photometry . . . 167
7.C Scattering the ground-truth redshifts . . . 168

8 Summary and conclusions 169

Introduction

Astronomers have been observing the Universe for centuries. Back in Ancient Greece, very well-known names as Anaxagoras or Ptolemy studied astronomical phenomena such as eclipses, the brightness of celestial objects or their rotational movement by observing the nearby sky. In 1929, galaxy surveys became a standard tool in astronomy, which entailed a change of model in astronomical studies from the analysis of single observations to a statistical one (Okamura, 2020).

Galaxy redshift surveys are a powerful tool to study the Universe. These map a region of the sky and locate the position and redshift of the objects inside. In the last decades, there has been a breakthrough in the amount and quality of galaxy survey's data, leading to unprecedented progress in our understanding of the Universe. As an example, the Palomar Observatory Sky Survey (POSS-I, Minkowski & Abell, 1963) which imaged 2/3 of the observable sky from Palomar Mountain to $i_{\mathrm{AB}} < 21$ in photographic plates back in 1950. In contrast, current modern surveys are observing hundred a million galaxies (The Dark Energy Survey Collaboration, 2005; de Jong et al., 2013) to fainter magnitudes $i_{\mathrm{AB}} \sim 24$.

Furthermore, in the next decades, the number of observed galaxies, the sky coverage, and the observation's depth will be extended by the next generation of ground and space telescopes. *Euclid* (Laureijs et al., 2011) will observe 15 000 deg^2, yielding photometry and photometric redshifts for about 10 billion sources. Also, LSST (Ivezić et al., 2019a) will image 20 billion galaxies to $i_{\mathrm{AB}} < 24$. These data will require fast and precise methods to turn astronomical images into photometry and ultimately into photometric redshift catalogues.

Galaxy surveys can be broadly classified as spectroscopic or photometric surveys. The former splits the light in narrow wavelength bins, enabling to determine very precise redshifts. However, it is time-consuming and the efficiency of obtaining redshifts is low. In contrast, photometric surveys image the sky using a few pass-band photometric filters at different wavelengths. This enables observing many objects simultaneously but at expense of a lower wavelength resolution, leading to less precise redshift measurements. While spectroscopic data are powerful for galaxy evolution studies, e.g. star formation and mergers (Robotham et al., 2014) and the environmental dependence of galaxy evolution (Alpaslan et al., 2015), large photometric data-sets are very suitable for large scale structure and gravitational lensing analysis (Kuijken et al., 2015; Elvin-Poole et al., 2018a).

Obtaining precise photometric redshift is crucial for most cosmological studies. This has prompted important efforts to improve the redshift estimation methods, leading to a wealth of techniques optimised for different science applications and types of data (e.g. Feldmann et al. 2006; Brammer et al. 2008; Eriksen et al. 2019). These techniques typically use the photometry measured from the astronomical images. Therefore, the data reduction process converting astronomical images into photometry catalogues is a key step in the determination of accurate photometric redshifts.

This thesis uses data from the Physics of the Accelerating Universe survey (PAUS, Martí et al., 2014), which is a unique imaging redshift survey taking data with a camera equipped with 40 narrow-band filters (Padilla et al., 2019a). Such a large number of photometric filters provides PAUS with a wavelength resolution in between broad-band photometry and spectroscopy, which enables reducing the photometric redshift uncertainty by a factor of around 15 with respect to typical broad-band imaging surveys (Eriksen et al., 2019; Eriksen et al., 2020; Soo et al., 2021).

Deep learning techniques have been undergoing an unprecedented revolution over the last few years. This has been prompted by an increasing amount of available data and computing power, together with a better theoretical understanding of the techniques. The development of Graphical Processing Units (GPUs) has been crucial for speeding up the computation of modern deep-learning algorithms, enabling the growth of a new deep-learning field devoted to the development of techniques applied to images. These techniques have also reached astronomy, where implementing deep learning tools to the increasing amount of astronomical images is a new promising venue (e.g. Pasquet et al., 2019; Zhang & Bloom, 2019; Arcelin et al., 2021).

In this thesis, we have developed an end-to-end deep learning pipeline to go from PAUS science images to photometry and the photometric redshift. Chapters 1, 2, and 3 are an introduction with useful information for understanding the thesis. The former (§ 1) presents the basic concepts for the understanding of neural networks and introduces the architectures and training methodologies used across the thesis. The second introductory chapter (§ 2) presents the PAU Survey, its camera and science goals. Finally, the last part of the introduction (§ 3) explains the general astronomical data reduction steps to reduce raw astronomical images to photometry and photometric redshift catalogues. The same chapter also introduces the PAUS data management pipeline. This consists of a de-trending code that converts raw images into reduced science images (§ 3.4.1) and a second part that estimates the photometry from such reduced images (§ 3.4.2).

Our deep-learning pipeline has been developed in multiple independent steps, each of them addressing a data reduction operation. Studying each step independently gives a better understanding of the data and the network's requirements, e.g. which information is relevant to improve the photometry and how to provide the network with such information. First. in Chapter 4, we have predicted the sky-background noise using CNNs, which is our first step in obtaining reliable photometry and photo-z. We introduce `BKGnet`, an algorithm to make accurate background noise predictions at the source location in the presence of nuisance artefacts and strongly varying background light. This is a published work in Cabayol-Garcia et al.

(2020), titled "The PAU Survey: Background light estimation with deep learning techniques".

Chapter 5 introduces `Lumos`, a CNN that predicts the probability distribution of the already background-light subtracted photometry. `Lumos` uses the experience from `BKGnet` to tackle a more complex data reduction operation. This work is published as "The PAU survey: Estimating galaxy photometry with deep learning" (Cabayol et al., 2021).

Some of the techniques used to extend `Lumos` to measure multi-band photometry were first tested in Chapter 6, which introduces a multi-task learning network to enable better broad-band photometric redshifts using PAUS data. This network predicts the photo-z and the PAUS narrow-band photometry simultaneously from the broad-band photometry, combining both tasks in the loss function. The method only uses PAUS data during the training phase to evaluate the accuracy of the network narrow-band photometry predictions. Therefore, we can estimate the photo-z of any galaxy with broad-band photometry, without requiring narrow-band observations. This work is currently undergoing *Euclid* internal review for publication in "The PAU Survey & *Euclid*: Improving broad-band photometric redshifts with multi-task learning" (Cabayol et al. in prep.).

Chapter 7 presents the last part of the photometric pipeline. This extends `Lumos` to predict the multi-band photometry and photo-z of any galaxy from its image observations. First, the network extracts a set of co-added features from all observations of a galaxy in a narrow band. Then, the features from all bands are used to predict the flux in each narrow band filter, in such a way that the network uses all the spectral energy distribution information encoded in the galaxy images to predict the photometry in a single band. This network uses the knowledge of Chapter 6 to implement a multi-task learning training that simultaneously predict the photometry photo-z. Both tasks share a set of network layers that capture data traits relevant for the two predictions. The photo-z prediction also inputs the co-added features from all narrow-bands as input, in such a way that it uses the information available in all the images of a galaxy. This last chapter builds on all the previous work presented, expanding the photometry pipeline to go end-to-end and using the multi-task learning techniques tested in Chapter 6. This work is in preparation as "The PAU Survey: Multi-band photometry and photo-z from narrow-band images with deep learning" (Cabayol et al. in prep.).

Part I

Concepts

Chapter 1

Deep learning background

1.1 Gentle introduction to machine learning

Artificial intelligence (AI) is a field focused on constructing complex machines that can process intellectual tasks normally performed by humans. The term ”artificial intelligence“ was coined back in 1956 at a conference at Dartmouth College (McCarthy et al., 2006). By that time, AI generated great expectations and money was invested in the field.

The first very simple Artificial Neural Network (ANN) was created in 1958 by Frank Rosenblatt and was named 'Perceptron'. Currently, ANNs concatenate several *layers* that are constructed by putting together collections of perceptrons. Each of these layers is typically responsible for learning a specific hidden pattern of the data.

The development of the perceptron created enthusiasm amongst the academic community, however, the computer power was limiting its extension to deeper (with more layers) ANNs and a single perceptron could only handle trivial versions of the problems they were supposed to solve. This led to disappointment and the interest and investment in the field dropped off. As a consequence, in 1973, the UK Parliament severely criticised the progress on AI, which triggered a cut in AI investment and the coming years (1974 to 1980) the research on AI was marginal (”AI winter“).

In the eighties, the interest in AI returned. In 1980, Yann LeCun developed an early version of a convolutional neural network (CNN, LeCun et al., 1989) that could recognise handwritten digits. This was successfully implemented in postal and banking services. Nevertheless and despite the current popularity of CNNs, by that time computer power was also limiting CNN's performance and this type of network architecture was left aside.

The strong limitations of ANN performance triggered that other types of AI had their heyday in the eighties. Expert systems were first introduced by Edward Feigenbaum in 1965. These rely on two components: a knowledge base that provides a set of rules to carry out a task and an inference engine that implements logical algorithms to the knowledge base to infer new rules. `MYCIN` (Buchanan & Shortliffe, 1984) is a classic expert system implementation developed to diagnose and recommend medical treatment, supporting clinicians in the early diagnosis of meningitis. MYCIN's knowledge relies on approximately 500 antecedent-consequent rules that enabled to recognise of $\sim$ 100 causes of bacterial infections. To make a decision, MYCIN starts with information such as e.g. the age, sex, and medical history of the patient, scaling to more specific questions when required. Another example of a successful expert system implementation is DeepBlue (Campbell et al., 2002). It is a chess computer that defeated the world chess champion, Kasparov, in 1998. DeepBlue consists of an early version of a supercomputer that estimates approximately 150 million possible chess movements per second and a decision tree calculating the best move. There were several successful expert system implementations however, managing the knowledge base and writing accurate expert system rules was difficult.

In the nineties, the hype in expert systems declined, which has two possible interpretations. The first one is that although expert systems provide deep, focused knowledge of a particular problem, this knowledge cannot be generalised to any other task,e.g. MYCIN cannot be implemented or easily adapted to the diagnosis of encephalitis. Therefore, these algorithms could not expand to a more general AI technology fast enough to keep the hype and AI moved on. Another possible interpretation is that expert systems were absorbed by other tools that used their technology as part of other offerings, leaving the standalone expert system out of the spotlight. Nevertheless, nowadays there is still some research on standalone expert systems, e.g. Ahmed & Mahmoud (2020).

In 2008 Fei-Fei Li set up ImageNet (Deng et al., 2009), which is a database of annotated images that provides a common image data-set to train and benchmark models. ImageNet quickly scaled to 11 million images in 2010 and currently contains more than 14 million examples. The setting of ImageNet was an AI's milestone that has eased the research in computer vision tasks. In 2012, AlexNet (Krizhevsky et al., 2012) beat any previous result in image recognition tasks using the ImageNet database. AlexNet is a CNN with eight layers; five convolutional layers followed by three fully-connected layers. The main result in Krizhevsky et al. (2012) was that the depth of the model (number of layers) was triggering the great network's performance. This is the origin of deep learning (DL), where the adjective "deep" refers to ANNs with a large number of layers. AlexNet was designed by Alex Krizhevsky under the supervision of Geoffrey Hinton, one of the pioneers of deep learning.

Nowadays, AI has become part of our daily existence. The explosion of smartphones and similar devices collecting huge amounts of data together with the development of applications using AI, e.g. voice and search assistants, triggered important companies like Google and Facebook to invest a lot of money in AI research. This revolution was boosted by several factors:

- **Big data**: Over the last years, the amount of available data has quickly increased thanks to devices like smartphones and computers. These gadgets daily generate huge amounts of data collected by services like e.g. Google, Facebook, YouTube, and Instagram. While traditional shallow ANNs do not benefit from having more data, deep ANNs can boost their performance by using very large data sets.
- **Processing Power**: Graphical Processing Units (GPUs) have emerged as technologies

to speed up computation, especially for deep learning algorithms requiring the computation of multiple parallel processes. GPUs were initially developed for accelerating graphics processing, however, these have become a crucial part of modern deep learning networks. GPUs are central to the increase in computer power (Oh & Jung, 2004). These were already used for the development of `Alexnet`.

Complex models trained with large data sets demand more computational power. While high-end GPUs can be very expensive, cloud services offer a cheaper alternative to increase computational power that many more people can access.

- **Open-source software**: Recently, several neural networks have been developed an provided as open-source software enabling a wider and standardised application of machine learning tools. Important examples are `Keras` (Chollet et al., 2015), `Tensorflow` (Abadi et al., 2015), and `PyTorch` (Paszke et al., 2017), where the last two have been developed by Google and Facebook, respectively.

1.2 Implementation of deep learning algorithms

Machine Learning (ML) is a branch of AI that learns how to solve a specific problem from data. In classical software, routines to perform a specific task are hand-coded with a specific set of instructions to perform such a task. Instead, machine learning algorithms iteratively learn from the data how to perform the task in a process called *training*.

To create algorithms that learn similarly to humans, ANN architectures are inspired by the structure of the human brain. As mentioned in § 1.1, the ANN computational unit is the *perceptron*, which is the analogue of a neuron, and ANNs are made of combinations of perceptrons named layers. The first layer of the network is the *input layer* and the last, the *output layer*. The layers in between are the so-called *hidden layers* (see Fig. 1.1).

Supervised ANNs model a problem by optimising a set of trainable parameters (the perceptrons and also technically named *weights*) to fit the data. This is done using a *training sample*, which is a data-set of input examples with a known solution. Given the training sample, the ANN optimises its trainable parameters to minimise the difference between the outcome prediction and the expected output. We can differentiate three stages in the ANN's training phase: forward propagation, backpropagation and weight optimisation.

The training starts with the *forward propagation*. At this stage, the input data ($x_i^{\rm in}$) propagates through all the network layers and the output layer provides a prediction for each of the input samples. This prediction can be of different types depending on the problem the network is addressing. If the network is a classifier, it predicts the class the input example belongs. In contrast, if the network is addressing a regression problem, the prediction is a value for the regression that can also be attached to other quantities such as the uncertainty or the covariance. In the case of a linear ANN, the forward propagation in one layer reads as

$$\vec{x'} = \phi(\mathbf{w} \cdot \vec{x} + \vec{b})\,, \tag{1.1}$$

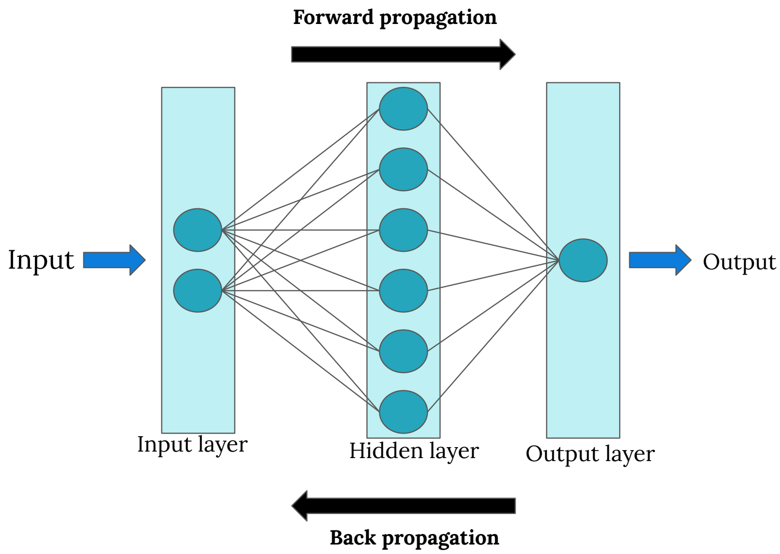

Figure 1.1: ANN composed of an input, a hidden and an output layer. Each circle represents a perceptron and has an associated weight w_i. The black lines are the connection between the perceptrons in one layer and those from the following one. This particular example has two variables as input at outputs a single prediction.

where $\vec{x'}$ is the signal after doing the forward propagation in the layer, $\vec{x}$ is the input vector to the layer, $\mathbf{w}$ is the weight matrix, and $\vec{b}$ is the *bias* term. After each layer, there is an *activation function* (ϕ) which is a non-linear function that maps the output of a layer to the input of the following one (see Fig. 1.1). This is required to produce non-linearities in the model. There are several common activation functions, e.g. the Sigmoid function or the hyperbolic tangent. Recently, the ReLU function (Nair & Hinton, 2010), which is

$$\phi(\vec{x}) = max(0, \vec{x}) , \tag{1.2}$$

has become the default activation function for many neural networks. The ReLU usually achieves better convergence performance and it is computationally more efficient than previous commonly used functions.

The main limitation of the ReLU function happens when many ReLU neurons only output zero values, which is known as the *dying ReLU* problem. As the slope in the negative range is zero, the dead neurons remain stuck providing zero values. Some variations of the ReLU function, e.g. the LeakyReLU (Xu et al., 2015) and the SELU (Klambauer et al., 2017), emerged in attempts to further optimisation of the dying ReLU problem.

After the forward propagation, the prediction is compared with the known true value (*label*) using a *loss function* that evaluates how well the algorithm models the data. There are many different loss functions and their choice depends on the task we are optimising. Typically classification problems use a cross-entropy loss function (Good, 1952), although there are also other options like e.g. the Kullback-Leibler divergence (Kullback & Leibler, 1951). On the other hand, regression networks predicting continuous values use e.g. the mean squared error or the mean absolute error. In § 1.2.2 we will describe more complicated loss functions that are used to predict the probability distribution of a continuous target quantity, e.g. Eq. 1.5, which is used to optimise Gaussian mixture density networks (§ 1.2.2).

The ultimate goal of a supervised machine learning algorithm is to efficiently minimise the loss function. *Backpropagation* (Kelley, 1960) is an optimisation method that consists in computing the contribution (*gradients*) of the ANN's weights $\mathbf{w}$ to the loss function $\mathcal{L}$ after each forward pass using the chain rule. After back-propagation, the optimiser uses the estimated gradients to update the parameters in a way that minimises the loss function (*weight optimisation*). This whole procedure takes place repeatedly and it is commonly implemented with a *gradient descent algorithm*. This algorithm reduces the loss function after each iteration while adapting the parameters to the data until finding the loss function global minimum. The gradients provide the direction with the steepest ascent in the loss function space. Therefore the optimisation must be done opposite to the gradient (this is why it is called gradient descent), i.e.

$$\vec{w} \leftarrow \vec{w} - \alpha \cdot \vec{\nabla}\mathcal{L}(\vec{w}) \tag{1.3}$$

where α is the so-called *learning rate*, which controls the variation of the model parameters. The gradients are smaller as the network approaches the minimum in the loss function space, where these are exactly zero.

Gradient descent computes the gradients using the full sample, which can be computationally expensive. *Stochastic Gradient Descent* is a variation of gradient descent that uses randomly shuffled and sampled data of a size smaller than the whole training sample (e.g. 100, 128, or 256 data examples) to estimate the gradients. Each of these groups of data is named *batch*. Gradients from batches are typically noisier than those estimated from the whole sample, thus the network takes longer to converge. Nevertheless, stochastic gradient descent is still computationally less expensive than typical gradient descent, which makes the former the commonly preferred algorithm to optimise neural networks.

Nowadays, there are many types of ANN, each of them used for different purposes. The simplest one is the *multi-layer perceptron* (MLP), also named *linear* network or *fully-connected network*. It consists of a concatenation of layers where all the perceptrons in one layer are connected to those in the following one. Figure 1.1 is an example of a three layers fully-connected neural network. It contains two input neurons that are fully connected to the six neurons in the hidden layer, which in turn are all connected to the output neuron. Besides MLP, in this thesis we have used CNNs and mixture density networks, which are explained in § 1.2.1 and § 1.2.2, respectively.

1.2.1 Convolutional Neural Networks

Convolutional neural networks have proven successful for a lot of image related applications, e.g. image classification (Sultana et al., 2019), image semantic segmentation (Liu et al., 2018),

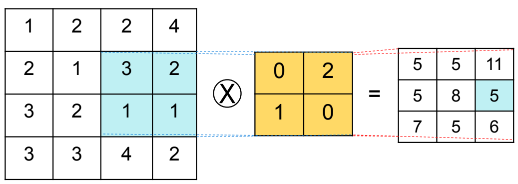

Figure 1.2: Example of convolution. The leftmost matrix corresponds to the input image. The middle yellow matrix is the convolutional kernel and the rightmost one is the output image.

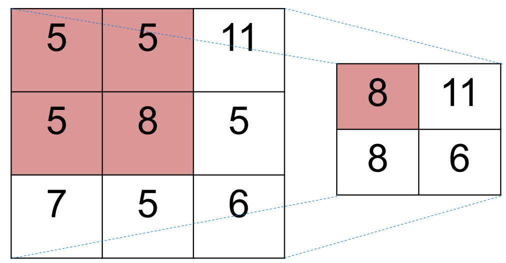

Figure 1.3: Example of the max-pooling operation. The left matrix is the input image and the right one is after the pooling.

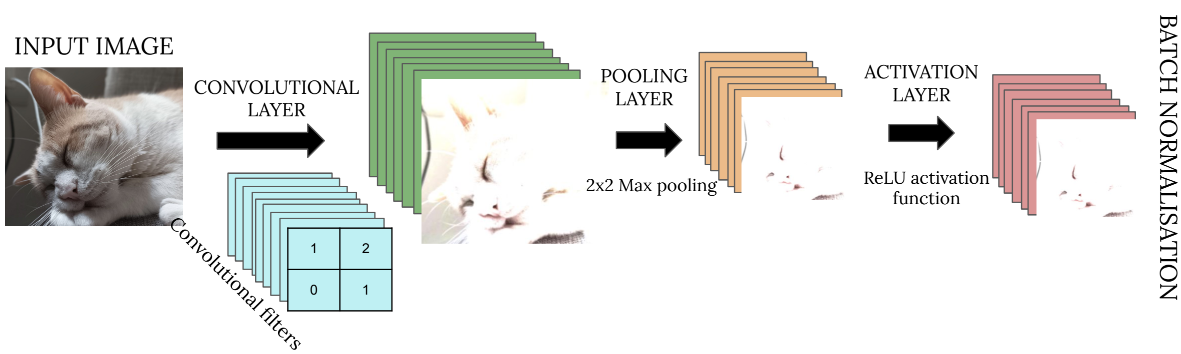

Figure 1.4: Example of a CNN composed of a convolutional layer, a pooling layer, a ReLU activation function. The batch normalisation comes after the activation layer.

and object detection (Zhao et al., 2018). CNNs are a type of ANN composed of *convolutional layers*. In contrast to a fully-connected ANN, where the input propagates linearly through the network (Eq. 1.1), the operation of a convolutional layer essentially consists in sliding the input image with *convolutional kernels*.

Convolutional kernels are (typically) 4-dimensional matrices of trainable parameters. When the kernel passes on a grid of pixels (of the same size as the kernel), each pixel is multiplied by the corresponding value in the kernel and the contribution of all pixels in the grid is added to a single number. The convolution over the complete image creates a new representation of the input data.

The left panel of Fig. 1.2 shows an example of convolution operation. The leftmost matrix represents the input image and the sun-seed yellow centred matrix is the convolutional kernel. The coloured grid on the input image is multiplied by the kernel and summed together, resulting in the bluish matrix value on the rightmost matrix. This procedure is repeated for each 4×4 group of pixels. Note that the convolution reduces the input image dimension from 4x4 to 3x3, as a 2x2 kernel can only slide 3 times in each direction over a 4x4 matrix. To avoid the dimensional reduction one can apply *padding*, which consists in adding extra rows and columns to the input image, enabling one more slide. The added values are commonly filled either with zeros or copying the pixels from the edge of the image.

Multiple convolutional kernels can be applied within a convolutional layer. Every kernel will create a new data feature map focusing on different data traits. Concatenating convolutional layers enables shallow layers, i.e. layers close to the input layer, to learn low-level features (e.g. edges and lines) while deeper layers learn more complicated features (e.g. shapes). As the number of convolutional layers and kernels per layer increases, so does the number of trainable parameters and the amount of data (and memory) that the network needs to handle.

Convolutional layers enable CNNs to use the local spatial coherence of images (i.e. the fact that spatially close pixels together have a meaning) to reduce the number of operations required to process an image. Furthermore, CNNs also learn from the order of their inputs. Considering every pixel in one image an input feature, the CNN sees where each of these pixels is located and uses this information to make predictions.

CNNs also use the spatial coherence of images to effectively reduce the dimension of the input features using the so-called *pooling layers* (Gholamalinezhad & Khosravi, 2020). Pooling layers apply any differentiable operation (e.g. the maximum or the average) to reduce a group of pixels in the feature map to a single value. Therefore, these layers down-sample feature maps by creating a smaller representation of each feature map separately. The right panel in Fig. 1.3 shows an example of 2x2 max-pooling, where each 2x2 pixel grid in the input image is replaced by its maximum value.

Pooling layers also help summarise the presence of features in the input image. Two images of the same object can present slightly different images as a result of e.g. rotation or cropping. After the convolutional layer, these images will result in different feature maps.

The pooling layer helps to regularise the differences between slightly different feature maps by applying operations over groups of nearby correlated pixels.

The last type of layer we introduce here is the *batch normalisation layer* (Ioffe & Szegedy, 2015). This layer is particularly helpful when training deep neural networks with lots of hidden layers. It is commonly implemented after the activation function and consists in re-scaling batch by batch the activated output of the previous layer, so that it has zero mean and unit variance (standardise). During the back-propagation process, weights are updated layer by layer. When doing so, we assume the weights in all the other layers are fixed. However, this is not the case since back-propagation iteratively updates all layers in the network, hindering the loss minimisation. Batch normalization helps coordinate the update of the different layers in the model, fastening the convergence and making the learning more robust.

Figure 1.4 presents an example of CNN composed of a convolutional layer, a max-pooling layer, the ReLU activation function layer, and batch normalisation. We can visualise that convolving the input image with the convolutional kernel highlights certain parts of the input image and smooths others. The pooling layer remarks even more on the features highlighted in the convolutional layer.

1.2.2 Mixture density networks

So far, the presented networks predict a single value. However, assessing the uncertainty of the predictions is often required for scientific applications. Mixture density networks (MDN, Bishop, 1994) predict the probability distribution of the prediction y given the data $\mathcal{D}$ as a weighted sum of k distributions that can be any sort of basis function, e.g. Gaussians functions, in such a way that

$$p(y|\mathcal{D}) = \sum_{i}^{k} \alpha_i N_i(\mu_i, \sigma_i)\,, \tag{1.4}$$

where $N_i(\mu_i, \sigma_i)$ is the i-th Gaussian component with mean μ and standard deviation σ. The α parameters are the so-called mixing coefficients, which give the relative contribution of each Gaussian component to the total probability distribution.

MDNs combine a neural network with a mixture density model. The neural network, which can be of any type (e.g. CNN, § 1.2.1), takes the input data $\mathcal{D}$ and converts it into a set of values that are modelled by the mixture model. The mixture model shapes the data using several distributions that can be written in a simple parametric form (e.g. a Gaussian, as in Eq. 7.3). Figure 1.5 presents a Gaussian MDN. The blue rectangles represent a 3-layers ANN that given two input values outputs the mean and standard deviation of N Gaussians (yellow points), together with the mixing coefficients α (Eq. 7.3). These output parameters build the probability distribution of the predicted quantity.

A Gaussian MDN is trained with a loss corresponding to the negative log-likelihood of a linear combination of Gaussian distributions, i.e.

$$\mathcal{L}_{\text{MDN}} = -\log\left(p(y|\mathcal{D})\right) = \sum_{i=1}^{k}\left[\log(\alpha_i) - \frac{(f_i-\mu_i)^2}{\sigma_i^2} - 2\log\left(\sigma_i\right)\right]. \tag{1.5}$$

This corresponds to maximising the likelihood function $\mathcal{L}(\mathcal{D}|\vec{\theta})$, where $\vec{\theta}$ are the $\vec{\mu}, \vec{\sigma}$, and $\vec{\alpha}$ parameters modelling the Gaussians in Eq.7.3.

1.2.3 Multi-task learning

Deep learning algorithms consist of training a single or an ensemble of models to accurately perform a single task (e.g. predicting the redshift). Multi-task learning (MTL) is a training methodology that aims to improve the performance on a single task by training the model on multiple related tasks simultaneously (Caruana, 1997). A pedagogical example is a network used to classify images of cats and dogs. If the same network is simultaneously trained to classify the shape of the ears, e.g. spiky or rounded, the network will learn correlations between the animal type and the ear shape, e.g. dogs mostly have rounded ears, in such a way that the ear shape predictions will help in the cat-dog classification.

There are two main types of MTL network architectures: soft- and hard-parameter sharing (Zhang & Yang, 2021). Hard-parameter sharing architectures are the most common type of MTL and that used in this thesis. This MTL implementation shares a set of hidden layers among tasks, while each task also implements task-specific layers after the shared ones. Figure 1.6 shows an example of three task hard-parameter sharing MTL, which is built of three shared layers (blue layers) and a single task-specific layer per task (tomato-red layers). Sharing hidden layers forces the network to learn representations that generalise for all tasks. Although the example (Fig. 1.6) only has a single task-specific layer per task, this is commonly extended to several. On the other hand, in soft-parameter sharing architectures, each task has its model and there are no shared layers. The distance between the parameters of the different models is regularised to keep them similar.

MTL has already been successfully applied to fields such as e.g. video processing where Song et al. (2020) implements MTL to simultaneously predict the edge and the disparity maps in stereo video processing[1]. Other example implementations include Moeskops et al. (2017), where an MTL network is trained to simultaneously segment tissues in brain images, the pectoral muscle in breast images, and the coronary arteries.

1.3 Deep learning in astronomy

Astronomy is experiencing an explosive growth of data as a result of past, current and upcoming surveys (Mickaelian, 2016; Zhang & Zhao, 2015). For example, the Palomar Digital

[1]Stereo video is the practice of producing the illusion of 3D images in moving form. Disparity maps display the apparent pixel difference between a pair of stereo images, i.e. images of the same taken from different perspectives and edge maps indicate the position of the edges in the image.

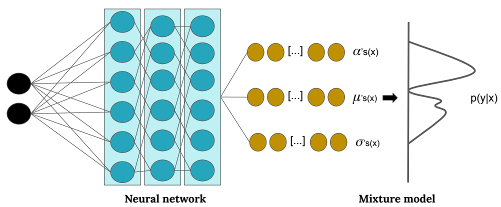

Figure 1.5: Mixture density network scheme. The first part represents a neural network extracting features from the input data while the second is a mixture model constructing the output's probability distribution from the network's output features.

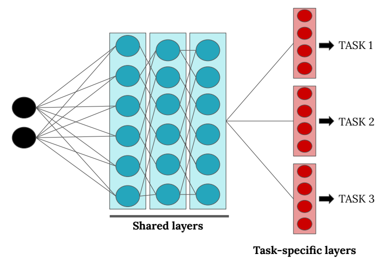

Figure 1.6: Multi-task learning scheme. This particular example corresponds to a hard parameter sharing architecture, where all tasks share a set of common layers (blue layers). The red layers represent the task-specific layers.

Sky Survey (DPOSS, Djorgovski et al., 1998) and the Two Micron All-Sky Survey (2MASS, Skrutskie et al., 2006) generated 3 TB and 10 TB of data respectively. This already increased to 40 TB for the Sloan Digital Sky Survey (SDSS, Ahumada et al., 2020) and it is expected to rise to 40 PB and 200 PB for The Panoramic Survey Telescope and Rapid Response System (PanSTARRS, Magnier et al., 2020) and the Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezić et al., 2019a). The improvement of technology has enabled the construction of larger and more powerful telescopes and cameras contributing to the rapid increment of astronomical data.

Furthermore, astronomical data embraces different data types and complexities, including images, spectra, simulations, and time series. For example, SDSS observed the spectra of millions of galaxies and multi-colour images of one-third of the sky, the Dark Energy Survey (DES, The Dark Energy Survey Collaboration, 2005) imaged 5000 $\deg^2$ of the southern sky ($\sim$300 million galaxies) in five optical filters, and the Kilo-Degree Survey (KiDS, de Jong et al., 2013), imaged two areas of 750 $\deg^2$ in four optical filters and five near-infrared bands. Furthermore, other surveys such as e.g Gaia (Gaia Collaboration, 2018) also produce multi-temporal data. Gaia is accurately mapping the Milky Way measuring the motion of each star around the centre of the galaxy.

The increasing amount of astronomical data to analyse has fostered the implementation of data-driven tools to address astronomical data analysis. Furthermore, future surveys like LSST, the Dark Energy Spectroscopic Instrument (DESI, DESI Collaboration et al., 2016), and *Euclid* (Laureijs et al., 2011) will increase the number of observed astronomical objects by more than an order of magnitude, enhancing the need of fast, automated tools to process all the data. While training deep learning models can be time-consuming, evaluating them on data is a fast operation.

There are many different examples of deep learning implementations in astronomy. Traditionally, these worked at a catalogue level, but recently with the development of very powerful CNN, there has been an increasing interest in implementations at the image level. This opens a new research path full of possibilities such as e.g. automated classification. Training a deep learning model to classify astronomical objects from the images enables an automated real-time object classification (Narayan et al., 2018), also allowing a rapid follow-up of rare phenomena.

In this thesis, we have developed a deep learning end-to-end pipeline to measure the photometry and the photometric redshift of galaxies directly from the images. This enables a fast evaluation of both quantities, reducing the number of processing steps and exploiting the information available in the images.

1.3.1 Object classification

One of the most studied deep learning applications in astronomy is object classification, e.g. star-galaxy and galaxy morphology classification. Most traditional star-galaxy classifiers use summary information from catalogues. In Ball et al. (2006), SDSS colours (i.e. $u-g$, $g-r$, $r-i$, and $i-z$) are used to provide a classification for all 143 million photometric objects in the SDSS-DR3. Also, in Cabayol et al. (2019) objects from the COSMOS field are classified based on 40 narrow-band colours using a 1D CNN. Baqui et al. (2021) tests six different machine learning algorithms (e.g. K-nearest neighbours and decision trees) to distinguish stars and galaxies using 56 narrow-band filters and 4 $ugri$ broad-band filters. Approaches addressing the classification at the image level use both the photometry encoded in the image and the morphology of the source to predict the object type. One example is Kim & Brunner (2016), where a CNN is trained on SDSS images in five photometric bands $ugriz$ to $r < 23$.

Traditionally, galaxy classification has only relied on galaxy morphology. The common galaxy classification scheme was proposed by Hubble in 1936 and splits the galaxies into four broad types based on their morphology: elliptical, spiral, barred-spiral, and irregular, each of these classes containing several sub-classes (Hubble, 1922; Hubble, 1926; Hubble, 1927; Hubble & Tolman, 1935). For most of the 20th century, galaxy classification was tackled by visual inspection of a group of astronomers, (e.g. de Vaucouleurs et al., 1991). With modern surveys data, visual inspection of the entire catalogue is infeasible due to the large number of observed galaxies. Moreover, to quantify the error in the classification, galaxies require multiple independent classifications. Galaxy Zoo was created to solve this problem. It is a crowd-sourcing project to classify more than 60 million SDSS galaxies based on online citizen visual inspection (Lintott et al., 2008).

CNN offer an alternative to visual inspection (Khalifa et al., 2017; Domínguez Sánchez et al., 2018). Working directly on galaxy images, these networks can provide a morphological classification for millions of objects using all the information available in the image (e.g. morphology, photometry, and environment). The classification of bright galaxies (Zhu et al., 2019; Goddard & Shamir, 2020) is addressed with annotated data sets like Galaxy Zoo 2 (Willett et al., 2013) or catalogues with reliable known galaxy morphologies e.g. the Principal

Galaxy Catalogue (Paturel et al., 2003) or the Value-Added Galaxy Catalogue (Choi et al., 2010). Furthermore, CNNs also offer a solution for galaxy classification of deep data sets, where the faintest galaxies are hardly distinguished from the background and visual morphological inspection is not a possibility. This could potentially be addressed using galaxy image simulations to train the CNN, although there are not many examples in the literature yet.

1.3.2 Photometric redshift estimation

Machine learning has also been extensively applied to photometric redshift (photo-z) estimation and present an alternative to template based spectral energy distribution (SED) fitting methods (e.g. `LePhare`, Arnouts & Ilbert 2011; `BPz`, Benítez 2011; `ZEBRA`, Feldmann et al. 2006; `EAZY`, Brammer et al. 2008). A vast variety of machine learning algorithms has been used to tackle photo-z estimation like tree-based methods (e.g. Carliles et al. 2010; Gerdes et al. 2010; Carrasco Kind & Brunner 2013), support vector machines (SVM, e.g. Wadadekar 2005; Wang et al. 2008) and fully-connected ANN (e.g. Collister & Lahav 2004; Bonnett 2015a), the majority of them using photometric features to make redshift predictions.

Recently, efforts have also focused on determining photometric redshifts directly from astronomical images using CNN. D'Isanto & Polsterer (2018) compares the performance of traditional redshift estimation methods using photometric features with a deep CNN predicting photo-z from astronomical images. The paper presents a redshift precision comparable to the state of the art results on bright galaxies. Furthermore, Pasquet et al. (2019) determines the photometric redshifts of bright galaxies in the Main Galaxy Sample of the Sloan Digital Sky Survey at $z < 0.4$ with a CNN on the $ugriz$ images. In this thesis, we present a novel deep learning method to predict multi-band photometry and photo-z from images (§7).

1.3.3 Other applications

A decade ago, almost all the machine learning implementation examples would have related to object classification and photo-z estimation. Nowadays, the hype on machine learning has also reached astronomy, and machine learning implementations have been widespread in several other science cases.

Examples include galaxy deblending, which will become a crucial step in the data reduction for upcoming surveys like e.g. LSST. Traditionally, deblenders mostly relied on analytical modelling of the blended galaxies, which requires very accurate galaxy models. Recently, more robust deep learning deblending algorithms have also been developed (Boucaud et al., 2020; Arcelin et al., 2021). Neural networks have also been developed to correct shear measurements from nuisance effects (Tewes et al., 2019; Matilla et al., 2020), including e.g. instrument optics, blending, and unknown galaxy morphologies. Furthermore, Gupta et al. (2018) uses CNNs to derive cosmological constraints from weak lensing maps.

Chapter 2

The PAU Survey

In this chapter, we introduce galaxy surveys (§ 2.1), focusing on imaging surveys (§ 2.1.1) to introduce the PAU Survey (§ 2.2).

2.1 Galaxy surveys

A large fraction of the data collected from the universe arrives as electromagnetic radiation, e.g. low energy radio photons (Wilson, 2011; Lacy et al., 2020), very energetic gamma rays (Di Sciascio, 2019; Mazin, 2019), or optical astronomy (The Dark Energy Survey Collaboration, 2005; Martí et al., 2014; Benitez et al., 2014). Galaxy surveys are surveys of a portion of the sky that provide fundamental data basis of galaxies and their distribution in the Universe.

There are two widely used techniques to observe the Universe: spectroscopy and photometry. Spectrographs split the light in wavelength such that it is possible to measure the amount of light in small wavelength intervals. Spectroscopy measures the spectral energy distribution (SED), i.e. the amount of energy per second, unit area, and unit wavelength of any astronomical source using a spectrograph, which enables the estimation of very precise galaxy redshifts. In contrast, imaging surveys consist of imaging the sky using optical and near infra-red (NIR) photometric filters, which enables increasing the number of observed galaxies by ~ 2 orders of magnitude.

Ideally, galaxy surveys should cover wide sky areas with a fine angular and wavelength resolution. Unfortunately, astronomical observations are limited and a high wavelength resolution is commonly at expense of a fine angular resolution over a wide sky area (and vice versa). While spectroscopic surveys (e.g. Ahumada et al., 2020; Scodeggio et al., 2018; Driver et al., 2011) can provide very high-resolution spectra, they demand long exposure times. Moreover, spectroscopic surveys also require targeting the observations, which potentially causes target-selection effects due to e.g the surface brightness detection limit of the imaging data used to derive the targets. In contrast, imaging photometric surveys (e.g. de Jong et al., 2013; Ivezić et al., 2019a) present an alternative method that enables covering larger areas of the sky with better angular resolution but worsening significantly the wavelength resolution.

Advances in observational technology (i.e. telescopes and detectors) have enabled galaxy surveys to increase the collected data from very few galaxies to billions of them. Optical imaging sky surveys started in the pre-photography era with naked-eye observations. The first astronomical catalogue was set up by Messier in 1774 (Messier, 1774) and contained 110 astronomical objects. Other examples of pre-photography catalogues are e.g. the still-used New General Catalogue (Dreyer, 1888) and the Index Catalogue (Dreyer, 1895) by John Dreyer .

Photography and monitoring systems transformed sky surveys enabling a systematic coverage of large areas of the sky. In the first half of the 20th century, several sky surveys provided astronomical catalogues containing $\sim$thousands of objects, mostly stars. Some examples are the Smithsonian Astrophysical Observatory Catalog[1], which contained positions, proper motions, and magnitudes for over 250 000 stars and the Henry Draper Catalogue[2], containing the spectral type of $\sim$360 000 stars. Photography also enabled the monitoring of the Magellanic Clouds with the discovery of the crucial period-luminosity relations for Cepheids (Leavitt & Pickering, 1912) in 1912, later used for the Hubble discovery of the Universe expansion.

In the second half of the century, the development of Schmidt telescopes led to the POSS-I survey, a major milestone for galaxy surveys (Minkowski & Abell, 1963). POSS-I mapped about two-thirds of the observable sky from the Palomar Mountain providing catalogues such as the Morphological Catalog of Galaxies[3], of $\sim$30 000 galaxies. Furthermore, the first spectroscopic surveys were designed in the early eighties and provided the first evidence of the large scale structure in the nearby universe. The first Center for Astrophysics redshift survey (CfA, Geller & Huchra, 1983) was the first spectroscopic survey, which observed $\sim$2300 galaxy spectra down to $m_{\rm AB}$ $\sim$14.5.

The development of charged-coupled devices (CCD) brought unprecedented progress to astronomy with fully-digital sky surveys. CCDs are silicon chips made up of an array of light-sensitive diodes (pixels) settled in rows and columns that become charged when light hits them (Lesser, 2015). SDSS (Gunn et al., 1998; York et al., 2000) was the first CCD survey, which eventually covered 14 500 deg^2 and collected 116 TB of data (Alam et al., 2015). SDSS was fundamental in transforming astronomy and enabled a wide range of science applications. Technology advances also affected spectroscopic surveys with the development of multi-fibre spectrographs, opening to massive redshift surveys, e.g. 2dF (Colless et al., 2001a) and SDSS (Ahumada et al., 2020), which together provided more than a million galaxy redshifts.

2.1.1 Imaging surveys

The wavelength resolution of imaging surveys depends on the set of *photometric filters* (i.e. the *photometric system*). The photometric system is characterised by the number, width,

[1]https://heasarc.gsfc.nasa.gov/W3Browse/star-catalog/sao.html
[2]http://server6.sky-map.org/group?id=23
[3]https://heasarc.gsfc.nasa.gov/W3Browse/galaxy-catalog/mcg.html

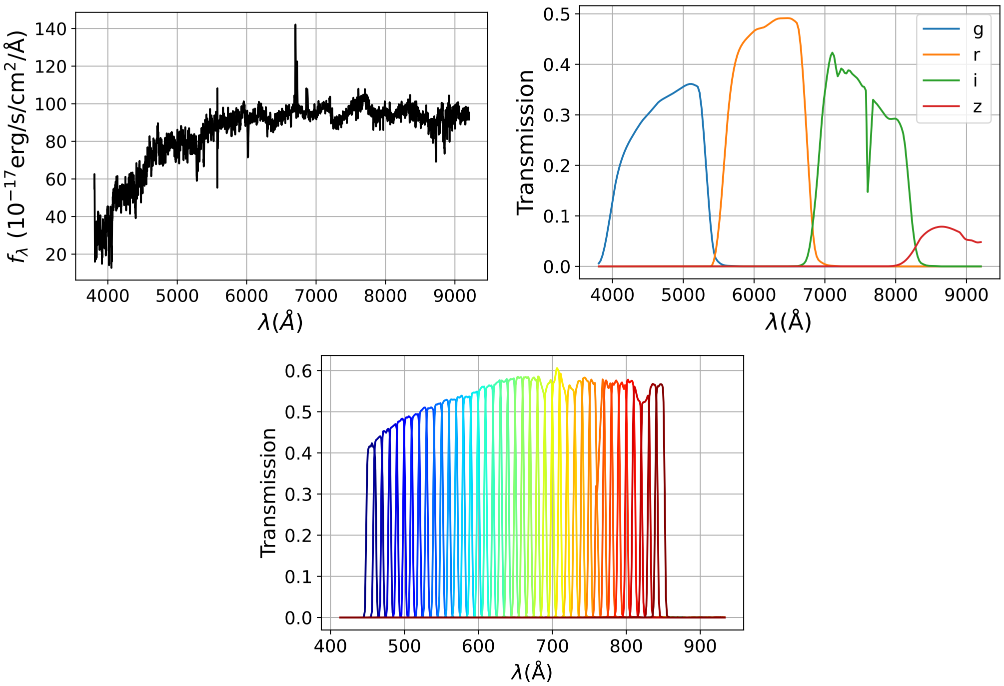

Figure 2.1: *Top left:* SDSS galaxy spectra. *Top right:* The SLOAN SDSS *griz* broad-band transmission curves. *Bottom:* The PAUS narrow-band transmission curves.

and wavelength coverage of the photometric filters. This includes broad-band systems with few photometric filters of width $\sim$1000Å (e.g. Honscheid & DePoy, 2008; Doi et al., 2010) and narrow-band systems (Molino et al., 2013; Padilla et al., 2016), which are made of a larger number of narrower photometric filters with $\sim$100Å width.

While broad-band photometric systems provide low spectral resolution, they enable observing large sky areas with great angular resolution. On the other hand, photometric systems with narrow-band filters increase the wavelength resolution but typically cover smaller sky areas since the survey needs to pass more times through the same sky region to cover the same wavelength range. Furthermore, narrow-band filters detect fewer photons than their broader counterparts for the same exposure time. This either yields in a signal-to-noise reduction, an increment of the exposure times required to observe a sufficient signal, or a trade-off between the two.

Figure 2.1 shows an example of galaxy spectra (upper-left) measured by SDSS. The right panel of the same figure presents the SDSS *griz* filter transmission curves (right) (Fukugita et al., 1996; Doi et al., 2010), while the bottom plot shows the PAUS narrow-band transmission curves (Casas et al., 2012). These plots evidence that broad-band imaging suffers a significant loss of wavelength resolution while narrow-band wavelength resolution is in-between broad-band imaging surveys and spectroscopy. Table 2.2.4 shows a few examples of spectroscopic and photometric surveys and their characteristics.

Astronomical images

The images captured by the CCD camera are named *raw images*. These are the primary source of data but are highly degraded by noise effects such as the turbulence of the atmosphere, charge inductions in the CCD electronics, and the telescope movement (Morganson et al., 2018). Other sources of noise such as e.g cosmic rays, very bright stars, and very massive galaxies can also affect the quality of the images.

Images are observed in a *field of view*, which defines the area of the sky that can be covered by the astronomical image. In ground-based telescopes, the atmosphere also affects the image by smearing out the light in a process named *seeing* (Trujillo et al., 2001). The seeing reduces the resolution of an astronomical image, lowering its mean surface brightness and increasing the observed radii. This is made evident in the image of a point-like source that should be captured by a single pixel, e.g. a distant star, spreading over a group of pixels.

The diffraction of the lens aperture and the fact that images are taken with discrete pixels (pixelisation) also contribute to the image spreading. The point spread function (PSF) quantifies the combination of all these effects (seeing, pixelisation, and telescope optics). The observed image is the galaxy image convolved with the PSF, which keeps the brightness of an object while spreading it on a larger group of pixels. Mathematically, the value of a pixel $\tilde{\mathcal{I}}_{x,y}$ placed at x and y convolved with the PSF is

$$\tilde{\mathcal{I}}_{x,y} = \sum_{i=-a}^{a} \sum_{j=-b}^{b} K_{i,j} \times \mathcal{I}_{x+i,y+j}, \tag{2.1}$$

where $\mathcal{I}$ is the galaxy image without PSF effects and $\boldsymbol{K}$ is the PSF kernel, which is assumed to have dimensions $(2a+1, 2b+1)$. The PSF can be defined by any mathematical function, e.g. a Gaussian. In astronomy, the PSF is most commonly modelled by the *Moffat* function, which is less sharp than a Gaussian (Eq. 3.15)

2.2 The PAU Survey

Large maps of galaxy tracers are a key ingredient for many cosmological studies. The Physics of the Accelerating Universe Survey (PAUS) is a 40 narrow-band imaging survey observing at the William Herschel Telescope, in La Palma (Spain). As of June 2022, PAUS has imaged 40 deg^2 of the sky. The observed astronomical fields are the COSMOS field (2 deg^2) and a fraction of CHFTLS wide fields W1 (10 deg^2), W2 (10 deg^2) and W3 (20 deg^2). Moreover, PAUS is also targeting the CHFTLS-W4 field, for which it still has very few observations.

2.2.1 History

Back in 1998, two independent teams of astronomers found that the Universe's expansion was speeding up (Riess et al., 1998; Perlmutter et al., 1999). This discovery had strong implications for the understanding of the Universe. Dark energy was postulated as the most

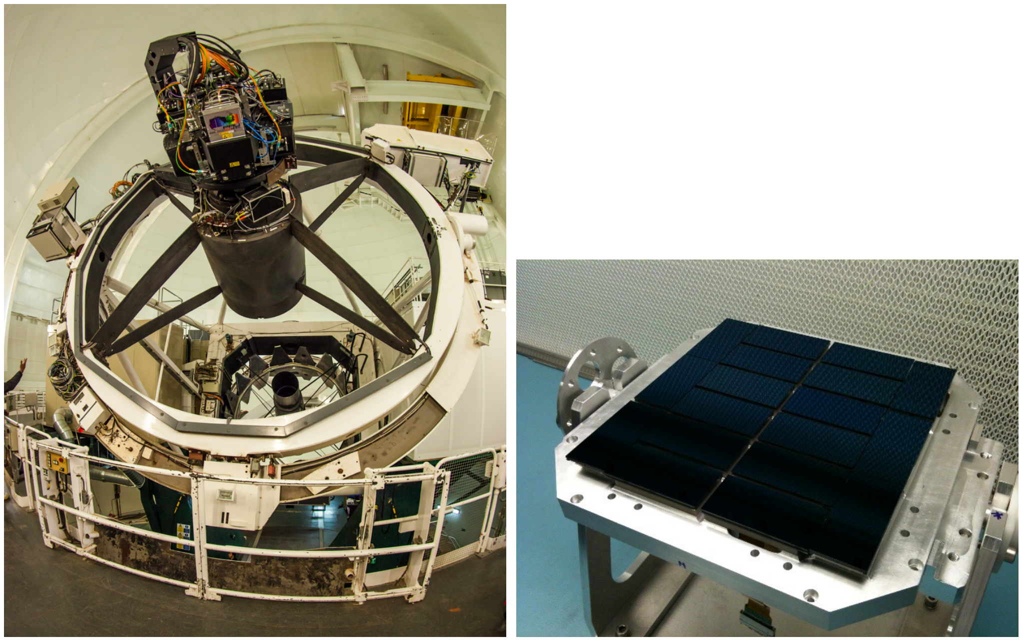

Figure 2.2: *Left:* PAUCam vessel and electronics installed at the primer focus of the WHT. *Right:* The eighteen Hamamatsu CCDs used for narrow and broad band imaging.

significant component of the cosmos and that responsible for the accelerated expansion, but its nature was unclear. In the following years, the scientific community focused on understanding the nature of dark energy and lots of efforts and funding were invested in its research. In 2005, the DES started as a cosmological survey using different probes to investigate the potential time-evolution of dark energy. This is a ground-based cosmological survey with a five broad-band photometric filters camera (DECam, Honscheid & DePoy, 2008) mounted on the Victor M. Blanco Telescope, at the Cerro Tololo Inter-American Observatory (CTIO) in Chile. DES is a large collaboration formed by institutions from all around the world, including both The Institut d'Altes Energies (IFAE) and the Institut de Ciències de l'Espai (ICE). While ICE had already been active in astronomical surveys, e.g. SLOAN (Margon, 1999), IFAE was at that time mainly involved in particle physics, especially in the ALEPH experiment (Wu, 1986). When ALEPH finished, IFAE wanted to diversify and joined the DES cosmology experiment. Initially, IFAE's main contribution was the readout electronics of the DECam.

In 2005, the Spanish government announced the Consolider Program, which was intended to foster research projects lead by Spanish institutions while stimulating the formation of research networks in the Spanish scientific community. The Consolider program was providing funds for five-year research projects, boosting new research lines. IFAE and ICE together with other Spanish institutions including the Universidad Autónoma de Madrid (UAM) and the Instituto de Física Teórica (IFT), presented a joint proposal that consisted in building an astronomical camera from scratch. Some of these institutions knew each other from DES, while others had been collaborating in the ALHAMBRA survey (Moles et al., 2008). The project was named The Physics of the Accelerating Universe (PAU[4]) and was selected by the Consolider Program in 2007. The survey name was later extended to PAUS.

ALHAMBRA (Moles et al., 2008) was a foregoing Spanish galaxy survey that consisted of 20 contiguous, equal width, medium-band photometric filters covering from 3500Å to 9700Å, plus the standard broad bands JHK near-infrared broad bands. A significant fraction of the PAUS collaboration had already participated in ALHAMBRA. This brought up the idea of extending PAUS as a natural continuation of ALHAMBRA, performing a large survey by constructing a camera with 40 narrow-band photometric filters.

2.2.2 PAUCam characteristics and construction

Before constructing the PAUS camera (from hereafter PAUCam), PAUS needed to locate a telescope where it could be installed. A potential was constructing a new telescope in Teruel, Spain. The location was initially characterised to study the viability of mounting PAUCam there. However, building the telescope, the camera, and the observatory was out of the scope of the Consolider program. In 2008, the Isaac Newton Group (ING) opened a call for visitor instruments for the William Herschel Telescope (WHT). PAU presented a proposal for mounting PAUCam at the WHT, which was accepted, achieving the first milestone for the PAUS project.

The construction of PAUCam started in 2010 and took five years. A major challenge was the WHT weight requirement. Instruments installed at the WHT prime focus cannot weigh > 270kg. This weight must include e.g. the vacuum pumps, the cryogenic system, the filters, the motors required to exchange the filter trays, and the electronic systems (including all the cable connections), which imposes a stringent limitation on the camera design. Astronomical cameras are typically built with aluminium, but this material would have exceeded the WHT weight limit.

Carbon fibre is a lighter alternative to aluminium, however, there were no previous cameras built with carbon fibre in cryogenics. This material is a mixture of carbon and epoxy, and type of epoxy must to be non-outgassing and therefore approved in a special list that NASA or ESA provides. After several tests at IFAE labs, it was decided to manufacture the camera vessel and many parts of the filter exchange system with carbon fibre. The PAUCam team had to work closely with the carbon fibre company to ensure they were using the right type of epoxy. Figure 2.2 shows the PAUCam carbon fibre vessel with all the electronics installed at the prime focus of the WHT (left).

Reducing the PAUCam weight also enables a faster camera installation, which is valuable since it is not the only instrument observing at the WHT. For this reason, PAUCam is also equipped with a dual cooling system to enable the cooling of the focal plane within hours. During regular operations, two cryo-tigers maintain the temperature of the narrow-band fil-

[4] In Catalan, PAU means peace.

ter trays, the CCDs, and the amplifiers. However, after the installation of the camera, the cryo-tigers would take ∼12 hours to cool the system to the operational temperature. Thus, the PAUCam is also equipped with a liquid nitrogen cooling system to ensure it is operational on the same day of the installation.

The PAUCam fully covers the telescope FoV with eighteen fully-depleted Hamamatsu CCDs of 2k × 4k pixels with a pixel size of 15 μm (Fig.2.2). A *mosaic image* is a sky exposure of the full 18-detector PAUCam mosaic (see the left panel on Fig. 3.2). The CCDs are divided into eight central CCDs used for narrow-band imaging and ten external CCDs for guiding calibration, and broad-band imaging. The CCD characterisation was done at IFAE labs by illuminating the CCDs with a Neon lamp multiple times to register the response to multiple wavelengths.

The camera has two optical elements: the entrance window and the photometric filters. The former permits the entrance of light while keeping the camera vessel in vacuum. The filters are divided into two differentiated sets:

i *The broad-band filter set*: This set is composed of six broad-band filters $ugrizY$ calibrated at the ICE-CSIC/IEEC/IFAE labs and with the same filter transmissions as the DECam photometric filters (Honscheid & DePoy, 2008).

ii *The narrow-band filter set*: It consists of 40 narrow-band filters spanning a wavelength range from 4550Å to 8450Å, which were calibrated at CIEMAT (Casas et al., 2016). The narrow-band filters have transmission curves with a FWHM of 1300Å and 1000Å of separation between consecutive bands. PAUCam distributes the 40 narrow-band filters in five independent filter trays. Each filter tray is equipped with eight narrow-band filters covering the eight central CCDs, which permits imaging with eight narrow-band filters simultaneously.

The broad- and narrow-band photometric filters are located between the last corrector lens and PAUCam's focal plane, at a minimum distance between the lenses and the filters. This is important since PAUCam is equipped with many small filters that have a structure in the filter trey. Consequently, larger distances between the tray and the focal plane increase the shade that the tray produces on the focal plane. The filter trays are in a vacuum environment and cooled to 250K.

The PAUCam camera has two prime control systems: the PAUCam Slow Control System (SC) and the PAUCam Control System (CCS). The SC is in charge of the telescope motion control, monitoring the sensors (e.g. temperature and pressure), and first safety reaction. It also controls the filters' jukebox movements, the shutter, the temperature monitoring, the vacuum sensor, and pump monitoring, and the slow control of the power supplies. On the other hand, the CCS controls all the PAUCam subsystems necessary to capture images. These subsystems include the Slow Control interface (SCi), in charge of sending commands to the SC; the Data Acquisition interface (DAQi), which is used to read out the PAUCam CCDs; the Online Quality analysis; the Guider, the Telescope Control System interface (TCSi), Storage, and Alarms.

2.2.3 PAUCam commissioning

In 2015 the camera was ready for commissioning and the commissioning team went to La Palma for two weeks. One of the main concerns was the grounding, which is unstable as La Palma resides in a volcanic region. All the electronics must be placed correctly, otherwise, insufficient grounding leads to a high readout noise. For this reason, they performed multiple grounding tests at IFAE labs before sending the camera. The very first image had a very high read-out noise of 70 electrons, which was reduced to 9-10 electrons after adjusting the grounding scheme. The readout noise has since remained stable.

The PAUCam commissioning team brought the camera to the telescope two weeks before the first night of observations. During those days, the team was working on the readout system and the software to build images in FITS files. They managed to take sky images already on the first night, however the the monitoring displays and the automatic scripting was not ready and they had to use command lines to e.g. execute the image, move the telescope, and save the FITS files. The observing system software managing and coordinating all the systems (e.g. the reading system, the filter trays, the mechanics, the telescope, temperature, and sensors) was developed while taking data.

During the first observation shifts, some issues in the operation of the camera were detected. The filter trays were stack from time to time, which required to move the camera in different position (e.g. telescope declination and rotator) and try to increase the torque of the motors until the tray was unstuck. Furthermore, a software error forced the observers to reboot the camera about twice a night. Most importantly, the image quality was degraded due to a significant amount of scattered-light (see §4.2). In PAUCam, scattered light was caused by the filters being installed slightly tilted to be perpendicular to incoming light. This caused a gap between the filters and the filter trays (Romanishin, 2014). Scattered-light was reducing the number of observed galaxies per image and their signal-to-noise.

In order to fix these shortcomings, the camera was shipped back to Barcelona. PAUCam was opened in 2015 to redesign the filter trays. Initially, these were inclined following the optical path and light was entering through since the filters did not have coating on their sides. In the camera intervention, the filters were modified to be completely parallel to the focal plane and CCDs. Furthermore, the borders of the filters were slightly increased to minimise light entering through their sides. The camera intervention drastically reduced scattered light by a factor of 4. The camera has worked smoothly until nowadays.

2.2.4 Science goals

Thanks to the narrow-band filter set, PAUS provides a low-resolution spectra ($\lambda/\Delta\lambda \sim 50$) for few million redshifts on the full sample (without selection effects) down to $i_{\rm AB} < 22.5$. This high wavelength resolution reduces the photo-z uncertainty by $\sim$15 times, which in turn reduces the error on other derived physical quantities such as e.g. the luminosity function. Given the uniqueness of the sample, PAUS has great potential for

- ☐ Studying intrinsic alignments at $z \sim 0.75$ due to increasing the galaxy density with sub-percent photo-z almost two orders of magnitude. This has been studied both on simulations (Stothert et al., 2018) and CFHT-W3 data (Johnston et al., 2021a).

- ☐ Galaxy evolution studies. Renard et al. (2021) performs Ly-α intensity mapping cross-correlating the spectroscopic Ly-α forest data with the background of narrow-band images from PAUS, while Renard et al. (2022) precisely measures the D4000 spectral break.

- ☐ Providing redshift calibration samples for large imaging surveys like DES, KiDS, LSST, and *Euclid*. Alarcon et al. (2021) uses a combination of 26 narrow-, intermediate-, and broad bands together with PAUS narrow bands to enable an unprecedented precision photo-z catalogue in the COSMOS field. Furthermore, § 6 introduces a deep learning methodology based on multi-task learning to improve broad-band photo-z in the wide fields using narrow-band photometry (Cabayol et. al in prep.).

- ☐ Improving and modelling of target selection for spectroscopic surveys as e.g. DESI (DESI Collaboration et al., 2016) or WEAVE (Dalton et al., 2014).

Survey	Type	Area (deg^2)	Mag. limit	Galaxies	Filters	Period
The Sloan Digital Sky Legacy Survey (SDSS York et al., 2000)	Spectrosopic	7500	-	1M	-	2000-2008
The Baryon Oscillation Spectroscopic Survey (BOSS Dawson et al., 2013)	Spectrosopic	10000	-	1.5M LRG 160000 Quasers	-	2008-2014
The Extended Baryon Oscillation Spectroscopic Survey (eBOSS Zhao et al., 2015)	Spectrosopic	16000	-	300000 LRG 189000 ELG 572000 Quasers	-	2014-2020
The Dark Energy Survey Instrument (DESI DESI Collaboration et al., 2016)	Spectrosopic	14000	-	4M LRG 18M ELG 1.4M Quasers	-	2021 -
(*Euclid*, Laureijs et al., 2011)	Spectrosopic Photometric	15000	YJH <24	Spectroscopic:50M Photometric :1.5B	$rizYJH$	2022 -
Canada-France-Hawai Telescope Legacy Survey (CFHTLS Cuillandre et al., 2012)	Photometric	171	i_{AB} <24		$ugriz$	2004-2009
The Kilo-Degree Survey (KiDS de Jong et al., 2013)	Photometric	1500	i_{AB} <24.2	90M	$ugri$	2011 -
The Dark Energy Survey (DES Abbott et al., 2018b)	Photometric	5000	r <24	300M	$grizy$	2013 -
The Physics of the Accelerating Universe Survey (PAUS Martí et al., 2014)	Photometric	100	$i_{\mathrm{AB}} \approx 23$	2M	40 narrow bands	2015 -
Large Synoptic Survey Telescope (LSST Ivezić et al., 2019a)	Photometric	20000	i_{AB} <23.9	4B	$ugrizy$	2022 -

Table 2.1: Summary of some of the most important spectroscopic and photometric galaxy surveys of the last two decades.

Chapter 3

Image processing, photometry and photometric redshifts

Reducing images to galaxy catalogues is a detailed and precise process that can be split in three main categories: image detrending (§ 3.1), photometry (§ 3.2), and further-processing (§ 3.3).

3.1 From raw to science images

Raw astronomical images contain several noise effects that need to be corrected before any science measurement. Converting raw images into science images requires several actions such as the bias subtraction (3.1.1), flat-fielding (3.1.2), dark-current correction (3.1.3), and artefact detection (3.1.4).

3.1.1 Bias subtraction

CCD detectors have an intrinsic additive electronic offset (bias current) that ensures all pixels collect a non-zero count. This noise is independent of the exposure time and is corrected by subtracting bias frames from the astronomical image. Bias frames are images taken with no light (shutter closed) and zero exposure time taken under the same camera conditions (e.g. the temperature and the gain) as the astronomical images to be corrected. Since the bias current fluctuates due to the read-out noise, the master-bias frame used for the correction combines between five and ten bias frames (Gilliland, 1992).

3.1.2 Flat fielding

The optical and CCD setup have spatially varying sensitivity across the CCD. This causes a multiplicative bias with different potential origins, such as dust and scratches in the optics and pixel sensitivity variations.

Flat frames aim to correct these effects and are obtained by imaging a uniform and isotropic source of light illuminating the dome screen (Massey & Jacoby, 1992; Zhou et al., 2007). The pixels' sensitivity also depend on the wavelength thus flat fields are taken for all photometric filters independently. Furthermore, as in the case of the bias frames, several flat fields are observed and combined into a single master flat frame. The flat-field corrected image ($\mathcal{I}'$) is

$$\mathcal{I}'(x, y) = I(x, y) \; / \; F(x, y) \,, \tag{3.1}$$

where $\mathcal{I}(x, y)$ is the bias-corrected image, x and y refer to the pixel coordinates, and $F(x, y)$ is the master flat frame.

The quality of flat frames is compromised by the presence of scattered-light residuals (§4.2.2). This effect cannot be corrected either with bias frames, since scattered light appears in the presence of light, or with flat frames, as scattered light is an additive effect. Flat fields have two main frequency components: a high pass pixel due to pixel sensitivity variations and a low-pass band caused by vignetting. The scattered-light frequency is in-between these two, enabling the correction of scattered-light residuals in the flat fields.

3.1.3 Dark current

Dark currents are additional charges in the CCD detectors caused by thermal fluctuations (Widenhorn et al., 2010). Since dark-current fluctuations quickly decay with temperature, these are avoided by cooling the CCD with (typically) liquid nitrogen (Bogget et al., 2014). Survey cameras like DECam (Honscheid & DePoy, 2008) and PAUCam (Padilla et al., 2016) implement this technology to reduce dark currents.

3.1.4 Cosmic rays and other spurious artefacts

Cosmic rays and saturated trails are other sources of CCD artefacts. The left panel of Fig. 3.1 shows an image affected by cosmic rays, which are thin-bright lines in the upper left corner of the cutout. The right panel contains saturated trails caused by close stars. These trails are bright horizontal and vertical lines crossing the stars.

Cosmic rays are high-energy particles origined outside the solar system that rain down on Earth. Their origin is still not fully known, although some hypotheses suggest that it is related to supernovae (Blasi, 2013). When a cosmic ray hits the CCD, this releases a cascade of electrons that causes a bright trace in the image. Sometimes, the decay of atoms used in the construction of the CCD can also trigger the same type of bright stripes.

Cosmic rays are typically very easy to recognize as they are very sharp and bright and appear randomly in the image. Nevertheless, removing them without harming other sources is difficult. There are already several algorithms to correct for cosmic-ray traces, which apply both traditional statistical methods and machine-learning techniques (Desai et al., 2016a; Pych, 2004; Zhang & Bloom, 2019).

Saturated trails are caused by an excess of photons illuminating a pixel region of the CCD. The amount of electrons that a pixel can store is limited. Therefore, if a pixel is illuminated with a very bright light source (or for too long exposure time), it gathers the

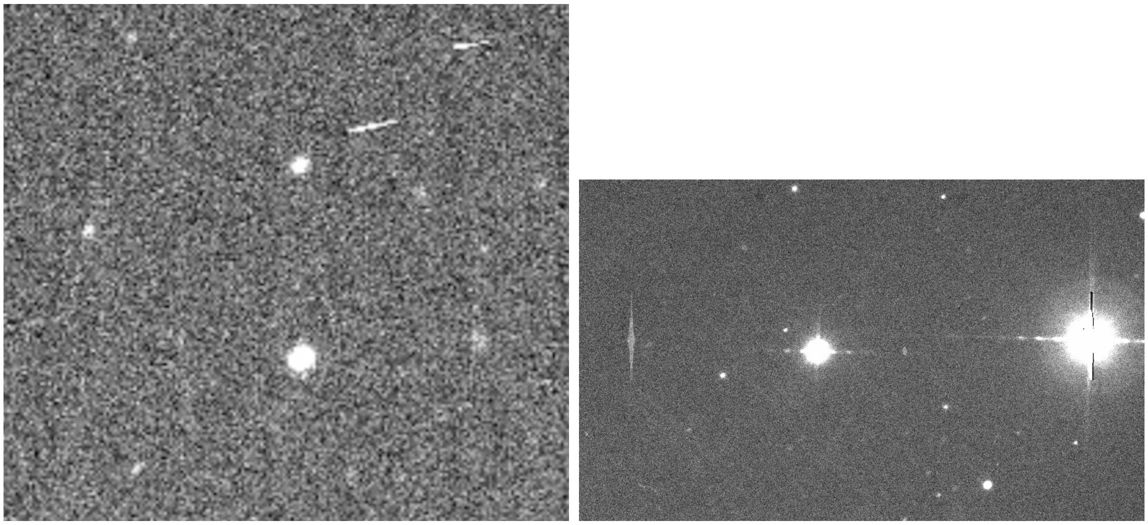

Figure 3.1: *Left:* Astronomical images with two cosmic rays in the upper right corner. *Right:* Bright star with two saturated trails crossing from side to side and from top to bottom. *Image credit: Serrano et al. (in prep).*

maximum number of electrons it can accumulate (i.e. the pixel saturates). As a consequence, during the image read-out, the extra electrons fill in the pixel row that contains the saturated pixel. This appears in the image as columns of very bright pixels containing the extra electrons from the saturated pixel. There are several techniques to remove these trails. For example, Desai et al. (2016b) implements an interpolation approach that uses the local PSF to remove artefacts, while Paillassa et al. (2020) uses a CNN to identify the contaminants. Dark currents are additional charges in the CCD detectors caused by thermal fluctuations (Widenhorn et al., 2010). Since dark-current fluctuations quickly decay with temperature, these are avoided by cooling the CCD with (typically) liquid nitrogen (Bogget et al., 2014). Survey cameras like DECam (Honscheid & DePoy, 2008) and PAUCam (Padilla et al., 2016) implement this technology to reduce dark currents.

3.2 Photometry

The amount of light passing through a photometric filter (Fig. 2.1) is the *flux* and the technique of measuring fluxes is *photometry*. We can broadly divide photometry into four fundamental steps:

Object detection: Detecting astronomical sources can be challenging, mostly due to the low signal-to-noise of astronomical images. Moreover, images of celestial objects do not have defined boundaries, and several sources can overlap in the same image region. There are multiple approaches to performing object detection. Some example implementations are smoothing the background to detect pixel fluctuations above a fixed threshold (Damiani et al., 1997), using principal component analysis to distinguish between objects and background noise (Andreon et al., 2000), and implement Bayesian techniques to detect objects based on prior knowledge (Guglielmetti et al., 2009).

Forced photometry is an alternative to detection photometry. It cross-matches astronomical images with external deeper catalogues to locate the target sources position (Ni et al., 2019). It is a strong alternative for surveys with low signal-to-noise images where sources are hard to detect. However, forced photometry is not very common since it requires re-observing a field with already existing deeper observations. This is the approach used in PAUS (Serrano et al. in prep.).

Centroiding: This consists in determining the centre of the target object in the image. The most straightforward measurement is drawing the 1D galaxy profile in the x and y dimensions and fitting them to a Gaussian. The centre of the object is estimated from the centre of the Gaussian best fitting the profile. This procedure is not optimal for low SNR objects, where objects are noisy and unresolved. In such cases, one potential solution is forced photometry, where both the detection and the centroiding are obtained from deeper, higher signal-to-noise images.

Background estimation: The night sky has an intrinsic brightness that produces a non-zero signal even if the telescope is not pointing to any source. Every pixel in the CCD captures light from the night sky, including those receiving light from target sources. Moreover, pixels also contain electronic noise.

Measuring the flux from the sources requires the estimation and subtraction of the sky background and electronic-noise contributions. There are several methods to estimate the sky background. One very widely implemented technique is placing an annulus centred at the target source and measuring the mean sky background ($\bar{B}$) as the average number of electrons in the pixels within this annulus, i.e.

$$\bar{B} = \frac{1}{M} \sum_{R_{\rm in}<r<r_{\rm out}} x(r)\,, \tag{3.2}$$

where M is the total number of pixels within the annulus. The standard deviation of the pixels within the background annulus is an estimator for the background error per pixel $\sigma_{\rm b/pix}$, which relates to the the error in the mean background ($\bar{\sigma_{\rm b}}$) as

$$\bar{\sigma}_{\rm b}^2 = \frac{\sigma_{\rm b/pix}^2}{M}\,. \tag{3.3}$$

The characteristics of the astronomical images, e.g. strongly-varying background noise or high read-out noise, might benefit from implementing different background estimation approaches. `SExtractor` (Bertin & Arnouts, 1996) is a very widely used algorithm that implements aperture photometry. It meshes the background and reconstructs a 'background map' with a background estimation at each particular mesh location.

DAOPHOT (Stetson, 1987) is an algorithm for crowded-field stellar photometry that measures the background with a least-squares fit to the data in the region around the source assuming Gaussian star profiles and BKGnet (§4 and Cabayol-Garcia et al. 2020), which is a CNN optimised to predict the background noise on images with strongly-varying backgrounds.

Flux estimation: The last step is to measure the light emitted by the source. Aperture photometry estimates the flux summing all the pixel contributions within a circular or elliptical aperture centred around the target object (Mighell, 1999). As previously mentioned, the flux within the aperture ($f_{\rm tot}$) contains light from both the target source and the sky background. It also contains electronic noise, which is also captured in the background annulus ($\bar{B}$, Eq. 3.2). Therefore, the total flux of the source ($f_{\rm src}$) is estimated by

$$f_{\rm src} = f_{\rm tot} - \bar{B}N\,, \tag{3.4}$$

where N is the number of pixels within the aperture.

The uncertainty in $f_{\rm src}$ ($\sigma_{\rm src}$) is

$$\sigma_{\rm src} = \sqrt{\frac{f_{\rm src}}{t_{\rm exp}} + \left(N + k\frac{N^2}{M}\right)\sigma^2_{\rm b/pix}} \tag{3.5}$$

and contains contributions from source shot noise (first term in the RHS of Eq. 3.5), the uncertainty in the background noise estimation (second term), and the error due to the background subtraction (third term). In Eq. 3.5, the factor k depends on the operation to estimate the background, e.g. $k = 1$ and $k = \frac{\pi}{2}$ for backgrounds estimated as the mean and median of the pixels within the annulus[1], respectively and $t_{\rm exp}$ is the exposure time.

The signal-to-noise of the flux measurement depends on the aperture size. Large apertures measure a larger fraction of source light, but these also capture more noise. There are different approaches to executing aperture photometry. As an example, the DES first-year analysis implements fixed-size apertures (Drlica-Wagner et al., 2018), i.e. the radius of the aperture is the same regardless of the size of the target object. In contrast, PAUS estimates the aperture size comprising a fixed fraction of galaxy light (e.g. typically 62.5%, see § 3.4.2).

3.2.1 Photometry calibration

Photometry produces flux measurements that still need to be corrected for effects such as e.g. extinction.[2] The *photometric calibration* is the process that converts the fluxes from instrumental counts to physical calibrated measurements of the energy from an astronomical

[1]http://wise2.ipac.caltech.edu/staff/fmasci/ApPhotUncert.pdf.

[2]Extinction is the dimming of light due to the collision of photons with the atmosphere and interstellar medium. Photons are scattered/absorbed by dust particles of similar size as the wavelength, which has a higher impact on the bluer part of the spectrum.

object before entering the atmosphere. The calibration step is crucial to improve the quality of the photometry. It compensates for the atmospheric extinction and other optical effects affecting light before hitting the CCD.

There are several example implementations of photometric calibration. As an example, the Dark Energy Survey first-year analysis (Abbott et al., 2018b) used the Southern SDSS $u'g'r'i'z'$ (Smith et al., 2005) and the SDSS Stripe 82 *ugriz* standard stars converted into DES *griz* system to calibrate the photometry (Tucker et al., 2007), since the SDSS and DES filter responses are similar (Doi et al., 2010; Honscheid & DePoy, 2008). Similarly, the Kilo-Degree Survey DR2 (KiDS, de Jong et al., 2015) estimates every night individual zero-points per CCD using the SDSS DR8 stars from the SA field as a reference. A relative calibration amongst CCD images is estimated by comparing the photometry of independent observations of the same source. Applying a minimisation algorithm, zero-point differences due to varying atmospheric extinction are derived from overlapping sources across exposures. Then, the photometric offsets are applied to the CCD with respect to the zero-point calculated for that night derived from the nightly SA observations.

3.3 Further-processing: co-adding flux measurements and derived properties

In this section, we introduce three further steps in obtaining photometric catalogues: co-addition (§ 3.3.1), photometric redshift estimation (§ 3.3.3), and object classification (§3.3.2).

3.3.1 Co-addition

Imaging surveys take several observations of the same source with the same photometric filter. The galaxy photometry can further optimise combining the information from all its exposures in a process named *co-addition*. The co-addition can be implemented either at the image level as part of the data pre-processing (Zackay & Ofek, 2017a,b) or at a flux-measurement level. Image co-addition increases the depth of the observations and removes artefacts such as cosmic rays. However, it is a non-trivial process that requires aligning the images and homogenising the PSF of the single-epoch observations.

DES implements image co-addition (Darnell et al., 2009). Their method remaps the flux values in the individual overlapping CCDs into a uniform pixel grid. The grid is constructed of artificial tiles of 1 deg on each side. For every tile, `SWarp` (Bertin et al., 2002) produces a co-added image. To homogenise the PSF, `PSFex` (Bertin, 2011) models the PSF of single epoch images.

Co-adding the measured fluxes does not require aligning the single epoch images or PSF smoothing (Serrano et al. in prep), thus it is simpler to implement. This co-addition method increases the SNR with respect to measurements from single observation images and enables the detection of transient artefacts. However, unlike co-added images, it does not enable deeper galaxy detection. Nevertheless, this is not problematic for surveys implementing forced photometry, e.g. PAUS, as they do not require source detection (§ 3.4.2).

3.3.2 Object classification

Accurate classification of astronomical objects, and particularly of foreground stars and background galaxies, is required to examine the nature of the universe. This is crucial for science applications such as intrinsic alignment studies (Troxel & Ishak, 2015), which demand very pure galaxy samples, i.e. samples without stars or quasars. Most conventional techniques differentiate extended sources (galaxies) from point-like ones (stars) using cuts in a magnitude-radius space (MacGillivray et al., 1976; Heydon-Dumbleton et al., 1989; Yee, 1991). This is sufficient for bright sources, but contamination from unresolved galaxies and noisy stars affects the purity of the samples at fainter magnitudes.

Machine learning classifiers are a powerful alternative in the astronomical object classification problem (§ 1.3.1 and Machado et al. 2016). There are several implementations to the star/galaxy classification problem. Examples include deep convolutional neural networks classifying images (Kim & Brunner, 2016) and boosted decision trees using catalogue image-derived morphology quantities (Sevilla-Noarbe & Etayo-Sotos, 2015). Also, a purely photometry based machine learning algorithm that uses the 40 flux measurements in the optical narrow bands has been developed using PAUS data (Cabayol et al., 2019).

3.3.3 Photometric redshift estimation

Redshift is the relative difference in wavelength between the observed (λ_{obs}) and emitted (λ_{em}) photon wavelengths

$$z \equiv \frac{\lambda_{\mathrm{obs}} - \lambda_{\mathrm{em}}}{\lambda_{\mathrm{em}}} \,. \tag{3.6}$$

There are three main sources of galaxy redshift: strong gravitational fields (gravitational redshift), the Doppler effect due to the relative velocity (v) of objects moving apart

$$1 + z = \sqrt{\frac{1 + v/z}{1 - v/z}} \,, \tag{3.7}$$

and the expansion of the universe (cosmological redshift)

$$1 + z = \frac{1}{a(t)} \,. \tag{3.8}$$

The scale factor $a(t)$ relates the physical distance with the comoving distance (Gray & Dunning-Davies, 2008).

Measuring precise photo-z is fundamental for most cosmological applications. In the last decades, there has been a huge effort to improve photo-z estimations, which has resulted in a wealth of different techniques (Zheng & Zhang, 2012b; Salvato et al., 2019a). These techniques aim to infer the redshift (z) or the probability distribution $P(z|f)$ given a vector of features (f), typically the galaxy photometry. Broadly, there are two techniques to infer photometric redshifts: template fitting and data-driven (machine learning) methods. The former technique introduces rest-frame template models (t) describing the galaxy spectral energy distribution (SED). Continuum and emission line SED templates are, in general, linearly combined with mixing coefficients (α) to shape a synthetic model $T(z)$

$$T(z) = \sum_{\mathrm{j}}^{n} \alpha_{\mathrm{j}} t_{\mathrm{j}}(z) \,. \tag{3.9}$$

These models are redshifted and convolved with the transmission curves from the photometric filters to provide a set of synthetic flux measurements at different redshifts (f_{model}). To estimate the redshifts, the synthetic fluxes are fitted to the observed galaxy fluxes (f_{obs})

$$\chi^2(z, A) = \sum_{b=1}^{N} \frac{(f^b_{\mathrm{obs}} - A f^b_{\mathrm{model}})^2}{\sigma^b_{\mathrm{obs}}} \,, \tag{3.10}$$

and summed over the band (b). A is a free factor in the modelling for the total flux and σ_{obs} is the uncertainty on the observed fluxes. Implementation examples include `HYPERZ` (Bolzonella et al., 2000), `ImpZ` (Babbedge et al., 2004), and `BCNz2` (Eriksen et al., 2019).

Photo-z estimation methods suffer from colour-redshift degeneracies that allow for multiple peaked photo-z probability distributions. Bayesian template-fitting methods use Bayesian priors to constrain the photometric redshifts with additional information. These methods compute *a priori* a set of model SEDs that encode the prior knowledge of the galaxies in the sample. Afterwards, the method determines the model with the lowest χ^2 to estimate the probability that the galaxy is observed at a given redshift and true SED template (t): $P(f|z, t)$

$$P(z|f) = \sum_{\mathrm{t}} P(z, t|f) = \sum_{\mathrm{t}} P(f|z, t) P(z, t) \,, \tag{3.11}$$

where $P(z, t)$ is the prior probability to have a galaxy with redshift z and template t. `BPZ` (Benítez, 2011) and `ZEBRA` (Feldmann et al., 2006) codes are examples of Bayesian template-fitting methods.

Machine learning methods attempt to learn the mapping between photometric space and redshift using a galaxy sample with known redshifts. Machine learning photo-z have been studied with several algorithms: decision trees and random forests (Carrasco Kind & Brunner, 2014b; Merloni et al., 2012); SVM (Zheng & Zhang, 2012a; Han et al., 2015); Gaussian processes: (Almosallam et al., 2015; Soo et al., 2021); and fully-connected neural networks: (Collister & Lahav, 2004; Cavuoti et al., 2015; Eriksen et al., 2020).

More recent studies have also implemented CNNs to predict photo-zs from astronomical images. As an example, Pasquet et al. (2019) implements a deep CNN on images from the SDSS Main Galaxy Sample, which contains 500 000 *ugriz* images with spectroscopic redshift $z_{\mathrm{s}} < 0.4$. This method predicts the best photo-z obtained so far on this galaxy sample. Another example is D'Isanto & Polsterer (2018), which also implements a CNN that combined with an MDN predicts the $p(z)$ for galaxies and quasars from SDSS DR9 *ugriz* images. Recently, Dey et al. (2021a) has implemented a capsule network to obtain photo-z also for the SDSS Main Galaxy Sample. To the best of our knowledge, photo-z from astronomical images has only been implemented on broad-band images, which have a significantly higher signal-to-noise than their narrow-band counterparts. Furthermore, most of the studies have been implemented in bright galaxies typically with low redshift.

3.4 PAUS data management

The data management of PAUS coordinates the transferring, archiving, calibration, data reduction, and distributing of the data products. The PAUCam generates about 350 GB of data every night, which requires storage, massive parallel processing, and a fast network connection. The Port d'Informació Científica (PIC) is the data center that provides the infrastructure to manage PAUS data (Tonello et al., 2019). While observing, the images transfer from the telescope's local archive to PIC, which takes about 3 hours. Once the data have been transferred, these are registered and the observation's metadata are stored in the PAU data management database (PAUdm).

PAUS data are permanently stored in magnetic tapes. These are cheaper than traditional spinning disks, occupy a small physical space, and are excellent for long term storage. However, random access to magnetic tapes is slow, e.g. accessing a file can take around one minute, as the volumes need to be searched and mounted. Data requiring fast access are pre-staged in a disk buffer. Databases provide easy access to large volumes of reduced data, in such a way that the data are sortable and easily searchable using queries on the data. The PAUS database (PAUdb) is managed using `Postgress`.

Once the data are transferred, the `Nightly` pipeline performs the basic data reduction and astrometry corrections (§3.4.1). The reduced science images produced in the `Nightly` pipeline are also stored in the archive-database system to proceed with the photometry measurements with `MEMBA` (§ 3.4.2).

3.4.1 From raw to science images: the `Nightly` pipeline

The left panel of Fig. 3.2 shows an example of a raw mosaic image with eight central and ten peripherical CCD observations. The right panel presents a single CCD science image of an uncommonly large extended source. The `Nightly` pipeline produces the astrometrically and photometrically calibrated science images from the raw images (§3.1). First, the pipeline generates the master flat and master bias images. The master flat (§3.1.2) is produced from the median average of five dome-flat exposures. Implementing the median operation reduces the impact of cosmic-ray hits and other spurious artefacts. On the other hand, the master bias (§3.1.1) is constructed by combining ten bias frames.

The next steps are the *overscan* and the *gain* corrections. The image readout process adds a base signal in all pixels, resulting in a bias value that varies amongst amplifiers. The

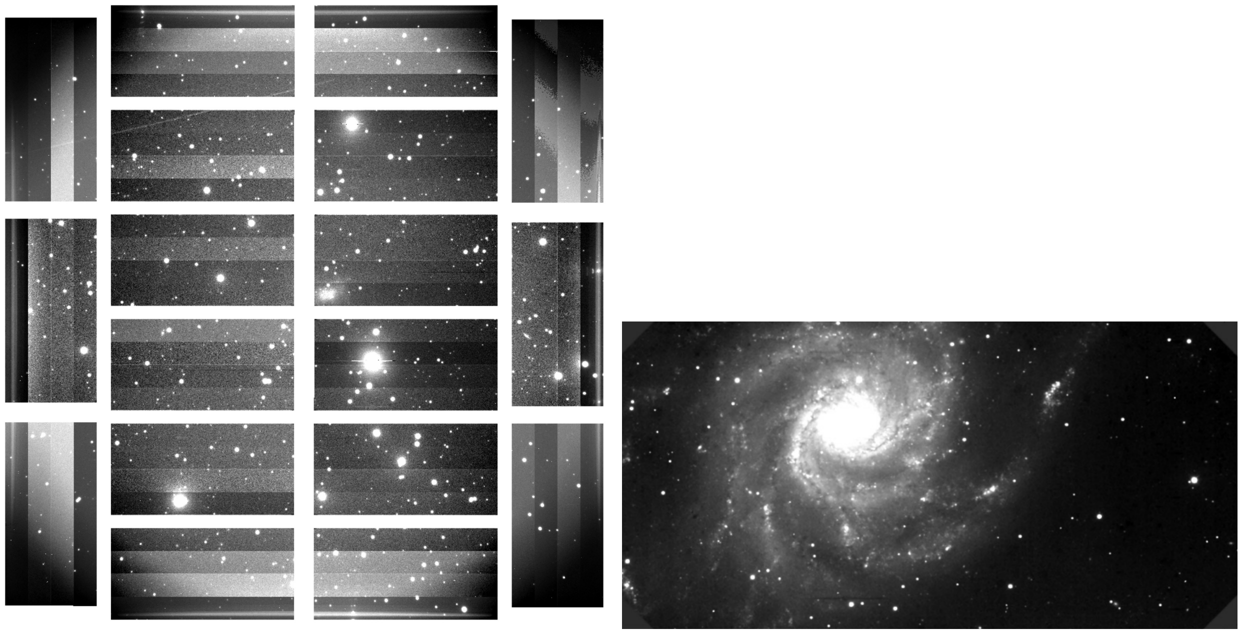

Figure 3.2: *Left:* A raw sky exposure of the full 18-detector PAUCam mosaic. All instrumental signatures are still present (e.g. saturated pixels, cross-talk, and scattered-light). The background has not been flattened with the master flat and the read-out regions from the 72 amplifiers can be easily identified. *Right:* Science corrected image of the extended source M101. This image has already been corrected from all the instrumental effecs.

overscan section is a CCD region that only includes bias, readout noise, and dark current, thus it can be used to estimate the overscan bias in each amplifier. The overscan value is computed row by row as the median of the pixel values in the overscan region and subtracted from the raw image. Then, the gain is estimated from the photon transfer curve analysis (Astier et al., 2019) and converts the science image from the readout digital units to electrons.

Cross-talk signals are trails of brighter pixels generated by current inductions amongst the four PAUS amplifiers (Freyhammer et al., 2001). This commonly happens when one of the amplifiers is reading a very bright (many times saturated) pixel that induces a charge on the other three amplifiers due to the closeness of the cables during the readout process. The configuration of the readout system causes that the three cross-talk signals appear mirrored with respect to the borders of the corresponding readout region. This enables predicting the position of the cross-talk signal from the location of a saturated pixel. Furthermore, the brightness of such pixel determines the amount of induced cross-talk signal. The induction ratio between amplifier i and j is

$$r_{ij}(k) = \mathrm{median}[f_j - b_i], \tag{3.12}$$

where k corresponds to an exposure image, f_j is the average signal at the mirrored position of saturated pixels in the j-th read-out region and b_i is the median background noise in the target read-out region. The ratios in each of the individual exposures are combined into a single ratio

$$r_{ij} = \frac{\sum_k r_{ij}(k) N_{\text{sat}}(k)}{\sum_k N_{\text{sat}}(k)}, \tag{3.13}$$

where $N_{\text{sat}}(k)$ is the number of saturated pixels in image k and i,j are two independent amplifiers. The `Nightly` pipeline uses the ratios in Eq. 3.13 to correct the cross-talk signals.

Scattered light was mostly corrected during the camera intervention in 2015. However, half of the images in the COSMOS field were taken before the intervention (§ 3.4.4). Furthermore, although drastically diminished, scattered-light is still present in PAUCam images after the intervention (§ 4.2). Therefore, scattered light requires additional correction methods to minimise its effect in scientific measurements.

PAUS has studied and implemented several methods to correct for scattered light. *Sky-flats* (§ 4.2.2) model the background averaging the background pixel values over a set of images, which includes effects from the dome-flat and scattered-light residuals. The application of sky-flat corrections is simple, but it requires having multiple images taken under similar conditions, which is not always possible. In this thesis, we propose a deep learning alternative to correct for scattered-light effects (§ 4 and Cabayol-Garcia et al. 2020).

The `Nightly` pipeline also performs World Coordinate System (WCS) and the astrometry calibration. The PAUCam's raw mosaic images include a base WCS (Calabretta & Greisen, 2002) specified in the header. However, the initial WCS solution estimated only from the mechanical layouts of the PAUCam is insufficiently accurate to calibrate the single-exposure images. PAUS uses `SCAMP` (Bertin, 2006) to compute a more accurate WCS solution from the measurements of stars in the focal covering the same sky-area (Gaia Collaboration, 2018).

The astrometry calibration is also implemented with `SCAMP`. Taking as input the pre-calibrated images provided in the WCS calibration, `SCAMP` is applied to single epoch mosaic images to provide an updated WCS calibration that includes the distortion and offset corrections. Such calibration is more precise than the initial WCS calibration and accurate enough to apply forced photometry (Serrano et al. in prep.).

Finally, `PSFex` models the PSF model across the focal plane. The PSF model can provide the PSF-FWHM at any position of the image, but the `MEMBA` forced-aperture measurements only use the mean PSF per image (§ 3.4.2). The detection of cosmic rays is implemented with `L.A.Cosmic` (van Dokkum, 2001), a Laplacian filtering algorithm that benefits from the cosmic rays not being blurred by the PSF.

3.4.2 Measuring photometry: the `MEMBA` pipeline

`MEMBA` is the PAUS pipeline to measure the photometry from the single epoch reduced images (§ 3.4.1). Since PAUS narrow-band images have very low signal-to-noise, `MEMBA` implements forced photometry (§3.2) using the `PhotUtils` library (Bradley et al., 2020). This requires a catalogue complete down to the PAUS magnitude limit to prevent target selection effects.

In the COSMOS field, PAUS uses a merged catalogue from Laigle et al. (2016) and the Zurich Structure & Morphology Catalog[3]. For the CFHTLS fields (W1, W3, and W4) the reference catalogue is Heymans et al. (2012), which combines photometric redshifts with PSF-matched photometry (Hildebrandt et al., 2012a), shear measurements with `Lensfit` (Miller et al., 2013), and weak lensing estimates with `THELI` (Erben et al., 2013). Finally, for the G09 KiDS field, the reference catalogue is provided by the KiDS DR4 (Kuijken et al., 2019).

The reference catalogue provides the location, shape, and scale of the object that `MEMBA` uses to target and estimate the photometric aperture. These apertures contain a fixed flux fraction for all sources, e.g. 62.5%. Therefore, the aperture size is calculated for each galaxy image independently since PSF varies from image to image and modifies the size of the target objects.

Assuming a Sérsic profile $I(\vec{r})$ with Sérsic index n and half-light radius r_0

$$I(\vec{r}) = I_0 \exp\left[-\left(\vec{r}/r_0\right)^{(1/n)}\right], \tag{3.14}$$

and a radial Moffat PSF (Ψ)

$$\Psi(\vec{r}; \alpha, \beta) = 2\left(\frac{\beta - 1}{\alpha^2}\right)\left[1 - \left(\frac{\vec{r}}{\alpha}\right)^2\right], \tag{3.15}$$

where $\vec{r}$ is the radial distance to the center of the profile, α is a scale factor, and β determines the overall shape of the PSF. The total aperture flux up to an aperture ratio A is calculated as

$$F(A) = 2\pi \int_0^A dr\; r I(r) \otimes \Psi(\vec{r}; \alpha, \beta), \tag{3.16}$$

where $\otimes$ is a convolution operation between the galaxy and the PSF profile.

`MEMBA` estimates the background-light contribution with an aperture-photometry annulus (§3.2) defined by an inner and outer radius of 30 and 45 pixels, respectively. As sources nearby the target galaxy can also affect the annulus measurements, `MEMBA` applies a σ-clipping to the annulus pixels. This also corrects for other potential spurious artefacts, such as cosmic rays.

The background-subtracted galaxy flux and its uncertainty are measured and estimated with Eq. 3.4 and Eq. 3.5, respectively. In § 5, we present an alternative deep learning-based method (`BKGnet`) to estimate the probability distribution of the background-subtracted photometry. This method provides higher signal-to-noise and has proven more robust towards optical effects as scattered-light (Cabayol et al., 2021).

[3]Zurich COSMOS catalogue https://irsa.ipac.caltech.edu/data/COSMOS/gator_docs/cosmos_morph_zurich_colDescriptions.html

PAUS takes, on average, between three and five observations of each source in every narrow-band filter. These exposure images are calibrated and combined. The PAUS calibration process fits stellar templates to SDSS broad-band data. These are then used to create synthetic narrow-band observations that are compared to PAUS observations of the same stars. Then, the zero-point (ZP) is the ratio between the PAUS stellar observations and the narrow-band synthetic data. The zero-points from every star in the image are statistically combined to a single zero-point per image and its associated uncertainty (σ_{ZP}).

The single-exposure flux measurement (f) is calibrated as

$$f_{\rm calib} = ZP \cdot f, \tag{3.17}$$

with an associated uncertainty

$$\sigma_{\rm calib} = \sqrt{\sigma_f^2 \sigma_{ZP}^2 + \sigma_f^2 ZP^2 + \sigma_{ZP}^2 f^2}. \tag{3.18}$$

The calibrated flux measurements are then co-added with an inverse variance weighting

$$f_{\rm coadd} = \frac{\sum_i f_{{\rm calib}_i / \sigma^2_{{\rm calib}_i}}}{\sum_i 1/\sigma^2_{{\rm calib}_i}}, \tag{3.19}$$

where i runs over observations of the same source. Assuming the measurement errors are independent, the associated uncertainty is estimated as

$$\sigma_{\rm coadd} = \frac{1}{\sum_i 1/\sigma^2_{\rm calib_i}}\,. \tag{3.20}$$

During the whole data reduction process, the `Nightly` and the `MEMBA` pipelines identify potential issues in the data reduction. While the `Nightly` flags problems at the pixel level, e.g. vignetted areas, saturated pixels, and cross-talk; `MEMBA` identifies potential problems at the catalogue level. Some examples are scattered light flagging, i.e. when the observation is in a region strongly affected by scattered light, galaxy observations too close to the edge of the image (edge detection), and sources in regions with strong optical distortion. The `MEMBA` co-added photometry is calculated excluding flagged observations.

3.4.3 PAUS photometric redshifts

The galaxy photometry in 40 narrow bands for the full sample (without selection effects) makes PAUS photometric sample unique data to obtain very precise photometric redshifts (§2.2.1 and §2.2.4). These have been estimated with template-based and machine-learning (§3.3.3) customed algorithms, as public implementations were not providing precise enough photo-zs.

Defining

$$\Delta z \coloneqq (z_{\rm p} - z_{\rm t}) \;/\; (1 + z_{\rm t})\,, \tag{3.21}$$

where $z_{\rm p}$ and $z_{\rm t}$ correspond to the photometric redshift and the true redshift, respectively, the photo-z scatter can be characterised by

$$\sigma_{68} \coloneqq \frac{1}{2}\left[Q_{84}(\Delta z) - Q_{16}(\Delta z)\right] , \tag{3.22}$$

with quantiles set to 84.1 and 15.9 percentage values. The σ_{68} definition is equivalent to the standard deviation (1-σ error) for a normal distribution, but it is less affected by outliers. In PAUS, outliers have a strict definition

$$|z_{\rm p} - z_{\rm t}| \,/\, (1 + z_{\rm t}) > 0.02 \,, \tag{3.23}$$

which is $\sim \times 10$ more stringent that typical broad-band photo-z outlier definitions.

On simulations, Martí et al. (2014) forecast a PAUS photo-z error of $\sigma_{68} \approx 0.0035$ for a selected 50% of the data to $i_{\rm AB} < 22.5$. This precision was obtained with `BPZ` (Benítez, 2011), however, this algorithm was far from reaching the required precision on PAUCam data.

So far PAUS has implemented three photo-z methods using different approaches that reach this photo-z precision, which are presented in the left panel of Fig. 3.3.

`BCNz2`: This algorithm is introduced in Eriksen et al. (2019). It is a template-based photometric redshift code that fits PAUS data to redshift dependent models constructed as a linear combination of SED templates (Eq. 3.9). The models include the emission lines as fixed amplitude ratios and added as two additional SEDs (see §4.5 in Eriksen et al. 2019).

`BCNz2` uses a combination of the PAUS 40 narrow bands and external broad bands. In the COSMOS field, these broad bands are the u^* band from the Canada-France Hawaii Telescope (CHFT/MegaCam) and B, V, r, i^+, z^{++} bands from Subaru, all available in Laigle et al. (2016). Because of potential problems with the photometry, the code includes a per galaxy scaling between the broad and narrow band fluxes. This calibration is introduced as an additional parameter to minimize the χ^2-fit. Moreover, a zero-point re-calibration per band is implemented by comparing the observed photometry with the model at the spectroscopic redshift. This zero-point determination runs iteratively 20 times. This algorithm achieves $\sigma_{68}/(1+z) \approx 0.0037$ for 50% of the galaxies with $i_{\rm AB} < 22.5$

`Delight`: Public code[4] (Leistedt & Hogg, 2017), adapted and tested on PAUS data in Soo et al. (2021), is a hybrid template-based and machine learning photometric redshift algorithm. `Delight` constructs a set of flux-redshift models from the training data guided by a template SED library and implements a Gaussian process to find the distribution over the possible functions that is consistent with the observed data. Similar to `BCNz2`, in the COSMOS field `Delight` also uses the 40 PAUS narrow bands and the $uBVriz$ broad bands. `Delight` achieves the photometric redshift precision of $\sigma_{68} < 0.0081$ for the full COSMOS sample and reaches the forecast photo-z precision of $\sigma_{68} < 0.0035$ for 60% of the objects with $i_{\rm AB} < 22.5$.

[4] `https://github.com/ixkael/Delight`

Deepz: Deep learning algorithm which first uses auto-encoder (Liou et al., 2014) to denoise the input photometry and extract relevant features. These features, together with the input photometry, are the input of a neural network that predicts the photometric redshift probability distribution (Eriksen et al., 2020). One important highlight of this model is that it uses galaxy simulations for training, which enables increasing the training sample, and later applies transfer learning to adapt the model to the data (Zhuang et al., 2019; Tan et al., 2018). In the transfer learning implementation, `Deepz` also takes advantage of single observations per galaxy and narrow band, which increases the training sample and reduces the impact of photometric outliers. `Deepz` reduces the photo-z scatter of `BCNz` and `Delight` by 50% for all the galaxy sample without quality cuts. Similarly, it also improves the photo-z precision for the best 50% of the sample and reduces the strict outlier fraction (Eq. 3.23) from 17% to 10%.

PAUS+26 external bands: This is a follow up work to `BCNz2`. Instead of selecting only the best fit linear combination of the SEDs, it integrates over the space of all SED amplitudes. It has been implemented in the COSMOS field using the 40 PAUS narrow bands and 26 external broad, intermediate and narrow bands (Alarcon et al., 2021). For these 66 bands, it achieves photometric redshifts with $\sigma_{68} \sim 0.003$ and $\sigma_{68} \sim 0.009$) for galaxies at magnitude $i_{\mathrm{AB}} \sim 18$ and $i_{\mathrm{AB}} \sim 23$, respectively.

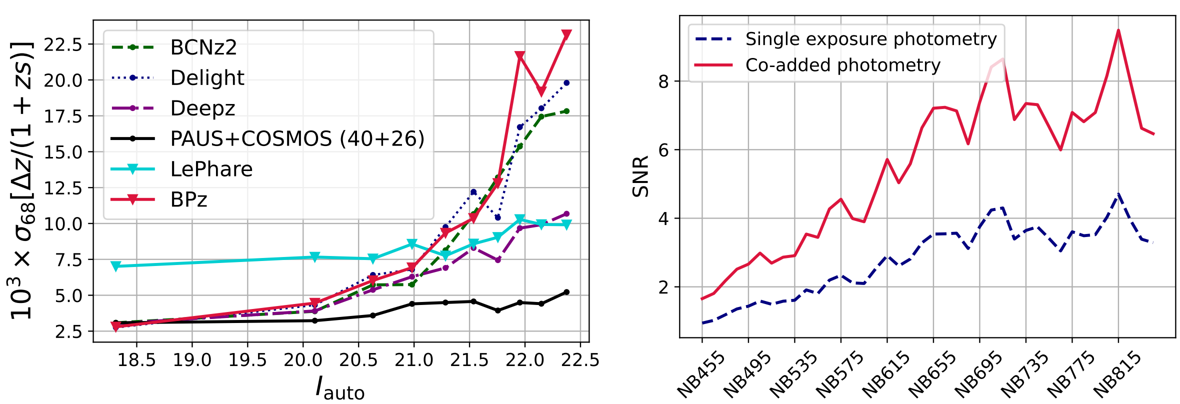

Figure 3.3: *Left:* PAUS photometric redshift precision with several photo-z codes. *Right:* Signal-to-Noise (SNR) of PAUS observations in the COSMOS field for the single aperture photometry measurements (blue) and the co-added photometry (red).

3.4.4 PAUS data in COSMOS

In this thesis, we have mostly used PAUS science images in the COSMOS field (§3.4.1). These data comprise 9749 images, 243 images in each narrow band. These data were taken in the semesters 2015B, 2016A, 2016B, and 2017B, with low efficiency due to bad weather. The camera was shipped back to Barcelona in 2016 (§2.2.3), where it was intervened to mitigate the effect of scattered light, which modified the noise patterns in the different CCDs. In the COSMOS field, half of the images (4928) were taken before the camera intervention and the other half (4821), after. The default exposure times are 70, 80, 90, 110 and 130 seconds from the bluest to the reddest filter tray.

The photometry catalogue contains ∼12,5 million measured galaxy fluxes, ∼ 5 observations per galaxy and narrow-band filter. These correspond to 64 476 observed galaxies to $i_{\mathrm{AB}}<23$ in the 40 narrow-band filters. The right panel on Fig. 3.3 shows the signal-to-noise obtained with `MEMBA` in the COSMOS field for unflagged objects to $i_{\mathrm{AB}}<22.5$. The blue line corresponds to single-exposure measurements, while in red line represents co-added observations.

Part II

Galaxy photometry and photo-zs with deep learning

Chapter 4

Background light prediction

4.1 Motivation

The positions, fluxes, and other properties of galaxies and stars can be determined by analysing images of the sky (§ 3). Modern imaging surveys can cover large areas of sky efficiently, resulting in measurements for large numbers of galaxies to faint magnitudes (e.g. DES DR1, Abbott et al., 2018b). Accurate photometry is crucial to ensure the analyses of future weak lensing surveys like LSST and *Euclid* are not dominated by systematic errors. Also, imaging surveys require accurate flux measurements to select samples of galaxies and infer their physical properties.

The determination of the sky background is a key step towards reliable photometry measurements (see § 3.2). The main source of background light is the intrinsic sky brightness, which varies due to a range of effects such as illumination by the moon, airglow, and light pollution. PAUS images are affected by scattered light, which is the result of light deflections from the instrument's optical path appearing at a different region of the detector (§2.2.2 and Romanishin 2014). Scattered light hinders accurate background-light predictions as it introduces spatial background-noise variations in the images.

Different approaches have been used to estimate the sky background, e.g. `DAOPHOT` (Stetson, 1987) and `SExtractor` (Bertin & Arnouts, 1996). `DAOPHOT` measures the background as the mode of the uniformly scattered pixels at a certain 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. Examples include Teeninga et al. (2015), which estimates the background at a location without nearby sources and Popowicz & Smolka (2015), based on the removal of small objects and the interpolation of missing pixels.

Over the last few years, deep learning algorithms have resulted in revolutionary advances in machine learning and computer vision (§ 1.1 and Voulodimos et al. 2018). Theoretical breakthroughs in training deep ANNs (Werbos, 1982) and CNN (LeCun et al., 1989; Lecun et al., 1998; Zeiler & Fergus, 2013), together with powerful and efficient parallel computing provided by GPUs (Krizhevsky et al., 2012), have led to groundbreaking improvements across a variety of applications. The number of deep learning projects in cosmology is quickly increasing (§ 1.3). This includes e.g. 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., 2018a).

Extracting the source photometry requires a significant amount of data engineering and parameter tweaking (§ 3). This can be particularly challenging for noisy sources. Deep learning has already been successfully implemented in different steps in source photometry extraction. Examples include source detection (Vafaei Sadr et al., 2019), cosmic-ray detection (Zhang & Bloom, 2019), and PSF modelling (Herbel et al., 2018b). Moreover, deep learning has also been used to directly estimate photometric redshifts from images (D'Isanto & Polsterer, 2018; Pasquet et al., 2019). Many of these algorithms implicitly require the network to understand galaxy photometry and thus, estimate the background light. Therefore, understanding how these networks learn the image processing steps can optimise the performance of more complicated deep learning algorithms such as galaxy classification and photo-z estimation from astronomical images.

Our goal is to develop and test a deep learning background-subtraction method using data from PAUS (§ 2). 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 $i_{\mathrm{AB}} < 22.5$ (left panel in Fig. 3.3) contains outliers when compared to the spectroscopic redshifts. Some of these outliers arise from noisy photometry, but others are due to the strongly varying background-noise pattern produced by scattered light. The excess of scattered light can alter the pixel values, potentially biasing the photometry of the galaxy.

In this chapter, we present `BKGnet`, a convolutional deep neural network capable of learning the underlying behaviour of scattered light and other distorting effects present in the PAUCam images (§ 3.4.1). `BKGnet` predicts the background light and its uncertainty at the location of the target source. Although `BKGnet` has been developed to improve PAUS photometry, it can potentially be implemented in other future imaging surveys such as LSST and *Euclid*. The code is available at `https://github.com/PAU-survey/bkgnet`.

The structure of this chapter is as follows. Section 4.2 presents the PAUCam images used to develop and evaluate `BKGnet` and characterises the scattered-light affecting the images. Section 4.3 introduces `BKGnet` and defines the training and testing process. In § 4.4 and § 4.5, we evaluate the background-light predictions with `BKGnet` on simulated and real PAUCam images, respectively. Finally, § 4.6 validates the network on target galaxies.

4.2 Modelling scattered-light

PAUCam images contain scattered light, which affects the edge regions of several CCDs. Scattered light increases the amount of background noise in the affected regions, distorting the expected statistics of the pixel values used to estimate the photometry. Moreover, the high background light also lowers the signal-to-noise of the photometry.

In 2016 the PAU camera was modified to mitigate the effect of scattered light (§ 2.2.3). Although this reduced the amount of scattered light, residuals remain. In the latest COSMOS data reduction, around 8% of exposures taken before the camera intervention are flagged as affected by scattered light, thus excluded from the sample (§ 3.4.2). After the intervention, this number was reduced to 5% of the exposures, in such a way that, on average, 7% of data in the COSMOS field are lost due to scattered light. In this section, we present the PAUCam scattered-light model we use throughout the paper.

4.2.1 The PAUS observations

In this chapter, we use PAUS data in the COSMOS field. The details on these data are in § 3.4.4. For details on the survey, the survey camera and the data reduction pipeline, see § 2. We use images that have already been corrected for instrumental effects with the `Nightly` pipeline (Serrano et al,. in prep., § 3.4.1). The background-light measurements from `MEMBA` require a (fairly) flat background for an accurate estimate (in app. 4.A we study the effect of a variable annulus). This assumption breaks down when either the annulus or source extraction regions are affected by scattered light. In addition, other artefacts, as e.g. undetected sources, cosmic rays, and cross-talk can potentially bias the background-noise estimation. Throughout this study we compare the background-light measurements from `BKGnet` with the measurements from the default `MEMBA` configuration.

4.2.2 Scattered-light templates

Figure 4.1 shows four PAUCam images in the narrow-band filter "NB685" before the camera intervention (first and second images on the left) and after the camera intervention (third and fourth images). These images present scattered light as a spatially varying amount of light near the edges of the CCD. The scattered-light pattern changes from before the camera intervention (two images on the left) to after (two images on the right). Furthermore, scattered light is also narrow-band dependent, meaning that each filter has its background-noise distinctive pattern.

One way to quantify and model scattered light is to estimate the amount of additional background that each pixel contains with respect to the median image background noise. Averaging the background intensity ratio from images taken with the same narrow-band filter, we obtain a model for the scattered-light intensity across the CCD. To do so, the first step is selecting the narrow-band images we are going to use. Then, for each of these images (I_{j}), we compute the median background level (μ_{BKG}) excluding the pixels from the edge of the CCD, as these are potentially affected by scattered light. After, we need to divide each image by its median background noise μ_{BKG} to obtain a background ratio per pixel

$$r_{\mathrm{j}} = I_{\mathrm{j}}/\mu_{\mathrm{BKG}}\,. \tag{4.1}$$

If the background were flat and followed Poisson statistics, the ratio in all pixels should fluctuate around unity. In contrast, pixels affected by scattered light will have a background

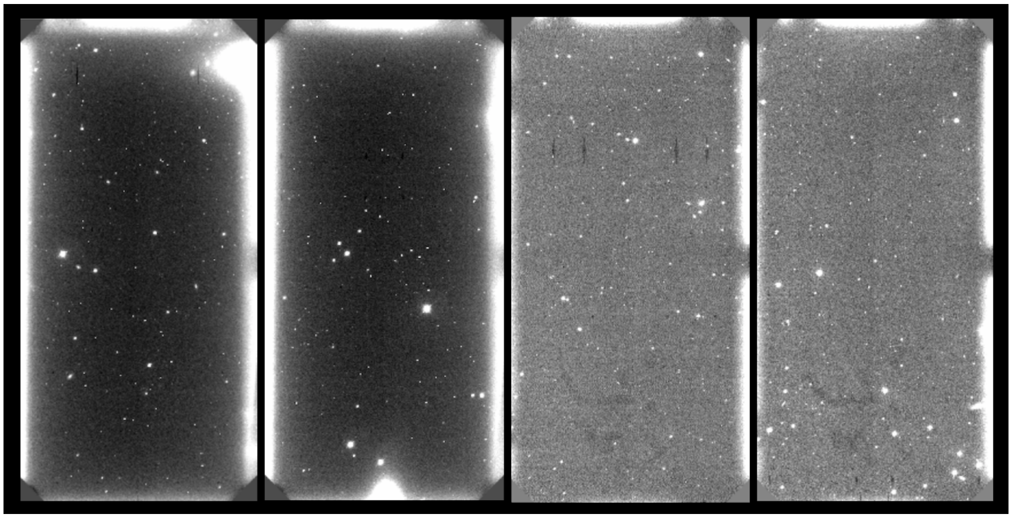

Figure 4.1: Images taken with the PAUCam, corresponding to the NB685 filter. *Left:* The first two images correspond to PAUCam images before the camera intervention. Notice that both exhibit the same scattered-light pattern. *Right:* The two images on the right correspond to PAUCam images after the intervention. 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.

ratio above unity. Therefore, we can understand this ratio as approximately the percentage of scattered light that affects the pixel.

To obtain a single pixel map per CCD, we combine the individual pixel ratios from several images using the median to get a *scattered-light template* (R_{SL}) from all the selected images

$$R_{\mathrm{SL}}(x, y) = \mathrm{median}_j[r_{\mathrm{j}}]\,. \tag{4.2}$$

Before combining the images, the sources need to be masked to ensure these are not affecting the statistics. The scattered-light template is many times referred as *sky-flat*.

The top panel in Figure 4.2 shows some examples of R_{SL} for the narrow-band filter "NB685". The plot draws the pixel values r of the background-ratio image, fixing the central row of the CCD and moving along the x-axis. Before the camera intervention (black dashed line), the plot shows a strong tendency of increasing the background at the edges of the CCD, which already starts 500 pixels away from the edge. In contrast, after the intervention (orange solid line), scattered light significantly reduces and appears < 100 pixels from the edge.

In the bottom panel in Fig. 4.2, we have combined the pixel ratios from all images available per narrow band (splitting in before/after the intervention) in a single scattered-light template (R_{SL}). The plot shows the mean value of each template per narrow band, which gives information about the amount of scattered light captured by the CCD. The effect of the intervention is evident since all the narrow-band filters reduce the mean of their scattered-light template.

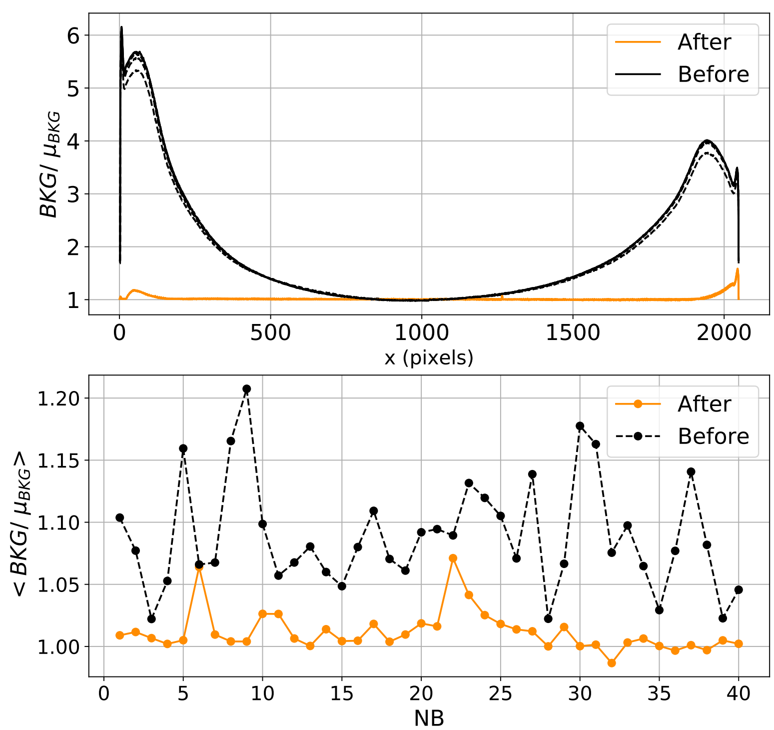

Figure 4.2: *Top*: Normalised background light content in each pixel as a function of the pixel position in the image for different images before (black dashed line) and after (orange solid line) the camera intervention. Each pixel value is divided by the mean background in the image. Regions without scattered-light should fluctuate around unity. Regions affected by scattered-light should be above unity. *Bottom*: Mean value of the normalised background curves considering all the images taken in that band, for the 40 narrow photometric bands.

4.2.3 Scattered-light templates as scattered-light correcting method

The scattered-light templates R_{SL} can be used to correct scattered light on PAUCam images if the modelling is sufficiently accurate. Assuming that all images taken with a narrow-band filter follow the same scattered-light pattern scaled by the image sky-background level, a way of correcting scattered-light is

$$\tilde{I}(x,y) = I(x,y) - (R_{\mathrm{SL}}(x,y) - 1)\mu_{\mathrm{BKG}}\,, \tag{4.3}$$

where we subtract from a science image $I(x,y)$ the scattered-light template scaled by the mean background of such image. Note that regions without scattered light should not be affected by the correction since we subtract $R_{\mathrm{SL}}(x,y) - 1$.

This correction has been implemented in Fig. 4.3. The left panel shows a science image taken with "NB685". In the middle panel, we see the same image once corrected with the scattered-light template. The scattered-light template used for the correction is in the rightmost panel. Visually, the scattered-light pattern in the original image (left) disappears after applying the correction (middle). However, external conditions, e.g. humidity, clouds, temperature, and moon) introduce fluctuations in the scattered-light pattern that are not properly modelled in the scattered-light template, leading to light residuals after the correction. A scattered-light template per night and narrow-band filter would be required to have a more accurate correction. Unfortunately, the number of images observed nightly with one PAUCam filter is insufficient to obtain accurate scattered-light modelling. Moreover, bright stars also contribute to scattered light and these cannot be corrected with the scattered-light templates.

Similarly to Fig 4.2, Fig. 4.4 shows the pixel values across the centre row of an image taken with "NB685" before (orange) and after (black) correcting the image with Eq. 4.3. The scattered-light template used for the correction combines all the available images in "NB685", without selecting images based on the observing conditions. The image without correction (solid orange line) displays two large scattered-light peaks at both edges of the image. These peaks are mostly corrected by the scattered-light template, however, scattered-light residuals are still present on both sides of the CCD after the correction.

4.3 `BKGnet`: A Deep Learning based method to predict the background

In this section we start by describing the `BKGnet` architecture (§ 4.3.1), the training and test samples (§ 4.3.2) and the training procedure (§ 4.3.3).

4.3.1 Neural network architecture

`BKGnet`[1] (Fig. 4.5) combines a CNN (§ 1.2.1) and a fully-connected neural network. The former contains five blocks of convolutional, pooling, and batch normalization layers, which

[1] https://gitlab.pic.es/pau/bkgnet

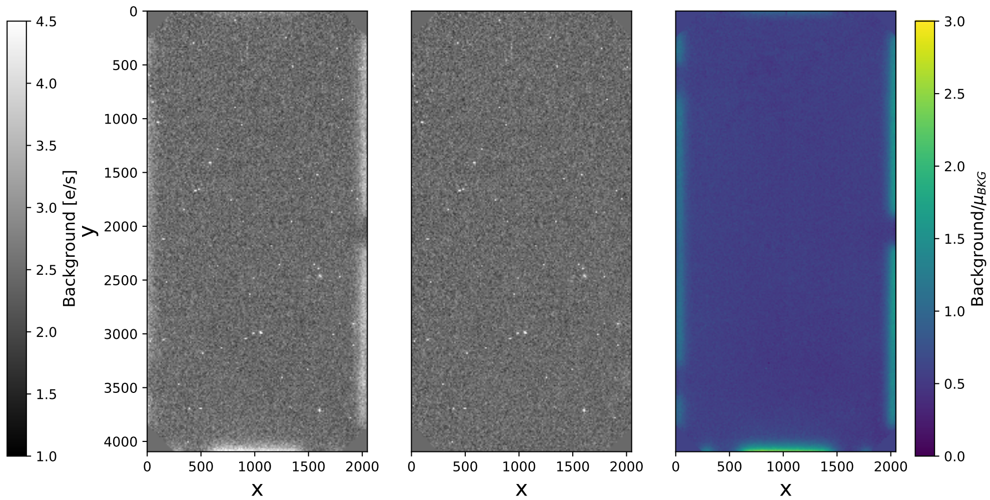

Figure 4.3: *Left:* Image taken in the NB685 filter showing a scattered-light pattern on the edges. *Middle:* Previous image corrected with the scattered-light template. *Right:* The scattered-light template generated with equation 4.2 including all images taken the same observation night as the original image.

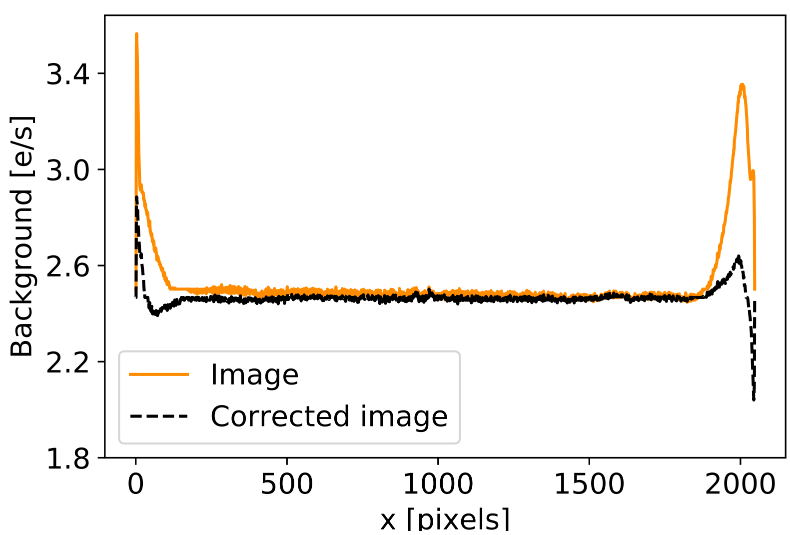

Figure 4.4: Background pixel values across the image. The original image (orange solid line) displays high peaks on the edges caused by scattered-light. After correcting with the scattered-light templates (dashed black line) the peaks are reduced, but some residuals remain. Pixel value unit are e/s.

are represented as red, yellow and blue layers in Fig. 4.5, respectively. The input images are 120×120 pixel cutouts centred at the target galaxy, which compromises having enough pixels whilst keeping the computing requirements (memory, training time) within reasonable limits.

The CNN extracts features from the input image, which are the input of the linear neural network. However, the network also requires other information not present in the images, e.g. the position of the galaxy in the CCD image, the narrow band filter, and a before/after intervention flag specifying when the galaxy was observed. These parameters are additionally provided to the linear network together with the target galaxy magnitude from a reference detection catalogue. The galaxy magnitude potentially contains information about the number of pixels affected by the galaxy.

The narrow-band filter and the intervention flag are discrete variables with forty possible values for the band parameter (1-40) and two for the intervention flag (0/1). The combination of these two effectively corresponds to 80 different scattered-light patterns (§4.2). An embedding layer encodes each of the eighty possible scattered-light patterns into ten trainable parameters that learn to characterise the pattern.

4.3.2 Data: training and test samples

`BKGnet` is trained with cutouts centred at empty positions, i.e. regions where there are no target sources, where we can estimate the ground-truth backgrounds noise (training labels) beforehand. In order to identify and select empty regions, we cross-correlate the sky coordinates of the cutout location with the sky coordinates of the sources in the COSMOS catalogue (Laigle et al., 2016). The training sample labels are computed as the mean background inside an eight-pixel radius circular aperture centred at the cutout. The label measurements have an associated uncertainty (σ_{label}) that directly depends on the size of the region to define the label. Assuming that the background is purely Poissonian, then

$$\sigma^2_{\text{label}} = \frac{N_{\text{a}}\bar{b}}{t_{\text{exp}}}\,, \tag{4.4}$$

where t_{exp} is the exposure time, $\bar{b}$ is the mean of the pixels inside the aperture, i.e. the background label, and N_{a} is the number of pixels within the circular aperture. The aperture radius is chosen to eight pixels to balance the uncertainty on the ground truth measurement and a precise label at the exact galaxy location.

`BKGnet` will evaluate the background of cutouts with galaxies at the centre. To resemble the training examples and the test cutouts, we simulate galaxies at the centre of the empty training cutouts. These simulated galaxies follow a Sérsic profile characterised by the Sérsic index (n) and the half-light radius (r_{50}), i.e. the radius that contains 50% of the light intensity (I_{50}). These parameters are drawn from the distribution of PAUS observed galaxies. As the simulated galaxies can potentially differ from the PAUS target galaxies, we also mask the central 16×16 pixels in both the training and test samples. Despite masking the central cutout region, including the simulated galaxy in the training examples is still beneficial since large and bright galaxies can potentially extend outside the masked area. This is evidenced

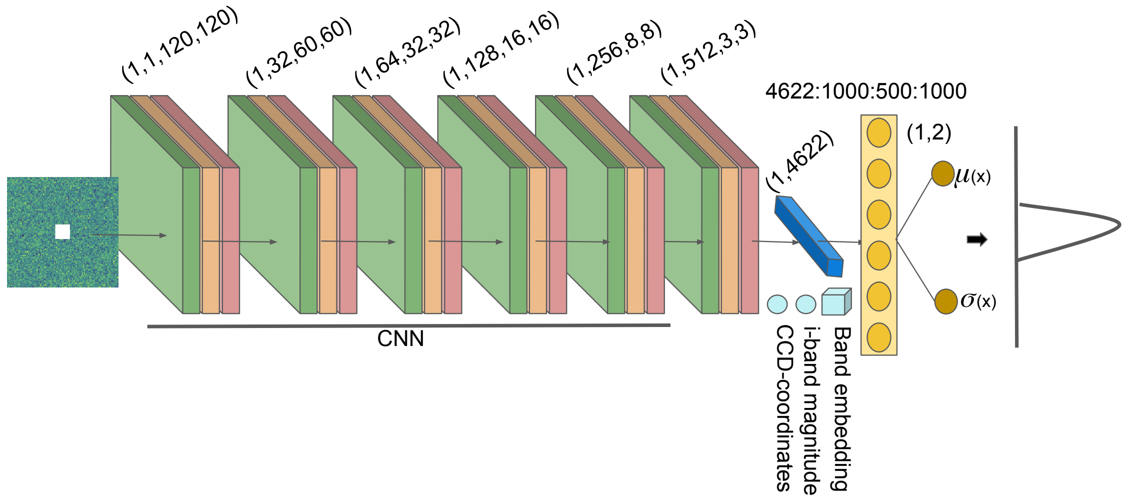

Figure 4.5: `BKGnet` scheme: The first set of layers corresponds to a Convolutional Neural Network to which one inputs the images. The CNN output, together with extra information are input to a linear neural network. The numbers on each of the convolutional layers represent the layer's dimension. The first number corresponds to the number of channels. The second and third numbers are the dimension of the cutout in that layer.

by testing very bright sources, where `BKGnet` fails without simulated centred galaxies in the training examples. As the label is estimated in an 8 pixels radius aperture, 16×16 pixels is the minimum area for which the network does not see the pixels used to estimate the ground-truth background.

The network is trained with the normalised cutouts, which speeds up the training convergence and helps in the network's regularisation. There are several methods to normalise the samples. We standardise image by image, subtracting the mean and dividing by the standard deviation. This is a widely used normalisation and the one performing better on our dataset.

We use all the PAUCam images in the COSMOS field to train and validate the network. There are 4928 PAUCam images before the intervention and 4821, after (§ 3.4.4). We sample 40 cutouts per science image, which provides 400,000 training cutouts split into 90% for training and the remaining 10% for validation.

4.3.3 Training process and loss function

To associate an uncertainty to each prediction, we have implemented a MDN (Bishop, 1994) with a single Gaussian component (§1.2.2). This assumes that the distribution $p\left(\vec{y}|f^{\vec{w}(\vec{x})}\right)$

is Gaussian, where $\vec{y}$ are the background labels, $\vec{x}$ are the input images and $f^{\vec{w}(\vec{x})}$ are the network background predictions. The loss function is defined as

$$\mathcal{L} = -\log\left(p(f^{\vec{w}(\vec{x})})\right) = \frac{\left(f^{\vec{w}(\vec{x})} - y\right)^2}{\sigma^2_{\text{bkgnet}}} + 2\log\left(\sigma_{\text{bkgnet}}\right), \tag{4.5}$$

where the second term on the right hand side prevents the network from predicting large uncertainties that minimise the mean-squared error term. Equation 1.5 shows the general loss function expression for a MDN with N Gaussian components .

The background-noise and its uncertainty (σ_{bkgnet}) are the mean and the standard deviation predicted by the network. The background labels y have an associated uncertainty (σ_{label}), therefore the network provides the uncertainty associated to the quantity $f^{\mathbf{w}}(\mathbf{x}) - y$, and the error on the prediction should be corrected by

$$\sigma_{\text{pred}} = \sqrt{\sigma^2_{\text{bkgnet}} - \sigma^2_{\text{label}}}\,, \tag{4.6}$$

where σ^2_{label} is defined in Eq. (4.4).

`BKGnet` is trained in 60 epochs with a batch size of 100 cutouts using the ADAM optimiser (Kingma & Ba, 2014) and a learning rate of 10^{-5}. The training takes about 2 hours using an NVIDIA TITAN V GPU.

4.4 Testing `BKGnet` on simulations

In this section, we test the performance of `BKGnet` on simulated data. Section 4.4.1 introduces the simulated images used for testing, while § 4.4.2 presents the `BKGnet` predictions on such images. Throughout the rest of the paper we compare the `BKGnet` predictions to those obtained with aperture photometry, which are estimated using the annulus method (§ 3.2) with the default `MEMBA` implementation (§ 3.4.2). The aperture background measurements are estimated from images with and without scattered-light corrections with templates (§ 4.2.3).

4.4.1 Simulated PAUCam background images

We use scattered-light templates (R_{SL}, § 4.2.2) to generate simulated PAUS background-noise images accurately mimicking the scattered-light pattern in each of the PAUS photometric filters. The simulated image $I_{\text{sim}}(x, y)$ can be expressed as

$$\text{I}_{\text{sim}}(x, y) = A \cdot \frac{t_{\text{exp}} \cdot R_{\text{SL}}(x, y) + P(t_{\text{exp}} \cdot R_{\text{SL}}(x, y))}{t_{\text{exp}}}\,, \tag{4.7}$$

where we multiply the scattered-light template by the exposure time (t_{exp}) to convert the simulated image to electrons. Then, we introduce Poisson noise $P(\cdot)$ and scale the template with a factor A to simulate a wide range of background levels. Finally, the simulated images is converted back to e/s.

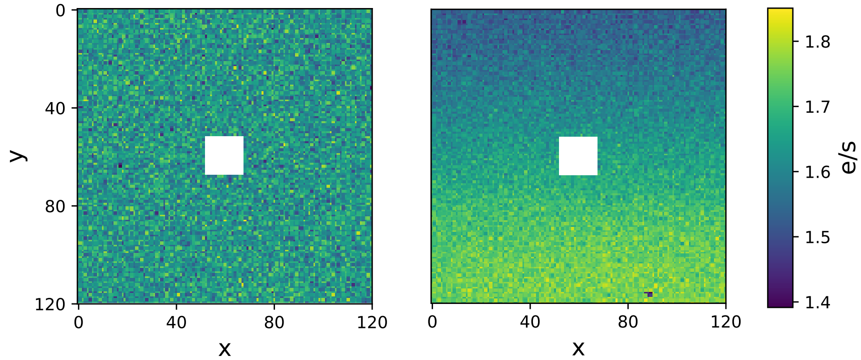

Figure 4.6: Image cutouts simulated with Eq. 4.7. *Left*: a cutout with a flat Poissonian background. *Right*: A cutout with a background with a gradient caused by scattered-light.

Figure 4.6 presents two examples of simulated cutouts generated with Eq. 4.7. On the left, the image presents a flat Poissonian background. In contrast, the cutout on the right shows a clear scattered-light background gradient. Both cutouts show a central 8×8 pixel masked region, blocking the light from the galaxy.

4.4.2 `BKGnet` predictions on simulations

We characterise the performance of `BKGnet` background predictions training and testing on simulated cutouts from "NB685" (§ 4.4.1) . Before the intervention, some of the CCDs contain a lot of scattered light, which is not adequate to test the network predictions. On the other hand, after the intervention, some of the CCDs barely contain scattered light and we could not test `BKGnet` predictions in scattered-light affected areas. We have chosen "NB685" as it is significantly affected by scattered light, both before and after the camera intervention, without being completely dominated by this effect.

Figure 4.7 compares `BKGnet` background predictions with the aperture photometry annulus approach (3.4.2). We have tested `BKGnet` with and without the CCD coordinates information to see the impact of this parameter and have a better understanding of the `BKGnet` underlying mechanism. The `BKGnet` performance improves significantly with the coordinate information (solid black line). This indicates that although the presence scattered light is encoded in the image, the CCD position includes additional essential information that `BKGnet` potentially uses to create something similar to the scattered-light templates (§ 4.2).

We use

$$\sigma_{68}[(b_{\text{pred}} - b_0)/b_0] \tag{4.8}$$

to quantify the precision in the background-light predictions, where b_{pred} is the `BKGnet` background prediction and b_0 is the ground-truth background noise. `BKGnet` achieves a $\sigma_{68} = 0.0038$ with information coming only with the cutouts. Including the coordinate information, this improves to $\sigma_{68} = 0.0022$, which corresponds to 70% more precise measurements. The aperture photometry estimates show tails on both sides of the distribution, and yields $\sigma_{68} =$

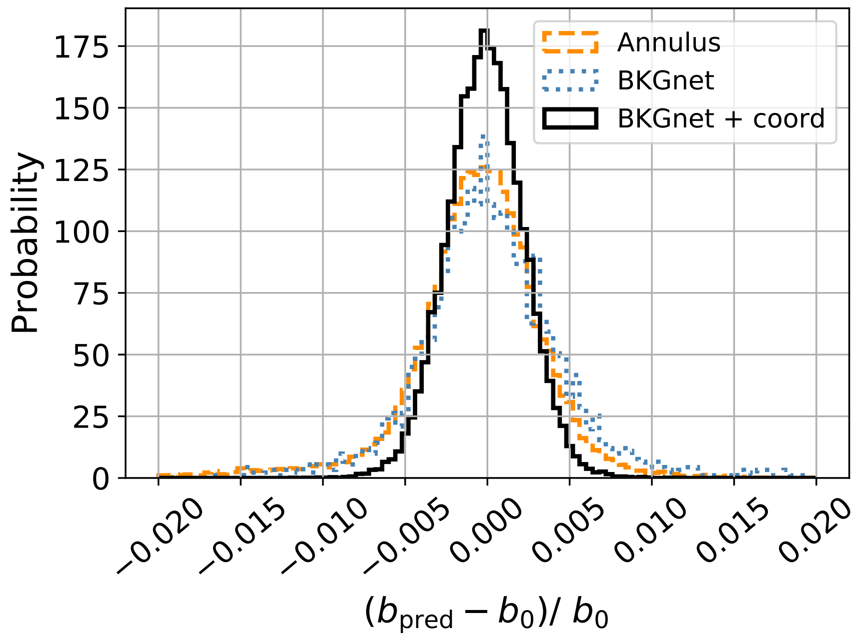

Figure 4.7: Relative error distributions for the `BKGnet` without (blue dotted) and with (solid black) coordinate information compared to the annulus predictions (dashed orange). b_0 is the background label and b_{pred} is the background prediction, with either annulus or `BKGnet`.

0.0033, which means `BKGnet` reduces the scatter by 42%.

For Fig. 4.8, we have predicted the background with `BKGnet` (second panel), aperture photometry (third panel) and a kNN algorithm (fourth panel) across the full CCD, centering the simulated cutout in each of the pixels consecutively. The leftmost panel in Fig. 4.8 shows the true spatial background map. The following panels present the relative error on the background-light prediction. For the annulus method (third panel), the precision is lower at the edges of the CCD, which indicates that scattered light causes the tails in Fig. 4.7. In contrast, `BKGnet` (second panel) shows a flat accuracy across the CCD, proving that it can detect and account for the presence of scattered light in its background-light predictions.

We have also implemented a k-nearest neighbors (kNN) (Cover & Hart, 2006), a support vector regression (SVR) (Drucker et al., 1996), a random forest (RF) (Breiman, 2001), and a neural network (NN) using their `scikit-learn` implementations (Pedregosa et al., 2011). Unlike CNNs, these algorithms are not suitable for images. The input provided is the narrow-band photometric filter, the coordinates of the cutout in the CCD, and the median background-light value of the cutout pixels. With this information, the algorithm could potentially correct the median background noise based on the narrow-band filter and the CCD position. The background predictions with the kNN (Fig. 4.8) are 3% biased in the flat-background regions (not affected by scattered-light), which does not happen neither with the annulus or `BKGnet`. The kNN also presents problems in scattered-light regions, where it provides background measurements ∼6 times less precise than in flat regions (for `BKGnet` this

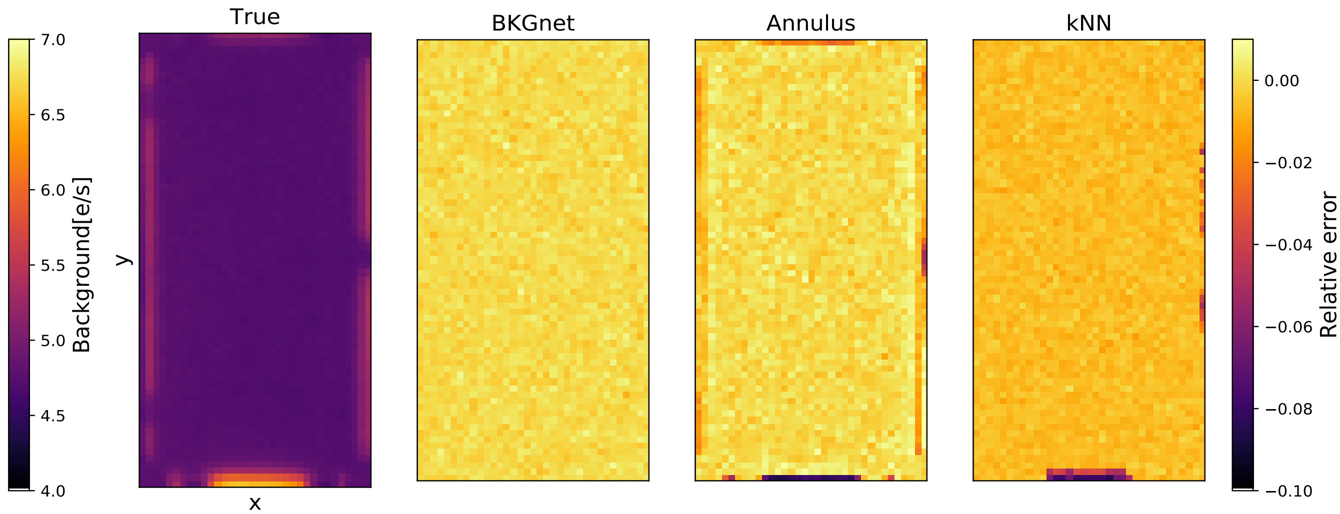

Figure 4.8: *Left*: CCD reconstruction with the ground-truth background noise labels used to train the network. . *Second:* Accuracy on the background prediction with `BKGnet` in the different image positions. We can see there are no spatial patterns. *Third*: Accuracy on the background prediction with the annulus in the different image positions. We can see there are no spatial patterns. *Right*: Accuracy on the background prediction with a kNN in the different image positions.

factor is only 1.2). 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 $\sim$4 times lower precision than in flat background regions. Finally, the RF and the SVR algorithms provide $\sim$6 and $\sim$4 times higher scatter than the `BKGnet` predictions, respectively, rendering these methods too imprecise.

4.5 `BKGnet` on PAUCam images

We have seen that `BKGnet` is able to accurately predict strongly scattered-light backgrounds on simple simulated cutouts including only background noise (§ 4.4.2). However, PAUCam images contain other distorting effects as e.g. cosmic rays, electronic cross-talk, read-out noise, and dark current (§ 3.1) that can potentially affect the background-noise pattern. Furthermore, pixel-to-pixel correlations are potentially introduced during the data reduction process, which are not intriduced in the simulated cutouts (§ 4.4.1). In this section we test `BKGnet` on PAUCam images to examine the impact of these real-life effects.

To evaluate the `BKGnet` performance, we use PAUCam cutouts centred at regions without galaxies, so that we can determine the ground-truth label in the same way as in the earlier simulated cutouts (§4.4.1). We use all the images available in COSMOS, splitting the data into images obtained before and after the camera intervention (§ 2.2.3), to balance the number of training cutouts from before and after the camera intervention. To avoid outliers

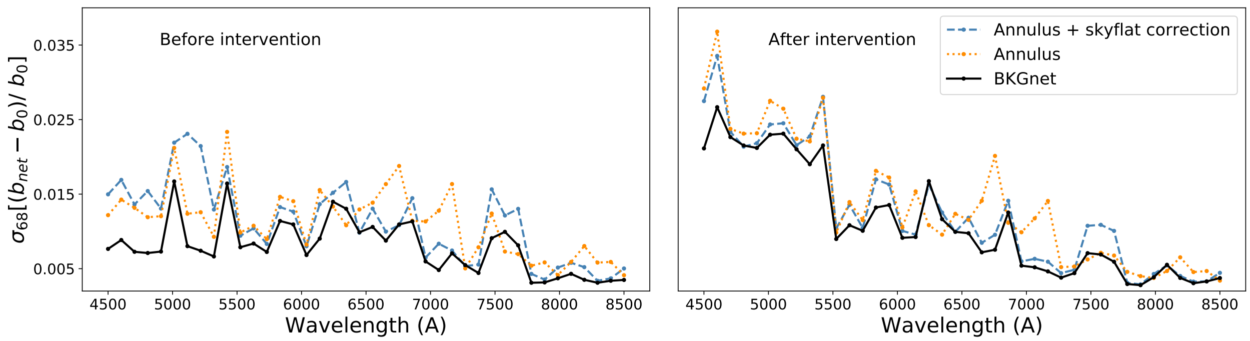

Figure 4.9: Precision in the background predictions with BKGnet and aperture photometry for the 40 PAUS narrow bands. *Left*: Before the intervention. *Right*: After the camera intervention.

in the training set, e.g. a cutout with a bright star covering most of the pixels or a bright object too close to the centre, all cutouts with a pixel containing more than 100 000 counts are excluded from the training sample. We also exclude 80 images (40 taken before and 40 after the camera intervention) used to evaluate the BKGnet performance. The test set is generated by systematically sampling cutouts consecutively in intervals of 60 pixels (instead of randomly selecting CCD positions). This ensures that we test all CCD regions, including regions affected by scattered light.

Figure 4.9 compares the precision in the BKGnet background predictions (black solid line), aperture photometry with an annulus (orange dotted line), and the annulus when the image has been previously corrected with the scattered-light template (blue dotted line, § 4.2.3). The left panel in Fig. 4.9 presents the predictions before the camera intervention. In several narrow bands (e.g. "NB455"), the annulus measurements do not benefit from the correction with the scattered-light template. In contrast, with BKGnet the background prediction precision improves the annulus background estimates(both with and without scattered-light template correction) in all narrow bands (black solid line). On average, BKGnet reduces σ_{68} by 37% compared to the scattered-light template and up to 50% if we only consider the first filter try (i.e. the eight bluer narrow bands). This indicates that the scattered-light correction with sky-flats is unstable before the camera intervention, requiring more flexible modellings.

On the other hand, after the camera intervention (right panel in Fig. 4.9) the annulus measurements benefit from the scattered-light template correction in all the narrow bands. This is expected since scattered light is more stable after the camera intervention (top panel in Fig. 4.2). Nevertheless, BKGnet improves the annulus+skyflat performance achieving an 18% more precise background measurements.

We cannot directly compare the left and right panels in Fig. 4.9 since the intervention changed the level of background noise captured by the CCDs. For instance, in the first filter tray, the background light is between 3 and 5 times higher before than after the intervention.

Table 4.1 presents the precision in the background-light measurement with the annulus

	BEFORE		AFTER	
	filtered	sources	filtered	sources
Annulus	0.011	0.011	0.014	0.014
+ sky-flat	0.011	0.011	0.011	0.013
BKGnet	0.008	0.008	0.011	0.011

Table 4.1: Average σ_{68} of the relative error in the background prediction across all the narrow bands for `BKGnet` trained before and after the camera intervention. We list the results for the data sets without filtering out cutouts affected by sources ('sources'), and if we remove these ('filtered').

method, the annulus method after correcting scattered light with sky-flats, and `BKGnet`. By default, cutouts with pixels above 100 000 counts are excluded from the training and test sample (column 'filtered' in the table). We have tested the impact of very bright objects in the cutouts by training and testing on a sample that has not been previously filtered from bright sources (column 'sources' in the table). The annulus and `BKGnet` precision are affected by bright sources neither before nor after the camera intervention. However, the correction with sky-flats worsens when the cutouts contain very bright sources, which indicates that the scattered-light template modelling is not sufficiently flexible towards bright-distorting effects. This only happens after the camera intervention, which suggests that, before the camera intervention, scattered light is the main source of bias and very bright sources are a second-order effect. `BKGnet` learns the underlying behaviour of scattered-light similarly to the scattered-light templates. However, the network also sees the cutout, enabling a more flexible correction in the presence of other artefacts (e.g. sources and cosmic rays).

`BKGnet` also provides an estimate for the uncertainty in its background measurements (σ_{net}). To validate such uncertainties, we use the distribution

$$\mathcal{G} = (b_{\text{pred}} - b_0)/\sigma_{\text{pred}}, \tag{4.9}$$

where b_{pred} is the background measurement and b_0 is the background label. If the uncertainties are consistent with the measurements, the distribution of $\mathcal{G}$ must be a Gaussian with zero mean and unit variance. Figure 4.10 shows the theoretical $N(0,1)$ (black solid line) and the $\mathcal{G}$ distribution for `BKGnet` (blue distribution) and the annulus (orange distribution) background measurements. While `BKGnet` predictions fit the theoretical $N(0,1)$ indicating that uncertainties are robust and consistent, the background measurements uncertainties with the annulus are underestimated by 47%.

4.6 `BKGnet` validation

The results presented in § 4.5 show that `BKGnet` yields to better background estimates compared to the annulus-based methods (see Fig. 4.9 and Tab. 4.1), while providing consistent background uncertainties σ_{pred} (Fig. 4.10). So far, these results validate cutouts without target galaxies. In this section we evaluate the performance of `BKGnet` background estimates on

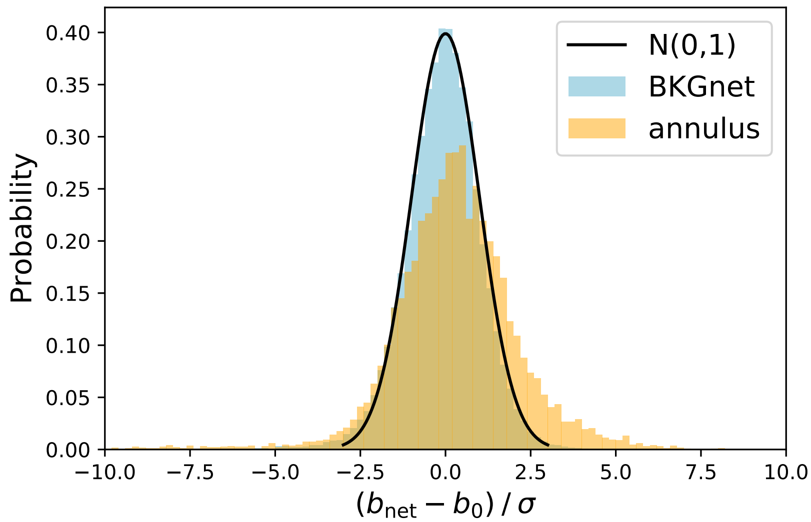

Figure 4.10: The distribution of $(b_{\mathrm{net}} - b_0)/\sigma$, where b_{net} is the background prediction and b_0 the true background and σ is the uncertainty in the prediction. We expect the distribution to be a Gaussian centered at zero with unit variance. We show the distribution for the annulus (orange) and `BKGnet` (blue) predictions.

PAUCam cutouts centred at target galaxy positions.

4.6.1 Generating the PAUS catalogue with `BKGnet` predictions

We have estimated the galaxy photometry of PAUS galaxies in the COSMOS field using aperture photometry flux measurements (see § 3.2) and subtracting the background noise estimated with `BKGnet` (Eq. 3.4). This catalogue is compared to that using `MEMBA` (§ 3.4.2), which uses the same aperture flux measurements and the background light estimated with an annulus. The catalogues contain ~ 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 flux uncertainty with the annulus background-light measurements is the estimated with Eq. 3.5. [2]. In contrast, when the background predictions are from `BKGnet`, the flux uncertainty is

$$\sigma_{\mathrm{src}} = f_{\mathrm{src}} + N(b_{\mathrm{pred}} + RN^2) + N^2\sigma^2_{\mathrm{pred}}\,, \tag{4.10}$$

where RN is the read-out noise, f_{src} is the flux aperture, and N is the number of pixels in the flux aperture.

There are three main contributions to the flux uncertainty: the uncertainty in the background-subtracted galaxy flux, the uncertainty in the background light, and the uncertainty intro-

[2] http://wise2.ipac.caltech.edu/staff/fmasci/ApPhotUncert.pdf

duced by the background subtraction (§ 3.2). Galaxy fluxes measured with `BKGnet` background estimates assume that the uncertainty in the background-subtracted galaxy flux and in the background light are captured by shot noise. The background-light uncertainty has an additional contribution from the read-out noise. The last term in Eq. 4.10 accounts for the error in the background-noise subtraction, which is the quantity directly predicted by the network $\sigma_{\rm pred}$.

In contrast, using the annulus background-light estimates, the background-light uncertainty is given by the mean-variance per pixel of the pixels within the annulus, which already accounts for other error contributions such as the read-out noise. The flux uncertainty contribution due to the background subtraction is determined by the subtraction of a background noise measured at a distance to the exact target galaxy position. This is explained in detail in § 3.2.

4.6.2 Validating the catalogues

A direct comparison between the `BKGnet` and the `MEMBA` photometry catalogues shows a 2% difference between their photometric fluxes. Also, the flux uncertainties are 4% lower with the aperture photometry from `MEMBA`. Scattered light only affects objects near the edges of the images. Therefore, for most of the galaxies in PAUS data, the background should be (approximately) flat so that we do not expect large differences between the `BKGnet` and the `MEMBA` catalogues.

To determine which catalogue provides more accurate photometry, we use the fact that PAUCam takes multiple observations of the same object in all narrow-band filters. Then, we can compare the flux measurement from different exposures of the same object, which should be compatible within errors after subtracting the background noise. This formulates as

$$D \equiv \frac{e_1 - e_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\,, \tag{4.11}$$

where e_i are different exposures of the same object and σ_i, ther 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.

Figure 4.11 shows the width of the duplicates distribution (Eq. 4.11) as a function of wavelength for the photometry with `BKGnet` background noise measurements (black line) and `MEMBA` annulus measurements (orange line). The solid lines correspond to catalogues containing all observations. In contrast, dashed lines present the same results excluding objects flagged as problematic due to strongly varying backgrounds (scattered-light, § 3.4.2). Dropping flagged objects does not significantly change the measurements for `BKGnet`, but we note a clear improvement for the `MEMBA` measurements. The improvement is particularly prominent for the NB755 filter (at 7500Å), which is affected by telluric absorption of O_2 in the atmosphere. Interestingly `BKGnet` learns how to deal with these objects, indicating robustness towards various sources of bias, not only scattered-light. Considering all narrow bands, we find $\langle \sigma_{68}[D] \rangle = 1.00$ for `BKGnet`, which is what we would expect for correct photometry.

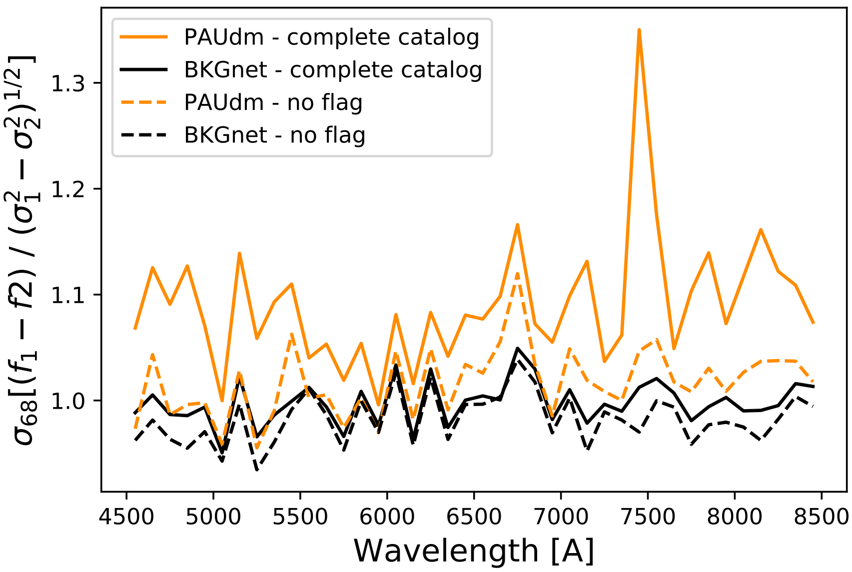

Figure 4.11: `BKGnet` validation with the duplicates distribution test. The plot shows the width of the distribution defined in Eq. 4.11 as a function of wavelength for the photometry with with `BKGnet` background measurements (black line) and the current `MEMBA` catalogue with annulus background predictions (orange line). The dashed line corresponds to the results excluding all objects flagged as affected by a strongly varying background in `MEMBA`. The solid line includes all objects.

On the other hand, the current `MEMBA` catalogue yields $\langle\sigma_{68}[D]\rangle = 1.10$, i.e. it underestimates the uncertainties.

Fig. 4.12 explores the robustness of the flux uncertainties with brightness showing $\sigma_{68}[D]$ as a function of the Subaru $i_{\rm AUTO}$ magnitude. The `MEMBA` measurements (orange dashed line) present a strong trend with magnitude, with 20% overestimated errors at the brightest end. To explore the origin of the trend, the blue dotted line considers the background predictions from `BKGnet` with errors from the annulus method and presents a similar trend with magnitude. This indicates that the annulus background uncertainties are triggering the trend with magnitude. Furthermore, `BKGnet` background predictions with the annulus uncertainties (blue dotted line) show better results than annulus background predictions with its own uncertainties (orange dashed line), indicating that the background measurements from `BKGnet` are more accurate than those estimated with an annulus.

To further validate the `BKGnet` catalogue we run `BCNz2` (Eriksen et al., 2019) using the fluxes determined using `BKGnet` background-light measurements. All objects flagged as problematic (e.g. scattered-light, vignetting, and cosmic rays) are excluded from the analysis in order to use exactly the same objects as in Eriksen et al. (2019). Nevertheless, as shown in Figs. 4.11 and 4.12, many of these objects would not need to be excluded from the data sample when their background noise is estimated with `BKGnet`. The photo-zs are compared to secure spectroscopic estimates from zCOSMOS DR3 (Lilly et al., 2007) with $i_{\rm AB} < 22.5$.

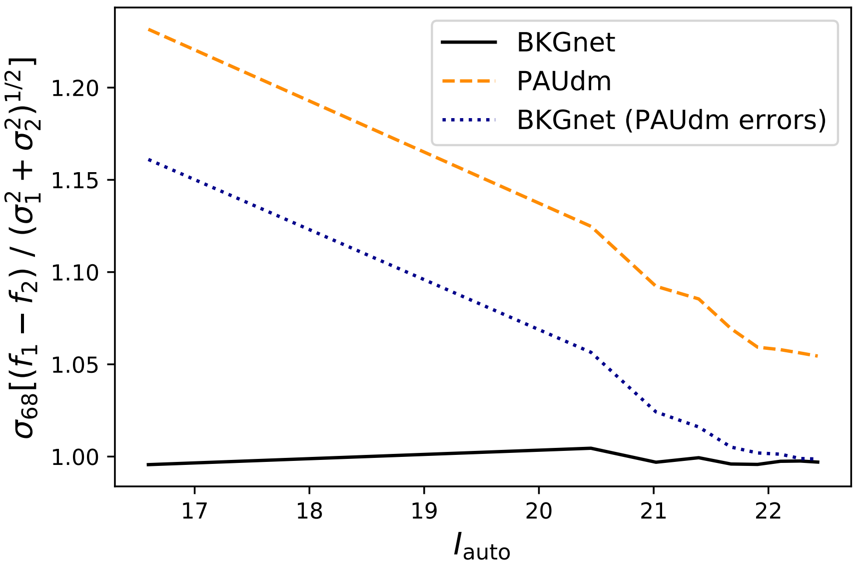

Figure 4.12: `BKGnet` validation with the duplicates distribution test. We plot the width of the distribution defined in Eq. 4.11 as a function of $i_{\rm AUTO}$ in the Subaru i-band for the catalogue generated with `BKGnet` (black solid line), the current `MEMBA` catalogue (aperture photometry with an annulus, orange dashed line) and a mixed catalogue with the predictions from `BKGnet` and the uncertainties from the annulus method (blue dotted line).

We split the sample based on a quality parameter defined as:

$$\mathrm{Qz} \equiv \frac{\chi^2}{N_\mathrm{f}-3}\left(\frac{z^{99}_{\rm quant} - z^{1}_{\rm quant}}{\mathrm{ODDS}(\Delta \mathrm{z} = 0.01)}\right), \tag{4.12}$$

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

$$\mathrm{ODDS} \equiv \int_{z_\mathrm{b}-\Delta z}^{z_\mathrm{b}+\Delta z} dz\ \mathrm{p}(z), \tag{4.13}$$

where z_b is the mode of the $p(z)$ and Δz defines a fixed redshift interval around the peak. PAUS has a strict outlier definition

$$|z_{\rm photo} - z_{\rm spec}|\ /\ (1+z_{\rm spec}) > 0.02\,, \tag{4.14}$$

while in broad-band photometry, a common outlier definition is $|z_{\rm photo} - z_{\rm spec}| > 0.15\,(1+z_{\rm spec})$, e.g. Ilbert et al. (2006); Bilicki et al. (2018).

Table 4.2 lists the outlier rate and the photometric redshift precision obtained with `BCNz2` on the `MEMBA` and `BKGnet` catalogues. To quantify the redshift precision we use σ_{68} (Eq.3.22). 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%) this improvement is small. A potential reason is that, in the full sample, outliers are dominated

	Outlier percentage		$10^3\sigma_{68}$	
Percentage	`BKGnet`	`MEMBA`	`BKGnet`	`MEMBA`
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

Table 4.2: Photo-z outlier rate and accuracy obtained with BCNz2 for the `BKGnet` and the `MEMBA` catalogues. The percentages correspond to the samples selected by the photo-z quality parameters Qz.

by photo-z outliers rather than outliers on the photometry itself. However, if we cut using the Qz parameter to get the best 20% and 50% of the sample, we notice that the outlier rate reduces significantly. These outliers should be dominated by photometric outliers. For the best 50% of the sample, we reduce the number of outliers by 25%. Furthermore, for the best 20% of objects, this improvement increases to 35%. This shows once more that `BKGnet` is a statistically accurate method that is also robust.

4.7 Conclusions

Imaging surveys need accurate background subtraction methods to obtain precise source photometry (§ 3.2). We have developed a deep learning method to predict the background light in astronomical images with strongly varying background noise (§ 4.3). The algorithm has been developed to predict the background on images taken with PAUCam. The edges of PAUCam images are affected by scattered-light (§ 4.2.2), 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 (§ 2.2.2 and Fig. 4.2).

Scattered light follows the same spatial pattern within the CCD and scales approximately linear with the background level (§ 4.2.2). We have constructed scattered-light templates and background ratio maps by combining images taken with the same narrow band and normalised by their background level. These scattered-light templates can be used to correct for scattered-light (§ 4.2.3). Nevertheless, background fluctuations due to external conditions (e.g. moon, seeing, airmass) can trigger nightly differences in scattered-light patterns. To accurately correct scattered-light with scattered-light templates, we would need to generate a scattered-light template per narrow band and night. Nevertheless, fluctuations during the night or a small number of available images in a narrow band can lead to inaccurate scattered-light corrections.

`BKGnet` is a deep learning-based algorithm that predicts the background and its associated uncertainty behind target sources accounting for scattered light and other distorting effects. `BKGnet` consists of a CNN followed by a linear neural network (§ 4.3.1). In the training set, we use empty cutouts, i.e. without a target galaxy, that enable the estimation of the ground-truth background light (§ 4.3.3). 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.

On empty regions with known ground-truth background, BKGnet improves over the annulus prediction once images have been corrected with the scattered-light template by 37% before the camera intervention and 17% after. The scattered-light template correction fails in many of the bands, especially on the bluer filter tray, which is affected the most by scattered-light (§ 4.4.2). The background-light uncertainties provided by BKGnet are also consistent (§ 4.5) while with the annulus method, these are underestimated by 47%. To test the network on PAUCam data, we have generated the PAUS photometry catalogue in the COSMOS field using BKGnet background-light measurements. The performance is evaluated by comparing the flux measurements of independent observations of the same object in the same band. The results show that BKGnet provides more robust uncertainties than aperture photometry, fixing a strong systematic trend with i-band magnitude (§ 4.6). The flux measurements are also more robust towards scattered-light and other sources of bias e.g. regions with high atmospheric absorption (Fig. 4.11). The BKGnet catalogue also reduces the BCNz2 photo-zs outlier rate by a 25% and 35% respectively for the best 50% and 20% photo-z samples, while the accuracy is not affected.

With BKGnet we have optimised 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 estimating the photo-zs directly from images intrinsically require estimating the object's photometry. Therefore, the understanding from BKGnet will also help to optimise such deep learning algorithms. Although the network has been tested with PAUCam images, the concept should also apply to future imaging surveys as *Euclid* and LSST.

4.A Variable annulus

Currently, the MEMBA pipeline estimates the background with an annulus of inner and outer radii fixed at 30 and 45 pixels from the source, respectively (§ 3.4.2). However, the target galaxy photometry can be optimised by adjusting the annulus radii galaxy-wise.

The annulus approach measures the background noise at a certain distance from the target source. This is not problematic for flat backgrounds, but if the background noise varies across the image, e.g. in the presence of scattered light, the target galaxy could be in a flat region while an annulus located far from the target source has scattered-light contributions. In this situation, the annulus would capture an amount of extra background that we define as $\Delta_{\rm B}$. On the other hand, an annulus set too close to the target source can capture light contributions from the source itself, especially if this is bright and large. We define the extra light contribution to the annulus from the target galaxy as $\Delta_{\rm F}$.

The relative error in the background prediction due to scattered-light and target-source contributions is

$$\Gamma \equiv \frac{|\Delta_{\rm F} + \Delta_{\rm B}|}{\sigma_{\rm b}}\,, \tag{4.15}$$

where $\sigma_{\rm b}$ is the error on the background subtraction. The annulus location minimising Γ is that providing the most optimal backgroud measurement. This location depends on the size and brightness of the target object, which is also affected by the PSF.

Figure 4.13 shows the histogram of minimum Γ measurements for galaxies at the center (orange) and at the borders (black) of the images. We evaluate Γ on a set of PAUCam galaxy image simulations (§ 4.4.1) for annulus radii from one to forty pixels from the target source, always keeping the annulus area such that $r_{\rm out} - r_{\rm in} = 15$ pixels.

Both histograms (centre and border target galaxy locations) show a large fraction of galaxies for which the annulus can accurately predict the background light (low gamma values). Galaxies at the centre of the image are in flat background regions where the annulus method provides accurate background-light measurements. Furthermore, many galaxies located at the border of the image are not affected by scattered light, also presenting accurate background-light measurements. For galaxies in scattered light regions, the annulus should locate very close to the target galaxy to minimise the effect of the varying background. This is only possible for target galaxies with $r_{50} \sim 1$ or 2 pixels.

The Γ distribution for galaxies at the border of the image presents a tail of measurements where the annulus estimation cannot be an optimal position. These are cases where the annulus captures background variations if this is set too far from the target source. However, the annulus also captures light from the target source if these two are too close. In other words, if the background variation is strong, the annulus will tend to get closer to the target source in order to minimize $\Delta_{\rm B}$. However, this is not possible for bright and large galaxies, since getting closer to the target increases $\Delta_{\rm F}$.

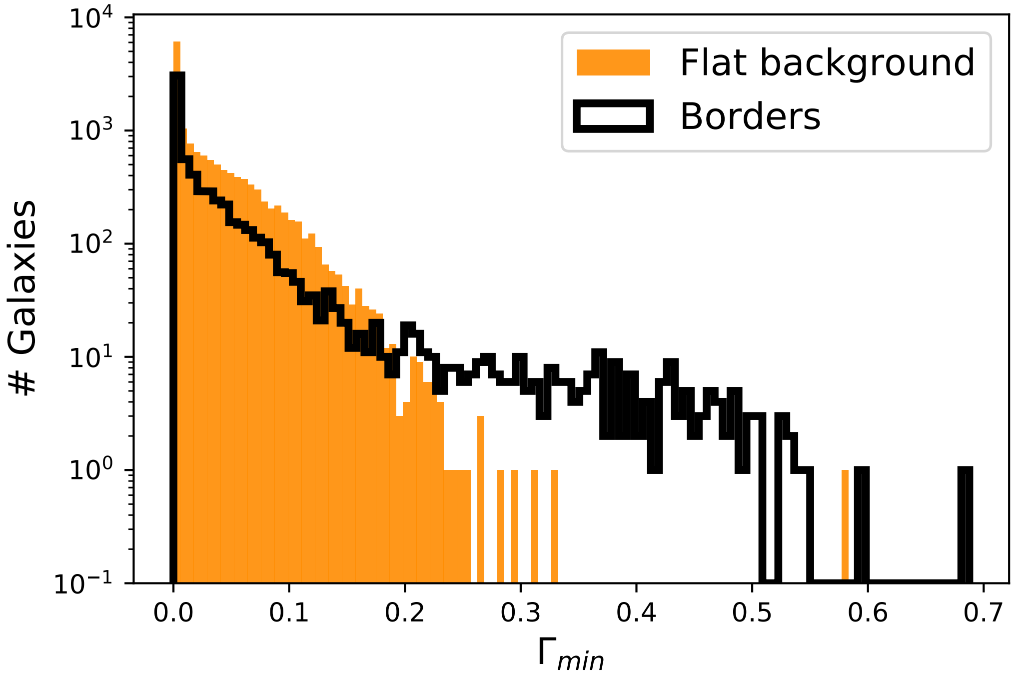

Figure 4.13: The $\Gamma_{\rm min}$ distribution for galaxies at the border of the CCD (scattered-light affected, black distribution) and at the center (flat background, filled orange) of the image.

Chapter 5

Single band photometry

5.1 Motivation

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. SDSS, 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 ≈ 2 orders of magnitude ahead in terms of number of objects. However, these are observed with a lower spectral resolution, which makes the redshift measurements less precise. Current and past imaging surveys, e.g. The KiDS, The Hyper Supreme-Cam Subaru (HSC, Aihara et al., 2018), and DES have detected hundreds of millions of galaxies and oncoming surveys like Euclid and LSST will increase this number to billions. Consequently, fast and precise methods to analyse and extract galaxy properties (e.g. flux, size, and shape) are needed.

There are many different algorithms to estimate galaxy photometry. One widely used example is `SExtractor` (Bertin & Arnouts, 1996), which applies aperture photometry (Ni et al., 2019) inspired by the Kron first moments algorithm (Kron, 1980). 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 (Heasley, 1999), which consists of fitting the galaxy image to a theoretical model and extracting its photometry. This includes the `GaaP` (Kuijken, 2008) algorithm, which estimates the total flux by fitting the pixelated galaxy images to polar shapelets splitting the galaxy image into components with explicit rotational symmetries (Refregier, 2003; Massey & Refregier, 2005).

There are many other examples as e.g. `ProFound` (Robotham et al., 2018), `T-PHOT` (Merlin et al., 2015) and `Tractor` (Lang et al., 2016), and each of them is adjusted to outperform the others on a particular data set. For instance, a photometry algorithm can be optimised to work very well on images with many blended galaxies (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 parametrisation. In contrast, machine learning techniques learn a model adapted to the data. Deep learning has proven to be very powerful in 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 concludes 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., 2018c), cosmic ray rejection (Zhang & Bloom, 2019) 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 chapter, we present `Lumos`[1], a deep learning based algorithm to extract the photometry from astronomical images. It consists of a CNN (§ 1.2.1) that works on input galaxy images and a MDN (§ 1.2.2) that outputs the probability distribution of galaxy fluxes. `Lumos` builds on `BKGnet` and estimates the probability distribution of the background-subtracted galaxy flux, which requires the implicit estimation and subtraction of the background noise. `Lumos` is also developed and tested using PAUS images (§ 2) although it can be adapted to any imaging survey, like e.g. *Euclid* or LSST.

The structure of this chapter is as follows. Section 5.2 presents the simulations we use for training. Then, § 5.3 introduces different flux estimation alternatives that we have compared to `Lumos` performance. In §5.4, we introduce `Lumos`, its architecture and the training procedure. Section 5.5 presents `Lumos` results on simulations, including validation of the flux probability distributions, a comparison with alternative flux estimation methods, and deblending tests. Finally, §5.6 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 §5.7.

5.2 Data

In this section, we present PAUS data (§ 5.2.1) and Teahupoo simulations (§ 5.2.2), the simulated galaxy images used throughout the paper.

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

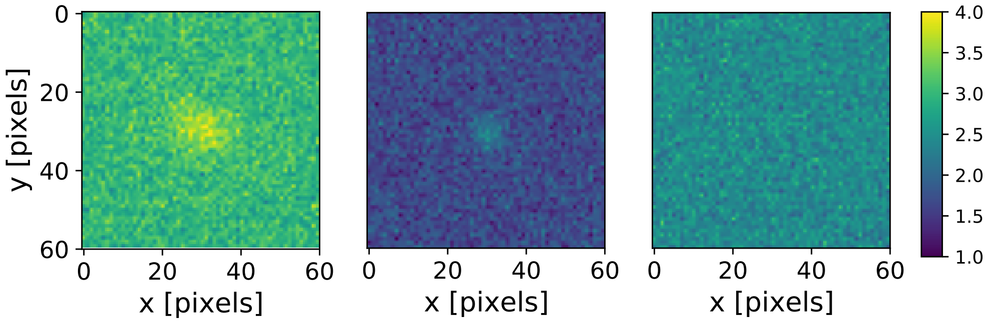

Figure 5.1: From left to right, TEAHUPOO galaxy images with i-band magnitudes $18.5, 20.3$ and 22.3, in PAUS "NB685". These are simulated with a exposure time of 90 seconds, the baseline PAUS exposure time in the "NB685" filter.

5.2.1 PAUS data

In this chapter, we again use the PAUS data taken in the COSMOS field (§ 3.4.4), which comprises a total of 9749 images, 243 images in each narrow band. The complete photometry catalogue comprises 64 476 galaxies to $i_{\mathrm{AB}} < 23$ in 40 narrow-band filters, which corresponds to around 12,5 million galaxy observations ($\sim$5 observations per galaxy and narrow-band filter). While observing COSMOS, the PAU camera was modified to mitigate the effect of scattered light (§2.2.2), which changed the noise patterns of PAUS images (§ 4.2). Half of the images in COSMOS were taken before the camera modification and the other half, after.

PAUS data reduction process consists of two pipelines: the `Nightly` (§ 3.4.1), which performs an instrumental de-trending processing, e.g. electronic and illumination biases, and `MEMBA` (§ 3.4.2), which applies forced aperture photometry to targets selected from an external detection catalogue. The deep learning network 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 intends to reduce the error propagation caused by errors in the detection catalogue profile parameters (e.g. the half-light radius and the Sérsic index).

5.2.2 Teahupoo simulations

We have constructed the TEAHUPOO[2] simulations, a set of PAUS-like galaxy image simulations. Three examples of TEAHUPOO galaxies with i-band magnitudes $i_{\mathrm{AB}} = 18.5, 20.3$ and 22.3 are shown in Fig. 5.1. Note that already at $i_{\mathrm{AB}} \approx 20$ it is hard to distinguish the galaxy from the background noise and with $i_{\mathrm{AB}} > 22$, the galaxy signal is visually masked by background fluctuations.

TEAHUPOO light profiles are modelled with a single Sérsic profile (Eq.3.14). We jointly sample the half-light radius, the Sérsic index (n_{s}), and the ellipticity from their distributions in the COSMOS field (Fig. 5.2), which are provided by Ilbert et al. (2009). This ensures that the correlation between the shape and the size of galaxies is represented in the training

[2]Named after the favourite sandwich of the author.

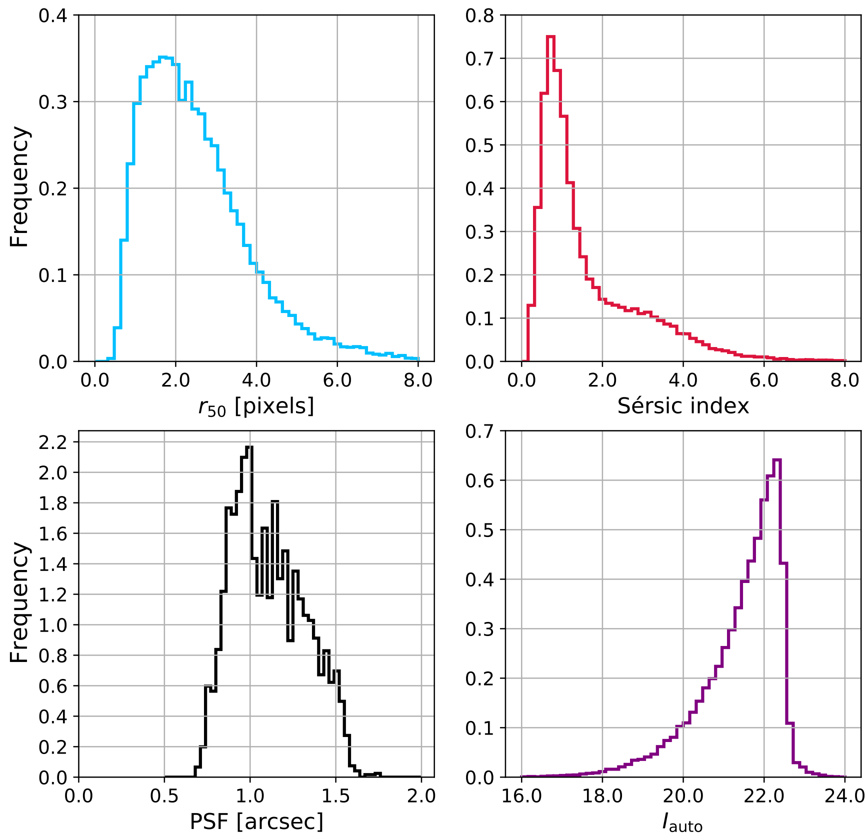

Figure 5.2: Distributions of the half-light radius (r_{50}) (top left), Sérsic index (n_{s}) (top right), the PSF-FWHM (bottom left), and the I_{auto} magnitude of PAUS galaxies in the COSMOS field (bottom right).

sample. Elliptical galaxies are simulated by elongating the half-light radius according to the b/a distribution.

Teahupoo image simulations are 60×60 pixels and are generated with `Astropy` (Astropy Collaboration et al., 2013; Price-Whelan et al., 2018). `Astropy` methods evaluate the galaxy profiles at the centre of each pixel instead of integrating along a pixel. This is problematic for small and steep galaxies (i.e. with high n_{s}), where the flux changes significantly along the pixel. To correct for this effect, galaxies are initially drawn in a 600x600 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 PAUCam images. These cutouts can contain artefacts such as e.g. other galaxies, cosmic rays, and crosstalk. This has proven very important for our network, as it learns how to make predictions when they are present (see e.g. §5.5.3 and §5.6.1). The background light noise patterns across the CCD are narrow-band dependent and changed when the PAUCam camera was modified (see §4.2 for more details). For this reason, the background cutouts are taken from a PAUCam image observed with the same narrow-band filter we are simulating. We also track if the image was observed before or after the camera intervention. The galaxy signal is also wavelength-dependent. Consequently, we independently sample the galaxy flux from forty flux distributions, one per narrow-band 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. 5.2 (bottom left 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 and electronic noise, are not required as the background cutout already includes them (§ 3.1). This is another benefit of using PAUCam background cutouts, as simulating realistic noise is often hard and could easily lead to differences between simulations and data.

As simulations use PAUS flux measurements and PAUCam background cutouts, outlier measurements in any of these two, e.g. background images with spurious effects and outlier flux measurements, might end up represented in the Teahupoo images. To reduce the number of affected Teahupoo galaxies, we clip the PAUS flux distribution at 0 and 1000 e^-/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 to filter out very poor observations. However this does not 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 to real survey images. In contrast, `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.

5.2.3 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 (§ 1.2). 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 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

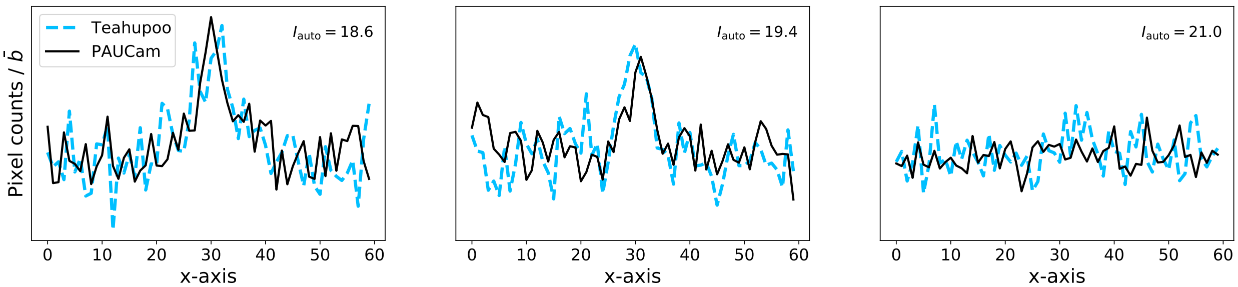

Figure 5.3: Comparison between single pixel row light profile for pairs of PAUCam galaxies (solid black line) and TEAHUPOO simulated galaxies (dashed light blue). TEAHUPOO galaxies are constructed to exactly mimic its real PAUCam pair. The plot shows the pixel value along the central row of pixels (crossing the source) divided by the mean image background noise. Therefore, the central peak at x = 30 corresponds to the galaxy light contribution. From left to right, the galaxy magnitudes are $i_{\mathrm{AB}} = 18.6$, 19.4 and 21.0.

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.

Figure 5.3 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 normalised with the mean background (excluding the source) along the central row of the cutout. Therefore, pixels without galaxy light contribution should fluctuate around unity, while pixels with higher values show 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 centre galaxy, the two galaxies also match reasonably well. However, it also exhibits a slight 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 are can barely distinguished from background-noise fluctuations (see also Fig. 5.1).

Several effects could bring variations between TEAHUPOO and PAUCam galaxies. For instance, inaccuracies in any step of the data reduction process (e.g. the photometric calibration, the astrometry, and the PSF measurement) are not represented in the simulations. Currently, the PSF is assumed constant across the PAUS image. This could potentially yield discrepancies between simulations and data. Additionally, if a single Sérsic function is not sufficient to model a PAU galaxy, that would also imply a difference between the two images. These discrepancies will propagate into larger errors in the flux estimation. However, inaccuracies in the calibration, the modelling, and 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 enables detecting inaccuracies in e.g. the astrometry and the PSF, which are also accounted for in the uncertainty measurement (see last paragraph in §5.6.1).

5.3 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 (§5.3.1), aperture photometry (§5.3.2), and a linear weighted sum of the galaxy pixels (§5.3.3).

5.3.1 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 $I(r) = I_\mathrm{e} R(r)$, where I_e is the profile amplitude and $R(r_\mathrm{i})$ corresponds to a Sérsic light profile at pixel i, we can fit the image to the theoretical profile with

$$\chi^2 = \sum_i \frac{(f_i - I(r_i))^2}{\sigma^2_{F,i}}, \tag{5.1}$$

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 $I(r_\mathrm{i})$ is the galaxy theoretical model in pixel i. Assuming Poisson errors, the previous equation becomes

$$\chi^2(I_e) = \sum_i \frac{\left(f_i - I_e R(r_i)\right)^2}{I_e R(r_i) + B}, \tag{5.2}$$

where B is the mean background per pixel. The total flux is measured as the I_e minimising Eq. 5.2. Note that the parameter I_e appears twice in the equation, which makes the closed form not feasible. Instead, we have minimised Eq. 5.2 with a Nealder-Mead algorithm from `SciPy` (Jones et al., 2001).

5.3.2 Aperture photometry

Aperture photometry (§ 3.2, 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 (§ 3.4.2 and Serrano et al. in prep.). This approach measures all the pixel contributions inside an aperture of radius R with subtraction of the background light (Eq. 3.4).

In PAUS, the apertures are elliptical, and their areas target a fixed amount of galaxy light (§ 3.4.2), in our case 62.5% of the flux. Therefore, obtaining the total flux requires scaling the measurement by 1/0.625. For 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 $R_\mathrm{in} = 30$ pixels, $R_\mathrm{out} = 45$ pixels centered at the targeted galaxy.

5.3.3 Weighted pixel sum

In aperture photometry, all pixels within the aperture contribute equally to the total flux, i.e. pixels at the galaxy border and at the centre of the galaxy contribute the same. This is not optimal in terms of total signal-to-noise, 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, some providing a higher signal-to-noise than others. Indeed aperture photometry is just a concrete 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 signal-to-noise (SNR), which can be written as

$$\mathrm{SNR} = \frac{\sum_i w_i m_i}{\sqrt{\sum_i w_i^2 (m_i + b_i)}}\,, \tag{5.3}$$

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 signal-to-noise (Eq. 5.3) as a function of the pixel weights w_i leads to

$$w_\mathrm{x} = \frac{\sum_\mathrm{i} m_\mathrm{i}}{\sum_\mathrm{i} m_\mathrm{i}^2/(m_\mathrm{i} + b_\mathrm{i})} \frac{1}{1 + b_\mathrm{x}/m_\mathrm{x}}\,, \tag{5.4}$$

where w_x is the optimal weight of pixel x (see Appendix 5.A for a more detailed derivation). This is the linear estimator providing the most precise unbiased flux measurements. However, obtaining the optimal weights in Eq. 5.4 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. Nonetheless, this methodology puts a limit on how well linear methods can measure galaxy fluxes, which can be used to benchmark the `Lumos` performance.

5.4 `Lumos`: Measuring fluxes with a CNN

In this section, we describe `Lumos`, our deep learning algorithm to measure the photometry of astronomical objects. We will discuss the network's input data (§5.4.1), its architecture (§5.4.2) and the training procedure (§5.4.3 and §5.4.4).

5.4.1 Input data

The `Lumos` input consists of two types of data, the most important input including two images of 60×60 pixels. The first one is made of an image cutout centered at the target galaxy and covering 16x16 arcsec of the night sky (given that PAUCam pixel scale is 0.263, § 2.2.2). Although most PAUS galaxies have a half-light radius between 1 and 3 pixels, which would not require such a large cutout, we already showed that the network needs larger stamps to accurately model the background noise fluctuations and scattered-light patterns (§ 4). The second image contains the convolved galaxy profile drawn using the parameters from an external detection 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 §5.2.2), while in the testing phase these of PAUCam 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:

1. 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.

2. 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 (Figs. 4.2 and 4.1).

3. The narrow band filter identification. The galaxy flux distribution is different for each narrow band filter. Furthermore, the scattered-light pattern also depends on the narrow-band filter. Therefore, the band provides valuable extra knowledge of the expected flux and background noise pattern.

4. A camera intervention flag. The camera was modified while observing the COSMOS field (see § 3.4.4). Therefore, the network also benefits from knowing if an image was taken before or after the camera intervention. As we did in `BKGnet`, this information is combined with that of the narrowband filter and given as an 80×10 trainable matrix (see § 5.4.2 for more details). In practice, the network effectively works for all intents and purposes as having 80 different narrowband filters instead of 40.

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

5.4.2 `Lumos` architecture

The `Lumos` architecture (see Fig. 5.4) has two differentiated parts, a CNN (§ 1.2.1) and a MDN (§ 1.2.2). The CNN works directly on the input images, and it builds with five blocks of convolution, pooling, and batch normalisation layers. These layers are represented in Fig. 5.4 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 narrow-band filter it was observed with, and the camera intervention flag (see §5.4.1). For the band information, the network uses an 80×10 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. The MDN in `Lumos` consists in four fully-connected layers with parameters 5000:1000:100:15 (§ 1.2). It outputs five mixing coefficients (α) together with five pairs of (μ, σ) parametrising the Gaussian components. This kind of network architecture has already been implemented to predict PAUS photo-z probability distributions in Eriksen et al. (2020).

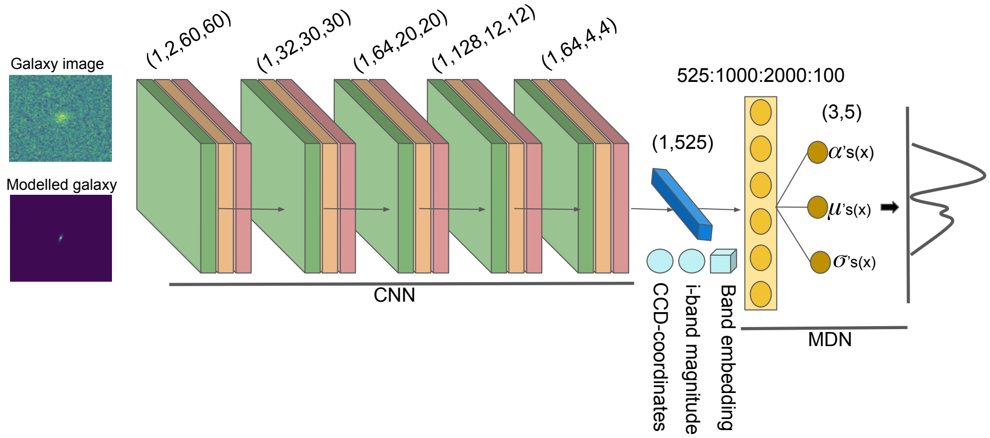

Figure 5.4: `Lumos` architecture. Orange cubes correspond to convolutional layers, blue cubes are pooling layers and the purple ones are batch normalisation layers. The vectors between two CNN blocks correspond to the dimension of the previous block's output and the input dimension of the following. The CNN's output is linearised (green stick), combined with external information (coordinates, NB filter and i-band magnitude) and input to a MDN. The MDN outputs a probability distribution of the total flux as a linear combination of five Gaussians.

The choice of loss function is a crucial step in the construction of a neural network (§ 1.2). `Lumos` combines two loss functions, the first as in Eq. 1.5, which is the most common loss function for Gaussian MDN, and a modified version of it in Eq. 5.6, which is 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 ∞, which changes the PDF normalisation, i.e.

$$\log\left(\mathcal{G}(x)\right) = \log\left(\frac{2}{\pi}\right) + \log\left(\exp\left(\frac{-\frac{1}{2}(x-\mu)^2}{\sigma^2}\right)\right) + \\ \log\left(\sigma \mathrm{Erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + \sigma\right), \tag{5.5}$$

which leads to the following loss function

$$\mathcal{L}_{\mathrm{MDN}} = \sum_{\mathrm{i}=1}^{k} \log\left(\alpha_{\mathrm{i}}\right) - \frac{(f_{\mathrm{i}} - \bar{f})^2}{\sigma_{\mathrm{i}}^2} - 2\log\left(\sigma_{\mathrm{i}}\right) - \\ \log\left(\mathrm{Erf}\left(\frac{\bar{f}}{\sqrt{2}\sigma_{\mathrm{i}}}\right) + 1\right), \tag{5.6}$$

where again $\bar{f}$ is the true flux and σ_{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 Eq. 1.5 on objects with $i_{\mathrm{AB}} < 20.5$ and Eq. 5.6 for the rest of the sample. Choosing a single loos function for all magnitudes also works, however combining the two improves the photometry for the faintest and the brightest galaxies.

5.4.3 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 discriminate between horses and zebras. Instead of training from scratch, the zebra-horse classifier starts with the optimised weights from the cat-dog classifier, in such a way that the network has already learnt to extract shared features, e.g. detecting edges and shapes. Such mutual features are commonly extracted in the shallower layers of the network, while deeper layers pick up more subtle data traits. For this reason, many times transfer learning only requires training deeper layers of the network, while the shallower ones are kept from the initial network.

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 not enough labelled data to train a network from scratch. A possible solution in such cases is to train the network on simulations and use a small labelled dataset to adapt the model to the data. This requires two consecutive trainings, one initial on simulations and an additional one on data with the network parameters from the training on simulations as a starting point. Domínguez Sánchez et al. (2019) uses transfer learning to adapt a morphology classifier trained on SDSS images to work on DES images.

`Lumos` is trained on TEAHUPOO image simulations (see §5.2.2 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 narrow-band filter to implement what we call *unsupervised transfer learning*.

We have collected a set of PAUCam observed galaxy pairs, with two independent images of the same galaxy observed with the same narrow-band 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 and 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 PAUCam data before evaluating the network on the test sample.

The unsupervised transfer learning loss function compares the probability distribution of two observations. Before comparing the PDFs, these need to be calibrated with the image zero-point (see §3.4.1 for more details). The PDFs are parametrised with five Gaussian distributions (as provided by `Lumos`, see §5.4.2), in such a way that each Gaussian component in the first observation compares to all the components in the second observation. The negative log-likelihood of the difference between the two predicted PDFs takes the form of

$$\begin{aligned}\mathcal{L}_{\mathrm{UTL}} = \sum_i \sum_j \log \alpha_i + \log \beta_j - \frac{1}{2}\frac{(f_i - f_j)^2}{\sigma_i^2 + \sigma_j^2} \\ -\sqrt{\sigma_{f_i}^2 + \sigma_{f_j}^2} - \log\left[1 + \mathrm{Erf}\left(\frac{f_i}{\sqrt{2}\sigma_i}\right)\right] \\ - \log\left[1 + \mathrm{Erf}\left(\frac{f_j}{\sqrt{2}\sigma_j}\right)\right] \\ + \log\left[1 + \mathrm{Erf}\left(\frac{f_i/\sigma_i^2 + f_j/\sigma_j^2}{\sqrt{1/\sigma_i^2 + 1/\sigma_j^2}}\right)\right],\end{aligned} \tag{5.7}$$

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

5.4.4 Training procedure

`Lumos` is trained with 500 000 simulated images constructed combining simulated galaxies and PAUCam cutouts (see §5.2.2). The sample is split into 90% for training and 10% 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 by a factor of 10 every 40 epochs. We use `Adam` (Kingma & Ba, 2014) as optimisation algorithm. The training takes about 20 hours with an NVIDIA TITAN V GPU.

The network is trained with a relatively large batch size of 500 galaxies. As simulations are constructed from PAUS flux measurements and PAUCam background cutouts, Teahupoo image simulations might contain some outliers. We have filtered the training sample to reduce outliers (see the last paragraph in §5.2.2), however, a few poor training examples can still be part of Teahupoo images. The large batch size reduces their effect on the overall loss function and, consequently, on the network's training.

After the supervised training, we apply the unsupervised transfer learning (§ 5.4.3). This part is trained with 20 000 galaxy pairs. A danger of applying unsupervised transfer learning is that nothing prevents the network from biasing the flux predictions, since Eq. 5.7 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 Eqs 1.5, 5.6 and 5.7 and therefore becomes

$$\mathcal{L} = \mathcal{L}_{\mathrm{MDN}} + \mathcal{L}_{\mathrm{UTL}} \,. \tag{5.8}$$

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. Consequently, the transfer learning only varies the parameters in the last linear layer instead of training all the network parameters. This reduces the training time and reduces the risk of overfitting. 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 iterations.

5.5 Lumos flux measurements on simulations

In this section we test Lumos on Teahupoo galaxies. We first validate the flux probability distributions (§5.5.1), followed by a comparison with other flux estimation methods (§5.5.2). Finally we test how well Lumos performs on blended galaxies (§5.5.3).

5.5.1 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 § 1.2.2 and §5.4.2). Figure 5.5 shows the predicted flux PDFs for two PAUCam galaxies (solid lines) and two Teahupoo galaxies mimicking them (dashed lines). The flux probability distributions on data and the simulations are very similar, providing additional confidence on the reliability of Teahupoo galaxies. The similarity between PAUS and Teahupoo galaxies is also tested in §5.2.3.

The predicted flux PDFs are not Gaussian, e.g. the faintest galaxy in Fig. 5.5 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 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 §5.6.3 for further discussion).

Probability Integral Transform (PIT) on simulations

The Probability Integral Transform (PIT, Dawid, 1984; Gneiting et al., 2005; Bordoloi et al., 2010) tests the quality of the probability distribution. It is defined by

$$\mathrm{PIT} \equiv \int_{-\infty}^{f^*} \mathrm{d}f\, \phi(f) \tag{5.9}$$

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

In Fig. 5.6, we have estimated the PIT value for 10 000 Teahupoo galaxies with known flux. The plot shows two distributions, one including (solid blue line) and another not including (dashed red line) the CCD coordinates of galaxies. When the training does not include

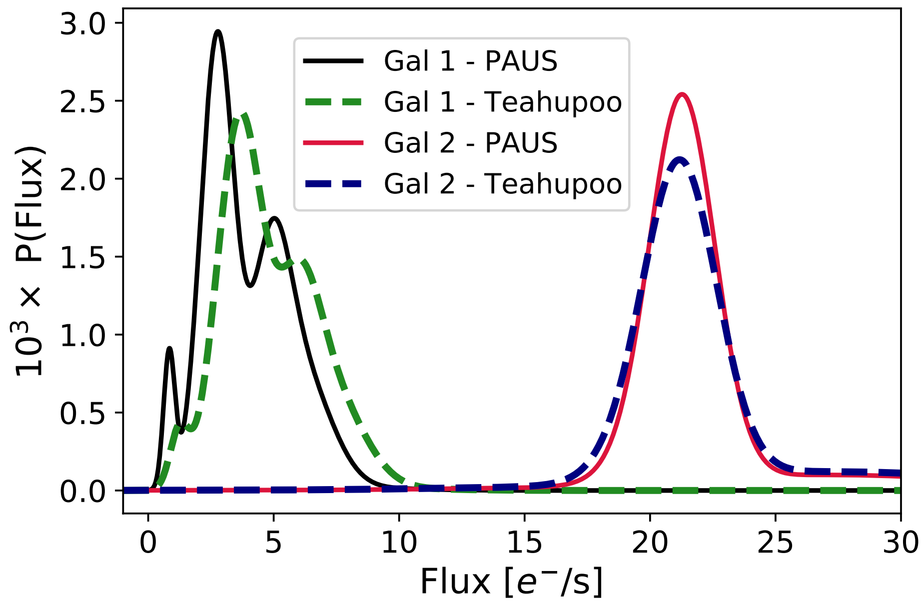

Figure 5.5: Flux probability distributions provided by Lumos for two PAUCam galaxies (Gal1, Gal2, solid lines) and their Teahupoo imitations (dashed lines).

the coordinates, the PIT distribution displays two peaks of outliers at the first and last bin of the histogram. These outliers correspond to galaxies with strongly varying background light that require accurate knowledge of the background noise patterns in the different narrowband 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 in scattered-light regions. This is consistent with the results in Fig. 4.7, where the CCD coordinates proved essential to predict accurate backgrounds (solid black line).

Single flux and flux uncertainty measurements

Although the flux PDF provides more information than single-value measurements, many applications require a single-flux measurement and its associated uncertainty. Flux point-like estimates can be computed with different statistical estimators, e.g. the mean, the median, and the peak. For (almost) Gaussian PDFs, these estimators lead to 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 this kind of PDFs commonly represents faint objects with the main peak very close to zero. Then, the secondary peaks and the tails shift the PDF towards higher fluxes (e.g. Fig. 5.17).

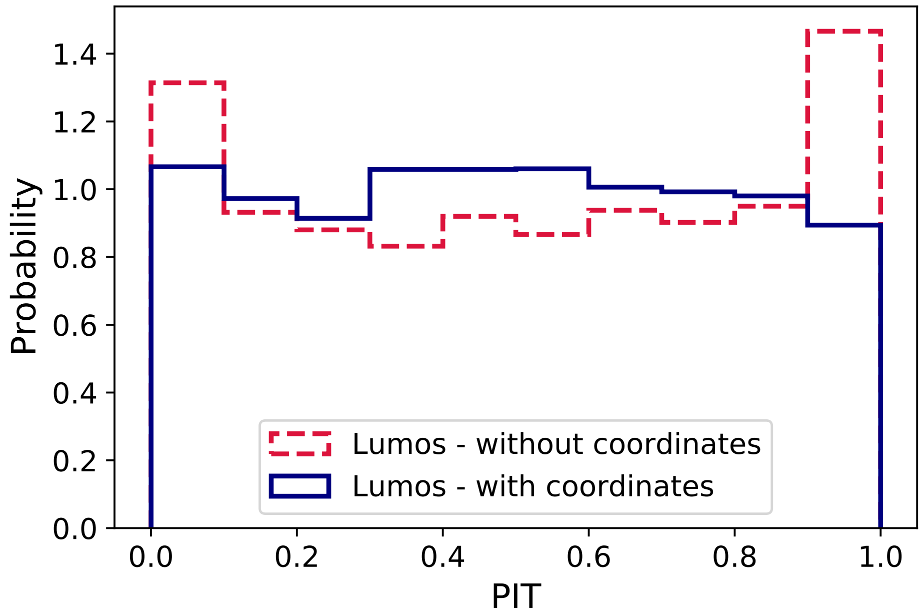

Figure 5.6: PIT distribution of the `Lumos` flux PDFs on a set of 10 000 Teahupoo galaxies. We have tested `Lumos` PDFs with (solid blue line) and without (red dashed line) including the CCD coordinates. If the CCD coordinates are not included, `Lumos` provides outliers (peaks at 0 and 1) corresponding to scattered-light affected objects.

For the flux uncertainty, the most straightforward estimator is the standard deviation. However, another possibility is σ_{68} (Eq. 3.22). For Gaussian distributions, these two estimators coincide. However, in the case of non-Gaussian PDFs, σ_{68} is more robust towards distributions with tails but provides higher uncertainties in the presence of multiple peaks (see §5.6.3 for further discussion).

Even though the median and σ_{68} are more robust with noisy PDFs, they require the explicit PDF construction from the predicted Gaussian parametrisation. This is time consuming, and 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

$$f = \sum_i \alpha_i \cdot \mu_i \,, \tag{5.10}$$

where α_i and μ_i are the mixing coefficient and the expected value of the i-th Gaussian component, respectively. The associated variance (σ_f^2) is then given by

$$\sigma_f^2 = \sum_i \left[\alpha_i \left(\sigma_i^2 + \ (\mu_i - \sum_j \alpha_j \mu_j)^2 \right) \right] , \tag{5.11}$$

where σ_i^2 is the variance of the i-th Gaussian component.

5.5.2 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, 2008). In this section, we will use simulations to compare the performance of `Lumos` with a profile fitting method (§ 5.3.1), aperture photometry (§ 5.3.2) and a linear weighted sum of pixels (§ 5.3.3). To quantify the quality of the flux measurements, we will use

$$\textbf{Bias}: \quad \text{Median}\left[(f - f_0)/f_0\right], \tag{5.12}$$

$$\textbf{Dispersion:} \quad \sigma_{68}\left[(f - f_0)/f_0\right], \tag{5.13}$$

where f_0 is the ground truth flux.

Figure 5.7 compares the bias (left panel) and the dispersion (right panel) in the flux predictions as a function of the $i_{\rm auto}$ 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% bias at the faint end. PAUS galaxies are already hard to distinguish from background fluctuations for magnitudes $i_{\rm AB} > 20$ (see Fig. 5.2), which could 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% biased. While the optimal weights (Eq. 5.4) ensure unbiased flux estimates, these also require perfect knowledge of the galaxy profile and the background light. On our simulations (§ 5.2.2), 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 (Eq. 5.4) seem to be very sensitive to inaccuracies in the background noise modelling. In contrast, aperture photometry and `Lumos` provide unbiased estimates to a 5% level up to magnitude 22.

In terms of dispersion, `Lumos` is the most precise method with $\sigma_{68} \approx 0.36$. This implies a 28% improvement with respect to the linear pixel weighting method, which is the second-best method with $\sigma_{68} \approx 0.47$. As expected, the optima-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 narrow-band filters. Considering flux measurements in each narrow-band filter independently, bluer bands with lower signal-to-noise have a higher dispersion. Furthermore, the flux measurements are unbiased to a 3% level in all narrow-band filters, and these also show unbiased for all galaxy sizes (r_{50} and PSFs.

5.5.3 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

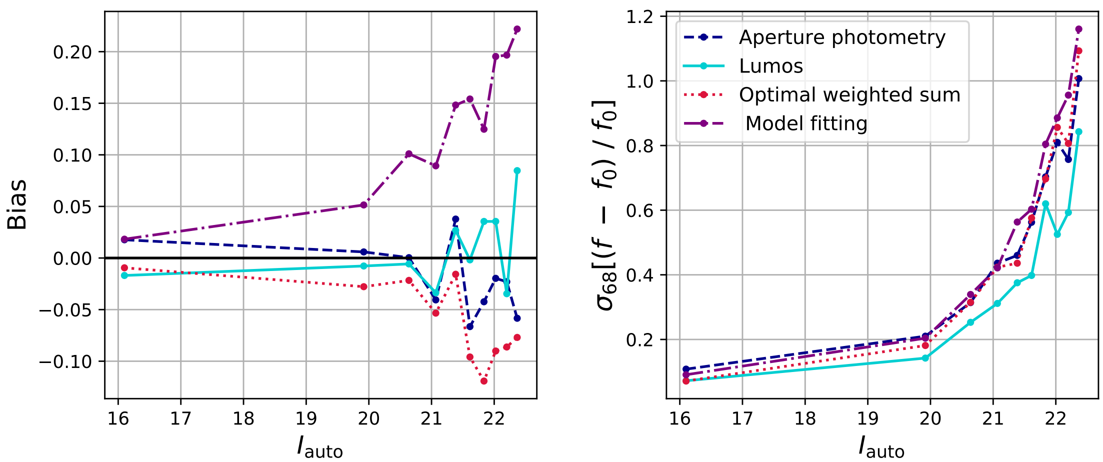

Figure 5.7: Comparison of the bias (left panel) and the dispersion (right panel) among the flux measurements with `Lumos`, aperture photometry, model fitting and optimal pixel weighting. These results include galaxy image simulations in the 40 PAUS narrowband filters. The I-band magnitude corresponds to the AUTO magnitude as measured by the HST-ACS on the COSMOS field.

lensing studies (Arcelin et al., 2021). Deblending will be a challenge for future ground-based photometric surveys such as LSST and *Euclid* (Laureijs et al., 2011) and it has recently been approached with deep learning techniques (Boucaud et al., 2020).

In this section, we test if `Lumos` can extract the galaxy photometry of blended the target galaxies. Even though `Lumos` is not explicitly trained to predict the photometry in the presence of other galaxies, the simulated galaxy images contain background cutouts from PAUCam images centred at random CCD positions (§5.2.2). Consequently, the training sample contains examples of blended galaxies. Machine learning algorithms 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 60×60 pixel realisations of the same target Teahupoo galaxy, located at the central pixel of the stamp. Each of these realisations also contains the same PAUS galaxy centred on a different pixel at a time. The PAUS galaxy moves across the image cutout in steps of one pixel, in such a way that it ends up covering all the pixels in the stamp. Therefore, the realisation where the PAUS galaxy is located at the central pixel corresponds to a total blending with the Teahupoo galaxy.

Figure 5.8 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 $i_{\rm AB} = 22$ and $i_{\rm AB} = 20$, respectively. With aperture photometry (dashed red line), the flux is considerably biased for all distances R. For $R < 15$, the bias measurement is caused by light from the overlapping source accounted inside the aperture. At larger R values, the background noise prediction is also affected by the overlapping source. In the PAUS aperture photometry pipeline, the background is estimated 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. 5.8 fluctuates around 2-10% for different distances R. Unlike aperture photometry, `Lumos` can distinguish between the two galaxies and consider the overlapping one a source of noise.

In Fig. 5.9, we explore 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 $i_{\mathrm{AB}} \approx 21$ galaxy at the centre (the same for the twenty realisations). In each of the cutouts, we have included a second source always located at five pixels from the centre, but which varies brightness among realisations. We have also applied a σ-clipping of the annulus to make a more robust background measurement. 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.

For the bottom plot, we have proceeded similarly. We have generated ten realisations of the same galaxy with magnitude $i_{\mathrm{AB}} \approx 21$ and in each of the cutouts, we have included a second source with $i_{\mathrm{AB}} \approx 22$, 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.

5.6 `Lumos` photometry on PAUS data

In this section, we present the `Lumos` photometry extracted from PAUCam images in the COSMOS field. First, we show single observation measurements (§5.6.1) and compare them to SDSS measurements (§5.6.2). We then discuss the co-added flux measurements (§5.6.3) and show the photometric redshift results with `Lumos` photometry (§5.6.4).

5.6.1 Single exposure measurements

PAUS has taken around 10 000 images in the COSMOS field (§3.4.4), which contain ten million galaxy observations. In this section we only show the results on the spectroscopic sample with $i_{\mathrm{AB}} < 22.5$, which are ≈ 3 million exposures from 15 000 galaxies. For the targeting, we use the Ilbert et al. (2008) catalogue, which is also the `MEMBA` detection catalogue (§ 3.4.2). Given a PAUCam image, `Lumos` matches it with the detection catalogue and creates the cutouts around the sources. The external catalogue also provides a value for the half-light radius, the ellipticity, and the Sérsic index, which are used to generate the modelled galaxy profiles (see §5.4.1).

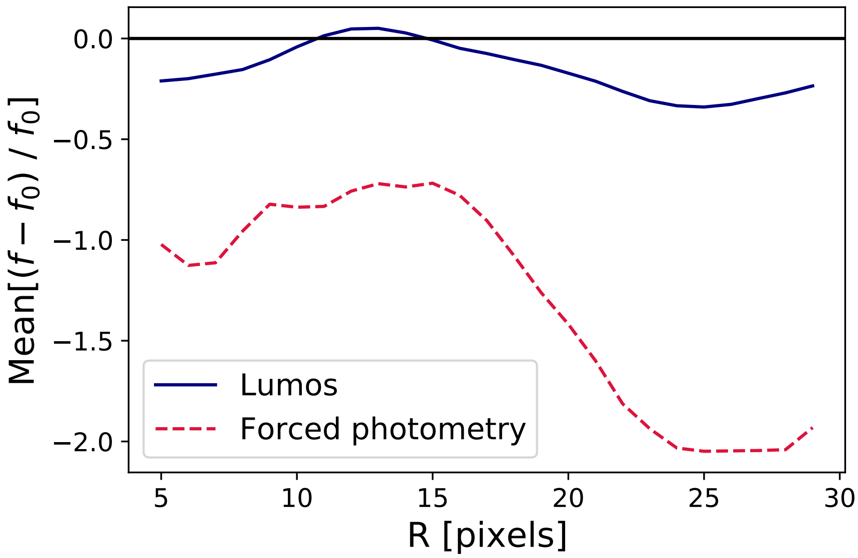

Figure 5.8: Accuracy in the flux measurement when the target galaxy ($i_{\mathrm{AB}=22}$) is blended with another source ($i_{\mathrm{AB}=20}$) as a function of the relative distance between TEAHUPOO target galaxy and the 'blending' PAUS galaxy (R). The solid blue line corresponds to `Lumos`, while the red dashed line is forced photometry.

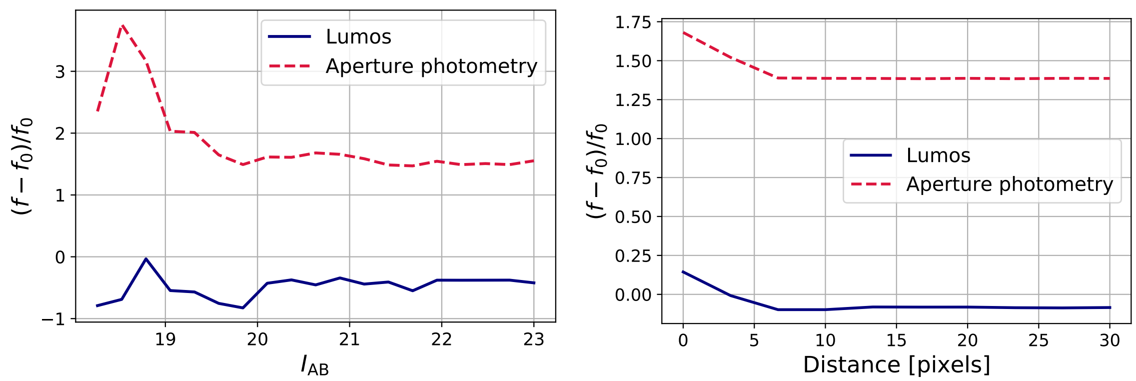

Figure 5.9: Accuracy in the flux predictions in the presence of overlapping sources for `Lumos` (solid blue line) and aperture photometry (dashed red line) as a function of *Top*: magnitude of the overlapping source. *Bottom* Distance in pixels between the target and the overlapping source.

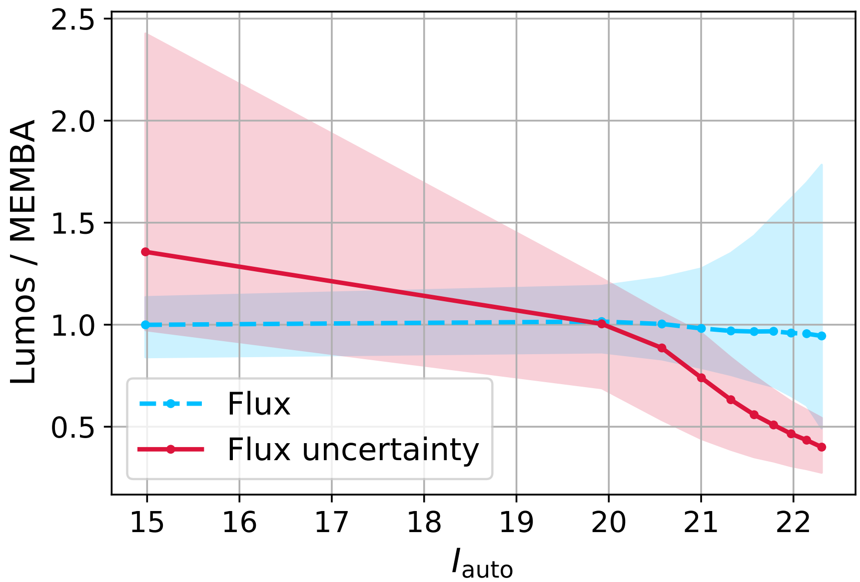

Figure 5.10: Flux and flux uncertainty ratios between Lumos and MEMBA photometry in equally populated magnitude bins. The shaded areas correspond to the 16th and 84th quantiles.

Flux and flux error measurements

In this section, we use flux and flux error point estimates calculated as the mean and the variance of the Lumos flux PDFs (§ 5.5.1 and Eqs. 3.4&3.5). For MEMBA, we use flux measurements from aperture photometry with the background subtraction from BKGnet, which has proven more accurate than that estimated with an annulus. Figure 5.10 shows the flux (dashed blue line) and flux error (solid red line) ratios between Lumos and MEMBA photometry in all narrow-band 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 $\approx 4\%$ less flux in the brightest ($i_{\rm AB} < 18$) and in the faintest ($i_{\rm AB} > 22$) bins. At the faintest end, the spread in the flux ratio increases, which is natural since these galaxies are noisier. Studying each narrow-band 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 ≈ 0.9 ratio between MEMBA and Lumos. We attribute this difference 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 photo-zs, therefore measuring similar fluxes with MEMBA and Lumos is a good first test.

Altogether, Lumos also provides 40% lower flux uncertainties than MEMBA displaying a lower error for 85% of the photometry measurements. The ratio between Lumos and MEMBA flux uncertainties (Fig 5.10) is not constant with magnitude. Lumos shows 60% lower errors for objects with $i_{\rm AB} > 22$. This number monotonically decreases to e.g. 30% at $i_{\rm AB} = 21$ and 10% at $i_{\rm AB} = 20.5$, while some of the brightest objects display lower errors with MEMBA. At the brightest end, the uncertainty ratio between Lumos and MEMBA has a large scatter. This is attributed to some bright galaxies with significantly large uncertainty in the Lumos measurement. While aperture photometry provides a purely statistical error, in `Lumos`, any additional source of error is also accounted for in the uncertainty estimate. For example, `Lumos` can capture inaccuracies in the profile parameters used to infer the photometry (e.g. $n_{\rm s}$ and r_{50}), as it sees both the galaxy image and the modelled galaxy. Therefore, the network can capture discrepancies or potential sources of inaccuracies and account for them in the photometry 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 large spread in the uncertainty ratio at the brightest end.

Colour histograms

Assuming that galaxies have an underlying distribution of colours, the width of the colour histograms estimates the uncertainty in the photometry measurements. Photometry uncertainties broaden the intrinsic width of the colour histogram. Consequently, the photometry providing narrower colour histograms is that with the lowest uncertainties. Using colour histograms to compare photometries was used in Wright et al. (2016), where they presented and implemented `LAMBDAR` to improve the Galaxy and Mass Assembly (GAMA, Driver et al., 2011) photometry .

Figure 5.11 shows the "NB785"-"NB795" colour distribution (more colour histograms can be found in Appendix 5.D). 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 σ_{68} and σ_{95} [3]. Concretely, `MEMBA` provides $\sigma_{68} = 0.26$, while `Lumos` results in $\sigma_{68} = 0.19$, which corresponds to a 30% lower effective width. Considering σ_{95}, `Lumos` reduces the width a factor of $\approx$3, from 0.74 to 0.41. This indicates that `Lumos` reduced the number of photometry outliers, which are not affecting σ_{68} but enlarge σ_{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. 5.22.

The other narrow bands also show narrower colour histograms with `Lumos` (Appendix 5.D, Fig. 5.11). Furthermore, the relative difference in σ_{95} is systematically higher than with σ_{68}. This is likely related with exposures with noisy photometry and outliers, which lay in the tails of the colour histograms (see Appendix 5.D for more details).

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 narrow-band filter.

[3] σ_{95} is equivalent to σ_{68} but considering the 2.5 and 97.5 quantiles, i.e. the width accounts for 95% of the data.

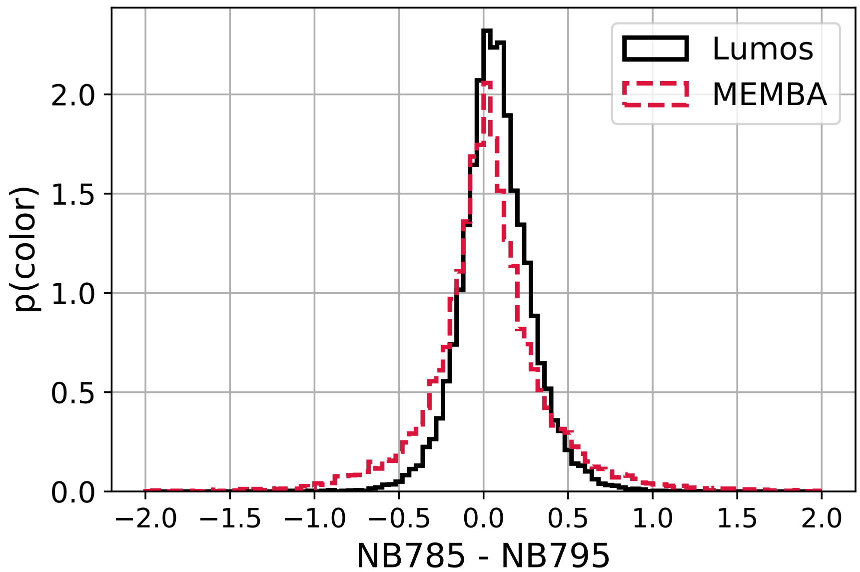

Figure 5.11: `MEMBA` (red dashed line) and `Lumos` (solid black line) colour histogram for narrowband colour (NB785-NB795). Note that the photometry with lower uncertainties is that displaying a narrower colour histogram, which in this case is the `Lumos` photometry

.

This is the same technique as used in the analysis of `BKGnet` uncertainties (§,4.5). Two observations of the same galaxy in the same narrow band must have compatible flux measurements. Therefore, the distribution

$$D \equiv \frac{(f_1 - f_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)}} \tag{5.14}$$

must be a Gaussian with zero mean and unit variance, where f_1, f_2 are the flux estimates of two independent observations of the same object and σ_1, σ_2 are their associated uncertainties.

Figure 5.12 shows the width of the D distribution in equally populated magnitude bins. To be less affected by outliers, we have estimated the width of D with σ_{68} and both `Lumos` and `MEMBA` display a quite constant unity $\sigma_{68}[D]$ across 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% underestimated errors at the bright end. This trend was fixed using `BKGnet` (see Fig. 4.12).

Figure 5.13 shows the distribution of the quantity defined in Eq. 5.14 with `Lumos` (solid black) and `MEMBA` (dashed red) photometries. As expected, both of them fit a Gaussian with zero mean and unit variance, however, the `MEMBA` photometry displays a tail of outliers not present in the `Lumos` distribution. 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% underestimated r_{50} for a particular galaxy, with aperture photometry

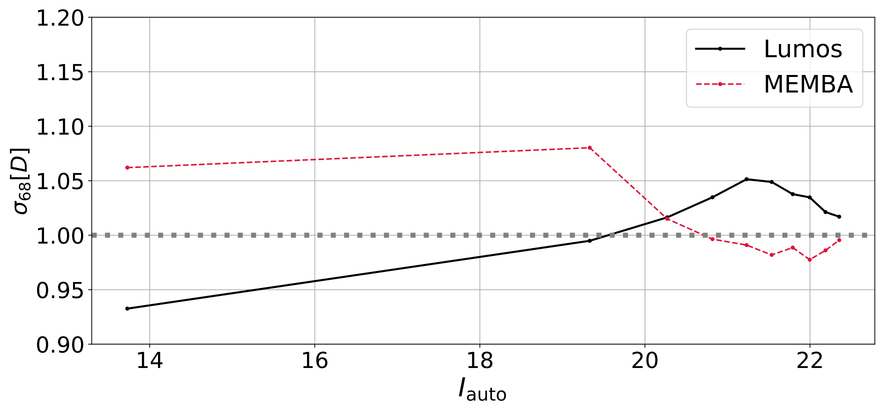

Figure 5.12: The width of the distribution in Eq. 5.14 for `Lumos` and `MEMBA` flux predictions in equally populated magnitude bins. Robust uncertainties must provide a unity width (marked by the thick grey dotted line).

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, and 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.

Figure 5.14 compares the median signal-to-noise per narrow band in `MEMBA` and `Lumos` photometries for galaxies with $i_{\mathrm{AB}} < 22.5$. The shaded areas correspond to the 16th and 84th quantiles of the signal-to-noise distribution. For the complete photometry catalogue, `Lumos` provides, on average, a 54% higher signal-to-noise. Furthermore, it gives a higher median signal-to-noise at all wavelengths, although the increment with respect to `MEMBA` is higher in bluer bands. For galaxies with $i_{\mathrm{AB}} > 22$, the signal-to-noise is 2.5 times higher in `Lumos`. The ratio increases to 3 taking into account only the bluest narrow band ("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 objects. Altogether, $\approx 85\%$ of the observations have higher SNR with `Lumos` photometry.

Observation's flagging

The `MEMBA` pipeline already provides an outlier flag for its measurements (§ 3.4.2). This is a discrete value that flags objects with problematic image reductions, e.g. saturated pixels, crosstalk, cosmetics, distortion, and undesirable artifacts near the target source such as scattered-light, cosmic rays and blending. `Lumos` uses the reduced PAUCam images (§ 3.4.1), which are affected by all these effects. However, as we already showed in § 4.5 and earlier in this paper (§5.5.3), `Lumos` deals with recurrent problems as scattered-light or blending. Figure 5.15 shows examples of scattered-light (left panel) and cosmic ray (right 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,

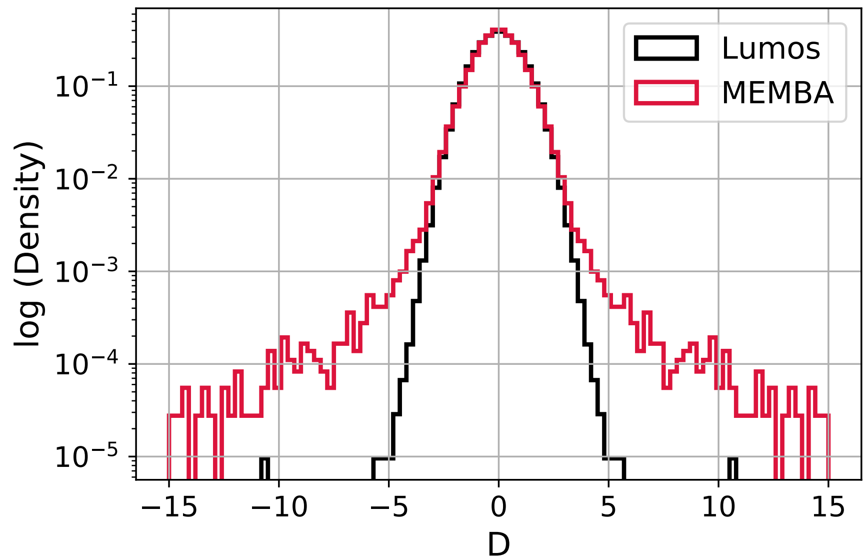

Figure 5.13: Distribution of D (Eq. 5.14) estimated with `Lumos` (solid black) and `MEMBA` (dashed red) photometries in logarithmic scale.

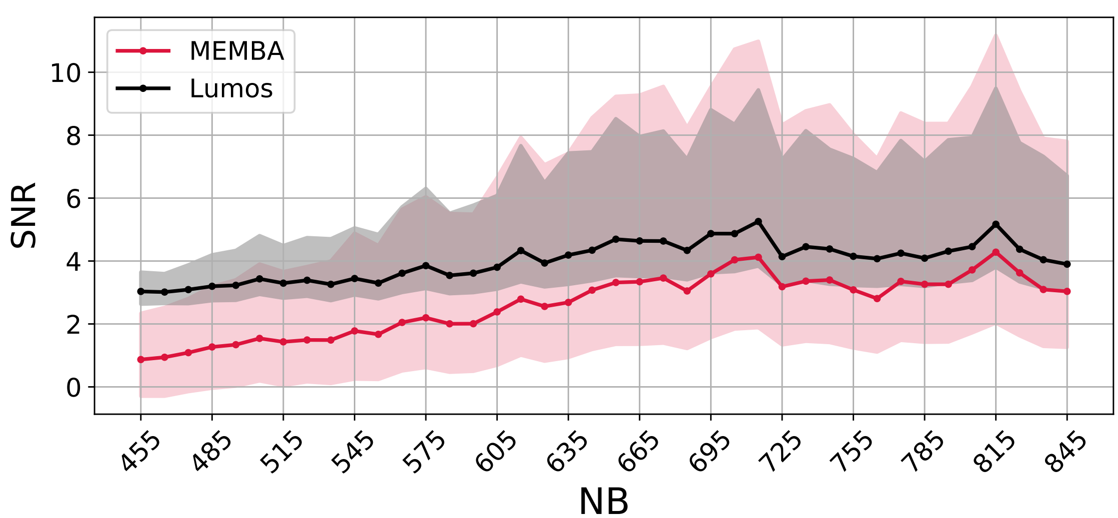

Figure 5.14: Median SNR per narrow band filter with `Lumos` and `MEMBA` flux measurements. Shaded areas are the 16-th and 84-th quantiles of the signal-to-noise (SNR) distribution.

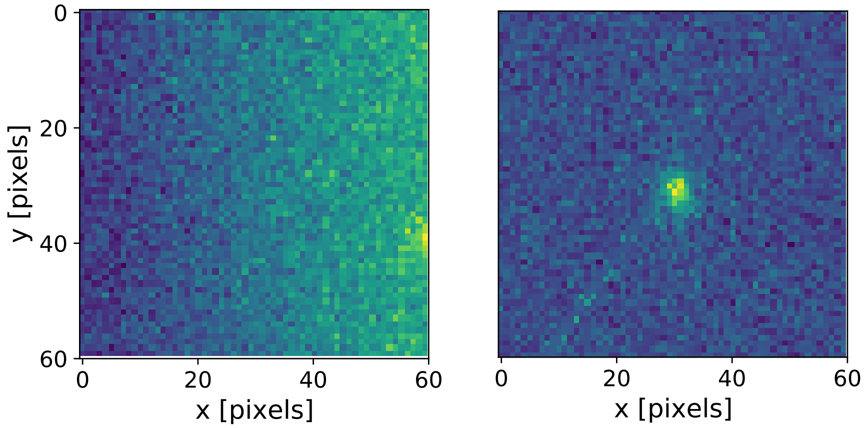

Figure 5.15: Observations affected by scattered-light (left panel) and cosmic rays (right panel). While `MEMBA` provides outlier flux measurements for both observations, `Lumos` estimates a flux close to that measured in other exposures of the same object.

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% of the observations are flagged. Within these flagged objects, 72% are observations affected by scattered light, 18% have image distortion effects, and the rest are other minority effects such as crosstalk, cosmic rays, and cosmetics. `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%. This reduction highlights that `Lumos` is more robust towards outliers in the photometry. This is particularly interesting since the network is not explicitly trained to deal with artefacts like cosmic rays and crosstalk signals. However, by using 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.

5.6.2 Comparison with SDSS spectroscopy

We have compared the `Lumos` flux measurements with synthetic PAUS photometry to have another validation of the flux estimates. The synthetic PAUS photometry convolves SDSS galaxy spectra with the PAUCam filter throughput. Unfortunately, the synthetic PAUS data corresponds to a bright sample with a magnitude limit $i_{\rm AB} < 20.5$, 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 minimising the χ^2 between

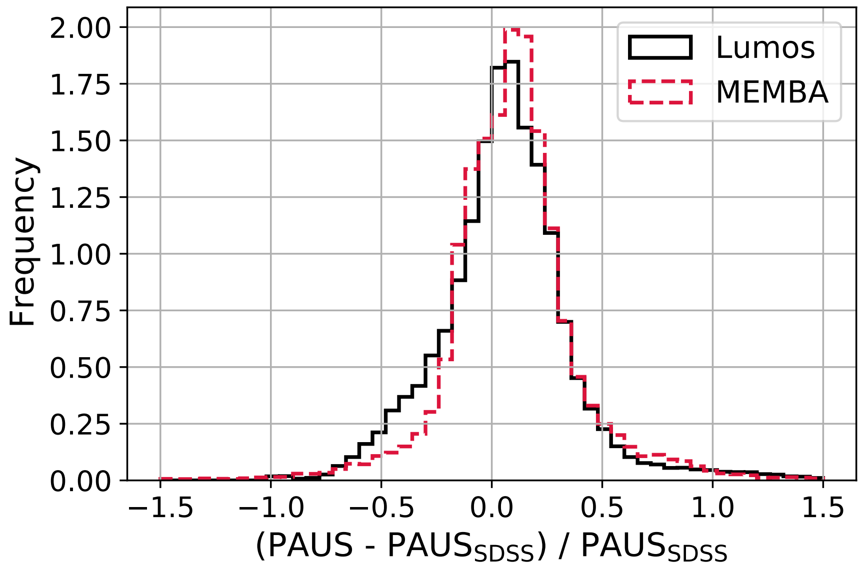

Figure 5.16: Comparison between PAUS flux measurements and SDSS measurements convolved with PAUCam filters (PAUS synthetic fluxes). The solid black line corresponds to `Lumos` measurements and the red dashed line to `MEMBA`.

PAUS observations and PAUS synthetic fluxes, i.e.

$$\chi^2 = \sum_i \frac{(f_{\mathrm{SDSS},i} - zp \cdot f_{\mathrm{SDSS_{PAUS}},\mathrm{i}})^2}{\sigma^2_{\mathrm{SDSS},i} + zp^2 \cdot \sigma^2_{\mathrm{SDSS_{PAUS}},i}}, \tag{5.15}$$

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

Figure 5.16 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. Nonetheless, `Lumos` shows a 5% bias while in `MEMBA` this goes to 10%. Furthermore, the number of observations at more than 5σ from the mean of the distribution reduces by 2 with `Lumos`, going from 3% to 1.5%.

5.6.3 Coadded flux measurements

The co-added flux measurements are constructed combining individual observations of the same galaxy in the same narrow band (§ 3.3.1). Co-adding exposures increases the signal-to-noise of the galaxy photometry and it is very helpful to reject wrong observations. Before co-adding individual observations, a zero-point calibration per image is required, which in our case is done relative to SDSS (§3.2.1). PAUS implements flux measurements co-addition (f_{coadd}) as a weighted sum of the individual observations

$$f_{\mathrm{coadd}} = \frac{\sum_i f_{\mathrm{i}}/\sigma_i^2}{\sum_i 1/\sigma_i^2}, \tag{5.16}$$

where the weights are the inverse variance of the observations and f_i and σ_i^2 are the flux measurement and its variance of the ith observation, respectively.

Lumos calibrates the Gaussian components individually with the photometric zero-point in such a way that when the these are combined, they already provide a calibrated PDF for the flux observation. With Lumos, combining point like estimates with Eq. 5.16 is still possible. Nevertheless, it can also generate co-added measurements combining the probability distribution of the individual galaxy observations. Figure 5.17 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 enables the calculation of the flux and flux uncertainty 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, σ_{68} and σ_{95}. From the two latest quantities, we construct a measurement of the PDFs gaussianity,

$$\eta \equiv \sigma_{95}/\sigma_{68} - 1\,, \tag{5.17}$$

which can be helpful to decide which are the best flux and flux error estimators. While for a sharp and peaked PDF ($\eta \gtrapprox 1$), the peak or the median would provide similar flux estimates, in non-gaussian PDFs as e.g. the galaxy in Fig. 5.17 these two estimators would give significantly different measurements. The PDF gaussianity also affects the flux uncertainty estimators. Broadly, galaxies with $\eta > 1$ have $\sigma_{68} < \sigma_{\rm std}$, while this is the opposite for galaxies with $\eta < 1$. The galaxy in Fig. 5.17 ($\eta = 0.56$) is an example f multiple peaked PDF where $\sigma_{68} > \sigma_{\rm std}$.

We have tested applying different flux and flux uncertainty estimators based on the η 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 σ_{68}, $\sigma_{\rm std}$ or $\sigma_{95}/2$ does not lead to a significant difference.

5.6.4 Photometric redshift estimates

Accurate photo-z estimates are crucial for many science applications. Improving the photometry signal-to-noise is expected to improve the photo-z estimates. In this section, we have tested the Lumos photometry by predicting the PAUS photo-z (§3.4.3) with BCNz2 (Eriksen et al., 2019) and Deepz (Eriksen et al., 2020). Testing the photo-zs has also been particularly helpful to find and fix some issues in the Lumos photometry missed with other validation tests. For example, several galaxies were exhibiting oscillating photometry. This kind of object was detected as a photo-z outlier, and we could trace that galaxy images with sub-pixel shifts with respect to the centre of the stamp triggered the oscillating pattern.

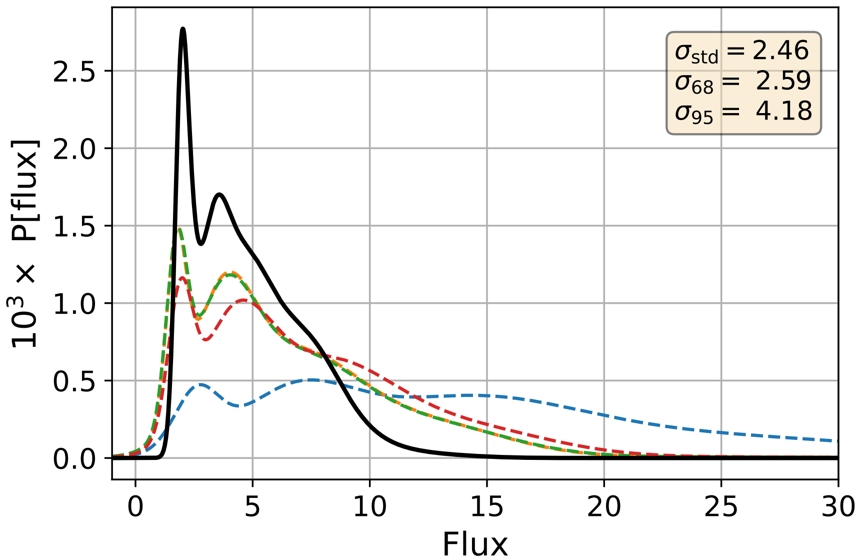

Figure 5.17: The co-added flux probability distribution (black solid line) constructed from its individual observations (colored dashed lines). The upper box displays the $\sigma_{\rm std}$, σ_{68} and σ_{95} of the co-added PDF.

Fig. 5.18 shows the photo-z dispersion with `BCNz2` or `Deepz` using `Lumos` photometry to $i_{\rm AB} < 22.5$. 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% for galaxies with $i_{\rm AB} > 20.5$. However, the right 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-zs with `Lumos` photometry are statistically better, however there are some high redshfit outliers that increase σ_{68}. Rejecting galaxies with spectroscopic redshift ($z_{\rm s}$) $z_{\rm s} > 0.8$, the faintest galaxies ($i_{\rm AB} > 22$) have a 14% lower photo-z dispersion with `Lumos` than with `MEMBA` photometry.

Photo-zs with `Deepz` are not showing this degradation at high redshift. For galaxies with $i_{\rm AB} > 20.5$, the `Deepz` photo-zs are between 10% and 20% more precise with `Lumos` photometry. Furthermore, at $i_{\rm AB} > 22$ and without any redshift cut, photo-zs are 15% 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, this effect only affects a small fraction of the brightest galaxies in the sample, which moreover are not those we are most interested in.

`Lumos` photometry reduces the outlier rate with both `BCNz2` and `Deepz`. With `BCNz2`, the

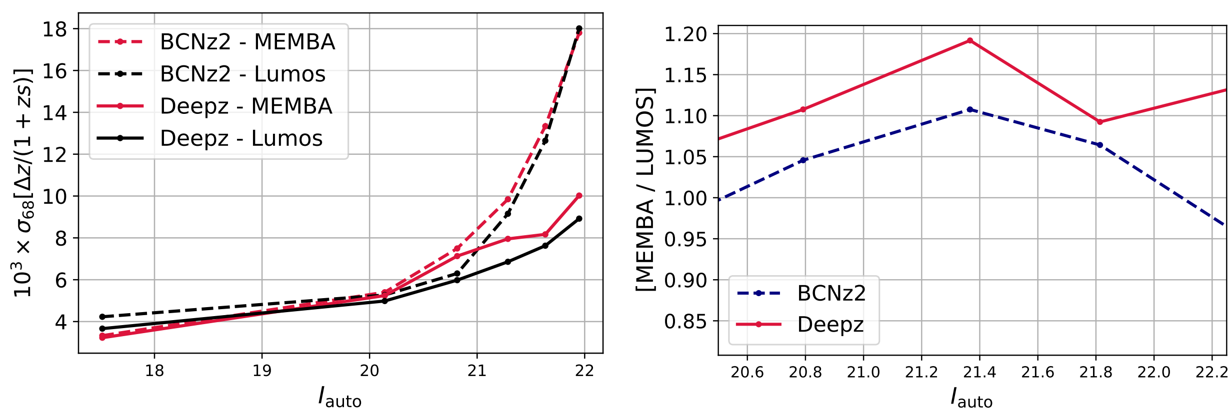

Figure 5.18: *Left:* Photo-z precision with `BCNz2` and `Deepz` using `Lumos` or `MEMBA` photometry. *Right*: Relative difference between photo-zs with `Lumos` or `MEMBA` photometry.

outlier rate is reduced by 5% in the complete catalogue. With `Deepz`, the improvement is greater, with 20% less outliers. This number increases to 23% for objects with $i_{\rm AB} > 22$.

To the best of our knowledge, there is not any photo-z code that could deal with flux PDFs, and both `BCNz2` and `Deepz` require point estimates for the flux and its uncertainty. Here, these quantities are the peak and σ_{68} of the co-added flux PDF. We have also tested other estimators, e.g. the median, the standard deviation, or choosing different estimators based on the η parameters (Eq. 5.17). However, the peak and σ_{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 signal-to-noise. In Appendix 5.C, we have used PAUS simulated mocks to test `BCNz2` performance with signal-to-noise and its behaviour with artificially injected issues in the sample photometry. The results suggest that the photo-z improvement should be ≈90% greater than what we see in 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.

5.7 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 (§ 5.4.2). `Lumos` is the evolution of `BKGnet` (§ 4.3), a deep learning method that predicts the background light of astronomical images with strongly varying noise patterns (§ 4.2). 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 (§ 5.2.2). TEAHUPOO galaxy images use PAUCam image cutouts for the background noise. Astronomical images contain distorting effects, e.g. scattered light and crosstalk, triggering inaccuracies in the photometry. Including PAUCam cutouts in the 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 observations affected by scattered light, cosmic rays, and other contaminating effects that require flagging with aperture photometry (§ 3.4.2).

Furthermore, we have tested the Lumos deblending capability on simulations (§5.5.3). Lumos can extract the target-galaxy photometry much better than aperture photometry without explicitly including blended galaxies in the training sample (Fig 5.8). While aperture photometry provides a catastrophic flux measurement for blended sources, Lumos can provide a flux with 2-10% accuracy, depending on the distance in pixels to the overlapping source. This is particularly interesting since Lumos is not developed 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 a MDN (Fig. 5.4). While most photometry algorithms provide a flux value and the 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 (§5.6.3). 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 a 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% from the baseline aperture photometry measurements (§5.6). Concerning uncertainties, our photometry errors are 40% lower than with aperture photometry. This translates into between 1.5 and 3 times higher signal-to-noise in Lumos than in MEMBA, with the largest improvement at the faint end (Fig. 5.14). We have run the BCNz2 and Deepz codes with Lumos photometry (§ 5.6.4), resulting in a reduction of the photo-z scatter with both (Fig 5.18). The photo-z improvement using Lumos photometry is greater with Deepz rather than with BCNz2. Deepz with Lumos photometry reduces the scatter by 10% on the full catalogue and 13% for galaxies with $i_{\mathrm{AB}} > 22$. Furthermore, it also reduces the outlier rate (Eq. 3.23) by $\approx$ 20%. Nevertheless, Appendix 5.C 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 and 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 LSST will observe much deeper galaxies with very low signal-to-noise, where `Lumos` could be a helpful tool to improve the photometry. Furthermore, `Lumos` is 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. This would require training `Lumos` with simulated galaxies mimicking the targeted survey. Nevertheless, one potential difficulty of the method applied to deeper surveys is the modelled galaxy profile input. While there exist previous deeper observations of PAUS galaxies, deeper surveys like LSST will observe galaxies for which there is no previous knowledge. Not using the modelled profile in the training barely affects the overall predicted flux measurements, however, this degrades the signal-to-noise by 15%.

`Lumos` supersedes `BKGnet` and provides background-subtracted flux measurements, which requires a measurement 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 an end-to-end pipeline that supersedes `Lumos`, providing multi-band galaxy photometry and the photometric redshift (§ 7).

5.A Flux estimation methods: derivations

This appendix derives the linear combination of pixel values giving and unbiased and optimal flux measurement (Eq. 5.4). The SNR of the measurement when combining pixels with mean m and weight w is

$$\mathrm{SNR} = \frac{\sum_i w_i m_i}{\sqrt{\sum_i w_i^2 (m_i + b_i)}}, \tag{5.18}$$

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

$$w_x = \lambda \frac{m_x}{(m_x + b_x)} \tag{5.19}$$

where λ is a constant. The flux measurement being unbiased means

$$\sum_i m_i = \sum_i w_i m_i. \tag{5.20}$$

Using this requirement, the pixel weights (Eq. 5.19) becomes

$$w_x = \frac{\sum_i m_i}{\sum_i m_i^2/(m_i + b_i)} \frac{1}{1 + b_x/m_x} \tag{5.21}$$

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.

5.B Forecasting the effect of errors on profile parameters

The algorithms described in §5.3 require information about the galaxy profile, making the flux measurement accuracy sensitive to errors on the profile parameters. In this section, we use Teahupoo galaxies to we quantify the effect that errors in the input profile parameters have on the flux measurements using a Fisher forecast formalism (Fisher, 1922).

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

$$\tilde{f} = \Phi(I, f, r_{50}, n_{\mathrm{s}}, PSF, e, b) \tag{5.22}$$

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). For instance, the aperture algorithm uses these quantities to scale the apertures and the profile-fitting method uses them to construct the galaxy model. We can estimate the propagation of these errors to the flux with a Fisher matrix formalism. The Fisher matrix is

$$\mathbf{FM}_{\mu\nu} = \frac{\partial \tilde{f}}{\partial \mu} \left(\sigma_{\tilde{f}}^{-2} \right) \frac{\partial \tilde{f}}{\partial \nu}, \tag{5.23}$$

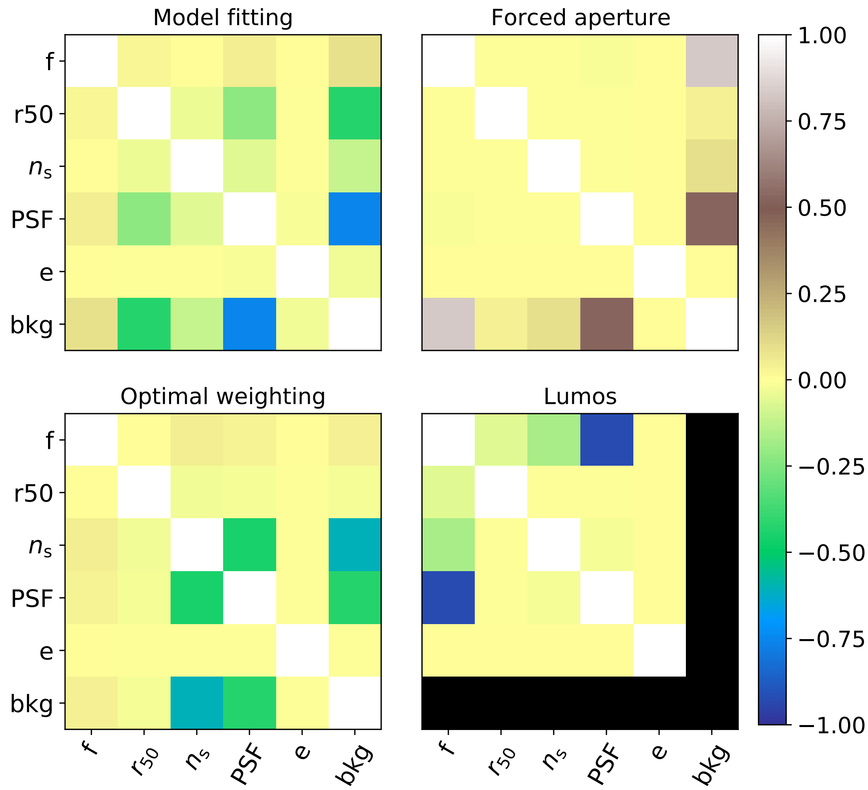

Figure 5.19: The correlation matrix for the parameters $f, r_{50}, n_{\rm s}, PSF, e$ and bkg for: *Top left:* The model fitting method, *Top right:* Forced aperture photometry, *Bottom left:* Optimal weighted pixel sum and *Bottom right:* `Lumos`.

where the indices μ and ν are the galaxy parameters the total flux depend on (see Eq. 5.22).

The covariance matrix of the flux measurements is the inverse of the Fisher matrix. Figure 5.19 shows the correlation matrices of the parameters f, r_{50}, $n_{\rm s}$, PSF, e, and b for the four flux estimation methods described in §5.3. The correlation matrix differs for different galaxy types (e.g. different morphology and brightness). We have constructed a common galaxy with r_{50}, $n_{\rm s}$, and PSF assigned to the mean of their distribution in the COSMOS field galaxies. The model-fitting method (top left panel) and the optimal weighting (bottom left panel) are those showing more correlation between the flux and the profile parameters. In contrast, the forced-aperture photometry and `Lumos` show a lower correlation, i.e. these methods are more robust since the effect of errors in the galaxy 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.

For the next test, we have assumed a 10% prior error in each of the input parameters, such that

$$\mathbf{FM}_{\rm comb} = \mathbf{FM} + \mathbf{FM}_{\rm priors}\,. \tag{5.24}$$

	r_{50}	$n_{\rm s}$	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	-

Table 5.1: Percentage of error in r_{50}, $n_{\rm s}$, 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.

$\mathbf{FM}_{\rm priors}$ is a diagonal matrix including the inverse prior variance of each parameter. The variance on the flux parameter is then

$$\sigma_f^2 = \left(\mathbf{FM}_{\rm comb}^{-1}\right)_{ff} \tag{5.25}$$

where the matrix subscripts ff denote selecting the row/column corresponding to the flux parameter.

Table 5.1 shows the percentage of error in the parameters that propagates to a 10% 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 amongst galaxy types, the results in Tab. 5.1 are an average of a hundred independent random galaxies. `Lumos` is the most robust method as it requires higher errors on the galaxy parameters to propagate to a 10% flux error. As expected from Fig. 5.19, the PSF is the parameter `Lumos` is more sensitive to, followed by the half-light radius. However, it is still less sensitive than the other methodologies in both cases.

`Lumos` uses the galaxy and the modelled galaxy images, which enables the comparison and detection of problematic profiles instead of blindly relying on the input parameters. Furthermore, errors in such parameters are more subtle and difficult to detect when these are encoded in a modelled profile rather than directly inputted into the method.

5.C Photometric redshifts with `BCNz2` on PAUS galaxy mocks

In this section, we run `BCNz2` on the PAUS galaxy mock to study 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 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 a SED are assigned to each of them. The SED templates also take into account the emission lines. H_α is computed form 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.

For this test, we use PAUS narrow bands, the CFHT u-band, and the Subaru $BVriz$ broad bands. We randomly select 10 000 noiseless galaxies and include Gaussian uncertainties to mimic the `Lumos` signal-to-noise (Fig. 5.20). For the broad bands, we generate uncertainties to match the signal-to-noise in Eriksen et al. (2019).

Figure 5.21 shows that, on average, the photo-zs on the PAUS-mock (black solid line) have 90% less scatter than on data (blue solid line). This is a large number considering that the signal-to-noise on both samples is similar. Consequently, the photo-zs appear limited by other factors than the photometry signal-to-noise.

The purple solid line shows the effect of the photometric calibration. For each flux measurement, we generate five individual observations scattering from a Gaussian centred at the co-added flux. Then, we assign a zero-point and a zero-point uncertainty from the distribution of zero-points in the PAUS COSMOS data to each observation. The individual observations are co-added to a single flux error using the scattered zero-points.

The median zero-point uncertainty in the PAUS data in COSMOS is $\approx$ 4%. Including this effect in the photo-zs (purple solid line in Fig. 5.21) degrades the photo-z precision by 40% with respect to perfect photometry (black solid line), particularly at the fainter end. Nevertheless, the calibration cannot fully explain the difference between the photo-z precision expected on simulations and that obtained on the data, since the photo-z precision is still significantly better on the PAUS mock compared to the results on data.

The coloured dotted lines in Fig. 5.21 shows the photo-z dispersion with additional effects in the photometry that could potentially lead to a photo-z degradation. All the dotted lines also incorporate the calibration effect simulated in the previous paragraph. In the green dotted, we have included an additional 20% error in 20% of the photometric zero points. This additional error is not accounted for in the final photometric error, thus it could potentially make an outlier from a correct photometry measurement. This error on the zero-points especially affects brighter galaxies and reduces the photo-z precision to $\sigma_{68} = 0.0050$.

In the orange dotted line, we have artificially injected 1.5% of outliers in the PAUS mock fluxes. These outliers directly affect the co-added flux measurement and therefore, a 1.5% of affected fluxes corresponds to a higher percentage of affected galaxies. Particularly, 45% of the galaxies in the PAUS mock have at least one affected NB flux measurement and $\approx$ 10% 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 signal-to-noise 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% was too few to explain the degradation of data and 2% was too much. With 1.5% of outliers, the photo-z

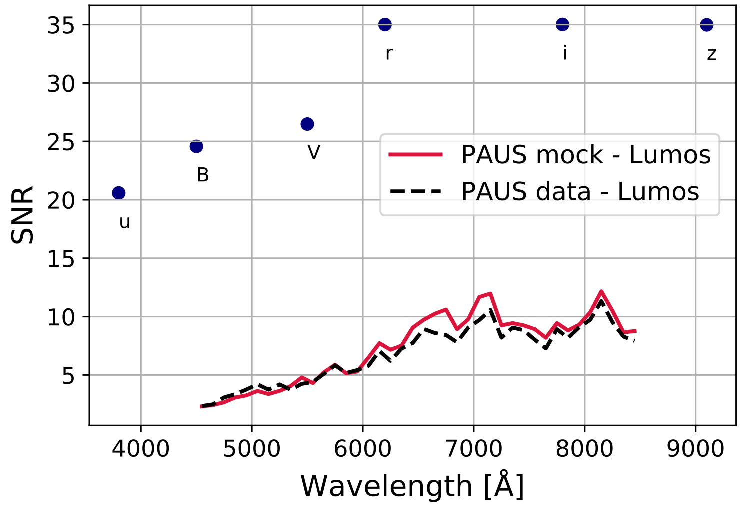

Figure 5.20: The SNR of the PAUS galaxy simulations used to run `BCNz2`. The solid red line corresponds to a PAUS mock with a SNR similar to that provided by `Lumos`. As a comparison, the dashed black line corresponds to the observed SNR on PAUS data with `Lumos` photometry.

precision degrades by 80%, 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.

5.D Colour histograms in the complete narrow band set

Colour histograms can be used to compare different photometries. Assuming an underlying galactic colour distribution, photometry uncertainties broaden such distribution. Consequently, the best photometry on a sample of galaxies is that providing narrower colour distributions. In Fig. 5.22 (§5.6.1), 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 narrow bands. Figure 5.22 shows the colour histogram of nine different narrow bands with the `Lumos` photometry (black solid line) and the `MEMBA` photometry (red dashed line).

Furthermore, Fig. 5.23 shows the relative difference in the effective width of the colour histograms with the photometries from `Lumos` and `MEMBA`. As in § 5.6.1, the effective widths are estimated with σ_{68} and σ_{95}. `Lumos` provides narrower colour histograms in all the narrow-band filters. The relative difference in σ_{68} oscillates between 30% and 40% in all narrow bands but the bluest, where it is ≈15%. With σ_{95}, the relative effective width is lower at the first eight bluer bands (≈30%) and increases to ≈70% for the rest of the bands. The relative difference in σ_{95} is systematically larger than with σ_{68}, which is likely related with very noisy measurements. Photometry measurements in the tails of the colour histograms will not affect the σ_{68} measurement, however these will be accounted in σ_{95}. Consequently, Fig. 5.23 would be showing that the number of outlier observations is lower in `Lumos` than in `MEMBA`.

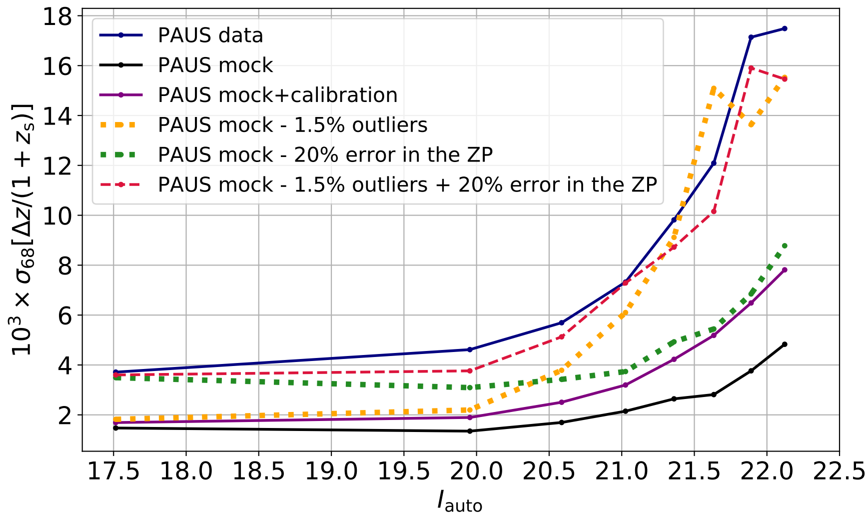

Figure 5.21: Photo-z dispersion as a function of i-band magnitude using `BCNz2` for a galaxy mock with `Lumos` SNR (black solid line) and PAUS data with `Lumos` photometry (blue solid line). The purple line includes the photometric calibration on the PAUS mocks. Dotted lines include outliers and calibration errors in the PAUS mock photometry. The red dashed line combines the effect of the two dotted lines.

5.E Photometry and photo-z correlations with galaxy parameters

Figure 5.7 only shows 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 photometry (§5.E.1) and the photo-z (§5.E.2) predictions as a function of other galaxy parameters as e.g. the galaxy size or ellipticity.

5.E.1 `Lumos` photometry correlation with galaxy parameters

Figure 5.24 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, all 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 dataset tend to have slightly underestimated flux predictions ($\approx$ 1-2%). This could be a consequence of the fixed stamp size, which could lead to a small leak of light.

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. We do not see any correlation between the photometry precision and the galaxy ellipticity.

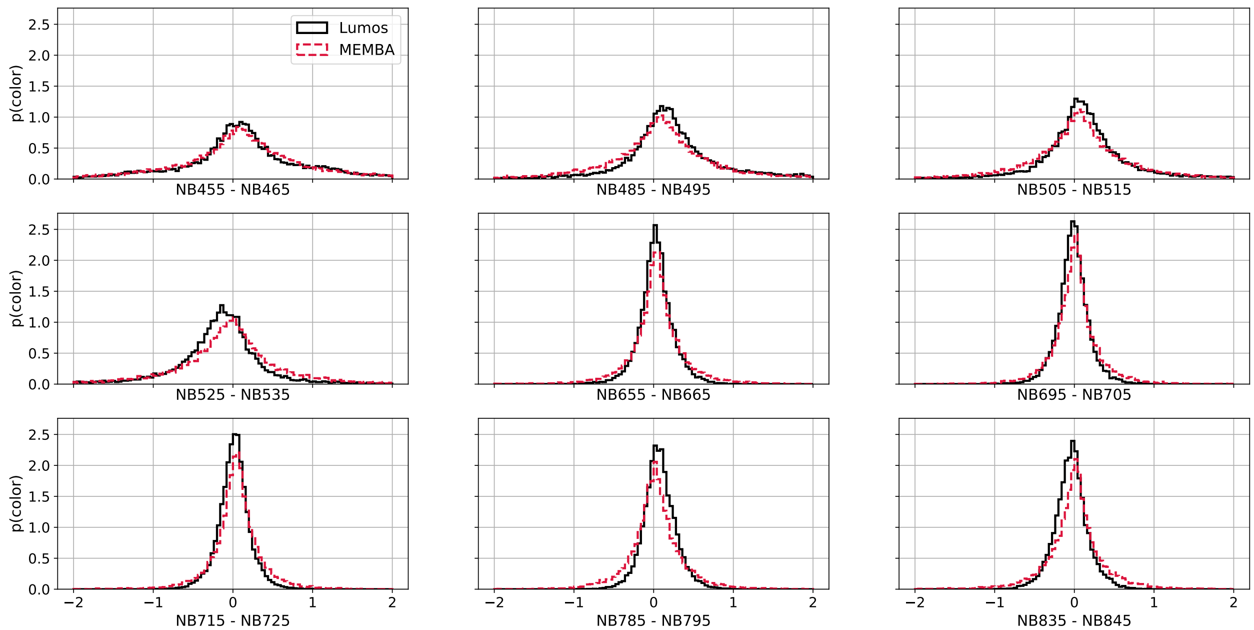

Figure 5.22: Colour histograms for nine different narrow band filters using `Lumos` (solid black line) or `MEMBA` (dashed red line) photometries.

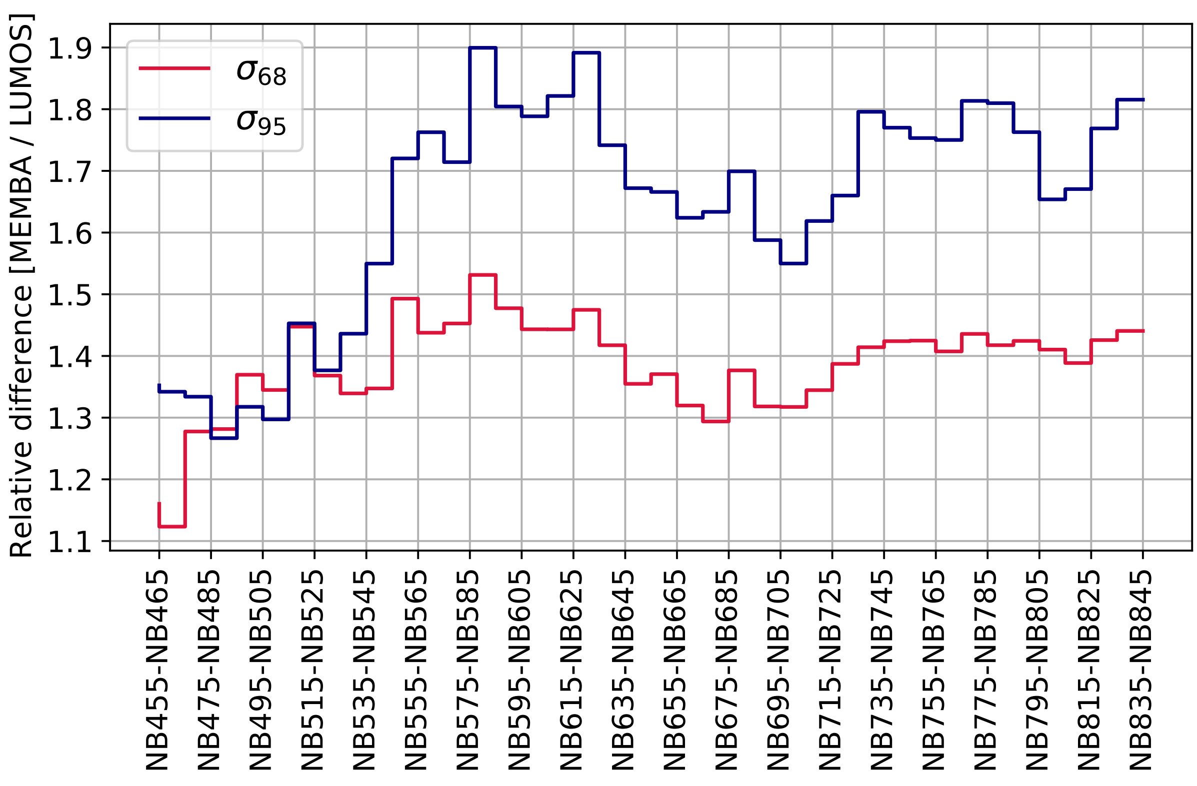

Figure 5.23: Relative difference between the effective width of the colour histograms of the PAUS photometry with `Lumos` and `MEMBA`. The effective width has been estimated with σ_{68} (blue line) and σ_{95} (red line).

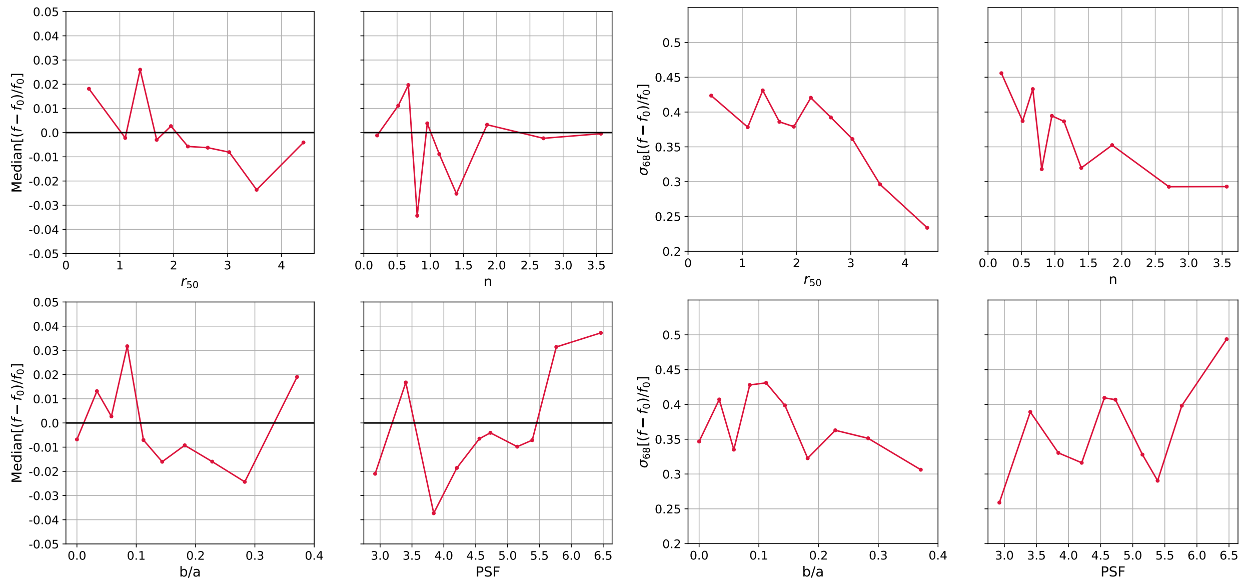

Figure 5.24: *Left: Bias in the photometry measurements as a function of the galaxy size* (r_{50}), galaxy shape (Sérsic index), ellipticity (b/a) and PSF. *Right*: Precision in the photometry as a function of the same parameters.

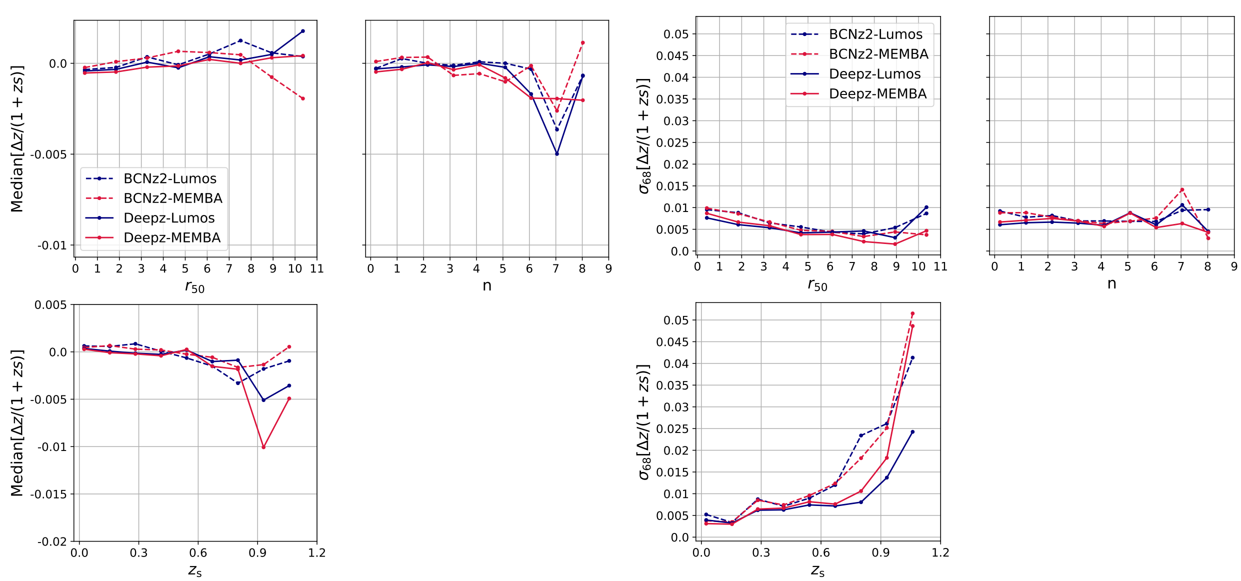

Figure 5.25: *Left*: Bias in the photo-z measurements as a function of the galaxy size (r_{50}), galaxy shape (Sérsic index) and the spectroscopic redshift (z_z). *Right*: Precision in the photo-z as a function of the same parameters.

5.E.2 Photo-z correlation with galaxy parameters

Figure 5.25 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-zs are also unbiased as a function of spectroscopic redshift, only presenting a $\approx$1% 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% 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 forth 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-zs for larger galaxies. This is potentially related to discrepancies between the training image simulations and the data (see §5.6.4). 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-zs for $z_{\rm s}$>1 with both the `BCNz2` and `Deepz` codes on the `Lumos` photometry. At high redshfits, the improvement with the `Lumos` photometry and the `Deepz` code is remarkable.

Chapter 6

Improving broadband photometric redshifts with multi-task learning

6.1 Motivation

Over the last few decades, multi-band wide imaging surveys have been driving discoveries, demonstrating the power of large data sets to enable precision cosmology. Obtaining precise photometric redshifts is crucial to exploit large galaxy imaging surveys (Salvato et al., 2019b) and are a limiting factor in the accuracy of cosmology measurements using galaxies (Knox et al., 2006). Current and upcoming imaging surveys like e.g. the Dark Energy Survey (DES, The Dark Energy Survey Collaboration, 2005), the Kilo-Degree Survey (KiDS, de Jong et al., 2013), *Euclid* (Laureijs et al., 2011), and the Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezić et al., 2019a) critically depend on robust redshift estimates to obtain reliable science results (Blake & Bridle, 2005).

With larger imaging surveys (as the quality and number of photometric observations increase), the photo-z performance requirements, both in terms of bias and precision, have become increasingly stringent in response to a need to reduce the uncertainties in the science measurements. As an example, the analysis of the first year of DES data (DES Y1) had a photo-z precision requirement $\sigma_{z_\mathrm{p}-z_\mathrm{s}} < 0.12$ (Sánchez et al., 2014), with $\sigma_{z_\mathrm{p}-z_\mathrm{s}}$ being the standard deviation of the residuals between the photometric redshift z_p and the spectroscopic redshift z_s (as a proxy of the true redshift). In order to exploit the constraining power of LSST, it is required that the mean fractional photo-z bias $\langle \Delta z \rangle < 0.003$, with $\Delta z := (z_\mathrm{p} - z_\mathrm{s})/(1 + z_\mathrm{s})$, and the scaled photo-$z$ scatter $\sigma_{\Delta z} < 0.02$ (Schmidt et al., 2020), which corresponds to around three times more precise photo-zs than DES Y1. Similarly, for *Euclid*, the scaled photo-z bias is required to be below 0.002 and $\sigma_{\Delta z} < 0.05$ (Laureijs et al., 2011).

The increasingly stringent requirements on the photo-z measurements have triggered extensive investigation efforts dedicated to improving photo-z estimation methodology. Therefore, there are many different photo-z codes, which can be classified into two main approaches: the so-called template-fitting methods, (e.g. `LePhare`: Arnouts & Ilbert 2011, `BPZ`: Benítez 2011, and `ZEBRA`: Feldmann et al. 2006); and data-driven (machine learning) methods (e.g.

ANNz: Collister & Lahav 2004, ANNz2: Sadeh et al. 2016, tpz: Carrasco Kind & Brunner 2013, Skynet: Bonnett 2015b, and spiderZ: Jones & Singal 2017). These methods commonly only use the measured photometry to produce photo-z estimates. Furthermore, there is a wealth of techniques to improve the photo-z performance, like including galaxy morphology (Soo et al., 2018), using Gaussian processes (Gomes et al., 2018; Leistedt & Hogg, 2017), implementing "pseudo-labeling" semi-supervised approaches to learn the underlying structure of the data (Humphrey et al. in prep.), and directly predicting the photo-z from the astronomical images (D'Isanto & Polsterer, 2018; Pasquet-Itam & Pasquet, 2018; Pasquet et al., 2019; Chong & Yang, 2019).

Broad-band photo-z performance is limited by the resolution and the wavelength coverage provided by the photometric filters. Narrow-band photometric surveys are in between spectroscopy and broad-band photometry (Benitez et al., 2014; Martí et al., 2014; Eriksen et al., 2019). These are imaging surveys with a higher wavelength resolution than broad-band surveys, but typically cover smaller sky areas due to the increased telescope time needed to cover the same wavelength range. In this chapter, we aim to use multi-task learning (MTL, Caruana, 1997) and narrow-band data to improve broad-band photo-z estimates. MTL is a machine-learning methodology in which the model benefits from predicting multiple related tasks together, e.g. a network that predicts the animal type (e.g. elephant, dog, dolphin, or unicorn) and its weight. In this example, the network learns correlations between each animal class and how heavy these are (e.g. an elephant is heavier than a dog), and such correlations are used to improve the final predictions on both tasks.

In astronomy, often data that could be helpful to improve the photo-z performance exist, e.g. photometry in several bands. However, such data are not always available for the complete wide field, preventing us from using it. With multi-task learning, we can benefit from these data to improve the photo-z predictions without explicitly providing it as input. Particularly, we have implemented an MTL neural network that predicts the photo-z and the narrow-band photometry of a galaxy from its broad-band photometry. The narrow-band data are used to provide ground-truth labels to train the auxiliary task of reconstructing the narrow-band photometry (Liebel & Körner, 2018). Therefore, we only need it to train the network, while we can evaluate the photo-z of any galaxy with only its broad-band photometry. In this way, the data available in certain fields can be exploited to improve the photo-z estimations in other fields.

We have tested the method with data from the Physics of the Accelerating Universe Survey (PAUS), which is a narrow-band imaging survey using the PAUCam instrument (Castander et al., 2012; Padilla et al., 2016, 2019a), a camera with 40 narrow bands covering the optical spectrum (Casas et al., 2016). The method could also be applied to other narrow-band surveys like the Javalambre Physics of the Accelerating Universe (JPAS, Benitez et al., 2014).

This chapter is structured as follows. In § 6.2, we present the data used throughout the chapter. Section 6.3 introduces multi-task learning and the method developed and tested in this work. In § 6.4, we show the performance of the photo-z method in the COSMOS field, including bias, scatter, outliers, and the photo-z distributions. The performance on a deeper galaxy sample is tested in § 6.5 using simulated galaxies. Finally, we use self-organising maps (SOM) to explore the photo-z distribution of COSMOS galaxies in colour-space (§ 6.6) and to have a better understanding of the underlying mechanism of our method (§ 6.7).

6.2 Data

In this section, we present the PAUS data used for the study (§ 6.2.1) and the photometric redshift galaxy sample (§ 6.2.2). The broad-band data and the spectroscopic sample are introduced in § 6.2.3 and § 6.2.4, respectively, while § 6.2.5 shows the galaxy simulations used in the paper.

6.2.1 PAUS data

In this chapter, we use PAUS narrow-band data (§ 2) in the COSMOS field, which comprises 64 476 galaxies to $i_{\mathrm{AB}} < 23$ in 40 narrow-band filters. This corresponds to approximately 12,5 million galaxy observations (5 observations per galaxy and narrowband filter) (§ 3.4.4). PAUS has developed two methods to extract the galaxy photometry: a forced aperture algorithm (`MEMBA`, § 3.4.2) and a deep learning-based pipeline (`Lumos`, Cabayol-Garcia et al., 2020; Cabayol et al., 2021, § 5). We have found that the methodology developed in this chapter is the same with both photometric approaches.

With a template-fitting algorithm, PAUS reaches a photo-z precision $\sigma_{68}/(1+z) = 0.0035(1+z)$ for the best 50% of the sample (Eriksen et al., 2019). Similar precision is obtained with `Delight` (Soo et al., 2021), a hybrid template-machine-learning photometric redshift algorithm that uses Gaussian processes. The PAUS photo-z precision was improved further with a deep learning algorithm that reduces the scatter by 50% compared to the template-fitting method (Eriksen et al., 2020, § 3.4.3). The excellent PAUS photo-zs precision allows for studies like intrinsic alignments and clustering (Johnston et al., 2021a), measuring galaxy properties (Tortorelli et al., 2021), and measuring the D4000 Å spectral break (Renard et al. in prep., § 2.2.4).

6.2.2 Photometric redshift sample

Throughout the chapter, we also use the high-precision photometric redshifts from Alarcon et al. (2021, PAUS+COSMOS hereafter). These photometric redshifts use a combination of the 40 PAUS narrow bands and 26 broad and intermediate bands covering the UV, visible, and near-infrared spectrum (see § 2 in Alarcon et al., 2021, for more details). The PAUS+COSMOS photo-zs reach a precision of $\sigma_z/(1+z) = 0.0036$ and $\sigma_z/(1+z) = 0.0049$ for galaxies at $i_{\mathrm{AB}} < 21$ and $i_{\mathrm{AB}} < 23$, respectively. These photo-zs are more precise and less biased than those from Laigle et al. (2016) (COSMOS2015 hereafter), which use a combination of 30 broad-, intermediate-, and narrow-band filters.

6.2.3 Broadband data

The broad-band data used in this chapter are from COSMOS2015, which includes the u-band from the Canada-France Hawaii Telescope (CHFT/MegaCam) and the Subaru $BVriz$ filters. We carry out a spatial matching of COSMOS2015 and PAUS galaxies within $1''$. Then, we apply a cut on magnitude $i_{\rm AB} < 23$ and on redshift $z < 1.5$, which results in a catalogue with around 33 000 galaxies of which approximately 9000 have spectroscopic redshifts. The redshift cut is prompted by the photo-z distribution in the PAUS+COSMOS catalogue, with very few galaxies with $z > 1.5$.

6.2.4 Spectroscopic galaxy sample

To train the neural network, one needs a galaxy catalogue with known redshifts. We use the zCOSMOS DR3 bright spectroscopic data (Lilly et al., 2007), which cover 1.7 deg^2 of the COSMOS field. The catalogue covers a magnitude range of $15 < i_{\rm AB} < 23$ and a redshift range of $0.1 < z < 1.2$. We only keep redshifts with a confidence class ($conf$) of $3 < conf < 5$, which leads to a catalogue with ~ 9400 galaxies. We extend the spectroscopic sample with a compilation of 2693 redshifts from Alarcon et al. (2021). This compilation includes redshifts from C3R2 DR1&DR2 (Masters et al., 2017, 2019), 2dF (Colless et al., 2001b), DEIMOS (Hasinger et al., 2018), FMOS (Kashino et al., 2019), LRIS (Lee et al., 2018), MOSFIRE (Kriek et al., 2015), MUSE (Urrutia et al., 2019), Magellan (Calabrò et al., 2018), and VIS3COS (Paulino-Afonso et al., 2018), with a quality cut to keep only those objects with a reliable measurement.

6.2.5 Galaxy mocks

In § 6.5 we also use the Flagship galaxy simulations described in Castander et al. (in prep.). These are *Euclid*-like galaxies generated using abundance matching between the halo mass function and the galaxy luminosity function taking into account the occupation of the haloes. Evolutionary population synthesis models are used to estimate the evolution of the galaxy colours with redshift. The simulated galaxies are compared to the COSMOS galaxies from COSMOS2015. The spectral energy distribution (SED), including its extinction of the best matching COSMOS galaxy, is assigned to each simulated galaxy. The SED library is the same is in COSMOS2015, which includes SED templates from Polletta et al. (2007) and and additional blue templates from Bruzual & Charlot (2003). Emission lines are then added to the SED of each galaxy. The Hα flux is computed from the rest-frame ultra-violet flux following Kennicutt (1998). The rest-frame fluxes do not contain absorption. The other emission-line fluxes are computed using observed relations. Finally, the SED is convolved with the filter transmission curves to produce the expected observed fluxes. This prescription is followed to generate both broad- and narrow-band photometry.

6.3 Multi-task neural network to improve broad-band photo-z

In this section, we describe MTL (§ 6.3.1) and present the networks and training procedures used throughout the paper (§ 6.3.2).

6.3.1 Multi-task learning

Deep learning algorithms consist of training a single model or an ensemble of models to accurately perform a single task, e.g. predicting the redshift. Multi-task learning is a training methodology that aims to improve the performance on a single task by training the model on multiple related tasks simultaneously (Caruana, 1997). One can think of MTL as a form of inductive transfer, where the knowledge that the network acquires from one task introduces an inductive bias to the model, making it prefer certain hypotheses over others. A simple pedagogical example is a network to classify cats and dogs. If we include a secondary task to classify the shape of the ears in e.g. spiky or rounded, the network will make correlations between the ear shapes and the animal class, in such a way that the predicted ears shape will also affect the cat-dog classification. This kind of network has already been successfully applied in other fields, such as e.g. video processing (Song et al., 2020) or medical imaging (Moeskops et al., 2017), where in the latter case a single network is trained to segment six tissues in brain images, the pectoral muscle in breast images, and the coronary arteries. There are also successful implementations in astrophysics. Examples include e.g. Parks et al. (2018), which characterises the strong HI Lyα absorption in quaser spectra simultaneously predicting the presence of strong HI absorption and the corresponding redshift z_{abs} and the HI column density. Also, Cunha & Humphrey (2022) describes `SHEEP`, a machine learning pipeline for the classification of galaxies, QSO, and stars from photometric data that benefits from predicting the photo-z and using its prediction as a new feature for model.

6.3.2 Model architecture and training procedures

Broadly, there are two types of MTL-network architectures called soft- and hard-parameter sharing (Zhang & Yang, 2021). In the former, each task has its parameters, which are regularised to be similar among tasks. For the latter, the hidden layers of the network are shared between tasks, while keeping task-specific layers separate. Hard-parameter sharing is the most common MTL architecture and it is the one used in this paper.

Figure 6.1 shows the two networks used in this paper. The top panel presents the baseline network, a single-task network mapping the broad-band photometry to the photometric redshifts. It concatenates six fully-connected layers with parameters 5:300:500:1000:500:300:1500, where the numbers correspond to the number of nodes in the layers. Therefore, the first contains five nodes, corresponding to the five consecutive colours obtained with the $uBVriz$ broad-bands. The last layer consists of 1500 redshift bins of redshift width 0.001 covering a redshift range $0 < z < 1.5$. We have selected this redshift range since there are very few higher redshift galaxies with PAUS photo-z and spectroscopic redshift. For every galaxy, the network outputs the probability that the redshift belongs to each redshift bin in such a way that it effectively predicts the redshift probability density function $p(z)$. Each layer is followed by a 2% dropout layer (Srivastava et al., 2014), a regularization method in which several nodes are randomly ignored during the training phase. Dropout is represented with the yellow-crossed circles in Fig. 6.1.

The bottom panel in Fig. 6.1 represents the MTL network introduced in this paper, which

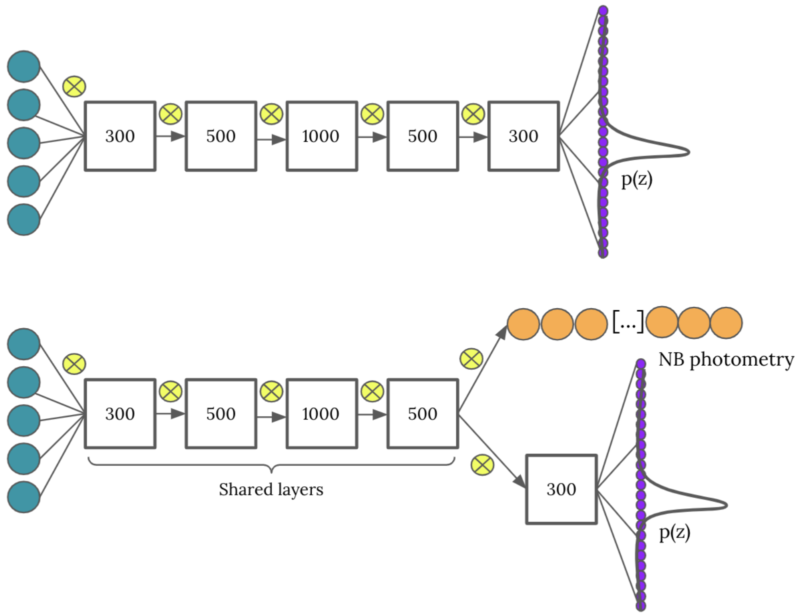

Figure 6.1: *Top:* Baseline network architecture. The input contains five colours that propagate through six fully connected layers. Each layer is followed by a dropout layer, which is represented by a yellow-crossed circle. *Bottom:* Multitask learning network. This builds on the baseline network and adds an extra output layer for the additional task of predicting the narrow-band photometry.

includes the additional task of predicting the PAUS narrow-band photometry using a hard parameter-sharing architecture. The core architecture is the same as that of the baseline network (upper panel) but this network contains an extra output layer for the additional task of predicting the narrow-band photometry.

Both networks train the photometric redshift prediction with a cross-entropy loss function (Good, 1952),

$$\mathcal{L}_{\mathrm{z}} := \sum_{c=1}^{1500} \left[\, p_c(z)\, \delta(z_{\mathrm{s}}) \,\right] \,, \tag{6.1}$$

which assumes the true-redshift probability distribution is a Delta function centred at the spectroscopic redshift bin $\delta(z_{\mathrm{s}})$. The ground-truth redshift labels are the spectroscopic redshifts as defined in § 6.2.4. The summation is over the redshift bins (1500 in our case) and $p_c(z)$ is the probability assigned by the network to redshift bin c. The cross-entropy loss is a standard loss function for classification problems. Initially, we also tested tackling the photo-z prediction as a regression problem using a mixture density network (§ 1.2.2, D'Isanto & Polsterer, 2018; Eriksen et al., 2020). Here this approach led to worse results and we decided on the classifier. The MTL network enables including information from the galaxy SED, while extending the training sample to galaxies without spectroscopic redshift but with narrow-band photometry. Predicting the photo-z and the narrow-band photometry simultaneously, the two tasks share internal representations, thus the non-spectroscopic galaxies indirectly affect the training of the photo-z prediction.

The training of the narrow-band is addressed with a least absolute deviation loss function

$$\mathcal{L}_{\rm NB} \coloneqq \frac{\sum_i \left|\mathrm{NB}_i^{\rm pred} - \mathrm{NB}_i^{\rm obs}\right|}{N-1}\,, \tag{6.2}$$

where $\mathrm{NB}_i^{\rm pred}$ and $\mathrm{NB}_i^{\rm obs}$ are the predicted and observed narrow-band colours in the i-th filter, respectively, and N is the number of narrow bands. We also tested other alternatives, e.g. the mean-squared error, but this was hindering the network's convergence and we decided on the absolute-mean error.

Consequently, the training methodologies are:

1. $z_{\rm s}$: This is the usual training that maps the broad-band photometry to photo-z using spectroscopic redshifts as ground-truth redshifts and a cross-entropy loss function (Eq. 6.1);

2. $z_{\rm s} + NB$: This methodology includes MTL. It maps the broad-band photometry to photo-z and narrow-band photometry, therefore the loss function is the mean of the combined cross-entropy loss (Eq. 6.1) and narrow-band reconstruction (Eq. 6.2) tasks for all galaxies (N) for which the loss is computed

$$\mathcal{L}_{\rm NB+z_s} \coloneqq \frac{1}{N}\sum_{j=1}^{N} \mathcal{L}_z^j + \mathcal{L}_{\rm NB}^j\,. \tag{6.3}$$

   We only use galaxies with spectroscopic redshift to train the photo-z predictions, while all galaxies with narrow-band observations train the narrow-band reconstruction. In general, one can also weight the two terms in the loss functions. Testing different values, we found the photo-z scatter to have a minimum in a wide range of values around equal weighting.

Furthermore, we considered two variants in the training procedure to explore the possibility of using high-precision photometric redshifts (§ 6.2.2) to train the networks:

3. $z_{\rm s} + z_{\rm PAUS}$: This is a variation of the $z_{\rm s}$ method. The training sample extends to galaxies having a high-precision photo-z estimate in the PAUS+COSMOS catalogue. For galaxies with spectroscopy, we use the spectroscopic redshift as ground-truth while for the rest of the training sample, the PAUS+COSMOS photo-z is used to train the network;

4. $z_{\rm s}$+NB+$z_{\rm PAUS}$: This is a variation of the $z_{\rm s} + NB$ method, and it also extends the training sample with galaxies with a high-precision photo-z estimate in the PAUS+COSMOS catalogue. In contrast to the $z_{\rm s} + NB$ method, here all galaxies are used to train the photo-z prediction and the narrow-band photometry reconstruction. The ground-truth redshift labels are the spectroscopic redshifts if available and otherwise, the PAUS+COSMOS photo-z.

The networks are implemented in `PyTorch` (Paszke et al., 2017), and all the training procedures use an `Adam` optimizer (Kingma & Ba, 2014) for 100 epochs with an initial learning rate of 10^{-3} that reduces by a factor of ten every 25 epochs.

6.4 Photo-z performance in the COSMOS field

In this section, we show the photo-z performance of our method on galaxies with $i_{\mathrm{AB}} < 23$ and $z < 1.5$ in the COSMOS field. We will study the effect that MTL has on the dispersion (§ 6.4.2) and the bias (§ 6.4.3) of the predicted photo-zs, while § 6.4.4 investigates the effect of MTL on the redshift distributions $N(z)$.

6.4.1 Photo-z performance metrics

To evaluate the accuracy and precision of the photo-z estimates, we define

$$\Delta z \coloneqq (z_{\mathrm{p}} - z_{\mathrm{t}}) \; / \; (1 + z_{\mathrm{t}}) \, , \tag{6.4}$$

where z_{p} and z_{t} are the photo-z and the ground-truth redshift, respectively. The bias and the dispersion are defined as the median and σ_{68} of Δz, respectively, where we define σ_{68} as

$$\sigma_{68} \coloneqq \frac{1}{2} \left[Q_{84}(\Delta z) - Q_{16}(\Delta z) \right] \, , \tag{6.5}$$

and $Q_{16}(\Delta z)$, $Q_{84}(\Delta z)$ are the 16th and 84th percentiles of the Δz distribution. We also include the metric

$$\sigma_{\mathrm{NMAD}} \coloneqq 1.4826 \times \mathrm{median} \left[\, |\Delta \mathrm{z} - \mathrm{median}(\Delta \mathrm{z})| \, \right] \tag{6.6}$$

used in the Euclid photo-z challenge paper (Desprez et al., 2020).

To evaluate the performance on the full COSMOS catalogue, we define the ground-truth redshift as the spectroscopic redshift if available and otherwise, as the PAUS+COSMOS photo-z (§ 6.2.2).[1]. If it is not specified by the method our networks are trained with spectroscopic redshifts only. For the performance evaluation, however, the PAUS+COSMOS photo-zs are also used but only to evaluate the photo-z of those galaxies from the full COSMOS catalogue that do not have a spectroscopic redshift estimate. The predicted photo-zs are defined as the mode of the redshift probability distribution provided by the network (§ 6.3.2).

In order to estimate the photo-zs of the complete COSMOS catalogue, the networks are trained independently ten times with ∼11 000 spectroscopic galaxies in each iteration, which roughly corresponds to 90% of the sample. Each network is used to evaluate the corresponding 10% excluded galaxies in such a way that the ensemble of networks evaluates the full COSMOS catalogue.

Including MTL extends the training sample to about 40 000 galaxies, which corresponds approximately 3.5 times more galaxies than in the spectroscopic sample. In order to evaluate

[1]The PAUS+COSMOS photo-zs used to evaluate the precision of non-spectroscopic galaxies (§ 6.2.2) also have an associated dispersion. This corresponds to approximately 4% lower photo-z scatter than that obtained for very bright galaxies and around 1% lower at the faintest end.

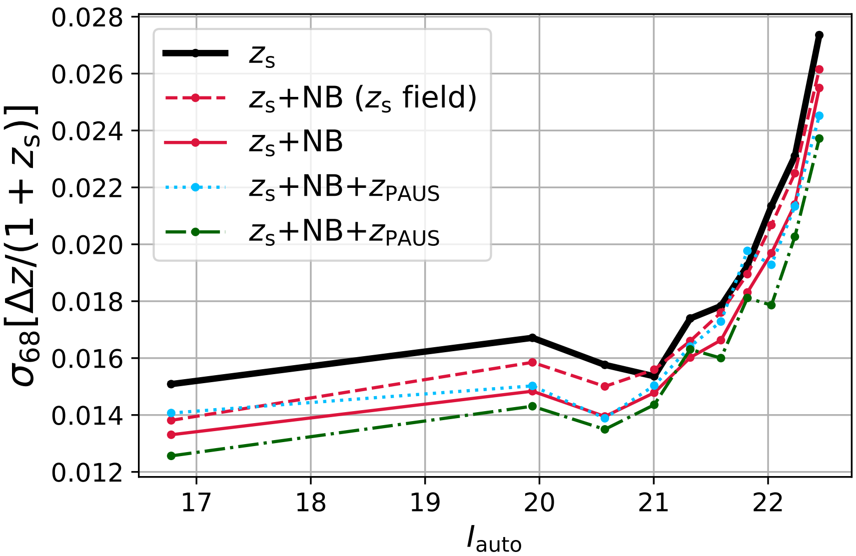

Figure 6.2: Photo-z dispersion in equally populated magnitude differential bins to $i_{AB} < 23$. Each line corresponds to a different training procedure (see § 6.3.2). While the black line corresponds to a baseline training, the other coloured lines include MTL (red and green lines) and data augmentation with photo-zs from the PAUS+COSMOS catalogue as ground-truth redshifts (blue and green lines).

the full COSMOS sample, we trained the network seven independent times with 85% of the spectroscopic galaxies and 85% of the non-spectroscopic sample. This corresponds to around 11 000 galaxies with spectroscopy and 25 000 without. We have ensured that the fraction of galaxies with spectroscopic redshifts in each iteration is similar by sampling without replacement the same number of spectroscopic galaxies in each iteration.

6.4.2 Photo-z dispersion

Table 6.1 presents the photo-z precision for the COSMOS spectroscopic sample, as well as the full sample using the four different training procedures presented in § 6.3.2. These results are presented in more detail in Fig. 6.2, which shows the photo-z dispersion in equally populated magnitude bins with the same four methodologies. The solid black line corresponds to the baseline network mapping broad-band photometry to photo-z (method z_s in § 6.3.2). This method is trained on the spectroscopic sample and provides a $\sigma_{68} = 0.021$ for the full sample, while being unbiased ($\Delta z < 0.01$). These are quite precise and accurate redshifts compared to other broad-band redshift estimates in the same field. In Hildebrandt et al. 2009, redshifts in the D2 CHFT deep field (Coupon et al., 2009), which overlaps with COSMOS, were estimated with the template-fitting code `BPz` (Benítez, 2011) using the CFHT *ugriz* filter set. Their photo-z precision is $\sigma_{68} = 0.0498$, while for the same galaxy sub-sample our network provides $\sigma_{68} = 0.0187$. Here neither the methodology nor the input data are the same, but these CFHT photo-z estimates are helpful to have a reference for our photo-z baseline network performance.

Both red lines in Fig. 6.2 use an MTL training (method $z_\mathrm{s} + NB$ in § 6.3.2). The dashed red line only exploits galaxies with spectroscopic redshifts to train the narrow-band reconstruction and the photo-z prediction. Therefore, we note that this does not fully correspond to the MTL methodology developed in this work, since the training sample does not extend to PAUS galaxies without spectroscopy. This methodology results in a precision of $\sigma_{68} = 0.020$, which is a 4% improvement with respect to the baseline methodology (black line, method z_s in § 6.3.2). The method has a larger effect at the bright end, while the photo-zs of fainter galaxies in the data set ($22 < i_\mathrm{AB} < 23$) are barely improved.

In contrast, the solid red line uses all galaxies with PAUS photometry to train the narrow-band reconstruction and only those with spectroscopy to train the photo-z prediction. This extends the training sample of the shared layers from around 12 000 to 30 000 galaxies, which results in a precision of $\sigma_{68} = 0.0173$, corresponding to a 9% improvement with respect to the baseline methodology (solid black line). Contrary to the dashed red line, the improvement is also significant at the faint-end, with 10% more precise photo-zs when MTL is included.

Making the distinction between these two training methodologies isolates the effect of the MTL. While in the case represented by the dashed red line the improvement in the photo-z prediction is due to including the PAUS narrow-band prediction as an auxiliary task, the combination of the auxiliary task and the explicit data augmentation causes the improvement evident in the solid red line. This implies that the improvement at the faint end is driven by the data augmentation, as applying an MTL training with the spectroscopic sample only barely improves the photo-z predictions of faint objects. Since PAUS galaxies have a low SNR in the narrow-bands at the faint end, MTL without data augmentation is not enough to improve the photo-z performance in this regime. In contrast, we obtain better photo-z predictions only with MTL training at the bright end, where the SNR is significantly higher.

The blue dotted line in Fig. 6.2 also corresponds to a direct mapping of the broad-band photometry to photo-zs. However, in contrast to the solid black line, this case is trained on an extended sample including galaxies without spectroscopic redshifts (method $z_\mathrm{s} + z_\mathrm{PAUS}$) in § 6.3.2), for which the PAUS+COSMOS photo-z measurement is used as a ground-truth redshift label in the training. It shows a precision of $\sigma_{68} = 0.0197$, which corresponds to a 8% improvement with respect to the baseline training. Note that this method does not use MTL, but its effect on the photo-z performance is similar to including it (the solid red line). In § 6.6 we discuss the underlying mechanism that causes MTL with PAUS to improve the photo-zs.

The best photo-z performance is achieved combining MTL and photo-z data augmentation with PAUS+COSMOS data (method z_s+NB+z_PAUS in § 6.3.2), which corresponds to the dotted green line in Fig. 6.2. This method gives a 16% improvement with respect to the baseline network, with a precision of $\sigma_{68} = 0.0185$.

	Precision (z_s sample)	Precision (COSMOS)	Outliers
z_s	1.70 (1.46)	2.14 (1.90)	1.2
$z_\mathrm{s} + NB$	1.54 (1.25)	1.97 (1.73)	0.8
$z_\mathrm{s} + z_\mathrm{PAUS}$	1.64 (1.43)	1.98 (1.76)	1.2
NB+$z_\mathrm{s} + z_\mathrm{PAUS}$	1.57 (1.30)	1.85 (1.60)	0.6

Table 6.1: Photo-z dispersion $\sigma_{68} \times 100$ for the different network configurations. The second column displays results restricted to the spectroscopic sample, while the third column shows the results for the full COSMOS to $i_\mathrm{AB} < 23$. For the full COSMOS sample results, the PAUS+COSMOS high-precision photo-zs are used as ground-truth redshifts when spectroscopy is not available. The numbers in parenthesis corresponds to the σ_NMAD. The fourth column presents the percentage of photo-z outliers on the full sample.

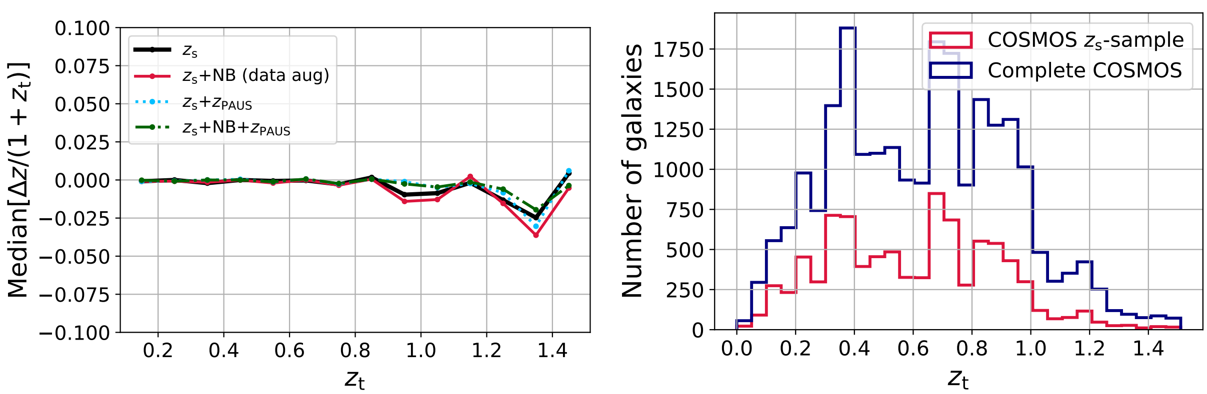

Figure 6.3: *Left*: Photo-z bias in equally spaced redshift bins. *Right*: Redshift distributions for the COSMOS spectroscopic sample (red line) and the full (spectroscopic and photo-z) COSMOS sample.

6.4.3 Photo-z bias and outlier rate

In this subsection, we show the bias for the photo-z predictions with the MTL networks and the baseline broad-band network. The left panel in Fig. 6.3 shows the photo-z bias in equally spaced redshift bins of width 0.1 in the redshift range $0.1 < z_\mathrm{t} < 1.5$. We have excluded the first redshift bin from the analysis since there are almost no galaxies with $z_\mathrm{t} < 0.07$, which caused a bias at very low redshift[2]. Overall, for $z_\mathrm{t} < 1.2$ the four methods presented in § 6.3.2 are unbiased at the level of < 0.01. At higher redshifts ($z_\mathrm{t} > 1.2$), the four methods show a $\sim$2% bias. The right panel of Fig. 6.3 suggests that this is likely to be caused by a lack of training examples at such redshifts, with very few spectroscopic training examples with $z_\mathrm{t} > 1.3$.

[2]There are training mechanisms to deal with unbalanced training samples like e.g. up-weighting the contribution of unbalanced class objects in the training or oversampling synthetic data from the unbalanced original ones (Yanminsun et al., 2011). However, the number of objects with $z < 0.07$ is too small to efficiently apply these techniques and there are very few galaxies affected.

In this chapter, we consider a galaxy to be an outlier if

$$|z_{\rm p} - z_{\rm t}| \; / \; (1 + z_{\rm t}) > 0.15\,. \tag{6.7}$$

In the spectroscopic sample, the baseline network yields 1.1% outliers, which reduces to 0.8% with the MTL using PAUS photometry, the training sample extension with PAUS+COSMOS photo-z, and the combination of both. The fraction of outliers in the full COSMOS sample is 1.2 for the baseline network and for the training sample extension with PAUS+COSMOS photo-zs ($z_{\rm s} + z_{\rm PAUS}$). The methodologies including MTL reduce the outlier fraction to 0.8% ($z_{\rm s} + NB$) and 0.6% ($z_{\rm s} + z_{\rm PAUS} + NB$). While in the spectroscopic sample extending the training sample and including MTL have a similar effect on the outlier fraction, in the full COSMOS sample MTL has a stronger impact.

6.4.4 Redshift distributions, N(z)

Unbiased redshift distributions, $N(z)$, are crucial for a variety of science applications, with the most stringent requirements being in weak lensing (e.g. Hildebrandt et al., 2012b; Hoyle et al., 2018). Broad-band photo-z commonly suffer from biases due to degeneracies between colours and redshift, (e.g. Newman et al., 2015; Masters et al., 2017) and as shown in Fig. 6.3, the baseline network exhibits a bias at high redshifts.

Figure 6.4 shows $N(z)$ in tomographic redshift bins for $0 < z_{\rm t} < 1.5$ spaced by 0.2. The last tomographic bin is defined from $1.2 < z_{\rm t} < 1.5$ so that the number of galaxies in the bin is increased. The ground-truth redshift defining the tomographic bins ($z_{\rm t}$) is a combination of the spectroscopic redshift (when it is available) and the PAUS+COSMOS photo-z elsewhere. The vertical solid grey line indicates the ground-truth median redshift of the tomographic bin, while the dashed coloured lines represent the median redshifts of the predicted photo-zs assigned to the bins.

MTL with photo-z data augmentation ($z_{\rm s} + z_{\rm PAUS} + NB$) always provides equal or more accurate $N(z)$ than the baseline network ($z_{\rm s}$, black line). As expected from Fig. 6.3, the $N(z)$ values exhibiting the largest bias are those with $z_{\rm t} > 1.2$, particularly the bin at $z_{\rm t} > 1.2$. In this bin, MTL together with the photo-z data augmentation ($z_{\rm s}$+NB+$z_{\rm PAUS}$, green line), significantly shifts the median of the $N(z)$ towards the PAUS+COSMOS result.

Commonly, redshift distributions require a bias correction to reach the accuracy requirements of cosmological measurements. Techniques such as clustering redshifts are applied to correct such biases (Ménard et al., 2013; Schmidt et al., 2013; Gatti et al., 2018; van den Busch et al., 2020; Hildebrandt et al., 2021). MTL reduces the bias of the $N(z)$ already at the photo-z prediction stage. Even if the MTL photo-zs still require some correction, the final redshift distributions would benefit from initially having less biased redshift distributions (if these redshift distributions are used to fit the clustering-z data points).

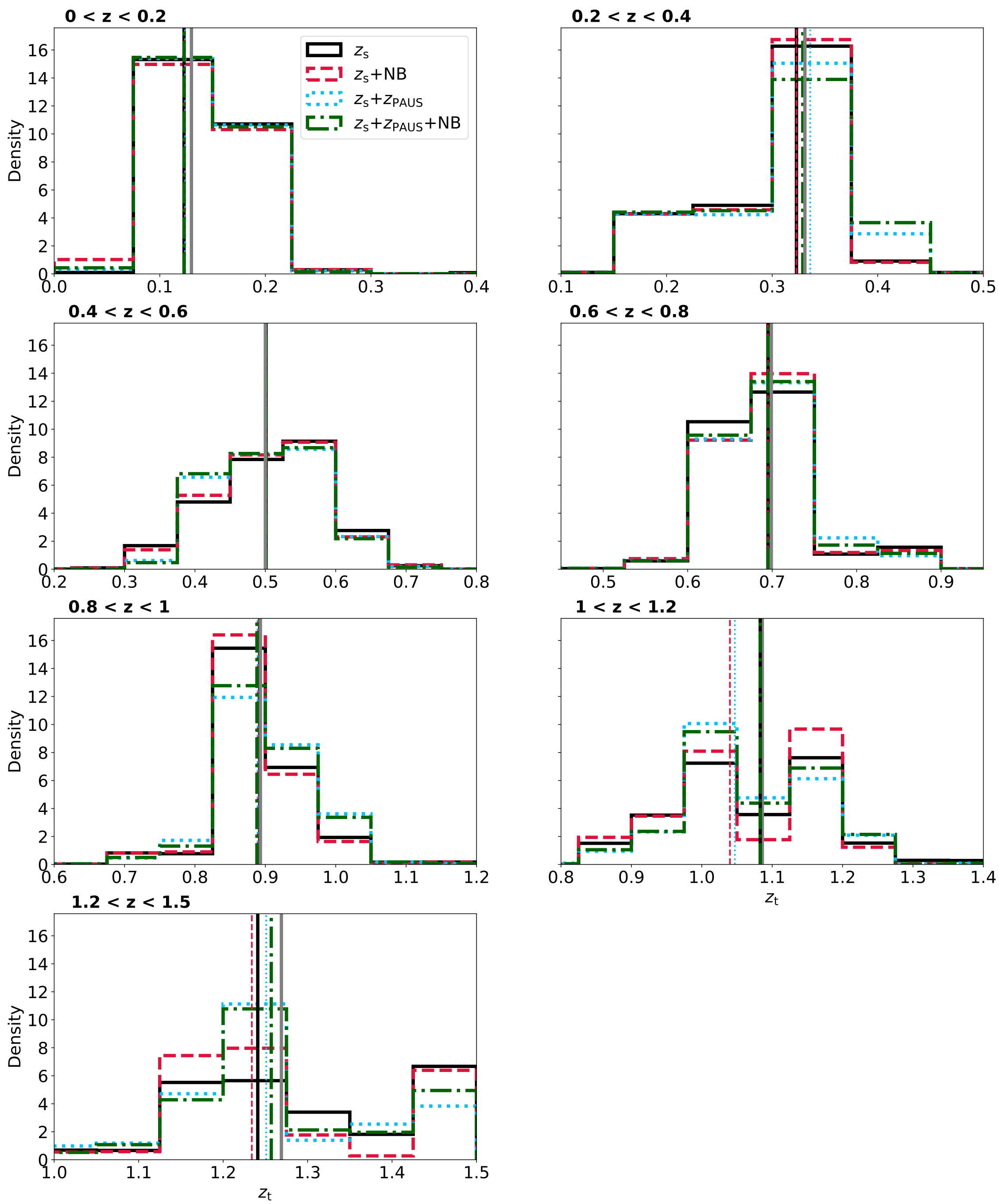

Figure 6.4: The $N(z)$ estimates of the full COSMOS sample divided into 7 tomographic bins over the redshift range $0 < z < 1.5$. Tomographic bins are defined using the spectroscopic redshifts and the PAUS+COSMOS high-precision photo-zs for galaxies without spectroscopy. The vertical solid grey lines indicate the median ground-truth redshift, while the other vertical lines indicate the median redshifts of the $N(z)$ estimates. Unseen lines are hidden by other overlapping lines.

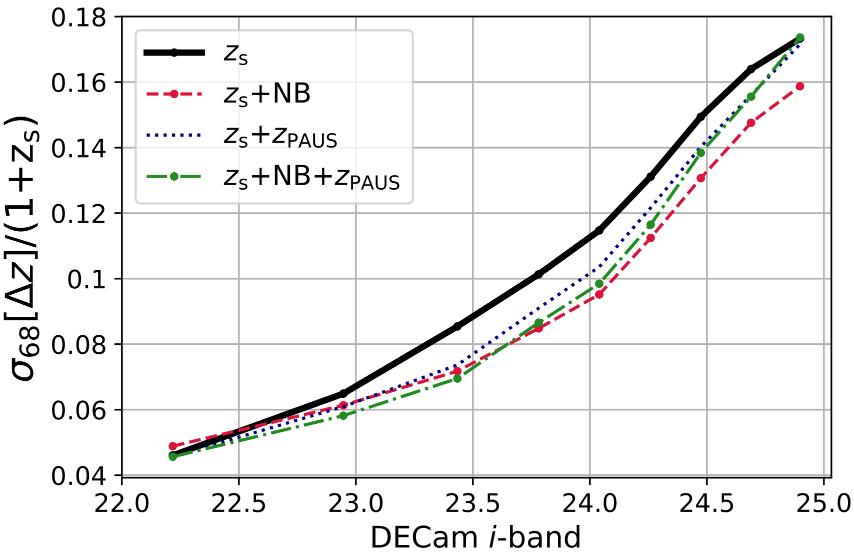

Figure 6.5: Photo-z dispersion in equally sized magnitude bins for 30 000 Flagship test galaxies with magnitudes $i_{AB} < 25$ for the methods presented in § 6.3.2. The training sample contains around 15 000 spectroscopic galaxies, extended to 30 000 with PAUS-like galaxies without spectroscopy, all of them to $i_{AB} < 23$.

6.5 Photo-z performance on deeper galaxy simulations

So far, all the networks have been trained and evaluated on samples within the same magnitude range $i_{AB} < 23$ (see § 6.4). However, if the MTL network developed in this paper aims to improve the photo-z estimates of future deeper broad-band surveys such as e.g. *Euclid* or LSST, the photo-z improvement it provides must hold for fainter galaxies. In the case of *Euclid*, observations will reach a limiting magnitude of 24.5 for the VIS instrument (Cropper et al., 2012; Amiaux et al., 2012) with 10 σ depth for extended sources, which corresponds to a similar depth in the i-band filter. Rubin will observe to a single exposure depth of $r_{AB} \sim 24.5$ and a co-added survey depth of $r_{AB} \sim 27.5$ (Ivezić et al., 2019b), where the depth in the r band and the i band are also similar.

Currently, there are no PAUS measurements beyond $i_{AB} = 23$, thus limiting the magnitude range of the MTL training sample. Although observing deeper with PAUS is technically feasible, it would require considerably more observing time. Therefore, the MTL network must provide reliable photo-z predictions for deep data samples, while it is trained on a shallower data sample. Nevertheless, we note that this problem is not exclusive to our MTL network, but it affects all photo-z machine learning algorithms. These are usually trained on relatively shallow spectroscopic samples and used to predict the photo-zs for much deeper data samples (Masters et al., 2017).

In this section, we explore how the MTL network performs for deep samples ($i_{AB} < 25$), while the training is limited to galaxies with $i_{AB} < 23$ using Flagship galaxy mocks (see § 6.2.5). The broad-bands used for this test are the CFHT u band, the $griz$ bands from DE-

Cam (Honscheid & DePoy, 2008), and the *Euclid*-NISP near-infrared $H_{\rm E}$, $J_{\rm E}$ and $Y_{\rm E}$ bands (Mauri et al., 2020)[3]. These are not the same bands that were used in the tests of the COSMOS field (see § 6.2.3 and § 6.4), but these bands were chosen to demonstrate the potential benefits for the *Euclid* photo-z estimation.

We trained the four methods presented in § 6.3.2 on a sample with 10 000 spectroscopic galaxies, which are augmented to 30 000 with PAUS-like galaxies without spectroscopic redshifts and limited to $i_{\rm AB} < 23$. These numbers were chosen to approximately match the number of spectroscopic and PAUS-like galaxies in the COSMOS field (see § 6.4). To simulate the performance of the approaches that extend the training sample with high-precision photo-zs (methods $z_{\rm s} + z_{\rm PAUS}$ and $z_{\rm s}$+NB+$z_{\rm PAUS}$ in § 6.3.1), we added a scatter to the true redshifts of the PAUS-like simulated galaxies, so that the precision resembles that of the PAUS+COSMOS photo-zs.

Figure 6.5 shows the photo-z dispersion of 30 000 simulated test galaxies to magnitude $i_{\rm AB} < 25$ in equally populated magnitude bins. The baseline network (black thick line) achieves an overall precision of $\sigma_{68} = 0.99$, which increases to $\sigma_{68} = 0.123$ for galaxies with $i_{\rm AB} > 23$. Training using photo-zs but without MTL ($z_{\rm s} + z_{\rm PAUS}$, dotted blue line) improves the precision to $\sigma_{68} = 0.090$ and $\sigma_{68} = 0.110$ for galaxies with $i_{\rm AB} > 23$. With $z_{\rm s} + NB$, the overall precision is $\sigma_{68} = 0.089$, which degrades to $\sigma_{68} = 0.107$ for galaxies with $i_{\rm AB} > 23$. Finally, combining MTL and the photo-z data augmentation ($z_{\rm s}$+NB+$z_{\rm PAUS}$, solid green line) provides the best photo-z performance with $\sigma_{68} = 0.086$ for the full sample, which increases to $\sigma_{68} = 0.107$ for galaxies with $i_{\rm AB} > 23$.

For all training methods the relative improvement with respect to the baseline network is larger at fainter magnitudes. As an example, the $z_{\rm s}$+NB+$z_{\rm PAUS}$ method (green line) provides a 4% improvement with respect to the ($z_{\rm s}$) network (black-thick line) for galaxies with $i_{\rm AB} < 23$. This improvement increases to 10% for galaxies with $23 < i_{\rm AB} < 25$. This indicates that by using narrow-band photometry as the auxiliary task, the network not only learns the colour-redshift relation, but also the underlying colour distribution of the sample, which in turn improves the redshift predictions for fainter galaxy samples, where the learning of the colour distribution proves to be more valuable (further discussion in § 6.7).

6.6 Photo-z in colour-space

MTL using PAUS photometry improves the photo-z performance even if the training sample does not include galaxies beyond the spectroscopic sample (see e.g. § 6.4). While the effect of increasing the training sample in machine learning algorithms has been extensively studied, we still need to understand why MTL with narrow-band photometry improves the photo-z estimates. In this section, we use SOMs (see Appendix 6.A) to explore the COSMOS photo-z performance in colour-space (§ 6.6.1). Furthermore, in § 6.6.2 and § 6.6.3 we study how MTL

[3] With the following $5\,\sigma$ limiting magnitudes: u: 25.25; g: 24.65; r: 24.15; i: 24.35; z: 23.95; $Y_{\rm E}$: 24.0, $J_{\rm E}$: 24, $H_{\rm E}$: 24.

with PAUS narrow bands breaks broad-band degeneracies potentially caused by emission lines.

6.6.1 MTL photo-z in colour-space

A SOM (SOM, Kohonen, 1982) is an unsupervised machine learning algorithm trained to produce a low-dimensional (typically two-dimensional) representation of a multi-dimensional space. A 2-dimensional SOM contains (N_x, N_y) cells, each of them with an associated vector of attributes, in our case colour vectors. Initially, each cell is represented with random colours, which during the training phase are optimised to represent the colour-space of the training sample. The SOM training also groups together cells representing similar colours, creating a colour-space map. Once trained, each galaxy is assigned to its closest cell in colour-space. Moreover, since the SOM clusters galaxies with similar galaxy colours it also clusters galaxies with similar redshifts (Masters et al., 2015; Buchs et al., 2019). Appendix 6.A contains a more detailed explanation of SOM algorithms. SOMs have already been used in different astronomical applications, such as the correction for systematic effects in angular galaxy clustering measurements (Johnston et al., 2021b) and for estimation and calibration of photometric redshifts (Carrasco Kind & Brunner, 2014a; Wright et al., 2020a,b; Hildebrandt et al., 2021).

To show the MTL performance in colour-space we trained a 60×70 SOM on the $uBVriz$ photometry from the COSMOS2015 catalogue (see § 6.2.3), and subsequently assigned a SOM cell to each galaxy in the catalogue. The choice of SOM dimension is based on previous works, where 60×70 cells was found to give a good balance between resolution in colour-space and the number of galaxies per cell. Figure 6.6 shows the predicted photo-zs in colour-space, with each column corresponding to a photo-z estimation method described in § 6.3.2. The first row shows the photo-z distribution, where each cell is coloured with the median photo-z of the galaxies it contains. The leftmost panel (($z_{\rm s}$), panel A) displays the photo-zs with the baseline network (($z_{\rm s}$) method), and the second (B) and third (C) panels include MTL in the training, i.e. $z_{\rm s}$+NB and NB+$z_{\rm s} + z_{\rm PAUS}$ methods, respectively, bottom panel on Fig. 6.1. The rightmost panel shows the ground-truth redshift distribution.

The three methods show a photo-z distribution in colour-space that is similar to that of the ground-truth redshifts. However, some differences can be seen in the plots in the second row (panels D, E, and F), which show the differences between the predicted and true-redshift colour-maps (e.g. panel D = panel A - $z_{\rm t}$). The network trained with only broad-bands (panel D) exhibits two regions with les accurate photo-zs. These regions are centred around coordinates $(5, 35)$ (yellowish spot) and $(55, 25)$ (bluish spot), and the redshift accuracy improves when MTL (panel E) or MTL+$z_{\rm PAUS}$ (panel F) are included in the training.

These regions are seen clearer in the third row of Fig. 6.6, which shows the photo-z precision (σ_{68}, Eq. 7.16). Comparing panels G and D, we note that the photo-z precision worsens in the same regions where photo-zs are less accurate, but this improves with MTL ($z_{\rm s} + NB$, panel H) and including the PAUS+COSMOS photo-zs (NB+$z_{\rm s} + z_{\rm PAUS}$, panel F). Finally, the fourth row shows the dispersion of the redshift distribution, i.e. the width of the $N(z)$, within

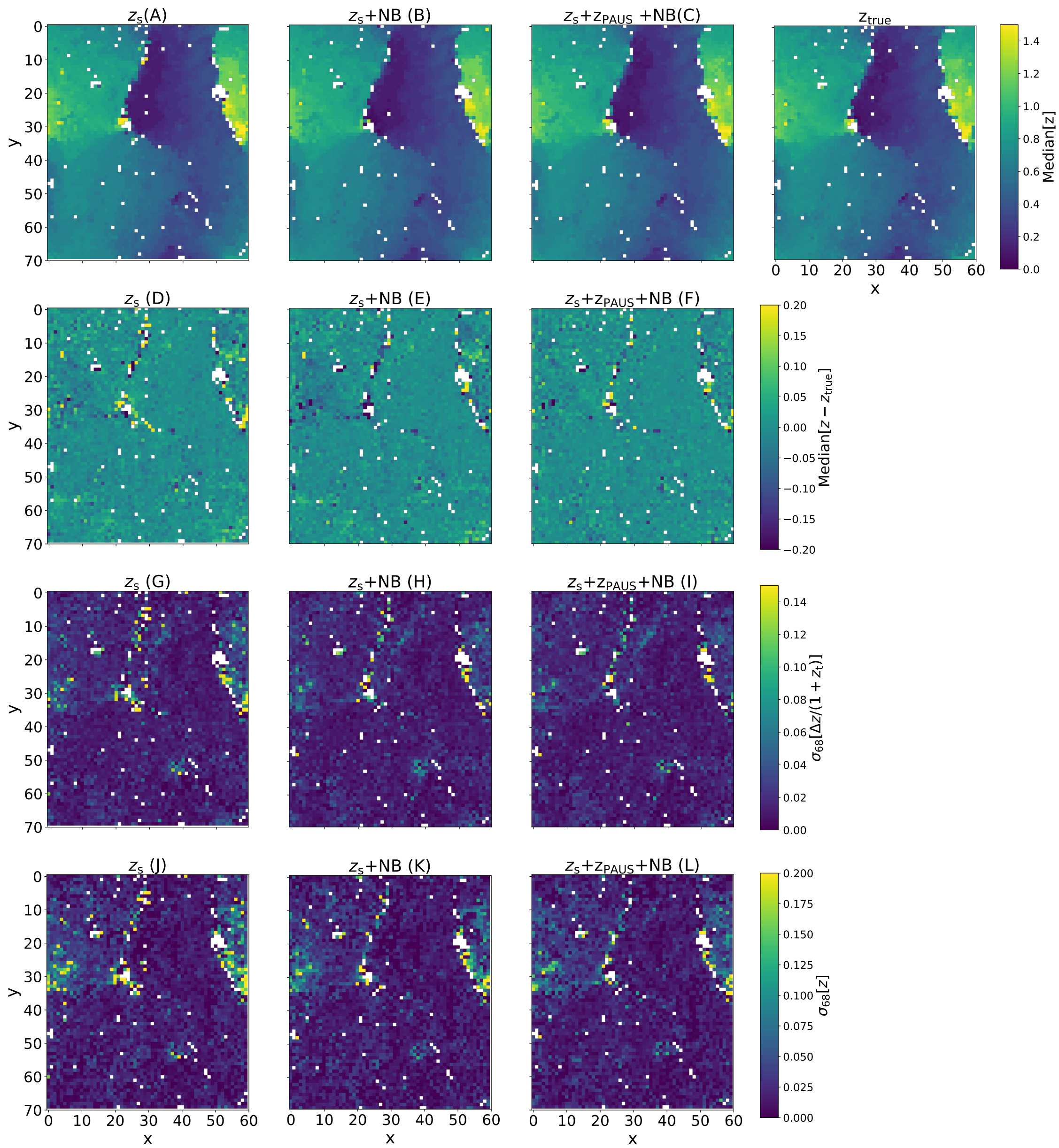

Figure 6.6: SOM maps showing the photo-z performance in the COSMOS field. The first row exhibits the median predicted photo-z in colour-space for the baseline network (first panel), including MTL training (second panel), with MTL and data augmentation with PAUS+COSMOS photo-zs (third panel) and the ground-truth redshift (fourth panel). The second row shows the bias in the photo-z predictions for the three training methods of the first row (three first panels). The third row follows the same scheme as the second but displays the photo-z precision. Finally, the fourth row shows the photo-z cell dispersion also following the same scheme. White cells correspond to empty cells, i.e. cells without any galaxy.

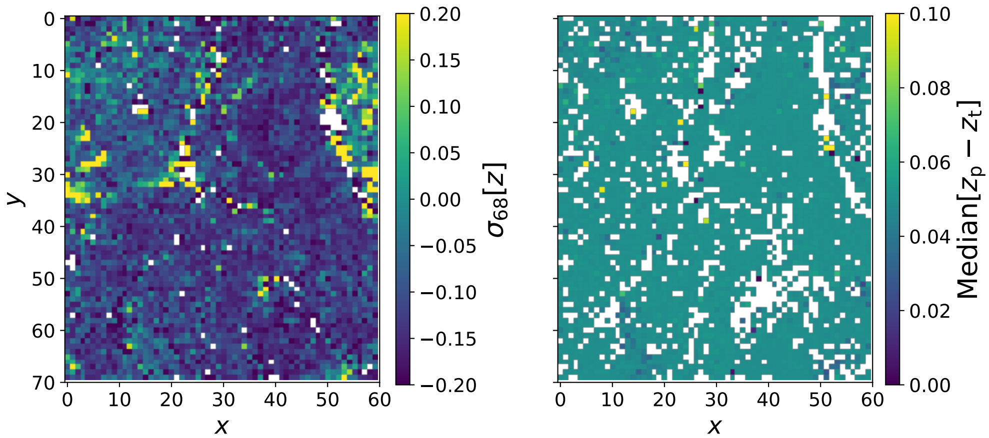

Figure 6.7: *Left:* Precision of the PAUS+COSMOS photo-zs in colour-space in the spectroscopic sample. *Right:* Accuracy of photo-z within the SOM cells for the complete catalogue.

SOM cells. This quantity is also higher for the clusters pointed out in panels D and G. However, contrary to the previous panels, the $z_\mathrm{s} + NB$ training (panel K), or the z_s+NB+z_PAUS (panel L) do not narrow the redshift distributions.

The fact that the photo-z accuracy and precision improve with MTL, while the width of the redshift distribution does not, suggests that galaxies from different populations, that is with different redshifts, are assigned to these cells. Figure 6.7 supports this hypothesis by showing that the PAUS+COSMOS photo-zs also exhibit a higher redshift dispersion (left panel) in the SOM cells within the problematic regions, while the PAUS+COSMOS photo-z accuracy is smooth across colour-space (right panel). Therefore, there are galaxies with different redshifts clustered together in broad-band colour-space.

6.6.2 Broad-band degeneracies in colour-space

SOM cells containing different galaxy populations can be the result of colour-redshift degeneracies in the broad-band photometry. Such broad-band degeneracies also cause the poor photo-z performance of the baseline network in the problematic colour-space regions. The photo-z performance improves with the MTL training (plot E in Fig. 6.6), which suggests that MTL with PAUS narrow-band photometry is able to break such broad-band colour-redshift degeneracies.

The inaccurate photo-z cluster in Fig. 6.6 is adjacent to an empty colour-space region, which shows up as a blank stripe separating two neighbouring galaxy populations. To understand which galaxies populate cells next to empty regions, we trained a SOM on a simulated galaxy sample (see § 6.2.5 for details on the mock) using the $uBVriz$ broad-band photometry. The top panel in Fig. 6.8 shows the median distance among the SOM vectors characterising each cell and its directly neighbouring cells (within a 3×3 square). Compared with the bottom panel in the same figure (where we have assigned each galaxy in the mock to a SOM cell), one can visually see that regions showing larger distances in the upper plot coincide

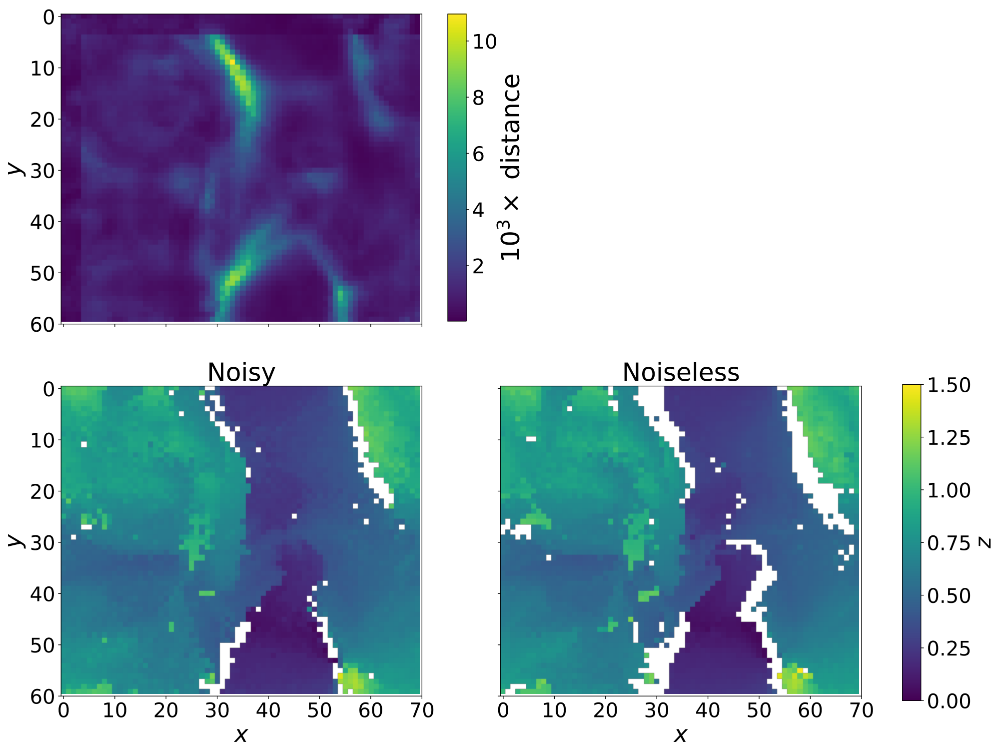

Figure 6.8: SOM trained on a galaxy simulated mock with the $uBVriz$ broad-bands. *Top:* Distance between every SOM cell vector and its 3×3 neighbours. *Bottom left*: Median photo-z in each SOM cell for noisy simulated galaxies. *Bottom right*: Median photo-z in each SOM cell for noiseless simulated galaxies.

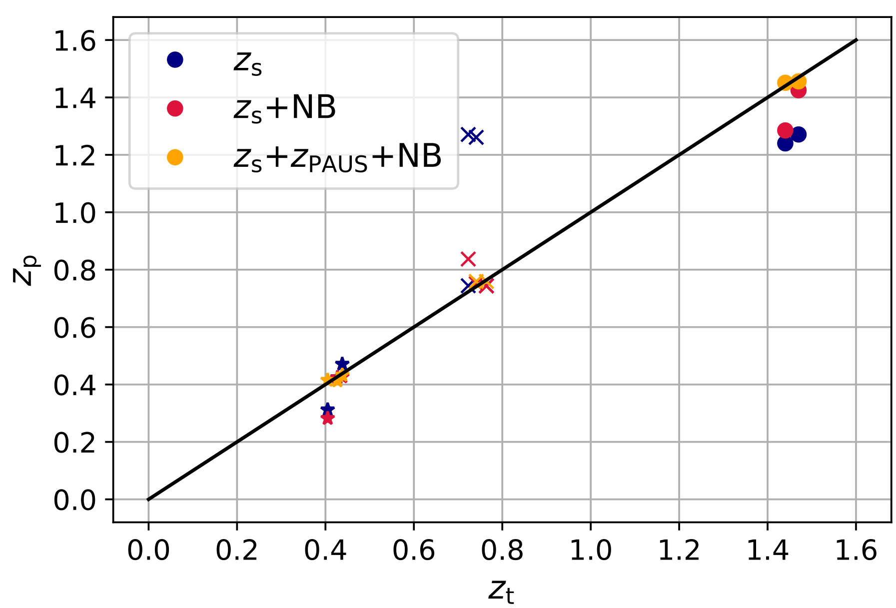

Figure 6.9: Photo-z scatter for galaxies in three independent SOM cells. The galaxies in each cell are represented with a different marker (stars, crosses and circles).

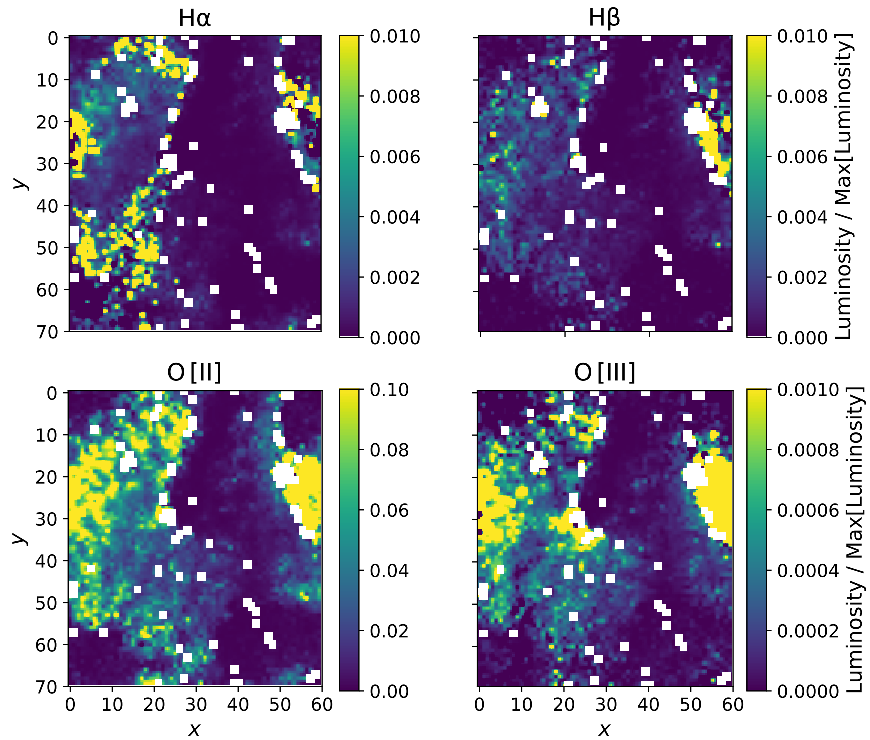

Figure 6.10: Emission line luminosity in colour space for Hα, Hβ, O[II] and O[III] as indicated in the title.

with empty regions (blank stripes) in the bottom ones. Therefore cells neighbouring empty colour-space regions represent noisier or outlier galaxies, whose colours differ from the rest of the galaxy sample.

To directly see the effect of noise in the SOM, the bottom row in Fig. 6.8 shows the colour-space redshift distribution for the noisy (left) and noiseless (right) colours of the same galaxies. Comparing the two panels demonstrates that the blank region between galaxy populations is broader in the noiseless case. When noise is included, cells on the edges of the empty regions in the right panel are populated. This, together with such cells being located further from the other cells in colour-space (top panel), indicates that cells neighbouring empty spaces describe a colour-space region that is not representative of the majority of the galaxy sample (e.g. very noisy galaxies or outliers), which can potentially cause broad-band colour-redshift degeneracies.

6.6.3 Emission line confusions

The SOM in Fig. 6.7 shows a region in colour-space that contains different galaxy populations, which indicates the presence of colour-redshift degeneracies. Figure 6.9 shows the photo-zs of the galaxies assigned to three different cells within such colour-space region. For the three cells (each of them represented with a different style marker), we plotted the predicted photo-

z ($z_{\rm p}$) and the true one ($z_{\rm t}$) with the baseline network ($z_{\rm s}$, blue), the network including MTL ($z_{\rm s}+NB$, red), and that including MTL and photo-z data augmentation ($z_{\rm s}$+NB+$z_{\rm PAUS}$, orange).

The first cell (marked with stars) contains galaxies with $z_{\rm t} \sim 0.4$. The baseline network (blue star) predicts a lower photo-z value $z_{\rm p} \sim 0.3$ for one galaxy, which is fixed with MTL+photo-z data augmentation (orange star). The second cell (marked with crosses) contains galaxies with $z_{\rm t} \sim 0.8$, that the baseline network estimates to be $z_{\rm p} \sim 1.2$. In contrast, the $z_{\rm s}+NB$ and the MTL+$\rm z_{PAUS}$ training methods are able to improve the photo-z estimates to values closer to the ground-truth. Lastly, the third cell (marked with dots) contains galaxies with redshifts $z_{\rm t} \sim 1.45$. The baseline network predicts these photo-zs around $z_{\rm p} \sim 1.25$, and again the $z_{\rm s}+NB$ and the MTL+$\rm z_{PAUS}$ training approaches are able to improve the photo-zs. Photo-z confusions from $z_{\rm t} \sim 0.8$ to $z_{\rm t} \sim 1.2$ and from $z_{\rm t} \sim 1.45$ to $z_{\rm t} \sim 1.25$ are recurrent, showing up at several SOM cells within the low photo-z performance cluster.

Figure 6.10 explores the mean Hα, Hβ, Oii, and Oiii emission line luminosity in colour space. The emission line luminosity is estimated as

$$L_{\rm el} \coloneqq 4\pi \; F_{\rm el} \; D_{\rm L}^2 \,, \tag{6.8}$$

where $F_{\rm el}$ is the emission line flux and $D_{\rm L}$ is the luminosity distance, which is estimated assuming Planck 2020 cosmology (Planck Collaboration et al., 2020). Emission line fluxes are taken from the photometry catalogue used for the PAUS+COSMOS photo-z (Alarcon et al., 2021), which were estimated by fitting the galaxy photometry to a template that modelled the emission line fluxes as a 10 Å wide Gaussian distribution.

Figure 6.10 shows strong emission lines at the low photo-z performance colour-space regions, e.g. the regions centred at $(5, 30)$ and $(55, 25)$. These results, together with the redshift confusions seen in Fig. 6.9, suggest that emission lines are likely to cause degeneracies in broad-band data.

Since a high ratio of Oiii to Hβ lines may indicate the presence of active galactic nuclei (AGN), we first verified that our galaxies do not host a Seyfert nucleus. The distribution of our sample on the "blue" emission-line diagnostic diagram (Lamareille, 2010) classify our sources as star-forming galaxies. Looking at the correlation of star-formation rates (SFR) and stellar masses, often called the main sequence (Whitaker et al., 2012), galaxies showing a photo-z mismatch from $z_{\rm t} \sim 0.8$ to $z_{\rm p} \sim 1.2$ occupy the starburst region (i.e. galaxies with enhanced star formation, Rodighiero et al., 2011). Furthermore, these two emission lines overlap at wavelengths between the i- and z-broad-band filters, which makes the emission line harder to detect.

Our findings suggest that some photometric features cause the photo-z mismatches. Emission lines have proven helpful to break colour-colour degeneracies and to improve the photo-z estimation (Csörnyei et al., 2021). Despite this, in some regions of colour parameter space emission line confusion is a potential cause for colour-redshift degeneracies.

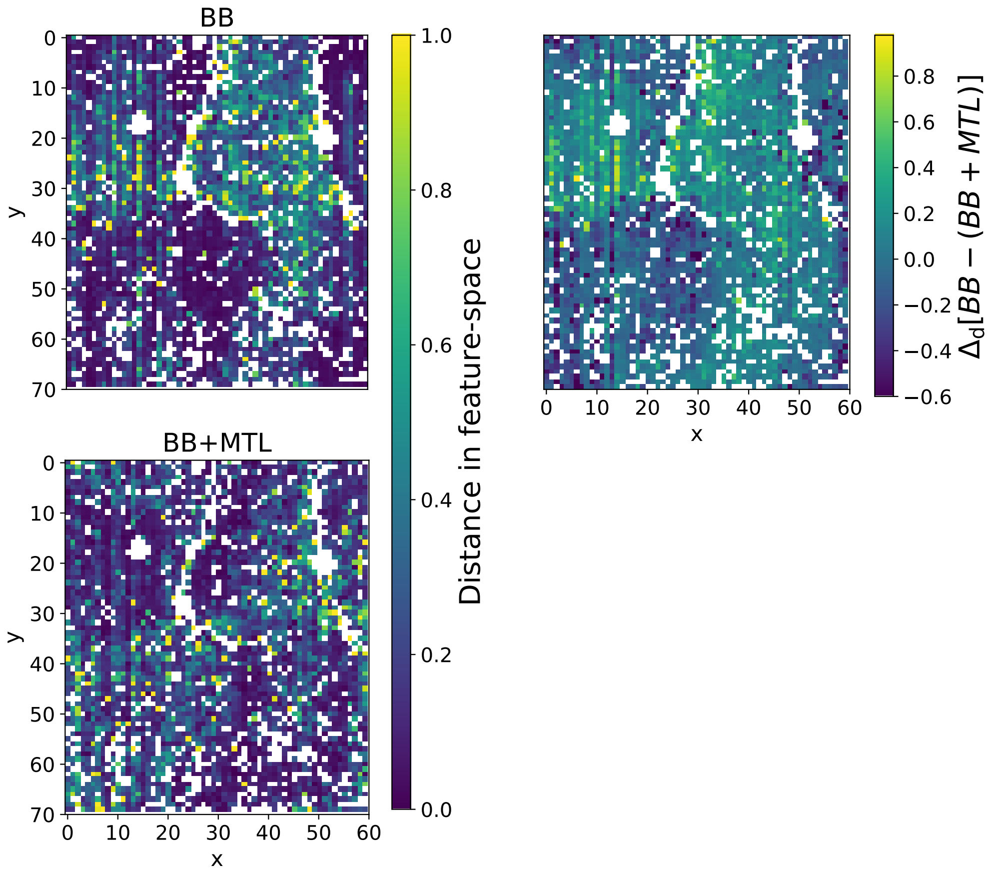

Figure 6.11: Distance in network's 2D feature space for COSMOS galaxies assigned to the same SOM cell. Feature space distances are normalised by the maximum of the distance evaluated in the plot. *Top left*: Distances when the network is trained only using the broad bands. *Bottom left*: Distances when the network training includes MTL with PAUS narrow-band photometry. *Top right*: Difference between the two left panels.

6.7 Understanding the MTL underlying mechanism

In this section, we aim to understand the underlying mechanism of MTL that improves the photo-z estimation. In § 6.7.1, we use a variation of our fiducial network to encode the galaxy photometry in a 2-dimensional space similar to a SOM, while in § 6.7.2 we study the impact of using other auxiliary tasks (other than predicting the narrow-band photometry) in the MTL network.

6.7.1 Underlying data representation in colour-space with MTL

For this test, we modify the fiducial network architecture (see § 6.3.2 and Fig. 6.1). The network architecture already encodes the information in a set of features, which then is used to both predict the photo-z and the narrow band fluxes. In the modified network we reduce the numbers of features to two numbers. Encoding the galaxy information in a 2D feature space simplifies its visualisation and brings it closer to the SOM colour-space representation, which we have already studied.

The galaxy representations in the 2D feature space must encode all the information needed to make the photo-z prediction. Furthermore, in the MTL network, those two numbers are also used to reconstruct the narrow-band photometry, thus these must also encode the relevant information for this task. Therefore, comparing the feature space representation of the baseline network (decoding only to the photo-z) and the MTL network (also predicting the narrow-band photometry) helps us to better understand why the MTL improves the photo-z estimates.

Figure 6.11 plots the mean mean-square distance in the 2D feature space among galaxies within the same SOM cell. The top left panel corresponds to distances assigned with the broad-band baseline network, while the bottom left panel corresponds to the MTL network. Galaxies assigned to the same SOM cell also cluster in the network's feature space, which indicates that the feature space encodes the input photometry in a similar way as the SOM.

As the network's feature space is not constrained, the network can encode the same galaxy differently in several independent trainings. Consequently, the coordinates assigned to each galaxy do not contain any valuable information by themselves and distances from different feature-space maps cannot be directly compared (e.g. the feature map of the ($z_{\rm s}$) network and that of the $z_{\rm s}+NB$). However, overlap in the feature-space coordinates that the network assigns to population of galaxies indicate degeneracies.

To compare the top and bottom left panels in Fig. 6.11, the results have been normalised to display distances between zero and unity. The top right panel in the figure shows the difference between the two left panels. Note that in the region with degenerate broad-band photometry $(55, 30)$, the feature space distance assigned by the MTL network is larger than in the broad-band only case. This suggests that MTL with PAUS narrow-bands assigns more distant feature space coordinates to galaxies with degenerate broad-band photometry, which effectively means that it is capable of learning a better representation breaking some degeneracies that the broad bands cannot break.

6.7.2 MTL with other galaxy parameters

So far in this paper, we explored how photo-z predictions benefit from MTL predicting PAUS narrow-band fluxes as an auxiliary task. However, MTL is a more general technique that could be exploited beyond narrow-band photometry reconstructions. While a conventional neural network training searches for the function (ϕ) that best predicts the photo-z (z) given the broad-band photometry (f), i.e. $\phi(z|f)$, with MTL the optimisation is extended to the function that best predicts the photo-z together with other related parameters ($x_{\rm i}$),

$$\phi(z, x_1, ..., x_{\rm N} \,|\, \theta)\,, \tag{6.9}$$

where $x_{\rm i}$ could be any galaxy parameter that correlates with the galaxy photo-z such as the galaxy type.

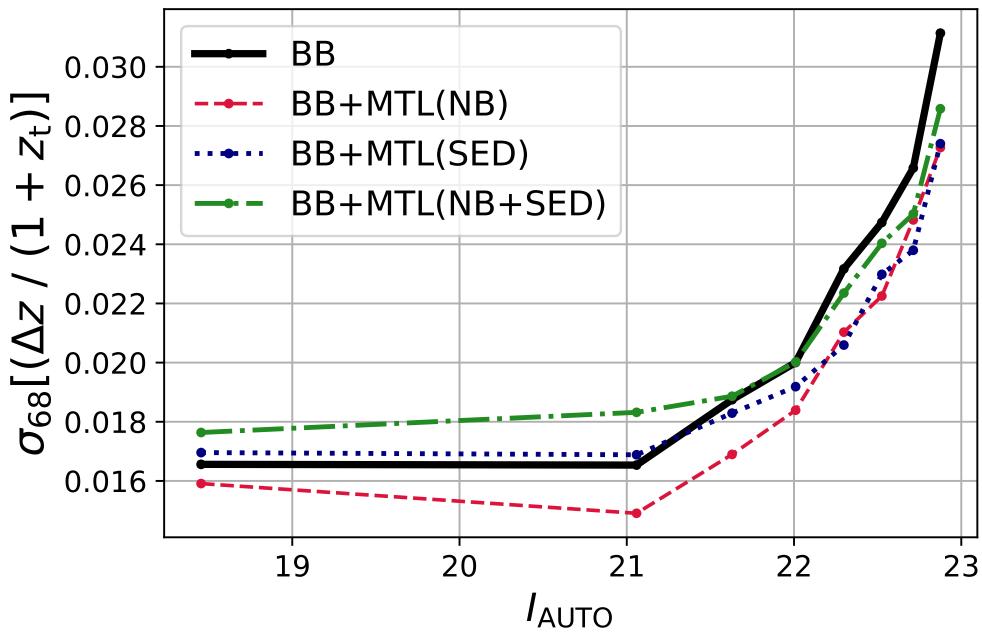

Figure 6.12: Photo-z precision in the COSMOS field including predicting the galaxy SED as auxiliary task. The galaxy SED prediction is addressed as a classification, where the true SED is a class between 1 and 47.

Template-fitting photo-z methods predict the joint probability distribution $p(z, t|f)$ of the redshift (z) and the galaxy type (t) and marginalise over the templates (Benítez, 2011). In principle, this is closely related to what MTL does when it is required to predict both quantities at the same time. The network looks for the function that better generalises the prediction of both parameters (e.g. type and redshift), but makes independent predictions in which it "marginalises" over the parameter it is not predicting.

Figure 6.12 shows the photo-z precision of data in the COSMOS field when the galaxy type is included as an MTL auxiliary task. The SED template is encoded as a discrete number between 1 and 47 as described in the COSMOS2015 catalogue. These correspond to 31 unique SEDs and 16 SEDs with different extinction laws. Including the SED template (dotted blue line) reduces the photo-z scatter with respect to the baseline network (solid black line). However, MTL using PAUS narrow-bands (dashed-red line) still provides better photo-z estimates. This result suggests that while the SED helps to produce a better representation of the data in colour-space (see § 6.7.1), PAUS narrow-band photometry contains information about the SED, as well as the emission lines or the extinction.

The dot-dashed green line in the same panel of Fig. 6.12 combines the SED and the narrow-band data using both as auxiliary tasks. We find that this degrades the photo-z performance with respect to using the SED or the narrow-band photometry solely. In theory, using both the narrow-band photometry and the SED number should benefit the network. However, the information available in these two tasks is highly correlated, which can hinder the predictions. Understanding this better is ongoing research and further study is deferred to future work.

We also explored MTL predicting galaxy parameters such as the SFR, the galaxy mass, and the $E(B-V)$ extinction parameter as auxiliary tasks (not shown). However, none of these parameters improved the predicted photo-zs. Furthermore, including the NIR photometry did not improve the photo-zs either.

6.7.3 Effect of narrow-band resolution

The improved photo-z from predicting the narrow-band photometry can potentially result from a better internal description of the galaxy SED type. We test this hypothesis by evaluating the performance of the networks using MTL for different resolutions of the output predicted photometry.

Figure 6.13 shows the photo-z precision of the MTL methods as a function of the number of predicted narrow bands, i.e. output photometry resolution. Assuming the MTL networks use the narrow band photometry to improve the internal representation of galaxies, increasing the output photometry resolution effectively corresponds to turning on this mechanism. To obtain lower resolution photometries, we take the mean of groups of consecutive narrow bands (e.g. 2, 4, 10). Then, we train the $z_{\rm s}$+NB and $z_{\rm s} + z_{\rm PAUS}$+NB methods several times to predict the photo-z and the photometry with a different number of bands in every training.

The horizontal flat lines in Fig. 6.13 indicate the photo-z precision for the the methods without MTL; $z_{\rm s}$ (dashed-dotted blue line) and $z_{\rm s} + z_{\rm PAUS}$ (solid red line). The dotted blue line and the dashed red line show the $z_{\rm s}$+NB and $z_{\rm s} + z_{\rm PAUS}$+NB performance for the different output photometry resolutions, respectively. As the output photometry resolution increases, the photometric redshift precision improves. This suggests that the MTL networks are using the narrow-band photometry prediction to improve the internal representation of the SED, and consequently the SED internal fitting, which has a direct impact in the photo-z prediction. The narrow-band photometry contains important additional information about the SED type and galaxy parameters, which are useful when predicting the redshift.

The $z_{\rm s}$+NB MTL produce for two bands predictions above the $z_{\rm s}$ line, which is the result without MTL. In this limit adding the photometry loss degrade the photo-z results. We trained this network several times to ensure the result was correct, obtaining the same degrading in all cases.

6.8 Discussion and conclusions

Photometric redshifts (photo-zs) are crucial to exploit ongoing and future large galaxy broad-band imaging surveys. While covering large sky areas, the broad-band spectral resolution limits the redshift performance through colour-redshift degeneracies. The PAU Survey (PAUS)

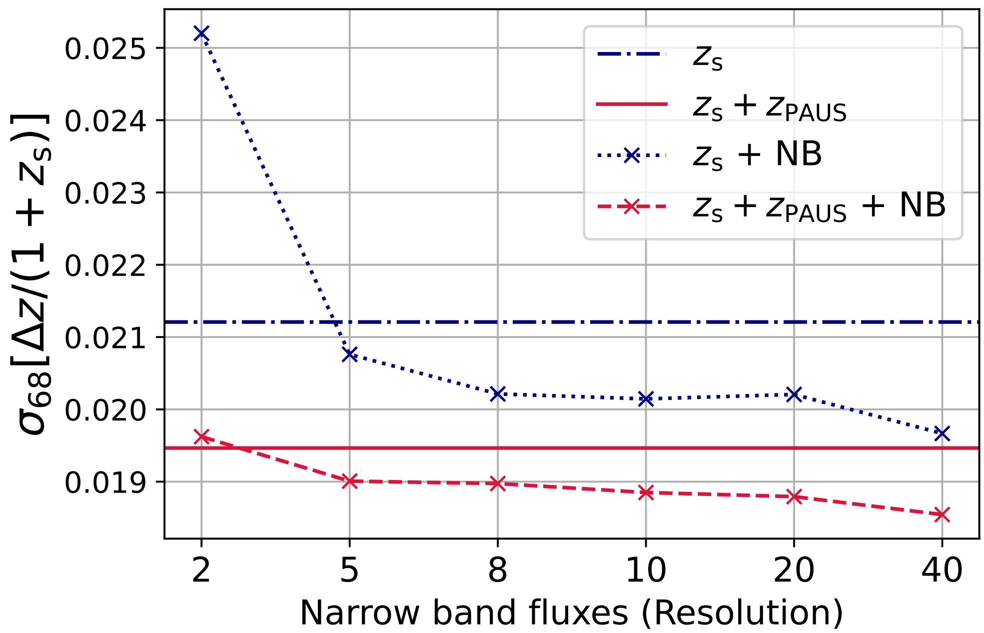

Figure 6.13: Photo-z precision as a function of number of bands in the predicted photometry for z_s +NB (blue dotted line) and $z_s + z_{\rm PAUS}$ +NB (dashed red line). The horizontal line corresponds to the z_s (dashed-dotted blue line) and $z_s + z_{\rm PAUS}$ (solid red line), where MTL is not enabled.

is a narrow-band imaging survey that can provide very precise photo-z measurements for a combination of wide and deep fields. In this paper we have introduced a new method to improve broad-band photo-z estimates, exploiting PAUS narrow-band data with deep learning techniques.

Multi-task learning is a machine-learning training methodology that aims to improve the performance and generalisation power of a network by training it on several related tasks simultaneously. This forces the model to share representations among related tasks, exploiting their commonalities and enabling the network to generalise better on the original task. We have implemented a multi-task learning network that predicts the photometric redshift and infers the narrow-band photometry simultaneously from the broad-band photometry (see § 6.3). The photo-z network is therefore forced to share parameters that are also used to predict the narrow-band photometry, which improves the internal colour-space representation of the data.

In the COSMOS field for galaxies to $i_{\rm AB} < 23$, our method reduces the photo-z scatter by approximately 16% (see § 6.4.2) and the number of photo-z outliers by about 40% (see § 6.4.3). The method also reduces the photo-z bias amongst high-redshift galaxies, where there is a lack of spectroscopic galaxies in the training sample, and improves the $N(z)$ distributions at these redshifts. We have also tested the potential of the method for fainter galaxies using *Euclid*-like galaxy simulations. For this, we have trained the network on a magnitude-limited sample with $i_{\rm AB} < 23$ and evaluated it on a sample with $i_{\rm AB} < 25$. The

MTL predicts up to 15% more precise photo-zs for galaxies with $24 < i_{\rm AB} < 25$ than the baseline network (see § 6.5).

We have used self-organising maps (SOMs) to study the photo-z performance in different colour-space regions, detecting a region containing galaxies with degenerate photometry-redshift mappings. This region has a larger photo-z variation within the SOM cells, suggesting that more than one galaxy population is assigned to the same colour-space location (see left panel in Fig. 6.7). This correlation results in a photo-z mismatch between two galaxy populations, which affects broad-band photo-z estimates. Our MTL network breaks some degeneracies using PAUS narrow-band data to learn the underlying colour-space distribution of galaxies.

This chapter explores how to exploit data from narrow-band photometric surveys like PAUS to improve the broad-band photo-z estimates using machine learning. The key point of using MTL, instead of, e.g. just using the narrow-band photometry to obtain more precise photo-zs, is that it only requires narrow-band photometry for the training galaxies, while the photo-z of any galaxy can be evaluated with only the broad-band data. This enables exploiting fields where we have narrow-band data to obtain better photo-zs in other fields where these are not available. PAUS photometry in the COSMOS field is publicly available so that current and future weak lensing surveys, like *Euclid* or the LSST, could readily benefit from this methodology to improve their photo-z estimates. Moreover, MTL is a general machine learning mechanism that enables fields with different types of photometry to be exploited for improving photo-z predictions. While PAUS narrow-band photometry is a clear candidate, other surveys like J-PAS (Benitez et al., 2014) or ALHAMBRA (Moles et al., 2008) provide more fields with interesting data to exploit for the benefit of photo-z.

6.A Self-organising maps

A self-organising map (SOM, Kohonen, 1982) is an unsupervised machine learning algorithm trained to produce a low-dimensional (typically two-dimensional) representation of a multi-dimensional space. A two-dimensional SOM contains $N_x \times N_y$ cells with an associated vector of attributes ($\vec{w}^k$), where $N_x(N_y)$ is the dimension of the SOM on the $x(y)$-axis, and k corresponds to the kth SOM cell. Each of these vectors has the same length as the input data.

The SOM training phase is an iterative process during which the SOM cells compete amongst themselves to represent the training data. Initially, the cell vectors ($\vec{w}^k$) are randomly sampled from a uniform distribution, and these are updated after each iteration step (t). In every training iteration, each galaxy vector of measured attributes $\vec{x}$ (e.g., in our case the galaxy colours), is compared to all the SOM cells' vectors via a χ^2 expression,

$$\chi^2 \left(\vec{w}^k(t), \vec{x}\right) = \sum_i \left[\frac{x_i - w_i^k(t)}{\sigma_i}\right]^2 , \tag{6.10}$$

where i sums over galaxy attributes and σ_i is the uncertainty associated with x_i. The evaluated galaxy is assigned to the cell with the lowest χ^2, which updates its associated vector of

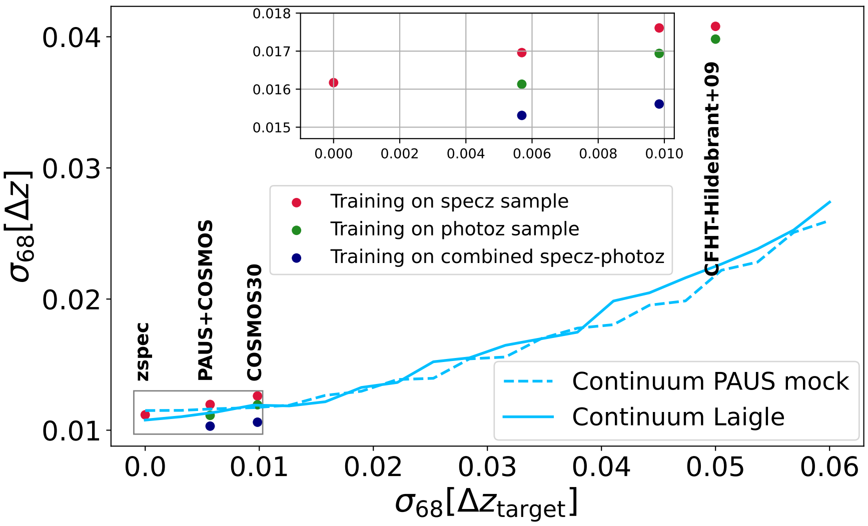

Figure 6.14: Photo-z performance as a function of the ground-truth redshift precision used for training the networks. The training redshifts are the spectroscopic redshifts, the PAUS+COSMOS photo-zs, COSMOS30, and a set of CFHT photo-zs in COSMOS. Red points correspond to training on the spectroscopic sample (around 6000 galaxies). In the green and red points, the training sample is extended to COSMOS galaxies with photo-z (around 15 000 galaxies). The blue lines show the expected photo-z performance as a function of target redshift precision. The true redshifts, spectroscopic redshift in the COSMOS2015 catalogue (blue solid line), and simulated redshift in the PAUS mock (blue dashed line) are scattered with precision in 0.001 bins. The top inset zooms the framed area in the main plot (lower left corner)

attributes $\vec{w}^k(t)$ according to the matched galaxy features

Furthermore, in the SOM training procedure, the vector of features from cells neighbouring the best matching cell are also updated, clustering together galaxies with similar attributes. This is implemented with a neighbouring function $H(t, d)$, which depends on the distance (d) between the best matching cell and the updated one. The neighbouring function is commonly implemented as a Gaussian kernel with an iteration-dependent variance $\sigma^2_{\mathrm{kernel}}(t)$. Therefore, the vector of attributes for a particular cell k after iteration $t+1$ is

$$\vec{w}^k(t+1) = \vec{w}^k(t) \;+\; \alpha(t)\; H\left(t, |\vec{w} - \vec{x}|\right)\; \left(\vec{x} - \vec{w}^k(t)\right)\;, \tag{6.11}$$

where $\alpha(t)$ is the learning rate. After a few iterations over the training sample, the result is a map of $(N_x \times N_y)$ vectors in a two-dimensional space grouping together cells with similar features while preserving the topology of the multi-dimensional space.

6.B Effect of training with photo-z as ground-truth targets

In this work, we have implemented and tested two training methodologies that rely on narrow-band photo-z estimates as ground-truth targets (see methods $z_{\mathrm{s}} + z_{\mathrm{PAUS}}$ and NB+$z_{\mathrm{s}} + z_{\mathrm{PAUS}}$

in § 6.3.2). Even if such photo-zs are overall very accurate, its implementation in the training could harm the photo-z performance since these are less precise than spectroscopic redshifts and could potentially include outliers. In this section, we explore the effect that less precise redshift labels (§ 6.B.1) and the presence of outliers (§ 6.B.2) have on the photo-z performance.

6.B.1 Effect of photo-z dispersion in the training redshifts

Figure 6.14 shows the photo-z precision of a set of 1000 spectroscopic galaxies for four independent broad-band networks (simply mapping colours to redshift), each of them trained with different ground-truth redshifts. The redshifts used for training are the spectroscopic redshifts (see § 6.2.4), the PAUS+COSMOS photo-zs (see § 6.2.2), the COSMOS30 photo-zs (Laigle et al., 2016), which combine 30 photometric filters and estimates the photo-z with `Lephare` (Arnouts & Ilbert, 2011), and a set of CFHT photo-zs from Hildebrandt et al. (2012b) combining six broad bands ($ugriz$) with photo-z estimated with `BPz` (Benítez, 2011). The input data are, in all cases, the CFHT u band and the $BVriz$ Subaru broad-band filters from COSMOS2015.

The red points in Fig. 6.14 show the redshift dispersion using a training sample of galaxies with spectroscopic redshift. We always keep the same training sample (which contains around 6000 galaxies) and change the labelled true redshifts in each independent training (spectroscopic catalogue, PAUS data, COSMOS30, and the CFHT catalogue). Using spectroscopic redshifts as ground-truth redshifts results in a dispersion of $\sigma_{68} = 0.016$. Replacing the spectroscopic redshift with the photo-z from PAUS+COSMOS, COSMOS30, or CFHT yields $\sigma_{68} = 0.017$, $\sigma_{68} = 0.018$, and $\sigma_{68} = 0.046$, respectively. As the ground-truth redshifts become less precise, the machine-learning photo-z performance degrades.

To obtain the green points (Fig. 6.14), we extended the training sample to all galaxies in the COSMOS sample with a photo-z estimate, which results in approximately 15 000 galaxies when the four catalogues are merged. Then, three independent networks are trained using the PAUS+COSMOS, the COSMOS-30, and the CFHT photo-zs as true redshifts (the spectroscopic redshift is not used even if it is available). This provides a precision of $\sigma_{68} = 0.016$, $\sigma_{68} = 0.017$, and $\sigma_{68} = 0.045$ for the PAUS+COSMOS, the COSMOS30, and the CFHT photo-zs, respectively. The three networks improve the photo-z precision with respect to training with spectroscopic redshifts only. Indeed, with the PAUS+COSMOS photo-z labels we already reach the photo-z precision with spectroscopic labels.

Finally, the blue points in the figure correspond to the networks trained with the same 15 000 photo-z galaxies as in the green points, but combining spectroscopic redshifts (if available) and photo-zs as ground-truth training redshifts. Combining spectroscopic redshifts with PAUS+COSMOS photo-zs yields $\sigma_{68} = 0.015$, which improves upon the precision obtained with spectroscopic redshifts only.

The light blue lines in Fig. 6.14 show the expected performance as a function of the ground-truth redshift precision. The solid line uses the COSMOS2015 $uBVriz$ broad bands and the dashed one uses simulated data from the PAUS mock described in § 6.5. In both cases, true

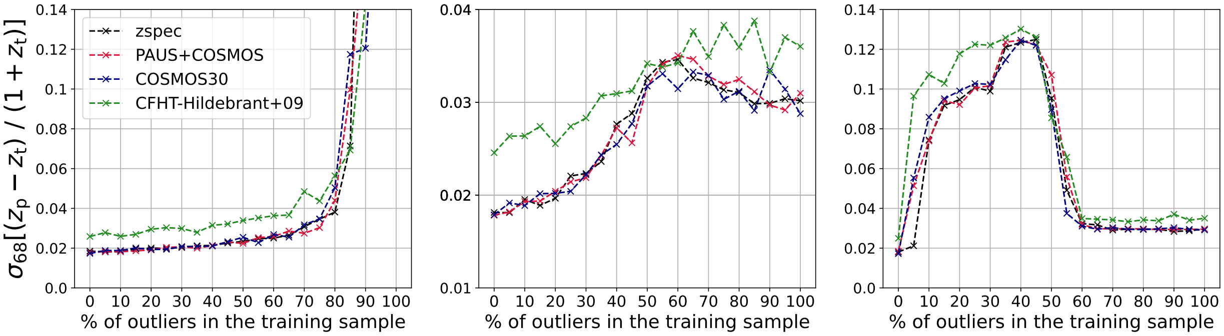

Figure 6.15: Effect of outliers and systematic errors in the ground-truth redshift sample used during training. The training sample consists of 5000 spectroscopic galaxies with photometry from COSMOS2015. Each coloured line uses a different sample of redshifts as true redshifts, i.e. spectroscopic redshifts (black), PAUS+COSMOS photo-zs (red), COSMOS30 photo-zs (blue), and CFHT photo-zs (green). The ground-truth redshift of the selected fraction of training galaxies is replaced by: *left*, a random redshift values sampled from $U(0, 1.5)$; *centre*, a 20% higher redshift; and *right*, redshifts modified with Eq. (6.12).

redshifts (spectroscopic or simulated) are scattered with the corresponding dispersion in the abscissa.

Both networks (solid and dashed lines) are trained with 15 000 galaxies to have a direct comparison with the previous results. We always use the scattered redshifts as ground-truth targets, in such a way that the lines should be compared with the green points since these are trained using only photometric redshifts. The results obtained with the PAUS+COSMOS and COSMOS30 match the expectation curves, but there is a significant mismatch with the CFHT photo-zs. This is potentially triggered by systematic errors or outliers in the CFHT photo-z not represented in the blue curves (see Appendix 6.B.2 for more details).

6.B.2 Effect of photo-z outliers in the training redshifts

Figure 6.14 showed a mismatch between the expected (solid blue curve) and photo-z performance training a network with the CFHT photo-zs as ground-truth redshifts (rightmost red point). However, the expectation assumes that the CFHT photo-zs are not affected by other effects such as systematic errors or catastrophic outliers.

Figure 6.15 shows the effect of outliers in the ground-truth targets of the training sample. The network is trained twenty independent times with 5000 COSMOS2015 spectroscopic galaxies including a fraction (monotonically increasing in each iteration) of labelled photo-z outliers. This procedure is repeated for the spectroscopic redshifts (black line), the PAUS+COSMOS photo-zs (red line), the COSMOS30 photo-zs (blue line), and the CFHT photo-zs (green line).

In the left panel, the artificial outlier redshifts are swapped with a random value sampled from a uniform distribution $U(0, 1.5)$ to simulate catastrophic outliers. The predicted photo-z

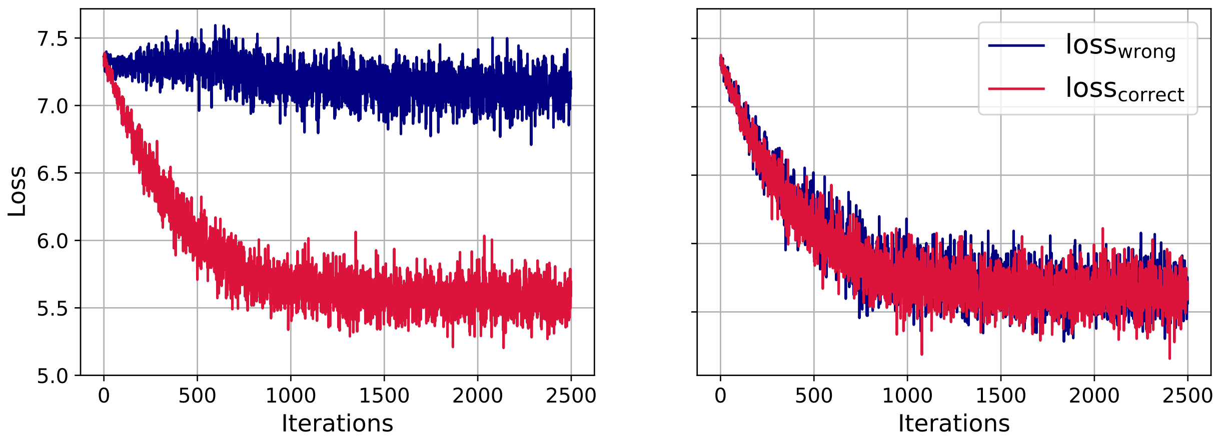

Figure 6.16: Training loss function for galaxies with wrong (blue) and corrected (red) target redshift. The training sample consists of 5000 spectroscopic galaxies with photometry from COSMOS2015. In the left panel, the modified target redshifts are randomly switched to a value drawn from $U(0, 1.5)$, while in the right panel the wrong redshift labels are generated with Eq. (6.12).

precision degrades as the fraction of target redshift outliers increases. This also affects the predicted $p(z)$, which become noisier and broader (not shown). However, unexpectedly the network can provide reasonable photo-z estimates with up to 80% of catastrophic outliers in the training sample. Furthermore, the network is able to make reliable photo-z predictions of galaxies that have been used in the training sample with wrong target redshift values. This result holds when either spectroscopic redshifts or any of the photo-zs are used for training.

The middle panel shows the effect of a systematic multiplicative shift in the training sample redshifts, where the selected targets are shifted to 20% higher redshifts. In this scenario, the predicted photo-z precision degrades faster than when outliers are random (left panel) but the network does never completely break. For an outlier fraction higher than 60%, the precision settles at $\sigma_{68} = 0.03$, but the bias rapidly increases. Finally, the rightmost panel presents the effect of a systematic shifting the redshift ($z_{\rm mod}$) so that Oiii is confused with Hα in the training redshifts, i.e.

$$z_{\rm mod} = \lambda_{\rm Oiii}/\lambda_{\rm H\alpha}\,(1 + z_{\rm t}) - 1\;, \tag{6.12}$$

where $z_{\rm t}$ is the galaxy redshift.

The training degrades and breaks much faster than in the two previous cases, where with around 40% of wrong target redshifts the network is not able to provide reliable predictions. As the fraction of affected target redshifts increases, the predicted $p(z)$ become more doubly peaked. Moreover, a plot of photo-z versus spec-z scatter displays two clear lines, one with the correct mapping and another shifted upwards (not shown), which is triggered by the training objects with the photo-z artificially shifted to confuse the emission lines. Again, the effect of outliers is similar regardless of the redshifts used for training (spectroscopic or different-precision photo-z).

Contrary to expectations, the left panel of Fig. 6.15 indicates that the network can learn the mapping between the galaxy photometry and redshifts with up to 80% of catastrophic outliers in the training sample. Given that the training sample is composed of 5000 galaxies, this means that the network can effectively learn the colour-redshift relations from 1000 galaxies, learning to ignore the remaining 4000 spurious galaxies.

Figure 6.16 shows the cost function evolution of a network trained with wrong target redshifts for half of the training while keeping the rest to the correct redshift values. The cost function is split in two; one for those objects with correct redshift (red) and another for those with wrong redshifts (blue). In the left panel, the modified target redshifts are switched to a random value from a uniform distribution $U(0, 1.5)$, as in the left panel of Fig. 6.15. The cost function of galaxies with the correct target redshift decreases, which indicates that the network is learning from them. In contrast, the cost function of incorrectly labelled galaxies remains constant along the training, showing that the network is not learning anything from them. Therefore, the network is effectively only learning from galaxies with correct target redshifts. Randomly swapping redshifts to different values breaks any correlation between the photometry and the redshifts. Hence, the network is only learning the colour-redshift mapping from galaxies with the correct target redshift. Nevertheless, having a large fraction of wrong labels adds noise to the training, broadening the predicted $p(z)$.

The right panel of Fig. 6.16 shows the loss function for the correct and the wrongly labelled training galaxies separately when the incorrect redshift labels are generated with Eq. (6.12). This introduces a new colour-redshift relation that forces the network to learn both from galaxies with wrong and correct target redshifts. This can also be noted in the $p(z)$ behaviour, which presents a double-peaked distribution (not shown). Hence, Figs. 6.15 and 6.16 indicate that having catastrophic outliers in the training sample labels effectively adds noise to the photo-z predictions. In contrast, a systematic bias in the training sample targets produces a bias in such predictions.

Chapter 7

Multi-band photometry and photometric redshifts

7.1 Motivation

Imaging galaxy surveys require precise photometry and photometric redshifts measurements (photo-zs) for a extensive set of science applications, e.g. weak gravitational lensing (Hoekstra & Jain, 2008; Abbott et al., 2018a; Giblin et al., 2021), galaxy clustering (Elvin-Poole et al., 2018b; Giocoli et al., 2021), and intrinsic alignments (Joachimi & Schneider, 2010; Johnston et al., 2019; Samuroff et al., 2019). In recent years, the amount and quality of galaxy survey data have drastically increased, requiring more precise and efficient methods for data analysis. Many efforts have been devoted to this task, resulting in a vast number of optimised methods.

`SExtractor` (Bertin & Arnouts, 1996) is a very widely used algorithm to measure the photometry, which implements aperture photometry (Mighell, 1999). Other examples include `GaaP` (Kuijken, 2008), a model-fitting algorithm that extracts the photometry fitting the galaxy image to polar shapelets (Refregier, 2003; Massey & Refregier, 2005), `Lumos`, (Cabayol et al., 2021), a deep learning network measuring the photometry probability distribution, `T-PHOT` (Merlin et al., 2015), and `Tractor` (Lang et al., 2016).

There are two well-known techniques to estimate photo-zs: template-based spectral energy distribution (SED) fitting and data-driven methods. Template-based methods fit a set of redshifted SED templates to the galaxy photometry using a χ^2 minimisation, e.g. `BPz` (Benítez, 2011) and `LEPHARE` (Arnouts & Ilbert, 2011). In contrast, machine learning techniques use a collection of known-redshift galaxies to train a model to map the galaxy photometry to redshifts, e.g. (Collister & Lahav, 2004; Bonnett, 2015b). Both techniques have known limitations, e.g. the quality of fitting templates and the depth of the training sample (Salvato et al., 2019b).

In astronomy, deep learning has already been implemented to several problems, including object classification (Kim & Brunner, 2016; Khalifa et al., 2017; Cabayol et al., 2019), weak lensing (Tewes et al., 2019; Matilla et al., 2020), and the data reduction process (Zhang & Bloom, 2019; Boucaud et al., 2020). Deep learning is also extensively applied to the photo-z

estimation problem. Traditionally, photo-z algorithms have relied on multi-band photometry, i.e. the measured flux from the galaxy exposure images in each band (Feldmann et al., 2006; Eriksen et al., 2020). However, some deep learning approaches implement convolutional neural networks (CNNs) directly on the galaxy images (Pasquet-Itam & Pasquet, 2018; D'Isanto & Polsterer, 2018; Pasquet et al., 2019). This opens a new field of study where neural networks can use all the information available in the images to improve the photo-z predictions.

In broad-band surveys, the number and width of the photometric filters limit the photo-z performance. In contrast, narrow-band imaging surveys provide an alternative that enables high precision photo-z estimates for a large sample of galaxies. The Physics of the Accelerating Universe Survey (PAUS) is an imaging survey observing with a 40 narrow-band photometric filters camera (PAUCam, Padilla et al., 2019b) covering the optical wavelength range from 4500 Å to 8500 Å. Eriksen et al. (2019) presents `BCNz`, the first custom photo-z algorithm for PAUS data. It is a template-fitting algorithm that reached a precision of $\sigma_{\mathrm{z}} = 0.0037(1{+}z)$ for the best selected 50% of the sample with $i_{\mathrm{AB}}{<}22.5$. Typically, the photo-z precision with broad-band optical photometry is of the order of $\sigma_{\mathrm{z}} = 0.05(1{+}z)$ (Hildebrandt et al., 2009). Similar results are obtained with a hybrid machine learning and template-fitting method running a Gaussian process algorithm (Soo et al., 2021). Furthermore, a deep learning algorithm that implements transfer learning from simulations to PAUS data (Eriksen et al., 2020) improves the photo-z precision by 50%.

In Cabayol et al. (2021), we developed `Lumos`, a deep learning based algorithm to predict the flux probability distribution of single-exposure galaxy images. On PAUS observations in the COSMOS field, `Lumos` increases the photometry signal-to-noise by a factor of 2 compared to an aperture photometry algorithm and has proven more robust towards distorting effects as e.g. blended galaxies, cosmic rays, and scattered-light (Cabayol et al., 2021). `Lumos` is built on `BKGnet` (Cabayol-Garcia et al., 2020), which predicts the astronomical background light in the presence of strong varying noise.

This chapter presents `Aczio`[1], a deep neural network that culminates the work of `Lumos` and `Deepz` (Chapters § 4 and 5, respectively) predicting both the galaxy multi-band photometry and the photometric redshift directly from their images, in contrast to using fluxes. Recently, there have been multiple works on estimating the photo-z directly from the images observations(D'Isanto & Polsterer, 2018; Pasquet et al., 2019; Dey et al., 2021b). However, to the best of our knowledge, our method is the first to predict both the photometry and the photo-z simultaneously from the galaxy images. This enables a fast and precise evaluation of the photo-z, drastically reducing the data-reduction process while also having photometry measurements. This is important since science applications as e.g. galaxy evolution studies also require precise photometry measurements. Furthermore, measuring the photometry and the photo-z simultaneously, we are implementing a multi-task learning network (Caruana, 1997), where the network can use the correlation between both quantities to improve their measurements. In Cabayol et al. (in prep), we already tested the benefits of predicting the

[1]Name inspired in the Harry Potter charm "Accio", which summons an object towards the caster. It can summon objects in direct line of sight of the caster, as well as things out of view.

photometry and the photo-z simultaneously using broad-band data.

This chapter is structured as follows. In § 7.2 we present the PAUS data (§ 7.2.1) and the simulations used for training (§ 7.2.3). Section 7.3 presents the method, including the network's architecture (§7.3.1 and §7.3.2) and the training procedure (§7.3.3). Then, § 7.4 shows the photometry obtained with `Aczio` on simulations (§ 7.4.1) and on the PAUS images in COSMOS (§ 7.4.2). In § 7.5, we present the `Aczio` photo-zs in the COSMOS field. Finally §7.6 concludes with a discussion on this work.

7.2 Data

In this section, we present the PAUS data (§ 7.2.1) and the TEAHUPOO image simulations used throughout the chapter (§ 7.2.3).

7.2.1 PAUS data

This chapter also focuses on data from the COSMOS field (Lilly et al., 2009, ,§ 3.4.4). These are the data used in the BKGNET and `Lumos` papers and also in all the photo-z studies published so far (Eriksen et al., 2019; Eriksen et al., 2020; Soo et al., 2021; Alarcon et al., 2021). The complete photometry catalogue in COSMOS contains 64 476 galaxies to $i_{\rm AB} < 23$ in 40 narrow-band filters. This corresponds to around 12,5 million galaxy observations, with approximately five observations per galaxy and narrow-band filter. PAUS data reduction process consists of two pipelines: the NIGHTLY and the MEMBA pipelines (§ 3.4.1&§ 3.4.2). The NIGHTLY pipeline is the first data-reduction step. It first performs instrumental de-trending, correcting the electronic and illumination biases. The MEMBA pipeline implements aperture photometry to targets selected from a deeper detection catalogue. Then, it combines the photometry of independent observations of the same galaxy in the same narrow-band filter using an inverse-variance weighted sum.

7.2.2 Spectroscopic sample

We use the spectroscopic sample to train and evaluate the performance of `Aczio`. The spectroscopic redshifts are those compiled in (Alarcon et al., 2021, ,from hereafter PAUS+COSMOS) which includes the zCOSMOS DR3 bright spectroscopic data (Lilly et al., 2009), redshifts from C3R2 (DR1&DR2 Masters et al., 2017, 2019), 2dF (Colless et al., 2001b), DEIMOS (Hasinger et al., 2018), FMOS (Kashino et al., 2019), LRIS (Lee et al., 2018), MOSFIRE (Kriek et al., 2015), MUSE (Urrutia et al., 2019), Magellan (Calabrò et al., 2018), and VIS3COS (Paulino-Afonso et al., 2018). Lilly et al. (2009) defines a confidence class ($conf$) for the spectroscopic redshifts, where the. For the zCOSMOS sample, the most secure redshifts are those with $3 < conf < 5$ (Lilly et al., 2009). The rest of the spectroscopic redshifts are already selected to keep only objects with a very reliable measurement.

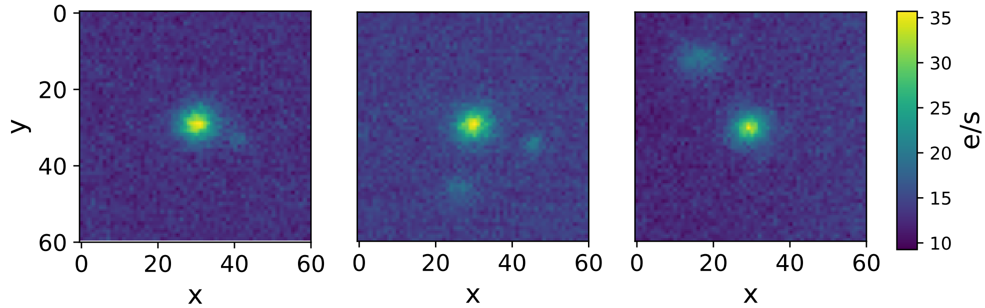

Figure 7.1: Three TEAHUPOO observations images of the same galaxy bright galaxy ($i_{\mathrm{AB}} < 19$) in the same narrow-band filter.

7.2.3 PAUS image simulations

The TEAHUPOO simulations are a set of single-observation PAUS-like galaxy images specifically built to train `Lumos`, a CNN to extract the photometry from single exposure images. These image simulations combine a randomly selected 60×60 PAUCam cutout and a single-Sérsic modelled galaxy profile (Sérsic, 1968) simulated with `Astropy` (Price-Whelan et al., 2018). In this paper, we extend the TEAHUPOO image simulations to combine bulge and disk modellings. The bulge is defined by a Sérsic profile with Sérsic index $n_{\mathrm{s}} > 1$, i.e.

$$I_{\mathrm{b}}(R) = I_{\mathrm{e}} \exp\left\{-(2n_{\mathrm{s}} - 1/3)\left[\left(\frac{R}{R_{50}}\right)^{1/n_{\mathrm{s}}} - 1\right]\right\}, \tag{7.1}$$

where I_{e} and r_{50} are the the surface brightness and the half-light radius. In contrast, the disk profile ($I_{\mathrm{d}}(R)$) corresponds to an exponential galaxy profile, i.e. the Sérsic profile (Eq. 7.1) with $n = 1$. The two components are combined into the galaxy profile (I) as

$$I(R) = B/T \cdot I_{\mathrm{b}}(R) + (1 - B/T) \cdot I_{\mathrm{d}}(R)\,, \tag{7.2}$$

where B/T is the bulge fraction and determines the relative contribution of each profile type. The TEAHUPOO galaxies are also rotated with the rotation angle θ and elongated with the ellipticity parameter $e = b/a$, where a and b are the major and minor axis of the galaxy, respectively.

The galaxy parameters (e.g. n_{s}, θ, and e) are generated with the same pipeline as the Flagship simulations (Castander et al in prep.). To simulate the galaxy brightness, we use noiseless fluxes that include the milky-way extinction. While the image simulations are 60×60 pixels, we initially draw the galaxies in a 600×600 grid, which is later reduced by a factor of ten. The higher resolution grid increases the simulated galaxy resolution and enables galaxies with sub-pixels shifts from the centre of the simulated stamp. For `Lumos`, this proved important to reduce the number of photometry outliers on PAUCam data.

In order to include background noise in our galaxy images, we stack the simulated galaxies on 60×60 PAUCam cutouts randomly selected from an observed PAUCam image. The background cutout must be from the same narrow-band filter as the simulated galaxy, since the noise patterns and the background level varies amongst filters. These cutouts include all sources of noise present in the real images, e.g. scattered light, transient effects, cross-talk signals, and read-out noise. Covering all these effects in the training simulations was crucial to make good photometry predictions on PAUCam data with `Lumos`. There, we showed that the network was capable of deblending target galaxies by just randomly including examples in the training sample. The galaxy is convolved with the same point spread function (PSF) as detected in the original PAUCam image before stacking the simulated profile and the background noise.

In this work, we also extend the Teahupoo simulations to include multiple observations of the same galaxy in the same narrow-band filter. Figure 7.1 shows an example of three simulated and calibrated observations of a bright galaxy ($i_{\mathrm{AB}} < 19$) in the PAUS narrow-band "NB685". In the simulations we have so far, observations of the same galaxy can potentially contain other sources at the different locations. This a weakness of our simulations that will soon address.

To simulate how the galaxy looks before the zero-point correction, we divide the noiseless flux by the zero point measured in the background cutout image. This results in three single-observation flux measurements of the same galaxy in the same narrow-band filter, from which we generate three Teahupoo image simulations. The simulated images (galaxy+background cutout) are then re-scaled with the same zero point as previously used to derive the single-observation flux measurement. To account for the uncertainty in the zero-point measurement, we include a 5% scatter to the zero points before calibrating the single-observation images, which is a reasonable number based on the observations. Furthermore, we also include a 2% of zero-point outliers to increase the robustness of the method.

7.3 Network architecture and training procedure

In this section, we introduce `Aczio`, a deep neural network to predict the multi-band photometry and the photo-z. `Aczio` has two different parts; a convolutional neural network (CNN, LeCun et al., 1989; Lecun et al., 1998) to extract information from the galaxy images (§ 7.3.1) and two mixture-density networks (MDN, Bishop, 1994) to predict the probability distribution of the flux measurements and the photo-zs (§ 7.3.2). Section § 7.3.3 describe the `Aczio`.

7.3.1 Feature extraction with a convolutional neural network

Convolutional neural networks (CNNs) have proven successful for multiple image-related applications, e.g. image classification (Sultana et al., 2019), image semantic segmentation (Liu et al., 2018), and object detection (Zhao et al., 2018). Figure 7.2 shows the `Aczio` architecture. The blue framed area corresponds to the feature extractor. It is composed of a 2D CNN (blue cube) followed by a multi-layer perceptron (MLP, blue circles) and a 1D CNN

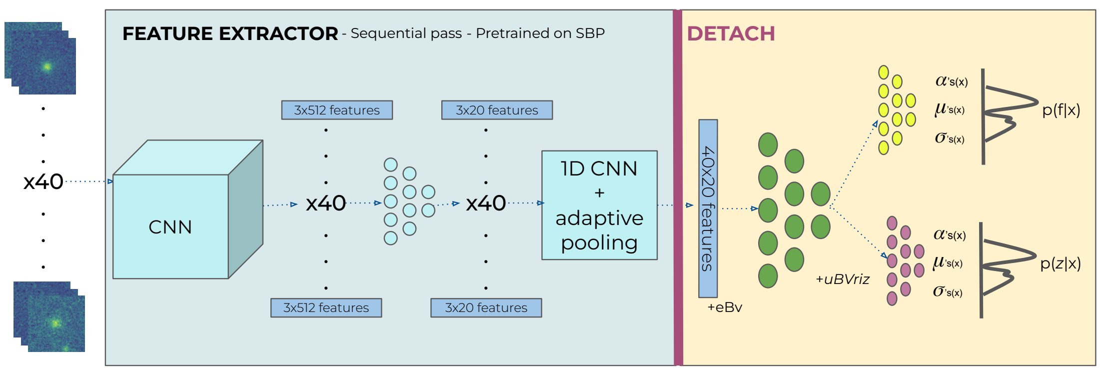

Figure 7.2: `Aczio` architecture. The blue shaded area corresponds to the "feature extractor" network. It is composed of a 2D-CNN (blue cube), a multi-layer perceptron (blue circles), and a 1D-CNN with an adaptive pooling layer (blue square). The second part of the network corresponds to a MTL-MDN network that predicts the photo-z and the photometry in one band. These two tasks share a set of common layers (green circles) and then have two sets of task-specific layers for the photo-z prediction (red circles) and the photometry (yellow circles).

with adaptive pooling (blue square). The 2D CNN is composed of five blocks of convolutional and pooling layers. After each block, there is a ReLU *activation function* (Nair & Hinton, 2010). We provide the 2D CNN with N tensors of size $(N_{\rm obs}, 60, 60)$, where $N_{\rm obs}$ is the number of observations per galaxy and narrow-band filter. In the training phase, the number of observations per galaxy and narrow band is fixed to three. The input images pass sequentially through the 2D CNN, which extracts 512 features per image. The MLP reduces the 512 features to 20 so that the network keeps the most relevant features. These 20 features encode the relevant image properties required for predicting the photometry and photo-z. We assume that the network should, at least, encode the main profile properties, e.g. the size, shape, rotation, ellipticity, and total flux, together with background-noise properties as the mean, the presence of nearby sources, cosmic-rays, or the position and brightness of scattered light. Reducing the number of features is also important to avoid running out of GPU memory. The training sample contains 120 simulated observations per galaxy, which extracting 512 features would correspond to 614 400 features per galaxy.

After the MLP, the features from different observations of the same galaxy and narrow-band filter concatenate into a tensor of size ($N_{\rm obs}$, 20). This tensor is the input of the 1D CNN that combines the multiple-observation features into a single set of 20 co-added features. To deal with the varying number of astronomical observations varies per galaxy and narrow-band filter, we use an adaptive-pooling layer (Liu et al., 2016) after the 1D CNN. Adaptive-pooling layers fix the output size given any input size. This allows us to fix the output size to 20 features while enabling variable $N_{\rm obs}$ in the input. Adaptive pooling is a dynamic way of dealing with the variable number of input galaxy exposures. Other alternatives would be reducing the number of input images to the lowest number of observations per galaxy and band or zero-padding the input to have the same number of observations per galaxy.

`Aczio` extracts 20 numbers per galaxy and narrow band with the feature extractor. The features from all bands are then concatenated into an 800 features tensor (Ψ). This tensor contains the SED and morphology information available in all the observations of a galaxy, which `Aczio` uses to make photometry and photo-z predictions.

7.3.2 Photometry and photo-zs with MDN

We use MDNs to estimate the photometry and the photo-z. MDNs predict the probability distribution (p) of the prediction (y) given the data ($\mathcal{D}$) as a weighted sum of k distributions that can be any basis function. Gaussians functions are a widely implemented basis, i.e.

$$p(y|\mathcal{D}) = \sum_{i}^{k} \alpha_i N_i(\mu_i, \sigma_i)\,, \tag{7.3}$$

where $N_i(\mu_i, \sigma_i)$ is the i-th Gaussian component with mean μ_i and standard deviation σ_i, and α_i is the so-called mixing coefficient, which give the relative contribution of the i-th Gaussian component to the probability distribution. MDNs enable assessing the uncertainty of the predictions, often required for scientific applications.

We use the log-likelihood of the MDN probability distribution as loss function of both the photometry and the photo-z predictions, i.e.

$$\mathcal{L}_{\rm MDN} = -2\log\left(p(y|\mathcal{I})\right) = \sum_{i=1}^{k}\left[\log(\alpha_i) - \frac{(\mu_i - y)^2}{\sigma_i^2} - 2\log\left(\sigma_i\right)\right]. \tag{7.4}$$

where the $\vec{\mu}$ and $\vec{\sigma}$ are the mean and standard deviation parametrising the Gaussians, k is the number of Gaussian components in the mixture model, and $\vec{\alpha}$ are the mixing coefficients. The ground-truth photometry labels are the noiseless fluxes corrected for extinction. For the photo-zs, we use the true redshift from the simulations.

The tensor of features Ψ (§ 7.3.1) is the input of the MDN (on a cream-yellow background in Fig. 7.2). We also include the milky-way extinction $E(B-V)$ to the set of features, which is a valuable parameter for both the photometry and the photo-z predictions. The MDNs predict the photometry in one band and the photo-z of the galaxy. These two tasks share a set of common layers (green circles). Then, the output of the shared layers is the input of two different task-specific blocks, one for the photometry predictions (yellow circles) and another for the photo-z estimation (purple circles). In the later case, we also concatenate the $uBVriz$ broad-band fluxes to the output of the task-common layers. The task-specific MDNs output the corresponding probability distributions as a combination of five Gaussian distributions. The choice of number of Gaussians is based on the Bayesian information criterion (BIC), which evaluates the trade-off between model performance and complexity.

Simultaneously predicting the photometry and the photo-z, we are implementing a machine learning technique named *multi-task learning* (MTL, § 6.3.1, Caruana, 1997). This consists in improving the algorithm's performance on a task by simultaneously training the model on multiple related tasks. MTL forces the model to share representations amongst the photometry and the photo-z predictions, exploiting their commonalities to constrain better the measurements. Our MTL networks use correlations between the photometry and the photo-z to improve the model's performance. In Chapter 6, we already tested MTL techniques with the photometry and the photo-z implementing a MTL network to improve the broad-band photo-z estimates by simultaneously predicting the narrow-band colour of the galaxy (Cabayol et al. in prep.).

The MDN provides the photo-z and the galaxy photometry in one narrow band. `Aczio` contains 40 equal MDNs, each of them predicting the photometry in one narrow-band filter from the same tensor of input features Ψ (§ 7.3.1). We tested predicting the photometry in the 40 narrow bands and the photo-z with a single MDN, combining the all predictions in the same loss function. However, this did not provide the expected results and we leave this optimisation for future work. Some of the difficulties of

7.3.3 Training procedure

First, we train a 2D CNN to predict single-band photometry from multiple exposures. This is not `Aczio`, it is essentially the `Lumos` network but also combining individual observations of the same galaxy and narrow band. This network is composed of a feature extractor as that in the blue shaded area in Fig. 7.2 and an MDN to predict single-band photometry. The output of the feature extractor is the input of the MDN, which predicts the co-added photometry of the three exposures of the galaxy. We use this network to pretrain the parameters of feature extractor before training the network end-to-end from images to photometry and photo-z. Once the 2D CNN is trained, we only keep the feature extractor trained parameters and reject the MDN ones.

The training sample is a tensor with $(N_{\mathrm{gal}}, N_{\mathrm{obs}}, 60, 60)$, where N_{gal} are 480 000 images of galaxies in all narrow bands. N_{obs} are three independent observations of the same galaxy and narrow band, which is the minimum of exposures PAUS observes per galaxy and narrow band. The 60×60 is the size of the image simulations or PAUCam cutouts. We add an additional 5% pixel noise to the input images that prevents overfitting making the training more robust. Furthermore, we also include a 2% of outliers on the fly by scaling the images with a wrong zero-point.

PAUS has no astronomical data sets with observed annotated photometry, therefore we use image simulations (§ 7.2.3) to train `Aczio`. In order to facilitate and ensure a good transition to PAUCam data, we also implement a training technique named *domain adaptation* (Csurka, 2017). This training method aims to align the latent space representations of two different data domains. In our case, the latent space data are the galaxy features extracted with the CNN, and we aim to align the PAUS simulations and data domains. Domain adaptation forces the model to learn domain-invariant features present in data and simulations, thus improving the model's robustness to discrepancies between the two. This technique has already been implemented in astronomy to e.g. to increase the robustness of galaxy morphology classifier against adversarial examples (Ćiprijanović et al., 2021) and to adapt models trained on simulations to data (Ćiprijanović et al., 2020).

We implement domain adaptation with a Maximum Mean Discrepancy loss (MMD, Gretton et al., 2012). This is a kernel distance-based method that minimises the distance between the image features from simulations (x_{s}) and from PAUCam data (x_{d}) mapping them into a higher-dimensional Hilbert space that preserves the statistical features of the original latent spaces. The MMD loss is

$$\mathcal{L}_{\mathrm{MMD}} = k(x_{\mathrm{s}}, x_{\mathrm{d}}) = \exp\left(\frac{|x_{\mathrm{s}} - x_{\mathrm{d}}|}{2\sigma^2}\right), \tag{7.5}$$

where the σ parameters determine the width of the kernel. We use $\sigma = 1$, although we have studied the effect of varying the σ parameters and there are not major differences amongst different σ values.

The loss we use for this first network is

$$\mathcal{L} = 10 \cdot \mathcal{L}_{\mathrm{MMD}} + \mathcal{L}_{\mathrm{MDN}}, \tag{7.6}$$

where we weight the MMD loss by ten so both losses have the same order of magnitude. Besides the 480 000 image simulations, we also use PAUCam cutouts to train the domain adaptation, corresponding to 12 000 PAUCam galaxies to $i_{\mathrm{AB}} < 22.5$ in the spectroscopic sample. Each galaxy and narrow band has a varying number of observations, which is addressed with the Adaptive-pooling layer (§ 7.3.1). The network is trained with an `ADAM` optimiser (Kingma & Ba, 2014) for 500 epochs with an initial learning rate of 10^{-4} that decays a factor of ten every 200 epochs.

Once the feature extractor is trained, its output tensor of features (Ψ) is detached from the training. This means that the training of the MDN does not back-propagate into these features and, therefore, the CNN parameters do not vary. The detached tensor Ψ is the input of the 40 MDN predicting the photometry and the photo-z (§ 7.3.2). Detaching the MDNs from the CNN implies that, once we have the features from each image, we do not need the images anymore. Therefore, the size of the input data drastically reduces, so as the training time. Furthermore, if we were not detaching the features, training 40 MDNs for each narrow band independently would imply that the parameters would update based on a single band in each iteration, which could severely bias the predictions in other narrow bands.

Normalising the input is common practice in machine learning algorithms. The normalisation constrains the amplitude of the predictions, which otherwise enters strongly in the loss function, hindering the training's convergence. The normalisation can be directly implemented to the input, the ground-truth labels, or to both. We are not normalising the photo-z MDNs, but we have tested several options for the photometry predictions, where the range of possible output fluxes is wider. Some of the normalisations tested are e.g. standardising the input cutouts or normalising the cutouts dividing by the maximum pixel value, and we obtain the best results when normalising only the training labels. We have studied two training labels normalisation: dividing by the i_{AB} magnitude and dividing by a fixed 100 value. Both perform similarly on simulations, but the former does not generalise well to data. Therefore, the fiducial photometry `Aczio` training outputs the Gaussian distributions $p(f/100|\mathcal{D})$, which need to be re-scaled to physical fluxes. This indicates that the network benefits from reducing the range of the predicted fluxes, but discrepancies between the i-band magnitudes on simulations and data affect the photometry re-scaling on PAUCam data.

7.4 Multi-band photometry measurements

In this section, we validate the `Aczio` photometry predictions on simulated and PAUCam images. In §7.4.1 we use simulations to evaluate the contribution of the morphology and the SED to the photometry measurement. Then, we validate the flux measurements comparing to `MEMBA` and `Lumos` photometry algorithms and PAUS synthetic SDSS observations (§7.4.2). Section 7.4.3 evaluates the robustness of the errors, and ensure realistic signal-to-noise estimates with the `Aczio` photometry (§7.4.4). In § 7.4.5, we use colour histograms to evaluate the precision of the `Aczio` photometry and § 7.4.6 tests the propagation of errors at the image level to the photometry measurements.

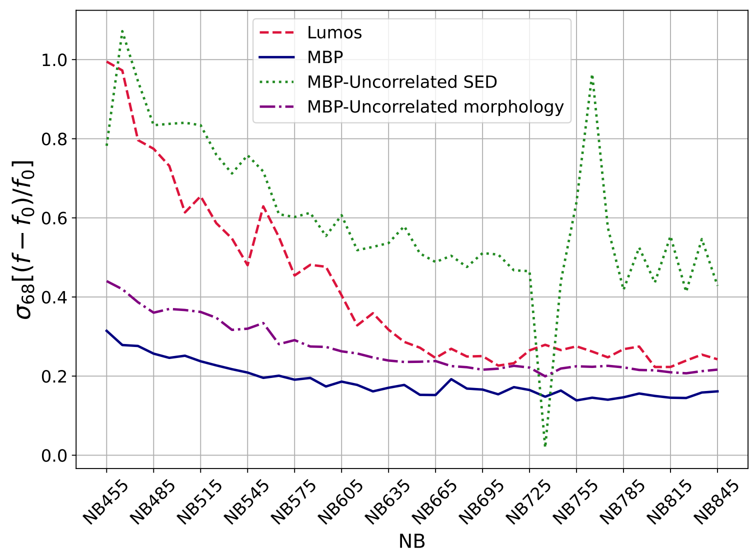

Figure 7.3: Precision in the photometry obtained with: *Dashed red*: Single-band photometry, *Solid blue*: Multi-band photometry. *Dashed-dotted purple*: Multi-band photometry with uncorrelated morphologies. Every observation of the same galaxy is simulated with a different morphology. *Dotted green*: Multi-band photometry with uncorrelated SEDs. The photometry of the galaxy in the different narrow-band filters does not follow any SED.

7.4.1 `Aczio` photometry on simulated data

The PAUS force-aperture photometry algorithm `MEMBA` (§ 7.2.1) measures the flux for single-exposure images and combines the individual flux measurements from several exposures into a co-added flux estimate. Similarly, `Lumos` predicts the flux probability distributions from single observations, which are combined into a co-added flux probability distribution. The construction of this co-added flux probability distribution is a time-consuming process that we side-step with `Aczio`, which directly measures the co-added flux probability distribution per narrow-band ($p(f)$). This also enables a faster and more flexible combination of individual exposures, in which the network can potentially down-weight noisier observations or problematic observations and rely mostly on those with higher signal-to-noise (see § 7.4.6).

There are multiple ways of reducing the co-added flux probability distribution to a single flux measurement (f) and its uncertainty, e.g. the mean, the median, and the peak of the distribution (σ_f). We use the same estimators as in `Lumos`, the mean flux

$$f = \sum_i \alpha_i \cdot \mu_i \,, \tag{7.7}$$

with an associated variance

$$\sigma_f^2 = \sum_i \left[\alpha_i \, \left(\sigma_i^2 + \, (\mu_i - \sum_j \alpha_j \mu_j)^2 \right) \right] \,. \tag{7.8}$$

Figure 7.3 compares the precision of the single-band flux measurements (`Lumos`) and `Aczio` on simulations for the different PAUS narrow bands. The simulated test set contains 1000 galaxies with three observations per galaxy and narrow band (§ 7.2.3). The original `Lumos` network measures single-band photometry from single-exposure images. For this test, we modify the original `Lumos` network to directly predict the co-added single-band photometry from multiple observations, as in `Aczio`. Therefore, the main difference between `Lumos` and `Aczio` is that the former only uses the information encoded in the images it is predicting, while `Aczio` can use the information in all the narrow-band images of the target galaxy to estimate the photometry in one band (§ 7.3.1).

Comparing `Aczio` (blue solid line) and `Lumos` (dashed red line), the former improves the `Lumos` photometry by a factor of $\sim$3 at bluer wavelengths. The difference between the two methods reduces at the red end, although `Aczio` still provides about 40% more precise photometry. Bluer-band images have a significantly lower signal-to-noise than the redder ones. Therefore, at the blue end the network benefits more from seeing other higher signal-to-noise images that it can use to improve the flux estimates. The `Aczio` additional narrow-band information provide extra knowledge to the network. For instance, the underlying SED distribution could help constraining the photometry in one band given the others.

The green dotted line in Fig. 7.3 study the contribution of the SED information to the `Aczio` photometry measurements. There, we train `Aczio` with a set of simulated galaxies with uncorrelated SED information. The photometry in each narrow band is randomly sampled from the true distribution of fluxes, therefore fluxes of consecutive bands are not correlated. Although `Aczio` is still capable of making photometry predictions, these are very inaccurate and present lower precision than the `Lumos` ones (dashed-red line). This indicates that `Aczio` relies on the underlying galaxy SED to infer the fluxes. The input features from uncorrelated fluxes contain, at least, the information `Lumos` uses for its photometry predictions, i.e. the information from the band it is predicting. However, the additional features we provide are potentially adding noise, since the underlying SED fluxes do not correlate. Tests presented in Appendix 7.A suggest that the presence of irrelevant features in the input hinders the prediction of precise photometry. The region with $\lambda \sim 7500$Å corresponds to a wavelength range telluric absorption of O_2 in the atmosphere, and was also problematic for the background-noise predictions in `BKGnet`.

`Aczio` could also use the morphology correlation to facilitate the photometry measurements. The networks need to internally estimate the galaxy size to measure the photometry. While a single-band photometry network is forced to use the size measurement from the narrow-band image it is predicting, `Aczio` can use the morphology inferred from higher signal-to-noise observations. The dotted-dashed purple line in Fig. 7.3 studies the contribution of the morphology to the photometry measurements. For that, we train the network on simulations with uncorrelated galaxy morphology, i.e. the simulated galaxies follow an underlying SED, but the size and shape of the galaxies change amongst bands. The improvement with respect to single-band photometry (dashed red line) is still very significant, however, the photometry scatter is $\approx$40% higher with respect to `Aczio` (solid blue line). This indicates that the morphology from other narrow-band images also has an impact on the

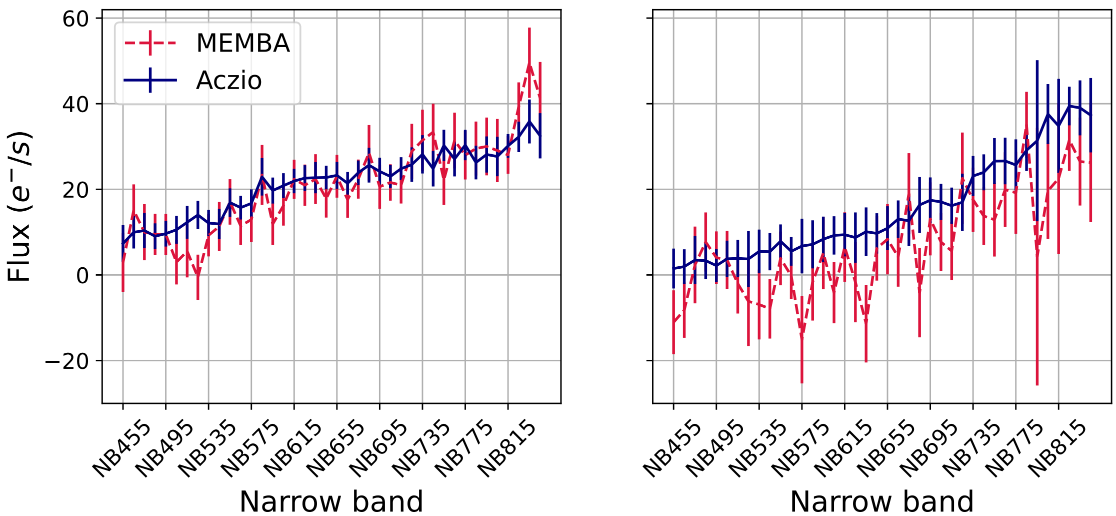

Figure 7.4: *Left*: Photometry of a PAUS galaxy with $i_{rmAB} = 22.1$ with `MEMBA` (dashed red line) and `Aczio` (solid blue line). *Right:* PAUS galaxy with 40% of the observations with a scattered light flagging. The `MEMBA` measurements (dashed red line) reject all problematic observations, with `Aczio` (solid blue line) we use all the images available for the galaxy.

`Aczio` photometry, however, the underlying SED information has a stronger effect.

On the same set of simulations, the `Aczio` photometry (blue solid line in Fig. 7.3) results in a relative bias $b_f < 3\%$ in the photometry measurements in all narrow-bands independently, where (b_f)

$$b_f = \text{Median}[(f - f_0)/f_0]\,, \tag{7.9}$$

and f and f_0 are the predicted and true fluxes, respectively.

7.4.2 `Aczio` fluxes on PAUCam images

In this section, we first compare the `Aczio` flux estimates with Eq.7.7, with the `Lumos` and `MEMBA` catalogues in COSMOS. Then, we compare our multi-band photometry measurements with synthetic PAUS photometry to further validate the flux estimates. We use ∼14 000 galaxies from the COSMOS spectroscopic sample (§ 7.2.3).

We first compare the `Aczio` individual flux measurements with the `MEMBA` and `Lumos` catalogues using the flux ratio

$$r = f_{\text{Aczio}}/f_{\text{x}}\,, \tag{7.10}$$

where f_{x} can be the flux of either `MEMBA` or `Lumos`. For the complete PAUS sample in COSMOS, `Aczio` measures 17% higher fluxes than `MEMBA` and `Lumos`. The image training simulations include milky-way extinction. We provide the network with the $E(B-V)$ values so that it directly learns to predict the extinction corrected photometry. In contrast, the `MEMBA` and `Lumos` photometries are not extinction corrected. To see the contribution of the

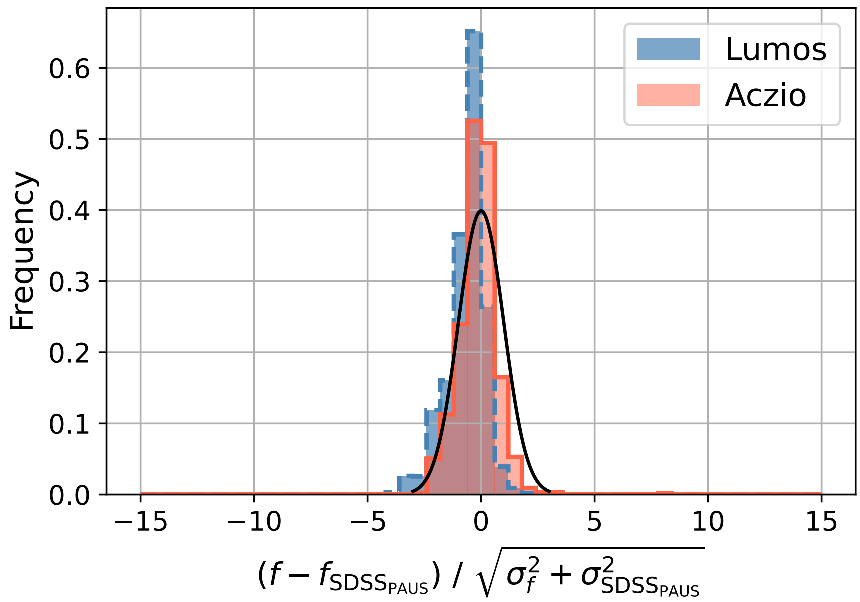

Figure 7.5: Comparison between PAUS synthetic flux measurements from SDSS spectra and the `Aczio` (solid red) and `Lumos` (dashed blue) predicted photometries.

extinction correction, we have trained and evaluated the `Aczio` photometry without correcting for extinction. In this case, `Aczio` measures 10% higher fluxes than `Lumos` and `MEMBA`. The ratio is constant in wavelength and increases to 1.15 at the faint end (i_{AB}>22).

The left panel on Fig. 7.4 shows an example of PAUS galaxy with the co-added photometry measured with `MEMBA` (dashed red line) and `Aczio`. For most of the bands, both measurements are compatible within error. The `MEMBA` fluxes are more noisy and oscillate around the `Aczio` fluxes, which show more stable. Also, the error bars in the `Aczio` photometry are lower than the `MEMBA` ones. The panel on the right shows the photometry for a galaxy with flagged outlier observations, which is discussed in § 7.4.6.

We also compare `Aczio` photometry with synthetic PAUS photometry to validate the flux estimates. The synthetic photometry is constructed convolving SDSS galaxy spectra with the PAUCam filters throughput and it corresponds to a bright sample with a magnitude limit $i_{\mathrm{AB}} < 20.5$. We match PAUS observations with SDSS spectra by sky position, pairing observations within 1”. Comparing PAUS photometry and SDSS spectra requires scaling the synthetic PAUS fluxes with a multiplicative scaling amplitude (A) that takes into account the observing conditions and the different data reduction pipelines. The scaling zero point is obtained by minimising the χ^2 between PAUS observations and PAUS synthetic fluxes, i.e.

$$\chi^2 = \sum_{\mathrm{x}} \frac{\left(f_{\mathrm{SDSS,x}} - A \cdot f_{\mathrm{SDSS_{PAUS,x}}}\right)^2}{\sigma^2_{\mathrm{SDSS,x}} + A^2 \cdot \sigma^2_{\mathrm{SDSS_{PAUS,x}}}}, \tag{7.11}$$

where SDSS labels the observed SDSS photometry, $\mathrm{SDSS_{PAUS}}$ is the SDSS-PAUS synthetic flux and the sum (x) is over the gri broad bands. This minimisation provides a median

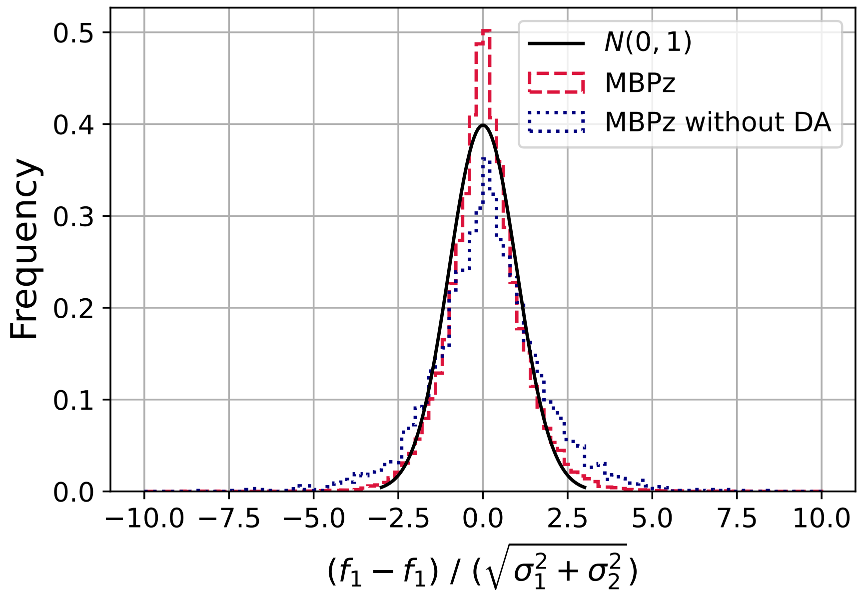

Figure 7.6: The $\mathcal{C}'$ distribution with `Aczio` photometry estimates for $\sim$14 000 PAUCam galaxies in the COSMOS field (dashed red). These should distribute as a Gaussian $N(0, 1)$, plotted in black solid. The blue dotted line present the same results for `Aczio` trained without domain adaptation.

multiplicative scaling amplitude of 1.64 with a $\sigma_{68} = 1.01$.

The `Aczio` photometry has a 2% relative error when compared to the synthetic photometry. Also, the ratio between the two is 0.97, indicating 3% lower fluxes in our multi-band photometry. Figure 7.5 shows the difference between the synthetic PAUS photometry and that predicted by `Aczio` (solid red) and `Lumos` (dashed blue), relative to the photometry uncertainties. If the two measurements were uncorrelated, the histograms should distribute as a Gaussian with zero mean and unit variance. The `Aczio` distribution is more Gaussian, while the `Lumos` is slightly skewed towards underestimated flux predictions. However, these two measurements are not necessarily uncorrelated and the width of both histograms is around $\sigma_{68} = 0.75$.

7.4.3 Flux uncertainties

To validate the uncertainties associated with the fluxes derived from the $p(f)$ (Eq. 7.8), we define

$$\mathcal{C} \equiv (f - f_0)/\sigma_f \,, \tag{7.12}$$

where f is the network's flux measurement, f_0 is the true flux, and σ_f is uncertainty associated to the flux measurement. As for the comparison with SDSS fluxes, if the flux uncertainties are robust, $\mathcal{C}$ should distribute as a Normal distribution with zero mean and unit variance $N(0, 1)$. In this case, we are comparing `Aczio` flux measurements to the true-flux value, thus the are no sources of correlated errors that affect the $\mathcal{C}$ distribution. We find $\sigma_{68}[\mathcal{C}] \approx 1.10$ on the simulated test set. Splitting in equally populated magnitude bins, flux uncertainties are $\sim 10\%$ overestimated for the 10% brightest and faintest galaxies, while the rest of the sample has robust uncertainties within a 5% error.

To validate the uncertainties on PAUCam data, we select four observations of each galaxy per narrow band and randomly split them in pairs. From each independent pairs of images, we evaluate with `Aczio` two uncorrelated flux measurements (f_1, f_2) of the same galaxy and narrow band with uncertainties (σ_1, σ_2). Then, we define the distribution of

$$\mathcal{C}' = (f_1 - f_2)/\sqrt{\sigma_1^2 + \sigma_2^2}\,, \tag{7.13}$$

where both the fluxes and the uncertainties are calibrated to be compared. Similarly to Eq. 7.12, for two uncorrelated photometry estimates $\mathcal{C}'$ should distribute as a Gaussian with zero mean and unit variance. However, despite measuring the photometries from independent pairs of observations, images overlapping in sky position use the same stars to derive the scaling zero points. Therefore, the photometric calibration is a source of correlated errors.

Figure 7.6 shows the distribution of $\mathcal{C}'$ for the photometry with `Aczio` of $\sim 14\,000$ PAUCam galaxies (dashed red line). It provides an overall width of $\sigma_{68}[\mathcal{C}'] = 0.95$. As on simulations, the 10% brightest and faintest galaxies have $\sim$15% overestimated uncertainties while the rest of the sample has robust uncertainties within a $\sim$5% error.

7.4.4 Signal-to-noise on PAUCam data

Figure 7.7 shows the signal-to-noise of $\approx$14 000 PAUCam galaxies in ten equally-populated magnitude bins for the aperture-photometry method `MEMBA` (dashed red line), the deep neural network `Lumos`, and `Aczio`. In all cases, we are showing the signal-to-noise of the co-added fluxes.

`Aczio` provides higher signal-to-noise than `MEMBA` and `Lumos` in all magnitude bins. Comparing to `MEMBA`, `Aczio` increases the signal-to-noise by 25% at the bright end and by a factor of three at the faintest end. Overall, `Aczio` provides around 2 higher signal-to-noise than `MEMBA`. In the case of `Lumos`, the signal-to-noise increment is lower, although still very significant. At the bright end, `Aczio` gives approximately 21% higher signal-to-noise, and this goes up to two times better photometry at the faintest end. Overall, `Aczio` provides 1.7 times higher signal-to-noise than `Lumos`.

7.4.5 Colour histograms

Assuming that galaxies have an underlying distribution of colours, noise in the observations smooths such colour distribution. The observed colour histograms correspond to the underlying colour distribution convolved with the noise. Consequently, the best photometry on a galaxy sample is that with the narrowest colour histograms, thus with lower scatter in the fluxes. We already used colour histograms to evaluate the `Lumos` photometry and these were

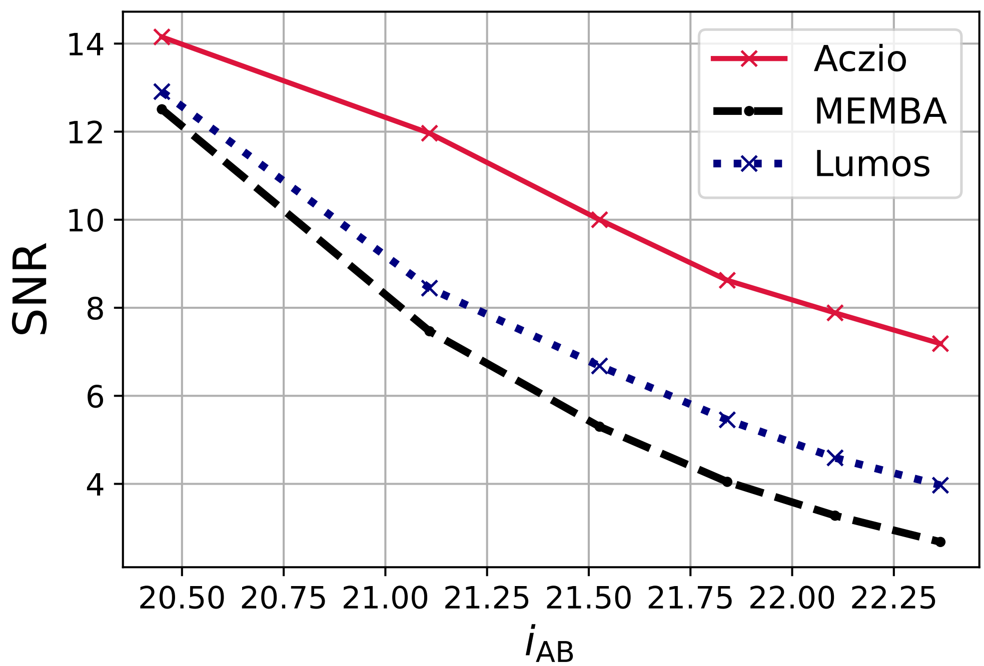

Figure 7.7: Signal-to-noise (SNR) of the photometry of ∼14 000 PAUCam galaxies with: *black solid*: `Aczio`; *dashed dotted blue:* `Lumos`; and *dashed red*: `MEMBA`. The results are presented in equally populated magnitude bins.

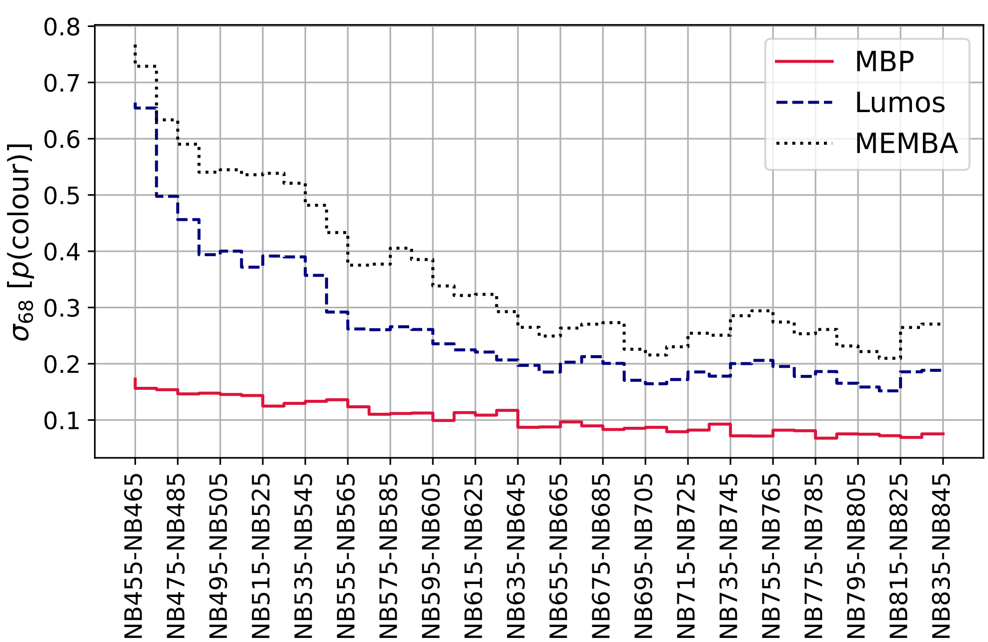

Figure 7.8: Width (σ_{68}) of the colour histograms from the `MEMBA` (dotted black line), `Lumos` (dashed blue line), and `Aczio` (solid red line) photometries.

also employed to evaluate the `LAMBDAR` photometry algorithm (Wright et al., 2016) on data from the Galaxy and Mass Assembly (GAMA, Driver et al., 2011).

Figure 7.8 presents the width (σ_{68}) of the colour histograms with `MEMBA` (dotted grey line), `Lumos` (dashed blue line), and `Aczio` (solid red line). `Lumos` already reduces the colour histograms width with respect to `MEMBA`. `Aczio` drastically improves both `Lumos` and `MEMBA` photometries. At bluer wavelengths, where observations are fainter, `Aczio` narrows the colour histograms by a factor of $\approx$4 with respect to `MEMBA` and $\sim$3 in the case of `Lumos`. This indicates that the colours with the `Aczio` is much more precise than the single-band photometry either with `MEMBA` or `Lumos`.

We have also estimated the width with σ_{95}, which is equivalent to σ_{68} but with the 2.5 and 97.5 percent quantiles. Unlike σ_{68}, which measures the core of the distribution, σ_{95}, traces the tails of the colour histograms. The ratio between σ_{95} and σ_{68}

$$\mathcal{S} = \sigma_{95}[\mathrm{p(colour)}]/\sigma_{68}[\mathrm{p(colour)}] \tag{7.14}$$

indicates the length of the distribution's tails, and thus it is an estimator of photometry outliers. At blue wavelengths, $\mathcal{S}$ is ~ 5 times lower with `Aczio` than with `Lumos`. Comparing to `MEMBA`, this factor increases to ~ 7. The difference reduces at redder wavelength images, with ~ 2 lower $\mathcal{S}$ values with `Aczio` than with `Lumos` and ~ 4 times lower than `MEMBA` (not shown). Overall, `Aczio` presents tighter colour histograms that the other two photometric methods, and which also indicate a reduction in the number of photometric outliers.

7.4.6 Robustness of the method

In this sub-section, we study the robustness of `Aczio` in the presence of distorting effects such as e.g. cosmic rays and scattered light. For that, we have select PAUCam galaxies with observations flagged as problematic and see the impact that these have on the `Aczio` co-added flux.

Scattered light is the result of light deflecting from the instrument optical path appearing at a different region of the detector. The PAUCam images are affected by scattered light, which is the cause numerous outlier observations. In `BKGnet`, we showed that scattered light is not a random effect. It follows a noise pattern in all PAUS observations that can be predicted and corrected for. `BKGnet` predicts the scattered light contribution from the galaxy image, the narrow band it has been taken with, and the galaxy coordinates. `Aczio` has all these information and, furthermore, it has access to other observations of the same galaxy, which are potentially not contaminated by scattered light.

The right panel on Fig. 7.4 shows the 40 PAUS co-added fluxes predicted with `Aczio` and `MEMBA`. This particular galaxy has 40% of its exposures flagged as strongly affected by scattered light. The `MEMBA` fluxes are estimated from the unflagged observations only, while we keep all observations in the `Aczio` measurements. The `MEMBA` photometry presents a noisy photometry, most likely due to the high number of observations rejected. Instead, the `Aczio` flux measurements using all observations are not affected by scattered light and provide reliable photometry.

We proceed similarly with other effects as cosmic rays with similar findings. `Aczio` does not need to reject observations with cosmic rays and knows how to predict robust photometry in its presence. We have also tested how errors in the image cosmetics propagate to the flux measurement. While aperture photometry measures very negative (thus with a overestimated noise subtraction), the `Aczio` co-added flux measurement is not affected.

7.5 Photo-z measurements

In this section, we validate the `Aczio` photo-z predictions on PAUS data. In § 7.5.1, we introduce transfer learning, a technique to adapt the photo-zs trained on simulations to data. Section 7.5.1 presents the main photo-z results in the COSMOS field, including the scatter and the photo-z distributions. Then, § 7.5.2 evaluates the photo-z probability distributions and § 7.5.3 studies the photo-z performance in colour space.

7.5.1 Photo-zs in COSMOS

In this sub-section, we study the performance of `Aczio` on PAUS data in the COSMOS field. To evaluate the accuracy and precision of the photo-z estimates, we define

$$\Delta z \coloneqq (z - z_{\mathrm{s}}) \;/\; (1 + z_{\mathrm{s}})\,, \tag{7.15}$$

where z and z_{s} are the photo-z and the spectroscopic redshifts, respectively. The bias and the dispersion are defined as the median and σ_{68} of Δz, respectively, where we define σ_{68} as

$$\sigma_{68} \coloneqq \frac{1}{2}\left[Q_{84}(\Delta z) - Q_{16}(\Delta z)\right]\,, \tag{7.16}$$

and $Q_{16}(\Delta z)$, $Q_{84}(\Delta z)$ are the 16th and 84th percentiles of the Δz distribution.

Transfer learning

Transfer learning (TL, Pan & Yang, 2010) is a machine learning technique to adapt the knowledge gained training one task to a different but related problem (Torrey & Shavlik, Torrey & Shavlik; Pan & Yang, 2010). This technique can be also implemented to adapt a model trained with some data, to perform well on different data set, e.g. astronomical observations from different surveys (Kim et al., 2021). Also, many times annotated training sets are not sufficiently large to train the neural network from scratch. Then, one possibility is to first train using large simulations and then implement transfer learning using the annotated smaller data set (Eriksen et al., 2020).

There are no data set of observed annotated PAUS photometry thus we use simulated PAUCam images to train the photometry and the photo-z predictions with `Aczio` (§ 7.2.3).

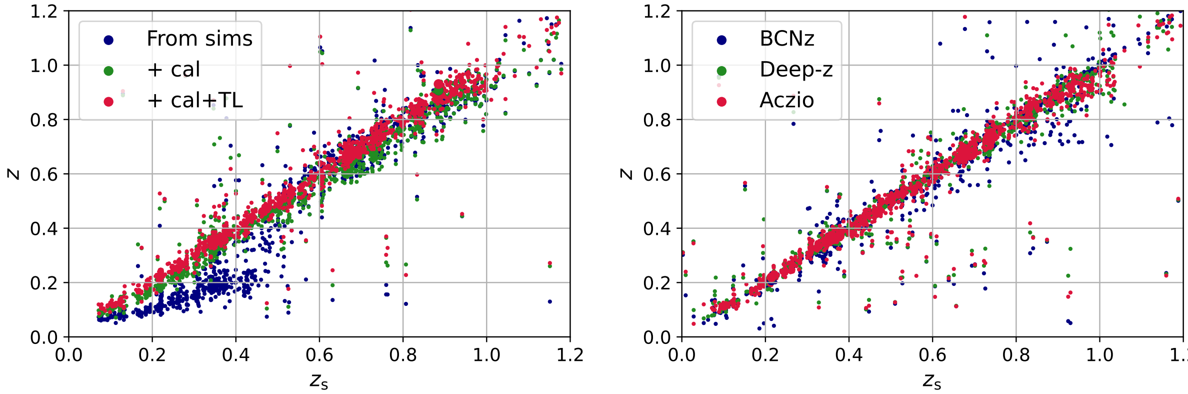

Figure 7.9: Photo-z scatter plot with `Aczio` (blue), `BCNz` (red), and `Deepz` (green). *Left:* Before implementing transfer learning. *Right:* After implementing transfer learning.

After that, we apply transfer learning to the `Aczio` photo-z prediction using the COSMOS spectroscopic sample. For that, we use PAUCam cutouts as input, together with the COSMOS $uBVriz$ photometry and the galactic extinction $E(B-V)$ from the dust map in Schlegel et al. (1998). For the ground-truth redshifts, we use the spectroscopic redshifts in COSMOS2015 selecting only very secure redshift with a the quality parameter (§ 7.2.1. We can only implement transfer learning to the photo-z predictions, since we do not have a smaller set of annotated photometry. Therefore, all the network parameters affecting the photometry predictions are detached from the transfer learning training. Otherwise, the photometry predictions could severely be affected.

On simulations, we obtain a photo-z precision of $\sigma_{68} \approx 0.017$. Evaluating the photo-$z$s on PAUS without transfer learning, we find $\sigma_{68} \approx 0.05$, which is driven by a clear photo-z systematic bias at $z_s < 0.4$ (blue scatter points in the left panel of Fig. 7.9). Before implementing transfer learning, we calibrate the COSMOS $uBVriz$ broad-band photometry implementing offsets (Δm), i.e.

$$m' = m + \Delta m\,, \tag{7.17}$$

where m' are the calibrated broad-band magnitudes and m are the $uBVriz$ magnitudes as in the COSMOS2015 catalogue. We estimate the Δm parameters minimising the photo-z scatter on the PAUS sample in COSMOS as a function the offsets. For that, we implement the Nealder-Mead minimisation algorithm (Jones et al., 2001) from `SciPy` (Jones et al., 2001). With the broad-band calibration, we reach a $\sigma_{68} \approx 0.02$, fixing the systematic bias at $z_s < 0.4$ (green scatter points in the left panel of Fig. 7.9).

After calibrating the broad bands, we train the transfer learning on 12 000 spectroscopic galaxies. We train it for 200 epochs with an initial learning rate of 10^3, decreasing a factor of ten every 75 epochs. Transfer learning increases the photo-z precision to $\sigma_{68} \approx 0.013$ (red scatter points in the left panel of Fig. 7.9).

The right panel in Figure 7.9 shows the photo-z scatter for the template-fitting algorithm `BCNz`, the deep neural network from derived photometry `Deepz`, and `Aczio`. All methods provide unbiased photo-z predictions to the sub-percent level. Without any quality cut on

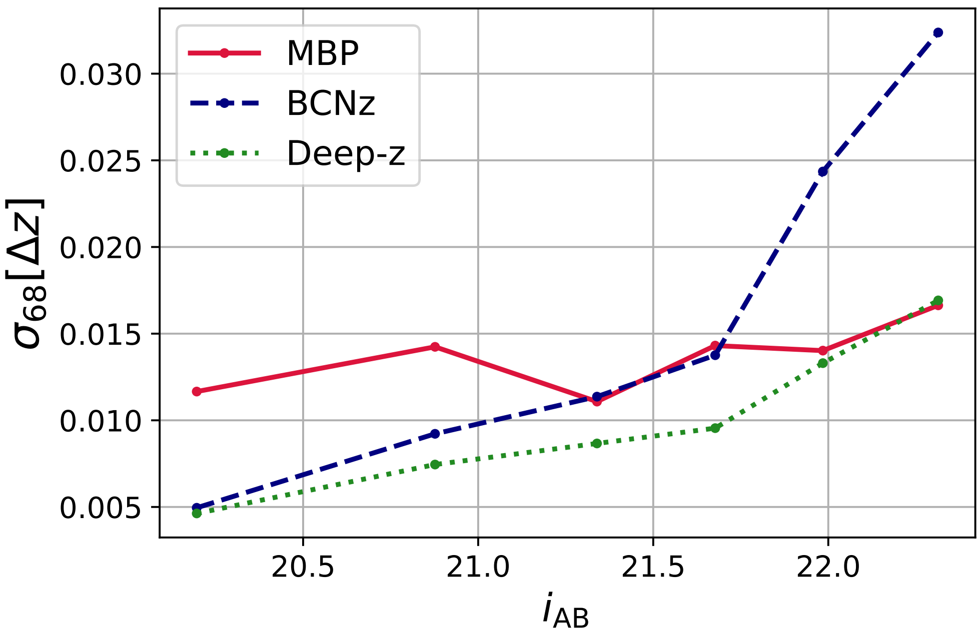

Figure 7.10: Photo-z precision in equally populated magnitude bins with `Aczio` (solid red), `BCNz` (dotted blue), and `Deepz` (dashed green).

	COSMOS	COSMOS (3<$conf$<5)	COSMOS $i_{\mathrm{AB}} > 20$
`BCNz`	0.0138	0.0083	0.016
`Deepz`	0.0098	0.0068	0.010
`Aczio`	0.0136	0.0111	0.013

Table 7.1: Photo-z precision in the COSMOS field for the template-fitting code `BCNz`, the deep neural network working on the derived photometry `Deepz`, and the deep convolutional neural network working on images `Aczio`.

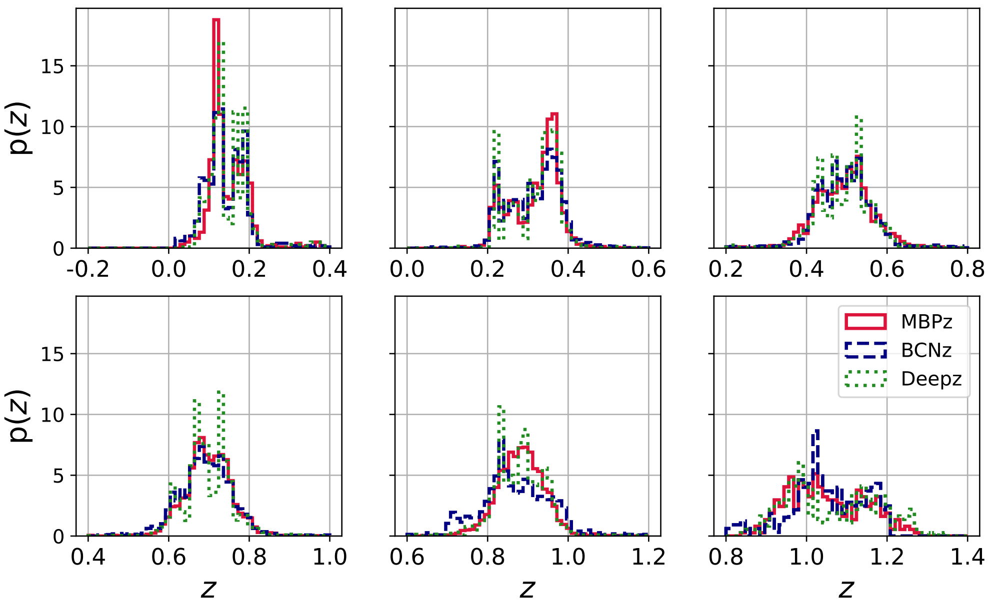

Figure 7.11: Redshift distributions for the COSMOS spectroscopic sample with $0 < z < 1.5$ with `Aczio`, `BCNz`, and `Deepz`. The redshift bins have a width of $\Delta z = 0.2$, except the last one, which expands $1.2 < z < 1.5$.

the spectroscopic sample, we obtain $\sigma_{68} \approx 0.0138$ with `BCNz`, $\sigma_{68} \approx 0.00976$ with `Deepz`, and $\sigma_{68} \approx 0.0136$ with `Aczio`. Lilly et al. (2009) defines a quality parameter ($conf$) based on the confidence in the spectroscopic redshift measurement. Cutting on $3 < conf < 5$, we obtain $\sigma_{68} \approx 0.0083$, $\sigma_{68} \approx 0.0068$, and $\sigma_{68} \approx 0.0111$ for `BCNz`, `Deepz`, and `Aczio`, respectively. All these values are summarised in Tab. 7.1.

The analysis is extended in Fig. 7.10, which shows the photo-z precision in equally populated magnitude bins to $i_{\mathrm{AB}} < 22.5$ for `Aczio` (solid red), `BCNz` (dotted blue), and `Deepz` (dashed green). Although the overall precision with `BCNz` and `Aczio` is similar, we can see that for $i_{\mathrm{AB}} > 21.5$, `Aczio` increases by 25% the `BCNz` photo-z precision and gets close to the `Deepz` one. In contrast, at the bright, `Aczio` provides poor photo-z estimates compared to the other two methods. One possible explanation is a discrepancy between the simulated and the PAUCam images that affect more bright galaxies since these are more resolved. Fixing this problem is still work in progress.

Figure 7.11 shows the redshift distribution (N(z)) in six equally spaced tomographic redshift covering of 0.2 redshift with covering the range $0 < z_{\mathrm{s}} < 1.5$. The last bin is wider and includes $1.2 < z_{\mathrm{s}} < 1.5$, so that we ensure a sufficient number of galaxies in the bin. The galaxies in each bin are selected with the spectroscopic redshift. All redshift bins except the $1.2 < z_{\mathrm{s}} < 1.5$ present distributions centred in the corresponding redshift range. In the last tomographic bin, we observe galaxies with underestimated photo-z with the three methods. The photo-z in the redshift range $1 < z_{\mathrm{s}} < 1.2$ has a better $N(z)$ distribution with `Aczio` than with the other algorithms, especially comparing to `BCNz`.

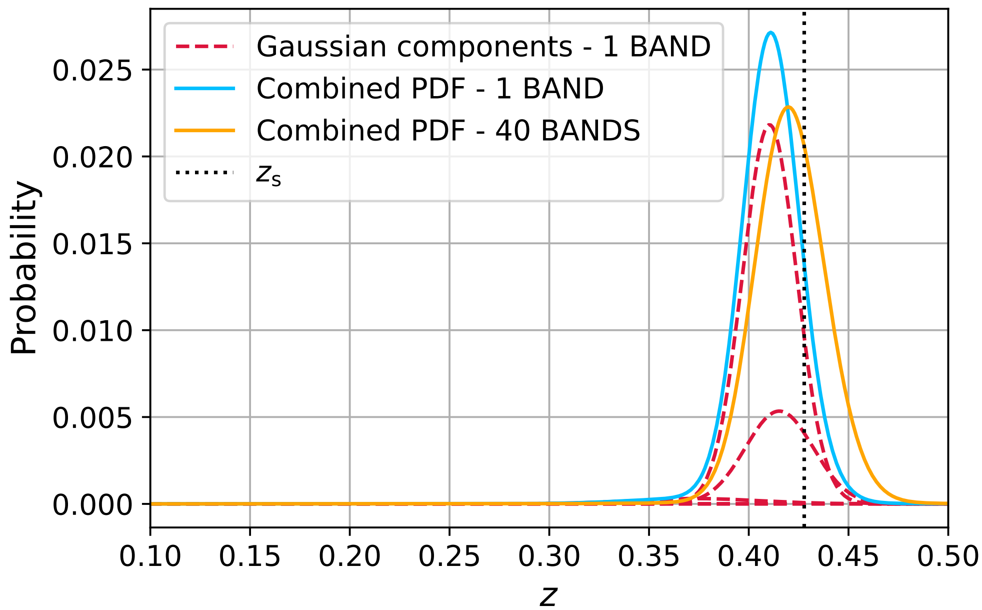

Figure 7.12: Example of redshift probability distribution with true redshift at $z_s = 0.43$ (black vertical line). The dashed red lines correspond the five Gaussian components constructing the redshift probability distribution of a single MDN. These are added into a single $p(z)$ (solid blue line). Both the red dashed lines and the blue one correspond to the output of one of the 40 MDNs. The photo-z measurements from each of the 40 MDNs combine in the orange solid line.

7.5.2 Photo-z probability distributions

`Aczio` provides both the flux and photo-z probability distributions as a combination of five Gaussian distributions (§ 7.3.2). Figure 7.12 shows an randomly selected example of a photo-z probability distribution predicted with `Aczio`. The vertical black dotted line is the spectroscopic redshift of the galaxy. The network contains 40 MDNs, each of which predicts the photometry in one narrow band and the photo-z of the galaxy as a linear combination of five Gaussian distributions. The dashed red distributions in Fig. 7.12 correspond to the five Gaussian components constructing the $p(z)$ of a single MDN. The Gaussian components are already weighted by the mixing coefficients (§ 7.3.2). In this case, there are two components contributing significantly to the $p(z)$, while the other add up to the tails.

The relaxing sea-blue solid line corresponds to the addition of the five Gaussian components of the mixture model weighted by the mixing coefficients. Then, the orange line is the combination of the 40 $p(z)$s, one per MDN.

For the photometry, where there are no annotated data for a direct validation of the predictions, we compared independent flux measurements of the same galaxy and narrow band (§ 7.4.3). In this section, we characterize the calibration and the sharpness of predicted photo-z probability distributions (p(z)) using the probability integral transform (PIT, Dawid, 1984; Gneiting et al., 2005; Bordoloi et al., 2010) and the continuous ranked probability score (CRPS, Hersbach, 2000; Polsterer et al., 2021) metrics. To validate the p(z) measurements,

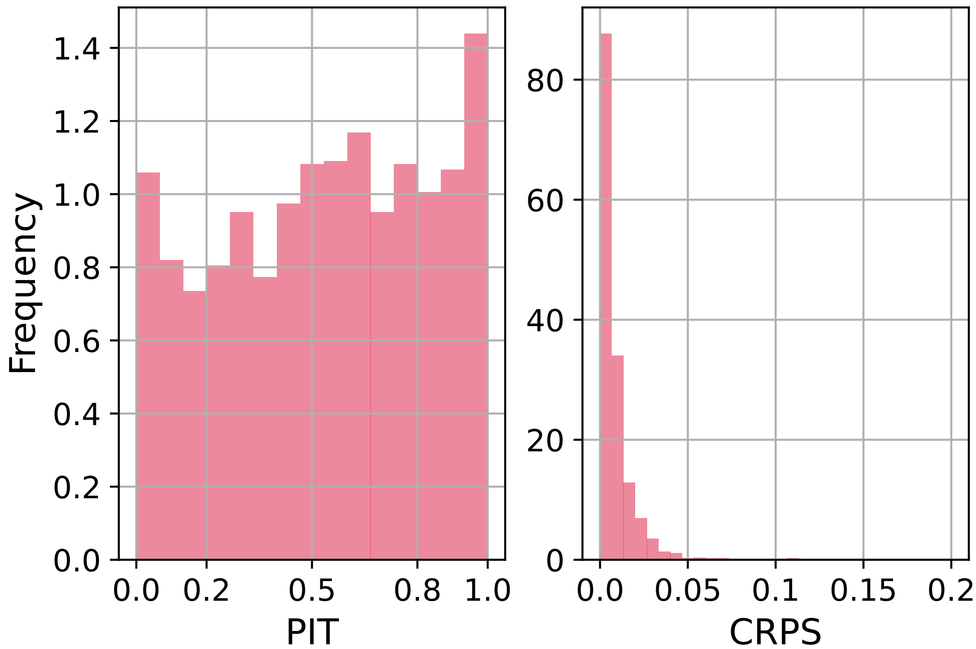

Figure 7.13: *Left:* Histogram of the PIT values for a sample of 3000 COSMOS galaxies with spectroscopic redshift. *Right:* Histogram of the CRPS values for the same sample of galaxies.

we use the COSMOS spectroscopic redshifts as true redshifts.

The PIT metric is defined by

$$\mathrm{PIT} \equiv \int_{-\infty}^{z_\mathrm{s}} \mathrm{d}z\, \mathrm{p}(z) \tag{7.18}$$

where z_s is the spectroscopic redshift. When the p(z) faithfully represents the true redshifts, the PIT distribution is the uniform distribution U[0,1]. Contrary, PIT histograms presenting peaks at the edges, i.e. around zero and unity, indicate the presence of outlier measurements. Also, PIT histograms more populated at the centers than on the edges denote over-dispersed probability distributions. The left panel of Fig. 7.13 contains the PIT distribution for a random subset of 3000 COSMOS galaxies. The PIT histogram is close to a U[0,1] with few outliers at unity, corresponding to underestimated $p(z)$s.

The CRPS is defined as

$$\mathrm{CRPS}(\mathrm{p}(z), z_\mathrm{s}) \equiv \int_{-\infty}^{\infty} \left(\mathrm{CDF}\left(p(z)\right) - \mathcal{H}(z - z_\mathrm{s})\right)^2 \mathrm{d}z\,, \tag{7.19}$$

where CDF $(p(z))$ is the photo-z cumulative distribution and $\mathcal{H}$ is the Heaviside step function. Robust photo-z probability distributions should have cumulative distributions close to unity when integrated up to the true redshift. Therefore, the integrated quantity in the CRPS score should be small in the limit where z reaches z_s. The lower the CRPS score is, the better our $p(z)$ capture the true redshift. The right panel on Fig. 7.13 shows the distribution of CRPS values, which presents the expected shape with a mean value of 0.013.

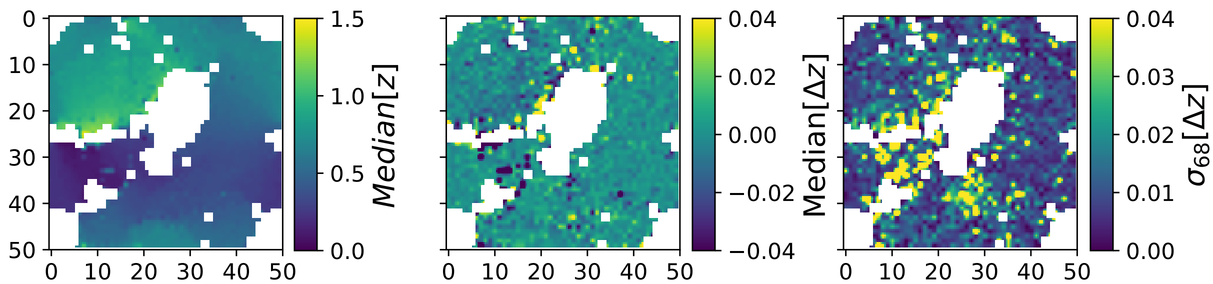

Figure 7.14: Colour-space distribution of the photo-zs measured with `Aczio`. The left-most panel presents the photo-z. The panel in the centre shows the photo-z bias and the right-most panel, the photo-z scatter.

7.5.3 Colour space

We study the colour-space distribution of `Aczio` photo-z measurements. For that, we use a self-organising map (SOM, Kohonen, 1982), an unsupervised machine learning algorithm trained to produce a low-dimensional (typically two-dimensional) representation of a multi-dimensional space. A two-dimensional SOM contains $N_x \times N_y$ cells with an associated vector of attributes ($\vec{w}^k$), where N_x and (N_y) correspond to the dimension of the SOM on the x and y-axis, respectively, and k corresponds to the kth SOM cell. The SOM training phase is an iterative process during which the SOM cells compete amongst themselves to represent the training data (see e.g. (Wright et al., 2020a; Myles et al., 2021)). Once the SOM is trained, the cells are optimised to represent a particular type of galaxy. Cells representing similar galaxies are also close-by in the SOM space.

We have trained a 50×50 SOM using the $uBVriz$ photometry from COSMOS2015. The size of the SOM is a compromise between colour-space resolution and the occupation of the cells. Also, we train the SOM with the broad bands because previous studies showed that training the SOM with 40 noisy narrow bands requires a very large training sample, not currently available.

Figure 7.14 shows the photo-z statistics in colour space. The first left-most panel corresponds to the photo-z distribution in colour space, where we can see that population with the same photo-z cluster. The centred panel shows the photo-z bias, where we do not appreciate any colour-space region with biased photo-zs. In contrast, there are some locations in colour-space with lower photo-z precision, but these also show lower precision with `BCNz` and `Deepz`, indicating that the issue could potentially be in the broad-band photometry or in the galaxy images.

7.6 Conclusions

Imaging surveys require accurate galaxy photometry and precise photometric redshifts for most of the science analysis. In this work, we are developing `Aczio`, a deep convolutional neural network to predict both the multi-band photometry and the galaxy redshift directly from its image observations. Aczio is built on Lumos (§ 5), a CNN that predicts single-band photometry and Deepz, a deep linear network that map photometry to redshifts. Aczio surpasses Lumos predicting the multi-band photometry using some of the techniques implemented in Deepz.

We have developed Aczio with data from the PAUS survey. Since PAUS does not have any set of annotated observed photometry, Aczio is trained on Teahupoo galaxies, image simulations specially built for this work. While we already used Teahupoo galaxies in Lumos (§ 5.2.2), we have extended these simulations to include bulge and a disk galaxy modelling and multiple calibrated observations of the same galaxy and narrow band (§ 7.2.3). One of the main characteristics of Teahupoo galaxies is that these use PAUCam image cutouts for the background noise. This enables the network to learn from examples how to make reliable predictions in the presence examples of distorting effects as e.g. blended galaxies and cosmic rays (§ 7.4.6).

At bluer wavelengths, the multi-band photometry from Aczio reduces the Lumos single-band photometry scatter by a factor of three. At redder bands with higher signal-to-noise, the improvement is 40%. In contrast to Lumos, Aczio uses all image observations of a galaxy in any narrow band to predict the photometry in each of the bands. This enables Aczio to use the SED information available in the images to improve the photometry in all bands (§ 7.4.1). Having multiple estimates of the morphology also contributes to the improvement in the photometry estimates, although this has a lower impact compared to the SED.

On PAUS data, Aczio doubles and triples the signal-to-noise at the faint end of Lumos and MEMBA, respectively (§ 7.4.4). We compare the Aczio photometry of pairs of independent observations of the same galaxy and narrow band to test the robustness of the uncertainty measurements, finding the signal-to-noise measurements are realistic (§ 7.4.3). Also, we validate the Aczio photometry with SDSS convolved spectra finding a 2% relative error.

So far, we reach an overall photo-z precision better than BCNz in the COSMOS spectroscopic sample, although we have still not achieved the Deepz performance. At the faint end, Aczio provides very precise photo-z estimates, but it shows issues predicting the photo-z of bright galaxies (§ 7.5.1). We have tested the calibration and sharpness of the photo-z probability distributions using the PIT and the CRPS scores. The former indicates that the $p(z)$s are well calibrated. We also obtain a CRPS value compatible to other photo-z studies.

This is still work in progress and in our near-future actions, we aim to e.g. testing different input normalisations, implementing better data augmentation techniques, optimising the Teahupoo image simulations, test the CRPS as loss function, and study the photo-z performance dependence with brightness and redshift to understand the trend and beat down the precision at the bright end.

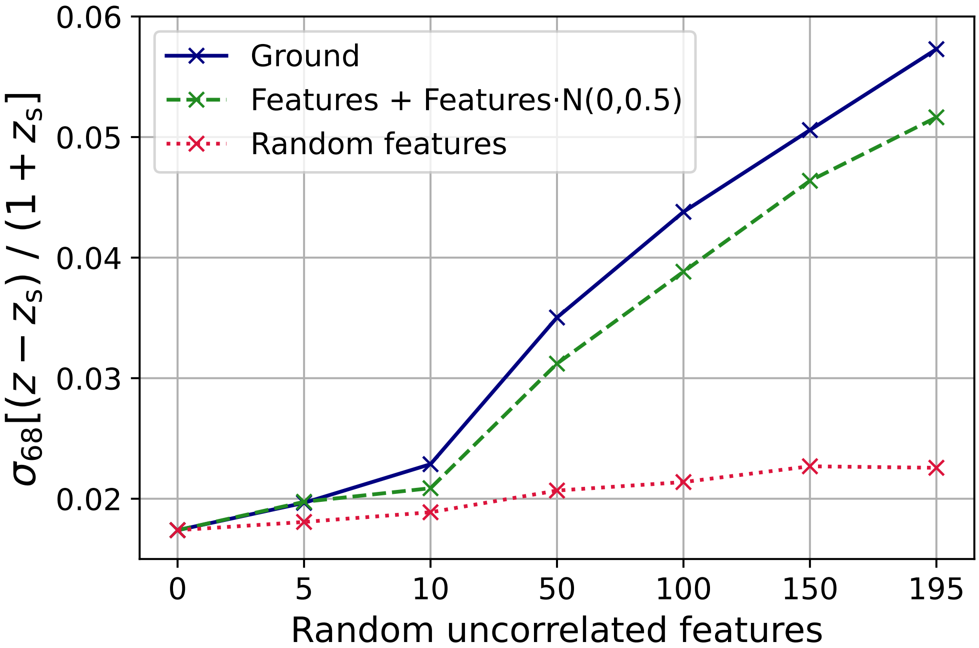

Figure 7.15: Photo-z precision for different number of uncorrelated random features in the input. In the dashed-green and red-dotted line, we implement two distinct data augmentation techniques in the training.

7.A Effect uninformative features in the input of a neural network

`Aczio` estimates the photometry and the photo-z from 800 features extracted from all the observed images of a galaxy (§ 7.3.2). These features contain all the SED and morphology information available in the images. The network uses the information from images taken with different narrow-band filters to improve the photometry measurement in one band (§ 7.4.1). However, it can potentially happen that some of these information (e.g. the features from a distant narrow band) are much less relevant for the photometry prediction in a given narrow band.

In this section, we study the effect of training a neural network when a fraction of the input data features is uninformative. When this fraction is small, the network should be capable of capturing the relevant features and ignore the others. However, a dominant fraction of irrelevant features could potentially hinder the training, adding too much noise for the network to learn from the valuable data.

To estimate the photometry and the photo-z of a galaxy, from each narrow-band filter the network extracts 20 features which potentially contain information about the galaxy photometry and morphology. Every galaxy is then defined by 800 features (40×20) that encode its SED and morphological information. `Aczio` uses these features to predict the photometry in each band and the photo-z of the galaxy (§ 7.3).

In § 7.4.1, we trained our network on galaxy simulations with uncorrelated SEDs (dotted green line in Fig. 7.3), obtaining worse precision than single-band photometry. In the uncorrelated SED case, the input of the network contains the information required to make a prediction at least as accurate as the single-band photometry estimate. However, the photometric information from a narrow band that the network is not predicting is completely uninformative. These features could represent up to 97% of the input (780 out of 800 features). A potential explanation for the worsening of the photometry performance is that the uninformative features add too much noise to the training, hindering the predictions.

To test the effect of uninformative input features, we train a linear neural network to predict the photo-z from the $uBVriz$ galaxy colours in the COSMOS field. The broad-band photometry is from Laigle et al. (2016) and the ground truth redshifts are the spectroscopic redshifts from Ilbert et al. (2008). Starting from the 5 broad-band colours, we have repeated the network's training including additional random features to the input. The solid blue line in Fig. 7.15 shows the photo-z performance as a function of number of random input features in the fiducial case. The photo-z precision has a 20% degradation when adding 5 or 10 additional uncorrelated features. Moreover, this reduction quickly increases for larger numbers of additional random input features. The point corresponding to 195 random features contains the same fraction of uninformative input data as the uncorrelated SED photometry (dotted-green line in Fig. 7.3 and shows a three-times lower photo-z precision than the fiducial case. This indicates that irrelevant input features can degrade the performance even if the input contains the information needed to meet a accuracy level.

In machine learning, *data augmentation* is a technique to increase the training sample by modifying the already existing training samples. Data augmentation is widely spread and has already been implemented in several astronomy projects, e.g. to improve the photo-z performance (Hoyle et al., 2015; Eriksen et al., 2020) and implemented in the galaxy detection and classification (González et al., 2018). We have implemented two data augmentation techniques in order to reduce the effect of the uncorrelated noisy features.

We have first added 50% of noise to the uncorrelated features on the fly while training. This is implemented scattering the random features (θ) with

$$\theta' = \theta + \theta \cdot N(\mu = 0, \sigma = 0.5), \tag{7.20}$$

where $N(\mu, \sigma)$ corresponds to a Gaussian distribution with mean μ and standard deviation σ. This data augmentation methodology has a modest impact on the photo-z precision (dashed green line in Fig. 7.15). The network sees noisy realisations of the same underlying uncorrelated features, so it could still try to learn from those.

The second data augmentation technique consists in producing the uncorrelated noisy features on the fly, in such a way that the network sees completely different noise features in each training iteration thus it cannot learn anything from those. This technique has a very strong impact on the photo-z precision (dotted red line in Fig. 7.15), unfortunately it cannot

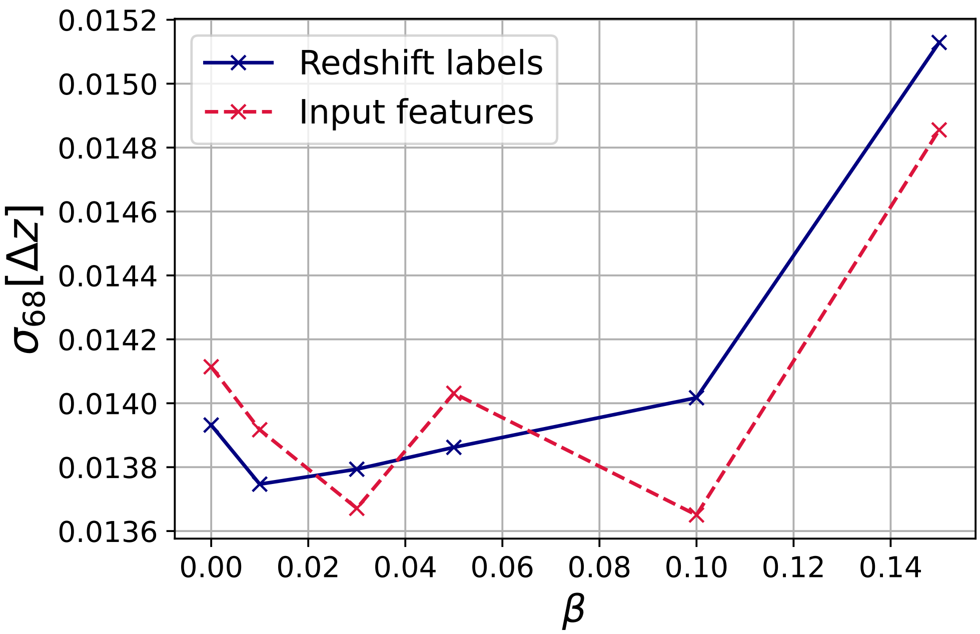

Figure 7.16: Photo-z precision as a function of the scatter added into the input features (dashed red) and the redshift labels (solid blue). The β parameter is defined in Eq. 7.21.

be implemented on data, where we do not know which features are not of interest and we could potentially reject helpful information coming from e.g. the morphology in other bands.

7.B Simulated broad-band photometry

The MDN predicting the photo-zs also uses the $uBVriz$ broad band photometry on top of the information from the narrow-band images. When training `Aczio` on simulations (§ 7.2.3&§ 7.3.3), we can include different levels of noise to the broad-band photometry. We have studied three cases: using the noiseless broad-band photometry, broad-band photometry with a fixed signal-to-noise of 35, and realistic errors. Testing on simulations, the former provides much better photo-zs that the other two options, however also has a worse generalisation when applied to data. The other two methods (fixed and realistic signal-to-noise) present a similar performance both on the simulations and the data. At the end of the day, as we also implement transfer learning to the photo-z predictions using the spectroscopic sample, the choice of noise level between these last options at the first stage of the photo-z training does not have any effect. For simplicity, we use a fixed signal-to-noise on the simulated broad bands.

7.C Scattering the ground-truth redshifts

Label smoothing is a machine-learning technique implemented on supervised neural networks that perturbs the ground-truth targets to make the model less certain of its predictions (Müller et al., 2019). This helps regularising the model and prevents from overfitting. Commonly, label smoothing prevents overfitting and is commonly used classification tasks We implement label smoothing in the transfer learning training. For that, we scatter the ground-truth spectroscopic redshift ($z_{\rm s}$) with a fixed error fraction β, so that

$$z_{\rm s}' = z_{\rm s} + z_{\rm s} \cdot \beta \cdot N(0,1)\,, \tag{7.21}$$

where $N(0,1)$ is a Normal distribution with zero mean and unit variance. The blue solid line in Fig. 7.16 shows the photo-z precision for different values of β, presenting a minimum for $\beta = 0.01$.

We also study the effect of scattering the input features of the photo-z MDNs in the transfer learning phase. We modify the input features following the same recipe as in Eq. 7.21. The dashed red line in Fig. 7.16 shows the photo-z precision for different scatter fractions, showing a minimum at $\beta = 0.03$.

Chapter 8

Summary and conclusions

Astronomy is a science with thousands of years of history. Already in the second millennium B.C, ancient Babylonian astronomers identified the planets in the solar system. Thousands of years after that, in 1774, Messier catalogued extragalactic objects. Photographic plates transformed the field enabling more astronomical observations. These were further extended with the development of the charged coupled devices (CCDs, Chapter 2), which enabled systematic observations of the sky using statistical tools (Chapter 3). In parallel, artificial intelligence and deep learning have revolutionised the creation of algorithms, allowing us to develop statistical tools hard to code by hand through training on large datasets (Chapter 1).

In this thesis, we have explored implementing deep learning techniques to the astronomical data reduction. More specifically, we have focused on developing techniques to reduce data from the PAUS narrow-band survey (Chapter 2). This entails deep-learning methods to estimate the photometry and predict the photo-z directly from astronomical images. We address the photometry and the photo-z estimation in three steps: background-light estimation (Chapter 4), single-band flux measurements (Chapter 5), and multi-band photometry and photo-zs predictions (Chapter 7).

In Chapter 4, we have optimised the background subtraction task with `BKGnet`, a CNN that predicts the background and its associated uncertainty behind target sources. This study focuses on the background-noise prediction in the presence of scattered light, which follows a band-dependent pattern that can be predicted and corrected. We tested modelling scattered light by combining background-ratio maps from several images taken with the same narrow band. We tested modelling scattered-light maps can correct for scattered light, they are not sufficiently accurate, since changes in the observing conditions cause fluctuations in the scattered-light patterns that the correcting template does not capture.

`BKGnet` predicts more precise and accurate background-noise estimates than traditional annulus-based methods (§ 3). In the presence of scattered light, it estimates more accurate background estimates than the correction with scattered-light templates. On PAUCam data, `BKGnet` provides more robust uncertainties than background-annulus estimates, fixing a strong systematic trend with i-band magnitude. Furthermore, `BKGnet` also reduces the `BCNz2` photo-zs outlier relative fraction by a 25% and 35% for a selected 50% and 20% photo-z samples, respectively, while the accuracy is unaffected.

In Chapter 5, we introduce `Lumos`, a CNN built on `BKGnet` that predicts the background-subtracted flux probability distribution. This requires an intrinsic background-noise subtraction that we address based on the `BKGnet` experience. To train `Lumos`, we have developed `Teahupoo`, a set of PAUCam image simulations. `Teahupoo` images use PAUCam image cutouts as background noise, which ensures that the training sample contains examples of distorting effects such as scattered light, cosmic rays, and crosstalk. This has proven helpful for predicting reliable flux measurements on PAUCam observations previously flagged as problematic. `Lumos` measures three times higher signal-to-noise photometry than `MEMBA` for faint galaxies. This leads to a reduction of the photo-z scatter with both `BCNz` and `Deepz`. For `Deepz`, using the `Lumos` photometry reduces the photo-z scatter by 10% on the full catalogue with respect to previous results with `MEMBA` photometry.

Chapter 6 introduces a method to exploit data from narrow-band photometric surveys like PAUS to improve the broad-band photo-zs using multi-task learning. Similar to many other cases, these additional data is only available for a small fraction of the wide-field survey area. Our suggested multi-task approach allows us to improve the broad-band photo-z training using narrow-band data that is only available for a small fraction of the sky.

In the COSMOS field to $i_{\mathrm{AB}} < 23$, our method reduces the photo-z scatter by approximately 16% with respect to a neural network mapping galaxy colours to redshift. Furthermore, it also reduces the number of photo-z outliers by $\approx$ 40%. We studied the potential of the method for the *Euclid* survey using the Flagship simulations, training the network on 30 000 simulated galaxies to $i_{\mathrm{AB}} < 23$, while evaluating the redshift of galaxies to $i_{\mathrm{AB}} < 25$. For galaxies with $24 < i_{\mathrm{AB}} < 25$, the network predicts up to 15% more precise photo-z than the traditional neural network. PAUS photometry in the COSMOS field is publicly available, thus future weak-lensing surveys like *Euclid* and LSST could benefit from this methodology to improve their photo-z estimates.

Finally, in Chapter 7 we have extended `Lumos` to both predict multi-band photometry and the redshift from PAUS narrow-band images. This method allows us to use all the available observed narrow-band images of a galaxy to predict the photometry of a single band. By doing so, we reduce the photometry scatter by three compared to `Lumos` single-band photometry. Most of the improvement comes from using SED information from all the narrow-band images. We also find photo-z measurements better the `BCNz` template-fitting photo-z, and comparable to `Deepz` on the faint end. The photo-z predictions are still work in progress and expected to improve in the near future.

In this theses, we have presented techniques for the data reduction of astronomical images, together with a method to both estimate the photometry and galaxy redshift directly from the images. While this thesis has focused on the PAU Survey data, we envision these approaches to be applicable for other current and future surveys, like *Euclid* and LSST. Traditionally, data reduction techniques have been dominated by classical algorithms, often requiring a significant amount of hand-tuning. Deep learning sidesteps this directly learning from data, but in our case puts requirements on accurate simulations. We consider the combined prediction of photometry and redshift a new and interesting contribution. The `Aczio` network compresses the information from PAUS images from the different wavelengths into one set of features describing the galaxy. This work can be potentially extended to also predict more quantities, like galaxy morphology, SED type, and equivalent width of emission lines. Also, adapting the method to work directly from the raw telescope images could further reduce the need of data pre-processing.

Bibliography

Abadi M., et al., 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, http://tensorflow.org/

Abbott T. M. C., et al., 2018a, Physical Review D: Particles, Fields, Gravitation and Cosmology,, 98, 043526

Abbott T. M. C., et al., 2018b, ApJS, 239, 18

Ahmed I. M., Mahmoud A. M., 2020, arXiv e-prints, p. arXiv:2003.05104

Ahumada R., et al., 2020, Astrophys. J. Suppl., 249, 3

Aihara H., et al., 2018, PASJ, 70, S8

Alam S., et al., 2015, The Astrophysical Journal Supplement Series, 219, 12

Alarcon A., et al., 2021, MNRAS, 501, 6103

Almosallam I. A., Lindsay S. N., Jarvis M. J., Roberts S. J., 2015, MNRAS, 455, 2387

Alpaslan M., et al., 2015, The Astronomical Journal, 451, 3249

Amiaux J., et al., 2012, in Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave. p. 84420Z (arXiv:1209.2228), doi:10.1117/12.926513

Andreon S., Gargiulo G., Longo G., Tagliaferri R., Capuano N., 2000, MNRAS, 319, 700

Arcelin B., Doux C., Aubourg E., Roucelle C., LSST Dark Energy Science Collaboration 2021, MNRAS, 500, 531

Arnouts S., Ilbert O., 2011, Astrophysics Source Code Library, p. ascl:1108.009

Astier P., Antilogus P., Juramy C., Le Breton R., Le Guillou L., Sepulveda E., 2019, A&A, 629, A36

Astropy Collaboration et al., 2013, A&A, 558, A33

Babbedge T. S. R., et al., 2004, MNRAS, 353, 654

Ball N. M., Brunner R. J., Myers A. D., Tcheng D., 2006, ApJ, 650, 497

Baqui P. O., et al., 2021, A&A, 645, A87

Benitez N., et al., 2014, pre-print (arXiv:1403.5237),

Benítez N., 2011, Astrophysics Source Code Library, p. ascl:1108.011

Bertin E., 2006, in Astronomical Data Analysis Software and Systems XV. p. 112

Bertin E., 2011, in Astronomical Data Analysis Software and Systems XX. p. 435

Bertin E., Arnouts S., 1996, Astronomy and Astrophysics Supplement Series, 117, 393

Bertin E., Mellier Y., Radovich M., Missonnier G., Didelon P., Morin B., 2002, in Astronomical Data Analysis Software and Systems XI. p. 228

Bilicki M., et al., 2018, A&A, 616, A69

Bishop C. M., 1994, Technical report, Mixture density networks

Blake C., Bridle S., 2005, MNRAS, 363, 1329

Blasi P., 2013, Astronomy & Astrophysics Reviews,, 21, 70

Bogget U. V., Vervenne V., Vinella R. M., van der Zanden K., Merken P., Vermeiren J. P., 2014, in Defense + Security Symposium.

Bolzonella M., Miralles J. M., Pelló R., 2000, A&A, 363, 476

Bonnett C., 2015a, MNRAS, 449, 1043

Bonnett C., 2015b, MNRAS, 449, 1043

Bordoloi R., Lilly S. J., Amara A., 2010, MNRAS, 406, 881

Boucaud A., et al., 2020, MNRAS, 491, 2481

Bradley L., et al., 2020, astropy/photutils: 1.0.0, doi:10.5281/zenodo.4044744, `https://doi.org/10.5281/zenodo.4044744`

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Breiman L., 2001, Mach. Learn., 45, 5

Bruzual G., Charlot S., 2003, The Astronomical Journal, 344, 1000

Buchanan B. G., Shortliffe E. H., 1984, Rule-based expert systems : the MYCIN experiments of the Stanford Heuristic Programming Project. Reading, Mass. : Addison-Wesley

Buchs R., et al., 2019, MNRAS, 489, 820

Cabayol-Garcia L., et al., 2020, MNRAS, 491, 5392

Cabayol L., et al., 2019, MNRAS, 483, 529

Cabayol L., et al., 2021, MNRAS, 506, 4048

Calabretta M. R., Greisen E. W., 2002, A&A, 395, 1077

Calabrò A., et al., 2018, ApJL, 862, L22

Campbell M., Hoane A., hsiung Hsu F., 2002, Artificial Intelligence, 134, 57

Carliles S., Budavári T., Heinis S., Priebe C., Szalay A. S., 2010, ApJ, 712, 511

Carrasco-Davis R., et al., 2018, pre-print (arXiv:1807.03869),

Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483

Carrasco Kind M., Brunner R. J., 2014a, MNRAS, 438, 3409

Carrasco Kind M., Brunner R. J., 2014b, MNRAS, 442, 3380

Caruana R., 1997, Machine Learning, 28, 41–75

Casas R., et al., 2012, in High Energy, Optical, and Infrared Detectors for Astronomy V. p. 845326, doi:10.1117/12.924640

Casas R., et al., 2016, in Ground-based and Airborne Instrumentation for Astronomy VI. p. 99084K, doi:10.1117/12.2232422

Castander F. J., et al., 2012, in Ground-based and Airborne Instrumentation for Astronomy IV. p. 84466D, doi:10.1117/12.926234

Cavuoti S., et al., 2015, MNRAS, 452, 3100

Choi Y.-Y., Han D.-H., Kim S. S., 2010, Journal of Korean Astronomical Society, 43, 191

Chollet F., et al., 2015, Keras, `https://github.com/fchollet/keras`

Chong De Wei K., Yang A., 2019, in European Physical Journal Web of Conferences. p. 09006, doi:10.1051/epjconf/201920609006

Ćiprijanović A., Kafkes D., Jenkins S., Downey K., Perdue G. N., Madireddy S., Johnston T., Nord B., 2020, arXiv e-prints, p. arXiv:2011.03591

Ćiprijanović A., et al., 2021, arXiv e-prints, p. arXiv:2112.14299

Colless M., et al., 2001a, Monthly Notices of the Royal Astronomical Society, 328, 1039

Colless M., et al., 2001b, MNRAS, 328, 1039

Collister A. A., Lahav O., 2004, PASP, 116, 345

Coupon J., et al., 2009, A&A, 500, 981

Cover T., Hart P., 2006, IEEE Trans. Inf. Theor., 13, 21

Cropper M., et al., 2012, in Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave. p. 84420V (`arXiv:1208.3369`), doi:10.1117/12.927241

Csörnyei G., Dobos L., Csabai I., 2021, MNRAS, 502, 5762

Csurka G., 2017, Domain Adaptation in Computer Vision Applications, 1st edn. Springer Publishing Company, Incorporated

Cuillandre J.-C. J., et al., 2012, in Observatory Operations: Strategies, Processes, and Systems IV. p. 84480M, doi:10.1117/12.925584

Cunha P. A. C., Humphrey A., 2022, arXiv e-prints,

DESI Collaboration et al., 2016, pre-print (arXiv:1611.00036),

D'Isanto A., Polsterer K. L., 2018, A&A, 609, A111

Dalton G., et al., 2014, in Ground-based and Airborne Instrumentation for Astronomy V. p. 91470L (arXiv:1412.0843), doi:10.1117/12.2055132

Damiani F., Maggio A., Micela G., Sciortino S., 1997, ApJ, 483, 350

Darnell T., et al., 2009, in Astronomical Data Analysis Software and Systems XVIII. p. 18

Dawid A. P., 1984, Journal of the Royal Statistical Society. Series A (General), 147, 278

Dawson K. S., et al., 2013, AJ, 145, 10

Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L., 2009, in 2009 IEEE conference on computer vision and pattern recognition. pp 248–255

Desai S., Mohr J. J., Bertin E., Kümmel M., Wetzstein M., 2016a, Astronomy and Computing, 16, 67

Desai S., Mohr J., Bertin E., Kummel M., Wetzstein M., 2016b, Astron. Comput., 16, 67

Desprez G., et al., 2020, Astronomy and Astrophysics, 644

Dey B., Andrews B. H., Newman J. A., Mao Y.-Y., Rau M. M., Zhou R., 2021a, arXiv e-prints, p. arXiv:2112.03939

Dey B., Andrews B. H., Newman J. A., Mao Y.-Y., Rau M. M., Zhou R., 2021b, arXiv e-prints, p. arXiv:2112.03939

Di Sciascio G., 2019, in Journal of Physics Conference Series. p. 012003 (arXiv:1904.06218), doi:10.1088/1742-6596/1263/1/012003

Djorgovski S. G., Gal R. R., Odewahn S. C., de Carvalho R. R., Brunner R., Longo G., Scaramella R., 1998, The Palomar Digital Sky Survey (DPOSS). Conference series Vol. 14, (arXiv:astro-ph/9809187)

Doi M., et al., 2010, AJ, 139, 1628

Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, MNRAS, 476, 3661

Domínguez Sánchez H., et al., 2019, MNRAS, 484, 93

Dreyer J. L. E., 1888, Memoirs of the Royal Astronomical Society, 49, 1

Dreyer J. L. E., 1895, Memoirs of the Royal Astronomical Society,, 51, 185

Driver S. P., et al., 2011, MNRAS, 413, 971

Drlica-Wagner A., et al., 2018, ApJS, 235, 33

Drucker H., Burges C. J. C., Kaufman L., Smola A., Vapnik V., 1996, Support Vector Regression Machines

Eckert K., et al., 2020, MNRAS, 497, 2529

Elvin-Poole J., et al., 2018a, Physical Review D: Particles, Fields, Gravitation and Cosmology,, 98, 042006

Elvin-Poole J., et al., 2018b, Physical Review D: Particles, Fields, Gravitation and Cosmology,, 98, 042006

Erben T., et al., 2013, Monthly Notices of the Royal Astronomical Society, 433, 2545

Eriksen M., et al., 2019, MNRAS, 484, 4200

Eriksen M., et al., 2020, MNRAS, 497, 4565

Feldmann R., et al., 2006, MNRAS, 372, 565

Fisher R. A., 1922, j-PHILOS-TRANS-R-SOC-LOND-A-CONTAIN-PAP-MATH-PHYS-CHARACTER, 222, 309

Fluri J., Kacprzak T., Refregier A., Amara A., Lucchi A., Hofmann T., 2018, Phys. Rev. D., 98, 123518

Freyhammer L., Andersen M., Arentoft T., Sterken C., Nørregaard P., 2001, Experimental Astronomy, 12, 147

Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748

Gaia Collaboration 2018, A&A, 616, A1

Gatti M., et al., 2018, MNRAS, 477, 1664

Geller M. J., Huchra J. P., 1983, The Astrophysical Journal Supplement Series, 52, 61

George D., Huerta E. A., 2018, Physics Letters B, 778, 64

Gerdes D. W., Sypniewski A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. H., Busha M. T., 2010, ApJ, 715, 823

Gholamalinezhad H., Khosravi H., 2020, arXiv e-prints, p. arXiv:2009.07485

Giblin B., et al., 2021, Astronomy & Astrophysics,, 645, A105

Gilliland R. L., 1992, in Howell S. B., ed., Astronomical Society of the Pacific Conference Series Vol. 23, Astronomical CCD Observing and Reduction Techniques. p. 68

Giocoli C., et al., 2021, Astron. Astrophys., 653, A19

Girshick R., 2015, in 2015 IEEE International Conference on Computer Vision (ICCV). pp 1440–1448, doi:10.1109/ICCV.2015.169

Gneiting T., Raftery A. E., Westveld A. H., Goldman T., 2005, Monthly Weather Review, 133, 1098

Goddard H., Shamir L., 2020, The Astrophysical Journal Supplement Series, 251, 28

Gomes Z., Jarvis M. J., Almosallam I. A., Roberts S. J., 2018, MNRAS, 475, 331

González R. E., Muñoz R. P., Hernández C. A., 2018, Astronomy and Computing, 25, 103

Good I. J., 1952, Journal of the Royal Statistical Society. Series B (Methodological), 14, 107

Gray R., Dunning-Davies J., 2008, arXiv e-prints, p. arXiv:0806.4085

Gretton A., Borgwardt K. M., Rasch M. J., Schölkopf B., Smola A., 2012, J. Mach. Learn. Res., 13, 723–773

Guglielmetti F., Fischer R., Dose V., 2009, MNRAS, 396, 165

Gunn J. E., et al., 1998, The Astronomical Journal, 116, 3040

Gupta A., Zorrilla Matilla J. M., Hsu D., Haiman Z., 2018, Phys. Rev. D., 97, 103515

Haigh C., Chamba N., Venhola A., Peletier R., Doorenbos L., Watkins M., Wilkinson M. H. F., 2021, A&A, 645, A107

Han B., Ding H., Zhang Y., Zhao Y., 2015, Proceedings of the International Astronomical Union, 11, 209–209

Hasinger G., et al., 2018, ApJ, 858, 77

Hausen R., Robertson B. E., 2020, ApJS, 248, 20

Heasley J. N., 1999, in Precision CCD Photometry. p. 56

Herbel J., Kacprzak T., Amara A., Refregier A., Lucchi A., 2018a, JCAP, 2018, 054

Herbel J., Kacprzak T., Amara A., Refregier A., Lucchi A., 2018b, JCAP, 2018, 054

Herbel J., Kacprzak T., Amara A., Refregier A., Lucchi A., 2018c, JCAP, 2018, 054

Hersbach H., 2000, Weather and Forecasting - WEATHER FORECAST, 15, 559

Heydon-Dumbleton N. H., Collins C. A., MacGillivray H. T., 1989, MNRAS, 238, 379

Heymans C., et al., 2012, The Astronomical Journal, 427, 146

Hildebrandt H., Pielorz J., Erben T., van Waerbeke L., Simon P., Capak P., 2009, A&A, 498, 725

Hildebrandt H., et al., 2012a, MNRAS, 421, 2355

Hildebrandt H., et al., 2012b, MNRAS, 421, 2355

Hildebrandt H., et al., 2021, A&A, 647, A124

Hoekstra H., Jain B., 2008, Annual Review of Nuclear and Particle Science, 58, 99

Honscheid K., DePoy D. L., 2008, pre-print (arXiv:0810.3600),

Hoyle B., Rau M. M., Bonnett C., Seitz S., Weller J., 2015, Monthly Notices of the Royal Astronomical Society, 450, 305

Hoyle B., et al., 2018, MNRAS, 478, 592

Hubble E. P., 1922, The Astrophysical Journal, 56, 162

Hubble E. P., 1926, The Astrophysical Journal, 64

Hubble E. P., 1927, The Observatory, 50, 276

Hubble E., Tolman R. C., 1935, The Astrophysical Journal, 82, 302

Ilbert O., et al., 2006, A&A, 457, 841

Ilbert O., et al., 2008, in Panoramic Views of Galaxy Formation and Evolution. p. 169

Ilbert O., et al., 2009, ApJ, 690, 1236

Ioffe S., Szegedy C., 2015, arXiv e-prints, p. arXiv:1502.03167

Ivezić Ž., et al., 2019a, ApJ, 873, 111

Ivezić Ž., et al., 2019b, The Astrophysical Journal, 873, 111

Joachimi B., Schneider P., 2010, Astronomy & Astrophysics,, 517, A4

Johnston H., et al., 2019, Astronomy & Astrophysics,, 624, A30

Johnston H., et al., 2021a, A&A, 646, A147

Johnston H., et al., 2021b, A&A, 648, A98

Jones E., Singal J., 2017, A&A, 600, A113

Jones E., Oliphant T., Peterson P., et al., 2001, SciPy: Open source scientific tools for Python, http://www.scipy.org/

Kashino D., et al., 2019, ApJS, 241, 10

Kelley H. J., 1960, Ars Journal, 30, 947

Kennicutt Robert C. J., 1998, ARA&A,, 36, 189

Khalifa N. E. M., Taha M. H. N., Hassanien A. E., Selim I. M., 2017, pre-print (arXiv:1709.02245),

Kim E. J., Brunner R. J., 2016, MNRAS, 464, 4463

Kim D.-W., Yeo D., Bailer-Jones C. A. L., Lee G., 2021, Astronomy & Astrophysics,, 653, A22

Kingma D. P., Ba J., 2014, pre-print (arXiv:1412.6980),

Klambauer G., Unterthiner T., Mayr A., Hochreiter S., 2017, arXiv e-prints, p. arXiv:1706.02515

Knox L., Song Y.-S., Zhan H., 2006, ApJ, 652, 857

Kohonen T., 1982, Biological Cybernetics, 43, 59

Kriek M., et al., 2015, ApJS, 218, 15

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, pp 1097–1105, http://dl.acm.org/citation.cfm?id=2999134.2999257

Kron R. G., 1980, ApJS, 43, 305

Kuijken K., 2008, A&A, 482, 1053

Kuijken K., et al., 2015, MNRAS, 454, 3500

Kuijken K., et al., 2019, A&A, 625, A2

Kullback S., Leibler R. A., 1951, The Annals of Mathematical Statistics, 22, 79

Lacy M., et al., 2020, PASP, 132, 035001

Laigle C., et al., 2016, ApJS, 224, 24

Lamareille F., 2010, A&A, 509, A53

Lang D., Hogg D. W., Schlegel D. J., 2016, The Astronomical Journal, 151, 36

Laureijs R., et al., 2011, pre-print (arXiv:1110.3193), LeCun Y., Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W., Jackel L. D., 1989, Neural Computation, 1, 541

Leavitt H. S., Pickering E. C., 1912, Harvard College Observatory Circular, 173, 1

Lecun Y., Bottou L., Bengio Y., Haffner P., 1998, Proceedings of the IEEE, 86, 2278

Lee K.-G., et al., 2018, ApJS, 237, 31

Leistedt B., Hogg D. W., 2017, The Astrophysical Journal, 838, 5

Lesser M., 2015, Publications of the Astronomical Society of the Pacific, 127, 000

Liebel L., Körner M., 2018, arXiv e-prints,

Lilly S. J., et al., 2007, ApJS, 172, 70

Lilly S. J., et al., 2009, ApJS, 184, 218

Lintott C. J., et al., 2008, MNRAS, 389, 1179

Liou C.-Y., Cheng W.-C., Liou J.-W., Liou D.-R., 2014, Neurocomputing, 139, 84

Liu Y., Zhang Y.-M., Zhang X.-Y., Liu C.-L., 2016, Pattern Recognition, 55, 58

Liu X., Deng Z., Yang Y., 2018, pre-print (arXiv:1809.10198),

MacGillivray H. T., Martin R., Pratt N. M., Reddish V. C., Seddon H., Alexander L. W. G., Walker G. S., Williams P. R., 1976, MNRAS, 176, 649

Machado E., et al., 2016. , doi:10.1109/IJCNN.2016.7727189

Magnier E. A., et al., 2020, ApJS, 251, 6

Margon B., 1999, Philosophical Transactions of the Royal Society of London Series A, 357, 93

Martí P., Miquel R., Castander F. J., Gaztanaga E., Eriksen M., Sànchez C., 2014, MNRAS, 442, 92

Massey P., Jacoby G. H., 1992, in Astronomical CCD Observing and Reduction Techniques. p. 240

Massey R., Refregier A., 2005, MNRAS, 363, 197

Masters D., et al., 2015, Astrophys. J., 813, 53

Masters D. C., Stern D. K., Cohen J. G., Capak P. L., Rhodes J. D., Castander F. J., Paltani S., 2017, ApJ, 841, 111

Masters D. C., et al., 2019, ApJ, 877, 81

Matilla J. M. Z., Sharma M., Hsu D., Haiman Z., 2020, Phyus. Rev. D., 102, 123506

Mauri N., et al., 2020, J. Phys. Conf. Ser., 1342, 012122

Mazin D., 2019, in 36th International Cosmic Ray Conference (ICRC2019). p. 741 (arXiv:1907.08530)

McCarthy J., Minsky M., Rochester N., Shannon C. E., 2006, AI Mag., 27, 12

Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, arXiv e-prints,

Merlin E., et al., 2015, A&A, 582, A15

Merloni A., et al., 2012, arXiv preprint arXiv:1209.3114

Messier C., 1774, Mémoires de l'Académie Royale des Sciences

Mickaelian A. M., 2016, Baltic Astronomy, 25, 75

Mighell K. J., 1999, in Craine E. R., Crawford D. L., Tucker R. A., eds, Astronomical Society of the Pacific Conference Series Vol. 189, Precision CCD Photometry. p. 50

Miller L., et al., 2013, MNRAS, 429, 2858

Minkowski R. L., Abell G. O., 1963, in Strand K. A., ed., , Basic Astronomical Data: Stars and Stellar Systems. p. 481

Moeskops P., Wolterink J. M., van der Velden B. H. M., Gilhuijs K. G. A., Leiner T., Viergever M. A., Išgum I., 2017, arXiv e-prints,

Moles M., et al., 2008, AJ, 136, 1325

Molino A., et al., 2013, MNRAS, 441

Morganson E., et al., 2018, Publications of the Astronomical Society of the Pacific, 130, 074501

Müller R., Kornblith S., Hinton G., 2019, arXiv e-prints, p. arXiv:1906.02629

Myles J., et al., 2021, Mon. Not. Roy. Astron. Soc., 505, 4249

Nair V., Hinton G. E., 2010, in Proceedings of the 27th International Conference on International Conference on Machine Learning. ICML'10. Omnipress, Madison, WI, USA, p. 807–814

Narayan G., et al., 2018, The Astrophysical Journal Supplement Series, 236, 9

Newman J. A., et al., 2015, Astroparticle Physics, 63, 81

Ni Q., Timlin J., Brandt W. N., Yang G., 2019, Research Notes of the AAS, 3, 5

Oh K.-S., Jung K., 2004, Pattern Recognition, 37, 1311

Okamura S., 2020, in Boquien M., Lusso E., Gruppioni C., Tissera P., eds, Conference series Vol. 341, Panchromatic Modelling with Next Generation Facilities. pp 1–11, doi:10.1017/S1743921319004800

Padilla C., et al., 2016, in Ground-based and Airborne Instrumentation for Astronomy VI. p. 99080Z, doi:10.1117/12.2231884

Padilla C., et al., 2019a, AJ, 157, 246

Padilla C., et al., 2019b, AJ, 157, 246

Paillassa M., Bertin E., Bouy H., 2020, Astronomy & Astrophysics,, 634, A48

Pan S. J., Yang Q., 2010, IEEE Transactions on Knowledge and Data Engineering, 22, 1345

Parks D., Prochaska J. X., Dong S., Cai Z., 2018, The Astronomical Journal, 476, 1151

Pasquet-Itam J., Pasquet J., 2018, A&A, 611, A97

Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, A&A, 621, A26

Paszke A., et al., 2017, in NIPS 2017 Workshop on Autodiff. https://openreview.net/forum?id=BJJsrmfCZ

Paturel G., Petit C., Prugniel P., Theureau G., Rousseau J., Brouty M., Dubois P., Cambrésy L., 2003, Astronomy & Astrophysics, 412, 45

Paulino-Afonso A., Sobral D., Darvish B., Ribeiro B., Stroe A., Best P., Afonso J., Matsuda Y., 2018, A&A, 620, A186

Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825

Perlmutter S., et al., 1999, The Astrophysical Journal, 517, 565

Planck Collaboration et al., 2020, A&A, 641, A6

Polletta M., et al., 2007, The Astrophysical Journal, 663, 81

Polsterer K. L., D'Isanto A., Lerch S., 2021, arXiv e-prints, p. arXiv:2103.03780

Popowicz A., Smolka B., 2015, MNRAS, 452, 809

Price-Whelan A. M., et al., 2018, AJ, 156, 123

Pych W., 2004, Publications of the Astronomical Society of the Pacific, 116, 148

Refregier A., 2003, MNRAS, 338, 35

Renard P., et al., 2021, MNRAS, 501, 3883

Renard P., et al., 2022, arXiv e-prints, p. arXiv:2201.04411

Riess A. G., et al., 1998, The Astronomical Journal, 116, 1009

Robotham A. S. G., et al., 2014, The Astronomical Journal, 444, 3986

Robotham A. S. G., Davies L. J. M., Driver S. P., Koushan S., Taranu D. S., Casura S., Liske J., 2018, MNRAS, 476, 3137

Rodighiero G., et al., 2011, ApJL, 739, L40

Romanishin W., 2014, An Introduction to Astronomical Photometry Using Ccds. Createspace Independent Pub, https://books.google.es/books?id=0nbMoQEACAAJ

Rowe B. T. P., et al., 2015, Astronomy and Computing, 10, 121

Sadeh I., Abdalla F. B., Lahav O., 2016, PASP, 128, 104502

Salvato M., Ilbert O., Hoyle B., 2019a, Nature Astronomy, 3, 212

Salvato M., Ilbert O., Hoyle B., 2019b, Nature Astronomy, 3, 212

Samuroff S., et al., 2019, Monthly Notices of the Royal Astronomical Society, 489, 5453

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, The Astrophysical Journal, 500, 525

Schmidt S. J., Ménard B., Scranton R., Morrison C., McBride C. K., 2013, MNRAS, 431, 3307

Schmidt S. J., et al., 2020, MNRAS, 499, 1587

Scodeggio M., et al., 2018, A&A, 609, A84

Sérsic J. L., 1968, Bulletin of the Astronomical Institutes of Czechoslovakia, 19, 105

Sevilla-Noarbe I., Etayo-Sotos P., 2015, Astronomy and Computing, 11, 64

Skrutskie M. F., et al., 2006, The Astronomical Journal, 131, 1163

Smith J., et al., 2005, AAS, 37, 1379

Song X., Zhao X., Fang L., Hu H., 2020, International Journal of Computer Vision, 128

Soo J. Y. H., et al., 2018, MNRAS, 475, 3613

Soo J. Y. H., et al., 2021, MNRAS, 503, 4118

Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, Journal of Machine Learning Research, 15, 1929

Stetson P. B., 1987, Publications of the Astronomical Society of the Pacific, 99, 191

Stothert L., et al., 2018, MNRAS, 481, 4221

Suchyta E., et al., 2016, MNRAS, 457, 786

Sultana F., Sufian A., Dutta P., 2019, pre-print (arXiv:1905.03288),

Sánchez C., et al., 2014, MNRAS, 445, 1482

Tan C., Sun F., Kong T., Zhang W., Yang C., Liu C., 2018, Lecture Notes in Computer Science, p. arXiv:1808.01974

Teeninga P., Moschini U., Trager S. C., Wilkinson M. H. F., 2015, in 2015 IEEE International Conference on Image Processing (ICIP). pp 1046–1050, doi:10.1109/ICIP.2015.7350959

Tercan H., Guajardo A., Heinisch J., Thiele T., Hopmann C., Meisen T., 2018, Procedia CIRP, 72, 185

Tewes M., Kuntzer T., Nakajima R., Courbin F., Hildebrandt H., Schrabback T., 2019, A&A, 621, A36

The Dark Energy Survey Collaboration 2005, pre-print (arXiv:051034)

Tonello N., et al., 2019, Astronomy and Computing, 27, 171

Torrey L., Shavlik J., , Transfer Learning

Tortorelli L., et al., 2021, JCAP, 2021, 013

Troxel M. A., Ishak M., 2015, Physics Reports, 558, 1

Trujillo I., Aguerri J. A. L., Cepa J., Gutiérrez C. M., 2001, The Astronomical Journal, 328, 977

Tucker D. L., et al., 2007, in The Future of Photometric, Spectrophotometric and Polarimetric Standardization. p. 187 (arXiv:astro-ph/0611137)

Urrutia T., et al., 2019, A&A, 624, A141

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, Computational Intelligence and Neuroscience, 2018, 1

Wadadekar Y., 2005, PASP, 117, 79

Wang D., Zhang Y.-X., Liu C., Zhao Y.-H., 2008, CJAA, 8, 119

Werbos P. J., 1982, in System Modeling and Optimization. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 762–770

Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, ApJL, 754, L29

Widenhorn R., Dunlap J. C., Bodegom E., 2010, IEEE Transactions on Electron Devices, 57, 581

Willett K. W., et al., 2013, The Astronomical Journal, 435, 2835

Wilson T. L., 2011, pre-print (arXiv:1111.1183),

Wright A. H., et al., 2016, MNRAS, 460, 765

Wright A. H., Hildebrandt H., van den Busch J. L., Heymans C., 2020a, A&A, 637, A100

Wright A. H., Hildebrandt H., van den Busch J. L., Heymans C., Joachimi B., Kannawadi A., Kuijken K., 2020b, A&A, 640, L14

Wu W.-M., 1986, Int. J. Mod. Phys. A, 1, 117

Xu B., Wang N., Chen T., Li M., 2015, p. arXiv:1505.00853

Yanminsun Wong A., Kamel M. S., 2011, International Journal of Pattern Recognition and Artificial Intelligence, 23

Yee H. K. C., 1991, PASP, 103, 396

York D. G., et al., 2000, AJ, 120, 1579

Zackay B., Ofek E. O., 2017a, ApJ, 836, 187

Zackay B., Ofek E. O., 2017b, ApJ, 836, 188

Zeiler M. D., Fergus R., 2013, pre-print (arXiv:1311.2901),

Zhang K., Bloom J., 2019, The Journal of Open Source Software, 4, 1651

Zhang Y., Yang Q., 2021, IEEE Transactions on Knowledge and Data Engineering, pp 1–1

Zhang Y., Zhao Y., 2015, Data Science Journal, 14, 11

Zhao G.-B., et al., 2015, MNRAS, 457

Zhao Z.-Q., Zheng P., Xu S.-t., Wu X., 2018, pre-print (arXiv:1807.05511),

Zhao Z., Zheng P., Xu S., Wu X., 2019, IEEE Transactions on Neural Networks and Learning Systems, 30, 3212

Zheng H., Zhang Y., 2012a, Support vector machines for photometric redshift measurement of quasars. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8451, SPIE, doi:10.1117/12.925761, `https://doi.org/10.1117/12.925761`

Zheng H., Zhang Y., 2012b, in Radziwill N. M., Chiozzi G., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8451, Software and Cyberinfrastructure for Astronomy II. p. 845134, doi:10.1117/12.925314

Zhou X., et al., 2007, AJ, 127, 3642

Zhu X.-P., Dai J.-M., Bian C.-J., Chen Y., Chen S., Hu C., 2019, Astrophysics & Space Science, 364, 55

Zhuang F., Qi Z., Duan K., Xi D., Zhu Y., Zhu H., Xiong H., He Q., 2019, preprint (arXiv:1911.02685),

de Jong J. T. A., et al., 2013, The Messenger, 154, 44

de Jong J. T. A., et al., 2015, A&A, 582, A62

de Vaucouleurs G., de Vaucouleurs A., Corwin Herold G. J., Buta R. J., Paturel G., Fouque P., 1991, Third Reference Catalogue of Bright Galaxies

van Dokkum P. G., 2001, PASP, 113, 1420

van den Busch J. L., et al., 2020, A&A, 642, A200