Extracting photometric redshift from galaxy flux and image data using neural networks in the CSST survey

Xingchen Zhou,[1,2] Yan Gong [ORCID],[1,3]★ Xian-Min Meng,[1] Ye Cao,[1,2] Xuelei Chen [ORCID],[2,4,5] Zhu Chen [ORCID],[6] Wei Du,[6] Liping Fu [ORCID][6] and Zhijian Luo[6]

[1]*Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China*

[2]*University of Chinese Academy of Sciences, NAOC, Beijing 100049, China*

[3]*Science Center for China Space Station Telescope, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China*

[4]*Key Laboratory for Computational Astrophysics, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China*

[5]*Centre for High Energy Physics, Peking University, Beijing 100871, China*

[6]*Shanghai Key Lab for Astrophysics, Shanghai Normal University, Shanghai 200234, China*

ABSTRACT

The accuracy of galaxy photometric redshift (photo-z) can significantly affect the analysis of weak gravitational lensing measurements, especially for future high-precision surveys. In this work, we try to extract photo-z information from both galaxy flux and image data expected to be obtained by China Space Station Telescope (CSST) using neural networks. We generate mock galaxy images based on the observational images from the Advanced Camera for Surveys of *Hubble Space Telescope* (*HST*-ACS) and COSMOS catalogues, considering the CSST instrumental effects. Galaxy flux data are then measured directly from these images by aperture photometry. The multilayer perceptron (MLP) and convolutional neural network (CNN) are constructed to predict photo-z from fluxes and images, respectively. We also propose to use an efficient hybrid network, which combines the MLP and CNN, by employing the transfer learning techniques to investigate the improvement of the result with both flux and image data included. We find that the photo-z accuracy and outlier fraction can achieve $\sigma_{\mathrm{NMAD}} = 0.023$ and $\eta = 1.43$ per cent for the MLP using flux data only, and $\sigma_{\mathrm{NMAD}} = 0.025$ and $\eta = 1.21$ per cent for the CNN using image data only. The result can be further improved in high efficiency as $\sigma_{\mathrm{NMAD}} = 0.020$ and $\eta = 0.90$ per cent for the hybrid transfer network. These approaches result in similar galaxy median and mean redshifts 0.8 and 0.9, respectively, for the redshift range from 0 to 4. This indicates that our networks can effectively and properly extract photo-z information from the CSST galaxy flux and image data.

Key words: methods: statistical – techniques: image processing – techniques: photometric – galaxies: distances and redshifts – galaxies: photometry – large-scale structure of Structure.

1 INTRODUCTION

Galaxy photometric surveys provide an efficient tool to explore the evolution of the Universe and the properties of galaxies. By taking galaxy images in a few photometric bands and deriving the corresponding fluxes, numbers of important information can be analysed and obtained, such as magnitude, colour, morphology, type, size, etc. Then several powerful cosmological probes, e.g. weak gravitational lensing and angular galaxy clustering measurements, can be fulfilled in the cosmological studies. As we know, there are a number of ongoing and next-generation photometric surveys are running and will be started up in near future, e.g. the Sloan Digital Sky Survey[1] (Fukugita et al. 1996; York et al. 2000), Dark Energy

Survey[2] (D. E. S. Collaboration 2016; Abbott et al. 2021), the Legacy Survey of Space and Time (LSST) or Vera C. Rubin Observatory[3] (LSST Science Collaboration 2009; Ivezić et al. 2019), the Euclid Space Telescope[4] (Laureijs et al. 2011), and the Wide-Field Infrared Survey Telescope or Nancy Grace Roman Space Telescope[5] (Green et al. 2012; Akeson et al. 2019). These surveys are expected to observe huge amount of galaxies in wide and deep fields, and can provide extremely precise measurements on the cosmic large-scale structure (LSS), properties of dark energy and dark matter, galaxy formation and evolution, and so on.

The photometric redshift (photo-z) is a key quantity in photometric surveys. In cosmological studies, accurate photo-z measurements

★ E-mail: gongyan@bao.ac.cn

[1]http://www.sdss.org/

[2]https://www.darkenergysurvey.org/

[3]https://www.lsst.org/

[4]https://www.euclid-ec.org/

[5]https://www.stsci.edu/roman

can provide reliable galaxy location information along the line of sight, which is crucial for probing the LSS from the photometric surveys (Zhan et al. 2006; Banerji et al. 2008). Besides, the photo-*z* accuracy is one of the main systematics in weak lensing surveys, and accurate measurements of photo-*z* and its variance can effectively suppress the systematics and precisely extract the information of kinetic and dynamic evolution of the Universe (Ma, Hu & Huterer 2006; Abdalla et al. 2008; Mandelbaum et al. 2008; Hearin et al. 2010). Currently, there are two main methods used to derive galaxy photo-*z* from photometric data. One is called template fitting method, which uses typical galaxy spectral energy distributions (SEDs) to fit photometric data and capture spectral features in multibands for deriving photo-*z* (Lanzetta, Yahil & Fernández-Soto 1996; Fernández-Soto, Lanzetta & Yahil 1999; Bolzonella, Miralles & Pelló 2000). Another method can be named as training method, and it will extract photo-*z* by developing empirical relations between redshift and different galaxy properties, such as magnitude, colour, and morphology (Collister & Lahav 2004; Sadeh, Abdalla & Lahav 2016; Brescia et al. 2021). This method always can be fulfilled by neural networks using galaxy spectral data with measured spectroscopic redshift as training sample. Both methods have their own advantages. The template fitting method can be widely used at all redshifts if the selected galaxy SED templates are representative enough for the whole sample. On the other hand, the training method can obtain more accurate photo-*z* result if the training sample contains sufficiently large and high-quality spectroscopic galaxy data, which can cover the whole redshifts and all features of galaxies in the photometric survey.

Although obtaining qualified training sample would be a challenge for future next-generation photometric surveys that observing deep fields with large redshift coverage, fortunately, a number of high-quality spectroscopic galaxy surveys are currently running or planned, such as Dark Energy Spectroscopic Instrument[6] (Levi et al. 2019), Prime Focus Spectrograph (PFS; Tamura & PFS Collaboration 2016), Multi-Object Optical and Near-infrared Spectrograph[7] (Cirasuolo et al. 2020; Maiolino et al. 2020), 4-metre Multi-Object Spectroscopic Telescope[8] (de Jong et al. 2019), MegaMapper (Schlegel, Kollmeier & Ferraro 2019), Fiber-Optic Broadband Optical Spectrograph (Bundy et al. 2019), and SpecTel (Ellis & Dawson 2019). On the other hand, recently the machine learning algorithms, especially neural networks, are remarkably developed. Neural networks are currently widely involved in astronomical and cosmological studies, and numbers of network forms are proposed to solve different problems. For example, there are two widely used networks, i.e. multilayer perceptron (MLP) and convolutional neural network (CNN). The MLP is the simplest form of neural network, constructed by an input layer, several hidden layers, and an output layer (Haykin 1994), and the CNN proposed by Fukushima & Miyake (1982) and Lecun et al. (1998), can extract features from images by learning kernel arrays, and gain great success in detection and recognition tasks.

In this work, we focus on the training method by employing neural networks to explore photo-*z* accuracy in the photometric surveys of the China Space Station Telescope (CSST). The CSST is a next-generation 2-m space telescope, which is planned to launch around 2024 (Zhan 2011, 2018, 2021; Cao et al. 2018; Gong et al. 2019). It will be in the same orbit with the China Manned Space
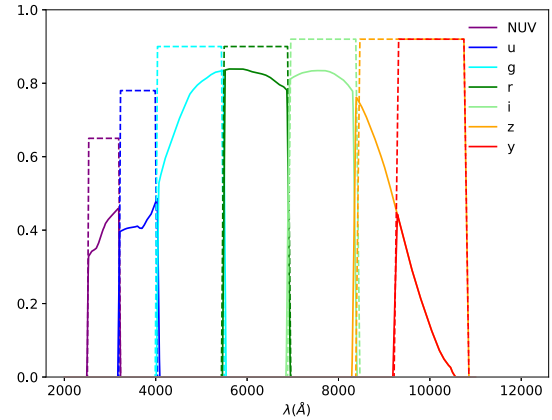
Figure 1. The intrinsic (dashed curves) and real (solid curves) transmissions of the CSST seven photometric filters. The real transmission curves include the effect of detector quantum efficiency. The details of the transmission parameters can be found in Cao et al. (2018) and Meng et al. (in preparation).

Station for maintenance and repair. It has seven photometric bands from near-ultraviolet (NUV) to near-infrared (NIR), i.e. *NUV*, *u*, *g*, *r*, *i*, *z*, *y*, covering wavelength range from ∼250 to ∼1000 nm, with 5*σ* magnitude limits as 25.4, 25.4, 26.3, 26.0, 25.9, 25.2, and 24.4, respectively. The intrinsic transmissions and real transmissions including detector quantum efficiency of the CSST seven photometric filters are illustrated in Fig. 1. The details of the transmission parameters can be found in Cao et al. (2018) and Meng et al. (in preparation). Totally 17 500 deg² sky area will be observed during its 10-yr mission period, and photometric and spectroscopic surveys will be simultaneously performed. The CSST will answer questions like the properties of dark matter and dark energy, the evolution of the LSS, formation of galaxies, and other important problems in cosmological and astrophysical studies. Many CSST scientific goals and surveys are related to photo-*z* measurements, e.g. weak gravitational lensing survey, which is the main CSST scientific driver.

In previous studies, the CSST photo-*z* accuracy has been preliminarily investigated by employing the SED template fitting and MLP neural network methods using the mock flux data directly generated from galaxy SED templates (Cao et al. 2018; Zhou et al. 2021). Although relevant background and instrumental noises are properly considered, their results are still ideal, which give photo-*z* accuracy ∼0.02–0.03 and outlier fraction ∼ 3 per cent for the SED fitting method, and ∼0.01–0.02 and ∼ 0.1 per cent for the neural network. In this work, we will simulate CSST mock galaxy images in the photometric surveys, and measure galaxy flux from these images with aperture photometry technique. Besides, since galaxy morphology should also contain photo-*z* information (Wadadekar 2005; Soo et al. 2018; Wilson et al. 2020), we try to estimate photo-*z* using galaxy images. The MLP and CNN are adopted to extract photo-*z* from the CSST galaxy flux and image data, respectively. We also propose a hybrid network using the transfer learning technique to efficiently derive photo-*z* information from both flux and image data by effectively combing the MLP and CNN.

This paper is organized as follows: in Section 2, we explain the details of generation of the mock image and flux data. In Section 3, we present architectures of the neural networks we use and details of training process. The results are shown in Section 4. We summarize our results in Section 5.

---

2 MOCK DATA

In order to simulate galaxy images of the CSST photometric survey as real as possible, we generate mock images based on the observations in the COSMOS field performed by the Advanced Camera for Surveys of *Hubble Space Telescope* (*HST*-ACS). The image simulation process will be explained in more details by Meng et al. (in preparation). This survey covers about 1.7 deg² in F814W band, which has similar spatial resolution as the CSST with 80 per cent energy concentration radius $R_{80} \simeq 0.15$ arcsec (Koekemoer et al. 2007; Massey et al. 2010; Bohlin 2016; Cao et al. 2018; Gong et al. 2019; Zhan 2021). The background noise of the COSMOS *HST*-ACS F814W survey is also quite low (expected to be ∼1/3 of the CSST survey), hence it can provide a good foundation for simulating the CSST galaxy images.

Note that, we only use the measurements from one *HST* band, i.e. F814W, to generate mock galaxy images without colour gradient considered. In real surveys, galaxy morphology will look different in different photometric bands. This would result in wavelength-dependent features and provide more information about galaxy properties, which can be helpful in the neural network training process and improve the accuracy of photo-*z* estimate. In addition, base on the simulation, the CSST point spread function (PSF) should be more regular or Gaussian-like than the PSF of *HST*. So the image distortion should be more easily corrected in CSST data processing. However, since the CSST field of view is much larger than the *HST* (about 300× larger), precisely understanding and measuring the variances of the PSFs at different locations of the focal plane would be a challenging task, which may affect the image calibration. In this work, for simplicity, we ignore the colour gradient and PSF effects on galaxy images.

To obtain high-quality images as 'ideal' images for following mock image production, we select the central 0.85 × 0.85 deg² area of this survey including ∼192 000 galaxies. Then, we rescale the pixel size of the COSMOS *HST*-ACS F814W survey from 0.03 arcsec to 0.075 for matching the CSST pixel size. Next, we extract a square stamp image for each galaxy from the survey area, with a galaxy in the centre of the stamp image. The size of the square stamp is 15 times of the galaxy's semimajor axis, which means the galaxy stamp images have different sizes. The galaxy's semimajor axis and other relevant galaxy morphological information can be obtained in the COSMOS weak lensing source catalogue (Leauthaud et al. 2007). In addition, we also mask all sources, with signal-to-noise ratio (SNR) >3σ, in the stamp image except for the galaxy in the centre, and replace them by the CSST background noise as we show later, which can be helpful for the neural networks to extract galaxy information.

By making use of galaxy SEDs, we can rescale the galaxy images of the COSMOS *HST*-ACS F814W survey to the CSST flux level. We first need to find the corresponding SEDs for galaxies by fitting their flux data. Here, the COSMOS2015 catalogue[9] is adopted to match galaxies in *HST*-ACS F814W survey, which contains 220 000 galaxies with measures of galaxy redshift, magnitude, size, and so on (Laigle et al. 2016). To avoid additional errors, we use the 31 SED templates from LEPHARE code (Arnouts et al. 1999; Ilbert et al. 2006), which are also used in the COSMOS2015 catalogue, to fit the galaxy flux data. Since the CSST has large wavelength coverage from NUV to NIR, for galaxies detected at $z \gtrsim 2$, we also extend

the wavelength range of these templates from ∼900 to ∼90 Å using the BC03 method (Bruzual & Charlot 2003). The details can be found in Cao et al. (2018). We select high-quality photometric data with reliable photo-*z* measurement and about 100 000 galaxies are selected. Besides, when fitting the flux data with the SED templates and finding the best-fitting SEDs, we also include dust extinction and emission lines, such as Ly α, H α, H β, [O II], and [O III].

After obtaining the best-fitting SED for each galaxy, we can calculate the theoretical flux data observed by the CSST. The theoretical flux in $e^- s^{-1}$ or electron counting rate for a band j can be estimated as

$$C_{\text{th}}^j = A_{\text{eff}}^j \int S(\lambda) \tau_j(\lambda) \frac{\lambda}{hc} d\lambda. \tag{1}$$

Here, A_{eff}^j is the effective telescope aperture area, h and c are the Planck constant and speed of light, $S(\lambda)$ is the galaxy SED, and $\tau_j(\lambda) = T_j(\lambda)Q_j(\lambda)M_j(\lambda)$ is total system throughput, where $T_j(\lambda)$, $Q_j(\lambda)$, and $M_j(\lambda)$ are the intrinsic filter transmission, detector quantum efficiency, and total mirror efficiency, respectively (Cao et al. 2018; Zhou et al. 2021). Then, the theoretical flux electron counts in j band can be expressed as

$$F_{\text{th}}^j = C_{\text{th}}^j t_{\text{exp}} N_{\text{exp}}^j, \tag{2}$$

where $t_{\text{exp}} = 150$ s is the exposure time, and N_{exp}^j is the number of exposure that $N_{\text{exp}}^j = 2$ for *u*, *g*, *r*, *i*, and *z* bands, and 4 for *NUV* and *y* bands. For the measured galaxy flux F_{obs}^j in the *HST*-ACS F814W survey, we collect the flux of galaxy in the centre of stamp image with photometry aperture size of 2 times of the Kron radius (Kron 1980), i.e. $2 \times R_{\text{Kron}}$. Then, we can rescale the measured galaxy flux to the expected flux observed by the CSST, and produce the CSST mock galaxy images in a band. Note that, in the meantime, we also rescale the background of the stamp image by the same factor, i.e. rescaling the whole stamp image including both galaxy in the centre and background, which can be used in the following background correction.

Besides rescaling galaxy flux, we also adjust the background noise in the stamp image to the CSST level. Since the background noise in the *HST*-ACS F814W survey is expected to be much lower than the CSST survey, we need to add additional noise into the image in j band, and we have

$$N_{\text{add}}^j = \sqrt{\left(N_{\text{bkg,th}}^j\right)^2 - \left(N_{\text{img}}^j\right)^2}, \tag{3}$$

where N_{img}^j is the background noise of the rescaled image discussed above, and $N_{\text{bkg,th}}^j$ is the theoretical CSST background noise per pixel, which can be estimated by

$$N_{\text{bkg,th}}^j = \sqrt{(B_{\text{sky}}^j + B_{\text{dark}})t_{\text{exp}}N_{\text{exp}}^j + R_{\text{n}}^2 N_{\text{exp}}^j}. \tag{4}$$

Here, $B_{\text{dark}} = 0.02 \ e^- s^{-1} \text{pix}^{-1}$ is the dark current, $R_{\text{n}} = 5 \ e^- \text{pix}^{-1}$ is the read-out noise, and B_{sky}^j is the sky background in unit of $e^- s^{-1} \text{pix}^{-1}$, which is given by

$$B_{\text{sky}}^j = A_{\text{eff}}^j l_{\text{pix}}^2 \int I_{\text{sky}}(\lambda) \tau_j(\lambda) \frac{\lambda}{hc} d\lambda, \tag{5}$$

where l_{pix} is the pixel size in arcsec, and $I_{\text{sky}}(\lambda)$ is the surface brightness intensity of the sky background in units of erg s⁻¹cm⁻² Å⁻¹ arcsec⁻². We evaluate B_{sky}^j based on the measurements of the earthshine and zodiacal light for the 'average' sky background case given in Ubeda (2011). We find that B_{sky}^j are 0.0023, 0.018, 0.142, 0.187, 0.187, 0.118, and 0.035 e⁻s⁻¹pix⁻¹ for *NUV*, *u*, *g*, *r*, *i*, *z*, and

---

[9]Note that there are still some issues in the COSMOS2015 catalogue, and the COSMOS2020 catalogue can make relevant improvements and could be used in future analysis (Weaver et al. 2022).
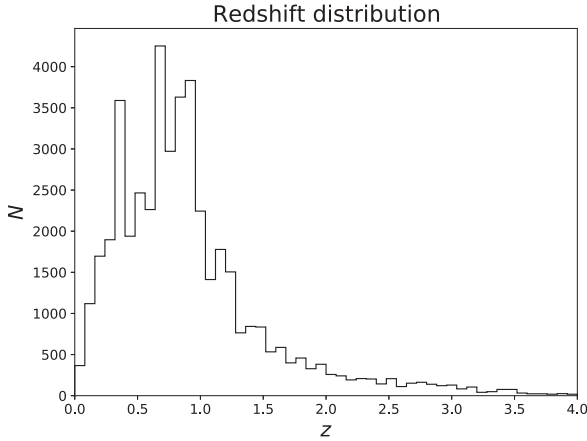
Figure 2. Galaxy redshift distribution of the sample selected from the COSMOS catalogue used in the neural networks. These sources have been selected with the SNR greater than 10 in the *g* or *i* bands. The distribution has a peak around $z = 0.7$, and can extend to $z \sim 4$, which is the same as the CSST photometric galaxy sample.

y bands, respectively, which are consistent with the results in Cao et al. ([2018](#)). Then, $N_{\mathrm{bkg,th}}^{j}$ are calculated as 10.65, 7.84, 9.93, 10.59, 10.59, 9.56, and 11.53 e^{-} for these bands. Thus, the additional noise N_{add}^{j} can be calculated by subtracting the rescaled image noise N_{img}^{j} as shown in equation ([3](#)), and added to pixels in stamp images by sampling from a Gaussian distribution with mean=0 and $\sigma = N_{\mathrm{add}}^{j}$ in band *j*. Then, we obtain the final mock CSST galaxy images for each band.

Next, we measure galaxy flux data from the CSST mock galaxy images using aperture photometry method. First, to obtain high-SNR source detections and morphological measurements, we stack the *g*, *r*, *i*, and *z* band images for each galaxy to create the detection images. Secondly, we measure the Kron radius along galaxy major- and minor-axis to find an elliptical aperture with the size $1 \times R_{\mathrm{Kron}}$, which can improve the SNR. Finally, the flux and error in *j* band can be calculated from electron number $N_{\mathrm{e^-}}^{j}$ and error $\sigma_{\mathrm{e^-}}^{j}$ measured within the aperture. Note that if a galaxy in a band is very faint, the measured flux could be negative due to background noise. This effect will not affect the training process of our neural networks, and as we show in Section [3](#), we will rescale it and reserve the information. The SNR of measured galaxies in band *j* can be estimated as

$$\mathrm{SNR}_j = \frac{F_{\mathrm{obs}}^{j}}{\sqrt{F_{\mathrm{obs}}^{j} + N_{\mathrm{pix}}(B_{\mathrm{sky}} + B_{\mathrm{dark}})t_{\mathrm{exp}}N_{\mathrm{exp}}^{j} + R_{\mathrm{n}}^{2}N_{\mathrm{exp}}^{j}N_{\mathrm{pix}}}}, \quad (6)$$

where F_{obs}^{j} is the observed electron counts, and N_{pix} is the number of pixels covered by a galaxy for the CSST, which can be derived from the COSMOS2015 catalogue.

In Fig. [2](#), we show the redshift distribution of the galaxy sample selected from the COSMOS catalogue (the selection details can be found in the next section). Since the selected galaxy sample has high-quality photo-*z* measurement, we assume that these photo-*z*s can be seen as spectroscopic redshifts (spec-*z*s), which can be used in the neural networks for method validation purpose. We can see that the distribution has a peak around $z = 0.6$–0.7, and can extend to $z \sim 4$, which is consistent with the CSST galaxy redshift distribution in previous studies (Cao et al. [2018](#); Gong et al. [2019](#); Zhou et al. [2021](#)). In Fig. [3](#), the CSST mock galaxy stamp images at different redshifts have been shown. We can find that the low-*z* galaxies have higher SNR with low backgrounds, while high-*z* galaxies can be

dominated by the background noise, especially in the bands with low transmissions. This indicates that it may be necessary to employ the machine learning technique to extract information from these bands. The corresponding measured galaxy fluxes from mock images by aperture photometry in the seven CSST bands are shown in Fig. [4](#). The SEDs are rescaled to the levels of flux data as comparison. As we show, our measured mock flux data match the corresponding SEDs very well, and can correctly represent the features.

3 NEURAL NETWORKS

The MLP and CNN are used to derive the photo-*z* from the mock flux and image data obtained in the last section, respectively. We also explore the effective way of combining these two networks for testing the improvement of the photo-*z* accuracy by including both flux and image data. The details of architectures and training processes of these networks are discussed in this section. All networks we construct are implemented by KERAS[10] with TENSORFLOW[11] as backend.

3.1 Network architecture

The flux data of a galaxy observed by the CSST would contain seven discrete data points in the seven CSST photometric bands, and we adopt the MLP to predict photo-*z* from these data. The MLP is the simplest form of neural networks, consisting of an input layer, several hidden layers, and an output layer. Every layer has several units or perceptrons, and the hidden layers can learn the internal relationship between input data points. Layers are connected by weights and biases, which can be initialized by random distributions and optimized in the training process.

We use the measured band flux, colour, and error as the inputs of the MLP (Zhou et al. [2021](#)). In order to speed up the training process, we need to rescale these data and make the network effectively extract the relevant information. A useful way we find for rescaling is to take the logarithm of these data. However, as we discussed in the last section, since the flux and error are measured by the aperture photometry and because of the fluctuations of noises, we may encounter some negative fluxes, especially in the CSST *NUV*, *u*, and *y* bands with relatively low transmissions. To fix this issue, we propose to use the following function

$$f(x) = \begin{cases} \log(x) & x > 0, \\ -\log(-x) & x < 0. \end{cases} \quad (7)$$

This function can naturally convert a negative value to be a logarithmic one, and could effectively reserve its information which can be useful to derive photo-*z* as we find.[12] Before rescaling, we perform normalization-like process for the fluxes and errors to speed up the training process of the network. For the flux data, we divide it by a fixed value that is close to the measured flux. Here, we take this fixed value as the flux-limit for each CSST band, which is derived from the magnitude limit shown in Section [1](#). Note that our result is not sensitive to the selection of this fixed value in each band. For the error, we divide it by its corresponding flux to obtain a relative error. To include more information, we also construct a colour-like

---

[10]https://keras.io

[11]https://www.tensorflow.org

[12]We also try to set the negative flux to be zero or use its measured upper-limit information, but we find that the negative flux also contains useful information when deriving photo-*z*, and we should reserve it in the rescaling process.
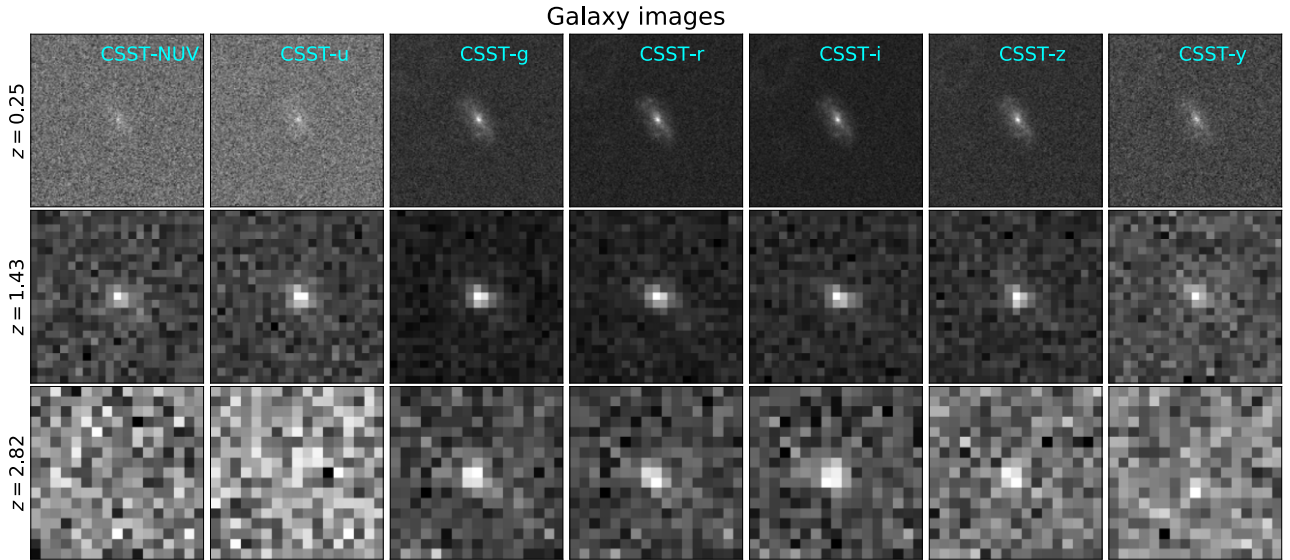
Galaxy images

Figure 3. Examples of simulated galaxy images in the seven CSST photometric bands. We notice that noises on the *NUV*, *u*, and *y* bands are larger than others, since these bands have lower transmissions. Many sources, especially at high redshifts, are overwhelmed by the background noises in some bands, which may indicate that neural networks are necessarily needed to extract information from these sources.
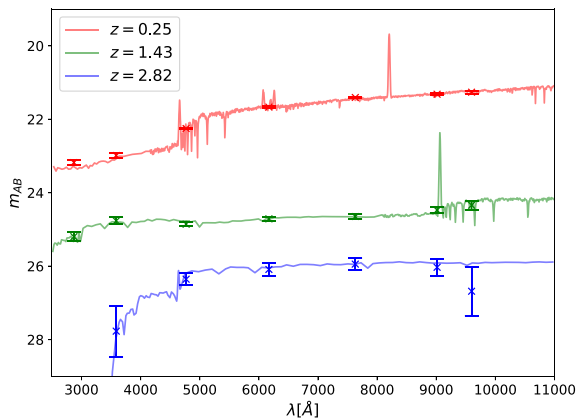
Figure 4. The flux data directly measured from the mock CSST galaxy images shown in Fig. 3 by aperture photometry in the seven photometric bands. The SEDs of the corresponding galaxy images are also shown in solid curves as comparison.

quantity, which is obtained by calculating the ratio of fluxes between two bands. Relative error and colour-like quantity are also rescaled by equation (7). For simplicity, we call these rescaled normalized flux, relative error, and colour-like quantity as flux, error, and colour in our following discussion.

Therefore, our MLP has 20 inputs, which contain seven fluxes, seven errors, and six colours. We construct six hidden layers with 40 units or neurons in each layer, which means a classic structure with n: $2n$:...$2n$: 1 is applied where n is number of data elements. We find that less number of hidden layers cannot get accurate results, while more layers result in lower efficiency and would not further improve the accuracy. Except for the first hidden layer, the BATCHNORMALIZATION layer is applied to reduce overfitting problem (Ioffe & Szegedy 2015), and every hidden layer is activated by the Rectified Linear Unit (RELU) non-linear function with a form

$y = \max(0, x)$ (Nair & Hinton 2010). The output of the MLP is the photometric redshift. The details of the architecture of the MLP are shown in the blue dash–dotted box of Fig. 5 and Table 1.

We adopt the CNN to analyse galaxy images. The mock CSST galaxy images observed in the seven photometric bands can be seen as 2d arrays or matrices in seven channels. The CNN can process multiple-channel arrays, which is suitable in our case for extracting photo-z from galaxy images. Note that the galaxy images we obtain are in different sizes due to the stamp slicing procedure based on semimajor axis mentioned in Section 2, but the CNN can only process images in the same sizes. So we crop centrally the images larger than a size of threshold $S_{threshold}$ and pad the images smaller than $S_{threshold}$ with random normal noise. Padding with random noise instead of 0 or fixed values can mimic the observations more realistically. These random normal noises are typical backgrounds derived from equation (4) in Section 2. The $S_{threshold}$ is set to be 32 pixels in our work, and other sizes with $S_{threshold} = 16$ and 64 are also explored. We find that smaller $S_{threshold}$ loses too much information since most galaxies occupy pixels larger than 16, and larger $S_{threshold}$ will be dominated by random padding noises that make the CNN cannot effectively concentrate on extracting information from galaxy in the centre of image. We also make use of the inception blocks to extract image features parallelly (Szegedy et al. 2014). The inception block is widely used for predicting photometric redshift through galaxy images, and the networks constructed with it display excellent performance (see e.g. Pasquet et al. 2019; Henghes et al. 2021). Inception block is illustrated in dashed blue box in Fig. 5. This block uses kernels of different sizes, i.e. (3 × 3) and (5 × 5), to extract image features parallelly. The (1 × 1) kernels are used to reduce number of channels to increase computation efficiency.

In our CNN, the inputs are (32 × 32 × 7) images. First, we apply a convolutional layer with 32 kernels of size (3 × 3) and stride size 2 to extract features and downsample images to 16. Then, we structure three inception blocks obtaining feature images of size 2. After these blocks, we use global average pooling to vectorize these features to 72 values (Lin, Chen & Yan 2013), and then apply a fully connected
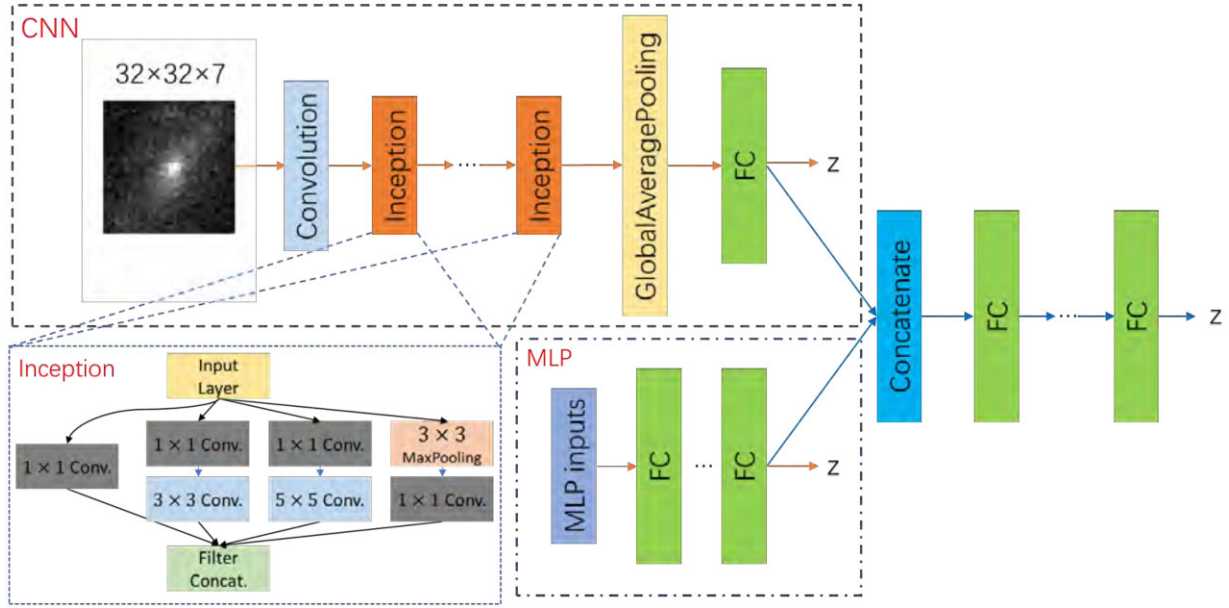
Figure 5. Architecture of the MLP, CNN, and hybrid networks. The MLP structure is shown in blue dash–dotted box. The inputs are rescaled fluxes, colours and errors, and totally six hidden layers are structured. The CNN structure is displayed in the dashed black box. The inputs are ($32 \times 32 \times 7$) images, convolved, and downsampled by the convolutional layer. Then, three inception blocks are structured to obtain the features of size 2, which is flattened by global average pooling to a vector of size 72. Next, the fully connected layer with 40 units is applied, and then the photo-z can be obtained. The inception blocks are illustrated in the dashed blue box. We use (3×3) and (5×5) kernels to extract features parallelly, and (1×1) kernels are adopted to reduce the number of features for increasing efficiency of computation. The hybrid network is constructed by concatenating the vectors extracted by the MLP and CNN from its fully connected layer. Totally, six fully connected layers with 80 units are structured, and after each layer the BATCHNORMALIZATION and RELU activation function are applied.

Table 1. Details of MLP architecture.

Layers	Output status[a]	Number of params.[b]
Input	20	0
FC[c]	40	840
ReLU	40	0
FC[c]	40	1640
BatchNormalization	40	160[d]
ReLU	40	0
	...[e]	
Output	1	41

[a]Number of data points or neurons.
[b]Total number of parameters: 9881.
[c]FC: fully connected layer.
[d]Half of them are non-trainable parameters.
[e]4 repeats of FC + BatchNormalization + ReLU.

Table 2. Details of CNN architecture.

Layers	Output status[a]	Number of params.[b]
Input	(32, 32, 7)	0
Conv2D	(16, 16, 32)	2048
BatchNormalization	(16, 16, 32)	128[c]
ReLU	(16, 16, 32)	0
Inception	(8, 8, 72)	9824
Inception	(4, 4, 72)	11 744
Inception	(2, 2, 72)	11 744
GlobalAveragePooling	72	0
FC[d]	40	2920
ReLU	40	0
Output	1	41

[a]Format: (dimension, dimension, channel) or number of neurons.
[b]Total number of parameters: 38 449.
[c]Half of them are non-trainable parameters.
[d]FC: fully connected layer.

layer of 40 units. Finally, the network outputs are photo-z values. In addition, after each convolutional layer, BATCHNORMALIZATION layer and LEAKYRELU activation function are applied (Maas et al. 2013). The architecture of our CNN is illustrated in dashed black box in Fig. 5, and the details of the CNN architecture are shown in Table 2.

Since the CNN can extract photo-z from the galaxy image mainly using morphology information, it is necessary to explore the improvement of the photo-z accuracy using both galaxy flux and image data. Hence, we construct a hybrid network to combine the MLP and CNN for including all data, i.e. galaxy fluxes, errors, colours, and images. In the hybrid network, the CNN and MLP parts are the same as the ones described above. Their fully connected layers are concatenated, and this means the features extracted by the CNN and MLP from galaxy images and flux are combined together. The

constructed vector by the combination is of size 80 (i.e. 40 from the CNN and 40 from the MLP). Then, we structure six fully connected layers with 80 units in each layer, and the BATCHNORMALIZATION and RELU activation function are applied after each layer. Then, the network outputs the photo-z values. The illustration of the hybrid network is shown in Fig. 5.
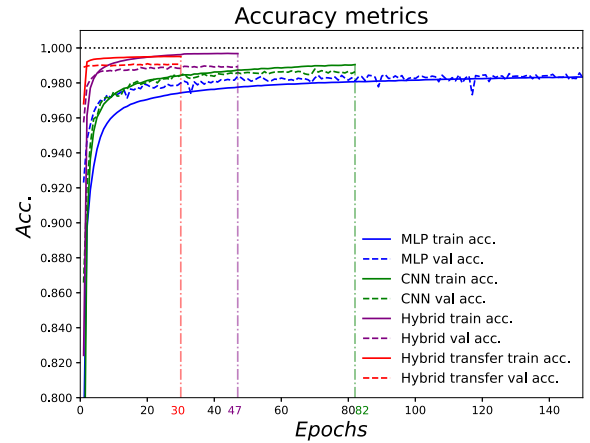
3.2 Training

The photo-z data of the CSST will be mainly used in the analysis of the weak gravitational lensing surveys, which needs high-quality sample with accurate photo-z information. Therefore, following previous works (Cao et al. 2018; Zhou et al. 2021), we select approximately 40 000 high-quality sources with the SNR in g or

i band larger than 10 from the original data set (containing $\sim$100 000 galaxies). So these 40 000 sources have both image and flux data, and also are assumed to have accurate spec-*z* measurements, although currently they only have high-quality photo-*z* measurements in the COSMOS catalogue. In the future real CSST survey, we will use the real spec-*z* sample obtained from future spectroscopic surveys as the training sample. We should notice that this selection criterion is conservative, and large amounts of galaxies with low SNRs will be removed, which can be used in weak lensing survey. However, since the systematical errors are the main problem to affect the accuracy of next-generation weak lensing measurements, and photo-*z* uncertainty is one of the main components, we will try to suppress the systematics with high priority, even losing a fraction of galaxies. Since the quality of galaxy shape measurement also should be considered as an important criterion in weak lensing surveys, other potentially better selection criteria will be explored in future works. Then, we divide the above sample into the training and testing data, and randomly save 10 000 sources for testing and the rest of 30 000 data for training, which means the ratio of the training and testing data is roughly 3: 1. In order to investigate the effect of the number of training data on the photo-*z* estimation, we also try ratios roughly 1: 1 and 1: 3 for the training and testing samples.

Note that we have assumed all the samples we use to train the networks have measured spec-*z* provided by future spectroscopic surveys as discussed in the introduction. Galaxies in these samples can fully represent the features of galaxies in the CSST photometric survey, such as the redshift and magnitude distributions, galaxy types, and so on. In real surveys, this assumption may not hold, and selection effect of spectroscopic samples will influence the estimates on photo-*z*. Since the result derived from the neural networks is highly dependent on the training sample, if there is a kind of galaxies that is not covered by the spec-*z* training set, the networks probably would not provide proper photo-*z* estimates for these galaxies. However, we may fix this problem by analysing the photometric and spectroscopic samples and finding the missing galaxies in the spec-*z* training set by some means, e.g. adopting the self-organizing map suggested in Masters et al. (2015). Then, we can try to cover these galaxies on purpose with the help of available spectroscopic surveys.

In our MLP, we use the mean absolute error (MAE) as our loss function, which calculates the average of absolute difference between predicted redshifts and true redshifts. The Adam optimizer is adopted to optimize learnable weights (Kingma & Ba 2014). Adam optimizer is an efficient stochastic optimization method requiring only first-order gradients, and it can adjust learning rate automatically in the training process for every trainable weight in network. In addition, we define an accuracy metric based on outlier percentage defined as $|z_{\text{true}} - z_{\text{pred}}| < 0.15(1 + z_{\text{true}})$, and the network will monitor this metric when training. We set the initial learning rate to be 10^{-4}, and use MODELCHECKPOINT callback to save the model with highest validation accuracy. The maximum number of epochs is set to be 150 in the training process. In order to reduce the effect of the statistical noise, the MLP training data are augmented by random realizations based on flux errors with the Gaussian distribution. This method has found to be effective for eliminating the effect of statistical noise in the training process (Zhou et al. 2021). Here, we use 50 realizations for every training source. We find that more realizations do not improve the results significantly. In Fig. 6, we show the accuracy versus epoch for both training and validation samples in blue solid and dashed curves, respectively. We can see that the validation accuracy can reach $\sim$0.98. The validation samples are mainly used to tune the hyper-parameters of a network. On the

Figure 6. Accuracy metrics of different networks in training (solid) and validation (dashed). Blue, green, purple, and red lines represent MLP, CNN, hybrid, and hybrid transfer, respectively. Validation accuracy reaches approximately 0.98 for the MLP. For CNN, the training stops at the 82nd epoch, since the validation accuracy does not improve in previous 10 epochs, and the maximum accuracy can be higher than MLP ones. Hybrid and hybrid transfer networks can reach highest accuracy trained only 47 and 30 epochs, respectively. The vertical dotted lines show the epochs when training stops.

phase of tuning networks, we randomly select validation sample as 10 per cent of training sample.

In the CNN, we also use MAE and Adam optimizer with initial learning rate setting to be 10^{-4}, and the learning rates are reduced by 10 through REDUCELRONPLATEAU callback when validation accuracy does not improve in 5 epochs. To make the training efficient, we use EARLYSTOPPING callback to shut down training if no improvement occurs in 10 epochs, and save the model with the highest validation accuracy. To augment the training data, galaxy images for training are rotated and flipped, resulting in 8 $\times$ original data size. In Fig. 6, the green lines show the accuracy curves for the CNN training (solid) and validation (dashed) samples. We can find that they stop at 82nd epoch, meaning previous 10 epochs have no improvement in validation accuracy, and the model is saved at the 72nd epoch.

For the hybrid network, settings of the loss function, optimizer, and callbacks are the same as the corresponding CNN described above. The hybrid network inputs include both flux and image data for a galaxy, i.e. flux, error, and colour for the MLP and a corresponding galaxy image for the CNN. The inputs of the MLP part in the hybrid network are also augmented by 50 realizations based on flux errors. For the CNN part, we need to input the corresponding galaxy image given a random flux realization. To make the training process efficient and extracting more information, we input a galaxy image by randomly rotating or flipping. This means that, for a galaxy, the inputs of the hybrid network contain 50 realizations of the flux data and the corresponding 50 randomly rotating or flipping galaxy images.

To investigate the possibility that if we can further improve the network efficiency and accuracy, we also propose and explore a hybrid transfer network, which is inspired by the techniques from the transfer learning. Transfer learning is a technique focusing on using knowledge gained from one problem to solve a different but related problem (West, Ventura & Warnick 2007). Here, we borrow this idea and combine the features learned by MLP and CNN together. As we show in the following discussion, this network can have better training efficiency and obtain higher photo-*z* accuracy. Since our

MLP and CNN have already been well trained and can provide good photo-z accuracies, we can directly freeze the weights obtained by the MLP and CNN and transfer them to the MLP and CNN parts of the hybrid network, i.e. constructing a hybrid transfer network. In this way, the MLP and CNN parts are restricted to learn features that is useful to predict photo-z, and the concatenated features in the joint fully connected layers can improve the estimation. Note that the architecture of the hybrid transfer network is identical to the hybrid network, except for applying a different training strategy inspired by transfer learning. Besides, in the MLP and CNN parts, we just freeze the first five fully connected layers in the MLP and layers before the global average pooling layer in the CNN, and the last fully connected layers in the MLP part and the one in the CNN part are still involved in the training process to include more flexibility for better extracting photo-z information.

In Fig. 6, the purple and red lines show the accuracy curves of the hybrid and hybrid transfer networks, respectively. We notice that validation curves (purple and red dashed curves) of the two networks can reach high accuracy in early epochs, and the training stops at the 47th and 30th epochs for the hybrid and hybrid transfer networks, respectively. This indicates that the hybrid network can be optimized efficiently, and the hybrid transfer network even performs better.

4 RESULT

To assess the accuracy of extracted photo-z from the networks, we define catastrophic outliers to be $|\Delta z|/(1 + z_{\text{true}}) < 0.15$, where $\Delta z = z_{\text{pred}} - z_{\text{true}}$, and adopt the normalized median absolute deviation (NMAD; Brammer, van Dokkum & Coppi 2008), which is calculated as

$$\sigma_{\text{NMAD}} = 1.48 \times \text{median} \left(\left| \frac{\Delta z - \text{median}(\Delta z)}{1 + z_{\text{true}}} \right| \right). \qquad (8)$$

This scatter or deviation can naturally suppress the weights of the outlier redshift, and provide a proper estimation of the photo-z accuracy. We also calculate the average of $|\Delta z|/(1 + z_{\text{true}})$ as MAE in the analysis.

In the upper-left panel of Fig. 7, we show the result from the MLP using the flux data. We find that the outlier fraction $\eta = 1.43$ per cent, $\sigma_{\text{NMAD}} = 0.023$ for the testing data. This result has been improved by a factor 2 for the outlier fraction compared to that using the SED template-fitting method shown in Cao et al. (2018). On the other hand, the result is worse than that given in Zhou et al. (2021). This is because in this work the flux data we use are obtained by directly measuring galaxy images with aperture photometry, instead of adopting the data estimated from SED templates used in Zhou et al. (2021), which are more realistic and including more noises. We can see that the MLP tends to predict larger redshifts at $z \lesssim 0.5$ and smaller when $z \gtrsim 1.5$. We also note that predictions have little outliers in redshift range $0.5 < z < 1.5$, which is assuring high photo-z accuracy for most of galaxies observed by the CSST (see Fig. 2).

The CNN result is shown in the upper-right panel of Fig. 7, and we have $\eta = 1.21$ per cent and $\sigma_{\text{NMAD}} = 0.025$. We can find that this result has smaller outlier fraction and a bit larger σ_{NMAD} compared to the MLP using the flux data. Since our galaxy flux data are measured from galaxy images by aperture photometry, in principle, besides the morphology information, the images also abstractly include the information of fluxes and colours. Hence, if the CNN can successfully extract all or part of these information, it is capable to obtain a comparable and even better photo-z result than that using the flux data only. Based on the current result, it seems that our CNN can really extract some flux information from galaxy images in addition

to morphology. In Fig. 7, we can see that the outlier distribution from the CNN is similar to the MLP, but it can further suppress the outlier fraction in the whole redshift range.

In the lower-left and lower-right panels of Fig. 7, we show the results from the hybrid and hybrid transfer networks, respectively. We find that our hybrid network applying transfer learning technique indeed can mildly improve the photo-z estimation with $\eta = 0.90$ per cent and $\sigma_{\text{NMAD}} = 0.020$ compared to the hybrid network with $\eta = 1.03$ per cent and $\sigma_{\text{NMAD}} = 0.020$. This indicates that both our hybrid and hybrid transfer network could provide accurate photo-z estimates by including galaxy flux and image data, and the hybrid transfer network performs even better with higher efficiency and accuracy. This implies that the features from trained CNN and MLP are probably better than features directly learned from hybrid when predicting photo-z. In Fig. 7, we notice that the hybrid transfer network can further suppress the outliers by 37 and 26 per cent compared to the MLP and CNN cases, respectively, especially in $z \lesssim 0.5$ and $z \gtrsim 1.5$. In addition, the results from hybrid or hybrid transfer networks are obviously better than the MLP only and CNN only cases, which means the CNN indeed can extract useful morphological information from galaxy images to improve the photo-z estimation.

The results shown above are obtained by using a ratio of $r = 3 : 1$ for the training and testing data. We also test if smaller training sample can severely affect our photo-z estimation, since the network training is restricted to the spectroscopic data providing accurate redshifts and we may not have large qualified training sample in real observations. Here, we try two training-testing ratios $r = 1 : 1$ and $1 : 3$, which means 20 000 and 10 000 training data are randomly selected, respectively, and the rest are saved for testing. After training and testing processes, the networks can give $\eta = 1.67, 1.51, 1.25, 1.06$ per cent and $\sigma_{\text{NMAD}} = 0.026, 0.027, 0.022, 0.021$ for the MLP, CNN, hybrid, and hybrid transfer networks, respectively, for $r = 1 : 1$ or 20 000 training sample case. The results become $\eta = 2.39, 1.98, 1.29$, and 1.29 per cent, and $\sigma_{\text{NMAD}} = 0.028, 0.030, 0.023$, and 0.023 for $r = 1 : 3$ or 10 000 training sample case. We can note that decreasing training data indeed result in worse predictions, but not significantly. The hybrid and hybrid transfer network seem mostly immune to this ratio in these four networks, and still can provide $\eta \sim 1$ per cent and $\sigma_{\text{NMAD}} \sim 0.02$, that is similar as the $r = 3 : 1$ or 30 000 training sample case. On the other hand, the results become obviously worse for the MLP and CNN. This indicates that relatively more data are needed to sufficiently train the MLP and CNN, compared to the hybrid and hybrid transfer networks. We also should note that the absolute number of training data determines the performance of trained model instead of the ratio of training and testing sample. As long as we have large enough spec-z data that can represent the features of all galaxies in photometric survey, the neural networks can be well trained and applied to derive photo-z, no matter how large of the ratio of spectroscopic and photometric data. Besides, the size of the testing sample should also be large enough, the results otherwise can be affected by statistical effects like cosmic variance.

In Fig. 8, the percentages of outliers as a function of photometric redshift and spectroscopic redshift are illustrated in the top and bottom panels, respectively. The redshift bin size is set to be $\Delta z = 0.5$. The average outlier fractions and errors derived from 10 mock flux testing data sets using MLP are also shown, which can indicate the statistical effects of number of sources in the training sample at different redshifts. Although currently it is hard for us to generate enough mock images to derive errors in the CNN case, the errors should be similar as the MLP case, since they have similar photo-z estimation results. This is also representative for the cases using the
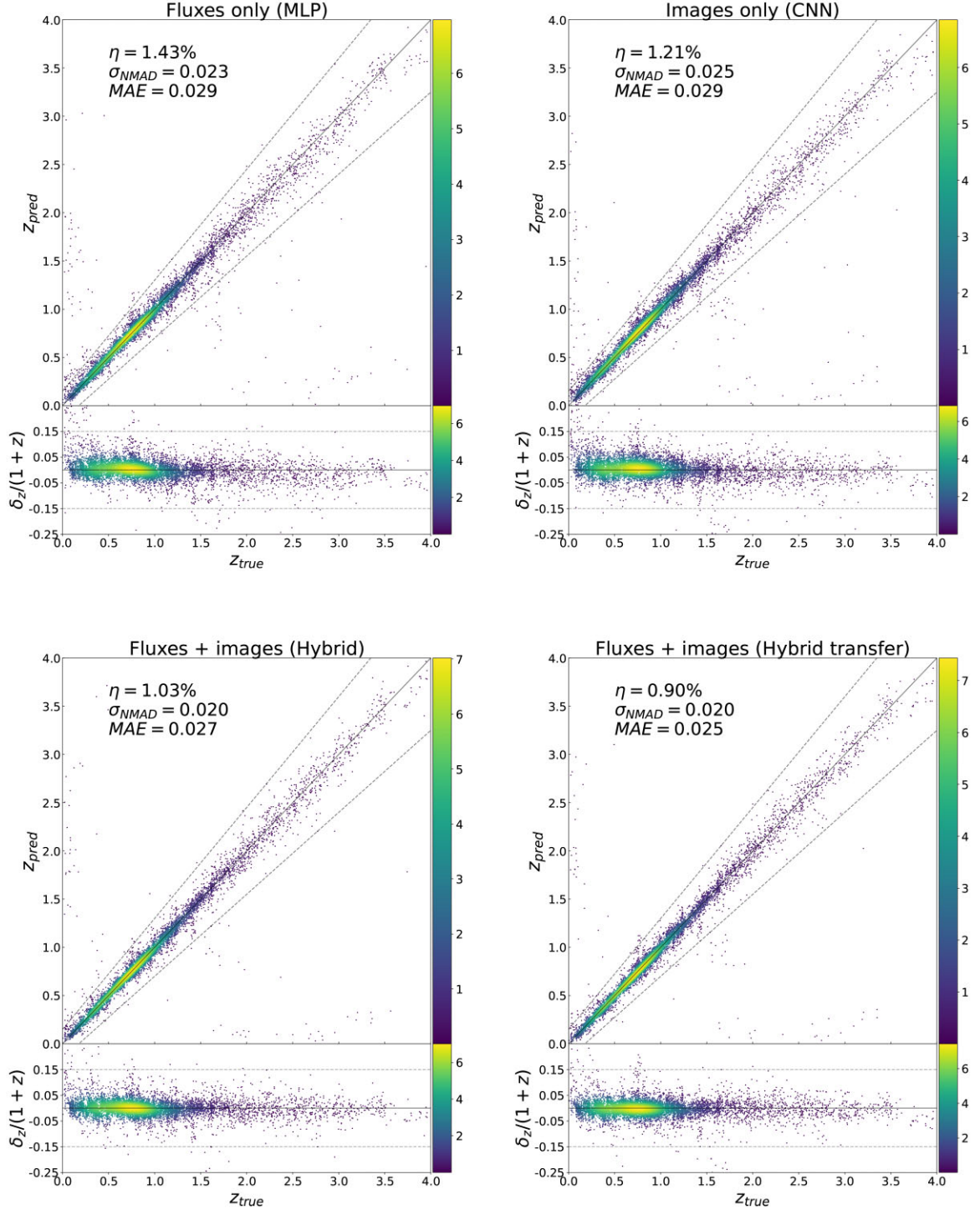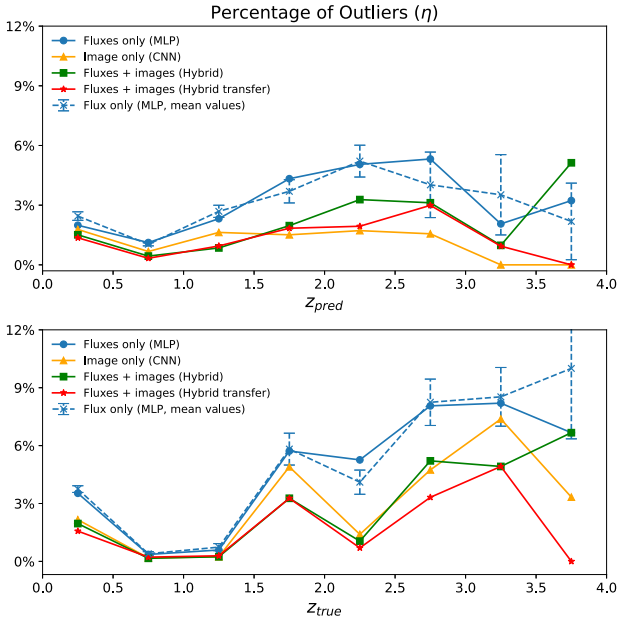
Figure 7. Results of the MLP, CNN, hybrid, and hybrid transfer networks when the ratio of training to testing data roughly equals to 3: 1. The colour bar at right of each panel denotes the number density of sources per pixel of the testing data. The MLP using flux data can provide lower photo-z scatter, while the CNN using image data has smaller outlier fraction. The hybrid network can improve the results of both accuracy and outlier fraction, and the hybrid transfer network can even further suppress the outliers and has the best photo-z result.

hybrid and hybrid transfer networks. In the bottom panel, we can see that, generally speaking, the outlier fraction is increasing as redshift increases. In $z = 0.5$–1.5, all of the four networks give the lowest outliers, that is probably due to large training sample in this range (see Fig. 2). In $z = 3.5$–4, there are few galaxies and the results

are dominated by statistical errors, which lead to large scatters in the results between the four networks. The results can be heavily disturbed if shown as a function of photo-z as indicated in the top panel. In this case, we still have small outliers in $z = 0.5$–1, but it is not visibly increasing but is relatively flat as photo-z increases.

Figure 8. The percentages of outliers as a function of photometric redshift (top panel) and spectroscopic redshift (bottom panel). The blue, orange, green, and red curves are for the MLP, CNN, hybrid, and hybrid transfer networks, respectively. The average outlier fractions and errors derived from 10 mock flux testing data sets using MLP are shown in blue dashed curves and data points with error bars.

5 SUMMARY AND CONCLUSION

We explore the photo-z accuracy of the CSST photometric observations using neural networks. Four neural networks, i.e. MLP, CNN, hybrid, and hybrid transfer, are adopted and tested to extract the photo-z information from the CSST mock flux and image data. We first simulate the CSST photometric observations to generate mock galaxy images for the seven bands, which are created based on the *HST*-ACS observations and the COSMOS catalogue considering the CSST instrumental effects. Then, we measure fluxes and errors in each band from galaxy images using aperture photometry method. Since the photo-z data would mainly be used in the analysis of the CSST weak gravitational lensing surveys, about 40 000 high-quality sources with SNR > 10 in the g or i band are selected as the training and testing samples in our neural networks.

The MLP and CNN are used to predict photo-z from fluxes and images, respectively. Since the flux measured by aperture photometry could be negative for faint galaxies especially in the bands with low transmissions, we rescale the measured galaxy flux, error, and colour, which also can speed up the training process of the networks. These flux data are inputted into the constructed MLP with six fully connected layers for deriving photo-z. The mock galaxy images in CSST seven photometric bands are cropped or padded into the same size as the inputs for the CNN. The inception blocks are employed to extract image features parallelly with different sizes of kernels. To investigate the improvement of photo-z accuracy using both flux and image data, we develop a hybrid network by properly combining the MLP and CNN. We also propose a hybrid transfer network, adopting the transfer learning techniques, which may have higher efficiency and accuracy.

In the MLP training process, to effectively suppress statistical noises of flux data, we randomly generate 50 realizations from Gaussian distribution for each galaxy flux data based on errors

as the training data. We also augment galaxy images by rotating and flipping to produce more training data by a factor of 8 in the CNN training process. In the hybrid network, we randomly input a corresponding rotating or flipping galaxy image in the CNN part given a flux realization, and can efficiently provide accurate photo-z result. In the hybrid transfer network, we freeze the weights of the MLP and CNN parts as the ones separately trained by the flux and image data. We find that the MLP, CNN, hybrid, and hybrid transfer networks can derive accurate photo-z results with similar deviation 0.020–0.025 and outlier fraction around 1 per cent. The CNN outlier fraction is better than the MLP, since our CNN can effectively extract both flux and morphology information from galaxy images. The hybrid (hybrid transfer) network can offer best result with 28 and 15 per cent (37 and 26 per cent) improvements compared to the MLP and CNN, respectively. The effect of training and testing ratio is also explored, and we find that smaller training sample would not affect the networks significantly.

Note that we only predict photo-z values without uncertainties or probability distributions. In future work, we will try to use probabilistic deep learning methods, such as Bayesian neural networks (Mackay 1995; Neal 2012; Blundell et al. 2015; Gal & Ghahramani 2015; Wilson & Izmailov 2020) or other approaches to output both photo-z values and uncertainties, and even their probability distributions.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

REFERENCES

Abbott T. M. C. et al., 2021, ApJS, 255, 20
Abdalla F. B., Amara A., Capak P., Cypriano E. S., Lahav O., Rhodes J., 2008, MNRAS, 387, 969
Akeson R. et al., 2019, preprint (arXiv:1902.05569)
Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540
Banerji M., Abdalla F. B., Lahav O., Lin H., 2008, MNRAS, 386, 1219
Blundell C., Cornebise J., Kavukcuoglu K., Wierstra D., 2015, preprint (arXiv:1505.05424)
Bohlin R. C., 2016, AJ, 152, 60
Bolzonella M., Miralles J. M., Pelló R., 2000, A&A, 363, 476
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Brescia M., Cavuoti S., Razim O., Amaro V., Riccio G., Longo G., 2021, Frontiers Astron. Space Sci., 8, 70
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Bundy K. et al., 2019, BAAS, 51, 198

Cao Y. et al., 2018, MNRAS, 480, 2178
Cirasuolo M. et al., 2020, The Messenger, 180, 10
DES, Abbott T.DES, et al., 2016, MNRAS, 460, 1270
Collister A. A., Lahav O., 2004, PASP, 116, 345
de Jong R. S. et al., 2019, The Messenger, 175, 3
Ellis R., Dawson K., 2019, BAAS, 51, 45
Fernández-Soto A., Lanzetta K. M., Yahil A., 1999, ApJ, 513, 34
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Fukushima K., Miyake S., 1982, Pattern Recognit., 15, 455
Gal Y., Ghahramani Z., 2015, preprint (arXiv:1506.02142)
Gong Y. et al., 2019, ApJ, 883, 203
Green J. et al., 2012, preprint (arXiv:1208.4012)
Haykin S., Haykin S. S, HAYKIN S. A, 1999, Neural Networks: A Comprehensive Foundation. Prentice Hall, Hoboken, New Jersey
Hearin A. P., Zentner A. R., Ma Z., Huterer D., 2010, ApJ, 720, 1351
Henghes B., Pettitt C., Thiyagalingam J., Hey T., Lahav O., 2022, MNRAS, 512, 1696
Ilbert O. et al., 2006, A&A, 457, 841
Ioffe S., Szegedy C., 2015, preprint (arXiv:1502.03167)
Ivezić Ž. et al., 2019, ApJ, 873, 111
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Koekemoer A. M. et al., 2007, ApJS, 172, 196
Kron R. G., 1980, ApJS, 43, 305
Laigle C. et al., 2016, ApJS, 224, 24
Lanzetta K. M., Yahil A., Fernández-Soto A., 1996, Nature, 381, 759
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Leauthaud A. et al., 2007, ApJS, 172, 219
Lecun Y., Bottou L., Bengio Y., Haffner P., 1998, Proc. IEEE, 86, 2278
Levi M. et al., 2019, BAAS, 51, 57
Lin M., Chen Q., Yan S., 2013, preprint (arXiv:1312.4400)
LSST, Abell P. ALSST, et al., 2009, preprint (arXiv:0912.0201)
Ma Z., Hu W., Huterer D., 2006, ApJ, 636, 21
Maas A. L., Hannun A. Y., Ng A. Y., 2013, Proc. icml, 30, 3
Mackay D. J. C., 1995, Netw., Comput. Neural Syst., 6, 469
Maiolino R. et al., 2020, The Messenger, 180, 24
Mandelbaum R. et al., 2008, MNRAS, 386, 781

Massey R., Stoughton C., Leauthaud A., Rhodes J., Koekemoer A., Ellis R., Shaghoulian E., 2010, Pixel-based correction for Charge Transfer Inefficiency in the Hubble Space Telescope Advanced Camera for Surveys, Instrument Science Report ACS 2010-01
Masters D. et al., 2015, ApJ, 813, 53
Nair V., Hinton G. E., 2010, Proceedings of ICML, 27, 807
Neal R., 2012, Bayesian Learning for Neural Networks. Lecture Notes in Statistics. Springer, New York
Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, A&A, 621, A26
Sadeh I., Abdalla F. B., Lahav O., 2016, PASP, 128, 104502
Schlegel D., Kollmeier J. A., Ferraro S., 2019, BAAS, 51, 229
Soo J. Y. H. et al., 2018, MNRAS, 475, 3613
Szegedy C. et al., 2014, preprint (arXiv:1409.4842)
Tamura N., PFS Collaboration, 2016, in Skillen I., Balcells M., Trager S., eds, ASP Conf. Ser. Vol. 507, Multi-Object Spectroscopy in the Next Decade: Big Questions, Large Surveys, and Wide Fields. Astron. Soc. Pac., San Francisco, p. 387
Ubeda L., et al., 2011, ACS Instrument Handbook for Cycle 20 v. 11.0, Vol. 11, STScI, Baltimore, p. 11
Wadadekar Y., 2005, PASP, 117, 79
Weaver J. R. et al., 2022, ApJS, 258, 11
West J., Ventura D., Warnick S., 2007, presentation at "A Theoretical Foundation for Inductive Transfer", Brigham Young University, College of Physical and Mathematical Sciences
Wilson A. G., Izmailov P., 2020, preprint (arXiv:2002.08791)
Wilson D., Nayyeri H., Cooray A., Häußler B., 2020, ApJ, 888, 83
York D. G. et al., 2000, AJ, 120, 1579
Zhan H., 2011, Sci. Sin. Phys. Mech. Astron., 41, 1441
Zhan H., 2018, in 42nd COSPAR Scientific Assembly, Pasadena, California, E1.16–4-18
Zhan H., 2021, Chin. Sci. Bull., 66, 1290
Zhan H., Knox L., Tyson J. A., Margoniner V., 2006, ApJ, 640, 8
Zhou X. et al., 2021, ApJ, 909, 53

This paper has been typeset from a TEX/LATEX file prepared by the author.