Research in Astron. Astrophys. **2009** Vol. **9** No. **4**, 390–400
http://www.raa-journal.org *http://www.iop.org/journals/raa*

***R**esearch in*
***A**stronomy and*
***A**strophysics*

Photometric redshifts of galaxies from SDSS and 2MASS *

Tao Wang[1], Jia-Sheng Huang[2] and Qiu-Sheng Gu[1]

[1] Department of Astronomy, Nanjing University, Nanjing 210093, China; *twang.nju@gmail.com*
[2] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138

Received 2008 June 16; accepted 2008 July 17

Abstract In order to find the physical parameters which determine the accuracy of photometric redshifts, we compare the spectroscopic and photometric redshifts (photo-z's) for a large sample of $\sim$ 80 000 SDSS−2MASS galaxies. Photo-z's in this paper are estimated by using the artificial neural network photometric redshift method (ANNz). For a subset of $\sim$40 000 randomly selected galaxies, we find that the photometric redshift recovers the spectroscopic redshift distribution very well with rms of 0.016. Our main results are as follows: (1) Using magnitudes directly as input parameters produces more accurate photo-z's than using colors; (2) The inclusion of 2MASS ($J, H, K_{\rm s}$) bands does not improve photo-z's significantly, which indicates that near infrared data might not be important for the low-redshift sample; (3) Adding the concentration index (essentially the steepness of the galaxy brightness profile) as an extra input can improve the photo-z's estimation up to $\sim$ 10 percent; (4) Dividing the sample into early- and late-type galaxies by using the concentration index, normal and abnormal galaxies by using the emission line flux ratios, and red and blue galaxies by using color index ($g - r$), we can improve the accuracy of photo-z's significantly; (5) Our analysis shows that the outliers (where there is a big difference between the spectroscopic and photometric redshifts) are mainly correlated with galaxy types, e.g., most outliers are late-type (blue) galaxies.

Key words: galaxies: distances and redshifts — methods: statistical — techniques: photometric

1 INTRODUCTION

The redshift is very basic and important in the study of extragalactic astronomy. Usually, we measure redshifts of galaxies by using spectra of sufficiently high resolution. However, for faint sources in the deep survey, it is time-consuming and sometimes impossible to observe their spectra, even with a large telescope. Thus, it is desirable to be able to measure an object's redshift from photometry alone. Redshifts measured in this way are called photometric redshifts, or photo-z's.

Photo-z's essentially convert photometric observables (e.g., colors, magnitudes and morphology) into estimates of the physical properties of galaxies (e.g., redshift and type). In particular, it shows great advantages for an effective analysis of those ongoing and planned sky surveys, such as the Sloan Digital Sky Survey (SDSS; York et al. 2000), which will lead to an exponential increase in both quality and quantity of data. Moreover, photometric redshifts have been widely used as an efficient and effective means of studying the statistical properties of galaxies and their evolution (Csabai et al. 2003; Abadalla et al. 2007).

* Supported by the National Natural Science Foundation of China.

Photo-z techniques can be traced back to Baum (1962), who combined nine photometric bands to form low-resolution spectral energy distributions (SEDs) for elliptical galaxies. The coarse SEDs traced the steep 4000 Å break feature, which is an excellent tool for redshift determination since it produces a large difference in flux where two passbands straddle it at a given redshift. In recent years, with the development of large-scale deep sky surveys, great progress in both technique and methodology has been made in estimating photo-z's (Wray & Gunn 2007). Basically, there are two approaches to estimating the photometric redshift: template-matching and empirical training-set methods. In the template-matching method (Koo 1985; Loh & Spillar 1986; Gwyn & Hartwick 1996; Sawicki, Lin & Yee 1997; Connolly et al. 1999; Benítez 2000; Budavári et al. 1999; Budavári et al. 2000; Csabai et al. 2000; Bolzonella et al. 2000; Xia et al. 2002; Mobasher et al. 2007), the colors of the given galaxy are compared with those of templates at different redshifts for determining the redshift of the given object. For a particular target galaxy, the photo-z is determined by the most closely matching template spectrum which has a minimum value of χ^2. The training-set method maps out an empirical relation between photometric observables and redshifts using a training set of galaxies with spectroscopic redshifts and photometry (Connolly et al. 1995; Brunner et al. 1997; Wang et al. 1998; Way & Srivastava 2006). A more detailed description of this method is presented by Csabai et al. (2003, and references therein). It is well established that the template method is more versatile when no spectroscopic data are available. However, the empirical technique shows the advantage of being automatically constrained by the properties of galaxies in the real universe and of requiring no additional assumptions about their formation and evolution. Such a method has proved to be more reliable and produce better photometric redshifts for cases in which moderately large training sets are available (Firth et al. 2003; Wadadekar 2005).

In this work, we use artificial neural networks (ANNz, Collister & Lahav 2004), an interpolative technique based on the training-set method, to estimate the photometric redshift for a large galaxy sample derived from a cross-match of SDSS and 2MASS. We investigate the performances of ANNz for different training sets and different input parameters, including magnitudes, colors, concentration indices and emission lines, etc. Furthermore, we study the causes of outliers (which have a big difference between spectroscopic and photometric redshifts) in detail.

This paper is organized as follows. In Section 2, we briefly describe the photometric catalog generated from SDSS and 2MASS. A concise description of the ANNz is given in Section 3. We then make a full photometric redshift analysis with the method to assess the accuracy of the photo-z obtained in each of the scenarios considered in Section 4, and outliers for each case are discussed in Section 5. We present our conclusions and future improvements for increasing the accuracy and applicability of the method in Section 5. All magnitudes quoted in this paper are in the AB system.

2 CATALOG GENERATION

The Sloan Digital Sky Survey (SDSS) is an imaging and spectroscopic survey, which covers more than a quarter of the whole sky, that constructs the first comprehensive digital map of the universe (York et al. 2000). The survey uses a dedicated, wide-field 2.5 m telescope located at Apache Point, New Mexico. Imaging is carried out in five broad bands (u, g, r, i, z), spanning the range from 3000 to 10 000 Å (Fukugita et al. 1996; Stoughton et al. 2002).

The Two Micron All Sky Survey (2MASS) project is designed to close the gap between our current technical capability and our knowledge of the near-infrared sky. It uses two highly-automated 1.3-m telescopes, one is at Mt. Hopkins, Arizona, and the other is located at CTIO, Chile (Cutri et al. 2003). Each telescope is equipped with a three-channel camera, with each channel consisting of a 256×256 array of HgCdTe detectors, capable of observing the sky simultaneously at J(1.25 mm), H(1.65 mm) and $K_{\rm s}$ (2.17 mm), to a 3σ limiting sensitivity of 17.1, 16.4 and 15.3 mag, respectively. Jarrett et al. (2000) presented detailed information on the extended source catalog. Table 1 describes the broadband filters and their wavelength ranges of the SDSS and 2MASS catalogs.

The NYU Value-Added Galaxy Catalog (NYU-VAGC; Blanton et al. 2005) is essentially an "extended main sample maintained for the study of galaxy formation and evolution based on a set of publicly-released surveys matched to the SDSS. Using the NYU-VAGC data, we generate about 180 000

Table 1 Survey Filters and Characteristics

Bandpass	Survey	$\lambda_{\rm eff}$(Å)	$\Delta\lambda$(Å)
u	SDSS	3551	600
g	SDSS	4686	1400
r	SDSS	6165	1400
i	SDSS	7481	1500
z	SDSS	8931	1200
J	2MASS	12500	1620
H	2MASS	16500	2510
$K_{\rm s}$	2MASS	21700	2620

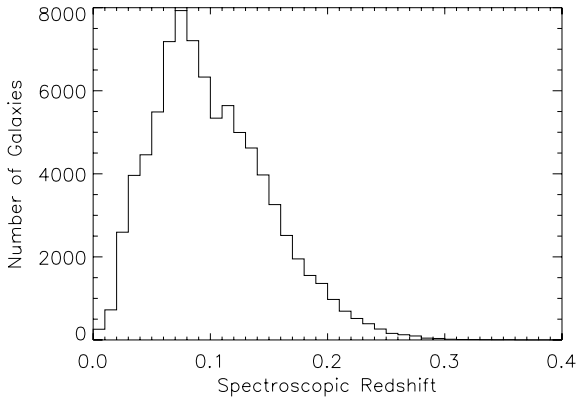

Fig. 1 The redshift distribution for ∼84 697 galaxies of the whole sample selected from the cross-match of SDSS and 2MASS.

galaxies that have both optical (SDSS) and near-infrared (2MASS) photometric data with known spectroscopic redshifts from SDSS DR4. From these galaxies, we selected the objects satisfying the following criteria: 1) the spectroscopic redshift confidence is greater than 0.95; 2) redshift warning flag of SDSS data is 0; 3) the AB magnitude $r < 17.5$ (derived from the Petrosian magnitude, Galactic extinction corrected). 4) The extended source catalog confusion flag of 2MASS, "cc_flg," is required to be zero in all three band passes. With these qualifications, we obtained a final sample of 84 697 galaxies. In Figure 1, we plot the redshift distribution for the whole sample.

3 ARTIFICIAL NEURAL NETWORKS

In this work, we use ANNz (Collister & Lahav 2004), a training-set method using Artificial Neural Networks, to estimate redshifts for the catalog described in Section 2. The method has been shown to produce competitive results compared to both template-fitting and other training-set methods available. It requires three sets of data: training, validation and testing sets. The training set and validation set are used to train the network and must include an estimate for the actual redshifts of galaxies (such as the spectroscopic redshift), while the testing set would usually contain galaxies for which no other estimate of redshift is available. Given a suitable training set of galaxies for which we have both photometry, m, and a spectroscopic redshift, $Z_{\rm spec}$, the ANNz is trained by minimizing the cost function

$$E = \sum_k (z_{\rm phot}(w, m_k) - z_{{\rm train},k})^2, \quad (1)$$

with respect to the free parameters (weights), w, where $Z_{\rm phot}(w, m_k)$ is the network output for the given input and weight vectors, and the sum is over the galaxies in the training set.

ANNz uses an iterative quasi-Newton method to perform this minimization. Details of the algorithm and regularization can be found in Bishop (1995) and Lahav et al. (1996). After each training iteration, the cost function is also evaluated on a separate validation set. Training terminates after a chosen number of iterations and the final weights chosen for the ANN are those from the iteration at which the cost function is minimal on the validation set. This is useful for avoiding over-fitting, which refers to the tendency of an algorithm with many adjustable parameters to fit to the noise in the training set data, if the training set is small.

The remaining freedom left in a neural network analysis is the architecture of the network. A simple architecture is easier to minimize but may not provide the best fit to the data, while a complicated one may remain stuck in a local minimum of the cost function more easily and hence not provide the best solution to the problem. Here, we use a network with architecture N:2N:2N:1 (i.e. which has N inputs, two hidden layers, each with 2N, and only one output estimating the redshift) which has been shown to work well on photometric data. For details about the architecture, see Collister and Lahav (2004, and references therein).

4 RESULTS AND DISCUSSION

4.1 Photometric Redshifts

We randomly select training, validation and testing sets from our main sample of 84 697 galaxies, which contain 30 000, 20 000 and 34 697 galaxies, respectively. Since our methods are training-set based, we make the spectroscopic training set fully representative of the photometric sample to be analyzed, i.e., to have similar statistical properties and magnitude/redshift distributions. In Figure 2, we plot the distributions of spectroscopic redshifts, g-magnitudes and colors for the three sets, which suggests that they match very well.

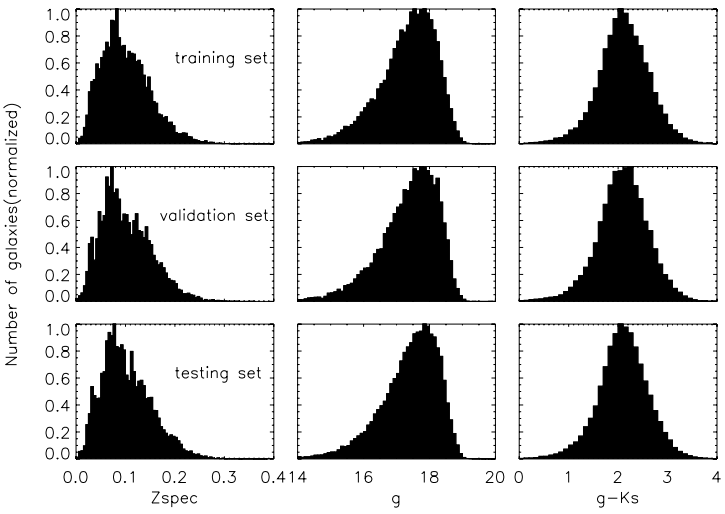

Fig. 2 The distribution of redshifts, g-magnitude and $g-K_{\rm s}$ color for the training, validation and testing sets. We can see that the three sets coincide with each other very well.

Table 2 Summary of ANNz Cases

Case	Inputs/Description	Number	$\sigma_{\rm rms}$	σ_{68}	Outliers (δ_z >0.03)
A1	$ugriz$	34697	0.0173	0.0132	2238
A2	$ugriz + JHK_{\rm s}$	34697	0.0169	0.0128	2050
A3	$u-g, g-r, r-i, i-z, z-J, J-H, H-K_{\rm s}$	34697	0.0190	0.0141	2846
B	$ugriz + JHK_{\rm s}, CuCgCrCiCz$	34697	0.0161	0.0124	1800
B1	late-type ($Cr \leq 3$)	23057	0.0182	0.0145	1538
B2	early-type ($Cr > 3$)	11640	0.0138	0.0099	340
C1	$ugriz + JHK_{\rm s}$, blue ($g-r \leq 1$)	25409	0.0178	0.0141	1772
C2	$ugriz + JHK_{\rm s}$, red ($g-r > 1$)	9288	0.0156	0.0103	356

Table 3 Comparison of Photo-z Errors from Different Techniques

Method	$\sigma_{\rm rms}$	Data Set	Input Parameters	Source
CWW templates	0.067	SDSS-EDR	$ugriz$	Csabai et al. (2003)
Bruzual-Charlot templates	0.055	SDSS-EDR	$ugriz$	Csabai et al. (2003)
Hybrid	0.035	SDSS-EDR	$ugriz$	Csabai et al. (2003)
Polynomial	0.032	SDSS-EDR	$ugriz$	Csabai et al. (2003)
Support Vector Machine	0.027	SDSS-DR2	$ugriz$	Wadadekar (2005)
Support Vector Machine	0.023	SDSS-DR2	$ugriz + r50 + r90$	Wadadekar (2005)
Kernel Regression	0.019	SDSS-DR5	color	Wang et al. (2008)
ANNz	0.017	SDSS-DR4	$ugriz$	this work
ANNz	0.016	SDSS-DR4	$ugriz + JHK_s, CuCgCrCiCz$	this work

To test the quality of the photo-z estimates, we use three photo-z performance metrics. The first two metrics are the photo-z bias, δ_z and the photo-z rms scatter, σ_z:

$$\delta_z = |Z_{\rm spec} - Z_{\rm phot}|/(1 + Z_{\rm spec}), \tag{2}$$

$$\sigma_z = \langle (Z_{\rm spec} - Z_{\rm phot})^2 \rangle^{\frac{1}{2}}. \tag{3}$$

The third performance metric, denoted by σ_{68}, is the range containing 68% of the test set objects in the distribution of δ_z. This metric is useful because the probability distribution function $P(\sigma_z)$ is in general non-Gaussian and asymmetric (for a Gaussian distribution, σ and σ_{68} coincide). Explicitly, σ_{68} is defined by the value of δ_z such that 68% of the objects have $\delta_z < \sigma_{68}$.

The ANNz is very flexible in the sense that it is easy to change the input parameters, the training set and the network configurations. We tried a variety of combinations of possible input photometric observables to see their effects on the photo-z quality. We calculated photo-z's using galaxy magnitudes, colors and the concentration indices for some or all of the passbands. Table 2 provides a summary of the performance results of the different cases. Moreover, we compare our results to those achieved by other methods in Table 3.

4.2 Impact of Different Input Parameters

First, in case A, we compare the photo-z accuracy for different choices of filters. Three of these combinations are listed. In case A1, the training and photo-z estimation are carried out using only the five optical bands $ugriz$ from SDSS, while in case A2, we use $ugriz$ (SDSS)+$JHK_{\rm s}$ (2MASS) as the input parameters. Then, we use only the seven colors $u-g, g-r, r-i, i-z, z-J, J-H$ and $H-K_{\rm s}$ in case A3. We find that using magnitudes as input parameters derives better photo-z accuracy than using colors. On the other hand, we find that for low redshift galaxies, the addition of IR data only yields a small improvement in the photo-z accuracy. The reason is that the most important information in determining photometric redshifts is the 4000 Å break for the low-redshift galaxies. The mean redshift of

our galaxies is about 0.1, and the 4000 Å break lies in the g band. The inclusion of IR data here plays a less important role for the estimation of photo-z's. However, at high-redshift, Abdalla et al. (2007) and Banerji et al. (2007) have shown that IR data are very important for determining photo-z's.

In case B, we draw attention to the effects of introducing additional observables (concentration indices) into our estimates of the photo-z. The concentration index Ci in passband i is defined as the ratio of $PetroR90$ and $PetroR50$, which are the radii that encircle 90% and 50% of the Petrosian flux, respectively. Early-type (E and S0) galaxies, with centrally peaked surface brightness profiles, tend to have high values of the concentration index, while late-type spirals, with quasi-exponential light profiles, typically have lower values. Previous studies have shown that the concentration parameter correlates well with galaxy morphological type (Morgan 1958; Shimasaku et al. 2001; Yamauchi et al. 2005; Park & Choi 2005), and we used it to help break the degeneracy between redshifts and galaxy types. Concentration indices are perhaps the best parameter to be used for classifying the morphology of galaxies (Shimasaku et al. 2001). In this case, we combine the eight magnitudes with the concentration indices Cu, Cg, Cr, Ci and Cz in the $ugriz$ filters as input parameters. We find that using concentration indices in addition to magnitudes really helps reduce the photo-z scatter by a few percent as well as the number of outliers.

Consequently, in cases B1 and B2, we split the whole sample (including the training and validation sets) into early-type and late-type galaxies using concentration indices, and estimate photo-z's respectively. Shimasaku et al. (2001) proved that a choice of $Cr = 3$ leads to a good dichotomous classification of galaxies into early and late types and we choose that value for our classification. Then, in case C, we divide the same catalog into red and blue galaxies using the criteria: $g - r = 1$ (one reason for this criteria is to keep the number of red galaxies compatible with early-type galaxies selected via the concentration index). Our results show that the accuracy of the photometric redshifts is also dependent on galaxy types, with smaller scatter for early-type or red galaxies.

In Figures 3 and 4, we plot photo-z's and $Z_{\rm phot}$ for galaxies in the testing set, and also true spectroscopic redshifts, $Z_{\rm spec}$, for all the cases we considered. In Figure 3, we show the results for the same testing set (34 697 galaxies) using different input parameters: the top row shows results for cases A1 and A2, which utilizes $ugriz$ and $ugriz + JHKs$, and the bottom row shows results for cases A3 and B using colors and $ugriz\ JHKs + CuCgCrCiCz$ as input parameters, respectively. In Figure 4, we plot photo-z vs. spectroscopic redshift for galaxies of different types: results for early-type and late-type galaxies are shown on the top, and red and blue galaxies on the bottom. In each panel of Figures 3 and 4, the values of the corresponding global photo-z performance metrics $\sigma_{\rm rms}$ are shown. Furthermore, the solid line in each figure traces $Z_{\rm phot} = Z_{\rm spec}$, i.e., the line for a perfect photo-z estimator.

4.3 Outliers and Errors in the Photo-z

In order to understand which physical parameters determine the accuracy of Photo-z, we identify outliers with large differences between photometric and spectroscopic redshifts, which will help us to learn how to reduce systematic effects for photometric redshifts and have a comprehensive understanding of how different bands and input parameters can help the photo-z estimates. For this purpose, first we perform a Gaussian fit for $\delta_z = |Z_{\rm phot} - Z_{\rm spec}|/(1 + Z_{\rm spec})$ for the case A2 which uses $ugriz + JHK_{\rm s}$ and has a better accuracy of photo-z's. As plotted in Figure 5, we define galaxies with $\delta_z > 3\sigma \sim 0.03$ as outliers. For a detailed analysis, we also perform the gaussian-fit by splitting the testing sample into 9 bins using spectroscopic redshifts, which we list in Figure 6. Then, in Figure 7, we plot photo-z's vs. spectroscopic redshifts for A2 after removing the outliers and the result shows that the 'cleaning' of outliers is effective. Hence, we apply it to all the scenarios we considered, and list the total number of outliers for each scenario in Table 2. Since it correlates with $\sigma_{\rm rms}$ and σ_{68}, we also recommend taking the number of outliers as a criterion for estimating the accuracy of the photo-z techniques.

Furthermore, we analyze the composition of outliers, and our analysis shows that the distributions of the outliers for different cases are almost the same. The distributions of the outliers on $g - r$ color and Cr, the concentration index in r band, are listed in Figure 8. As noted, the outliers in both cases are mostly concentrated in $g - r < 1$ or $Cr < 3$, i.e. blue or late-type galaxies. In particular, for case B, the

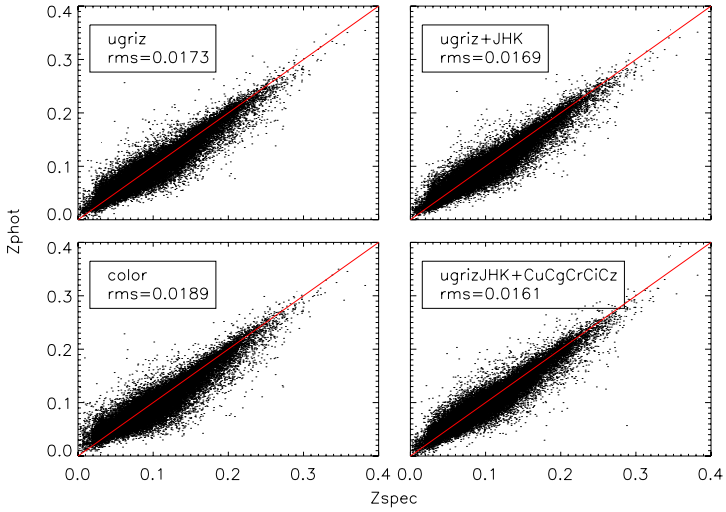

Fig. 3 Photo-z's performance for the same testing set using different input parameters: optical bands, optical and near-infrared bands, colors and eight bands combined with concentration indices. We assess the difference between bands and colors as inputs in this figure, and we also verify how the inclusion of IR data would enhance the photometric redshift estimation.

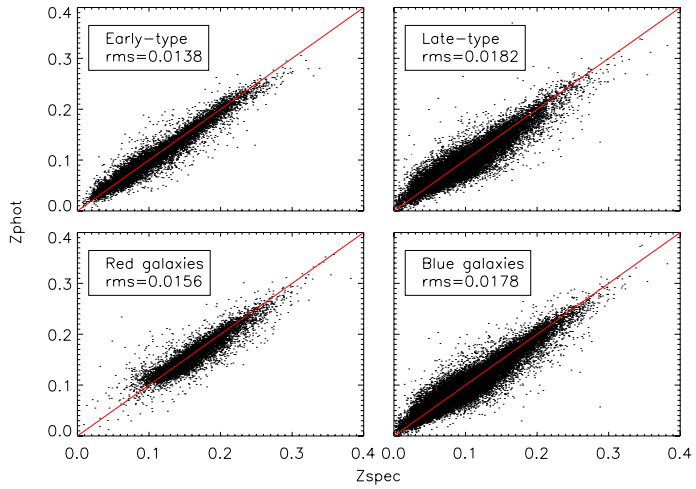

Fig. 4 Photo-z's performance for galaxies of different types. Top panels show results for early-type and late-type galaxies, while bottom panels show red and blue galaxies, respectively.

number of galaxies with $g-r>1$ out of 1800 outliers is only 362 while there are $\sim 10\,000$ red galaxies in the whole sample. Identically, the number with $Cr>3$ in outliers is only 327 while it is 11 640 in the whole sample. We attribute the reason to be that it is naturally hard to get photo-z's of those young blue galaxies from their relatively featureless continuum.

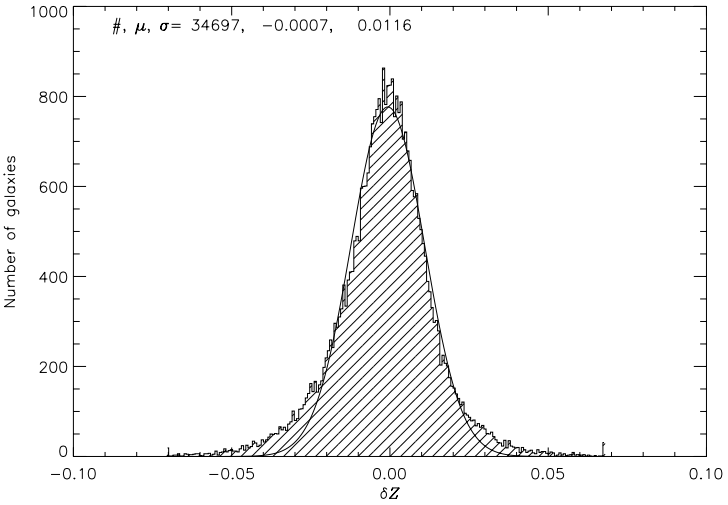

Fig. 5 Gaussian fit for $\delta_z = |Z_{\rm phot} - Z_{\rm spec}|/(1 + Z_{\rm spec})$ for the case A2 which uses $ugriz + JHK_{\rm s}$ as inputs. Utilizing the fit result σ=0.0116, we define galaxies with $\delta_z > 3\sigma$, i.e., $\delta_z > 0.03$ as outliers in all the cases. $\#$, μ and σ denote the number of galaxies, the mean value and the standard deviation, respectively.

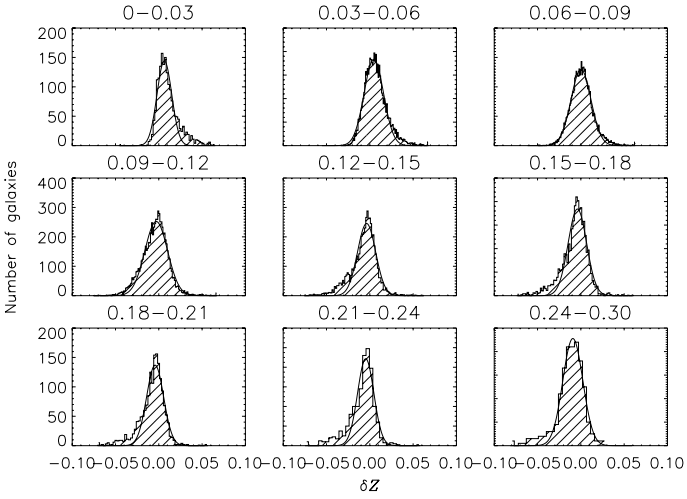

Fig. 6 Gaussian-fit for the distribution of δ_z for different redshift bins of the testing set. From the top left to the bottom right, each corresponds to redshift bins $0 - 0.03, 0.03 - 0.06, 0.06 - 0.09, 0.09 - 0.12, 0.12 - 0.15, 0.15 - 0.18, 0.18 - 0.21, 0.21 - 0.24$ and $0.24 - 0.3$, respectively.

When a neural network is trained, we obtain an estimate for the error of each of the photometric redshifts predicted. This error is obtained the following way. For every scenario, the inputs of a neural network have noise associated with them. We can assess the variance that this noise would introduce into the output of the network by changing the inputs according to the error. This will lead to the following

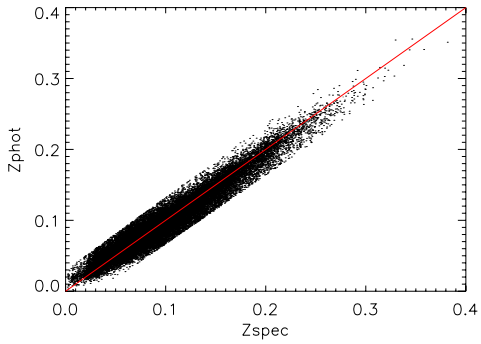

Fig. 7 Photo-z's versus spectroscopic redshifts after removing outliers with $\delta_z > 0.03$ for the case A2 that uses $ugriz+JHK_{\rm s}$ as inputs. We can see that such a criterion can exclude outliers very effectively.

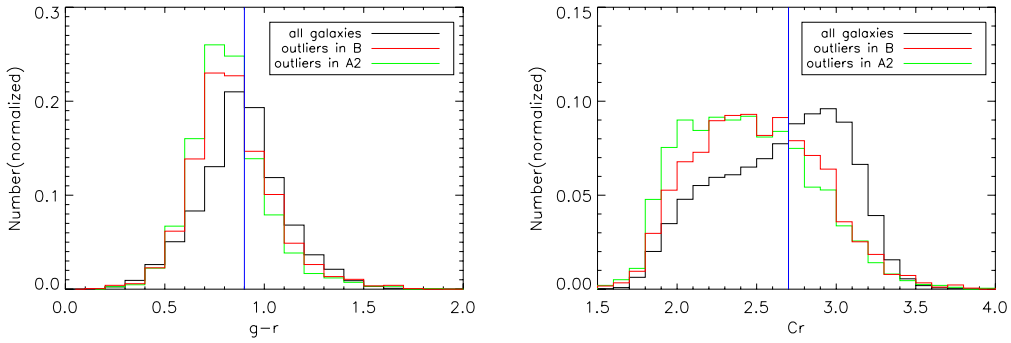

Fig. 8 The distributions of outliers on $g-r$ and Cr for cases A2 and B. The black line denotes the whole sample while the red and green lines denote outliers for A2 ($ugriz+JHK_{\rm s}$) and B($ugriz+JHK_{\rm s}+CuCgCrCiCz$), respectively. We can see that the fraction of outliers in both cases is much higher than the whole sample in $g-r<0.9$ and $Cr<2.7$, and that most outliers are concentrated in $g-r<1$ and $Cr<3$, i.e. blue or late-type galaxies.

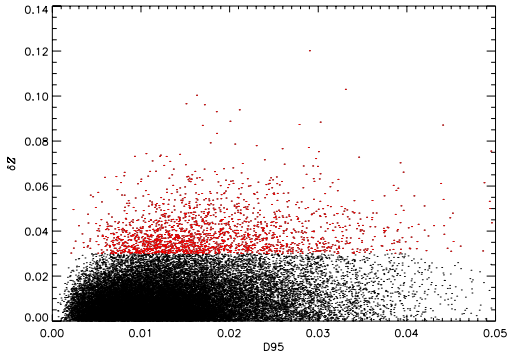

Fig. 9 D95 vs. $\delta z=|Z_{\rm spec}-Z_{\rm phot}|/(1+Z_{\rm spec})$ for the whole sample (black) and the outliers (red) in case A2.

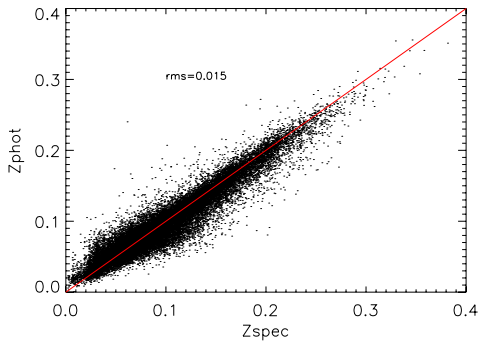

Fig. 10 Estimating photo-z's for case A2 after removing outliers using our criterion, we find that the accuracy increased by 10%.

error estimate:

$$\sigma_z^2 = \sum \left(\frac{\partial z}{\partial m_i}\right)^2 (\sigma_{m_i})^2. \quad (4)$$

Mobasher et al. (2007) proposed a D_{95} method to enable the identification of outliers in derived photo-z's. The parameter D_{95} is defined as

$$D_{95} = \frac{\sigma_{95}}{1 + Z_{\rm phot}}, \quad (5)$$

where σ_{95} is the 95% confidence interval of σ_z and $Z_{\rm phot}$ is the estimated photometric redshift. Therefore, D_{95} can be calculated independently from any knowledge about spectroscopic redshift. If the error distribution is Gaussian, then, by definition, $\sigma_{95} = 2\sigma_z$. Their results show that D_{95} provides a useful and practical measure to identify the fraction of outliers. We find that the contamination is very serious, as shown in Figure 9. Here, we improved the method by only removing those late-type and blue galaxies instead of all the galaxies with a D_{95} cut. For example, in case A2, we remove galaxies with $Cr < 3$ and $g - r < 1$ whose $D_{95} > 0.02$, about 4583 galaxies which contain 599 outliers ($\delta_z > 0.03$), i.e. by removing 10% of galaxies (4583/34 697) of the whole sample we remove $\sim 30\%$ outliers (599/2050). After applying this criterion to remove outliers, we estimate the photo-z's and find that the accuracy increased by 10%, as plotted in Figure 10.

5 CONCLUSIONS

In this work, we use ANNz to estimate photometric redshifts for a large sample of 84 697 galaxies from SDSS and 2MASS. We compare the photometric redshift quality for different choices of survey depths and different choices of input parameters. After considering the different cases, the photometric redshifts are well determined after introducing the concentration index and dividing the sample into different types. We also examine the role of near infrared data in constraining photometric redshifts and find that for low-redshift galaxies, the NIR data does not play an important role. Our results suggest that it is important to remove outliers in the photo-z estimation. After removing those galaxies whose $Cr < 3$ and $g - r < 1$ under a D_{95} cut, we can improve the accuracy of photometric redshifts significantly.

Acknowledgements This work is supported by the Program for New Century Excellent Talents in University (NCET), the National Natural Science Foundation of China under grants 10221001 and 10633040 and the National Basic Research Program (973 program No. 2007CB815405). Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, Cambridge University, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington. This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

References

Abdalla, F. B., Amara, A., Capak, P., et al. 2008, MNRAS, 387, 969
Banerji, M., Abdalla, F. B., Lahav, O., et al. 2008, MNRAS, 386, 1219
Baum, W. A. 1962, in IAU Symp. 15, Problems of Extra-Galactic Research, ed. G. C. McVittie (New York: Macmillan Press), 390
Benítez, N. 2000, ApJ, 536, 571
Bishop, C. M. 1995, Neural Networks for Pattern Recognition (New York: Oxford Univ. Press)
Blanton, M. R., et al. 2005, AJ, 129, 2562
Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, A&A, 363, 476
Brunner, R. J., Connolly, A. J., & Szalay, A. S. 1999, ApJ, 516, 563
Budavári, T. 2000, AJ, 120, 1588
Budavári, T., Szalay, A. S., Connolly, A. J., et al. 1999, in ASP Conf. Ser. 191, Photometric Redshifts and High Redshift Galaxies, eds. R. J. Weymann, L. J. Storrie-Lombardi, M. Sawicki et al. (San Francisco: ASP), 19
Collister, A. A., & Lahav, O. 2004, PASP, 116, 345
Connolly, A. J., Budavari, T., Szalay, A. S., et al. 1999, in ASP Conf. Ser. 191, Photometric Redshifts and High Redshift Galaxies, eds. R. J. Weymann, L. J. Storrie-Lombardi, M. Sawicki et al. (San Francisco: ASP), 13
Csabai, I., et al. 2003, AJ, 125, 580
Csabai, I., Connolly, A. J., Szalay, A. S., et al. 2000, AJ, 119, 69
Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, VizieR On-line Data Catalog: II/246
Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, ApJ, 513, 34
Firth, A. E., Lahav, O., & Somerville, R. S. 2003, MNRAS, 339,1195
Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748
Gwyn, S. D. J., & Hartwick, F. D. A. 1996, ApJ, 468, L77
Jarrett, T. H, Chester, T., Cutri, R., et al. 2000, AJ, 119, 2498
Koo, D. C. 1985, AJ, 90, 418
Lahav, O., Naim, A., Sodré, L., et al. 1996, MNRAS, 283,207
Loh, E. D., & Spillar, E. J. 1986, ApJ, 303, 154
Mobasher, B., et al. 2007, ApJS, 172, 117
Park, C., & Choi, Y.-Y. 2005, ApJ, 635, L29
Sawicki, M. J., Lin, H., & Yee, H. K. C. 1997, AJ, 113, 1
Shimasaku, K., et al. 2001, AJ, 122, 12
Stoughton, C., et al. 2002, AJ, 123, 485
Wadadekar, Y. 2005, PASP, 117, 79
Wang, D., Zhang, Y.-X., Liu, C., et al. 2008, ChJAA (Chin. J. Astron. Astrophys.), 8, 119
Wang, Y., Bahcall, N., & Turner, E. L. 1998, AJ, 116, 2081
Way, M. J., & Srivastava, A. N. 2006, ApJ, 647, 102
Wray, J. J., & Gunn, J. E. 2008, ApJ, 678, 144
Xia, L., et al. 2002, PASP, 114, 1349
Yamauchi, C., et al. 2005, AJ, 130, 1545
York, D. G., et al. 2000, AJ, 120, 1579