The PAU Survey & *Euclid*: Improving broad-band photometric redshifts with multi-task learning[*]

L. Cabayol[1][†], M. Eriksen[1,2], J. Carretero[2,3], R. Casas[3,4], F. J. Castander[3,4], E. Fernández[1], J. García-Bellido[6],
E. Gaztanaga[3,4], H. Hildebrandt[7], H. Hoekstra[10], B. Joachimi[5], R. Miquel[1,8], C. Padilla[1], A. Pocino[3,4], E. Sánchez[9],
S. Serrano[3,4], I. Sevilla-Noarbe[9], M. Siudek[1,11], P. Tallada-Crespí[2,9]

Affiliations are listed at the end of the paper.

ABSTRACT

Current and future imaging surveys require estimating photometric redshifts (photo-z) of millions of galaxies. Improving the photo-z quality is a major challenge to advance our understanding of cosmology. In this paper, we explore how the synergies between narrow-band photometric data and large imaging surveys can be exploited to improve broad-band photometric redshifts. We use a multi-task learning (MTL) network to improve broad-band photo-z estimates by simultaneously predicting the broad-band photo-z and the narrow-band photometry from the broad-band photometry. The narrow-band photometry is only required in the training field, which enables better photo-z predictions also for the galaxies without narrow-band photometry in the wide field. This technique is tested with data from the Physics of the Accelerating Universe Survey (PAUS) in the COSMOS field. We find that the method predicts photo-z that are 14% more precise down to magnitude $i_{AB} < 23$ while reducing the outlier rate by 40% with respect to photo-zs estimated solely from broad bands. Furthermore, MTL significantly reduces the photo-z bias for high-redshift galaxies, improving the redshift distributions for tomographic bins with $z > 1$. Applying this technique to deeper samples is crucial for future surveys like *Euclid* or LSST. For simulated data, training on a sample with $i_{AB} < 23$, the method reduces the photo-z scatter by 15% for all galaxies with $24 < i_{AB} < 25$. We also study the effects of extending the training sample with photometric galaxies using PAUS high-precision photo-zs, which further reduces the photo-z scatter.

Key words. Galaxies: photometry, Methods: data analysis, Surveys, Cosmology: observations

1. Introduction

Over the last 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. 2019) and are a limiting factor in the accuracy of cosmology measurements (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, LSST Science Collaboration et al. 2009) 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 $\sigma_{z_p-z_s} < 0.12$ (Sánchez et al. 2014), with $\sigma_{z_p-z_s}$ 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 scaled photo-z bias $\langle \Delta z \rangle < 0.003$, with $\Delta z := (z_p - z_s)/(1 + z_s)$, 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. 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 in 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 2015, 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; Soo et al. 2021), and directly predicting the photo-z from the astronomical images (Pasquet-Itam & Pasquet 2018; Pasquet et al. 2019; Chong & Yang 2019).

The 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 typ-

ically cover smaller sky areas due to the increased telescope time needed to cover the same wavelength range. In this paper, 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 equipped with the PAUCam instrument (Castander et al. 2012; Padilla et al. 2016, 2019), 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).

The paper is structured as follows. In Sect. 2, we present the data used throughout the paper. Section 3 introduces multi-task learning and the method developed and tested in this work. In Sect. 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 Sect. 5 using simulated galaxies. Finally, we use self-organising maps (SOM) to explore the photo-*z* distribution of COSMOS galaxies in colour-space (Sect. 6) and to have a better understanding of the underlying mechanism of our method (Sect. 7).

2. Data

In this section, we present the PAUS data (Sect. 2.1) and the photometric redshift galaxy sample (Sect. 2.2). The broad-band data and the spectroscopic sample are introduced in Sect. 2.3 and Sect. 2.4, respectively, while Sect. 2.5 shows the galaxy simulations used in the paper.

2.1. PAUS data

PAUS data are taken at the William Herschel Telescope (WHT), at the Observatorio del Roque de los Muchachos in La Palma (Canary Islands). Images are taken with the PAUCam instrument (Castander et al. 2012; Padilla et al. 2019), an optical camera equipped with 40 narrow bands covering a wavelength range from 4500 to 8500 Å (Casas et al. 2016). The narrow-band filters have 130 Å Full-Width-Half-Maximum (FWHM) and separation

between consecutive bands of 100 Å. They are mounted in five trays with eight filters per tray that can be exchanged and placed in front of the CCDs. The narrow-band filter set effectively provides a high-resolution photometric spectrum ($R \sim 50$). This allows PAUS to measure high-precision photo-*z*s to faint magnitudes ($i_{AB} < 23$) while covering a large sky area (Martí et al. 2014).

With a template-fitting algorithm, PAUS reaches a photo-*z* precision $\sigma_z/(1 + z) = 0.0035$ 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 in Eriksen et al. (2020). Furthermore, with a combination of PAUS narrow bands and 26 broad and intermediate bands covering the UV, visible, and near infrared spectral range, Alarcon et al. (2021) presented an unprecedented precise photo-*z* catalogue for COSMOS (Scoville et al. 2007) with $\sigma_z/(1 + z) = 0.0049$ for galaxies with $i_{AB} < 23$. The excellent PAUS photo-*z* precision enables studies of intrinsic galaxy alignments and 3D galaxy clustering (Johnston et al. 2021a), measuring galaxy properties (Tortorelli et al. 2021) or measuring the D4000 Å spectral break (Renard et al. in prep.).

PAUS has been observing since the 2015B semester, and as of 2021B, PAUS has taken data during 160 nights. It partially covers the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) fields[1] W1, W2, W3, and the full COSMOS field[2]. In the W2 field, so far PAUS has observed in the overlapping region with the GAMA 9-hour field[3] (G09). Currently, PAUS data have a 40 narrow-band coverage of 10 deg^2 in each, W1 and G09, 20 deg^2 in W3, and 2 deg^2 in COSMOS. The PAUS data are stored at the Port d'Informació Científica (PIC), where the data are processed and distributed (Tonello et al. 2019). This paper uses data from the COSMOS field (Scoville et al. 2007), which were taken in the semesters 2015B, 2016A, 2016B, and 2017B. The complete PAUS photometric catalogue in COSMOS comprises 64 476 galaxies to $i_{AB} < 23$ in 40 narrow-band filters. This corresponds to approximately 12,5 million galaxy observations (5 observations per galaxy and narrow-band filter).

PAUS has developed two methods to extract the galaxy photometry: a forced aperture algorithm (`MEMBA`) and a deep learning-based pipeline (`Lumos`, Cabayol-Garcia et al. 2020; Cabayol et al. 2021). In this study, we have found that the resulting photo-*z* performance with both photometric approaches is very similar. In the COSMOS field, the parent detection catalogue is provided by Ilbert et al. (2009) and the photometry calibration is relative to SDSS stars (Castander et al. in prep.). A brief description of the photometric calibration can be found in Eriksen et al. (2019).

2.2. Photometric redshift sample

Throughout the paper, we will 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 Sect. 2 in Alarcon et al. 2021, for more details). The PAUS+COSMOS

---

[1]http://www.cfht.hawaii.edu/Science/CFHTLS_Y_WIRCam/cfhtlsdeepwidefields.html

[2]http://cosmos.astro.caltech.edu/

[3]https://www.astro.ljmu.ac.uk/ ikb/research/gama_fields/

photo-zs reach a precision of $\sigma_z/(1+z) = 0.0036$ and $\sigma_z/(1+z) = 0.0049$ for galaxies at $i_{AB} < 21$ and $i_{AB} < 23$, respectively. These photo-zs are more precise and less biased than those from Ilbert et al. (2009), COSMOS30 hereafter, which use a combination of 30 broad-, intermediate-, and narrow-band filters.

2.3. Broad-band data

The broad-band data used in this paper are from Laigle et al. (2016, COSMOS2015 hereafter), 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 $i_{AB} < 23$ and on redshift $z < 1.5$, which results in a catalogue with around 33 000 galaxies of which approximately 9000 have spectroscopic redshift.

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 $15 < i_{AB} < 23$ and a redshift range of $0.1 < z < 1.2$. We extend that 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. 2001), 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.

2.5. Galaxy mocks

In Sect. 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 Ilbert et al. (2009). The spectral energy distribution (SED) including its extinction of the best matching COSMOS galaxy is assigned to each simulated galaxy. 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 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.

3. Multi-task neural network to improve broad-band photo-z

In this section, we describe MTL (Sect. 3.1) and present the networks and training procedures used throughout the paper (Sect. 3.2).
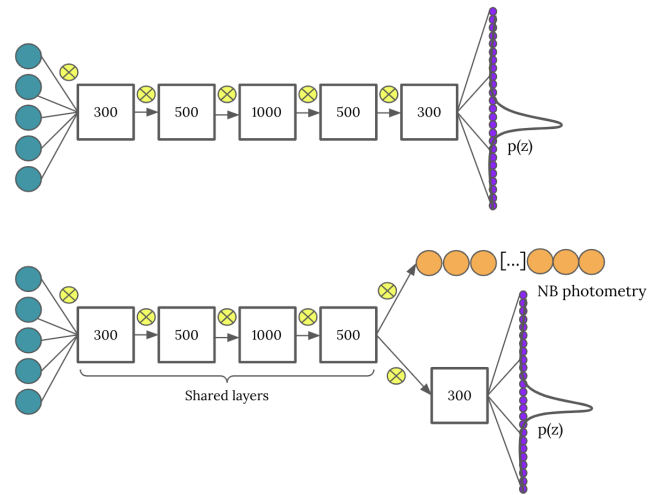

Fig. 1: *Top:* Baseline network architecture. The input contains five colours that propagate through six fully connected layers. Each layer is followed by a dropout layer, which is represented by a yellow-crossed circle. *Bottom:* Multitask learning network. It builds on the baseline network and adds an extra output layer for the additional task of predicting the narrow-band photometry.

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 to 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.

3.2. Model architecture and training procedures

Broadly, there are two types of MTL-network architectures: 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 1 shows the two networks used in this work. 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 $0 < z < 1.5$. 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. 1.

The bottom panel in Fig. 1 represents the MTL network introduced in the paper, which 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 with an extra output layer for the additional task of predicting the narrow-band photometry.

The photometric redshifts are trained with a cross-entropy loss function (Good 1952)

$$\mathcal{L}_z := \sum_{c=0}^{1500} \left[p_c(z) \log p_c(z) \right],\qquad(1)$$

where 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 MTL network also uses a least absolute deviation loss function to optimise the narrow-band photometry reconstruction

$$\mathcal{L}_{NB} := \frac{\sum_i \left| NB_i^{pred} - NB_i^{obs} \right|}{N},\qquad(2)$$

where NB_i^{pred} and NB_i^{obs} are the predicted and observed narrow-band colours in the i-th filter, respectively, and N is the number of narrow bands minus one.

We can identify four training methodologies:

1. BB: 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. 1);

2. BB+MTL: This training procedure includes MTL. It maps the broad-band photometry to photo-z and narrow-band photometry representations. The training sample can contain galaxies without spectroscopy. All galaxies (with and without spectroscopy) are used to optimise the shared parameters in the hidden layers and the PAUS narrow-band photometry output layer. The photo-z prediction layer is only trained with spectroscopic galaxies. The loss function

$$\mathcal{L}_{BB+MTL} := \mathcal{L}_z + \mathcal{L}_{NB}\qquad(3)$$

combines a cross-entropy loss for the photo-z prediction (Eq. 1) and a least absolute deviation to optimise the narrow-band reconstruction (Eq. 2).

Furthermore, we considered two variants in the training procedure to explore the possibility of using high-precision photometric redshifts (Sect. 2.2) to train the networks:

3. BB+PAUS$_z$: This is a variation of the BB method. The training sample extends to galaxies with 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. BB+MTL+PAUS$_z$: This is a variation of the BB+MTL 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 BB+MTL method, here all galaxies are used to train the photo-z prediction and the narrow-band photometry reconstruction;

The networks are implemented in `PyTorch` (Paszke et al. 2017), and all the training procedures use an `Adam` optimizer (Kingma & Ba 2015) for 100 epochs with an initial learning rate of 10^{-3} that reduces a factor of ten every 25 epochs.

4. photo-z performance in the COSMOS field

In this section, we show the photo-z performance of our method on galaxies with $i_{AB} < 23$ and $z < 1.5$ in the COSMOS field. We will study the effect that MTL has on the dispersion (Sect. 4.2) and the bias (Sect. 4.3) of the predicted photo-zs, while Sect. 4.4 investigates the effect of MTL on the redshift distributions $N(z)$.

4.1. Photo-z performance metrics

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

$$\Delta z := (z_p - z_t) / (1 + z_t),\qquad(4)$$

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

$$\sigma_{68} := \frac{1}{2} \left[Q_{84}(\Delta z) - Q_{16}(\Delta z) \right],\qquad(5)$$

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

The ground-truth redshift is defined as the spectroscopic redshift if available and as the PAUS+COSMOS photo-z (Sect. 2.2) otherwise [4].

If it is not specified by the method our networks are trained with spectroscopic galaxies 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 (Sect. 3.2).

In order to estimate the photo-zs of the complete COSMOS catalogue, the networks are trained independently nine times with 8000 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 33 000 galaxies, which represents approximately 3.5 times more galaxies in the training sample. In order to evaluate 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 5000 galaxies

---

[4]The PAUS+COSMOS photo-zs used to evaluate the precision of non-spectroscopic galaxies (see Sect. 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.
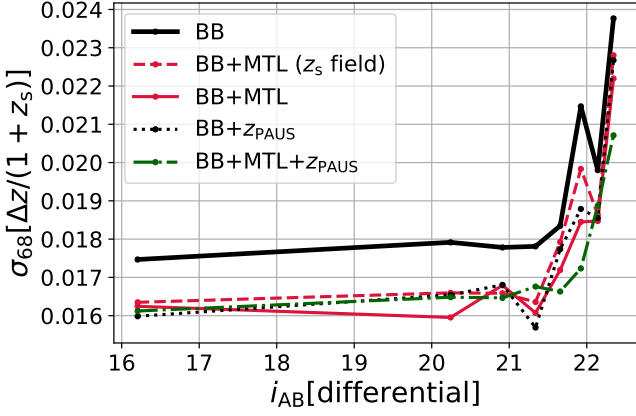
Fig. 2: Photo-z dispersion in equally populated magnitude differential bins to $i_{AB} < 23$. Each line corresponds to a different training procedure (see Sect. 3.2). While the black line corresponds to a baseline training, the other coloured lines include MTL (red and green lines) and data augmentation with photo-zs from the PAUS+COSMOS catalogue as ground-truth redshifts (blue and green lines).
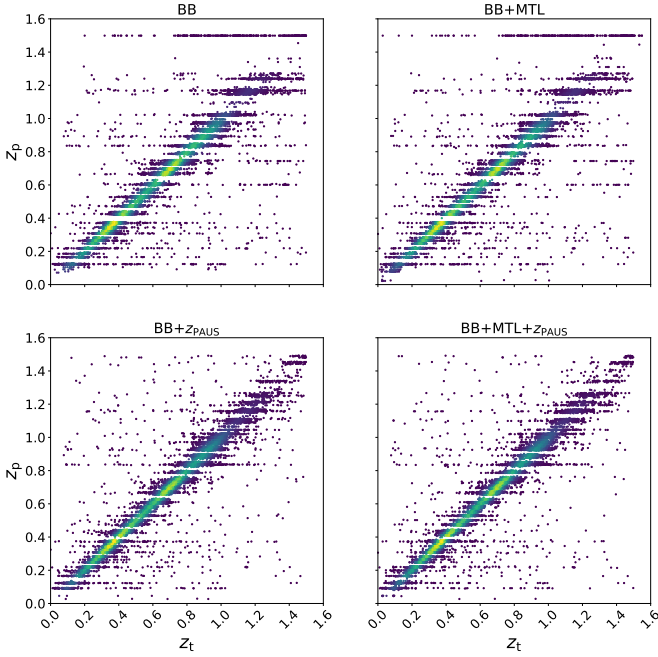

Fig. 3: Ground-truth redshift (z_t, spectroscopic or PAUS+COSMOS) versus photo-z (z_p) scatter plot for the four networks presented in Sect. 3.2.

4.2. Photo-z dispersion

Table 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 Sect. 3.2. These results are presented in more detail in Fig. 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 BB in Sect. 3.2). This method is trained on the spectroscopic sample and provides a $\sigma_{68} = 0.022$ 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. 2 use a MTL training (method BB+MTL in Sect. 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 $\sigma_{68} = 0.021$, which is a 4% improvement with respect to the baseline methodology (black line, method BB in Sect. 3.2). The method has a larger effect at the bright end, while the photo-zs of fainter galaxies in the data set ($22 < i_{AB} < 23$) 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 9000 to 30 000 galaxies, which results in a precision of $\sigma_{68} = 0.0191$ corresponding to a 14% 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 dotted black line in Fig. 2 also corresponds to a direct mapping of the broad-band photometry to photo-z. However, in contrast to the solid black line, this case is trained on an extended sample including galaxies without spectroscopic redshifts (method BB + PAUS$_z$ in Sect. 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 $\sigma_{68} = 0.0195$, which corresponds to a 12% improvement with respect to the baseline training. Note that this method does not use MTL, but its effect in the photo-z performance is similar to including it (the solid red line). In Sect. 6 we discuss the underlying mechanism that causes MTL with PAUS to improve the photo-zs.

With spectroscopy and 20 000 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.

Table 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 $i_{AB} < 23$. 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 fourth column presents the fraction of photo-z outliers on the full sample.

	z_s sample	COSMOS	Outliers
BB	2.00	2.19	2.6
BB+MTL	1.80	1.91	2.2
BB+z_{PAUS}	1.78	1.95	1.5
BB+MTL+z_{PAUS}	1.89	1.89	1.5

The best photo-z performance is achieved combining MTL and photo-z data augmentation with PAUS+COSMOS data (method BB + MTL + PAUS$_z$ in Sect. 3.2), which corresponds to the dotted green line in Fig. 2. This method gives a 16% improvement with respect to the baseline network, with a precision of $\sigma_{68} = 0.0189$.

To complete the analysis, Fig. 3 shows the photo-z vs. true-z scatter plot for the four networks already presented. In the top panels, the broad-band-only (BB) and MTL network (BB+MTL) show stripes around particular photo-z values at high-redshift, which is significantly reduced in the bottom panels (methods extending the training sample with PAUS galaxies, i.e. BB+z_{PAUS} and BB+MTL+z_{PAUS}). Also, the BB method assigns high-redshift galaxies to the highest redshift bin ($z = 1.5$), causing an empty region at $z_p > 1.2$. This is partially fixed in the BB+MTL method and drastically improved in the methodologies including z_{PAUS}. These effects are potentially triggered by the increment of training examples, which is more remarkable at high-redshift. Furthermore, both top panels also present empty photo-z stripes at e.g. $z \sim 0.4$ and $z \sim 0.6$, which again fade when the training sample is extended by including the PAUS+COSMOS photo-zs (bottom panels).

4.3. Photo-z bias and outlier rate

In this subsection, we show the reduction in bias for the photo-z predictions with the MTL networks with respect to the baseline broad-band network. The left panel in Fig. 4 shows the photo-z bias in equally spaced redshift bins of width 0.1 in the redshift range $0.1 < z_t < 1.5$. We have excluded the first redshift bin from the analysis since there are almost no galaxies with $z_t < 0.07$, which caused a bias at very low redshift[5].

Overall, for $z_t < 1$ the four methods presented in Sect. 3.2 are unbiased at the level of < 0.01. At higher redshifts ($z_t > 1$), methods without data augmentation, i.e. the baseline network (solid black line) and the MTL training on the spectroscopic sample (dashed red line), show a bias whose absolute value increases with redshift. The right panel of Fig. 4 suggests that this is likely to be caused by a lack of training examples at such redshifts, with very few spectroscopic training examples with $z_t > 1.3$. Indeed, such a bias is significantly reduced when the training sample is extended to PAUS galaxies without spec-

---

[5]There are training mechanisms to deal with imbalanced training samples like e.g. up-weighting the contribution of imbalanced class objects in the training or oversampling synthetic data from the imbalanced 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.
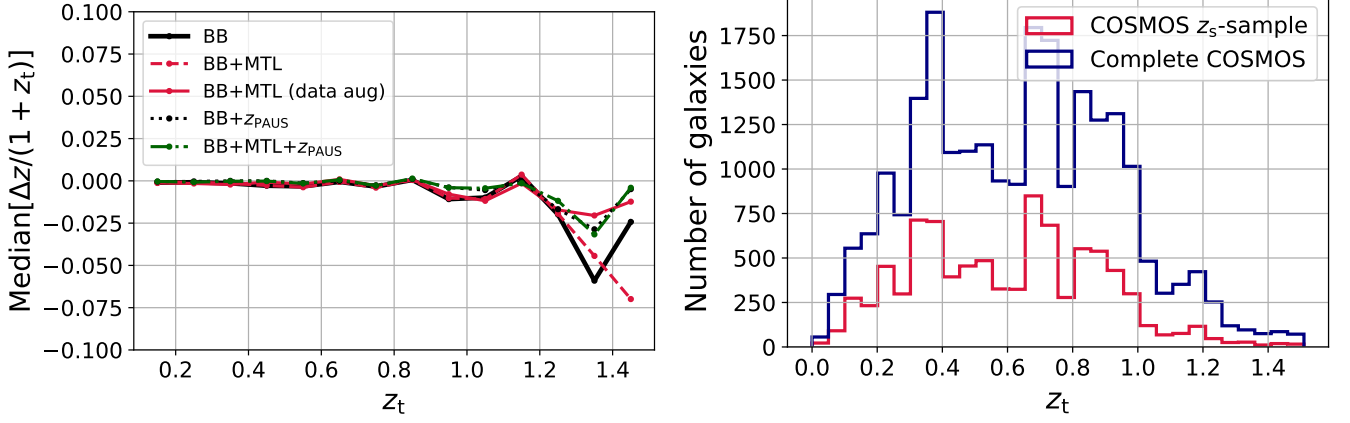
troscopy (solid red, dotted black, and dot-dashed green lines), supporting the hypothesis that the bias is caused by a lack of training examples.

The dashed red line (BB+MTL) is particularly interesting since it includes the MTL training but it only trains on the spectroscopic sample (as the BB method, black line). This indicates that the MTL is indeed working and capable of reducing the photo-z bias by training the auxiliary task of predicting the narrow-band photometry.

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

$$|z_p - z_t| / (1 + z_t) > 0.15 . \qquad (6)$$

The baseline network yields 2.6% outliers, which reduces to 2.2% with the MTL with PAUS photometry and keeps decreasing to 1.5% when we include the PAUS+COSMOS photo-zs (with or without the MTL). While for some galaxies, MTL is sufficient to correct photo-z outliers, including the PAUS+COSMOS photo-zs in the training has a stronger impact on the outliers. Indeed, including the photo-z helps particularly at the faint-end where it reduces the number of outlier galaxies with $i_{AB} > 22$ by 70%. The PAUS+COSMOS photo-zs are also very helpful for high-redshift galaxies, where they reduce the rate of outliers from 8.5% to 3%. Similar to the results on the photo-z bias, this improvement is triggered by the larger number of training galaxies at such redshifts (right panel on Fig. 4).

4.4. Redshift distributions, N(z)

Unbiased redshift distributions, $N(z)$, are crucial for a variety of science applications, with the most stringent requirements in weak lensing (e.g. Hildebrandt et al. 2012; 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. 4, the baseline network exhibits a bias at high redshifts.

Figure 5 shows the $N(z)$ in tomographic redshift bins from $0 < z_t < 1.5$ spaced by 0.2. The last tomographic bin is defined from $1.2 < z_t < 1.5$ so that the number of galaxies in the bin is increased. The ground-truth redshift defining the tomographic bins (z_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 and photo-z data augmentation always provide equal or more accurate $N(z)$ than the baseline network (BB, black line). As expected from Fig. 4, the $N(z)$ exhibiting the largest bias are those with $z_t > 1$, particularly the bin at $1 < z_t < 1.2$. In this bin, MTL together with the photo-z data augmentation (BB+MTL+z_{PAUS}, green line), reduces the bias from 0.05 to < 0.01.

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).

Fig. 4: *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.

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 Sect. 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 $10\,\sigma$ 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. 2019), 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-z 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 Sect. 2.5). The broad-bands used for this test are the CFHT u-band, the *griz* bands from DECam (Honscheid & DePoy 2008), and the *Euclid*-NISP near-infrared H, J and Y bands (Mauri et al. 2020)[6]. These are not the same bands that were used in the tests of the COSMOS field (see Sect. 2.3 and Sect. 4), but these bands were chosen to demonstrate the potential benefits for the *Euclid* photo-z estimation.

We trained the four methods presented in Sect. 3.2 on a sample with 15 000 spectroscopic galaxies, which are augmented to 30 000 with PAUS-like galaxies without spectroscopic redshifts and limited to $i_{AB} < 23$. These numbers were chosen to approximately match the number of spectroscopic and

PAUS-like galaxies in the COSMOS field (see Sect. 4). To simulate the performance of the approaches that extend the training sample with high-precision photo-zs (methods BB+z_{PAUS} and BB+MTL+z_{PAUS} in Sect. 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 shows the photo-z dispersion of 30 000 simulated test galaxies to magnitude $i_{AB} < 25$ in equally populated magnitude bins. The baseline network (black-thick line) achieves an overall precision of $\sigma_{68} = 0.104$, which increases to $\sigma_{68} = 0.166$ for galaxies with $i_{AB} > 23$. Training using photo-zs but without MTL (BB+z_{PAUS}, dotted blue line) improves the precision to $\sigma_{68} = 0.097$ and $\sigma_{68} = 0.148$ for galaxies with $i_{AB} > 23$. With BB+MTL, the overall precision is $\sigma_{68} = 0.010$, which degrades to $\sigma_{68} = 0.152$ for galaxies with $i_{AB} > 23$. Finally, combining MTL and the photo-z data augmentation (BB+MTL+z_{PAUS}, solid green line) provides the best photo-z performance with $\sigma_{68} = 0.097$ for the full sample, which increases to $\sigma_{68} = 0.146$ for galaxies with $i_{AB} > 23$.

For all training methods the relative improvement with respect to the baseline network is larger at fainter magnitudes. As an example, the BB+MTL+z_{PAUS} method (green line) provides a 4% improvement with respect to the BB network (black-thick line) for galaxies with $i_{AB} < 23$. This improvement increases to 10% for galaxies with $23 < i_{AB} < 25$ and up to a 15% for galaxies with $24 < i_{AB} < 25$. 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 Sect. 7).

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. Sect. 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 A) to explore the COSMOS photo-z performance in colour-space (Sect. 6.1). Furthermore, in Sect. 6.2 and Sect. 6.3 we study how MTL with PAUS

---

[6]With the following $5\,\sigma$ limiting magnitudes: u :25.25; g :24.65; r: 24.15; i :24.35; z: 23.95; Y: 24.0, J: 24, H: 24.
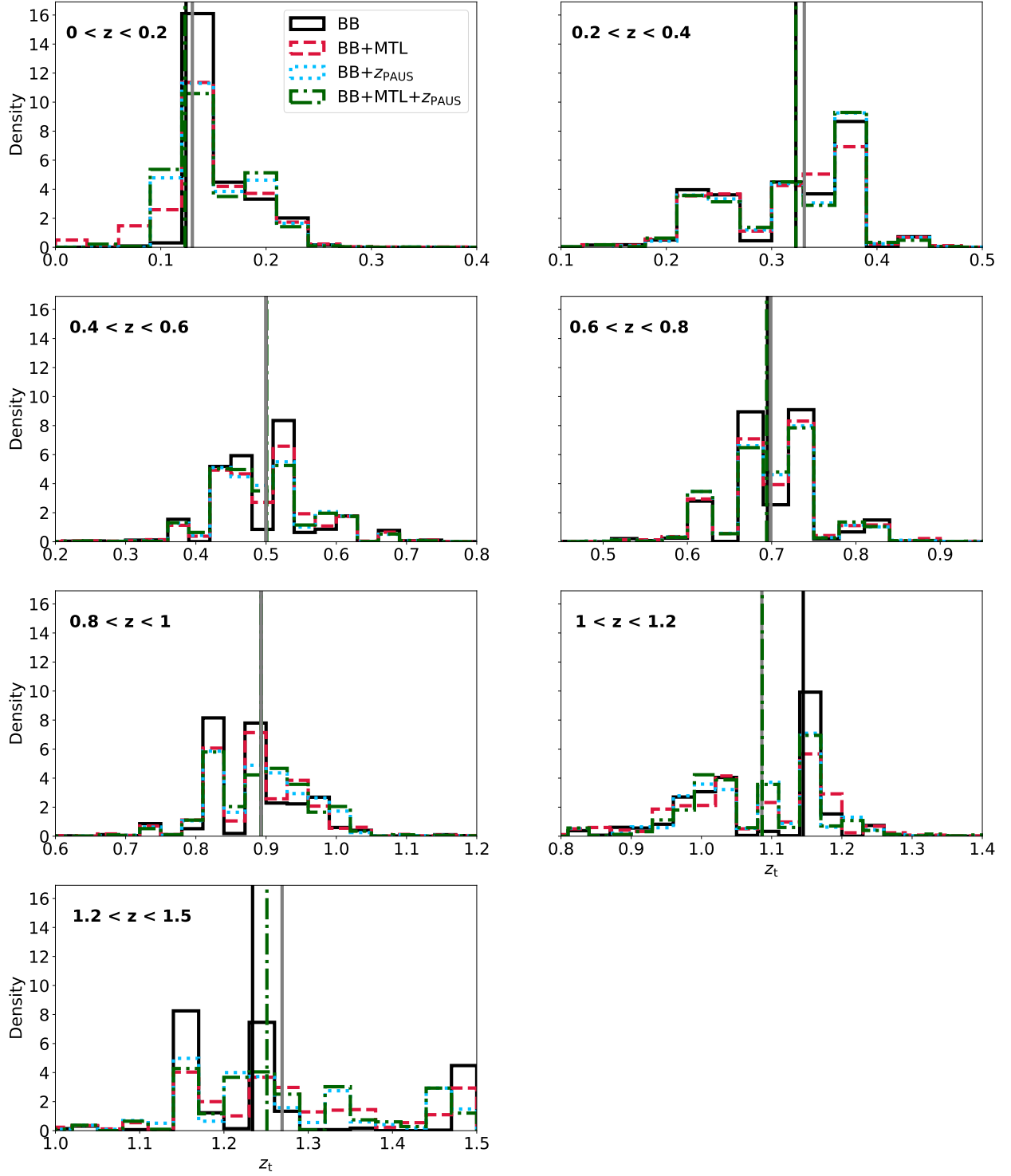
Fig. 5: The $N(z)$ estimates of the full COSMOS sample divided in 7 tomographic bins over the redshift range $0 < z < 1.5$. Tomographic bins are defined using the spectroscopic redshifts and the PAUS+COSMOS high-precision photo-zs for galaxies without spectroscopy. The vertical solid grey lines indicate the median ground-truth redshift, while the other vertical lines indicate the median redshifts of the $N(z)$ estimates.

narrow-bands breaks broad-band degeneracies caused by confusing emission lines.

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
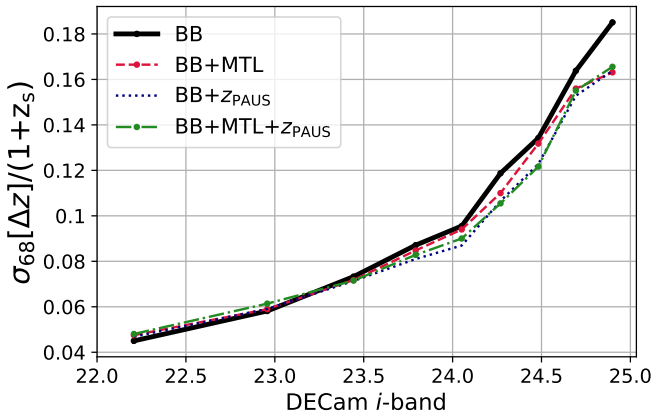
Fig. 6: Photo-z dispersion in equally sized magnitude bins for 30 000 Flagship test galaxies with magnitudes $i_{AB} < 25$ for the methods presented in Sect. 3.2. The training sample contains around 15 000 spectroscopic galaxies, extended to 30 000 with PAUS-like galaxies without spectroscopy, all of them to $i_{AB} < 23$.

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, as the SOM clusters galaxies with similar galaxy colours it also clusters galaxies with similar redshifts (Masters et al. 2015; Buchs et al. 2019). Appendix 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 2014; 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 Sect. 2.3), and subsequently assigned a SOM cell to each galaxy in the catalogue. Figure 7 shows the predicted photo-zs in colour-space, with each column corresponding to a photo-z estimation method described in Sect. 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 (BB, panel A) displays the photo-zs with the baseline network (BB method), and the second (B) and third (C) panels include MTL in the training (BB+MTL and BB+MTL+z_{PAUS} methods, respectively (bottom panel on Fig. 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 on 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_t). The network trained with only broad-bands (panel D) exhibits two regions with inaccurate 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_{PAUS} (panel F) are included in the training.

The third row in Fig. 7 shows the photo-z precision (σ_{68}, Eq. 5). Comparing panels G and D, we note that the photo-z pre-

cision worsens in the same regions where photo-zs are inaccurate, but this improves with MTL (BB+MTL, panel H) and including the PAUS+COSMOS photo-zs (BB+MTL+z_{PAUS}, panel F). Finally, the fourth row shows the dispersion of the redshift distribution, the width of the $N(z)$, within SOM cells. This quantity is also higher at the clusters pointed out in panels D and G. However, contrary to the previous panels, the BB+MTL training (panel K), or the BB+MTL+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 8 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.

6.2. 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. 7), 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. 7 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 Sect. 2.5 for details on the mock) using the $uBVriz$ broad-band photometry. The top panel in Fig. 9 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 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. 9 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.

6.3. Emission line confusions

The SOM in Fig. 8 shows a region in colour-space that contains different galaxy populations, which indicates the presence of colour-redshift degeneracies. Figure 10 shows the photo-zs of
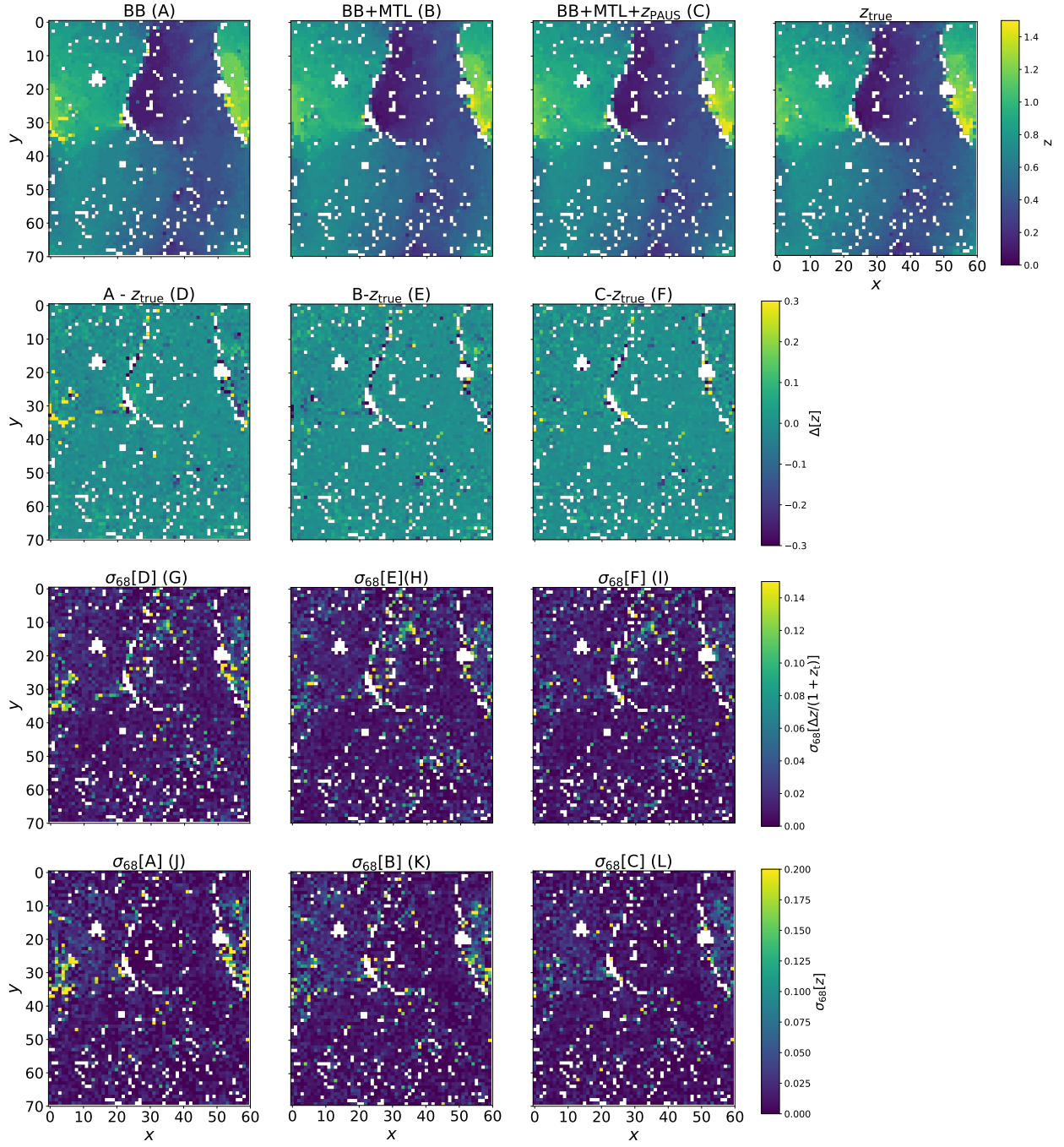
Fig. 7: SOM maps showing the photo-z performance in the COSMOS field. The first row exhibits the median predicted photo-z in colour-space for the baseline network (first panel), including MTL training (second panel), with MTL and data augmentation with PAUS+COSMOS photo-zs (third panel) and the ground-truth 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.

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_p) and the true one (z_t) with the baseline network (BB, blue), the network including MTL (BB+MTL, red), and that including MTL and photo-z data augmentation (BB+MTL+z_PAUS, orange).

The first cell (marked with stars) contains galaxies with $z_\mathrm{t} \sim 0.4$. The baseline network (blue star) predicts a lower photo-z value $z_\mathrm{p} \sim 0.3$ for one galaxy, which is fixed with MTL+photo-z data augmentation (orange star). The second cell

(marked with crosses) contains galaxies with $z_\mathrm{t} \sim 0.8$, that the baseline network estimates to be $z_\mathrm{p} \sim 1.2$. In contrast, the BB+MTL and the MTL+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 $z_\mathrm{t} \sim 1.45$. The baseline network predicts these photo-zs around $z_\mathrm{p} \sim 1.25$, and again the BB+MTL and the MTL+z_PAUS training approaches are able to improve the photo-zs. Photo-z confusions from $z_\mathrm{t} \sim 0.8$ to $z_\mathrm{t} \sim 1.2$ and from $z_\mathrm{t} \sim 1.45$ to
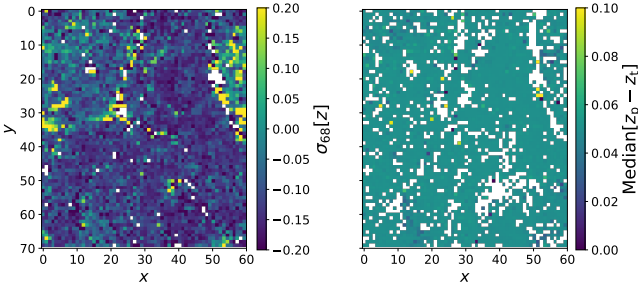
Fig. 8: *Left:* Precision of the PAUS+COSMOS photo-*z*s in colour-space in the spectroscopic sample. *Right:* Accuracy of photo-*z* within the SOM cells for the complete catalogue.
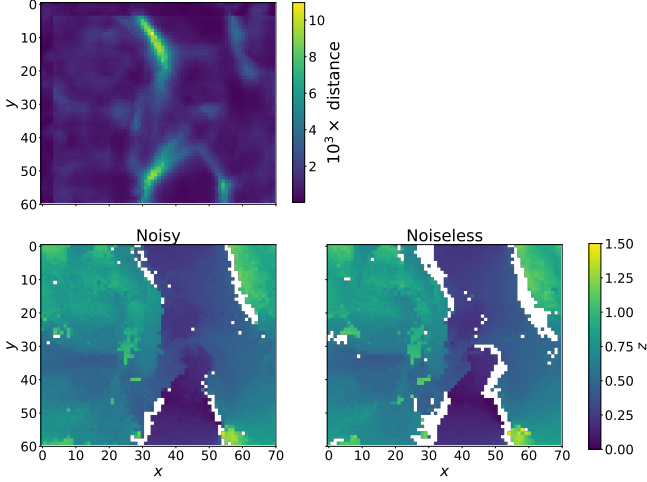

Fig. 9: SOM trained on a galaxy simulated mock with the *uBVriz* broad-bands. *Top:* Distance between every SOM cell vector and its 3×3 neighbours. *Bottom left*: Median photo-*z* in each SOM cell for noisy simulated galaxies. *Bottom right*: Median photo-*z* in each SOM cell for noiseless simulated galaxies.

$z_t \sim 1.25$ are recurrent, showing up at several SOM cells within the low photo-*z* performance cluster.

Figure 11 explores the mean Hα, Hβ, [O II], and [O III] emission line luminosity in colour-space. The emission line luminos-
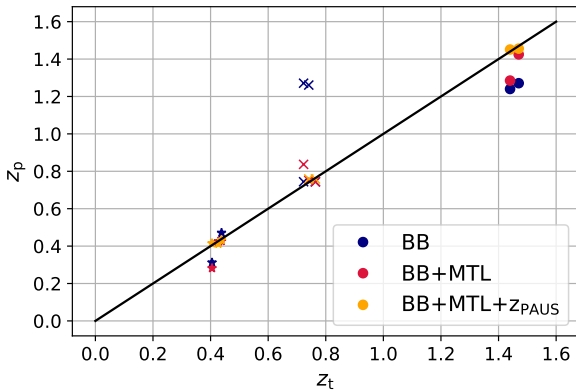
Fig. 10: Photo-*z* scatter for galaxies in three independent SOM cells. The galaxies in each cell are represented with a different marker (stars, crosses and circles).
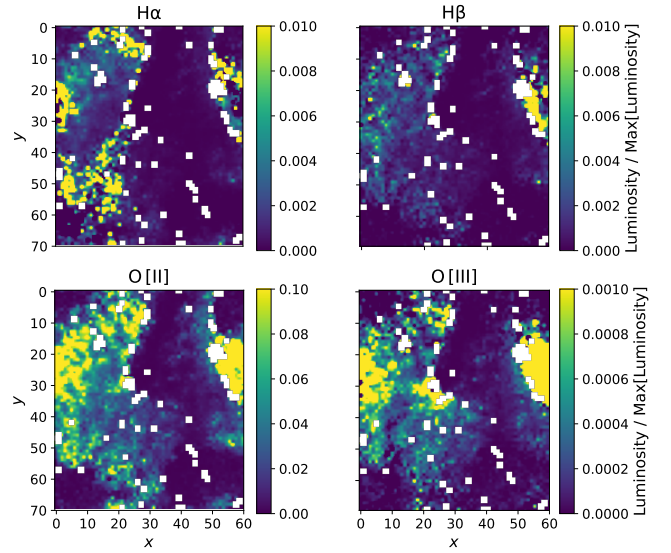
Fig. 11: Emission line luminosity in colour space for Hα, Hβ, [O II] and [O III] as indicated in the title.

ity is estimated as

$$L_{el} := 4\pi \, F_{el} \, D_L^2 \,, \tag{7}$$

where F_{el} is the emission line flux and D_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 11 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. 10 suggest that emission lines are likely to cause degeneracies in broad-band data.

Since a high ratio of [O III] 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, e.g. (Whitaker et al. 2012), galaxies showing a photo-*z* mismatch from $z_t \sim 0.8$ to $z_p \sim 1.2$ 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.

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 Sect. 7.1, we
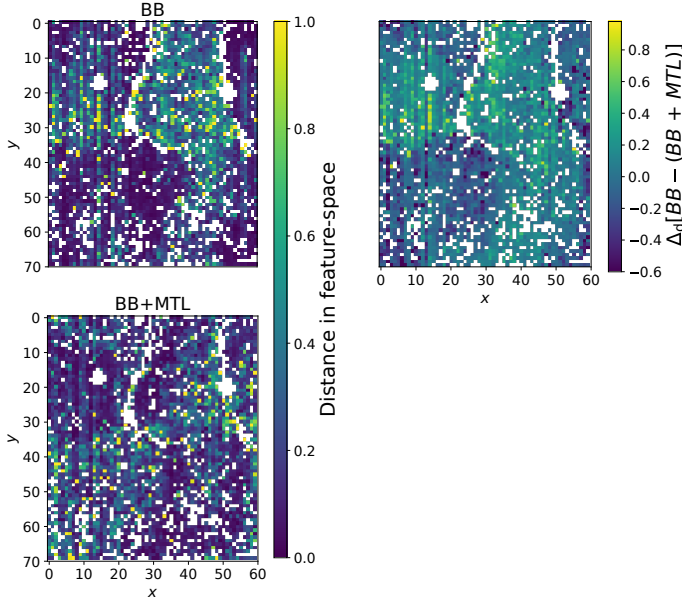
Fig. 12: Distance in feature space for COSMOS galaxies assigned to the same SOM cell. Feature space distances are normalised by the maximum of the distance evaluated in the plot. *Top left*: Distances when the network is trained only using the broad-bands, *bottom left*: Distances when the network training includes MTL with PAUS narrow-band photometry and *top right*: Difference between the two left panels.
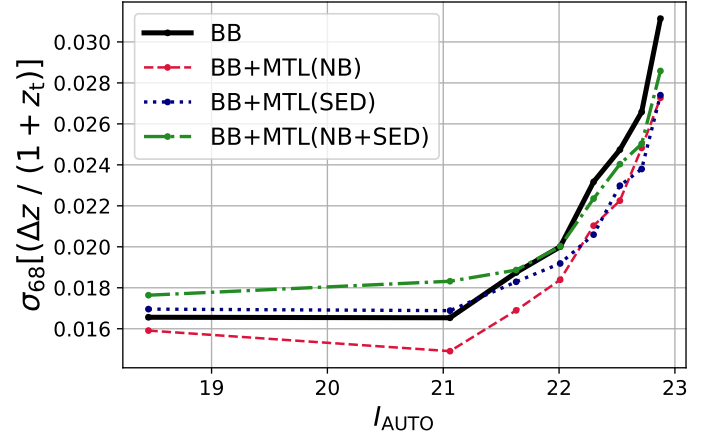

Fig. 13: Photo-z precision in the COSMOS field including predicting the galaxy SED as auxiliary task. The galaxy SED prediction is addressed as a classification, where the true SED is a class between one and 47.

748 use a variation of our fiducial network to encode the galaxy
749 photometry in a 2-dimensional space similar to a SOM, while
750 in Sect. 7.2 we study the impact of using other auxiliary tasks
751 (other than predicting the narrow-band photometry) in the MTL
752 network.

7.1. Underlying data representation in colour-space with MTL

755 For this test, we will modify the fiducial network architec-
756 ture (see Sect. 3.2 and Fig. 1). The modified architecture has an
757 encoder-decoder-like architecture where the minimum number
758 of nodes is found in an intermediate layer. Particularly, our net-
759 work encodes the galaxy information in two numbers that are
760 used to predict the photo-z (2D feature space). This is the only
761 change we introduced since our fiducial network also encodes
762 the photometry in a feature space with more dimensions. Encod-
763 ing the galaxy information in a 2D feature space simplifies its
764 visualisation and brings it closer to the SOM colour-space map,
765 which we already studied.

766 The galaxy representations in the 2D feature space must en-
767 code all the information needed to make the photo-z prediction.
768 Furthermore, as in the MTL network, those two numbers are also
769 used to reconstruct the narrow-band photometry. Hence, these
770 must also encode the relevant information for this task. There-
771 fore, comparing the feature space representation of the baseline
772 network (decoding only to the photo-z) and the MTL network
773 (also predicting the narrow-band photometry) helps us to better
774 understand why the MTL improves the photo-z estimates.

775 Figure 12 shows the mean distance in feature space among
776 galaxies within the same SOM cell. The top left plot corresponds
777 to the broad-band baseline network, while the bottom left panel
778 includes the PAUS narrow-band MTL. The first thing we note

779 is that galaxies assigned to the same SOM cell also cluster in
780 feature space, which indicates that the feature space encodes the
781 input photometry in a similar way as the SOM.

782 As the 2D feature space is not constrained during the train-
783 ing, the coordinates assigned to each galaxy do not contain any
784 valuable information by themselves (that is, the network is al-
785 lowed to encode the same galaxy differently in several indepen-
786 dent trainings). Therefore, distances from different feature space
787 maps cannot be directly compared (e.g. the feature map of the
788 BB network and that of the BB+MTL), but given a trained net-
789 work the relative distances between galaxies in feature space
790 contains information about the similarity of these.

791 To compare the top and bottom left panels in Fig. 12, the
792 results have been normalised to display distances between zero
793 and unity. The top right panel in the figure shows the difference
794 between the two left panels. Note that in the region with degener-
795 ate broad-band photometry with coordinates around $(55, 30)$, the
796 feature space distance assigned by the MTL network is larger
797 than in the broad-band only case. This suggests that MTL with
798 PAUS narrow-bands places degenerate galaxies in more distant
799 feature space regions, which effectively means that it is capable
800 of capturing some differences that the BB method does not.

7.2. MTL with other galaxy parameters

801 So far in this paper, we explored how photo-z predictions bene-
fit from MTL predicting PAUS narrow-band fluxes as an auxil-
iary 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_i)

$$\phi(z, x_1, ..., x_N | \theta), \tag{8}$$

802 where x_i could be any galaxy parameter that correlates with the
803 galaxy photo-z such as the galaxy type.

804 Template-fitting photo-z methods predict the joint probabil-
805 ity distribution $p(z, t|f)$ of the redshift (z) and the galaxy type (t)
806 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 13 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 one and 47 extracted from Ilbert et al. (2009). 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 Sect. 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 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.

8. Discussion and conclusions

Photometric redshifts (photo-z) are crucial to exploit ongoing and future large galaxy broad-band imaging surveys. While covering a large area, the broad-band spectral resolution limits the redshift performance through colour-redshift degeneracies. The PAU Survey (PAUS) 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 Sect. 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 to $i_{AB} < 23$, our method reduces the photo-z scatter by approximately 16% (see Sect. 4.2) and the number of photo-z outliers by $\sim 40\%$ (see Sect. 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 $i_{AB} < 23$ and evaluated it on a sample with $i_{AB} < 25$. The MTL predicts up to 15% more precise photo-zs for galaxies with $24 < i_{AB} < 25$ than the baseline network (see Sect. 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. 8). 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 paper 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.

Data availability

The PAUS raw data is publicly available through the ING group. A few reduced images are publicly available at https://www.pausurvey.org. The Flagship catalogue is a property of the *Euclid* Consortium.

References

Alarcon, A., Gaztanaga, E., Eriksen, M., et al. 2021, MNRAS, 501, 6103
Amiaux, J., Scaramella, R., Mellier, Y., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8442, Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, ed. M. C. Clampin, G. G. Fazio, H. A. MacEwen, & J. Oschmann, Jacobus M., 84420Z
Arnouts, S. & Ilbert, O. 2011, Astrophysics Source Code Library, ascl:1108.009
Benítez, N. 2011, Astrophysics Source Code Library, ascl:1108.011
Benitez, N., Dupke, R., Moles, M., et al. 2014, arXiv e-prints [arXiv:1403.5237]
Blake, C. & Bridle, S. 2005, MNRAS, 363, 1329
Bonnett, C. 2015, MNRAS, 449, 1043
Buchs, R., Davis, C., Gruen, D., et al. 2019, MNRAS, 489, 820
Cabayol, L., Eriksen, M., Amara, A., et al. 2021, MNRAS, 506, 4048
Cabayol-Garcia, L., Eriksen, M., Alarcón, A., et al. 2020, MNRAS, 491, 5392
Calabrò, A., Daddi, E., Cassata, P., et al. 2018, ApJ, 862, L22
Carrasco Kind, M. & Brunner, R. J. 2013, MNRAS, 432, 1483
Carrasco Kind, M. & Brunner, R. J. 2014, MNRAS, 438, 3409
Caruana, R. 1997, Machine Learning, 28, 41–75
Casas, R., Cardiel-Sas, L., Castander, F. J., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 99084K
Castander, F. J., Ballester, O., Bauer, A., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, 84466D
Chong, De Wei, K. & Yang, A. 2019, in European Physical Journal Web of Conferences, Vol. 206, European Physical Journal Web of Conferences, 09006
Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039
Collister, A. A. & Lahav, O. 2004, PASP, 116, 345
Coupon, J., Ilbert, O., Kilbinger, M., et al. 2009, A&A, 500, 981
Cropper, M., Cole, R., James, A., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8442, Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, ed. M. C. Clampin, G. G. Fazio, H. A. MacEwen, & J. Oschmann, Jacobus M., 84420V
Csörnyei, G., Dobos, L., & Csabai, I. 2021, MNRAS, 502, 5762
de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Experimental Astronomy, 35, 25
Eriksen, M., Alarcon, A., Cabayol, L., et al. 2020, MNRAS, 497, 4565
Eriksen, M., Alarcon, A., Gaztanaga, E., et al. 2019, MNRAS, 484, 4200
Feldmann, R., Carollo, C. M., Porciani, C., et al. 2006, MNRAS, 372, 565
Gatti, M., Vielzeuf, P., Davis, C., et al. 2018, MNRAS, 477, 1664
Gomes, Z., Jarvis, M. J., Almosallam, I. A., & Roberts, S. J. 2018, MNRAS, 475, 331
Good, I. J. 1952, Journal of the Royal Statistical Society. Series B (Methodological), 14, 107
Hasinger, G., Capak, P., Salvato, M., et al. 2018, ApJ, 858, 77
Hildebrandt, H., Erben, T., Kuijken, K., et al. 2012, MNRAS, 421, 2355
Hildebrandt, H., Pielorz, J., Erben, T., et al. 2009, A&A, 498, 725
Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, A&A, 647, A124
Honscheid, K. & DePoy, D. L. 2008, International conference on high energy physics (ICHEP08) [arXiv:0810.3600]
Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Johnston, H., Joachimi, B., Norberg, P., et al. 2021a, A&A, 646, A147
Johnston, H., Wright, A. H., Joachimi, B., et al. 2021b, A&A, 648, A98
Jones, E. & Singal, J. 2017, A&A, 600, A113
Kashino, D., Silverman, J. D., Sanders, D., et al. 2019, ApJS, 241, 10
Kennicutt, Robert C., J. 1998, ARA&A, 36, 189
Kingma, D. P. & Ba, J. 2015, 3rd International Conference for Learning Representations, San Diego [arXiv:1412.6980]
Knox, L., Song, Y.-S., & Zhan, H. 2006, ApJ, 652, 857
Kohonen, T. 1982, Biological Cybernetics, 43, 59
Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, ApJS, 218, 15
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Lamareille, F. 2010, A&A, 509, A53
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints [arXiv:1110.3193]
Lee, K.-G., Krolewski, A., White, M., et al. 2018, ApJS, 237, 31
Liebel, L. & Körner, M. 2018, arXiv e-prints [arXiv:1805.06334]
Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, ApJS, 172, 70
LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints [arXiv:0912.0201]
Martí, P., Miquel, R., Castander, F. J., et al. 2014, MNRAS, 442, 92
Masters, D. et al. 2015, Astrophys. J., 813, 53
Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2017, ApJ, 841, 111
Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, ApJ, 877, 81
Mauri, N. et al. 2020, J. Phys. Conf. Ser., 1342, 012122
Ménard, B., Scranton, R., Schmidt, S., et al. 2013, arXiv e-prints [arXiv:1303.4722]
Moeskops, P., Wolterink, J. M., van der Velden, B. H. M., et al. 2017, arXiv e-prints [arXiv:1704.03379]
Moles, M., Benítez, N., Aguerri, J. A. L., et al. 2008, AJ, 136, 1325
Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015, Astroparticle Physics, 63, 81
Padilla, C., Ballester, O., Cardiel-Sas, L., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 99080Z
Padilla, C., Castander, F. J., Alarcón, A., et al. 2019, AJ, 157, 246
Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., & Fouchez, D. 2019, A&A, 621, A26
Pasquet-Itam, J. & Pasquet, J. 2018, A&A, 611, A97
Paszke, A., Gross, S., Chintala, S., et al. 2017, NIPS 2017 Workshop Autodiff Submission
Paulino-Afonso, A., Sobral, D., Darvish, B., et al. 2018, A&A, 620, A186
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, ApJ, 739, L40
Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, PASP, 128, 104502
Salvato, M., Ilbert, O., & Hoyle, B. 2019, Nature Astronomy, 3, 212
Schmidt, S. J., Malz, A. I., Soo, J. Y. H., et al. 2020, MNRAS, 499, 1587
Schmidt, S. J., Ménard, B., Scranton, R., Morrison, C., & McBride, C. K. 2013, MNRAS, 431, 3307
Scoville, N., Abraham, R. G., Aussel, H., et al. 2007, ApJS, 172, 38
Song, X., Zhao, X., Fang, L., & Hu, H. 2020, International Journal of Computer Vision, 128
Soo, J. Y. H., Joachimi, B., Eriksen, M., et al. 2021, MNRAS, 503, 4118
Soo, J. Y. H., Moraes, B., Joachimi, B., et al. 2018, MNRAS, 475, 3613
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, Journal of Machine Learning Research, 15, 1929
Sánchez, C., Carrasco Kind, M., Lin, H., et al. 2014, MNRAS, 445, 1482
The Dark Energy Survey Collaboration. 2005, arXiv e-prints [astro-ph/0510346]
Tonello, N., Tallada, P., Serrano, S., et al. 2019, Astronomy and Computing, 27, 171
Tortorelli, L., Siudek, M., Moser, B., et al. 2021, J. Cosmology Astropart. Phys., 2021, 013
Urrutia, T., Wisotzki, L., Kerutt, J., et al. 2019, A&A, 624, A141
van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, A&A, 642, A200
Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, ApJ, 754, L29
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., et al. 2020b, A&A, 640, L14
Yanminsun, Wong, A., & Kamel, M. S. 2011, International Journal of Pattern Recognition and Artificial Intelligence, 23
Zhang, Y. & Yang, Q. 2021, IEEE Transactions on Knowledge and Data Engineering, 1

Affiliations

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

[2] Port d'Informació Científica, Carrer de Can Magrans, s/n , 08193 Bellaterra (Barcelona), Spain

[3] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain

[4] Institut d'Estudis Espacials de Catalunya (IEEC), 08193 Barcelona, Spain

[5] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

[6] Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, 28049 Madrid, Spain

[7] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB),
German Centre for Cosmological Lensing, 44780 Bochum, Germany

[8] Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain

[9] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

[10] Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA, Leiden, the Netherlands

[11] National Centre for Nuclear Research, ul. Pasteura 7, 02-093, Warsaw, Poland

Appendix 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 ($\mathbf{w}^k$), 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 ($\mathbf{w}^k$) 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 $\mathbf{x}$ (e.g., in our case the galaxy colours), is compared to all the SOM cells' vectors via a χ^2 expression,

$$\chi^2\left(\mathbf{w}^k(t), \mathbf{x}\right) = \sum_i \left[\frac{x_i - w_i^k(t)}{\sigma_i}\right]^2 , \qquad (A.1)$$

where i sums over galaxy attributes and σ_i is the uncertainty associated to x_i. The evaluated galaxy is assigned to the cell with the lowest χ^2, which updates its associated vector of attributes $\mathbf{w}^k(t)$ 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 $\sigma_{\text{kernel}}^2(t)$. Therefore, the vector of attributes for a particular cell k after iteration $t + 1$ is

$$\mathbf{w}^k(t + 1) = \mathbf{w}^k(t) + \alpha(t) \, H(t, |\mathbf{w} - \mathbf{x}|) \left(\mathbf{x} - \mathbf{w}^k(t)\right) , \qquad (A.2)$$

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.

Appendix 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 BB+z$_{\text{PAUS}}$ and BB+MTL+z$_{\text{PAUS}}$ in Sect. 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 (Sect. B.1) and the presence of outliers (Sect. B.2) have on the photo-z performance.

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

Figure B.1 shows the photo-z precision of a set of 1000 spectroscopic galaxies for four independent BB networks (simply
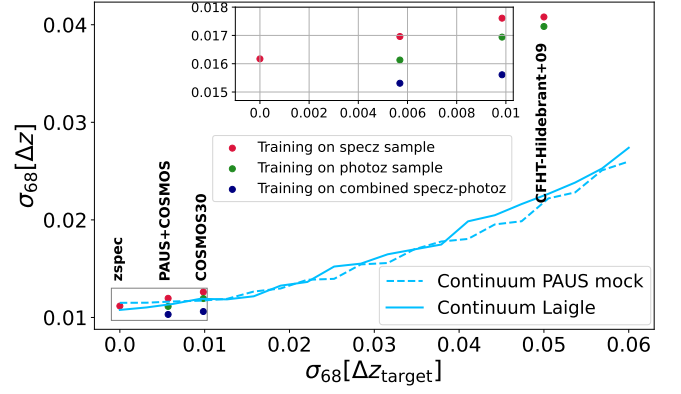
Fig. B.1: Photo-z performance as a function of the ground-truth redshift precision used for training the networks. The training redshifts are the spectroscopic redshifts, the PAUS+COSMOS photo-zs, COSMOS30, and a set of CFHT photo-zs in COSMOS. Red points correspond to training on the spectroscopic sample (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 Laigle 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)

mapping colours to redshift), each of them trained with different ground-truth redshifts. The redshifts used for training are the spectroscopic redshifts (see Sect. 2.4), the PAUS+COSMOS photo-zs (see Sect. 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. (2012) 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. B.1 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. B.1), 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, or 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. In-

deed, 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. B.1 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 Sect. 5. In both cases, true 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 B.2 for more details).

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

Figure B.1 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 B.2 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 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_{mod}) so that [O III] is confused with Hα in the training redshifts, i.e.

$$z_{\mathrm{mod}} = \lambda_{[\mathrm{O\,III}]}/\lambda_{\mathrm{H}\alpha} \, (1 + z_{\mathrm{t}}) - 1 \,, \qquad (\mathrm{B.1})$$

where z_{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 vs. 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. B.2 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 B.3 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. B.2. 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. B.3 shows the loss function for the correct and the wrongly labelled training galaxies separately when the incorrect redshift labels are generated with Eq. (B.1). 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 $p(z)$, which present a double-peaked distribution (not shown). Hence, Figs. B.2 and B.3 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.
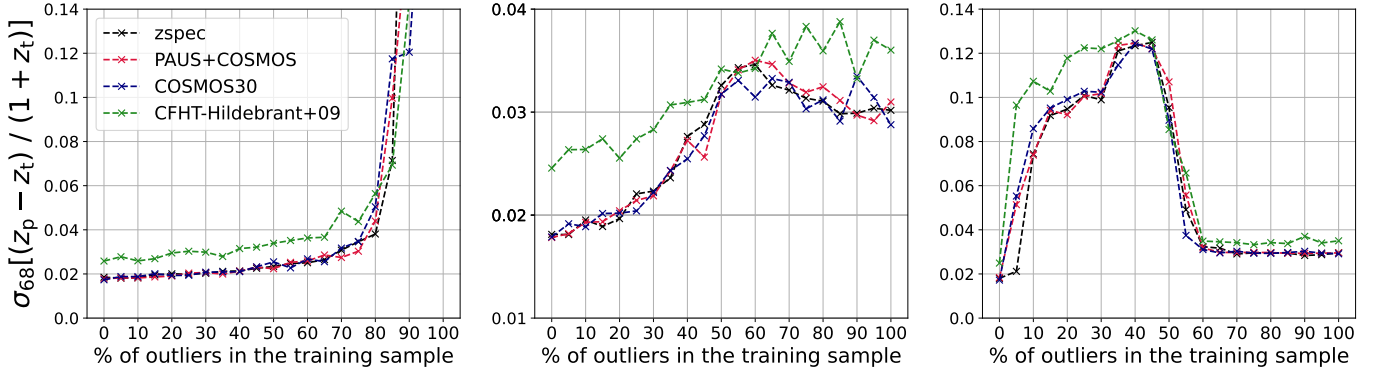
Fig. B.2: Effect of outliers and systematic errors in the ground-truth redshift sample used during training check that I captured the intended meaning of the statement. The training sample consists of 5000 spectroscopic galaxies with photometry from COSMOS2015. Each coloured line uses a different sample 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; *right:* redshifts modified with Eq. (B.1).
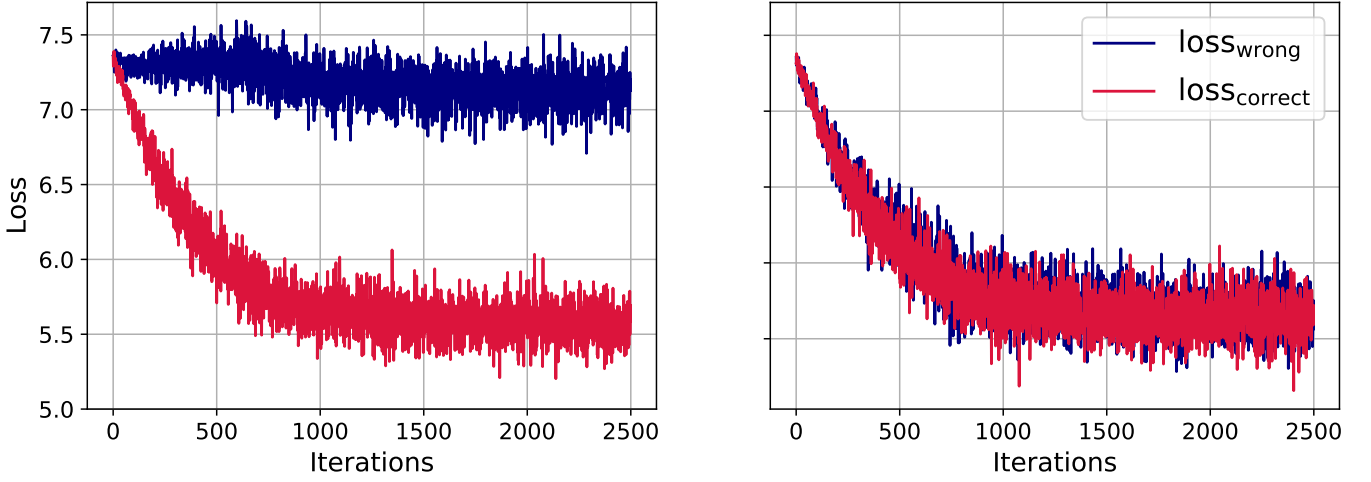

Fig. B.3: Training loss function for galaxies with wrong (blue) and corrected (red) target redshift. The training sample consists of 5000 spectroscopic galaxies with photometry from COSMOS2015. In the left panel, the modified target redshifts are randomly switched to a value drawn from $U(0, 1.5)$, while in the right panel the wrong redshift labels are generated with Eq. (B.1).