CrossMark

Spectroscopic and Photometric Redshift Estimation by Neural Networks for the China Space Station Optical Survey (CSS-OS)

Xingchen Zhou[1,2], Yan Gong[1,3,10] , Xian-Min Meng[1], Xin Zhang[1], Ye Cao[1,2], Xuelei Chen[2,4,5] , Valeria Amaro[6], Zuhui Fan[7,8], and Liping Fu[9]

[1] Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, People's Republic of China; gongyan@bao.ac.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
[3] Science Center for China Space Station Telescope, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, People's Republic of China
[4] Key Laboratory for Computational Astrophysics, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, People's Republic of China
[5] Centre for High Energy Physics, Peking University, Beijing 100871, Peoples Republic of China
[6] School of Physics and Astronomy, Sun Yat-sen University, Zhuhai Campus, Guangzhou 519082, Peoples Republic of China
[7] South-Western Institute for Astronomy Research, Yunnan University, Kunming 650500, Peoples Republic of China
[8] Department of Astronomy, School of Physics, Peking University, Beijing 100871, Peoples Republic of China
[9] Shanghai Key Lab for Astrophysics, Shanghai Normal University, Shanghai 200234, People's Republic of China

Abstract

The estimation of spectroscopic and photometric redshifts (spec-z and photo-z) is crucial for future cosmological surveys. It can directly affect several powerful measurements of the universe, such as weak lensing and galaxy clustering. In this work, we explore the accuracies of spec-z and photo-z that can be obtained by the China Space Station Optical Surveys, which is a next-generation space survey, using a neural network. The one-dimensional Convolutional Neural Networks and Multi-Layer Perceptron (MLP, the simplest form of an artificial neural network) are employed to derive spec-z and photo-z, respectively. The mock spectral and photometric data used for training and testing the networks are generated based on the COSMOS catalog. The networks have been trained with noisy data by creating Gaussian random realizations to reduce the statistical effects, resulting in a similar redshift accuracy for data with both high and low signal-to-noise ratios. The probability distribution functions of the predicted redshifts are also derived via Gaussian random realizations of the testing data, and then the best-fit redshifts and 1σ errors also can be obtained. We find that our networks can provide excellent redshift estimates with accuracies of $\sim$0.001 and 0.01 on spec-z and photo-z, respectively. Compared to existing photo-z codes, our MLP has a similar accuracy but is more efficient in the training process. The fractions of catastrophic redshifts or outliers can be dramatically suppressed compared to the ordinary template-fitting method. This indicates that the neural network method is feasible and powerful for spec-z and photo-z estimations in future cosmological surveys.

Unified Astronomy Thesaurus concepts: Cosmology (343); Large-scale structure of the universe (902)

1. Introduction

Understanding the properties of dark matter and dark energy and the formation and evolution of the cosmic large-scale structure (LSS) are important goals for cosmological observations. The ongoing and next-generation ground and space telescopes, such as the Dark Energy Spectroscopic Instrument (DESI)[11], the Large Synoptic Survey Telescope[12] (Ivezic et al. 2008; Abell et al. 2009), the Euclid space telescope[13] (Laureijs et al. 2011), and the Wide Field Infrared Survey Telescope[14] (WFIRST), are dedicated to answering these questions to ultrahigh precision. By then, billions of galaxies will be detected in enormous spatial volumes over large redshift ranges via spectroscopy and imaging photometry, and the kinematic and dynamical evolution of the universe can be precisely measured. All of these goals are tightly related to the measurements of galaxy redshifts. Obtaining accurate spectroscopic and photometric redshifts (spec-z and photo-z) is a strict requirement for these future cosmological surveys.

The China Space Station Telescope (CSST) is a 2 meter space telescope focusing on the next-generation cosmological observations. It is scheduled to launch in 2024 and will be in the same orbit as the China Manned Space Station (Zhan 2011, 2018; Cao et al. 2018; Gong et al. 2019). The China Space Station Optical Survey (CSS-OS) is one of the main missions of the CSST, and it is planned to observe 17,500 deg^2 in about 10 years, covering the optical and near-IR bands from $\sim$250 nm to 1000 nm. It can simultaneously perform photometric and slitless spectroscopic surveys over a large field of view and at high spatial resolution. The scientific goals of the CSS-OS include exploring the properties of dark matter and dark energy, the evolution of the LSS, galaxy formation and evolution, galaxy clusters, etc. Several kinds of cosmological and astrophysical observations will be performed by the CSST to achieve these goals, such as weak and strong gravitational lensing, two- and three-dimensional (2D and 3D) galaxy clustering, galaxy and active galactic nucleus surveys, and so on.

Weak lensing and galaxy clustering are powerful probes in cosmological studies, which could provide stringent constraints on dark matter, dark energy, and the evolution of the LSS. Since the accuracies of spec-z and photo-z can directly affect the measurements of the shear correlation function of weak lensing and 2D and 3D galaxy power spectra, high accuracies are crucial

---

[10] Corresponding author.
[11] https://www.desi.lbl.gov/
[12] https://www.lsst.org/
[13] https://www.euclid-ec.org/
[14] https://wfirst.gsfc.nasa.gov/

for the CSS-OS. Spec-z and photo-z usually can be derived by fitting galaxy spectral energy distribution (SED) templates. However, this method is highly dependent on the quality of the selected SED templates, such that accurate spec-z and photo-z data may not be obtained if the adopted SED templates are not sufficiently representative of the target population. Besides, poor spectroscopic and photometric measurements can also lead to inaccurate spec-z and photo-z fitting in this method. This is particularly the case for high-redshift and faint galaxies, slitless spectroscopic surveys with low spectral resolution (like CSST), and photometric bands with small filter quantum transmission efficiency (e.g., CSST NUV and y bands).

That said, we can try to find empirical relations between galaxy properties (e.g., SED, band magnitude, color, morphology, etc.) and redshift by exploring a huge amount of galaxies with known redshifts. This can be achieved by training neural networks with appropriate training data sets, which can be obtained by processing high-quality data from ongoing or future spectroscopic surveys, such as the Multi-Object Optical and Near-infrared Spectrograph (MOONS;[15] Cirasuolo et al. 1919; Maiolino et al. 2020), Prime Focus Spectrograph (Tamura et al. 2016), DESI, 4 meter Multi-Object Spectroscopic Telescope[16], MegaMapper (Schlegel & Kollmeier 2019), Fiber-Optic Broadband Optical Spectrograph (Bundy 2019), SpecTel (Ellis 2019), and so on. These powerful spectroscopic surveys would get large amounts of high-quality galaxy SEDs at $z = 0 \sim 4$, which can be used as training data for the CSST spectroscopic and photometric surveys.

In recent years, the development of machine learning has been remarkable, especially for neural network algorithms. Neural networks have been widely applied in astronomy and cosmology studies, especially the Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN). The MLP is the simplest form of the Artificial Neural Network (ANN), which consists of an input layer, several hidden layers, and an output layer. The CNN was first proposed by Fukushima et al. (1970) and Lecun et al. (1998), and it is a kind of neural network applied widely in daily life and in scientific research. CNNs can process data with multiple arrays, such as images with RGB channels, and they have great success in detection, segmentation, and recognition applications. In this work, the 1D CNN and MLP are adopted to derive spec-z and photo-z for the CSS-OS, respectively. The Adam optimizer is adopted to automatically update the network learning rate (Kingma & Ba 2014), which is more efficient for network training than the previous machine-learning codes used in redshift estimation (e.g., Collister & Lahav 2004; Gerdes et al. 2010; Samui & Pal 2017).

We make use of the COSMOS galaxy catalog (Capak et al. 2007; Ilbert et al. 2009) to generate mock spectroscopic and photometric data for training and testing the neural networks. The probability distribution function (PDF), best fit (the value at the peak of PDF), and 1σ error of the predicted redshift can be obtained in our analysis. In this study, we try to address a number of important questions about the neural networks applied to the CSS-OS, e.g., the accuracy that the CSS-OS spec-z and photo-z can achieve using this method, how many training data are needed in this process, can the redshifts be correctly derived from low-quality data, and so on. The

answers to these questions will help determine the detailed strategy of extracting spec-z and photo-z using neural networks in the processing of real data from CSS-OS in the future.

This paper is organized as follows: in Section 2, we discuss the generation of mock data we used for training and testing. In Section 3, we show the architectures of the adopted neural networks and details of the training process for the CSST spec-z and photo-z predictions. The results are presented in Section 4. We summarize our results in Section 5.

2. Mock Data

In this section, we discuss the generation of spectroscopic and photometric mock data for the CSS-OS. The mock data need to correctly represent the galaxy observations of the CSS-OS, which should have similar properties, e.g., redshift and magnitude distributions, and galaxy types. Here, the COSMOS galaxy catalog is used to create the mock data, which contains about 220,000 galaxies with $i^+ \leqslant 25.2$ in ~ 2 deg^2, including measures of galaxy redshift, magnitude, size, dust extinction, best-fit SED, and so on (Capak et al. 2007; Ilbert et al. 2009; Cao et al. 2018; Gong et al. 2019). Since the galaxy redshifts and best-fit SEDs can be seen as high-quality data obtained by a number of powerful spectroscopic galaxy surveys (as listed in the introduction), with the help of the information given by the catalog, we can generate the training and testing samples for both the CSST spectroscopic and photometric surveys by considering the CSST observational and instrumental effects. As mentioned in the following discussion, the main selection criteria of the CSST spectroscopic and photometric mock data are the magnitude limit and signal-to-noise ratio (S/N).

2.1. Spectroscopic Mock Data

The CSST will perform the spectroscopic survey using slitless gratings. It has three bands, GU, GV, and GI, as shown as dashed curves in the two panels of Figure 1. The AB magnitude 5σ limits of the three bands are 23.1, 23.4, and 23.5 for point sources, respectively, with a spectral resolution $R = \lambda/\Delta\lambda \gtrsim 200$ (Gong et al. 2019). Based on these survey parameters, we can select the galaxy samples and generate the mock data from the COSMOS catalog.

First, in order to obtain the galaxy SEDs with emission lines, we refit the galaxy photometric data of the COSMOS catalog using the *LePhare*[17] code (Arnouts et al. 1999; Ilbert et al. 2006) by switching on the emission line option. There are 10 bands used in the fitting process, including u^* (CFHT), B_J (Subaru), V_J (Subaru), g^+ (Subaru), r^+ (Subaru), i^+ (Subaru), i^* (CFHT), z^+ (Subaru), J (UKIRT), and K (CFHT); (Ilbert et al. 2009). These 10 bands contain point-spread function (PSF) matched photometry data and have similar detection limits. A number of emission lines are involved, e.g., the Lyα, Hα, Hβ, [O II], and [O III] lines. The redshift is fixed in this process to improve the SED fitting accuracy, since we have assumed that the given redshifts here are accurate and can be seen as the real redshifts. The templates of the extinction curves and other initial parameter set-ups are the same as Ilbert et al. (2009). We have checked the fitting results with those provided by the COSMOS catalog, and we find that they are in good agreement. Examples of derived SEDs from the 10 bands are shown in Figure 2.
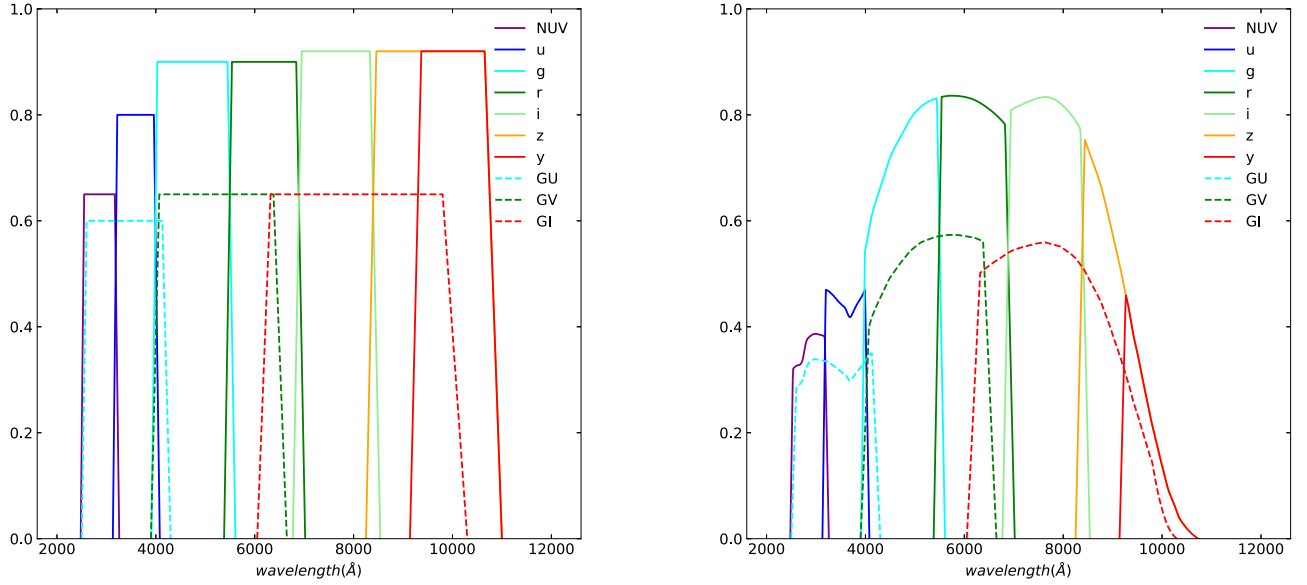
Figure 1. Left panel: the transmission curves of the CSST seven photometric (solid) and three spectroscopic (dashed) bands, covering the near-UV to the near-IR (Cao et al. 2018; Gong et al. 2019). Right panel: the total transmission curves considering quantum efficiency (Cao et al. 2018).
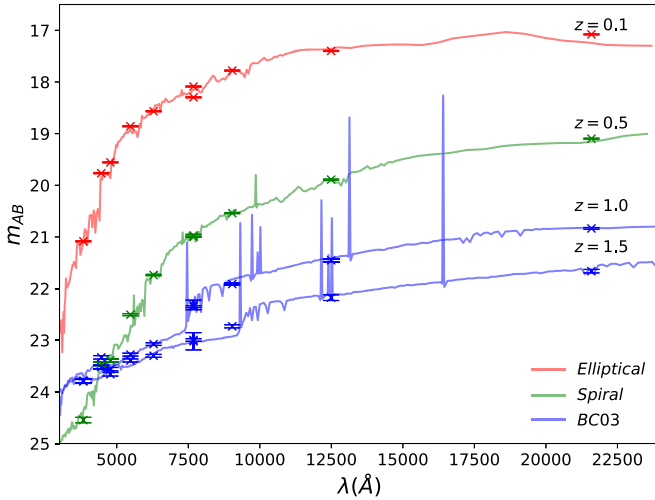

Figure 2. Examples of the derived SEDs with emission lines by refitting the photometric data of the COSMOS catalog using the *LePhare* code. For comparison, the photometric data in the 10 bands are also shown, i.e., u^* (CFHT), B_J (Subaru), V_J (Subaru), g^+ (Subaru), r^+ (Subaru), i^+ (Subaru), i^* (CFHT), z^+ (Subaru), J (UKIRT), and K (CFHT); (Ilbert et al. 2009). The red, green, and blue curves and data points are for elliptical, spiral, and young blue star-forming galaxies (denoted as BC03, which are fitted by the method given in Bruzual & Charlot 2003), respectively.

Second, we convolve the obtained SEDs with the CSST spectroscopic total transmission curves (dashed lines in the right panel of Figure 1), and select the galaxy samples as extended sources with $GV \lesssim 22.5$, which is about one magnitude shallower than the 5σ point-source limit of the GV band. We then obtained about 19,500 galaxies and corresponding SEDs. The mock galaxy redshift distribution of the CSST spectroscopic survey is shown in the left panel of Figure 3. As can be seen, it has a peak at $z = 0.3$–0.4 and can extend to $z = 1.5$. We find that both the number density and the redshift distribution are consistent with the results derived from the zCOSMOS survey (Lilly et al. 2007, 2009) given in Gong et al. (2019).

Third, we need to consider instrumental effects and add noise to the SEDs. This requires an analysis and simulation of the observational process. Since the CSST spectroscopic observation is performed by slitless gratings, the galaxy light will be dispersed and form multiorder spectra with low wavelength resolution on the detector. Only the first-order spectrum is considered in this work. In addition, galaxy spectra have 2D features on the detector, which are along and perpendicular to the dispersion direction. In addition, the PSF also needs to be included to simulate the observed spectrum. Hence, the dispersed galaxy spectral image on the detector can be estimated as (Ubeda et al. 2011; Cao et al. 2018)

$$\mathrm{SI}(x, y, \lambda) = I(x, y) \otimes \mathrm{SED}_{\mathrm{pix}}(x, y, \lambda) \otimes \mathrm{PSF}(x, y), \quad (1)$$

where x and y denote the position on the detector, and λ is the observed wavelength. $I(x, y)$ is the normalized galaxy surface energy distribution function, which can extend and redistribute a galaxy 1D SED into a 2D strip based on the galaxy spatial energy profile. $\mathrm{PSF}(x, y)$ is the point-spread function, which is assumed to be a 2D Gaussian distribution with the radius of 80% energy concentration $R_{\mathrm{EE80}} \lesssim 0\farcs 3$ (about 4 pixels on the detector; Gong et al. 2019). $\mathrm{SED}_{\mathrm{pix}}(x, y, \lambda)$ is the observed detector-pixelized SED with low CSST spectral resolution. It can be derived in a few steps. We first convolve the high-resolution SED obtained from the COSMOS catalog with the CSST spectroscopic total transmissions of the GU, GV, and GI bands. Then, we degrade the "observed" high-resolution SED to the CSST low-resolution SED. For the CSST spectroscopic survey, a spectral resolution unit $\Delta\lambda$ is designed to cover on average about 4 pixels on the detector for all three spectroscopic bands. The CSST spectral resolution is defined as $R = 241$ at 337.5 nm for the GU band, 263 at 525 nm for the GV band, and 270 at 810 nm for the GI band. Therefore, the $\Delta\lambda$ above are 1.4 nm, 2.0 nm, and 3.0 nm for the GU, GV, and GI bands, respectively, by which the SED will be assigned as a 1D pixel array. In this way, we can find the relation between
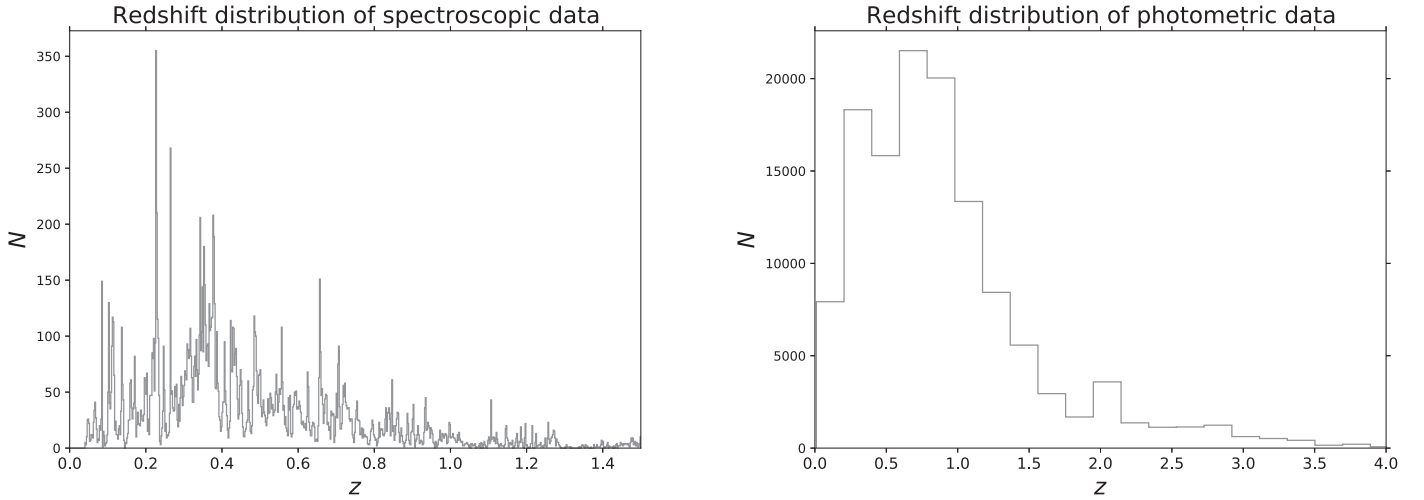
Figure 3. The galaxy redshift distributions of the CSST spectroscopic (left panel) and photometric (right panel) surveys derived from the COSMOS catalog. These redshift distributions are consistent with previous studies given by Cao et al. (2018) and Gong et al. (2019). The redshift bin size is $dz = 0.002$ in the left panel, which matches the dz adopted in the training and testing processes of the CNN as discussed in Section 3.

pixels (or positions on the detector) and wavelengths, and then $\mathrm{SED}_{\mathrm{pix}}(x, y, \lambda)$ can be obtained.

After estimating the observed spectral image $\mathrm{SI}(x, y, \lambda)$, we can evaluate the noise per spectral resolution unit. The S/N can be estimated by (Bohin et al. 2011, Ubeda et al. 2011)

$$\mathrm{S/N} = \frac{C_s t}{\sqrt{C_s t + N_{\mathrm{pix}}(B_{\mathrm{sky}} + B_{\mathrm{det}})t + N_{\mathrm{pix}} N_{\mathrm{read}} R_n^2}}, \quad (2)$$

where C_s is the electron counting rate in $\mathrm{e}^- \, \mathrm{s}^{-1}$, which can be estimated by

$$C_s = A_{\mathrm{eff}} \int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} S_{\mathrm{obs}}(\lambda) \tau(\lambda) \frac{\lambda}{hc} d\lambda, \quad (3)$$

where h and c are the Planck constant and speed of light, respectively, A_{eff} is the effective telescope aperture area for a band, $S_{\mathrm{obs}}(\lambda)$ is the observed SED, and $\tau(\lambda) = m_{\mathrm{eff}}^s(\lambda) T_s(\lambda)$ is the system throughput for the CSST spectroscopic survey. Here, $T_s(\lambda)$ is the total transmissions of the CSST spectroscopic filters. The term m_{eff}^s is the mirror efficiency, and it is about 0.6 for GU, 0.8 for GV, and 0.8 for GI. It is the sum of the signal from all pixels for a spectral resolution unit $\Delta\lambda = \lambda_{\mathrm{max}} - \lambda_{\mathrm{min}}$. Since the regions that generate emission lines in a galaxy are usually small, for simplicity, we assume that most galaxies can be approximated as point sources in the CSST spectroscopic survey. We find that this assumption will not affect our results significantly. Then, the number of pixels N_{pix} in a spectral resolution unit can be estimated as 4 (along the dispersion direction) $\times$ 8 (perpendicular to the dispersion direction) = 32 within the R_{EE80}.[18] The exposure time t is given by $150\,\mathrm{s} \times 4 = 600\,\mathrm{s}$ (Gong et al. 2019). B_{sky} is the sky background, which is given by

$$B_{\mathrm{sky}} = A_{\mathrm{eff}} \int I_{\mathrm{sky}} l_p^2 \tau(\lambda) \frac{\lambda}{hc} d\lambda, \quad (4)$$

where I_{sky} is the surface brightness of the sky background in $\mathrm{erg}\,\mathrm{s}^{-1}\,\mathrm{cm}^{-2}\,\mathring{A}^{-1}$, which mainly includes zodiacal light and earthshine components (Ubeda et al. 2011), and $l_p = 0''074$ is the

pixel scale of the detector. We find that B_{sky} is 0.018, 0.212, and 0.288 $\mathrm{e}^-\,\mathrm{s}^{-1}\,\mathrm{pixel}^{-1}$ for the GU, GV, and GI bands, respectively. $B_{\mathrm{det}} = 0.02\,\mathrm{e}^-\mathrm{s}^{-1}\,\mathrm{pixel}^{-1}$ is the detector dark current, $N_{\mathrm{read}} = 4$ is the number of detector readouts, and $R_n = 5\,\mathrm{e}^-\,\mathrm{pixel}^{-1}$ is the read noise. More details can be found in Cao et al. (2018) and Gong et al. (2019). Lastly, the SED noise per resolution unit can be obtained by calculating the denominator of Equation (2).

In Figure 4, we show the S/N distributions of the three CSST spectroscopic bands and the average total S/N for all three bands of all galaxies, $\overline{\mathrm{S/N}_{\mathrm{tot}}}$, at z in the top left and right panels, respectively. In the top left panel, we find that the peaks of the S/N distributions for the three bands are around $\mathrm{S/N} = 1$, and there are a few samples at $\mathrm{S/N} > 6$. In the top right panel, $\overline{\mathrm{S/N}_{\mathrm{tot}}}$ can be as high as 4 at $z \sim 0$ and slowly declines toward ~ 1 at $z = 1.5$. This indicates that most galaxies have relatively low S/N in the CSST spectroscopic survey, and it is probably hard to extract their redshifts accurately . This is more obvious in Figure 5, where we show examples of galaxy slitless spectra for the CSST spectroscopic survey. As can be seen, the noise can be very strong and dominant for the high-z galaxy samples, and the emission lines are easily submerged by the noise at low S/N. This makes it quite challenging for the ordinary SED template-fitting method to derive accurate redshifts from the data, and it is necessary to explore other methods (like neural networks) here.

Note that the spectra measured by slitless spectroscopic surveys can suffer from an overlapping problem, such that a dispersed multiorder spectrum may overlap with parts of another nearby spectrum, especially in crowded fields. This effect can cause trouble in the spectrum extraction and can suppress the spectral quality and the fraction of usable spectra. However, it is not a main issue in this work, since we assume that the galaxy spectra have already been obtained by previous data processing stages, and regardless of overlapping, the spectral quality can be reflected by the S/N. A few spectra suffering severe overlapping (e.g., in crowded fields) can be simply discarded in the analysis.

2.2. Photometric Mock Data

The CSST has seven photometric imaging bands covering a similar wavelength range as the spectroscopic survey from the

---

[18] Based on the instrumental design, the CSST R_{EE80} takes about 4 pixels, so that the diameter takes 8 pixels.
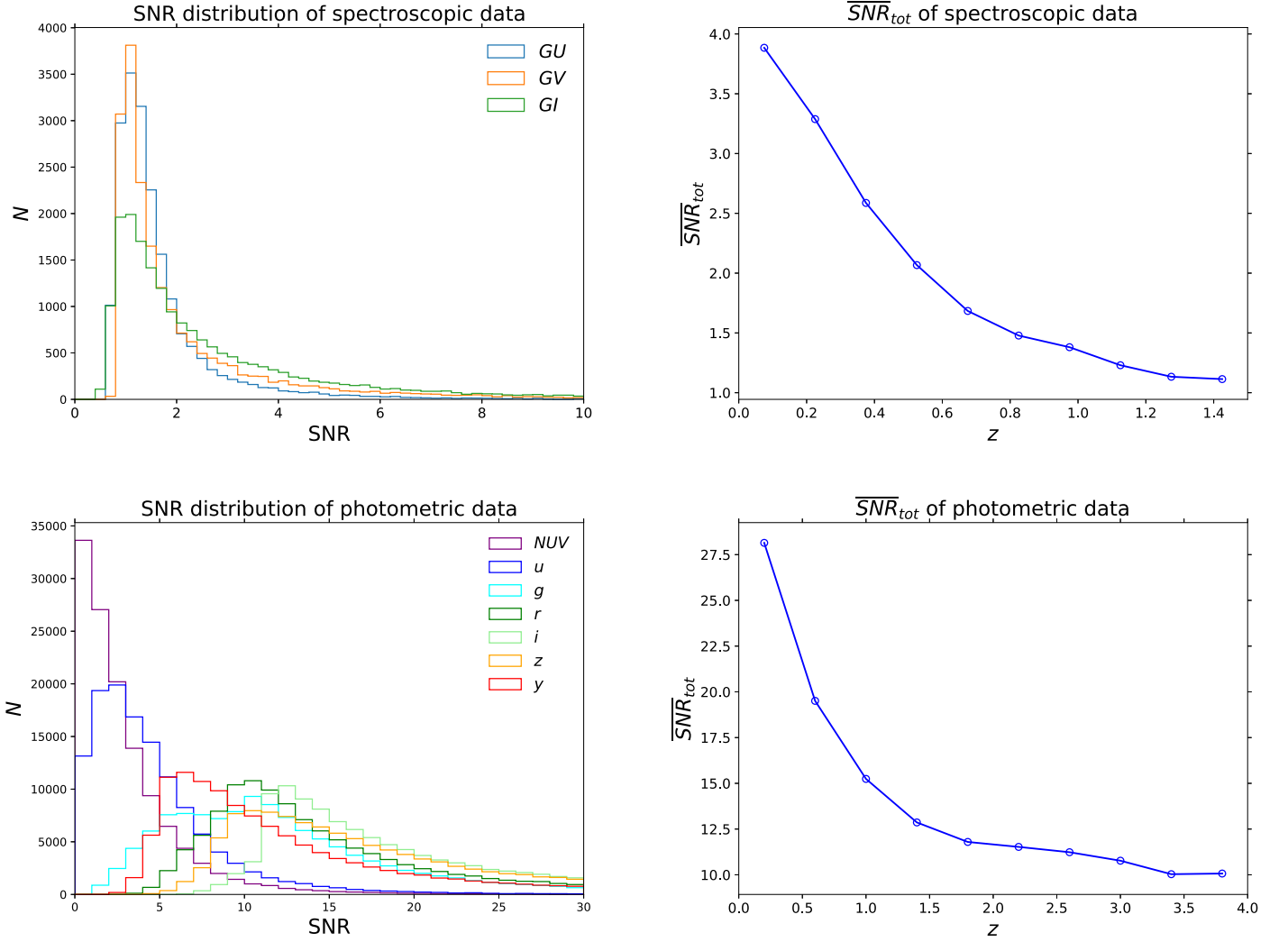
Figure 4. Top panels: the S/N distributions of the mock data for the CSST three spectroscopic bands, and the distribution of the average total S/N of all galaxies for all the three bands at *z*. Bottom panels: the same cases for the CSST photometric survey.
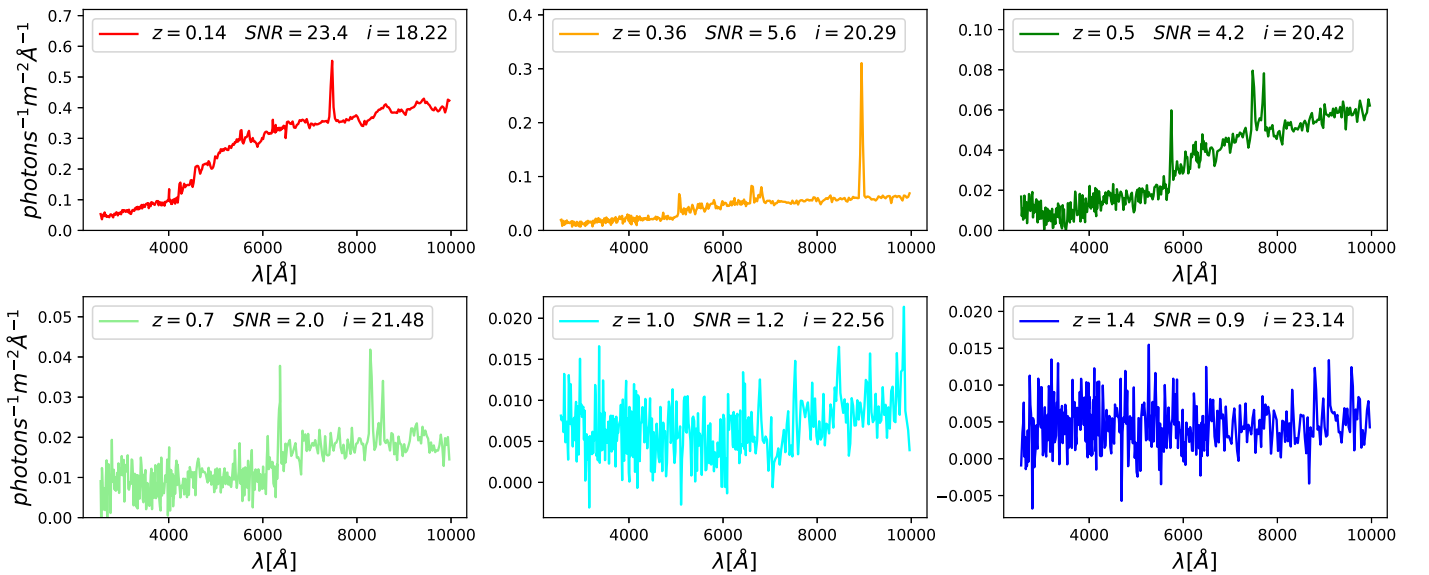

Figure 5. Examples of mock slitless spectra measured by the CSST spectroscopic survey at different redshifts. As can be seen, the spectra with low S/N are dominated by noise; the emission lines are not obvious and are even submerged by noise. In these cases, the SED template-fitting method may not work well, and other methods, like neural networks, need to be considered.
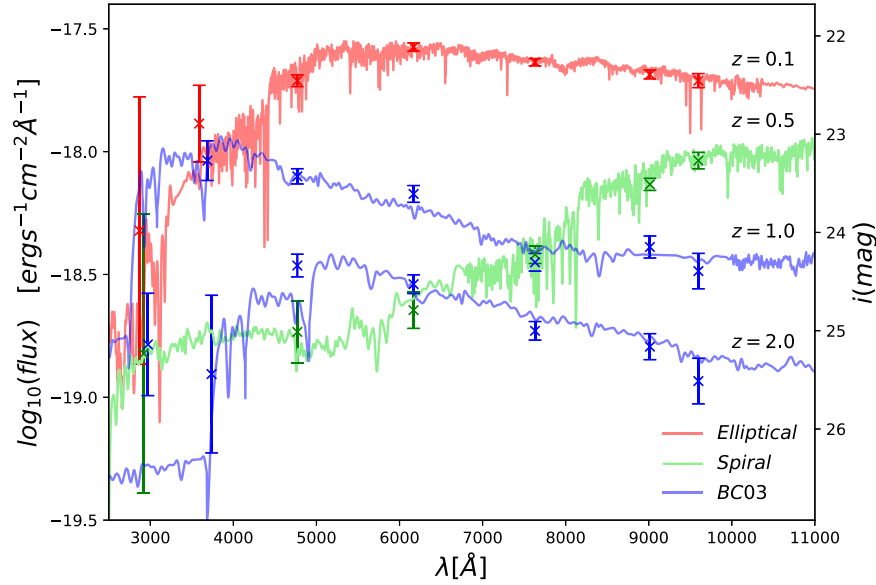
Figure 6. Examples of mock flux data at $z = 0.1$, 0.5, 1.0, and 2.0 for the seven bands of the CSST photometric survey. For comparison, the corresponding SEDs are also shown. The flux data are shifted by a random draw from a Gaussian distribution with $\sigma = \sigma_{\rm m}$.

near-UV to near-IR, *NUV*, *u*, *g*, *r*, *i*, *z*, and *y*, as shown in solid lines in Figure 1. The magnitude 5σ limit for a point source can be as high as $\sim$26 AB mag for the *g*, *r*, and *i* bands, and is about 24.5–25.5 for the other bands. The spatial resolution of the CSST photometric survey is $0.''15$ within 80% energy concentration for the *NUV*, *u*, *g*, and *r* bands, $0.''16$ for the *i* band, and $0.''18$ for the *z* and *y* bands. More details can be found in Cao et al. (2018) and Gong et al. (2019).

We follow Cao et al. (2018) to generate the mock flux data and errors for the seven photometric bands. First, we reproduce the best-fit galaxy SEDs given in the COSMOS catalog and extend their wavelength range from $\sim$900 Å to $\sim$90 Å to match the CSS-OS large wavelength coverage (Cao et al. 2018). Second, we add the dust extinction from interstellar dust in galaxies and absorption by the intergalactic medium (IGM) to the galaxy SEDs we obtained. Five extinction laws are considered for the interstellar dust origin, including the ones derived from the Milky Way (Allen 1976; Seaton 1979), the Large Magellanic Cloud (Fitzpatrick 1986), the Small Magellanic Cloud (Prevot et al. 1984; Bouchet et al. 1985), and a starburst galaxy (Calzetti et al. 2000). For the IGM absorption, we use the attenuation laws computed by Madau (1995). Third, we rescale the obtained SEDs based on the Subaru i^+ band magnitude, and convolve them with the CSST filter transmission curves (solid lines in the right panel of Figure 1) to get the mock CSST photometric flux data. Note that, similar to the CSST spectroscopic survey, the SED here is the observed SED without instrumental effects, and can be seen as high-quality spectral data obtained by powerful spectroscopic surveys.

Next, we can estimate the flux errors for the seven bands. The same S/N formula shown in Equation (2) is used in the calculation. Here, the electron count rate can be estimated by Equation (3), but with $\tau(\lambda)$ replaced by $\tau_{\rm p}(\lambda) = m_{\rm eff}^{\rm p}(\lambda)T_{\rm p}(\lambda)$, which is the system throughput for the CSST photometric survey. $T_{\rm p}(\lambda)$ is the total transmission of the CSST photometric filters. The term $m_{\rm eff}^{\rm p}$ is the mirror efficiency, and it is about 0.5 for *NUV*, 0.7 for *u*, and 0.8 for the other bands. Wavelengths $\lambda_{\rm min}$ and $\lambda_{\rm max}$ cover the whole band range. The exposure time $t = 150\,{\rm s} \times 2 = 300\,{\rm s}$, the number of detector pixels covered by

an object $N_{\rm pix}$ is derived from the galaxy size, the number of detector readout $N_{\rm read} = 2$, and the values of $B_{\rm det}$ and $R_{\rm n}$ are the same as for the CSST spectroscopic survey. The sky background $B_{\rm sky}$ here can be estimated by Equation (4) using $\tau_{\rm p}(\lambda)$. We find that $B_{\rm sky}$ are 0.003, 0.018, 0.156, 0.200, 0.207, 0.123, and 0.036 $\rm e^-\,s^{-1}pixel^{-1}$ for *NUV*, *u*, *g*, *r*, *i*, *z*, and *y* bands, respectively. Then, the photometric error can be calculated as $\sigma_{\rm p} \simeq 2.5\log_{10}(1 + 1/{\rm S/N})$ mag (Pozzetti et al. 1996, 1998; Bolzonella et al. 2000). We assume the systematic error $\sigma_{\rm sys} = 0.02$, which can be seen as a systematic floor for the photometric surveys (e.g., see Astier et al. 2006; Sullivan et al. 2006). The total magnitude error is given by $\sigma_{\rm m} = \sqrt{\sigma_{\rm ph}^2 + \sigma_{\rm sys}^2}$ (Cao et al. 2018).

Since accurate photo-z is needed in the weak lensing and 2D galaxy clustering analysis, only high-quality photometric data are selected in real observations. Here, we select the flux data with ${\rm S/N} \geqslant 10$ for *g* or *i* band, and about 126,000 galaxies are obtained (Cao et al. 2018). In the right panel of Figure 3, the redshift distribution of the selected mock galaxy sample for the CSST photometric survey is shown, which has a peak at $z = 0.7$ and can extend to $z = 4$. In the bottom panels of Figure 4, we show the S/N distributions for the seven photometric bands, and the average total S/N (i.e., $\overline{\rm S/N}_{\rm tot}$) for all bands and all galaxies at z. In the bottom left panel, we find that the peaks of the S/N distributions are less than 5 for the *NUV* and *u* bands, $\sim$7 for the *y* band, and $\sim$10–13 for the *g*, *r*, *i*, and *z* bands. In the bottom right panel, $\overline{\rm S/N}_{\rm tot}$ can be greater than 25 at $z < 0.5$ and declines to about 10 (the S/N cut for *g* or *i* band) at $z = 4$. In Figure 6, we show examples of mock flux data for the seven CSST photometric bands at $z = 0.1$, 0.5, 1.0, and 2.0. The corresponding observed SEDs are also shown for comparison. We can see that, except for the *NUV* and *u* bands, the flux data have relatively small errors, and they can well represent the features of galaxy SEDs.

3. Neural Network

After obtaining the mock data of the CSST spectroscopic and photometric surveys, we can explore the estimation and
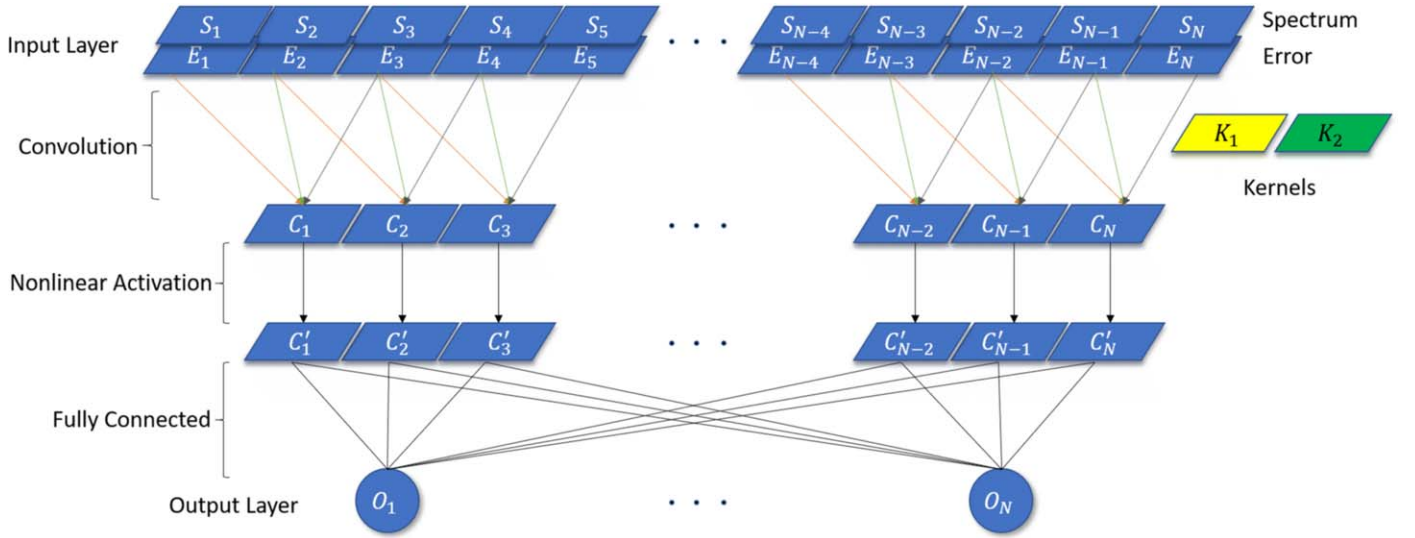
Figure 7. The architecture of 1D CNN used in the CSST spec-z analysis. The input layer contains two channels, spectral data and errors. Two convolutional layers are considered, which are operated by two and four kernels, respectively, to obtain feature maps. Two fully connected layers are included to perform the spec-z analysis as a classification task. The output layer contains 750 neurons for classification, which can provide the probability for each redshift interval $dz = 0.002$ from $z = 0$ to 1.5. Only one convolutional layer and one fully connected layer are shown here as an illustration.

accuracy of the spec-z and photo-z using the neural network. The CNN and MLP are adopted in the spec-z and photo-z analysis, respectively, and the details of their architectures and training processes are discussed in this section.

3.1. Network Architecture

The CNN and MLP adopted in this work are implemented by Keras[19], a high-level neural network application programming interface (API) running on top of TensorFlow.[20] We use 1D CNN to perform the spec-z analysis for the CSST slitless spectroscopic survey. The CNN has advantages of capturing local information and reducing the model complexity and overfitting problem, so it is suitable and widely used in the extraction of feature information from images. Since galaxy spectra can be seen as 1D images, we apply the 1D CNN in the spec-z estimation.

In Figure 7, the architecture of the 1D CNN adopted in this work is shown. The 1D CNN is mainly composed of an input layer, convolutional layers, fully connected layers, and an output layer. In the input layer, we make use of two channels, i.e., spectral data and errors. As we can see in the later discussion, the errors as input data can significantly improve the accuracy of the spec-z. In total, we have 369 data points for each galaxy spectrum, which has wavelength intervals of 14 Å, 20 Å, and 30 Å for the GU, GV, and GI bands, respectively.

We set two 1D convolutional layers, and this is found to be good enough to derive the results (i.e., more layers get similar results). For the first and second convolutional layers, two and four kernels are employed to extract useful features of the data, respectively. All kernels have the same 1×5 spatial dimension. To avoid losing marginal information, we set the same padding on the first convolutional layer to obtain a feature map with the same number of elements as the input data. After each convolutional layer, the Rectified Linear Unit (ReLU) is applied as a nonlinear activation function (Nair & Hinton 2010),

which is given by $y = \max(0, x)$, to implement nonlinearity, reduce the vanishing-gradient problem, and speed up the training process. Next, the pooling layer is used to downsample the feature map. The max pooling method is adopted here, such that the maximum value is chosen in two adjacent elements. Then we obtain a feature map with a dimension of 90×4. Next, flattening is applied to get a 1D vector for connecting with the fully connected layer by trainable weights.

We structure two fully connected layers, and the ReLU is applied after each. In the training process, in order to reduce overfitting, a dropout layer is added with a keeping rate of 0.5 after each fully connected layer (Srivastava et al. 2014), which means half of the neurons are temporarily turned off in training. As we discuss later, the dropout can affect the accuracy in the training and validation process. The fully connected layer also can be seen as a classifier, which can deal with the spec-z analysis as a classification task (Stivaktakis et al. 2019). In detail, the CSST spectroscopic redshift range $z = 0$–1.5 is divided into 750 bins or classes with redshift interval $dz = 0.002$.[21] Correspondingly, in the output layer, 750 neurons are assigned with the multiclass activation function, i.e., softmax function, which takes the exponential form $S_i = \exp(V_i) / \sum_j \exp(V_j)$. Here, V_i is the value of ith neurons of output layer. The softmax function can give a probability for each class, and the sum is equal to 1. The details of layers, e.g., number of data points, neurons, and parameters, can be found in Table 1.

The MLP is adopted to derive photo-z from the CSST photometric survey, since the CSST photometric data only have seven data points for each galaxy and are much simpler than the spectroscopic data. We show the MLP architecture and details in Figure 8 and Table 2. In the input layer, we consider the magnitudes, errors, and colors for the seven CSST

---

[21] We notice that, as shown in the left panel of Figure 3, in some of these narrow redshift bins at $z \gtrsim 1.3$, the number of galaxies may be close to zero, which can affect the training process and hence the spec-z estimate in these bins. However, since there are not many spectroscopic sample in this range for the CSST, and the data from other spectroscopic surveys can be used as training sample, this problem is solvable.

Table 1
Details of 1D CNN used in the CSST Spec-z Analysis

Layers	Output Status[a]	Number of Paras[b]
Input	(369,2)	0
Conv1D (2,5)[c]	(369,2)	22
ReLU	(369,2)	0
MaxPooling	(184,2)	0
Conv1D (4,5)[d]	(180,4)	44
ReLU	(180,4)	0
MaxPooling	(90,4)	0
Flatten	(360,1)	0
FC[e]	(256,1)	92,416
ReLU	(256,1)	0
FC[e]	(256,1)	65,782
ReLU	(256,1)	0
Output	(750,1)	192,750

Notes.
[a] The format is: number of data points or neurons, number of channels or dimensions.
[b] Total number of parameters: 351,024.
[c] Two convolution kernels with spatial dimension 1×5.
[d] Four convolution kernels with spatial dimension 1×5.
[e] FC: fully connected layer.

Figure 8. The architecture of the MLP used in the CSST photo-z estimation. The values included in the input layer are seven magnitudes and errors, and six colors. Two hidden layers are structured to predict photo-z given by the output layer.

Table 2
Details of the MLP used in the CSST Photo-z Analysis

Layers	Output Status[a]	Number of Paras[b]
Input	20	0
FC[c]	40	840
ReLU	40	0
FC[c]	40	1640
ReLU	40	0
Output	1	41

Notes.
[a] Number of data points or neurons.
[b] Total number of parameters: 2521.
[c] FC: fully connected layer.

photometric bands, which give 20 data points for each galaxy. Two hidden layers are used here. Since we apply the classic structure with an n: $2n$: $2n$: 1 feature, where n is the number of data elements, there are 40 neurons in each hidden layer. The ReLU is also applied after each hidden layer. Then, the photo-z will be obtained in the output layer.

3.2. Training Process

As we discussed in the last section, the majority of the CSST slitless spectral data may have relatively large noise and low S/N (i.e., $S/N < 3$), especially for galaxies at high redshifts (see the top panels of Figure 4). This means that there will be large deviation of the spectral data points from the real spectral curve, such that multiple measurements may result in distinctly different shapes for the same spectrum or spectral type due to statistical reason. This problem can be serious for the ordinary SED template-fitting method, which may lead to different spec-z predictions for the same galaxy with large errors. However, since all of these measured spectral data with different shapes have statistical origins, we can generate mock data to simulate this feature and input them into the neural networks as a training sample to reduce this effect. Looking from another aspect, these realizations also can be seen as spectra measured from different galaxies with similar spectral types at the same redshift, which can dramatically expand the training sample.

Since the spectral data with high S/N and resolution are needed to generate the training data for the CSST slitless spectroscopic survey, it could be still hard to achieve the requirements for even the future powerful spectroscopic surveys at high 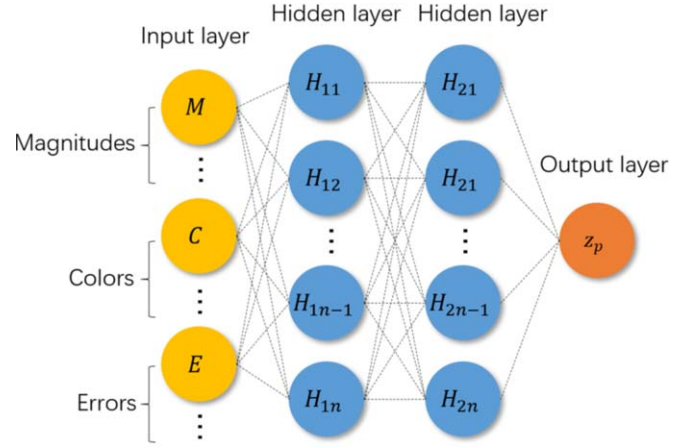redshifts $z > 1$. Therefore, here we make a conservative assumption that the redshift distribution of the high-quality spectral data used to generate the training data is similar to that of the CSST spectroscopic survey. This means we assume there are not many data that can be used to train the network at $z > 1$, and we will generate the training data directly from our mock CSST spectral catalog. If more high-quality spectral data at high-z can be used for training, the results will be undoubtedly improved.

We divide the mock spectral data into two parts, i.e., training and testing data sets. We randomly select 15,000 galaxies from the mock spectral data set as a training sample and the rest of the $\sim$4500 are for testing. Assuming that the deviation of a spectral data point at each wavelength follows a Gaussian distribution with the spectral error as the 1σ error, we randomly generate 50 realizations for each spectrum of the training set. We have checked that more realizations would not improve the result significantly, and the results would not considerably change if the number of realizations is greater than 30. In this training process, the categorical cross-entropy loss function is optimized by an Adam optimizer (Kingma & Ba 2014). It is an efficient method for stochastic optimization that only requires first-order gradients and can adjust the learning rate automatically. This makes our code more efficient than the current machine-learning codes used in redshift estimation, such as
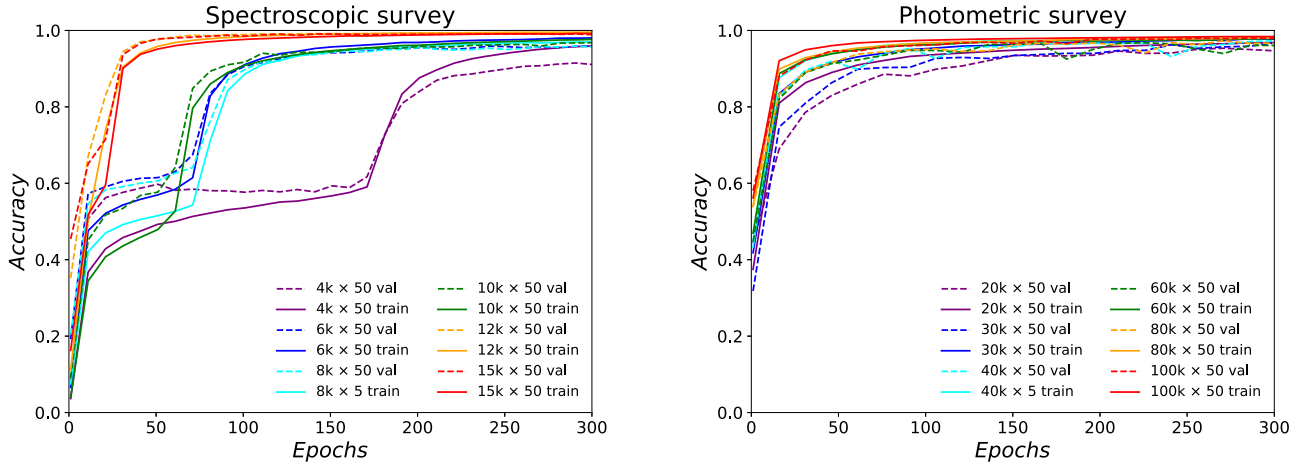
Figure 9. Left: the fraction of galaxies that can reach the accuracy criterion for the training and validation samples as a function of training epochs in the CSST spectroscopic survey. The results of different numbers of data from 4000×50 to $15,000 \times 50$ have been shown for the training (solid) and validation (dashed) data sets. A total of 50 realizations are generated from a Gaussian distribution based on the original data sets. Right: the results for the photometric survey, and the results for the numbers of training data are from $20,000 \times 50$ to $100,000 \times 50$ have been shown.

ANNz (Collister & Lahav 2004), ArborZ (Gerdes et al. 2010), and CuBANz (Samui & Pal 2017), in the training process. Based on the pre-running tests, we set the maximum epochs of the neural network to be 300 as the termination criteria. Our network can save the best model of these epochs with the highest accuracy during training, which will be applied to the testing sample.

To show the training process clearly, we define an accuracy criterion as $|z_{\mathrm{pred}} - z_{\mathrm{true}}|/(1 + z_{\mathrm{true}}) \leqslant 0.002$, where z_{pred} and z_{true} are the predicted and true redshifts, respectively. In the left panel of Figure 9, we show the fraction of the training and validation sample that can achieve the accuracy criterion as a function of training epochs for different sizes of training data. The validation sample used here is set to be the same as the testing sample, which contains 50 Gaussian random realizations of ~ 4500 original galaxy spectra. Note that the number of realizations of the validation data is not necessarily the same as that of the training data. No matter how many realizations of the validation data we created, the results would not change as long as the realizations of the training data are more than $30 \sim 50$. The results for training and validation data are shown in solid and dashed curves, respectively. We find that the accuracy fractions of all training data sets we check can approach 1 at epoch $\simeq 300$, which means that the spec-z can be efficiently derived to high accuracy by our network. More training data can speed up this process, and it converges when the number of training data reaches $\sim 12,000 \times 50$. That said, fewer training data will take more epochs, and may experience a "metastable state", but can reach higher accuracy in later epochs.

In addition, we also notice that the accuracy of the validation data can be higher than that of the training data in early epochs. This is mainly due to two reasons. The main reason is that the dropout layers are only applied in the training stage, and only half of the neurons are used to derive spec-z when training, while all neurons are used in the validation. Hence, in principle, the network is more powerful in the validation and can obtain a higher accuracy at the same epoch. The second reason is caused by the difference when calculating the training and validation accuracy. The training accuracy is estimated by averaging the accuracies of all batches in an epoch. Generally speaking, the training accuracy over the first batches are lower than those

over the last batches. However, the validation accuracy is obtained using the model generated at the end of epoch, resulting in higher accuracy. In the early epochs, these effects are more obvious, since the network is not well trained, while they can be effectively relieved in the late epochs.

For the CSST photometric survey, the spectral data at high redshifts, $z \sim 4$, could be found in the future high-resolution spectroscopic surveys. As for the spectroscopic survey, we assume the redshift distribution of the high-quality spectral data used to generate the photo-z training data is similar to that of the CSST photometric survey, and that it can be obtained from the CSST mock photometric catalog. We divide the mock data set of the CSST photometric survey into the training and testing samples. We randomly select 10,000 galaxy spectra as testing data, and the rest of the $\sim 116,000$ are employed as training data. Then, 50 Gaussian random realizations are created for each galaxy using the same method as for the spectroscopic survey, and more realizations would not change the results significantly. The accuracy criterion defined in the photometric survey is given by $|z_{\mathrm{pred}} - z_{\mathrm{true}}|/(1 + z_{\mathrm{true}}) \leqslant 0.05$. In the right panel of Figure 9, we show the accuracy of training and validation samples versus the training epochs for different numbers of training data. We find that the accuracies of all data sets approach unity at early epochs $\lesssim 100$, which means the network is efficient at finding the correct photo-z to high accuracy. Note that the problem of the training accuracy being less than the validation accuracy does not appear here, since no dropout layer is structured in the MLP.

4. Result

In order to assess the accuracy of the derived spec-z and photo-z, we define the catastrophic redshift or outlier as $|\Delta z|/(1 + z_{\mathrm{true}}) > D_{\mathrm{out}}$, where $\Delta z = z_{\mathrm{pred}} - z_{\mathrm{true}}$, and z_{pred} and z_{true} are the predicted and true redshifts, respectively. $D_{\mathrm{out}} = 0.02$ and 0.15 for the spec-z and photo-z analysis, respectively. The outlier fraction is denoted as η here. We also adopt the normalized median absolute deviation (NMAD, Brammer et al. 2008), and it is given by

$$\sigma_{\mathrm{NMAD}} = 1.48 \times \mathrm{median}\left(\left| \frac{\Delta z - \mathrm{median}(\Delta z)}{1 + z_{\mathrm{true}}} \right|\right). \quad (5)$$
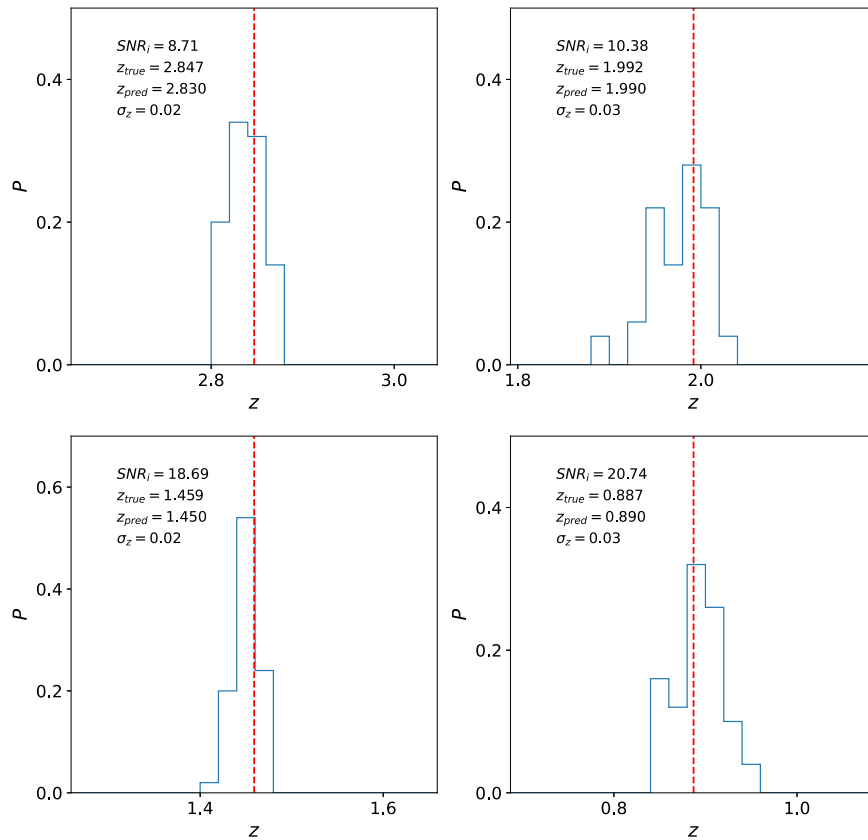
Figure 10. Examples of the photo-z PDFs derived from our neural networks for different i-band S/Ns. The true photo-z is denoted by the vertical red dashed line. We find that the width of the PDF is not sensitive to the S/N.

This definition has advantages in that it can naturally suppress the weights of catastrophic redshifts and can provide a proper estimate of the redshift accuracy. In addition, the PDFs, best fits, and 1σ errors of the predicted redshifts are also derived, based on the uncertainties of the testing data. We randomly generate 50 samples from a Gaussian distribution based on the flux and error at a wavelength for each testing data set, and input them into the networks to get 50 predicted redshifts. Then the PDF, best fit (the value at the PDF peak), and 1σ error of the redshift can be derived from the number distribution of these redshift points. Examples of photo-z PDFs for different i-band S/Ns are shown in Figure 10. The PDFs of spec-z are quite narrow spikes with width $dz = 0.002$. We find that 50 random samples are sufficient to produce the PDFs; more samples would not change the result significantly. Note that we do not consider the uncertainty from the neural network, since we have chosen the best network model, which gives the most accurate redshift estimate from a set of well-trained networks, and can effectively reduce this effect from the network.[22]

In Figure 11, we show the best-fit spec-z results of the spectroscopic testing data for different numbers of training samples, which are randomly selected from the whole training data set. The 1σ errors, σ_z of the predicted spec-z, are shown in gray bars. Both the best fits and errors are derived from the spec-z PDFs obtained by the network. The root mean square

error (RMSE) for the whole sample is also calculated to further indicate the effect of an outlier. The results of the spectroscopic data with different S/N ranges are also shown and denoted by different colors. The fractions of the data with total $S/N < 1$, $1 \leqslant S/N < 3$, $3 \leqslant S/N < 5$, $5 \leqslant S/N < 10$, and $S/N \geqslant 10$ are 6.3%, 71.2%, 13.0%, 7.6%, and 1.9%, respectively. As we can see, most of the CSST slitless spectral data have S/N less than 3, which is challenging for redshift derivation. However, we find that the redshifts obtained from the low S/N data still have high accuracy, with $\sigma_{\rm NMAD} \simeq 0.001$, which is comparable to that of the high S/N data. This is because we have sufficiently considered the statistical effects resulting from data with large noise levels and have generated the corresponding noisy data to train the networks. This indicates that the neural network has advantages to derive spec-z from low S/N data, so long as we can generate a sufficient training sample and train it with real-enough noisy spectra.

We also find that more training data can give better results with smaller outliers, and the results will be good enough and would not notably change when the number of training data is greater than 10,000 with 50 Gaussian random realizations. We also test the case in which no errors are inputted into the 1D CNN for training, and find that $\sigma_{\rm NMAD} = 0.002$ and outlier fraction $\eta = 10.12\%$ for $12,000 \times 50$ case, which are much larger than those when errors are included. This implies that the errors contain useful information that can be helpful for network training to derive more accurate redshifts.

In Figure 12, the best-fit photo-z and 1σ errors, σ_z of the photometric testing data, extracted from the MLP for different numbers of training data have been shown. The photo-z best fits and errors are estimated from the photo-z PDFs derived by

---

[22] After performing a few tests, we find that the uncertainty from the networks is always smaller than that from the data, and it would not affect the results if the networks are well trained.
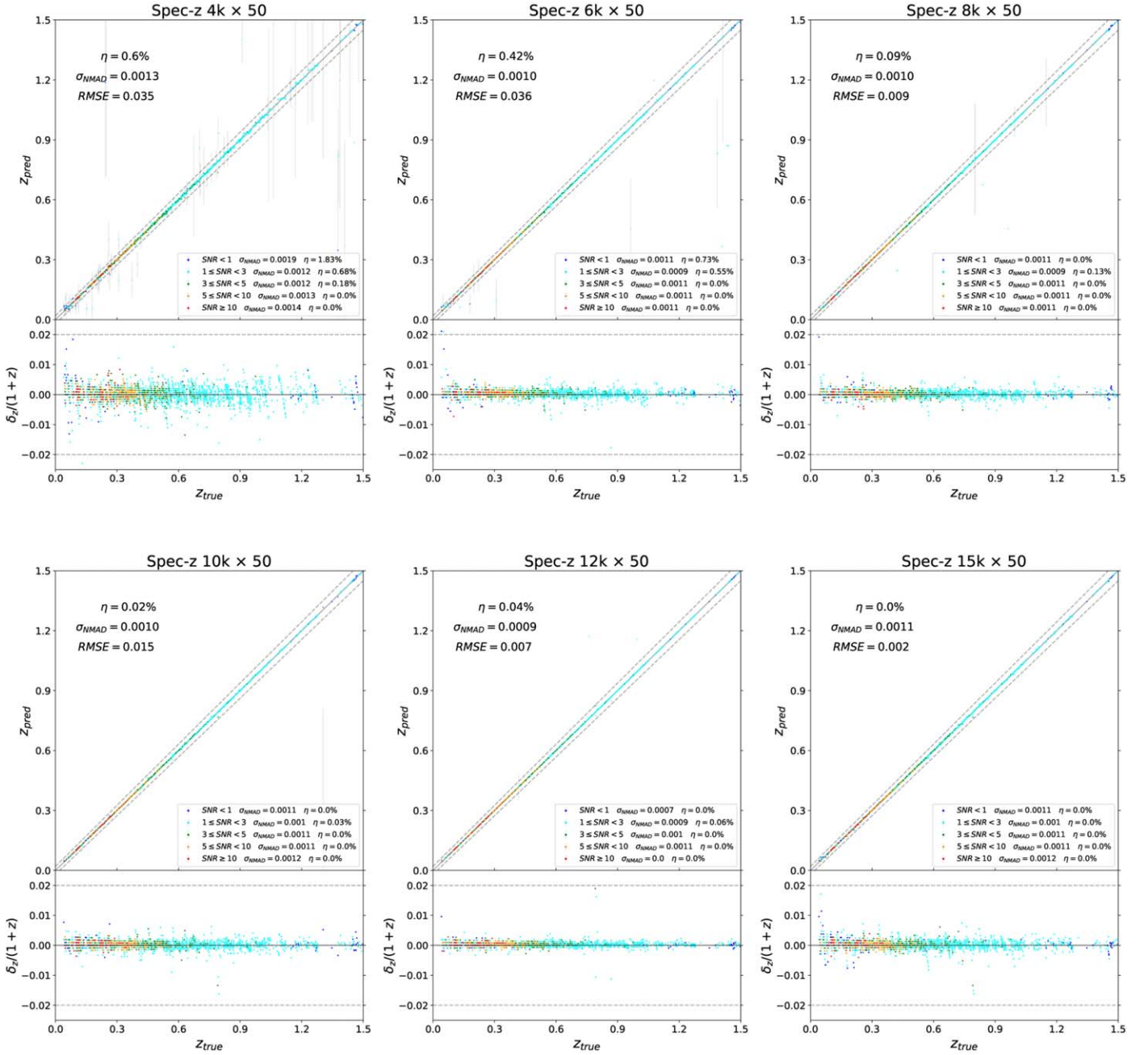
Figure 11. The best-fit spec-z and 1σ errors, σ_z derived from the spec-z PDFs, for the testing data, which are estimated by the 1D CNN for different numbers of training data, i.e., 50 Gaussian random realizations of 4000, 6000, 8000, 10,000, 12,000, and 15,000 original spectral data. The results derived from the data with different total S/N are also shown as blue (S/N < 1), cyan ($1 \leqslant$ S/N < 3), green ($3 \leqslant$ S/N < 5), orange ($5 \leqslant$ S/N < 10), and red (S/N $\geqslant$ 10) dots. The redshift outlier is denoted by dashed lines. We find that the 1D CNN can provide excellent results for spectra of the CSST spectroscopic survey with $\sigma_{\mathrm{NMAD}} \simeq 0.001$, as well as for data with low S/N. More training data tend to get a higher accuracy and smaller outlier fraction η, and it seems to saturate when the number of training data is greater than $\sim 10{,}000 \times 50$.

the network. The results of the data with different S/N ranges are also listed. The percentages of the data with S/N < 10, $10 \leqslant$ S/N < 20, and S/N $\geqslant$ 20 are found to be 2.4%, 53.0%, and 44.6%, respectively. Note that the S/N is only for the i band here; it could be less than 10, since we select the data with S/N $\geqslant$ 10 for the i or g band. We can see that the MLP can provide good predictions for photo-z, and the σ_{NMAD} is close to 0.01 for the number of training data greater than $60{,}000 \times 50$. In addition, the fraction of catastrophic redshifts or outliers is quite low, with $\eta < 0.5\%$ for all cases we explore. The RMSE is also given to show the effect of outliers. We find that our

MLP can provide a similar accuracy for photo-z estimation compared to that of exiting codes (e.g., Collister & Lahav 2004; Gerdes et al. 2010; Samui & Pal 2017). By comparing to the results from the SED template-fitting method given by Cao et al. (2018), we find that our results for σ_{NMAD} are improved by a factor of at least two, and the outlier is dramatically suppressed by more than one order of magnitude. In addition, as for the results for spec-z, the photo-z accuracy for the low-S/N data is similar to that for the high-S/N data, which proves that our method of generating the training data is feasible and effective. If errors are not included in the training process, we
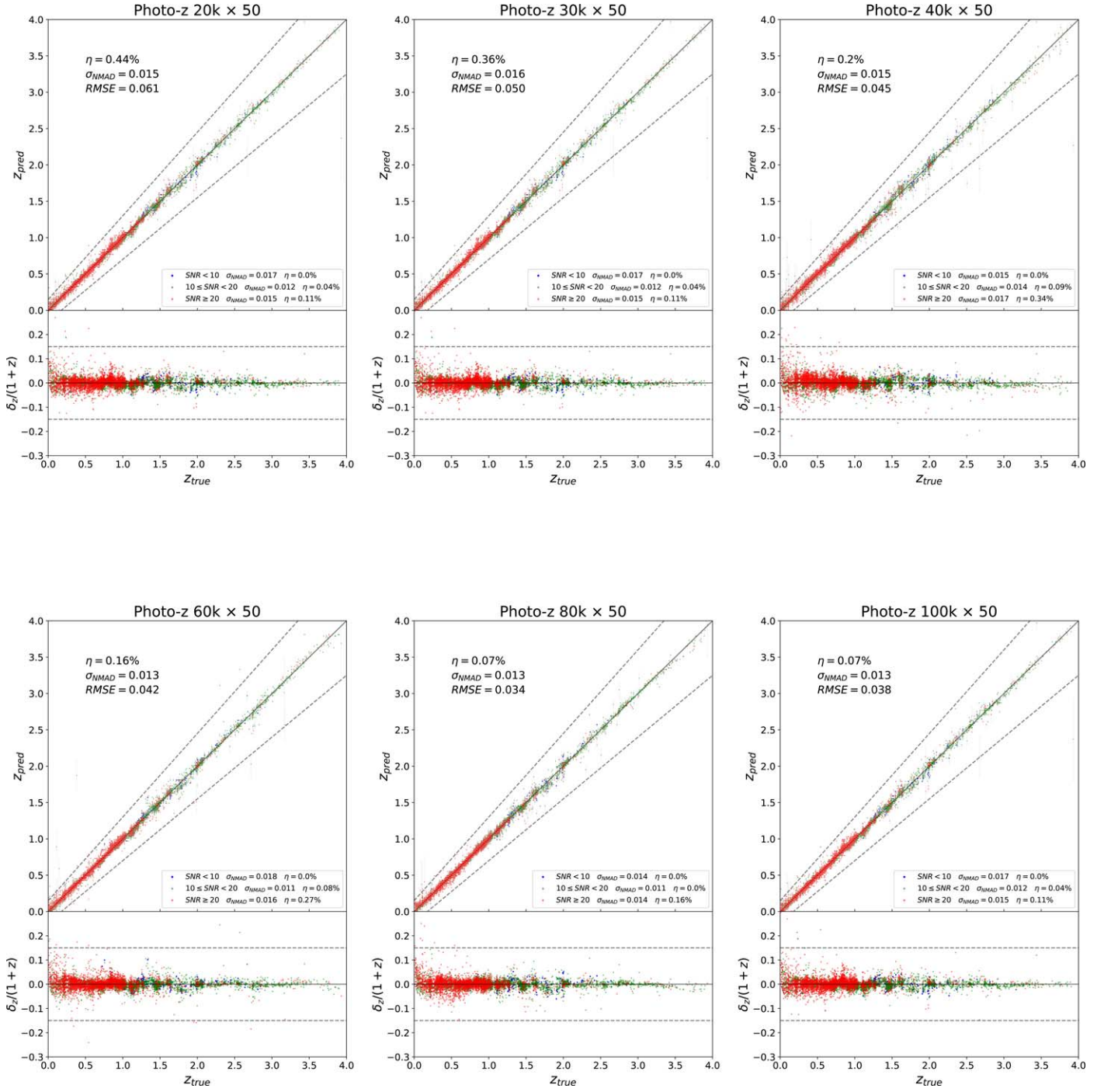
Figure 12. The best-fit photo-z and 1σ errors, σ_z derived from the photo-z PDFs, for the photometric testing data, which is estimated by the MLP for different numbers of training data sets, i.e., 50 Gaussian random realizations of 20,000, 30,000, 40,000, 60,000, 80,000, and 100,000 original photometric data. The results derived from the data with different S/N are also shown as blue (S/N < 10), green ($10 \leqslant$ S/N < 20), and red (S/N $\geqslant$ 20) dots. Note that the S/N is only for the i band here. The redshift outlier is denoted by dashed lines. We can find that the MLP we use can provide accurate photo-z estimates with $\sigma_{\mathrm{NMAD}} \simeq 0.01$, which is smaller than that of the SED template-fitting method by a factor of at least two. In addition, similar to the spec-z case, the results of the low-S/N data are of comparable accuracy to those of the high-S/N data, and basically more training data can obtain a higher accuracy and smaller outlier fraction η when the training data number is less than $\sim$80,000 $\times$ 50.

obtain $\sigma_{\mathrm{NMAD}} = 0.053$ and $\eta = 9.09\%$ for the $80{,}000 \times 50$ case, which proves again that the errors can be helpful to improve the redshift estimate accuracy in the training.

In Figure 13, we show the mean redshift errors $\overline{\sigma_z}$ (the mean 1σ error derived from the PDFs for all galaxies at z) as a function of redshift for different numbers of training data for the CSST spectroscopic and photometric surveys. We find that the mean errors are lower than 0.002 (except for the 4000×50 case) at $z \lesssim 1$ and 0.03 at $z \lesssim 3.5$ for spec-z and photo-z, respectively. At higher redshifts, $\overline{\sigma_z}$ increases quickly, due to the fact that fewer training data are generated and used in these ranges. The spec-z accuracy at $1 < z < 1.5$ actually can be further increased if more high-quality spectra can be used from ongoing or future high-resolution spectroscopic surveys.
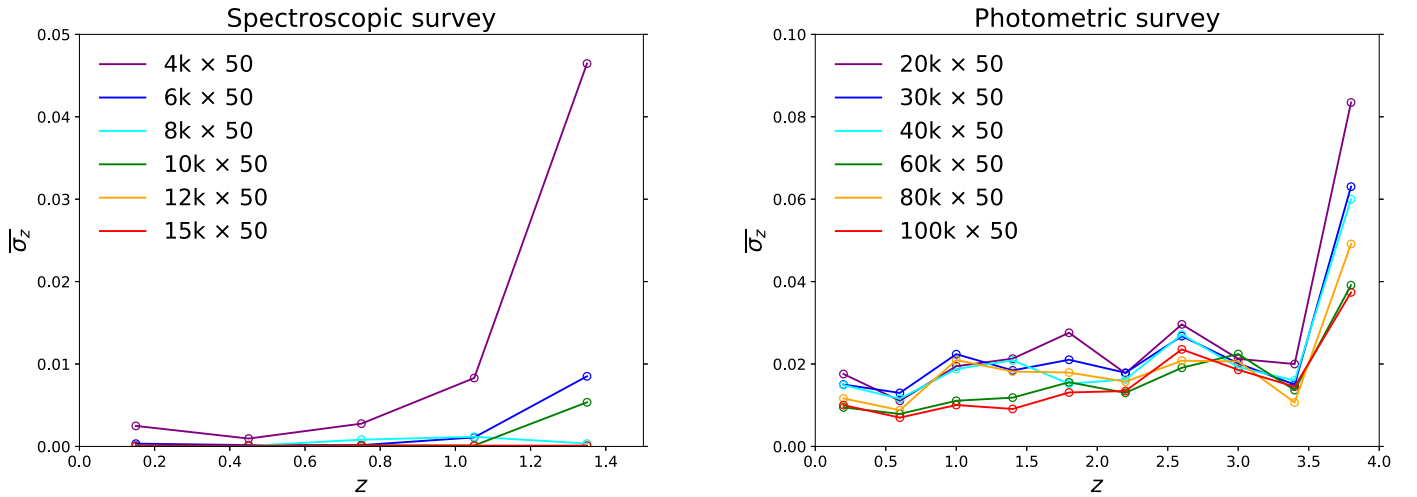
Figure 13. The mean redshift errors $\overline{\sigma_z}$ as a function of z for the spec-z (left panel) and photo-z (right panel) analysis. The redshift errors are averaged over redshift bins $\Delta z = 0.3$ and 0.4 for the spec-z and photo-z surveys, respectively. The results from different numbers of training data have been shown. We find that $\overline{\sigma_z}$ is basically stable and low at $z < 1$ and 3.5 for spec-z and photo-z, respectively, and increases quickly at higher redshifts.

5. Summary

We explore the accuracies of spec-z and photo-z that can be derived from the CSST slitless spectroscopic and photometric surveys by neural networks. The 1D CNN and MLP are adopted to extract spec-z and photo-z, respectively. We assume that high-quality galaxy spectra and redshifts obtained by ongoing and future powerful spectroscopic surveys can be used as training data. We generate the mock data for the CSST spectroscopic and photometric surveys based on the COSMOS catalog, considering the CSST observational and instrumental effects. About 20,000 and 126,000 galaxies are selected for the CSST spectroscopic and photometric surveys, respectively.

The 1D CNN and MLP we adopt are developed based on the Keras API. We have tested different architectures and choose to use two convolutional and two fully connected layers for the 1D CNN in the spec-z analysis, and two hidden layers for the MLP in photo-z estimate. The mock data are divided into training and testing data sets. About 4500 and 10,000 data are randomly selected as the testing data for the spec-z and photo-z analysis, respectively, and the rests are used as training data sets. In order to reduce the statistical effect on the observational data and generate enough training samples for the same galaxy type, we create 50 Gaussian random realizations of the original training data to train the networks. More realizations would not change the results significantly. After the validation process, we find that our networks can be well trained to get accurate redshift predictions.

We also try to derive the errors of predicted redshifts by generating random samples based on the testing data. The PDFs can be properly illustrated by about 50 realizations, and then the best fits and 1σ errors can be obtained. We find that the neural network can provide excellent estimates for both spec-z and photo-z, and which are significantly better than those of the ordinary SED template-fitting method, especially for noisy data. The redshift accuracy can achieve 0.001 and 0.01 for spec-z and photo-z, respectively, with extremely small outlier fractions. Our MLP can reach a similar photo-z accuracy with a more efficient training process compared to existing codes. In particular, by using our method, the results of low-S/N data are comparable to those of high-S/N data, which indicates the advantages of neural networks in future redshift analyses.

ORCID iDs

Yan Gong ⓘ https://orcid.org/0000-0003-0709-0101
Xuelei Chen ⓘ https://orcid.org/0000-0001-6475-8863

References

Abell, P. A., Allison, J., Anderson, S. F., et al. 2009, arXiv:0912.0201
Allen, C. W. 1976, Astrophysical Quantities (London: Athlone), 264
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Astier, P., Guy, J., Regnault, N., et al. 2006, A&A, 447, 31
Bohin, R., Mack, J., & Ubeda, L. 2011, Instrument Science Report ACS, 2011
Bolzonella, M., Miralles, J. M., & Pello, R. 2000, A&A, 363, 476
Bouchet, P., Lequeux, J., Maurice, E., Prevot, L., & Prevot-Burnichon, M. L. 1985, A&A, 149, 330
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bundy, K. 2019, arXiv:1907.07195
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cao, Y., Gong, Y., Meng, X.-M., et al. 2018, MNRAS, 480, 2178
Capak, P., Aussel, H., Ajiki, M., et al. 2007, ApJS, 172, 99
Cirasuolo, M., Fairley, A., Rees, P., et al. 1919, Msngr, 180, 10
Collister, A. A., & Lahav, O. 2004, PASP, 116, 345
Ellis, R. 2019, arXiv:1907.06797
Fitzpatrick, E. L. 1986, AJ, 92, 1068
Fukushima, K., Miyake, S., Ito, T., et al. 1970, ITSMC, SMC-13, 826
Gerdes, D. W., Sypniewski, A. J., McKay, T. A., et al. 2010, ApJ, 715, 823

Gong, Y., Liu, X., Cao, Y., et al. 2019, ApJ, 883, 203
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Ivezic, Z., Kahn, S. M., Tyson, J. A., et al. 2008, arXiv:0805.2366
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proc. IEEE, 86, 2278
Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, ApJS, 184, 218
Lilly, S. J., Le Fevre, O., Renzini, A., et al. 2007, ApJS, 172, 70
Madau, P. 1995, ApJ, 441, 18
Maiolino, R., Cirasuolo, M., Afonso, J., et al. 2020, Msngr, 180, 24
Nair, V., & Hinton, G. E. 2010, in Proc. 27th Int. Conf. Machine Learning (ICML-10), ed. J. Fürnkranz & T. Joachims (Madison, WI: Omnipress), 807
Pozzetti, L., Bruzual, G., & Zamorani, G. 1996, MNRAS, 281, 953
Pozzetti, L., Madau, P., Zamorani, G., et al. 1998, MNRAS, 298, 1133

Prevot, M. L., Lequeux, J., Prevot, L., Maurice, E., & Rocca-Volmerange, B. 1984, A&A, 132, 389
Samui, S., & Pal, S. S. 2017, NewA, 51, 169
Schlegel, D. J., & Kollmeier, J. A. 2019, arXiv:1907.11171
Seaton, M. J. 1979, MNRAS, 187, 73
Srivastava, N., Hinto, G., Krizhevsky, A., et al. 2014, JMLR, 15, 1929
Stivaktakis, R., Tsagkatakis, G., Moraes, B., et al. 2019, IEEE Transactions on Big Data, 6, 460
Sullivan, M., Howell, D. A., Perrett, K., et al. 2006, AJ, 131, 960
Tamura, N., Takata, N., Shimono, A., et al. 2016, Proc. SPIE, 9908, 99081M
Ubeda, L., et al. 2011, ACS Instrument Handbook, Version 11.0 (Baltimore: STScI)
Zhan, H. 2011, SciSn, 41, 1441
Zhan, H. 2018, in 42nd COSPAR Scientific Assembly (Vancouver: COSPAR), E1.16-4-18