

The PAU Survey and *Euclid*: Improving broadband photometric redshifts with multi-task learning[★]

L. Cabayol^{1,2}, M. Eriksen^{1,2}, J. Carretero^{1,2}, R. Casas^{3,4}, F. J. Castander^{4,3}, E. Fernández¹, J. Garcia-Bellido⁵, E. Gaztanaga^{3,4}, H. Hildebrandt⁶, H. Hoekstra⁷, B. Joachimi⁸, R. Miquel^{1,9}, C. Padilla¹, A. Pocino^{3,4}, E. Sanchez¹⁰, S. Serrano^{3,11}, I. Sevilla¹⁰, M. Siudek^{1,3}, P. Tallada-Crespi^{10,2}, N. Aghanim¹², A. Amara¹³, N. Auricchio¹⁴, M. Baldi^{15,14,16}, R. Bender^{17,18}, D. Bonino¹⁹, E. Branchini^{20,21}, M. Brescia²², J. Brinchmann²³, S. Camera^{24,25,19}, V. Capobianco¹⁹, C. Carbone²⁶, M. Castellano²⁷, S. Cavuoti^{22,28,29}, A. Cimatti^{30,31}, R. Cledassou^{32,33}, G. Congedo³⁴, C. J. Conselice³⁵, L. Conversi^{36,37}, Y. Copin³⁸, L. Corcione¹⁹, F. Courbin³⁹, M. Cropper⁴⁰, A. Da Silva^{41,42}, H. Degaudenzi⁴³, M. Douspis¹², F. Dubath⁴³, C. A. J. Duncan^{35,44}, X. Dupac³⁶, S. Dusini⁴⁵, S. Farrens⁴⁶, P. Fosalba^{3,4}, M. Frailis⁴⁷, E. Franceschi¹⁴, P. Franzetti²⁶, B. Garilli²⁶, W. Gillard⁴⁸, B. Gillis³⁴, C. Giocoli^{14,49}, A. Grazian⁵⁰, F. Grupp^{17,18}, S. V. H. Haugan⁵¹, W. Holmes⁵², F. Hormuth⁵³, A. Hornstrup⁵⁴, P. Hudelot⁵⁵, K. Jahnke⁵⁶, M. Kümmel¹⁸, S. Kermiche⁴⁸, A. Kiessling⁵², M. Kilbinger⁴⁶, R. Kohley³⁶, H. Kurki-Suonio⁵⁷, S. Ligori¹⁹, P. B. Lilje⁵¹, I. Lloro⁵⁸, E. Maiorano¹⁴, O. Mansutti⁴⁷, O. Marggraf⁵⁹, K. Markovic⁵², F. Marulli^{60,14,16}, R. Massey⁶¹, S. Mei⁶², M. Meneghetti^{63,14}, E. Merlin²⁷, G. Meylan⁶⁴, M. Moresco^{60,14}, L. Moscardini^{60,14,16}, E. Munari⁴⁷, R. Nakajima⁵⁹, S. M. Niemi⁶⁵, S. Paltani⁴³, F. Pasian⁴⁷, K. Pedersen⁶⁶, V. Pettorino⁴⁶, G. Polenta⁶⁷, M. Poncet³², L. Popa⁶⁸, L. Pozzetti¹⁴, F. Raison¹⁷, R. Rebolo^{69,70}, J. Rhodes⁵², G. Riccio²², C. Rosset⁶², E. Rossetti⁶⁰, R. Saglia^{17,18}, B. Sartoris^{18,47}, P. Schneider⁵⁹, A. Secroun⁴⁸, G. Seidel⁵⁶, C. Sirignano^{71,45}, G. Sirri¹⁶, L. Stanco⁴⁵, A. N. Taylor³⁴, I. Tereno^{41,72}, R. Toledo-Moreo⁷³, F. Torradeflot^{2,10}, I. Tutusaus⁷⁴, E. Valentijn⁷⁵, L. Valenziano^{14,16}, Y. Wang⁷⁶, J. Weller^{17,18}, G. Zamorani¹⁴, J. Zoubian⁴⁸, S. Andreon⁷⁷, V. Scottez^{55,78}, and A. Tramacere⁴³

(Affiliations can be found after the references)

Received 21 September 2022 / Accepted 13 January 2023

ABSTRACT

Current and future imaging surveys require photometric redshifts (photo-zs) to be estimated for millions of galaxies. Improving the photo-z quality is a major challenge but is needed 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 broadband photometric redshifts. We used a multi-task learning (MTL) network to improve broadband photo-z estimates by simultaneously predicting the broadband photo-z and the narrow-band photometry from the broadband photometry. The narrow-band photometry is only required in the training field, which also enables better photo-z predictions for the galaxies without narrow-band photometry in the wide field. This technique was tested with data from the Physics of the Accelerating Universe Survey (PAUS) in the COSMOS field. We find that the method predicts photo-zs that are 13% more precise down to magnitude $i_{AB} < 23$; the outlier rate is also 40% lower when compared to the baseline network. Furthermore, MTL 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 such as *Euclid* or LSST. For simulated data, training on a sample with $i_{AB} < 23$, the method reduces the photo-z scatter by 16% for all galaxies with $i_{AB} < 25$. We also studied the effects of extending the training sample with photometric galaxies using PAUS high-precision photo-zs, which reduces the photo-z scatter by 20% in the COSMOS field.

Key words. surveys – methods: data analysis – techniques: image processing – techniques: photometric – methods: observational

1. Introduction

Over the last few decades, multi-band wide imaging surveys have been driving discoveries, demonstrating the power of large datasets to enable precision cosmology. Obtaining precise photometric redshifts is crucial for exploiting large galaxy imaging surveys (Salvato et al. 2019), and they are a limiting factor in the accuracy of cosmology measurements that use galaxies (Knox et al. 2006). Current and upcoming imaging surveys such

as 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 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

[★] This paper is published on behalf of the Euclid Consortium.

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 fractional 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 photo- z s that are around three times more precise than in 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 methodologies. 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 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 for improving the photo- z performance, such as including galaxy morphology (Soo et al. 2018), using Gaussian processes (Gomes et al. 2018; Soo et al. 2021), implementing ‘pseudo-labelling’ semi-supervised approaches to determine the underlying structure of the data (Humphrey et al., in prep.), and directly predicting the photo- z from astronomical images (Pasquet-Itam & Pasquet 2018; Pasquet et al. 2019; Chong & Yang 2019).

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

In astronomy, data that could be helpful for improving the photo- z performance often exist, for example photometry in several bands. However, such data are not always available for the complete wide field, preventing us from using it. With MTL, we can utilise these data to improve the photo- z predictions without explicitly providing them as input. Particularly, we implemented an MTL neural network that predicts the photo- z and the narrow-band photometry of a galaxy from its broadband 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, and we can evaluate the photo- z of any galaxy with only its broadband photometry. In this way, the data available in certain fields can be exploited to improve the photo- z estimations in other fields.

We tested the method with data from the Physics of the Accelerating Universe Survey (PAUS). It is a narrow-band imaging survey carried out with the PAUCam instrument (Castander et al. 2012; Padilla et al. 2016, 2019), a camera with 40 narrow bands that cover the optical spectrum (Casas et al. 2016). The method could also be applied to other narrow-band surveys such as the Javalambre Physics of the Accelerating Universe Survey (J-PAS; Benitez et al. 2014).

The paper is structured as follows. In Sect. 2 we present the data used throughout the paper. Section 3 introduces MTL 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 (SOMs) to explore the photo- z distribution of COSMOS galaxies in colour space (Sect. 6) and to gain 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 broadband 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 a 130 Å full width at half maximum and a 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). In this work we use the full pass-band filter information¹.

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 three-dimensional galaxy clustering (Johnston et al. 2021a), as well as determining galaxy properties (Tortorelli et al. 2021) and measuring the D4000 Å spectral break (Renard et al., in prep.).

¹ Similar filter functions to the ones used in the paper are available at the PAUS website www.pausurvey.org

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) fields² W1, W2, and W3, as well as the full COSMOS field³. In the W2 field, so far PAUS has observed in the overlapping region with the GAMA 9-hour field⁴ (G09). Currently, PAUS data have a 40 narrow-band coverage of 10 deg² in each of W1 and G09, 20 deg² in W3, and 2 deg² 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 specifically 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).

Two methods for extracting the galaxy photometry have been developed for PAUS: 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 Laigle et al. (2016) and the photometry calibration is relative to the Sloan Digital Sky Survey (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 also use the high-precision photometric redshifts from Alarcon et al. (2021, PAUS+COSMOS hereafter). They were estimated with a template-fitting method modelling the spectral energy distributions (SEDs) as a linear combination of emission line and continuum templates to then compute the Bayes evidence by integrating over the linear combinations. In addition to the PAUS narrow bands, the PAUS+COSMOS catalogue uses 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 photo- z s 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- z s are more precise and less biased than those from Laigle et al. (2016), which use a combination of 30 broad-, intermediate-, and narrow-band filters.

2.3. Broadband data

The broadband data used in this paper are from Laigle et al. (2016, COSMOS2015 hereafter), which includes the u -band from the Canada-France-Hawaii Telescope (CFHT)/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 redshifts. The redshift cut is prompted by the photo- z distribution in the PAUS+COSMOS catalogue, with very few galaxies with $z > 1.5$ (Fig. 1).

² http://www.cfht.hawaii.edu/Science/CFHTLS_Y_WIRCAM/cfhtlsdeepwidefields.html

³ <http://cosmos.astro.caltech.edu/>

⁴ https://www.astro.ljmu.ac.uk/~ikb/research/gama_fields/

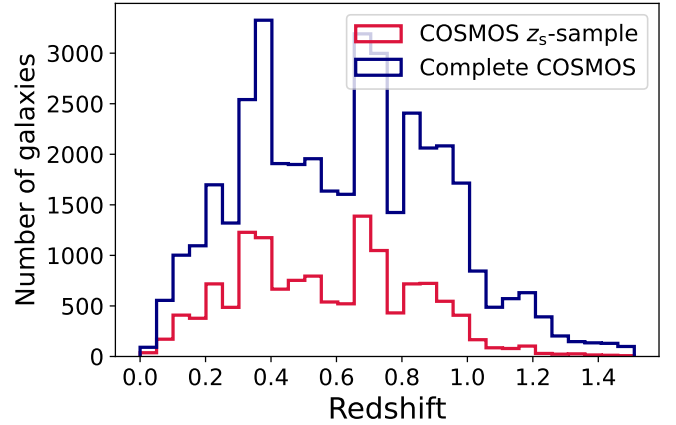


Fig. 1. Redshift distributions for the COSMOS spectroscopic sample (red line) and the full (spectroscopic and photo- z) COSMOS sample.

2.4. Spectroscopic galaxy sample

To train the neural network, one needs a galaxy catalogue with known redshifts. We used the zCOSMOS Data Release (DR) 3 bright spectroscopic data (Lilly et al. 2007), which cover 1.7 deg² 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 only keep redshifts with a confidence class (*conf*) of $3 < conf < 5$, which leads to a catalogue with ~ 9400 galaxies. We extended the spectroscopic sample with a compilation of 2693 redshifts from Alarcon et al. (2021). This compilation includes redshifts from C3R2 DR1 and 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.

Figure 1 shows the redshift distribution of the COSMOS spectroscopic sample (red) and the full PAUS sample in the COSMOS field (blue), where the redshift is defined as the spectroscopic redshift if this is available and as the PAUS+COSMOS photo- z otherwise. Including the PAUS+COSMOS photo- z is particularly relevant for galaxies with $z > 1$, where there are very few spectroscopic measurements.

2.5. Galaxy mocks

In Sect. 5 we also use the Flagship galaxy simulations described in Castander et al. (in prep.). The Flagship galaxy catalogue has been developed to study the performance of the *Euclid* mission. The mock catalogue populates the halos detected in the *Euclid* Flagship N-body simulation (Potter et al. 2017), which is a large two trillion particles simulation on a box of $3780 h^{-1} \text{Mpc}$, and a mass resolution of $m_p = 2.4 \times 10^9 h^{-1} M_\odot$. The N-body simulation uses a cosmological model with parameters similar to the *Planck* 2015 cosmology (Planck Collaboration XIII 2016). Halos are identified with the ROCKSTAR halo finder (Behroozi et al. 2013). Galaxies are assigned to the halos using a hybrid halo occupation distribution and abundance matching technique similar to the one used for the MICE catalogues described in Carretero et al. (2014). Galaxies are divided into central and satellites. Each halo contains a central and a number of satellites given by their halo occupation. Galaxies are also tagged in three colour types: blue, green, and red. The relative abundance

of central and satellites as a function of colour type and absolute magnitude is constrained by the observed colour-magnitude distribution and the clustering as a function of colour at low redshift. At higher redshift only observed colour distributions are used. Each galaxy is assigned a SED, including its extinction, from the COSMOS SED library (e.g. Ilbert et al. 2013), which includes SED templates from Polletta et al. (2007) and additional blue templates from Bruzual & Charlot (2003). In order to have a more continuous distribution of galaxy magnitudes and colours, the SED assigned to each galaxy is a linear combination of two consecutive templates in the COSMOS template library. Emission lines are then added to the SED of each galaxy. The $H\alpha$ flux is computed from the rest-frame ultra-violet flux of each galaxy template following Kennicutt (1998). The $H\beta$ fluxes are then re-adjusted to make them follow the Pozzetti et al. (2016) model 1 and model 3 distributions. The $H\beta$ flux is computed from the $H\alpha$ flux assuming case B recombination (Osterbrock & Ferland 2006). The other emission line fluxes ([OII], [OIII], [NII], and [SII]) are computed following relations obtained from observed distributions. The emission line fluxes are added to the continuum assuming a Gaussian distribution of width given by the galaxy magnitude and the Faber-Jackson or Tully-Fisher relation. Finally, the SEDs containing the emission lines are convolved with the filter transmission curves to produce the expected observed fluxes. This prescription is followed to generate both broad- and narrow-band photometry. The Flagship catalogue is a property of the Euclid Consortium and is available at CosmoHub⁵ (Carretero et al. 2017; Tallada et al. 2020), a web application based on Hadoop to interactively distribute and explore massive cosmological datasets.

3. Multi-task neural network to improve broadband photo-zs

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

3.1. Multi-task learning

Deep-learning algorithms consist of training a single or an ensemble of models to accurately perform a single task, for example 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, for example 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 video processing (Song et al. 2020) and 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, for example, Parks et al. (2018), which characterises the strong HI Ly α absorption in

⁵ <https://cosmohub.pic.es>

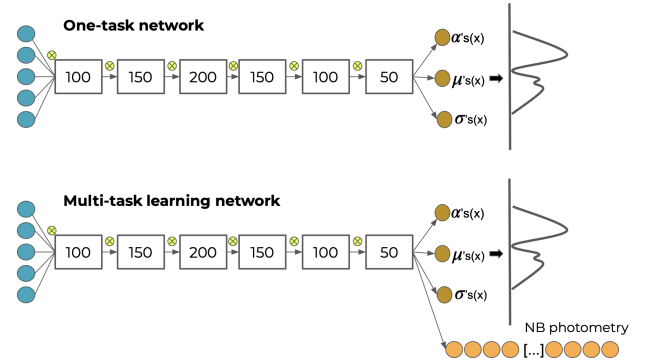


Fig. 2. Top: baseline network architecture. The input contains five colours that propagate through six fully connected layers. Each layer is followed by a dropout layer, which is represented by a yellow-crossed circle. Bottom: MTL network. This builds on the baseline network and adds an extra output layer for the additional task of predicting the narrow-band photometry.

quasar spectra simultaneously predicting the presence of strong HI absorption and the corresponding redshift z_{abs} and the HI column density. Also, Cunha & Humphrey (2022) describe SHEEP, a machine-learning pipeline for the classification of galaxies, quasi-stellar objects, and stars from photometric data. Broadly speaking, 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 amongst 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.

3.2. Model architecture and training procedures

In our analysis we used mixture density networks (MDNs) to predict the photo- z probability distribution as a linear combination of N independent Gaussians (D'Isanto & Polsterer 2018; Eriksen et al. 2020). The network predicts the mean and the standard deviation of N distributions, together with N additional mixing coefficients (α) weighting the relative importance of each Gaussian component to the combined probability distribution, so that $\sum_{i=0}^{i=N} \alpha_i = 1$.

Figure 2 shows the two MDNs used in this paper, both of them predicting the photo- z probability distribution $p(z)$ as the combination of three independent Gaussian distributions. The top panel presents the baseline network, a single-task network mapping the broadband photometry to the photometric redshifts. It concatenates six fully connected layers with parameters 5:300:500:1000:500:300:9, where the numbers correspond to the number of nodes in the layers. Therefore, the first contains five nodes, corresponding to the $uBVriz$ broadband colours. The last layer consists of nine output parameters corresponding to the mean (z), the standard deviation (σ_i), and the mixing coefficients, α , of the three Gaussians building the $p(z)$. Each layer is followed by a 2% dropout layer (Srivastava et al. 2014), a regularisation method in which several nodes are randomly ignored during the training phase.

The bottom panel in Fig. 2 represents the MTL network introduced in this paper, which includes the additional task of predicting the PAUS narrow-band photometry using a hard parameter-sharing architecture (Fig. 2). 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.

The photo- z loss function of both networks is the negative log-likelihood:

$$\mathcal{L}_z := \sum_{i=1}^N \left[\log(\alpha_i) - \frac{(z_i - z_s)^2}{\sigma_i^2} - 2 \log(\sigma_i) \right]. \quad (1)$$

The ground-truth redshift labels are the spectroscopic redshifts (z_s) as defined in Sect. 2.4 and the summation is over the Gaussian components. For some training configurations, we also used high-precision photo- z s (Sect. 2.2) as ground-truth labels to extend the photo- z training sample beyond the spectroscopic sample.

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. The two tasks share internal representations when predicting the photo- z and the narrow-band photometry simultaneously; 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,

$$\mathcal{L}_{\text{NB}} := \frac{\sum_i |\text{NB}_i^{\text{pred}} - \text{NB}_i^{\text{obs}}|}{N - 1}, \quad (2)$$

where $\text{NB}_i^{\text{pred}}$ 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. We also tested other alternatives, for example the mean-squared error, but this was hindering the network's convergence and we decided on the absolute-mean error. Another alternative was to predict the probability distribution of the narrow-band fluxes using a MDN as well, but this did not result in better photo- z estimations.

Consequently, there are the following two training methodologies. The first is z_s : This is the usual training that maps the broadband photometry to photo- z using spectroscopic redshifts as ground-truth redshifts and a negative log-likelihood loss function (Eq. (1)).

The second is $z_s + \text{NB}$: This methodology includes MTL. It maps the broadband photometry to photo- z and narrow-band photometry, and therefore the loss function is the mean of the combined negative log-likelihood loss (Eq. (1)) and narrow-band reconstruction (Eq. (2)) tasks for all galaxies (N) for which the loss is computed as

$$\mathcal{L}_{\text{NB}+z_s} := \frac{1}{N} \sum_{j=1}^N [\mathcal{L}_z^j + \mathcal{L}_{\text{NB}}^j]. \quad (3)$$

We only used galaxies with spectroscopic redshift to train the photo- z predictions, while all galaxies with narrow-band observations trained 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 (Sect. 2.2) to train the networks: the first is $z_s + z_{\text{PAUS}}$: This is a variation of the z_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.

The second is $z_s + \text{NB} + z_{\text{PAUS}}$: This is a variation of the $z_s + \text{NB}$ 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 $z_s + \text{NB}$ 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). All the training procedures use an Adam optimiser (Kingma & Ba 2015) for 100 epochs with an initial learning rate of 10^{-3} that reduces by a factor of ten every 50 epochs.

4. Photo- z performance in the COSMOS field

In this section we show the photo- z performance of our method on galaxies with $i_{\text{AB}} < 23$ and $z < 1.5$ in the COSMOS field. We study the effect that MTL has on the dispersion (Sect. 4.2) and the bias (Sect. 4.3) of the predicted photo- z s.

4.1. Photo- z performance metrics

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

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

where z_p and z_t are the mean predicted 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

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

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

$$\sigma_{\text{NMAD}} := 1.4826 \times \text{median} [|\Delta z - \text{median}(\Delta z)|] \quad (6)$$

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 (Sect. 2.2)⁶. 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- z s are also used, but only to evaluate the photo- z of galaxies from the full COSMOS catalogue that do not have a spectroscopic redshift estimate. The predicted photo- z s are defined as the mean of the redshift probability distribution provided by the network (Sect. 3.2).

In order to estimate the photo- z s of the complete COSMOS catalogue, the networks are trained independently ten times with $\sim 11\,000$ spectroscopic galaxies in each iteration, which roughly corresponds to 90% of the sample. Each network is used to

⁶ The PAUS+COSMOS photo- z s used to evaluate the precision of non-spectroscopic galaxies (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.

Table 1. Photo- z dispersion $\sigma_{68} \times 100$ for the different network configurations.

	Precision (z_s sample)	Precision (COSMOS)	Outliers (z_s sample)	Outliers (COSMOS)
z_s	2.00 (1.92)	2.43 (2.30)	0.6	1.1
z_s +NB	1.76 (1.68)	2.21 (2.04)	0.5	0.8
$z_s + z_{\text{PAUS}}$	1.68 (1.62)	2.03 (1.91)	0.5	0.6
$z_s + z_{\text{PAUS}}$ +NB	1.63 (1.58)	1.99 (1.86)	0.5	0.6

Notes. The second column displays results restricted to the spectroscopic sample, while the third column shows the results for the full COSMOS to $i_{\text{AB}} < 23$. For the full COSMOS sample results, the PAUS+COSMOS high-precision photo- z s are used as ground-truth redshifts when spectroscopy is not available. The numbers in parenthesis corresponds to the σ_{NMAD} . The fourth and fifth columns present the percentage of photo- z outliers on the COSMOS spectroscopic and complete sample, respectively.

evaluate the corresponding 10% of 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 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 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.

4.2. Photo- z dispersion

Table 1 presents the photo- z precision for the COSMOS spectroscopic sample and the complete COSMOS sample using the four different training procedures presented in Sect. 3.2. These results are presented in more detail in Fig. 3, which shows the photo- z dispersion in equally populated magnitude and redshift bins with the same four methodologies. The solid black line corresponds to the baseline network mapping broadband photometry to photo- z (method z_s in Sect. 3.2). This method is trained on the spectroscopic sample and provides a $\sigma_{68} = 0.020$ for the full sample. These are quite precise and accurate redshifts compared to other broadband redshift estimates in the same field. In Hildebrandt et al. (2009), redshifts in the D2 CFHT 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 having these CFHT photo- z estimates as a reference improves our photo- z baseline network performance. Here neither the methodology nor the input data are the same, but the CFHT photo- z s are a reference to compare the performance of our baseline network with.

In Fig. 3 we show the MTL training (method z_s +NB in Sect. 3.2) that 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 (see the bottom panel of Fig. 2) from around 12 000–30 000 galaxies. This method provides a precision of $\sigma_{68} = 0.0176$, corresponding to a 13% improvement with respect to the baseline methodology (solid black line). Moreover, the additional PAUS galaxies for the narrow-band reconstruction loss includes a more homogeneous colour-space coverage in the training sample. In Sect. 6 we discuss the underlying mechanism that causes MTL with PAUS to improve the photo- z s.

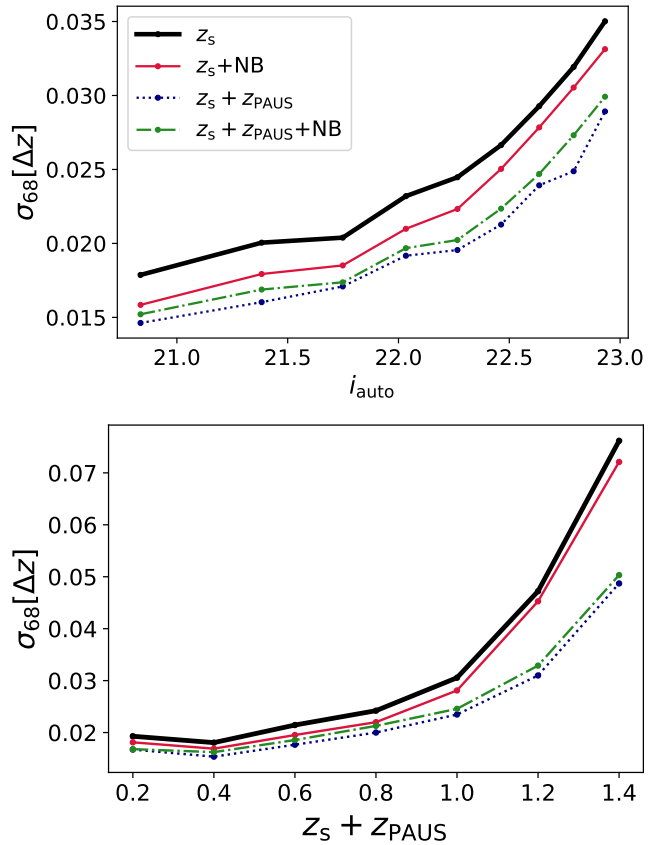


Fig. 3. Photo- z dispersion in equally populated magnitude differential bins to $i_{\text{AB}} < 23$ (top) and equally spaced redshift bins to $z < 1.5$ (bottom). Each line corresponds to a different training procedure (see Sect. 3.2). While the black line corresponds to a baseline training, the other coloured lines include MTL (red and green lines) and data augmentation with photo- z s from the PAUS+COSMOS catalogue as ground-truth redshifts (blue and green lines).

The blue dotted line in Fig. 3 also corresponds to a direct mapping of the broadband photometry to photo- z s. However, in contrast to the solid black line, this case is trained on an extended sample including galaxies without spectroscopic redshifts (method $z_s + z_{\text{PAUS}}$ 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 $\sigma_{68} = 0.0168$, which corresponds to a 18% improvement with respect to the baseline training.

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 Sect. 3.2), which corresponds to the

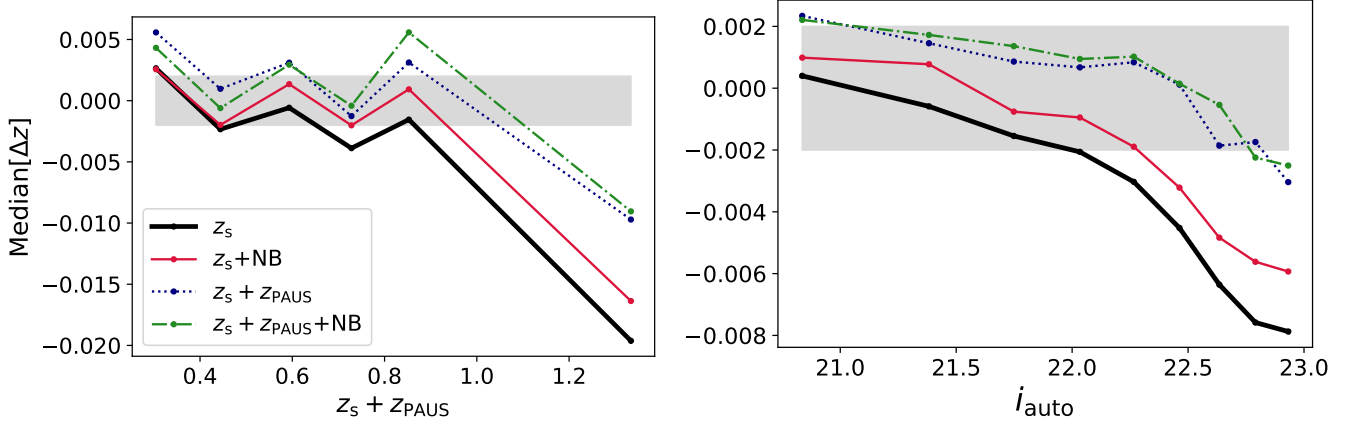


Fig. 4. Photo- z bias in equally populated redshift bins (left) and equally populated i -band magnitude bins (right). The grey area corresponds to the *Euclid* photo- z bias requirement of $\Delta z = 0.002$.

dotted green line in Fig. 3. This method gives a 22% improvement with respect to the baseline network, with a precision of $\sigma_{68} = 0.0163$.

In addition to uncertainties due to limited sample size (sample variance), our findings could also be affected by the intrinsic galaxy distribution being different at different parts of the sky (cosmic variance). To ensure that our results are not due to imprinting cosmic variance from the training to the test field, we tested our methods on two independent and spatially separated fields. These two fields are $\sim 2 \text{ deg}^2$ and contain galaxies from the Flagship simulations to $i < 23$ (Sect. 2.5). All the networks have been trained with 30 000 galaxies from the train field, and later evaluated on 20 000 galaxies different galaxies from the train and test fields (making sure that there is no overlap between the training and test galaxies in the train field). We estimated the sample variance of each of these fields by making 1000 bootstrap realisations and it is a sub-percent error. With the baseline z_s method, we obtain a 2% change in the photo- z precision between the train and test fields. Repeating the same test with the $z_s + z_{\text{PAUS}} + \text{NB}$ method, we obtain a 3% change in the photo- z precision between fields, similar to the baseline case. These changes are much lower than the photo- z improvement we obtain with the MTL implementations (e.g. 22% for the $z_s + z_{\text{PAUS}} + \text{NB}$ method), suggesting that such improvements are not caused by cosmic variance.

4.3. Photo- z bias and outlier rate

In this subsection we show the bias and the outlier rate for the photo- z predictions with the MTL networks and the baseline broadband network. The left panel in Fig. 4 shows the photo- z bias in equally populated redshift bins in the redshift range $0.1 < z_t < 1.5$. We 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 redshift⁷. The shaded area corresponds to the *Euclid* photo- z bias requirement < 0.002 (Laureijs et al. 2011). Overall, for $z_t < 1$ the four methods presented in Sect. 3.2 are unbiased at the level of < 0.002 . However, the z_s and the $z_s + \text{NB}$ are still showing a trend within the 0.2% bias range, where low-redshift galaxies tend to be biased positive and the

⁷ There are training mechanisms to deal with unbalanced training samples such as 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.

high-redshift ones, biased negative. In contrast, the $z_s + z_{\text{PAUS}}$ and the $z_s + z_{\text{PAUS}} + \text{NB}$ methods display a flatter bias with redshift.

At higher redshifts ($z_t > 1$), the baseline network photo- z s show a $\sim 2\%$ bias. Implementing MTL without increasing the photo- z training sample (solid red line) moderately improves the bias, but it is still far from the *Euclid* requirement. On the other hand, increasing the training sample with PAUS+COSMOS photo- z s produces a strong bias reduction (blue and green lines), decreasing the bias to $\sim 1\%$ for the highest-redshift galaxies. Figure 1 suggests that this is likely to be caused by a lack of training examples with spectroscopy at $z_t > 1$. The training sample at high-redshift is increased with the PAUS+COSMOS photo- z s.

The right panel of Fig. 4 shows the photo- z bias in equally populated i -band magnitude bins. Comparing to the right panel in the same figure, the bias binning in i -band magnitude is lower than that binning in redshift. For instance, binning in redshift the largest bias that the z_s method obtains is a $\sim 2.5\%$ for the highest-redshift galaxy bin. In contrast, binning in magnitude, galaxies in the faintest bin reach a 0.8% bias with the same method. This is partly because binning in magnitude, positive and negative biases in redshift cancel each other out. The photo- z s of galaxies with $i < 22$ are unbiased with the four methods. For galaxies with $i > 22$, the z_s method displays the largest bias, which is already reduced with the MTL method without data augmentation ($z_s + \text{NB}$). The methods extending the sample using the PAUS+COSMOS photo- z s (green and blue lines) reduce the bias of the z_s and the $z_s + \text{NB}$ methods.

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

$$|z_p - z_t| / (1 + z_t) > 0.15. \quad (7)$$

In the spectroscopic sample, the baseline network yields 0.6% outliers, which reduces to 0.5% 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 PAUS sample in COSMOS is 1.1% for the baseline network and for the training sample extension with PAUS+COSMOS photo- z s ($z_s + z_{\text{PAUS}}$). The methodologies including MTL reduce the outlier fraction to 0.8% ($z_s + \text{NB}$) and 0.6% ($z_s + z_{\text{PAUS}} + \text{NB}$). While in the spectroscopic sample extending the training sample and including MTL have a similar effect on the outlier fraction, in the full PAUS sample in COSMOS MTL has a stronger impact. The MTL methodologies are particularly reducing the number of high-redshift photo- z outliers.

In order to validate the predicted photo- z probability distributions $p(z)$, we use the probability integral transform (PIT;

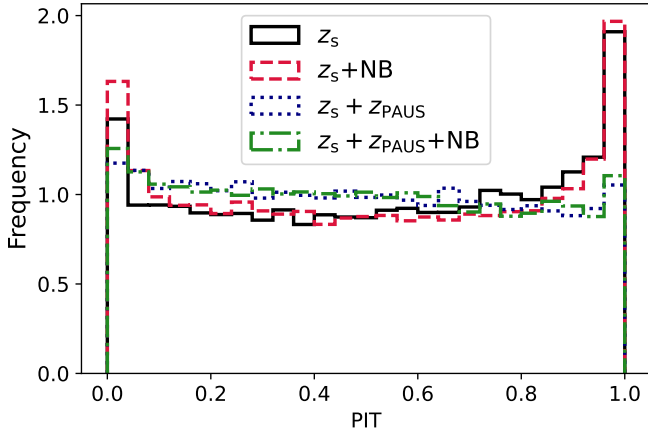


Fig. 5. PIT distribution for the COSMOS photo- z s predicted with the baseline z_s method (black), the z_s +NB method (red), the z_s + z_{PAUS} method (blue), and the z_s + z_{PAUS} +NB method (green). Including the PAUS+COSMOS photo- z s in the training reduces the number of outliers on the edges of the distribution.

Dawid 1984; Gneiting et al. 2005; Bordoloi et al. 2010), which is defined as

$$\text{PIT} \equiv \int_{-\infty}^{z_{\text{true}}} dz p(z), \quad (8)$$

where z_{true} is the true redshift. When the $p(z)$ faithfully represents the true redshift, the PIT distribution is the uniform distribution $U[0,1]$. Contrary, PIT histograms with peaks at the edges (i.e. around zero and unity) indicate the presence of outlier measurements. Also, PIT histograms more populated at the centres than on the edges denote over-dispersed probability distributions, while valleys at the centre of the histogram correspond to under-dispersed ones.

We measure the PIT distribution for the complete COSMOS sample using a combination of spectroscopic redshifts and high-precision photo- z as true redshift. Figure 5 shows the PIT distributions for the $p(z)$ measured with the baseline z_s method (black line), the MTL method (dashed red line), extending the training sample with high-precision photo- z (dotted blue line), and combining the training sample augmentations and MTL (green line). In all cases, the PIT distribution is approximately a $U[0,1]$ distribution, indicating that our networks predict robust probability distributions with reliable uncertainty measurements. The baseline and the z_s +NB methods display peaks on the edges of the distribution corresponding to outliers in the probability distributions. These peaks are reduced with the two methods using PAUS+COSMOS photo- z s in the training 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 $i_{\text{AB}} < 23$ (see Sect. 4). However, if the MTL network developed in this paper aims to improve the photo- z estimates of future deeper broadband surveys such as *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 10σ depth for extended sources, which corresponds to a similar depth

in the i -band filter. Rubin will observe to a single exposure depth of $r_{\text{AB}} \sim 24.5$ and a co-added survey depth of $r_{\text{AB}} \sim 27.5$ (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 $i_{\text{AB}} = 23$, 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 s for much deeper data samples (Masters et al. 2017).

In this section we explore how the MTL network performs for deep samples ($i_{\text{AB}} < 25$), while the training is limited to galaxies with $i_{\text{AB}} < 23$ using Flagship simulated 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* Near-infrared spectrometer and photometer (NISF) near-infrared bands H_E , J_E and Y_E (Euclid Collaboration 2022)⁸. These are not the same bands that were used in the tests of the COSMOS field (see Sects. 2.3 and 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 10 000 spectroscopic galaxies, which are augmented to 30 000 with PAUS-like galaxies without spectroscopic redshifts and limited to $i_{\text{AB}} < 23$. 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- z s (methods $z_s + z_{\text{PAUS}}$ and z_s +NB+ 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- z s.

The left panels in Fig. 6 show the photo- z bias of 30 000 simulated test galaxies to magnitude $i_{\text{AB}} < 25$ in equally populated magnitude bins (top) and in equally spaced redshift bins (bottom). The shaded areas correspond to the *Euclid* photo- z requirement of $\Delta z < 0.002$. We obtain a larger bias than the *Euclid* requirement with all methods, although those including MTL reduce the bias of fainter galaxies. Although we are not meeting the *Euclid* bias requirement, our aim is to advance the usage of machine-learning photo- z developing novel methodology, rather than providing the final pipeline. We hope the improvement and ideas seen in this paper can be helpful for further development of *Euclid* machine-learning algorithms.

The right panels in Fig. 6 show the photo- z precision for the same 30 000 simulated test galaxies to magnitude $i_{\text{AB}} < 25$ in magnitude (top) and redshift (bottom) bins. The baseline network (black thick line) achieves an overall precision of $\sigma_{68} = 0.076$, which increases to $\sigma_{68} = 0.085$ for galaxies with $i_{\text{AB}} > 23$. Training using photo- z s but without MTL ($z_s + z_{\text{PAUS}}$, dotted blue line) improves the precision to $\sigma_{68} = 0.0654$ and $\sigma_{68} = 0.080$ for galaxies with $i_{\text{AB}} > 23$. With z_s +NB, the overall precision is $\sigma_{68} = 0.067$, which degrades to $\sigma_{68} = 0.082$ for galaxies with $i_{\text{AB}} > 23$. Finally, combining MTL and the photo- z data augmentation (z_s +NB+ z_{PAUS} , solid green line) provides the best photo- z performance with $\sigma_{68} = 0.065$ for the full sample,

⁸ With the following 5σ limiting magnitudes: u : 25.25; g : 24.65; r : 24.15; i : 24.35; z : 23.95; Y_E : 24.0; J_E : 24; H_E : 24.

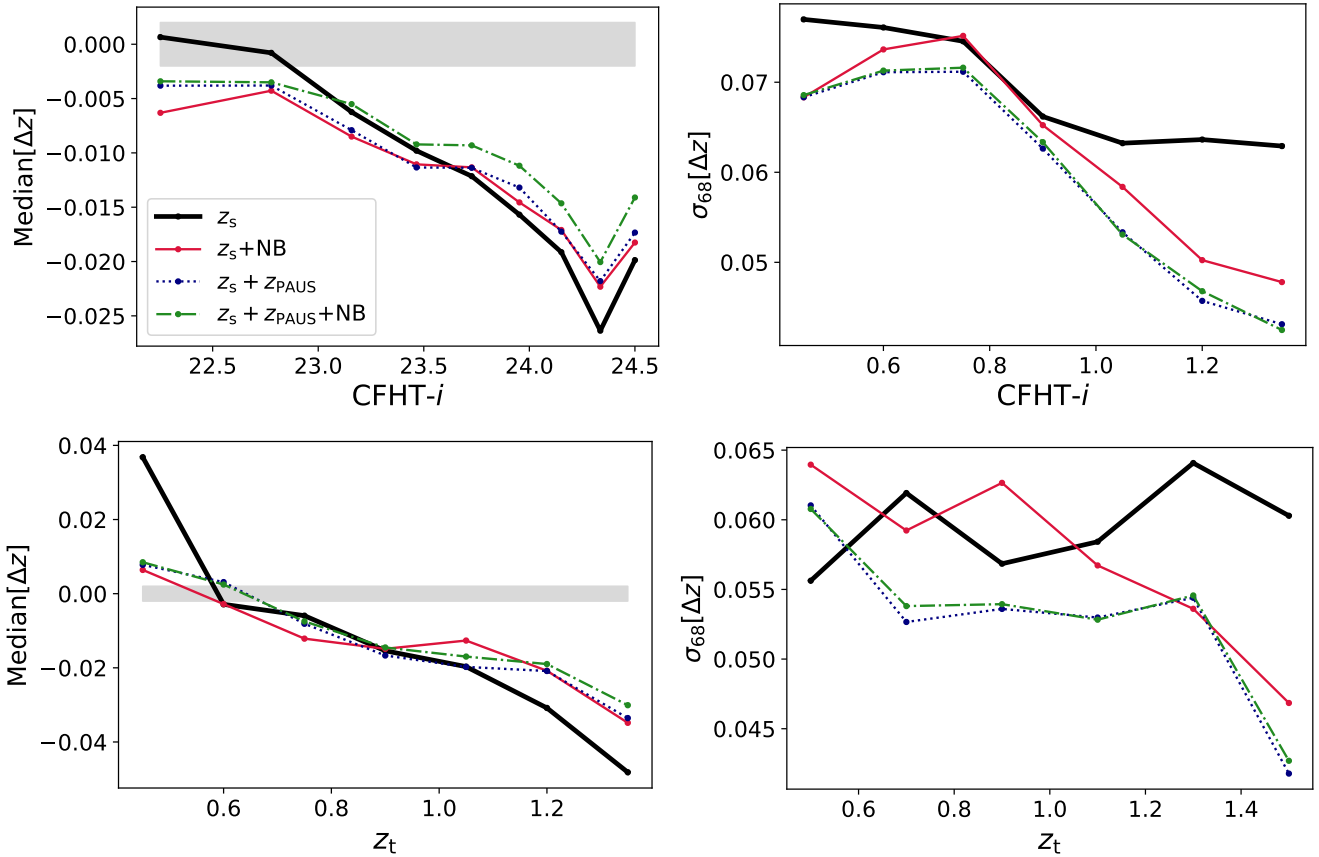
Photo- z s in Flagship

Fig. 6. Top: photo- z bias (left) and precision (right) in equally populated magnitude bins. Bottom: photo- z bias (left) and dispersion (right) in equally spaced spectroscopic redshift bins. The shaded grey areas indicate $\Delta z > 0.002$, corresponding to the *Euclid* requirement for the photo- z bias. All plots are for 30 000 Flagship test galaxies with magnitudes $i_{AB} < 24.5$ 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 $i_{AB} < 23$.

which increases to $\sigma_{68} = 0.079$ for galaxies with $i_{AB} > 23$. The best performance in terms of bias and precision is obtained with the $z_s+z_{PAUS}+NB$ method, which provides 16% more precise photo- z s than the baseline network for galaxies with $i_{AB} < 25$, which increases to 20% for $i_{AB} < 24$.

6. Photo- z in colour space

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 to explore the COSMOS photo- z performance in colour space (Sect. 6.1). Furthermore, in Sects. 6.2 and 6.3 we identify colour-space regions with strong emission lines where the broadband photo- z s precision is lower.

6.1. MTL photo- z in colour space

A SOM (Kohonen 1982)⁹ 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, N_y) cells, each of them with

⁹ <https://github.com/lauracabayol/SOM>

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). The appendices contain a more detailed explanation of SOM algorithms. Self-organising maps 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. 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 7 shows the predicted photo- z s 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

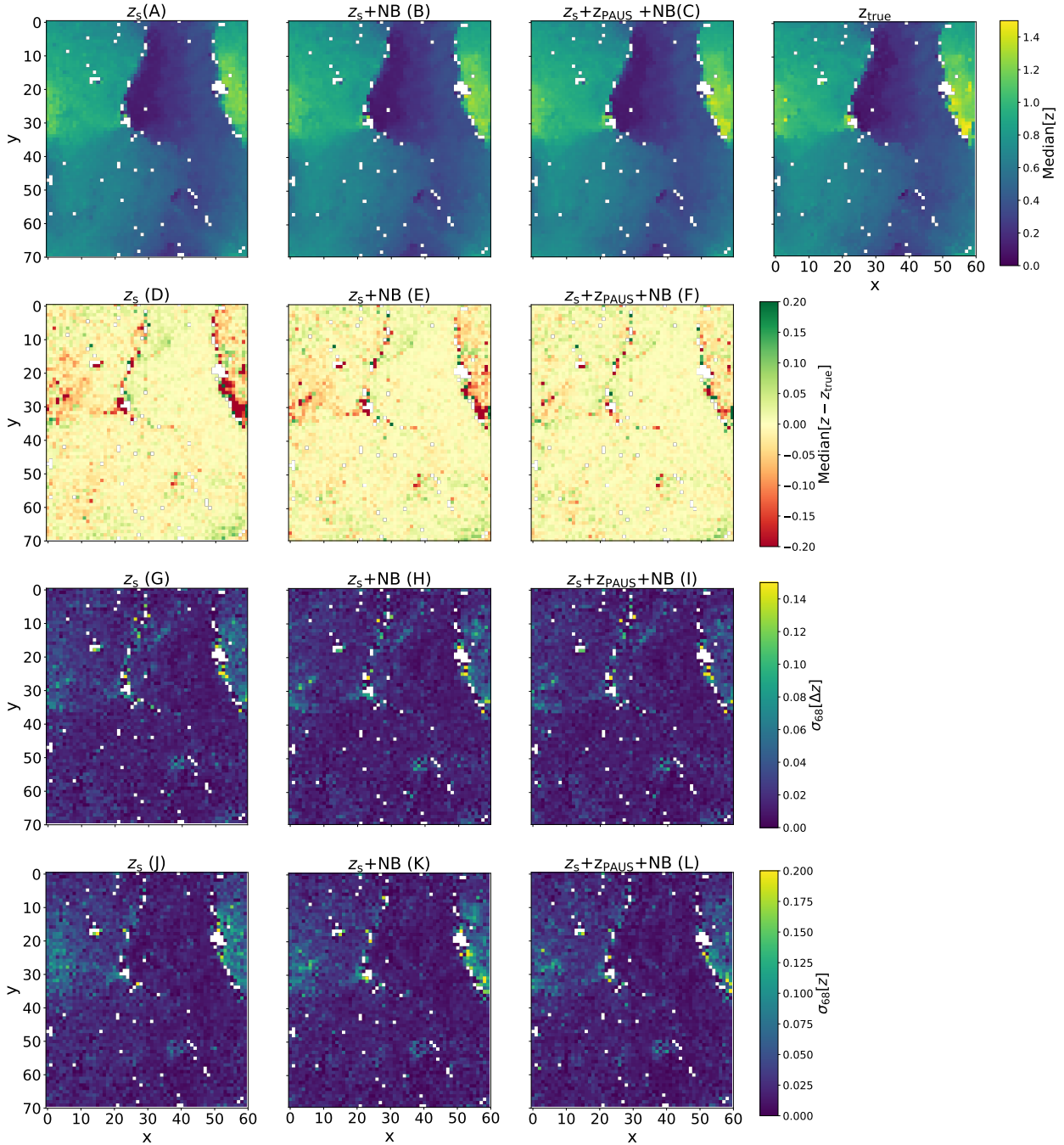


Fig. 7. SOMs showing the photo- z performance in the COSMOS field. The first row exhibits the median predicted photo- z in colour space for the baseline network (first panel), including MTL training (second panel), with MTL and data augmentation with PAUS+COSMOS photo- z s (third panel), and the ground-truth redshift (fourth panel). The second row shows the bias in the photo- z predictions for the three training methods of the first row (first three 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, that is, cells without any galaxy.

median photo- z of the galaxies it contains. The leftmost panel (z_s , panel A) displays the photo- z s with the baseline network (z_s method), and the second (B) and third (C) panels include MTL in the training (i.e. z_s +NB and NB+ z_s + z_{PAUS} methods, respectively; bottom panel of Fig. 2). 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_t). The network trained with only broad bands (panel D) exhibits two regions with less accurate photo- z s. These regions are centred around coordinates (5, 35) and (55, 25), and the redshift accuracy improves when MTL (panel E) or z_{PAUS} +MTL (panel F) are included in the training.

These regions are also spotted in the third row of Fig. 7, which shows the photo- z precision (σ_{68} , Eq. (5)). Comparing panels G and D, we note that the photo- z precision worsens in the same regions where photo- z s are less accurate, but this

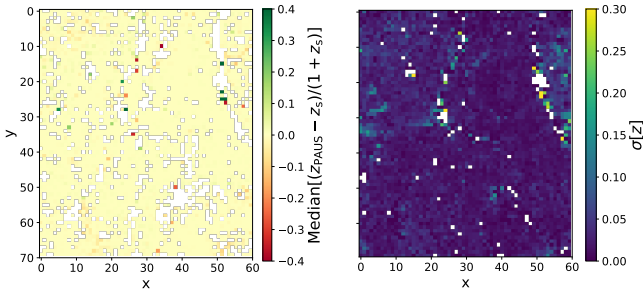


Fig. 8. Bias (left) and precision (right) of the PAUS+COSMOS photo- z s in the COSMOS spectroscopic sample.

moderately improves with MTL (z_s +NB, panel H) and including the PAUS+COSMOS photo- z s (NB+ z_s + z_{PAUS} , panel F). Finally, the fourth row shows the dispersion of the redshift distribution (i.e. the width of the $N(z)$) within 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_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, galaxies with different redshifts, are assigned to these cells. Figure 8 supports this hypothesis by showing that the PAUS+COSMOS photo- z s also exhibit a higher redshift dispersion (right panel) in the SOM cells within the problematic regions, while the PAUS+COSMOS photo- z accuracy is smooth across colour space (left panel). Therefore, there are galaxies with different redshifts clustered together in broadband colour space.

6.2. Broadband degeneracies in colour space

Self-organising map cells that contain different galaxy populations can be the result of colour-redshift degeneracies in the broadband photometry. Such broadband degeneracies also cause the worse photo- z performance of the baseline network in the problematic colour-space regions. The photo- z performance improves with the MTL training (panel E in Fig. 7).

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 $uBVriz$ broadband 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

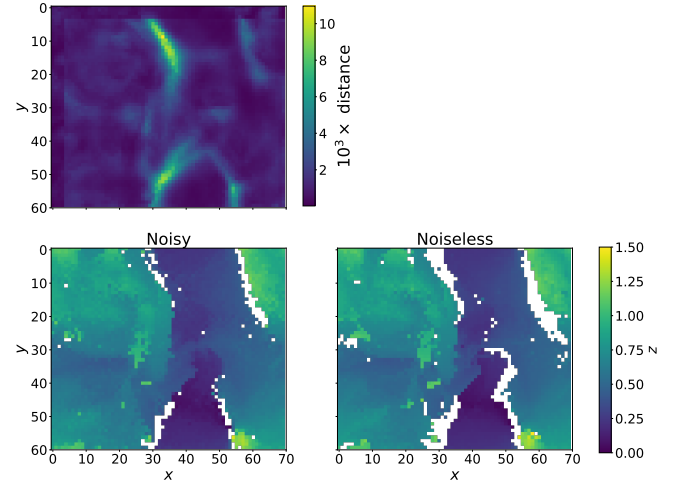


Fig. 9. SOM trained on a galaxy simulated mock with the $uBVriz$ broad bands. Top: distance between every SOM cell vector and its 3×3 neighbours. Bottom left: median photo- z in each SOM cell for noisy simulated galaxies. Bottom right: median photo- z in each SOM cell for noiseless simulated galaxies.

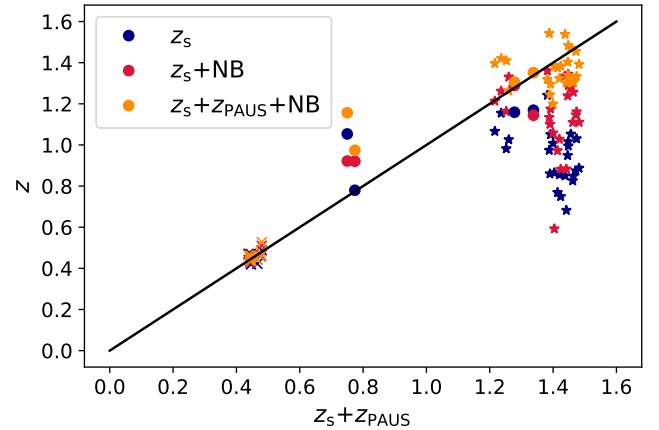


Fig. 10. Photo- z scatter for galaxies in three independent SOM cells. The galaxies in each cell are represented with a different marker (stars, crosses, and circles).

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 broadband 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 potential presence of colour-redshift degeneracies. Figure 10 shows the photo- z s of the galaxies assigned to three different cells within such a colour-space region. There are four different redshift populations assigned to the region: $z \sim 1.4$, $z \sim 0.4$, and $z \sim 1.2$, which is many times confused with galaxies at $z \sim 0.8$. 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 (z_s , blue), the network including MTL (z_s +NB, red), and that including MTL and photo- z data augmentation (z_s +NB+ z_{PAUS} , orange).

The first cell (marked with stars) contains galaxies with $z_t \sim 0.4$ and the three networks predict the correct redshift.

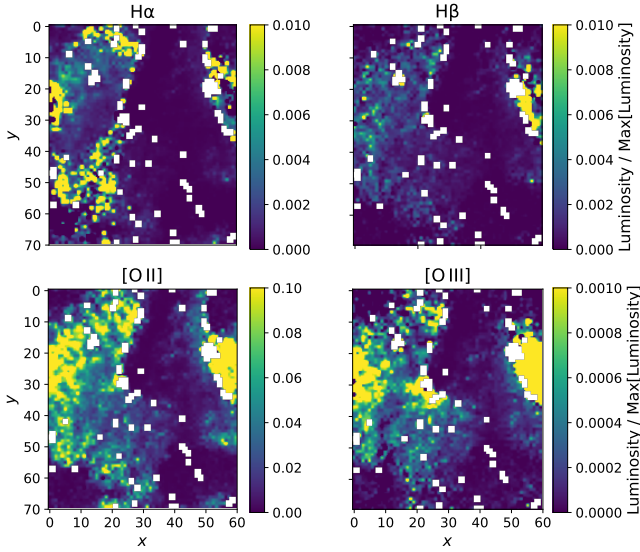


Fig. 11. Emission-line luminosity in colour space for $H\alpha$, $H\beta$, $[O II]$, and $[O III]$, as indicated in the title.

The second cell (marked with crosses) contains galaxies with $z_{\text{true}} \sim 0.8$ and $z_{\text{true}} \sim 1.2$. In general, the MTL network improve the photo- z prediction of these galaxies. Lastly, the third cell (marked with dots) contains galaxies with redshifts $z_t \sim 1.4$. The baseline network predicts these photo- z s around $z_p \sim 0.8$, and again the z_s +NB and the MTL+ z_{PAUS} training approaches are able to improve the photo- z s. Photo- z confusions from $z_t \sim 0.8$ to $z_t \sim 1.2$ and from $z_t \sim 1.45$ to $z_t \sim 1.25$ are recurrent, showing up at several SOM cells within the low photo- z performance cluster.

Figure 11 explores the mean $H\alpha$, $H\beta$, $[O II]$, and $[O III]$ emission-line luminosity in colour space. The emission-line luminosity is estimated as

$$L_{\text{el}} := 4\pi F_{\text{el}} D_L^2, \quad (9)$$

where F_{el} is the emission-line flux and D_L is the luminosity distance, which is estimated assuming *Planck* 2020 cosmology (Planck Collaboration VI 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 \AA wide Gaussian distribution.

Figure 11 shows strong emission lines at the low photo- z performance colour-space regions, for example 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 broadband data.

Since a high ratio of $[O III]$ to $H\beta$ lines may indicate the presence of active galactic nuclei, 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 and stellar masses, often called the main sequence (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 -broadband 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 use a variation of our fiducial network to encode the galaxy photometry in a 2-dimensional space similar to a SOM, while in Sect. 7.2 we study the impact of using other auxiliary tasks (other than predicting the narrow-band photometry) in the MTL network.

7.1. Underlying data representation in colour space with MTL

For this test, we modify the fiducial network architecture (see Sect. 3.2 and Fig. 2). In the modified network, we reduce the input dimension to two features, which are used to predict the photo- z and reconstruct the narrow-band colours. Encoding the galaxy information in a two-dimensional feature space simplifies its visualisation and brings it closer to the SOM colour-space representation, which we have already studied (Sect. 6).

The galaxy representations in the two-dimensional 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 and thus 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.

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 (e.g. the feature map of the z_s network and that of the z_s +NB) cannot be directly compared. However, the overlap of different redshift populations in feature-space indicate potential degeneracies.

In Fig. 12 we plot the 50%, 68%, and 95% contours of the feature-space coordinates for the z_s (left panel) and the z_s +NB (right panel) methods. These are drawn using a test set of 70 000 Flagship galaxies (Sect. 2.5) to $i_{\text{AB}} < 25$, while the methods train on galaxies to $i_{\text{AB}} < 23$ (Sect. 5). We draw the contours for a selection the redshift bins to show the separation of high-redshift galaxies (red, blue, and green contours), where the MTL method significantly reduces the photo- z scatter with respect to the z_s method (Fig. 6, top and bottom left panels). We have also plotted the contours of a distant redshift population (purple contours) to show that this is further in feature space than the others.

There is a significant overlap amongst high-redshift populations in the z_s case (left panel). Particularly, the core of the green and blue contours overlap with the red-contour galaxies. We expect some overlap since the three contours are consecutive in redshift; however, the z_s +NB method shows a cleaner separation between the three redshift populations. This indicates that the z_s +NB has a better internal representation of the galaxies, where different redshift populations are further in feature space. The narrow-band reconstruction loss (Eq. (2)) adds the

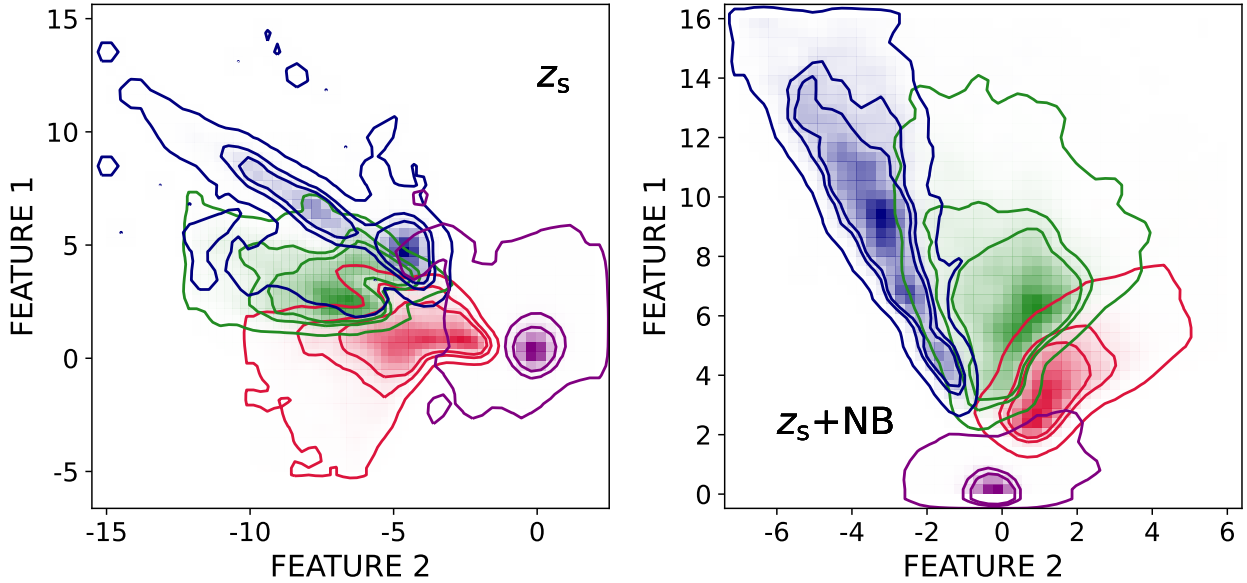


Fig. 12. Contours of the two-dimensional feature space coordinates for the z_s (left) and z_s+NB (right) methods. The features from each of the methods are from independent training and cannot be compared. We can only compare the overlap of the different populations.

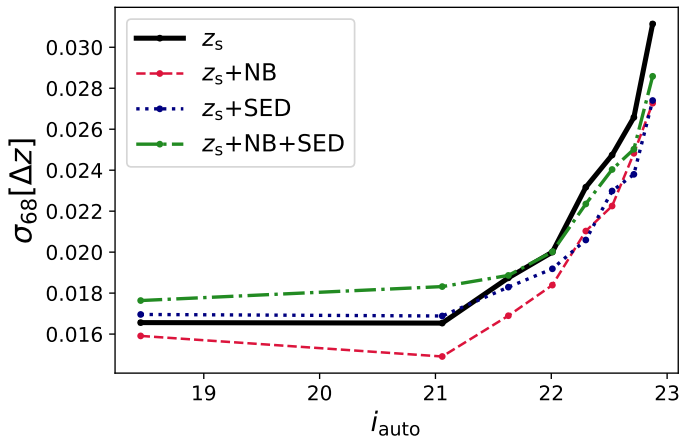


Fig. 13. Photo- z precision in the COSMOS field when the auxiliary task of predicting the galaxy SED is included in the training. The galaxy SED prediction is addressed as a classification, where the true SED is a class between 1 and 47.

low-resolution SED information to the training, which can potentially lead to an improved internal representation of galaxies in the two-dimensional feature space. Furthermore, MTL methods also include this information for galaxies without spectroscopic redshift, which effectively acts as a data augmentation technique. This is particularly important for high-redshift galaxies, for which we have very few examples in the spectroscopic sample (Fig. 1).

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 broadband photometry (f), namely $\phi(z|f)$, with MTL the optimisation is extended to the function that best predicts the photo- z together

with other related parameters (x_i),

$$\phi(z, x_1, \dots, x_N | \theta), \quad (10)$$

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

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 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 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 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.

Figure 13 also shows the photo- z performance when both the SED and the narrow-band data are used as auxiliary tasks (green dashed-dotted line). 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 star-formation rate, the galaxy mass, and the $E(B-V)$ extinction parameter as auxiliary tasks (not shown). However, none of

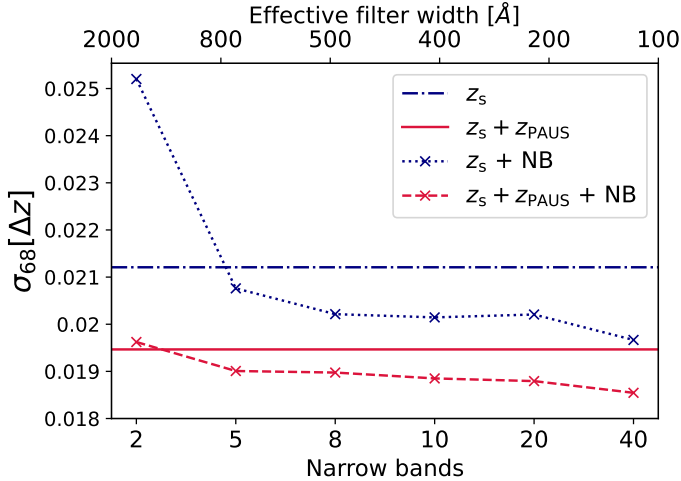


Fig. 14. Photo- z precision as a function of number of bands in the predicted photometry for $z_s + \text{NB}$ (dotted blue line) and $z_s + z_{\text{PAUS}} + \text{NB}$ (dashed red line). The horizontal line corresponds to the z_s (dashed-dotted blue line) and $z_s + z_{\text{PAUS}}$ (solid red line), where MTL is not enabled.

these parameters improved the predicted photo- z s. Furthermore, including the near-infrared photometry did not improve the photo- z s either.

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 14 shows the photo- z precision of the MTL methods as a function of the number of predicted narrow bands (i.e. the 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, and 10). Then, we train the $z_s + \text{NB}$ and $z_s + z_{\text{PAUS}} + \text{NB}$ methods several times to predict the photo- z and the narrow-band photometry with a different resolution in every training.

The horizontal flat lines in Fig. 14 indicate the photo- z precision for the methods without MTL; z_s (dashed-dotted blue line) and $z_s + z_{\text{PAUS}}$ (solid red line). The dotted blue line and the dashed red line show the $z_s + \text{NB}$ and $z_s + z_{\text{PAUS}} + \text{NB}$ 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 $z_s + \text{NB}$ MTL recovering two-band photometry leads to predictions above the z_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.

8. Discussion and conclusions

Photometric redshifts are crucial for exploiting ongoing and future large galaxy broadband imaging surveys. While covering large sky areas, the broadband spectral resolution limits the redshift performance through colour-redshift degeneracies. The 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 for improving broadband photo- z estimates, using deep-learning techniques on PAUS narrow-band data.

Multi-task learning is a machine-learning 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 implemented an MTL network that simultaneously predicts the photometric redshift and infers the narrow-band photometry from the broadband 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 for galaxies to $i_{\text{AB}} < 23$, our method reduces the photo- z scatter by approximately 20% (see Sect. 4.2) and the number of photo- z outliers by from ~ 1.1 to $\sim 0.6\%$ (see Sect. 4.3). We also tested the potential of the method for fainter galaxies using *Euclid*-like galaxy simulations. For this, we trained the network on a magnitude-limited sample with $i_{\text{AB}} < 23$ and evaluated it on a sample with $i_{\text{AB}} < 25$. The MTL predicts up to 16% more precise photo- z s for galaxies with $i_{\text{AB}} < 25$ than the baseline network (see Sect. 5).

We used SOMs to study the photo- z performance in different colour-space regions, detecting a region that contains 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 the left panel in Fig. 8). This correlation results in a photo- z mismatch between two galaxy populations, which affects broadband photo- z estimates. Our MTL network improves the photo- z s in the degenerated colour-space regions 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 such as PAUS to improve broadband photo- z estimates using machine learning. The key point of using MTL instead of, for example, just using the narrow-band photometry to obtain more precise photo- z s is that it only requires narrow-band photometry for the training galaxies and the photo- z of any galaxy can be evaluated with only the broadband data. This enables us to exploit fields where we have narrow-band data to obtain better photo- z s in other fields where these are not available. PAUS photometry in the COSMOS field is publicly available, so current and future weak lensing surveys, such as *Euclid* or the LSST, can 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 in order to improve photo- z predictions. While PAUS narrow-band photometry is a clear candidate, other surveys such as 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 estimations.

Acknowledgements. The PAU Survey is partially supported by MINECO under grants CSD2007-00060, AYA2015-71825, ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, MDM-2015-0509, PID2019-111317GB-C31 and Juan de la Cierva fellowship and LACEGAL and EWC Marie Skłodowska-Curie grant No 734374 and no.776247 with ERDF funds from the EU Horizon 2020 Programme, some of which include ERDF funds from the European Union. IEEC and IFAE are partially funded by the CERCA and Beatriu de Pinos program of the Generalitat de Catalunya. Funding for PAUS has also been provided by Durham University (via the ERC StG DEGAS-259586), ETH Zurich, Leiden University (via ERC StG ADULT-279396 and Netherlands Organisation for Scientific Research (NWO) Vici grant 639.043.512), Bochum University (via a Heisenberg grant of the Deutsche Forschungsgemeinschaft (Hi 1495/5-1) as well as an ERC Consolidator Grant (No. 770935)), University College London, Portsmouth support through the Royal Society Wolfson fellowship and from the European Union's Horizon 2020 research and innovation programme under the grant agreement No 776247 EWC. The results published were also funded by the Polish National Agency for Academic Exchange (Bekker grant BPN/BEK/2021/1/00298/DEC/1), the European Union's Horizon 2020 research and innovation programme under the Maria Skłodowska-Curie (grant agreement No 754510) and by the Spanish Ministry of Science and Innovation through Juan de la Cierva-formacion program (reference FJC2018-038792-I). The PAU data centre is hosted by the Port d'Informació Científica (PIC), maintained through a collaboration of CIEMAT and IFAE, with additional support from Universitat Autònoma de Barcelona and ERDF. We acknowledge the PIC services department team for their support and fruitful discussions. CosmoHub has been developed by the Port d'Informació Científica (PIC), maintained through a collaboration of the Institut de Física d'Altes Energies (IFAE) and the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) and the Institute of Space Sciences (CSIC&IEEC), and was partially funded by the "Plan Estatal de Investigación Científica y Técnica y de Innovación" program of the Spanish government. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research. The Euclid Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of *Euclid*, in particular the Academy of Finland, the Agenzia Spaziale Italiana, the Belgian Science Policy, the Canadian Euclid Consortium, the French Centre National d'Etudes Spatiales, the Deutsches Zentrum für Luft- und Raumfahrt, the Danish Space Research Institute, the Fundação para a Ciência e a Tecnologia, the Ministerio de Economía y Competitividad, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Nederlandse Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Romanian Space Agency, the State Secretariat for Education, Research and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. A complete and detailed list is available on the *Euclid* web site (<http://www.euclid-ec.org>). Data availability: The PAUS raw data are publicly available through the ING group. A few reduced images are publicly available at <https://www.pausurvey.org>. 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, *SPIE Conf. Ser.*, 8442, 84420Z
- Arnouts, S., & Ilbert, O. 2011, *Astrophysics Source Code Library* [[record ascl:1108.009](https://arxiv.org/abs/1108.009)]
- Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, *ApJ*, 762, 109
- Benítez, N. 2011, *Astrophysics Source Code Library* [[record ascl:1108.011](https://arxiv.org/abs/1108.011)]
- Benitez, N., Dupke, R., Moles, M., et al. 2014, *ArXiv eprints* [[arXiv:1403.5237](https://arxiv.org/abs/1403.5237)]
- Blake, C., & Bridle, S. 2005, *MNRAS*, 363, 1329
- Bonnett, C. 2015, *MNRAS*, 449, 1043
- Bordoloi, R., Lilly, S. J., & Amara, A. 2010, *MNRAS*, 406, 881
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Buchas, R., Davis, C., Gruen, D., et al. 2019, *MNRAS*, 489, 820
- Cabayol, L., Eriksen, M., Amara, A., et al. 2021, *MNRAS*, 506, 4048
- Cabayol-García, 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
- Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2014, *MNRAS*, 447, 646
- Carretero, J., Tallada, P., Casals, J., et al. 2017, in *Proceedings of the European Physical Society Conference on High Energy Physics*. 5-12 July, 488
- Caruana, R. 1997, *Mach. Learn.*, 28, 41
- Casas, R., Cardiel-Sas, L., Castander, F. J., et al. 2016, *SPIE Conf. Ser.*, 9908, 99084K
- Castander, F. J., Ballester, O., Bauer, A., et al. 2012, *SPIE Conf. Ser.*, 8446, 84466D
- Chong, De Wei, K., & Yang, A. 2019, *Euro. Phys. J. Web Conf.*, 206, 09006
- Collès, 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, *SPIE*, 8442, 84420V
- Csörnyei, G., Dobos, L., & Csabai, I. 2021, *MNRAS*, 502, 5762
- Cunha, P. A. C., & Humphrey, A. 2022, *A&A* 666, A87
- Dawid, A. P. 1984, *J. R. Statist. Soc. Ser. A*, 147, 278
- de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, *Exp. Astron.*, 35, 25
- Desprez, G., Paltani, S., Coupon, J., et al. 2020, *A&A*, 644, A31
- D’Ianto, A., & Polsterer, K. L. 2018, *A&A*, 609, A111
- Eriksen, M., Alarcon, A., Gaztanaga, E., et al. 2019, *MNRAS*, 484, 4200
- Eriksen, M., Alarcon, A., Cabayol, L., et al. 2020, *MNRAS*, 497, 4565
- Euclid Collaboration (Schirmer, M., et al.) 2022, *A&A* 662, A92
- 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
- Gneiting, T., Raftery, A. E., Westveld, A. H., & Goldman, T. 2005, *Monthly Weather Rev.*, 133, 1098
- Gomes, Z., Jarvis, M. J., Almosallam, I. A., & Roberts, S. J. 2018, *MNRAS*, 475, 331
- Hasinger, G., Capak, P., Salvato, M., et al. 2018, *ApJ*, 858, 77
- Hildebrandt, H., Pielorz, J., Erben, T., et al. 2009, *A&A*, 498, 725
- Hildebrandt, H., Erben, T., Kuijken, K., et al. 2012, *MNRAS*, 421, 2355
- 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)
- Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, *MNRAS*, 478, 592
- Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, *A&A*, 556, A55
- 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
- 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 eprints* [[arXiv:1110.3193](https://arxiv.org/abs/1110.3193)]
- Lee, K.-G., Krolewski, A., White, M., et al. 2018, *ApJS*, 237, 31
- Liebel, L., & Körner, M. 2018, *ArXiv eprints* [[arXiv:1805.06334](https://arxiv.org/abs/1805.06334)]
- Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJS*, 172, 70
- LSST Science Collaboration (Abell, P. A., et al.) 2009, *ArXiv eprints* [[arXiv:0912.0201](https://arxiv.org/abs/0912.0201)]
- Martí, P., Miquel, R., Castander, F. J., et al. 2014, *MNRAS*, 442, 92
- Masters, D., et al. 2015, *ApJ*, 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
- Ménard, B., Scranton, R., Schmidt, S., et al. 2013, *ArXiv eprints* [[arXiv:1303.4722](https://arxiv.org/abs/1303.4722)]
- Moeskops, P., Wolterink, J. M., van der Velden, B. H. M., et al. 2017, *ArXiv eprints* [[arXiv:1704.03379](https://arxiv.org/abs/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, *Astropart. Phys.*, 63, 81
- Osterbrock, D. E., & Ferland, G. J. 2006, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* (USA: University Science Books)
- Padilla, C., Ballester, O., Cardiel-Sas, L., et al. 2016, *SPIE*, 9908, 99080Z
- Padilla, C., Castander, F. J., Alarcón, A., et al. 2019, *AJ*, 157, 246
- Parks, D., Prochaska, J. X., Dong, S., & Cai, Z. 2018, *MNRAS*, 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., 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 VI. 2020, *A&A*, 641, A6
- Planck Collaboration XIII. 2016, *A&A*, 594, A13
- Polletta, M., Tاجر, M., Maraschi, L., et al. 2007, *ApJ*, 663, 81
- Potter, D., Stadel, J., & Teysier, R. 2017, *Comput. Astrophys. Cosmol.*, 4, 2
- Pozzetti, L., Hirata, C. M., Geach, J. E., et al. 2016, *A&A*, 590, A3
- 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, *Nat. Astron.*, **3**, 212
- Sánchez, C., Carrasco Kind, M., Lin, H., et al. 2014, *MNRAS*, **445**, 1482
- Schmidt, S. J., Ménard, B., Scranton, R., Morrison, C., & McBride, C. K. 2013, *MNRAS*, **431**, 3307
- Schmidt, S. J., Malz, A. I., Soo, J. Y. H., et al. 2020, *MNRAS*, **499**, 1587
- 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* (Berlin: Springer), 128
- Soo, J. Y. H., Moraes, B., Joachimi, B., et al. 2018, *MNRAS*, **475**, 3613
- Soo, J. Y. H., Joachimi, B., Eriksen, M., et al. 2021, *MNRAS*, **503**, 4118
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, *J. Mach. Learn. Res.*, **15**, 1929
- Tallada, P., Carretero, J., Casals, J., et al. 2020, *Astron. Comput.*, **32**, 100391
- The Dark Energy Survey Collaboration 2005, *Archiv e-print: [astro-ph/0510346]*
- Tonello, N., Tallada, P., Serrano, S., et al. 2019, *Astron. Comput.*, **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, *Int. J. Pattern Recog. Artific. Intell.*, **23**, 687
- Zhang, Y., & Yang, Q. 2021, *IEEE Transac. Knowledge Data Eng.*, **1**
- ¹ Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
e-mail: lcabayol@pic.es
- ² Port d'Informació Científica, Campus UAB, C. Albareda 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), Carrer Gran Capitá 2-4, 08034 Barcelona, Spain
- ⁵ Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain
- ⁶ Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany
- ⁷ Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
- ⁸ Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
- ⁹ Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Lluís Companys 23, 08010 Barcelona, Spain
- ¹⁰ Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
- ¹¹ Institut de Ciències de l'Espai (IEEC-CSIC), Campus UAB, Carrer de Can Magrans, s/n Cerdanyola del Vallés, 08193 Barcelona, Spain
- ¹² Université Paris-Saclay, CNRS, Institut d'astrophysique spatiale, 91405 Orsay, France
- ¹³ Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
- ¹⁴ INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
- ¹⁵ Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
- ¹⁶ INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- ¹⁷ Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
- ¹⁸ Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
- ¹⁹ INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy
- ²⁰ Dipartimento di Fisica, Università degli studi di Genova, and INFN-Sezione di Genova, via Dodecaneso 33, 16146 Genova, Italy
- ²¹ INFN-Sezione di Roma Tre, Via della Vasca Navale 84, 00146 Roma, Italy
- ²² INAF-Osservatorio Astronomico di Capodimonte, Via Moiarriello 16, 80131 Napoli, Italy
- ²³ Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal
- ²⁴ Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
- ²⁵ INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
- ²⁶ INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
- ²⁷ INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
- ²⁸ INFN section of Naples, Via Cinthia 6, 80126 Napoli, Italy
- ²⁹ Department of Physics "E. Pancini", University Federico II, Via Cinthia 6, 80126 Napoli, Italy
- ³⁰ Dipartimento di Fisica e Astronomia "Augusto Righi" - Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- ³¹ INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy
- ³² Centre National d'Études Spatiales, Centre spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
- ³³ Institut national de physique nucléaire et de physique des particules, 3 rue Michel-Ange, 75794 Paris Cedex 16, France
- ³⁴ Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
- ³⁵ Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
- ³⁶ ESAC/ESA, Camino Bajo del Castillo, s/n., Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
- ³⁷ European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
- ³⁸ Univ. Lyon, Univ. Claude Bernard Lyon 1, CNRS/IN2P3, IP21 Lyon, UMR 5822, 69622 Villeurbanne, France
- ³⁹ Observatoire de Sauverny, École Polytechnique Fédérale de Lausanne, 1290 Versoix, Switzerland
- ⁴⁰ Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
- ⁴¹ Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisboa, Portugal
- ⁴² Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
- ⁴³ Department of Astronomy, University of Geneva, ch. d'Ecogia 16, 1290 Versoix, Switzerland
- ⁴⁴ Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK
- ⁴⁵ INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
- ⁴⁶ Université Paris-Saclay, Université Paris Cité, CEA, CNRS, Astrophysique, Instrumentation et Modélisation Paris-Saclay, 91191 Gif-sur-Yvette, France
- ⁴⁷ INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
- ⁴⁸ Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
- ⁴⁹ Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Via Irnerio 46, 40126 Bologna, Italy
- ⁵⁰ INAF-Osservatorio Astronomico di Padova, Via dell'Osservatorio 5, 35122 Padova, Italy
- ⁵¹ Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029 Blindern, 0315 Oslo, Norway
- ⁵² Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109, USA
- ⁵³ von Hoerner & Sulger GmbH, Schloßplatz 8, 68723 Schwetzingen, Germany

- ⁵⁴ Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
- ⁵⁵ Institut d’Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France
- ⁵⁶ Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
- ⁵⁷ Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, 00014 University of Helsinki, Finland
- ⁵⁸ NOVA optical infrared instrumentation group at ASTRON, Oude Hoogeveensedijk 4, 7991PD Dwingeloo, The Netherlands
- ⁵⁹ Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
- ⁶⁰ Dipartimento di Fisica e Astronomia “Augusto Righi” – Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, 40129 Bologna, Italy
- ⁶¹ Department of Physics, Institute for Computational Cosmology, Durham University, South Road, DH1 3LE, UK
- ⁶² Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
- ⁶³ INFN-Bologna, Via Irnerio 46, 40126 Bologna, Italy
- ⁶⁴ Institute of Physics, Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
- ⁶⁵ European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
- ⁶⁶ Department of Physics and Astronomy, University of Aarhus, Ny Munkegade 120, 8000 Aarhus C, Denmark
- ⁶⁷ Space Science Data Center, Italian Space Agency, via del Politecnico snc, 00133 Roma, Italy
- ⁶⁸ Institute of Space Science, Str. Atomîștilor, nr. 409 Măgurele, Ilfov, 077125, Romania
- ⁶⁹ Instituto de Astrofísica de Canarias, Calle Vía Láctea s/n, 38204, San Cristóbal de La Laguna, Tenerife, Spain
- ⁷⁰ Departamento de Astrofísica, Universidad de La Laguna, 38206, La Laguna, Tenerife, Spain
- ⁷¹ Dipartimento di Fisica e Astronomia “G.Galilei”, Università di Padova, Via Marzolo 8, 35131 Padova, Italy
- ⁷² Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
- ⁷³ Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, 30202 Cartagena, Spain
- ⁷⁴ Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, 1211 Genève 4, Switzerland
- ⁷⁵ Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- ⁷⁶ Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
- ⁷⁷ INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
- ⁷⁸ Junia, EPA department, 41 Bd Vauban, 59000 Lille, France

Appendix A: Self-organising maps

A 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 k th 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,

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

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 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

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

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: Redshift distributions, $N(z)$, and scatter plots

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. 2012; Hoyle et al. 2018). broadband photo- z s commonly suffer from biases due to degeneracies between colours and redshift, (e.g. Newman et al. 2015; Masters et al. 2017).

Figure B.1 shows $N(z)$ in tomographic redshift bins for $0 < z_t < 1.5$ 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- z s assigned to the bins.

Multi-task learning with photo- z data augmentation ($z_s + z_{\text{PAUS+NB}}$) always provides equal or more accurate $N(z)$ than

the baseline network (z_s , black line). As expected from Fig. 4, the $N(z)$ values exhibiting the largest bias are those with $z_t > 1.2$, particularly the bin at $z_t > 1.2$. In this bin, MTL together with the photo- z data augmentation ($z_s + \text{NB} + z_{\text{PAUS}}$, 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- z s 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 B.2 shows the density scatter between the predicted photo- z s and the true redshift. Here, we are plotting the complete COSMOS sample, and we therefore use a combination of spectroscopic redshift and PAUS+COSMOS photo- z s as true redshift. The top left panel corresponds to baseline network (z_s method) and clearly shows higher photo- z scatter and more outliers in the high-redshift region with respect to the other methods. The MTL method (top right, $z_s + \text{NB}$) already reduces photo- z scatter and number of high-redshift outliers. The methods including additional PAUS+COSMOS photo- z s in the training sample (bottom panels) further improve the photo- z performance.

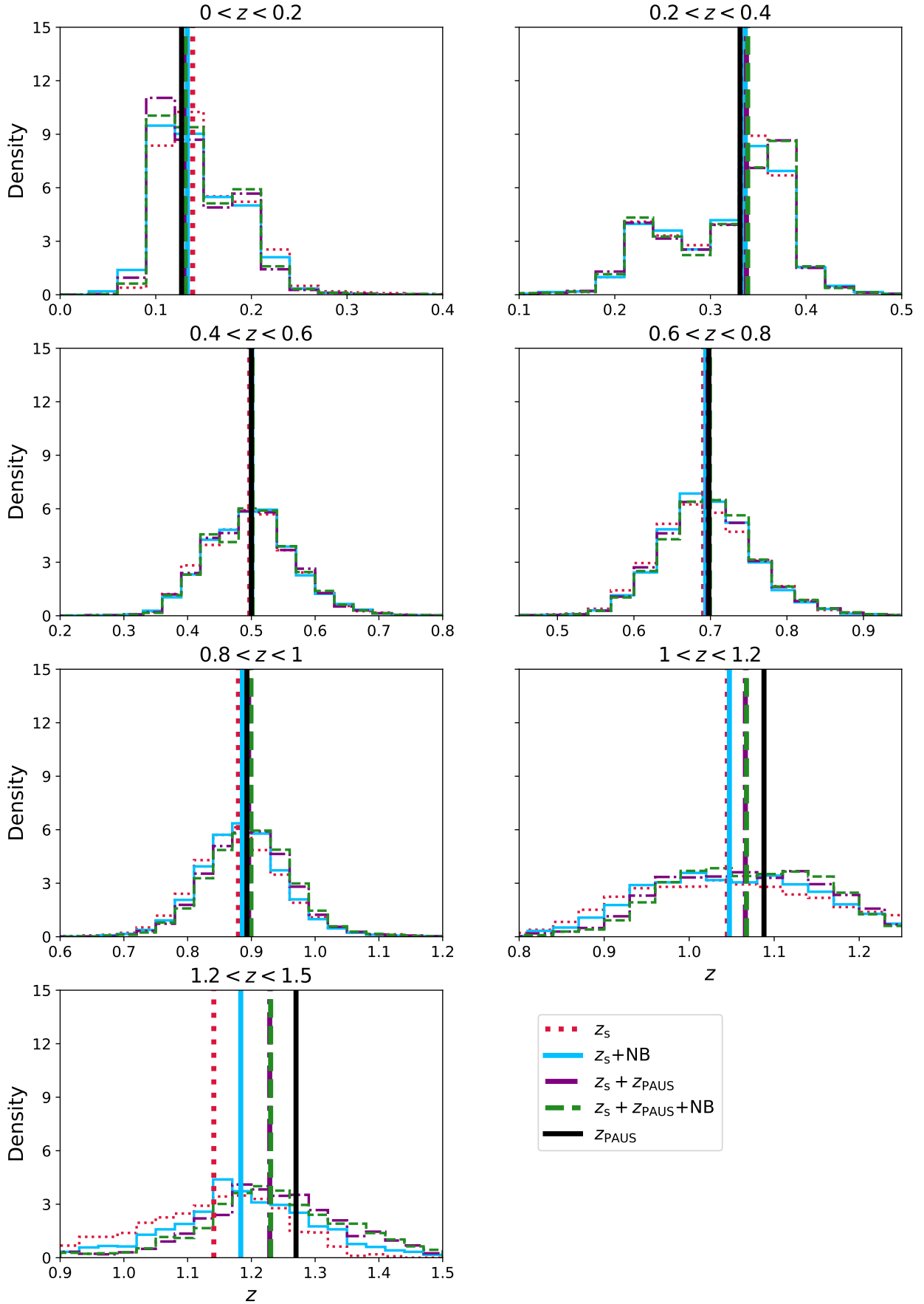


Fig. B.1. $N(z)$ estimates of the full COSMOS sample divided into seven tomographic bins over the redshift range $0 < z < 1.5$. Tomographic bins are defined using the spectroscopic redshifts and the PAUS+COSMOS high-precision photo- z s for galaxies without spectroscopy. The vertical solid black lines 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.

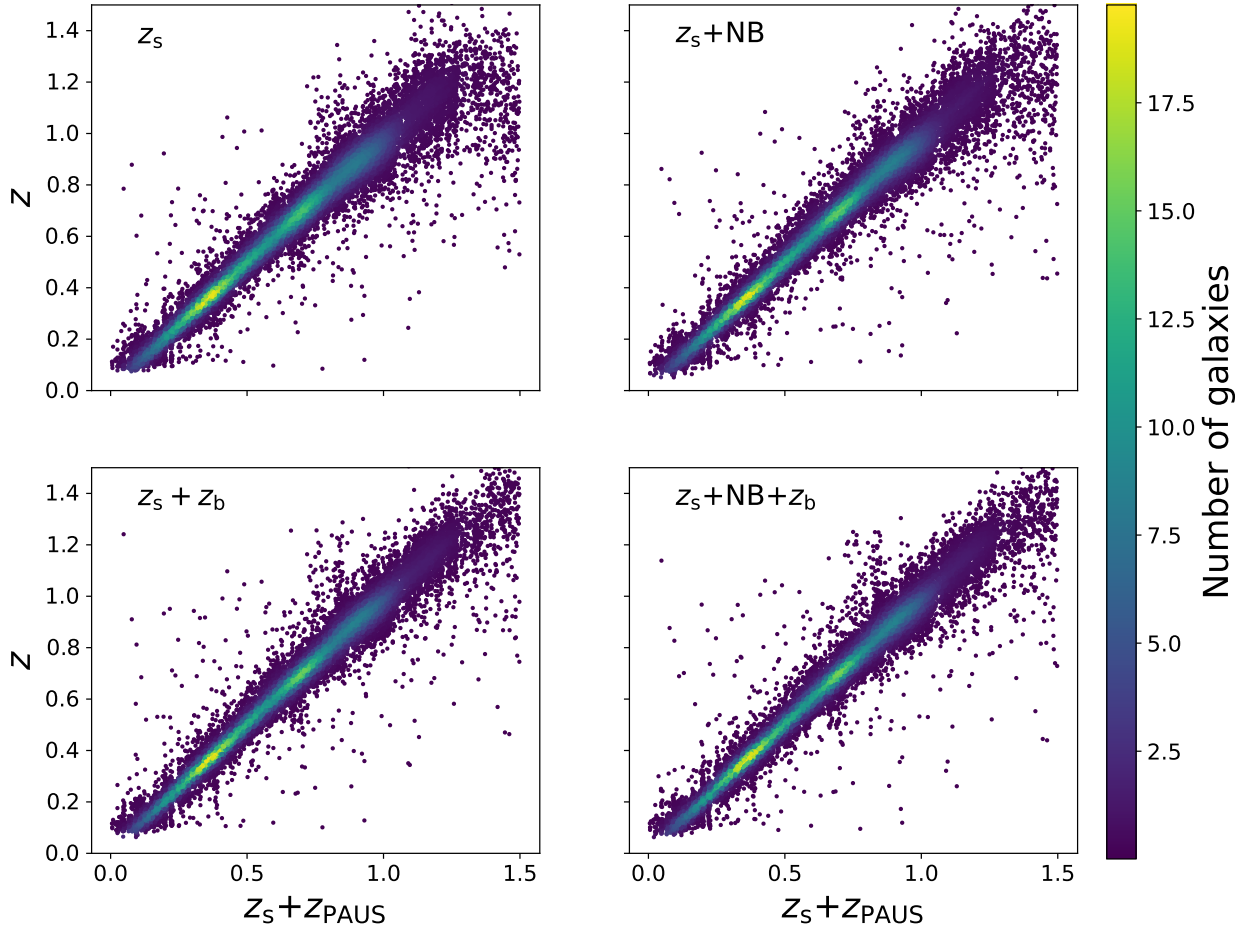


Fig. B.2. Scatter plot of the 1:1 relation between the predicted photo- z and the true redshift, which is a combination of spectroscopic redshift and PAUS+COSMOS photo- z s, in the complete COSMOS sample for the four methods in Sect. 3.

Appendix C: Effect of training with photo- z s 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 $z_s + z_{PAUS}$ and $NB + z_s + z_{PAUS}$ in Sect. 3.2). Even if such photo- z s 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. C.1) and the presence of outliers (Sect. C.2) have on the photo- z performance.

Appendix C.1: Effect of the dispersion in the ground-truth photo- z

Figure C.1 shows the photo- z precision of a set of 1000 spectroscopic galaxies for four independent broadband 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 Sect. 2.4), the PAUS+COSMOS photo- z s (see Sect. 2.2), the COSMOS30 photo- z s (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- z s 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 broadband filters from COSMOS2015.

The red points in Fig. C.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. C.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, and the CFHT photo- z s 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- z s, 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.

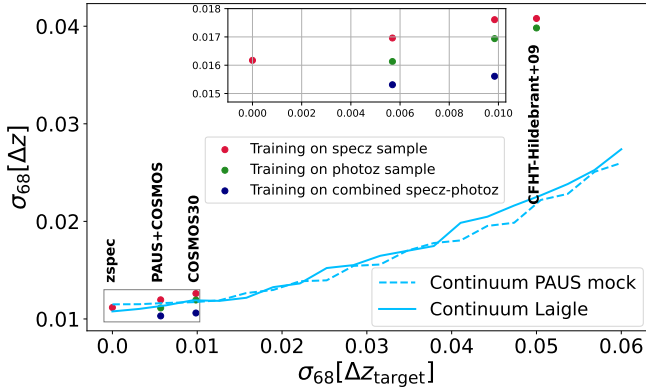


Fig. C.1. Photo- z performance as a function of the ground-truth redshift precision used for training the networks. The training redshifts are the spectroscopic redshifts, the PAUS+COSMOS photo- z s, COSMOS30, and a set of CFHT photo- z s in COSMOS. Red points correspond to training on the spectroscopic sample (around 6000 galaxies). The green and red points show the training sample extended to COSMOS galaxies with photo- z s (around 15 000 galaxies). The blue lines show the expected photo- z performance as a function of target redshift precision. The true redshifts, the spectroscopic redshift in the COSMOS2015 catalogue (solid blue line), and the simulated redshift in the PAUS mock (blue dashed line) are scattered with precision in 0.001 bins. The top inset zooms into the framed area in the main plot (lower-left corner)

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- z s as ground-truth training redshifts. Combining spectroscopic redshifts with PAUS+COSMOS photo- z s yields $\sigma_{68} = 0.015$, which improves upon the precision obtained with spectroscopic redshifts only.

The light blue lines in Fig. C.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- z s. This is potentially triggered by systematic errors or outliers in the CFHT photo- z not represented in the blue curves.

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

In Fig. C.1 showed a mismatch between the expected (solid blue curve) and photo- z performance training a network with the CFHT photo- z s as ground-truth redshifts (rightmost red point). However, the expectation assumes that the CFHT photo- z s are not affected by other effects such as systematic errors or catastrophic outliers.

Figure C.2 shows the effect of outliers in the ground-truth targets of the training sample. The network is trained 20 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- z s (red line), the COSMOS30 photo- z s (blue line), and the CFHT photo- z s (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, and 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- z s 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 H α in the training redshifts:

$$z_{\text{mod}} = \lambda_{\text{O III}} / \lambda_{\text{H}\alpha} (1 + z_t) - 1, \quad (\text{C.1})$$

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 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. C.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 C.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. C.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

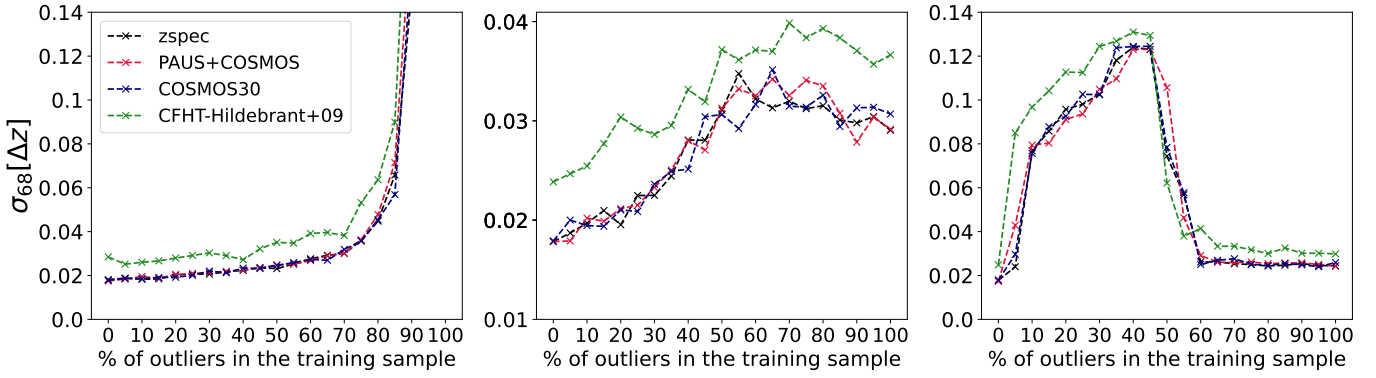


Fig. C.2. Effect of outliers and systematic errors on the ground-truth redshift sample used during training. The training sample consists of 5000 spectroscopic galaxies with photometry from COSMOS2015. Each coloured line uses a different sample of redshifts as true redshifts, i.e. spectroscopic redshifts (black), PAUS+COSMOS photo- z s (red), COSMOS30 photo- z s (blue), and CFHT photo- z s (green). The ground-truth redshift of the selected fraction of training galaxies is replaced by a random redshift value sampled from $U(0, 1.5)$ (left), a 20% higher redshift (centre), and redshifts modified with Eq. (C.1) (right).

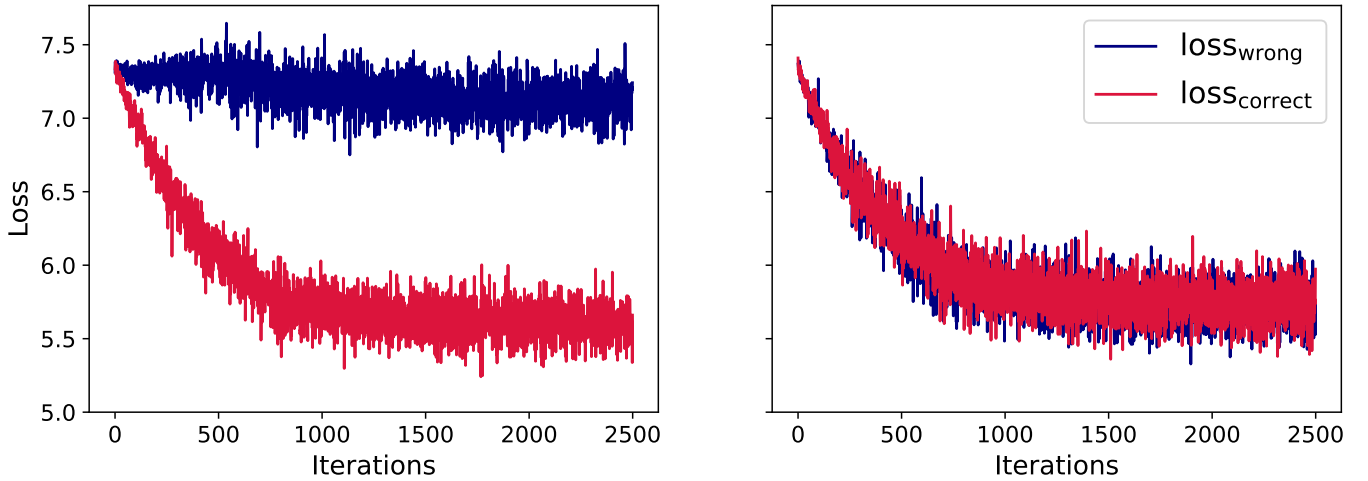


Fig. C.3. Training loss function for galaxies with a wrong (blue) and a corrected (red) target redshift. The training sample consists of 5000 spectroscopic galaxies with photometry from COSMOS2015. In the left panel, the modified target redshifts are randomly switched to a value drawn from $U(0, 1.5)$,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

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. C.3 shows the loss function for the correct and the wrongly labelled training galaxies separately when the incorrect redshift labels are generated with Eq. (C.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 $p(z)$ behaviour, which presents a double-peaked distribution (not shown). Hence, Figs. C.2 and C.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.

Appendix D: Robustness of the methods to outliers in the target redshifts

In this appendix we study the robustness of our training methodologies to outliers in the redshifts used as ground-truth to train the network. In Sect. 5, in order to simulate the PAUS+COSMOS photo- z s used to train the z_s+z_{PAUS} and the $z_s+z_{\text{PAUS}}+\text{NB}$ meth-

ods, we scatter the true redshift from the simulations to a similar precision of PAUS+COSMOS photometric redshifts. This process assumes that photometric redshift errors are purely Gaussian; however, real data also have non-Gaussian errors and photo- z outliers, which can be caused by, for example, noisy photometry, emission-line confusions, and other artefacts in the data.

Systematic outliers in the ground-truth target redshifts have a much stronger impact than random catastrophic outliers (Figs. C.3 and C.2). Therefore, we assume the most adverse scenario where all outliers are systematically shifted according to Eq. (C.1).

Figure D.1 shows the impact that outliers in the target redshifts have in the performance of the methods. We studied four different cases: a sample without outliers in the training sample (black solid line), with a 5% of outliers in the spectroscopic redshifts and 1% of outliers in the photo- z s (red dashed line), 5% of spectroscopic redshift outliers and 5% photo- z s outliers (blue dashed line), and 5% of spectroscopic redshift outliers and 10% photo- z s outliers (green dashed line).

Adding the 5% of spectroscopic redshift outliers already has an impact on the predicted photo- z bias of the baseline method (z_s , top left panel). The effect of spectroscopic redshift outliers

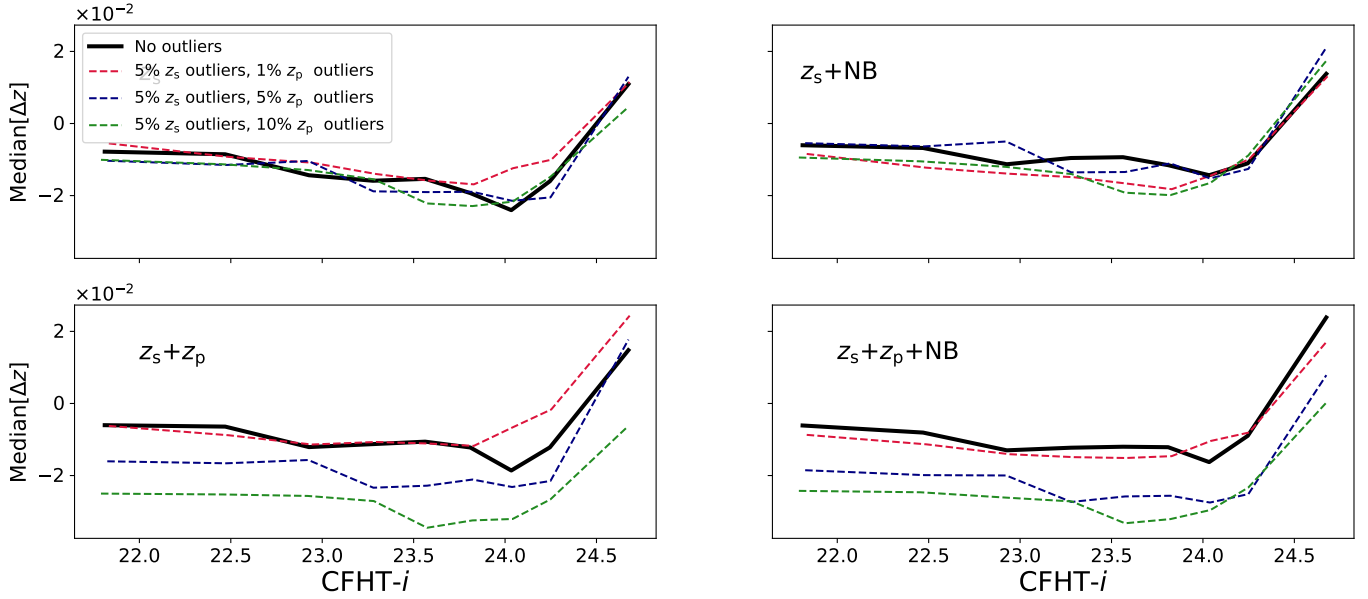


Fig. D.1. Effect on the photo- z predictions of different outlier rates in the target spectroscopic and high-precision photometric redshifts used as ground-truth targets to train the methods in Sect. 3. In all cases, outliers have been included following Eq.(C.1).

in the training sample is mitigated by the MTL network (top right panel), which is not affected by training-photo- z outliers since these are not used during the training (method $z_s + \text{NB}$, Sect. 3.2). We also observe that training with photo- z samples with up to $\sim 5\%$ of outliers also mitigates the effect of outliers in the spectroscopic sample. This is expected since adding more training data reduces the relative importance of an outlier in the training sample. However, the bottom plots also show that training samples with more than 10% of systematic photo- z outliers degrade the photo- z performance.

We also studied the robustness of the dispersion and the outlier rate to the target-redshift outliers and these two metrics are much less affected by the presence of target-redshift outliers.

Appendix E: Further studies of multi-task training

In this section we aim to give a more technical view of the network functioning. In Fig. E.1 we show the evolution of the photo- z prediction loss function (Eq. 1) with time for the baseline method (z_s , black line) and the MTL method ($z_s + \text{NB}$, red line). In both cases, the solid line corresponds to the training loss, while the dashed line is the validation loss. The networks have been trained for 100 epochs with an initial learning rate of 10^{-3} , which decreases to 10^{-4} after 50 epochs. In this test, unlike previous tests, we initialise the two networks with the same weights. Still we observe training with the two different losses leads to a lower photo- z loss (\mathcal{L}_z). We also find adding the narrow-band loss stabilises the photo- z loss in the validation sample, meaning the network better generalise with the additional narrow-band loss.

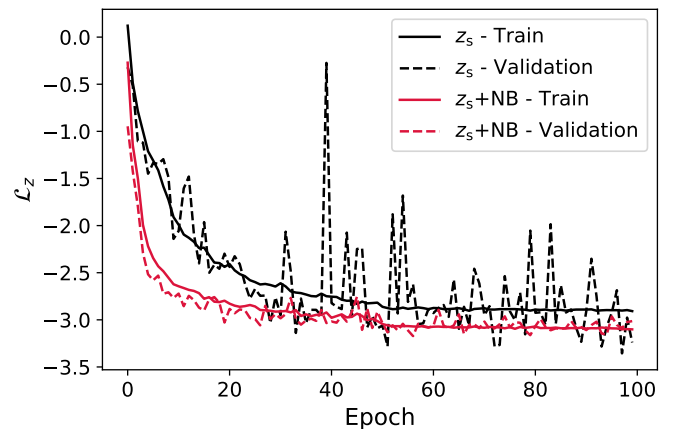


Fig. E.1. Training (solid lines) and validation (dashed lines) loss for the z_s (black) and $z_s + \text{NB}$ (red) methods. All methods are trained for 100 epochs with an initial learning rate of 10^{-3} and the same initial conditions.