MODEL-INDEPENDENTLY CALIBRATING THE LUMINOSITY CORRELATIONS OF GAMMA-RAY BURSTS

Shreeprasad Bhat [1], Shantanu Desai [1,2]
[1] Department of Artificial Intelligence, IIT Hyderabad
[2] Department of Physics, IIT Hyderabad

April 27, 2022

Abstract

Gamma-ray bursts (GRBs) detected at high redshift can be used to trace the Hubble diagram of the Universe. However, the distance calibration of GRBs is not as easily as that of type Ia supernovae (SNe Ia). For the calibrating method based on the empirical luminosity correlations, there is an underlying assumption that the correlations should be universal over the whole redshift range. In this paper, we investigate the possible redshift dependence of six luminosity correlations with a completely model-independent deep learning method. We construct a network combining the Recurrent Neural Networks (RNN) and the Bayesian Neural Networks (BNN), where RNN is used to reconstruct the distance-redshift relation by training the network with the Pantheon compilation, and BNN is used to calculate the uncertainty of the reconstruction. Using the reconstructed distance-redshift relation of Pantheon, we test the redshift dependence of six luminosity correlations by dividing the full GRB sample into two subsamples (low-z and high-z subsamples), and find that only the $E_p - E_\gamma$ relation has no evidence for redshift dependence. We use the $E_p - E_\gamma$ relation to calibrate GRBs, and the calibrated GRBs give tight constraint on the flat ΛCDM model, with the best-fitting parameter $\Omega_M = 0.307^{+0.065}_{-0.073}$. Subject headings: gamma-ray burst: general - cosmology: observations - distance scale.

1. INTRODUCTION

The accelerating expansion of the universe is first found from the fact that the luminosity of type Ia supernovae (SNe Ia) is dimmer than expected (Riess et al. 1998: Perlmutter et al. 1999). Since then, many SNe Ia datasets have been compiled to cosmological researches (Suzuki et al.|2012; Betoule et al. 2014 ; Scolnic et al. 2018). However, due to the limited luminosity, most observable SNe Ia are at redshift $z \leq 2$. With these low redshift data, cosmological models cannot be unambiguously distinguished (Zhu et al. 2008; Dolgov et al. 2014; |Tutusaus et al. 2016; Wei et al. 2016). Among these SNe Ia samples, the most up-to-date Pantheon compilation (Scolnic et al. 2018) is the largest sample and the redshift of the furthest SNe reaches up to $z \sim 2.3$, but the number of SNe whose redshift is larger than 1.4 is only six. Actually, the subtle difference between cosmological models in low redshift range would be remarkable in high redshift range. For various cosmological models, such as Λ CDM model, wCDM model, Chevallier-Polarsky-Linder (CPL) model, holographic dark energy (HDE) model and generalised chaplygin gas (GCG) model, the evolutions of dark energy equation of state are discrepant in high redshift (Chevallier & Polarski 2001; Wang et al. 2017: Pan et al. 2020: Escamilla-Rivera et al. 2020). Thus, it is important to extend the Hubble diagram to high redshift range. Gamma-ray bursts (GRBs), as the most energetic explosions in the universe, is detectable up to very high redshift (Cucchiara et al. 2011). Therefore, it is possible to use GRBs as standard candles to trace the Hubble diagram at high redshift. Combining GRBs with other standard candles, the cosmological parameters can be tightly constrained (Friedman & Bloom 2005: Wang & Dail 2006: Izzo et al. 2009: Liang et al. 2011: Bloom et al. 2003: Xu et al. 2005: Wei & Zhang 2009: Demianski & Piedipalumbo 2011; Cai et al. 2013; Chang et al. 2014; Lin et al. 2016). However, it is not easy to calibrate the distance of GRBs due to the lack of knowledge on the explosion mechanism.

Based on the correlations between various observables of the prompt or afterglow emission, several methods have been proposed to calibrate the distance of GRBs (Dai et al. 2004 Ghirlanda GhiselLini & Lazzati; Firmani et al. 2005; Schaefer 2007: Liang et al. 2008 . Wei 2010. Liu & Wei 2015). Most calibrating methods rely on one of the following 2dimensional empirical luminosity calibrations found in long GRBs: the correlation between spectrum lag and isotropic dimensional empirical luminosity calibrations found in long GRBs: the correlation between spectrum lag and isotropic peak luminosity ($\tau_{\rm lag} - L$ relation) (Norris et al.| 2000), the correlation between time variability and isotropic peak luminosity ($V - L$ relation) (Fenimore & Ramirez-Ruiz 2000), the correlation between peak energy and isotropic peak luminosity ($E_p - L$ relation) (Yonetoku et al.|2004), the correlation between peak energy and collimation-corrected energy ($E_p - E_\gamma$ relation) (Ghirlanda GhiselLini & Lazzati), the correlation between minimum

rise time of light curve and isotropic peak luminosity ($\tau_{\rm RT} - L$) (Schaeter 2007) and the correlation spectrum and isotropic equivalent energy(($E_p - E_{iso}$ relation) (Amati et al. 2002). What is noteworthy is that, all of the above correlations depend on a certain cosmological model, thus leading to the circularity problem when use the calibrated GRBs to constrain cosmological models. Several model-independent methods have been proposed by using distance ladders to calibrate GRBs (Liang et al. 2008: Wei & Zhang 2009 : Wei

- corresponding author: linhn@cqu.edu.cn 2010: Liu & Wei 2015). Using cubic interpolation or other approximations from SNe Ia dataset, one can first calculates the distance for low- $z(z < 1.4)$ GRBs to derive the empirical luminosity correlations. Then by extrapolating the correlations to high- $z(z > 1.4)$ GRBs, one can inversely derive the distance for high- z GRBs. However, in addition to the dependency of a certain approximation form, there is an underlying assumption in this method that the luminosity correlations are universal over all redshifts. Many works have been devoted to test the possible redshift dependence of luminosity correlations (Li 2007; Basilakos & Perivolaropoulos 2008 ; Wang et al. 2011; Lin et al. 2015, 2016). Assuming a flat Λ CDM model with different cosmological parameters, Basilakos & Perivolaropoulos (2008) and Wang et al. (2011) investigated the above six empirical luminosity correlations in four redshift bins, and found no statistically significant evidence for redshift evolution. While Lin et al. (2016) rechecked these correlation in two redshift bins, and found moderate evidence ($> 3\sigma$) for the redshift evolution in four out of six correlations.

All the above works to check the redshift dependence of GRB luminosity correlations still depend on cosmological model. In this paper, we will investigate the redshift dependence of these correlations with a model-independent method, i.e. the deep learning, which is one of the most exciting areas in machine learning. Deep learning is tackling large and highly complex machine learning tasks by training deep neural networks constructed with an input layer to receive the features, several hidden layers to transform the information from the previous layer, and an output layer to export the target, where each layer contains hundreds of nonlinear processing neurons (Aurelien 2017). Recently, deep learning method has been widely employed in cosmological researches, such as predicting galaxy morphology (Dieleman et al. 2015), learning the universe at scale (Mathuriva et al. 2018), constraining cosmological dark energy (Escamilla-Rivera et al. 2020), and so on. In this paper, we proposed a method to calibrate the luminosity correlations of GRBs using deep learning method, which is completely independent on the cosmological model. We first combine the Recurrent Neural Networks (RNN) and the Bayesian Neural Networks (BNN) to reconstruct the distance-redshift relation from the Pantheon sample up to the highest redshift of GRB dataset, then we test the redshift dependence of GRB luminosity correlations by dividing the full GRB sample into low- $z(z \leq 1.4)$ and high- $z(z > 1.4)$ subsamples. Comparing with previous works (Basilakos & Perivolaropoulos 2008 ; Wang et al. 2011; Lin et al. [2016), our method is model-independent and only relies on the training data, i.e. the Pantheon compilation. Without any assumption about the cosmological model or about the specific form of distance-redshift relation of SNe Ia, we directly calibrate the distance of GRBs from the Pantheon sample with deep learning.

The rest of the paper is organized as follows: In Section 2 we introduce the architecture of the RNN + BNN network. In Section 3. we use the network to reconstruct the distance-redshift relation from Pantheon supernovae data set. In Section 4 we use the reconstructed distance-redshift relation to test the possible redshift dependence of luminosity correlations of GRBs. Finally, discussions and conclusion are given in Section 5

1   GRB Dataset

The GRB data is taken directly from (Wang et al 2011). In Table 1, we list the variables of 116 GRBs that we use in fitting luminosity correlations. In addition to the GRBs included in the analysis of Xiao & Schaefer (2009), we add the GRB090423, which has the highest redshift so far. We use the typical spectral index $\alpha = -1$ and $\beta = -2.2$ for this burst (Salvaterra et al. 2009). The data of 116 GRBs used in our analysis. For pre-Swift GRBs, we take the values of P_{bolo} and S_{bolo} directly from Schaefer (2007). For those GRBs observed by Swift, we adopt the values of P and S from Swift website and calculate P_{bolo} and S_{bolo}. We use the F_{beam} value from Ghirlanda et al (2007). Other data are taken from Xiao and Schaefer (2009).

GRB	z	P_{bolo} [erg/cm^{2}s]	S_{bolo} [erg/cm^2]	F_{beam}	τ_{lag} [sec]	V	E_{peak} [keV]	τ_{RT} [sec]
970228	0.70	7.3E-6 ± 4.3E-7	$\cdots$	$\cdots$	$\cdots$	0.016 ± 0.010	115^{+38}_{-38}	$\cdots$
970508	0.84	3.3E-6 ± 3.3E-7	8.09E-6 ± 8.1E-7	0.0795 ± 0.0204	0.49 ± 0.02	0.018 ± 0.004	389^{+40}_{-40}	0.65 ± 0.07
970828	0.96	1.0E-5 ± 1.1E-6	1.23E-4 ± 1.2E-5	5.32E-03 ± 1.44E-03	$\cdots$	0.052 ± 0.005	298^{+30}_{-30}	0.36 ± 0.14
971214	3.42	7.5E-7 ± 2.4E-8	$\cdots$	$\cdots$	0.03 ± 0.05	0.048 ± 0.002	190^{+20}_{-20}	$\cdots$
980703	0.97	1.2E-6 ± 3.6E-8	2.83E-5 ± 2.9E-6	1.84E-02 ± 2.67E-03	0.69 ± 0.02	0.024 ± 0.001	254^{+25}_{-25}	3.00 ± 0.19
990123	1.61	1.3E-5 ± 5.0E-7	3.11E-4 ± 3.1E-5	2.41E-03 ± 6.90E-04	0.07 ± 0.01	0.059 ± 0.003	604^{+60}_{-60}	$\cdots$
990506	1.31	1.1E-5 ± 1.5E-7	$\cdots$	$\cdots$	0.04 ± 0.01	0.337 ± 0.001	283^{+30}_{-30}	0.13 ± 0.01
990510	1.62	3.3E-6 ± 1.2E-7	2.85E-5 ± 2.9E-6	2.13E-03 ± 3.19E-04	0.03 ± 0.01	0.118 ± 0.001	126^{+10}_{-10}	0.13 ± 0.01
990705	0.84	6.6E-6 ± 2.6E-7	1.34E-4 ± 1.5E-5	3.48E-03 ± 9.60E-04	$\cdots$	0.097 ± 0.004	189^{+15}_{-15}	0.62 ± 0.37
991208	0.71	2.1E-5 ± 2.1E-6	$\cdots$	$\cdots$	$\cdots$	0.023 ± 0.003	190^{+20}_{-20}	0.27 ± 0.01
991216	1.02	4.1E-5 ± 3.8E-7	2.48E-4 ± 2.5E-5	3.00E-03 ± 9.46E-04	0.03 ± 0.01	0.062 ± 0.003	318^{+30}_{-30}	0.09 ± 0.01
000131	4.50	7.3E-7 ± 8.3E-8	$\cdots$	$\cdots$	$\cdots$	0.056 ± 0.005	163^{+13}_{-13}	0.84 ± 0.39
000210	0.85	2.0E-5 ± 2.1E-6	$\cdots$	$\cdots$	$\cdots$	0.018 ± 0.002	408^{+14}_{-14}	0.45 ± 0.03
000911	1.06	1.9E-5 ± 1.9E-6	$\cdots$	$\cdots$	$\cdots$	0.122 ± 0.013	986^{+100}_{-100}	0.07 ± 0.22
000926	2.07	2.9E-6 ± 2.9E-7	$\cdots$	$\cdots$	$\cdots$	0.326 ± 0.034	100^{+7}_{-7}	$\cdots$
010222	1.48	2.3E-5 ± 7.2E-7	2.45E-4 ± 9.1E-6	0.0014 ± 0.0001	$\cdots$	0.143 ± 0.004	309^{+12}_{-12}	0.45 ± 0.01
010921	0.45	1.8E-6 ± 1.6E-7	$\cdots$	$\cdots$	1.00 ± 0.04	0.008 ± 0.006	89^{+22}_{-14}	4.31 ± 0.71
020124	3.20	6.1E-7 ± 1.0E-7	1.14E-5 ± 1.1E-6	4.10E-03 ± 1.09E-03	0.07 ± 0.06	0.266 ± 0.040	87^{+18}_{-12}	0.59 ± 0.17
020405	0.70	7.4E-6 ± 3.1E-7	1.10E-4 ± 2.1E-6	5.98E-03 ± 1.96E-03	$\cdots$	0.104 ± 0.007	364^{+90}_{-90}	0.48 ± 0.09
020813	1.25	3.8E-6 ± 2.6E-7	1.59E-4 ± 2.9E-6	1.14E-03 ± 2.92E-04	0.15 ± 0.01	0.164 ± 0.004	140^{+14}_{-13}	0.59 ± 0.05
021004	2.32	2.3E-7 ± 5.5E-8	3.61E-6 ± 8.6E-7	1.04E-02 ± 2.56E-03	0.71 ± 0.19	0.035 ± 0.067	80^{+53}_{-23}	1.23 ± 0.96
021211	1.01	2.3E-6 ± 1.7E-7	$\cdots$	$\cdots$	0.31 ± 0.01	0.006 ± 0.003	46^{+8}_{-6}	0.57 ± 0.01
030115	2.50	3.2E-7 ± 5.1E-8	$\cdots$	$\cdots$	0.44 ± 0.06	0.020 ± 0.020	83^{+53}_{-22}	0.70 ± 0.40
030226	1.98	2.6E-7 ± 4.7E-8	8.33E-6 ± 9.8E-7	2.72E-03 ± 6.82E-04	0.31 ± 0.22	0.033 ± 0.029	97^{+27}_{-17}	1.76 ± 1.15
030323	3.37	1.2E-7 ± 6.0E-8	$\cdots$	$\cdots$	$\cdots$	0.021 ± 0.338	44^{+90}_{-26}	$\cdots$
030328	1.52	1.6E-6 ± 1.1E-7	6.14E-5 ± 2.4E-6	1.96E-03 ± 4.92E-04	0.08 ± 0.08	0.024 ± 0.003	130^{+14}_{-13}	1.69 ± 0.81
030329	0.17	2.0E-5 ± 1.0E-6	2.31E-4 ± 2.0E-6	4.89E-03 ± 8.62E-04	0.15 ± 0.01	0.065 ± 0.002	68^{+2}_{-2}	0.66 ± 0.01
030429	2.66	2.0E-7 ± 5.4E-8	1.13E-6 ± 1.9E-7	5.76E-03 ± 2.79E-03	0.03 ± 0.17	0.220 ± 0.135	35^{+12}_{-8}	$\cdots$
030528	0.78	1.6E-7 ± 3.2E-8	$\cdots$	$\cdots$	12.56 ± 0.14	0.017 ± 0.010	32^{+5}_{-5}	2.13 ± 0.42
040924	0.86	2.6E-6 ± 2.8E-7	$\cdots$	$\cdots$	0.90 ± 0.01	0.060 ± 0.003	67^{+6}_{-6}	0.33 ± 0.17
041006	0.71	2.5E-6 ± 1.4E-7	1.75E-5 ± 1.8E-6	1.13E-03 ± 3.40E-04	$\cdots$	0.050 ± 0.002	63^{+13}_{-13}	1.28 ± 0.01
050126	1.29	1.07E-07 ± 1.56E-08	1.99E-06 ± 1.15E-07	$\cdots$	2.74 ± 0.02	−0.010 ± 0.065	47^{+23}_{-8}	1.58 ± 1.91
050223	0.59	1.18E-07 ± 1.66E-08	1.68E-06 ± 1.04E-07	$\cdots$	$\cdots$	0.111 ± 0.094	62^{+10}_{-10}	$\cdots$
050315	1.95	2.79E-07 ± 1.93E-08	7.52E-06 ± 2.07E-07	$\cdots$	$\cdots$	0.032 ± 0.016	39^{+7}_{-7}	1.97 ± 1.62
050401	2.90	1.74E-06 ± 9.09E-08	1.69E-05 ± 3.83E-07	2.20E-03 ± 7.52E-04	0.06 ± 0.02	0.187 ± 0.019	118^{+18}_{-18}	0.25 ± 0.16
050406	2.44	4.05E-08 ± 6.84E-09	1.41E-07 ± 1.77E-08	$\cdots$	$\cdots$	0.020 ± 0.274	25^{+35}_{-13}	$\cdots$
050408	1.24	1.1E-6 ± 2.1E-7	$\cdots$	$\cdots$	0.31 ± 0.02	0.082 ± 0.005	100^{+100}_{-50}	0.49 ± 0.02
050416A	0.65	5.41E-07 ± 3.24E-08	9.28E-07 ± 5.68E-08	1.45E-02 ± 8.38E-03	$\cdots$	0.021 ± 0.030	15^{+2}_{-3}	0.54 ± 0.06
050505	4.27	2.94E-07 ± 2.99E-08	5.23E-06 ± 2.29E-07	$\cdots$	0.71 ± 0.13	0.076 ± 0.031	70^{+140}_{-24}	0.60 ± 0.21
050525A	0.61	4.74E-06 ± 6.50E-08	2.44E-05 ± 2.14E-07	2.47E-03 ± 8.46E-04	0.12 ± 0.01	0.093 ± 0.003	81^{+1}_{-1}	0.32 ± 0.01
050603	2.82	8.01E-06 ± 2.42E-07	2.73E-05 ± 5.98E-07	$\cdots$	−0.01 ± 0.01	0.125 ± 0.014	344^{+52}_{-52}	0.19 ± 0.01
050730	3.97	1.02E-07 ± 1.58E-08	5.80E-06 ± 2.25E-07	$\cdots$	$\cdots$	0.027 ± 0.066	124^{+26}_{-26}	$\cdots$
050802	1.71	5.47E-07 ± 5.32E-08	5.24E-06 ± 2.50E-07	$\cdots$	$\cdots$	0.070 ± 0.036	121^{+28}_{-28}	2.03 ± 1.02
050814	5.30	1.04E-07 ± 2.24E-08	3.99E-06 ± 2.65E-07	$\cdots$	$\cdots$	−0.009 ± 0.180	60^{+24}_{-6}	$\cdots$
050820A	2.61	6.12E-07 ± 3.49E-08	1.09E-05 ± 4.66E-07	6.73E-03 ± 3.09E-03	$\cdots$	0.061 ± 0.033	246^{+76}_{-40}	1.01 ± 0.75
050824	0.83	7.92E-08 ± 1.44E-08	7.41E-07 ± 8.81E-08	$\cdots$	$\cdots$	0.289 ± 0.640	15^{+5}_{-5}	$\cdots$
050826	0.30	7.66E-08 ± 1.59E-08	1.12E-06 ± 1.19E-07	$\cdots$	$\cdots$	0.063 ± 0.105	105^{+47}_{-47}	1.11 ± 2.28
050908	3.35	9.83E-08 ± 1.20E-08	1.09E-06 ± 6.98E-08	$\cdots$	$\cdots$	−0.017 ± 0.046	41^{+9}_{-5}	1.10 ± 1.47
050922C	2.20	1.93E-06 ± 5.18E-08	5.09E-06 ± 1.03E-07	$\cdots$	0.06 ± 0.01	0.015 ± 0.003	198^{+38}_{-22}	0.13 ± 0.01
051016B	0.94	1.92E-07 ± 1.43E-08	4.31E-07 ± 3.39E-08	$\cdots$	$\cdots$	0.008 ± 0.030	24^{+7}_{-7}	$\cdots$
051022	0.80	1.1E-5 ± 8.7E-7	3.40E-4 ± 1.2E-5	0.0029 ± 0.0001	$\cdots$	0.088 ± 0.008	510^{+22}_{-20}	0.19 ± 0.04
051109A	2.35	8.30E-07 ± 8.83E-08	6.10E-06 ± 4.58E-07	$\cdots$	$\cdots$	−0.006 ± 0.025	161^{+130}_{-35}	0.70 ± 1.25
051111	1.55	7.61E-07 ± 3.65E-08	1.38E-05 ± 2.76E-07	$\cdots$	1.70 ± 0.07	0.009 ± 0.004	220^{+1703}_{-48}	1.80 ± 0.24
060108	2.03	1.22E-07 ± 1.16E-08	8.62E-07 ± 5.26E-08	$\cdots$	$\cdots$	0.006 ± 0.040	65^{+600}_{-10}	$\cdots$
060115	3.53	1.30E-07 ± 1.09E-08	3.76E-06 ± 2.01E-07	$\cdots$	$\cdots$	0.019 ± 0.029	62^{+19}_{-6}	1.11 ± 1.71
060206	4.05	4.41E-07 ± 1.63E-08	1.90E-06 ± 5.83E-08	$\cdots$	0.01 ± 0.03	0.007 ± 0.004	78^{+23}_{-8}	1.16 ± 0.18
060210	3.91	5.37E-07 ± 3.36E-08	1.97E-05 ± 6.39E-07	$\cdots$	0.15 ± 0.17	0.183 ± 0.033	149^{+400}_{-35}	0.73 ± 0.50
060223A	4.41	2.06E-07 ± 1.67E-08	1.51E-06 ± 6.54E-08	$\cdots$	$\cdots$	0.036 ± 0.021	71^{+100}_{-10}	0.41 ± 0.23
060418	1.49	1.49E-06 ± 4.85E-08	2.62E-05 ± 4.85E-07	$\cdots$	0.22 ± 0.03	0.104 ± 0.008	230^{+20}_{-20}	0.67 ± 0.08
060502A	1.51	3.72E-07 ± 2.81E-08	6.59E-06 ± 1.77E-07	$\cdots$	4.90 ± 0.11	0.004 ± 0.010	156^{+400}_{-33}	2.94 ± 1.19
060510B	4.90	9.51E-08 ± 1.12E-08	9.98E-06 ± 2.62E-07	$\cdots$	$\cdots$	0.110 ± 0.060	95^{+60}_{-30}	$\cdots$

GRB	z	$P_{\rm bolo}$ [erg/cm^{2}s]	$S_{\rm bolo}$ [erg/cm^2]	$F_{\rm beam}$	τ_{lag} [sec]	V	$E_{\rm peak}$ [keV]	τ_{RT} [sec]
060512	0.44	1.32E-07 ± 1.83E-08	6.04E-07 ± 6.34E-08	$\cdots$	$\cdots$	0.043 ± 0.173	22^{+6}_{-6}	$\cdots$
060522	5.11	8.73E-08 ± 1.45E-08	2.42E-06 ± 1.43E-07	$\cdots$	$\cdots$	0.034 ± 0.185	80^{+382}_{-12}	$\cdots$
060526	3.21	2.33E-07 ± 1.53E-08	3.01E-06 ± 2.40E-07	6.55E-03 ± 1.60E-03	0.17 ± 0.09	0.085 ± 0.030	25^{+5}_{-5}	0.38 ± 0.11
060604	2.68	5.10E-08 ± 1.19E-08	9.82E-07 ± 1.57E-07	$\cdots$	$\cdots$	0.080 ± 0.338	40^{+5}_{-5}	$\cdots$
060605	3.80	8.56E-08 ± 1.36E-08	1.58E-06 ± 1.24E-07	8.23E-04 ± 5.14E-05	$\cdots$	-0.013 ± 0.068	90^{+91}_{-12}	1.22 ± 0.72
060607A	3.08	2.66E-07 ± 1.50E-08	6.33E-06 ± 1.69E-07	$\cdots$	1.98 ± 0.11	0.025 ± 0.008	120^{+190}_{-17}	1.23 ± 0.68
060707	3.43	1.53E-07 ± 2.12E-08	3.41E-06 ± 1.96E-07	$\cdots$	$\cdots$	0.050 ± 0.054	63^{+13}_{-6}	$\cdots$
060714	2.71	2.30E-07 ± 1.42E-08	6.88E-06 ± 2.47E-07	$\cdots$	$\cdots$	0.125 ± 0.022	103^{+21}_{-16}	$\cdots$
060729	0.54	1.93E-07 ± 1.30E-08	6.43E-06 ± 3.16E-07	$\cdots$	$\cdots$	0.092 ± 0.041	61^{+9}_{-9}	$\cdots$
060814	0.84	1.83E-06 ± 4.44E-08	4.94E-05 ± 4.91E-07	$\cdots$	0.29 ± 0.03	0.040 ± 0.003	257^{+74}_{-35}	1.65 ± 0.24
060904B	0.70	4.37E-07 ± 2.28E-08	4.05E-06 ± 2.17E-07	$\cdots$	0.36 ± 0.09	0.003 ± 0.008	80^{+770}_{-12}	1.00 ± 0.16
060908	2.43	6.69E-07 ± 3.36E-08	7.68E-06 ± 1.85E-07	$\cdots$	0.26 ± 0.06	0.061 ± 0.008	151^{+112}_{-25}	0.52 ± 0.09
060926	3.21	1.56E-07 ± 1.22E-08	5.47E-07 ± 3.80E-08	$\cdots$	1.03 ± 0.11	0.148 ± 0.050	20^{+11}_{-11}	$\cdots$
060927	5.60	4.02E-07 ± 1.54E-08	2.37E-06 ± 8.67E-08	$\cdots$	0.12 ± 0.04	0.094 ± 0.010	72^{+15}_{-7}	0.46 ± 0.12
061007	1.26	7.20E-06 ± 1.11E-07	2.24E-04 ± 1.72E-06	$\cdots$	0.11 ± 0.01	0.066 ± 0.003	399^{+12}_{-11}	0.38 ± 0.02
061110A	0.76	9.79E-08 ± 1.35E-08	2.71E-06 ± 1.18E-07	$\cdots$	$\cdots$	-0.038 ± 0.050	90^{+13}_{-13}	$\cdots$
061110B	3.44	1.79E-07 ± 2.66E-08	6.12E-06 ± 3.38E-07	$\cdots$	0.24 ± 0.36	0.155 ± 0.064	517^{+53}_{-53}	0.79 ± 0.64
061121	1.31	8.04E-06 ± 1.07E-07	6.53E-05 ± 5.76E-07	$\cdots$	0.03 ± 0.01	0.050 ± 0.003	606^{+55}_{-44}	0.98 ± 0.19
061222B	3.36	2.29E-07 ± 3.15E-08	5.01E-06 ± 2.49E-07	$\cdots$	$\cdots$	0.024 ± 0.043	49^{+8}_{-8}	$\cdots$
070110	2.35	1.12E-07 ± 1.36E-08	4.04E-06 ± 1.64E-07	$\cdots$	$\cdots$	-0.010 ± 0.031	110^{+30}_{-30}	$\cdots$
070208	1.17	1.39E-07 ± 2.06E-08	1.06E-06 ± 1.46E-07	$\cdots$	$\cdots$	0.083 ± 0.211	51^{+10}_{-10}	$\cdots$
070318	0.84	4.10E-07 ± 2.12E-08	7.34E-06 ± 2.01E-07	$\cdots$	$\cdots$	0.037 ± 0.008	154^{+19}_{-19}	0.72 ± 0.24
070411	2.95	1.50E-07 ± 1.31E-08	6.29E-06 ± 2.19E-07	$\cdots$	$\cdots$	0.041 ± 0.029	83^{+11}_{-11}	$\cdots$
070506	2.31	1.67E-07 ± 1.38E-08	5.16E-07 ± 3.43E-08	$\cdots$	2.52 ± 0.04	0.010 ± 0.030	31^{+2}_{-3}	0.12 ± 0.06
070508	0.82	7.67E-06 ± 1.18E-07	7.26E-05 ± 6.15E-07	$\cdots$	0.04 ± 0.01	0.106 ± 0.003	233^{+7}_{-7}	0.20 ± 0.01
070521	0.55	2.09E-06 ± 5.26E-08	2.97E-05 ± 4.00E-07	$\cdots$	0.04 ± 0.01	0.116 ± 0.004	222^{+16}_{-12}	0.58 ± 0.06
070529	2.50	3.32E-07 ± 5.08E-08	7.44E-06 ± 4.31E-07	$\cdots$	$\cdots$	0.170 ± 0.091	180^{+52}_{-52}	$\cdots$
070611	2.04	1.45E-07 ± 2.25E-08	9.52E-07 ± 8.44E-08	$\cdots$	$\cdots$	0.053 ± 0.080	92^{+30}_{-30}	$\cdots$
070612A	0.62	2.77E-07 ± 4.24E-08	2.72E-05 ± 9.37E-07	$\cdots$	$\cdots$	0.032 ± 0.023	87^{+17}_{-17}	2.49 ± 1.48
070714B	0.92	3.24E-06 ± 1.46E-07	8.91E-06 ± 6.77E-07	$\cdots$	0.03 ± 0.01	0.164 ± 0.021	1120^{+473}_{-230}	0.45 ± 0.04
070802	2.45	6.38E-08 ± 9.69E-09	6.50E-07 ± 7.05E-08	$\cdots$	$\cdots$	-0.156 ± 0.150	70^{+25}_{-25}	$\cdots$
070810A	2.17	2.77E-07 ± 1.77E-08	1.59E-06 ± 8.43E-08	$\cdots$	1.09 ± 0.23	-0.006 ± 0.015	44^{+9}_{-9}	0.73 ± 0.22
071003	1.10	4.71E-06 ± 1.82E-07	6.73E-05 ± 1.48E-06	$\cdots$	0.38 ± 0.05	0.072 ± 0.007	799^{+75}_{-61}	0.88 ± 0.07
071010A	0.98	1.17E-07 ± 2.67E-08	4.97E-07 ± 6.05E-08	$\cdots$	$\cdots$	-0.076 ± 0.153	27^{+10}_{-10}	$\cdots$
071010B	0.95	9.20E-07 ± 2.18E-08	8.37E-06 ± 1.16E-07	$\cdots$	0.84 ± 0.04	0.010 ± 0.003	52^{+6}_{-8}	1.21 ± 0.03
071031	2.69	7.08E-08 ± 8.61E-09	2.19E-06 ± 1.92E-07	$\cdots$	$\cdots$	-0.038 ± 0.108	24^{+7}_{-7}	$\cdots$
071117	1.33	2.71E-06 ± 5.83E-08	7.97E-06 ± 2.02E-07	$\cdots$	0.60 ± 0.01	0.009 ± 0.003	278^{+143}_{-48}	0.20 ± 0.02
071122	1.14	6.76E-08 ± 2.06E-08	1.41E-06 ± 1.63E-07	$\cdots$	$\cdots$	0.391 ± 0.392	73^{+30}_{-30}	$\cdots$
080210	2.64	2.57E-07 ± 1.95E-08	4.17E-06 ± 1.41E-07	$\cdots$	0.53 ± 0.17	0.019 ± 0.013	73^{+15}_{-15}	0.57 ± 0.44
080310	2.43	1.83E-07 ± 1.72E-08	5.49E-06 ± 2.90E-07	$\cdots$	$\cdots$	0.038 ± 0.021	28^{+6}_{-6}	0.41 ± 0.55
080319B	0.94	1.55E-05 ± 1.91E-07	5.25E-04 ± 3.94E-06	$\cdots$	0.02 ± 0.01	0.031 ± 0.003	651^{+8}_{-8}	0.14 ± 0.01
080319C	1.95	2.22E-06 ± 7.79E-08	1.77E-05 ± 2.99E-07	$\cdots$	$\cdots$	0.042 ± 0.007	307^{+85}_{-56}	0.21 ± 0.12
080330	1.51	1.33E-07 ± 1.80E-08	8.77E-07 ± 1.26E-07	$\cdots$	$\cdots$	0.109 ± 0.060	20^{+9}_{-9}	$\cdots$
080411	1.03	1.04E-05 ± 1.31E-07	8.75E-05 ± 2.01E-07	$\cdots$	0.21 ± 0.01	0.167 ± 0.003	259^{+21}_{-16}	0.65 ± 0.01
080413A	2.43	1.22E-06 ± 2.65E-08	9.86E-06 ± 1.71E-07	$\cdots$	0.13 ± 0.03	0.078 ± 0.004	170^{+48}_{-24}	0.23 ± 0.03
080413B	1.10	3.17E-06 ± 8.25E-08	8.00E-06 ± 1.52E-07	$\cdots$	0.23 ± 0.01	0.004 ± 0.003	73^{+10}_{-10}	0.50 ± 0.03
080430	0.77	4.60E-07 ± 2.15E-08	3.01E-06 ± 1.53E-07	$\cdots$	0.68 ± 0.08	0.009 ± 0.004	80^{+15}_{-15}	0.76 ± 0.12
080516	3.20	2.77E-07 ± 2.80E-08	5.88E-07 ± 5.50E-08	$\cdots$	0.15 ± 0.01	0.168 ± 0.055	66^{+24}_{-24}	$\cdots$
080520	1.55	8.23E-08 ± 1.00E-08	1.59E-07 ± 3.00E-08	$\cdots$	$\cdots$	0.037 ± 0.098	12^{+5}_{-5}	$\cdots$
080603B	2.69	7.57E-07 ± 2.63E-08	7.02E-06 ± 1.78E-07	$\cdots$	0.08 ± 0.01	0.283 ± 0.010	85^{+55}_{-18}	0.22 ± 0.03
080605	1.64	5.99E-06 ± 1.10E-07	4.72E-05 ± 4.32E-07	$\cdots$	0.11 ± 0.01	0.057 ± 0.003	246^{+14}_{-11}	0.22 ± 0.01
080607	3.04	8.35E-06 ± 2.42E-07	1.00E-04 ± 0.00E+00	$\cdots$	0.04 ± 0.01	0.035 ± 0.003	394^{+35}_{-33}	0.18 ± 0.06
080707	1.23	1.68E-07 ± 1.02E-08	1.26E-06 ± 8.87E-08	$\cdots$	$\cdots$	0.093 ± 0.032	73^{+20}_{-20}	$\cdots$
080721	2.60	9.57E-06 ± 5.01E-07	5.99E-05 ± 3.04E-06	$\cdots$	0.13 ± 0.05	0.048 ± 0.009	485^{+41}_{-36}	0.09 ± 0.04
090423	8.2	2.17E-07 ± 1.55E-08	1.15E-06 ± 4.73E-08	$\cdots$	$\cdots$	$\cdots$	$48.6^{+3.8}_{-3.8}$	$\cdots$

Table 1: The data of 116 GRBs used in our analysis. For pre-*Swift* GRBs, we

# 2 THE ARCHITECTURE OF THE NEURAL NETWORK

Before testing the redshift dependence of luminosity correlations of GRBs, we firstly introduce the process of reconstructing the luminosity distance of SNe with deep learning. Based on the training Artificial Neural Networks (ANN), such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Bayesian Neural Networks

(BNN), deep learning is versatile, powerful and scalable in tackling complex problems such as classifying billions of images, recognizing speech, detecting subtle patterns in data, etc (Aurelien 2017). Recently, the application of deep learning in cosmological research is very extensive and successful (Dieleman et al. 2015; Mathuriva et al. 2018; Escamilla-Rivera et al. 2020). Following the work of Escamilla-Rivera et al. (2020), we reconstruct the distance moduli from the Pantheon compilation (Scolnic et al. 2018) with RNN+BNN. In this process, the reconstruction of distance only depends on the Pantheon dataset, and without any assumption on the cosmological model.

RNN is a class of nets which can predict the future from the complex sequential information without any model assumption, but is incapable of estimating the uncertainty of target. This shortcoming can be fixed up with BNN. Therefore, our neural network is composed of RNN and BNN, the details of which are described bellow.

RNN is one of the supervised learning algorithm by training the neural networks with the real data, reaching an ideal network characterizing the relationship between the target and the feature by minimizing the loss function (Aurelien 2017). In our work, the Pantheon data set is used as the training data, in which the redshift is the feature and the distance module is the target. In RNN, the activation not only flows from the input layer to the output layer, but also has connections pointing backward. The architecture of RNN is shown in Figure In the unrolled RNN, the neurons at each time step t receive the inputs as well as the outputs from the previous time step (Aurelien 2017). In the neural network, the loss function is used to depict the difference between the targets and the predicts. We adopt the Mean Squared Error (MSE) function as the loss function and find the minimum with the Adam optimizer.

Handling long sequences, the training of RNN will take a long time and the information of initial inputs will gradually fades away Aurelien 2017). Thus, we adopt the time step $t = 4$ to alleviate the long training time, and use the most popular basic cell called Long Short-Term Memory (LSTM) cell to solve the problem of information loss. RNN with LSTM cell is aware of what to store, throw away and read. The computations of LSTM are
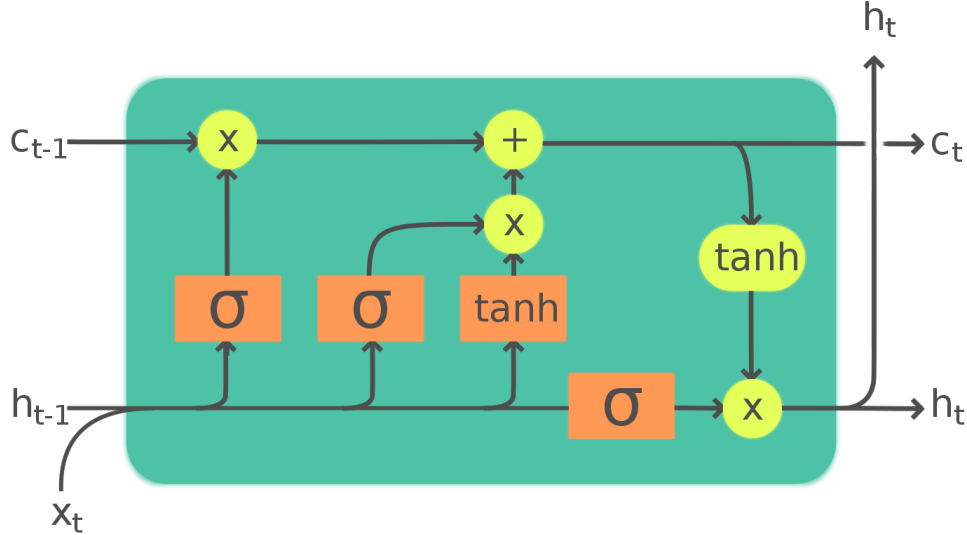
$$i^{<t>} = \sigma \left(W_{xi}^T \cdot x^{<t>} + W_{hi}^T \cdot h^{<t-1>} + b_i \right)$$
$$f^{<t>} = \sigma \left(W_{xf}^T \cdot x^{<t>} + W_{hf}^T \cdot h^{<t-1>} + b_f \right)$$
$$o^{<t>} = \sigma \left(W_{xo}^T \cdot x^{<t>} + W_{ho}^T \cdot h^{<t-1>} + b_o \right)$$
$$g^{<t>} = A_f \left(W_{xg}^T \cdot x^{<t>} + W_{hg}^T \cdot h^{<t-1>} + b_g \right)$$

$$c^{<t>} = f<t> \otimes c^{<t-1>} + i^{<t>} \otimes g^{<t>},$$
$$y^{<t>} = h^{<t>} = o^{<t>} \otimes A_f \left(c^{<t>} \right),$$

where σ is the sigmoid function that outputs a value between 0 and 1, t is the time step referring to the current

sequence (for example $t = 1$ for the first redshift). The superscript $<t>$ indicates a vector of steps t, the superscript T is the transpose of the matrix, the dot is matrix product and $\otimes$ is direct product. $x^{<t>}$ and $y^{<t>}$ are respectively the current input and output vectors. $h<t>$ and $c^{<t>}$ are respectively the short-term state and long-term state of LSTM cells. A_f is an activation function to make the network be capable of solving complex tasks by introducing the non-linearity to network. In our work, we use the tanh activation function, which is defined as

$$A_{f_{\mathrm{Tanh}}} = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$
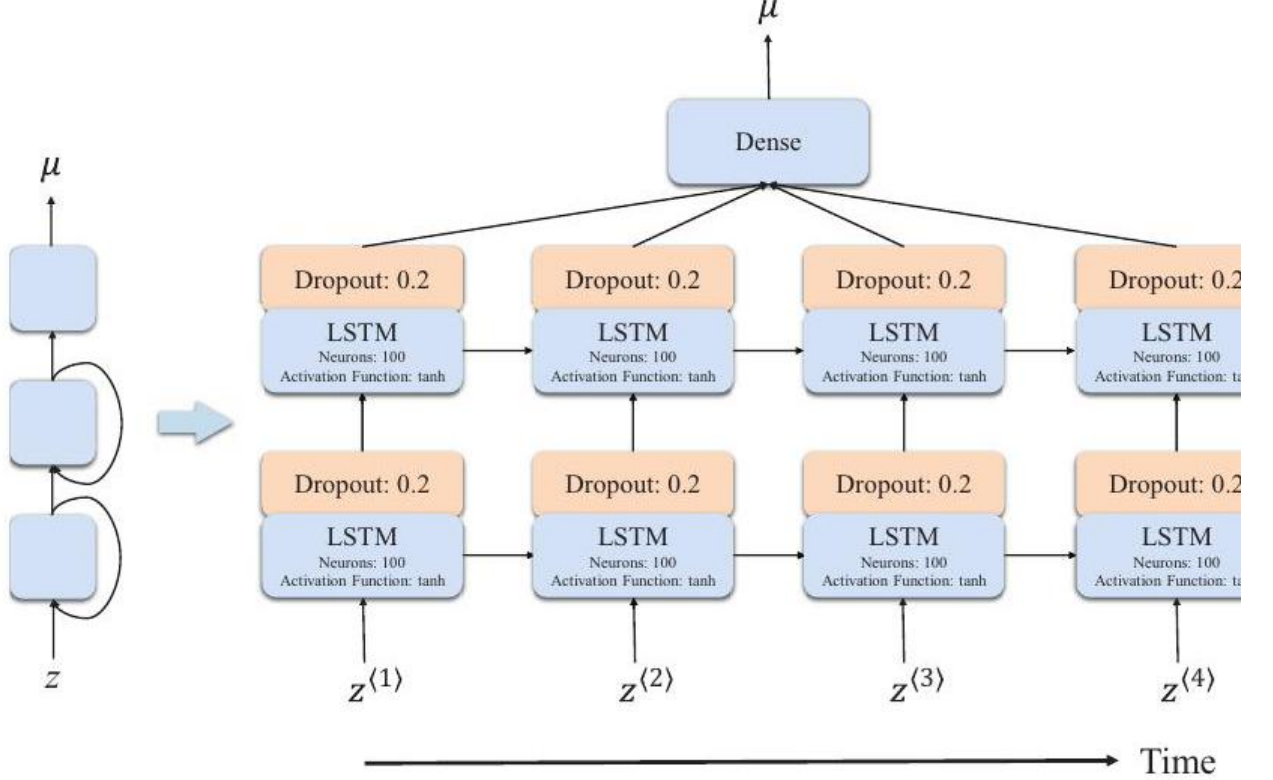
FIG. 1. - The architecture of our network with one hidden layer (left), unrolled through time step $t = 4$ (right). In the unrolled network, each column is one of the t time steps, while the three rows from bottom to top represent input layer, hidden layer and output layer, respectively. The first two layers with tanh activation function consist of LSTM cell containing 100 neurons, while the output layer is a fully-connected (dense) layer. To avoid overfitting, the dropout technique is employed between LSTM and its next layers, and we set the dropout rate to 0.2.

There are four connected layers playing different roles, where the main layer that outputs $g^{<t>}$ analyzes the current inputs $x^{<t>}$ and the previous state $h^{<t-1>}$, the rest three layers are gate controllers: (a) Input gate controlled by $i^{<t>}$ determines which parts of $g^{<t>}$ should be added to $c^{<t>}$, (b) Forget gate controlled by $f<t>$ determines which parts of $c^{<t>}$ should be abandoned, (c) Output gate controlled by $o<t>$ determines which parts of $c<t>$ should be output. It can be easily found that, these gate controllers are related to the logistic activation function σ, thus they would close the gate if output 0 and open it if output 1. W_{xi}, W_{xf}, W_{xo} and W_{xg} are the weight matrices of each of above four layers connecting to the input vector. W_{hi}, W_{hf}, W_{ho}, and W_{hg} are the weight matrices of each of layers connecting to the previous short-term state. b_i, b_f, b_o, and b_g are the bias terms for each of layer.

In a deep neural network, the training may suffer from overfitting due to a large number of its own hyperparameters. We can use the method called regularization to prevent it from overfitting. Dropout is one of the most popular regularization techniques, applying in some layers to reduce the overfitting risk (Aurelien 2017). In this way, some neurons has a probability of being ignored at every step controlled by dropout rate. Besides, it is also of benefit to estimate the confidence of the training in BNN.

BNN is a supplementary of RNN for calculating the uncertainty of the prediction. BNN is defined in terms of a prior distribution with parameters over the wights $p(\omega)$, which manifests a prior belief about parameters generating the observations. With a given dataset $\{\mathbf{X}, \mathbf{Y}\}$, we can achieve the posterior distribution of the parameters space $p(\omega \mid \mathbf{X}, \mathbf{Y})$. Thus the output of a new input point x can be anticipated by the integration

$$p\left(y^* \mid x^*, \mathbf{X}, \mathbf{Y}\right) = \int p\left(y^* \mid x^*, \omega\right) p(\omega \mid \mathbf{X}, \mathbf{Y}) \mathrm{d}\omega$$

A full BNN is extremely complex. Several works had shown that a network with a dropout is approximately equivalent to the Bayesian model (Gal & Ghahramani 2016a.b c; Louizos & Welling 2016). Introducing the Bayesian machinery

into the deep learning framework, Gal & Ghahramani (2016ab, c) developed a new framework casting dropout training in deep neural network as approximate Bayesian inference in deep Gaussian processes and successfully applied in RNN. Their results offer a Bayesian interpretation of the dropout technique, and verify that a network with a dropout is mathematically equivalent to the Bayesian model. When the RNN is well-trained and executed n times, the network is equivalent to BNN. Therefore, we employ the d ropout in the training and call the trained network n times to estimate the uncertainty of outputs, where the dropout is an approximation of the Gaussian processes and cooperates with the activation function to determine the confidence regions of prediction.

3   RECONSTRUCTING THE DISTANCE-REDSHIFT RELATION FROM PANTHEON

In order to reconstruct the Hubble diagram to high redshift range with our network, we use the latest Pantheon compilation (Scolnic et al. 2018) of SNe Ia as the training data. The Pantheon compilation consists of 1048 wellcalibrated SNe Ia in the redshift range $0.01 < z < 2.3$. The distance modulus of SNe Ia is given by

$$\mu_B\left(z; \alpha, \beta, M_b\right) = m - M_b + \alpha x(z) - \beta c(z),$$

where m is the apparent magnitude, M_b is the absolute magnitude, x and c are the stretch factor and color parameter respectively, α and β are the nuisance parameter. In the Pantheon dataset, the presented apparent magnitude has already been corrected for stretch and color, thus the stretch and color corrections are vanishing in eq. (9)). The absolute magnitude is fixed to $M_b = -19.36$. We obtain the distance moduli $\mu(z)$ from eq. (9) and sort the data points according to the redshift from low to high.

We construct the RNN+BNN network and train it with the package TensorFlow. For clarity, we present the corresponding hyperparameters in Figure 1 and list the steps to reconstruct data with our network as follow: (a) Data processing. The scale of data has an effect on training. Hence, we normalize the distance moduli of the sorted Pantheon data and re-arrange $\mu - z$ as sequences with the step number $t = 4$. (b) Building RNN. We build RNN with three layers, i.e. an input layer, a hidden layer and an output layer as described in Figure 1 The first two layers are constructed with the LSTM cells of 100 neurons. The redshifts $z^{<t>}$ and the corresponding distance moduli $\mu < t >$ are the input and output vectors, respectively. We employ the Adam optimizer to minimize the cost function MSE and train the network 1000 times. (c) Building BNN. We set the dropout rate to 0 in the input layer to avoid the lost of information, and to 0.2 in the second layer as well as the output layer (Bonjean 2020 ; Mangena et al. 2020). We execute the trained network 1000 times to obtain the distribution of distance moduli.

The highest redshift of GRB sample (Wang et al. 2011) is 8.2. Therefore, we reconstruct the distance moduli by training RNN+BNN up to $z = 8.5$. The result is shown in Figure 2 The red dots with error bars and the light-blue dots are the observational data points of Pantheon and the central values of reconstruction, respectively. The shaded regions are the 1σ and 2σ uncertainties of the reconstruction. The inside plot at the lower-right corner of Figure 2 is the MSE loss over the training epochs for both the train (blue) and test (orange) sets, which approach their minimum value when epoch near 400 . For comparison, we also plot the best-fitting curve of the Λ CDM model (black line). The result shows that, our reconstruction is excellently consistent with the Λ CDM model within 1σ confidence level. We want to emphasize again that our reconstruction of distance moduli is neither dependent on the cosmological model, nor has any assumption about the concrete form of the reconstructing function. It is the architecture itself makes the neural network become a universal approximator, not the specific activation function (Hornik 1991). We note that the Gaussian processes can also reconstruct the distance moduli without involving any model assumption (Lin et al. 2018). However, the reconstructed uncertainty of Gaussian processes is very large in the region where the data points are sparse. The advantage of RNN + BNN compared with the Gaussian processes is that, the former can reconstruct the curve precisely even far beyond the data points.

To investigate the influence of different choices of activation function on the reconstruction, we use Monte Carlo simulations. We first simulate a set of Pantheon-like sample, whose redshifts and distance uncertainties are the same to that of Pantheon sample, and the distance moduli are sampled from Gaussian distribution $\mu \sim G\left(\bar{\mu}, \sigma_\mu\right)$, where $\bar{\mu}$ is calculated using the fiducial Λ CDM model with $\Omega_M = 0.3$ and $H_0 = 70$ km/s/Mpc, and σ_μ is the uncertainty of the observed distance modulus of Pantheon sample. Then we replace the Pantheon data with the mock data and retrain the network, and then reconstruct the distance-redshift relation using the trained network. We also consider other three activation functions as a comparison with tanh function:

$$A_{f_{\text{Relu}}} = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$$

$$A_{f_{\text{Elu}}} = \begin{cases} \alpha\left(e^x - 1\right), & x \leq 0 \\ x, & x > 0 \end{cases}$$
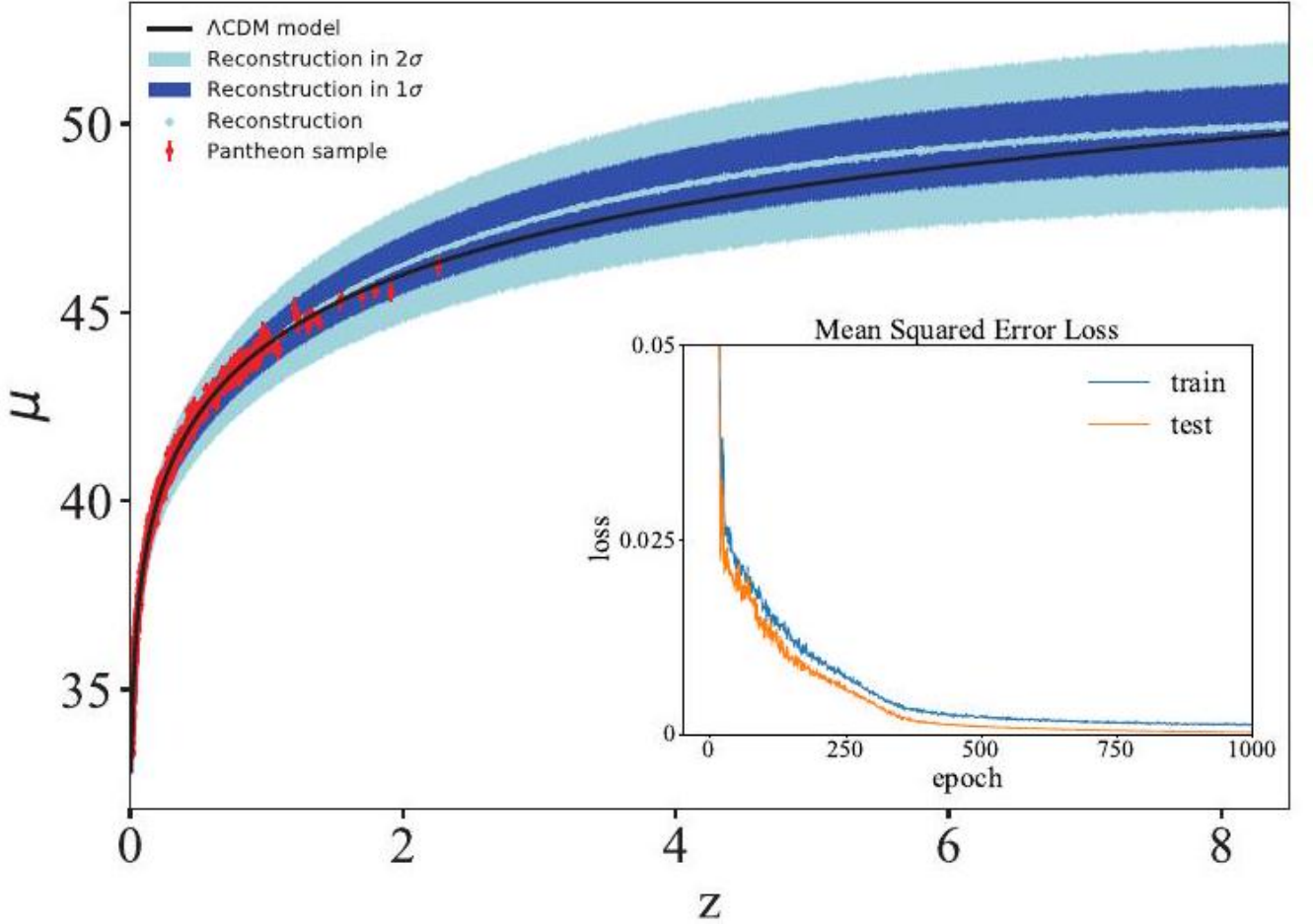
FIG. 2. - The reconstruction of distance moduli from Pantheon data set. The red dots with 1σ error bars are the Pantheon data points. The light-blue dots are the central values of reconstruction. The shaded regions are the 1σ and 2σ uncertainties. The black curve is the best-fitting curve of flat Λ CDM model. Inside plot: the Mean Squared Error (MSE) loss over the training epochs for both the train (blue) and test (orange) sets.

$$A_{f_{\text{Selu}}} = \begin{cases} \alpha\lambda\left(e^x - 1\right), & x \leq 0 \\ \lambda x, & x > 0 \end{cases}$$

where in Elu function $\alpha = 1$ (Clevert et al. 2015), and in Selu function $\alpha \approx 1.673$ and $\lambda \approx 1.051$ (Klambauer et al. 2017). The reconstructed results with four different activation functions are presented in Figure 3 It is shown that only with the tanh function our network can correctly reconstruct the distance-redshift relation up to $z \approx 8.5$ within 1σ uncertainty. For the rest three activation functions, the network couldn't correctly reconstruct the curve at high redshift. Therefore, we choose the tanh function rather than the rest three in our work. Note that the similarity between the elu and selu functions makes the reconstructed curves be similar to each other.

We fix the activation function to tanh to investigate the influence of other hyperparameters on our network. Firstly, we fix the number of layers to 3 and choose three different number of neurons $\{50, 100 150\}$ for comparison. We also fix the number of neurons to 100 and choose three different number of layers $\{2, 3, 4\}$ for comparison. The training losses of different cases are presented in Figure 4. For the same number of layers, the training loss of 50 neurons is distinctly higher than that of 100 and 150 neurons. The increase of neurons from 100 to 150 does not significantly reduce the training loss. For the same number of neurons, the training loss of 2 layers is much higher than that of 3 and 4 layers, but the training losses of 3 layers and 4 layers are not significantly distinguishable. Therefore, our choice of 100 neurons and 3 layers is appropriate to construct a well-trained network.

4   TESTING THE REDSHIFT DEPENDENCE OF LUMINOSITY CORRELATIONS

After reconstructing the distance-redshift relation from SNe, we can use it to calibrate the luminosity correlations of GRBs. The luminosity correlations of GRB can be expressed with the exponential form $R = AQ^b$, and further be

re-expressed with the linear form by taking logarithms,

$$y = a + bx \quad (y \equiv \log R, x \equiv \log Q, a \equiv \log A),$$

in which 'log' denotes the logarithm of base 10 . We definitely write the six luminosity correlations as follows

$$\log \frac{L}{\text{erg s}^{-1}} = a_1 + b_1 \log \frac{\tau_{\text{lag},i}}{0.1s},$$

$$\log \frac{L}{\text{erg s}^{-1}} = a_2 + b_2 \log \frac{V_i}{0.02},$$

$$\log \frac{L}{\text{erg s}^{-1}} = a_3 + b_3 \log \frac{E_{p,i}}{300\text{keV}}$$
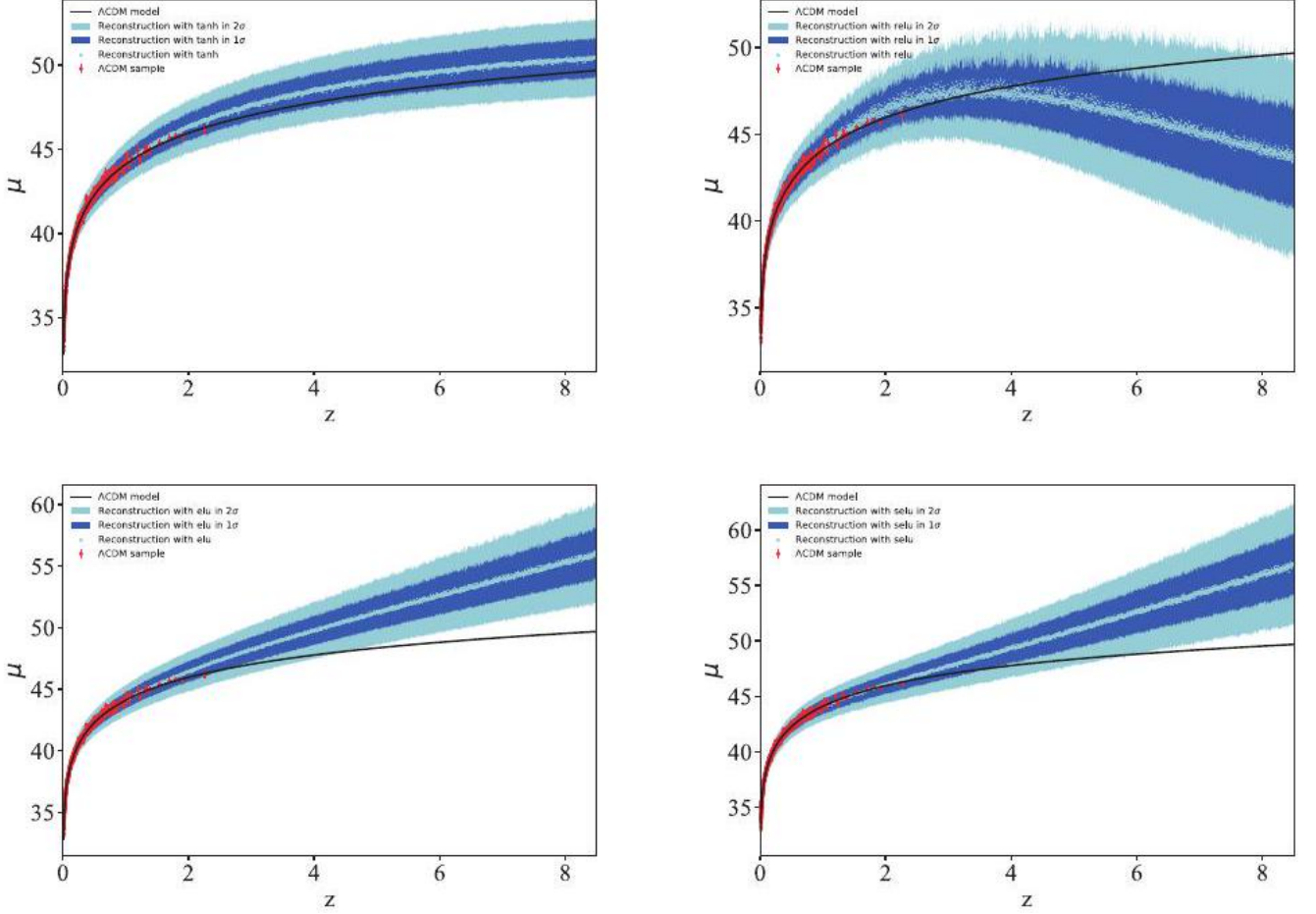
FIG. 3. - The reconstructions of distance-redshift relation from Pantheon-like sample (generated from fiducial Λ CDM model) with four different activation functions. Top-left: tanh; top-right: relu; bottom-left: elu; bottom-right: selu. This figure shows that the reconstruction depends on the activation functions.

$$\log \frac{E_\gamma}{\text{erg}} = a_4 + b_4 \log \frac{E_{p,i}}{300\text{keV}},$$

$$\log \frac{L}{\text{erg s}} = a_5 + b_5 \log \frac{\tau_{\text{RT},i}}{0.1 \text{ s}},$$

$$\log \frac{E_{\text{iso}}}{\text{erg}} = a_6 + b_6 \log \frac{E_{p,i}}{300\text{keV}},$$

Here, quantities with a subscript ' i ' indicate that they are measured in the comoving frame, which are related to the quantities in the observer frame by $\tau_{\text{lag},i} = \tau_{\text{lag}}(1+z)^{-1}, \tau_{\text{RT},i} = \tau_{\text{RT}}(1+z)^{-1}, V_i = V(1+z)$ and $E_{p,i} = E_p(1+z)$, where $\tau_{\text{lag}}, \tau_{\text{RT}}, V, E_p$ can be directly derived from the observations of spectrum or light curve of GRBs.

Assuming that GRBs radiate isotropically, the isotropic equivalent luminosity can be derived from the bolometric peak flux P_{bolo} by (Schaefer 2007)

$$L = 4\pi d_L^2 P_{\text{bolo}} ,$$

where d_L is the luminosity distance of GRB, which can be obtained from the reconstructed distance moduli of Pantheon presented in section B with the relation

$$\mu = 5\log\frac{d_L}{\text{Mpc}} + 25.$$

Hence, the uncertainty of L propagates from the uncertainties of P_{bolo} and d_L. The isotropic equivalent energy E_{iso} can be obtained from the bolometric fluence S_{bolo} by

$$E_{\text{iso}} = 4\pi d_L^2 S_{\text{bolo}} (1+z)^{-1},$$

the uncertainty of E_{iso} propagates from the uncertainties of S_{bolo} and d_L. If on the other hand, GRBs radiate in two symmetric beams, then we can define the collimation-corrected energy E_γ as

$$E_\gamma \equiv E_{\text{iso}} \, F_{\text{beam}} \, ,$$
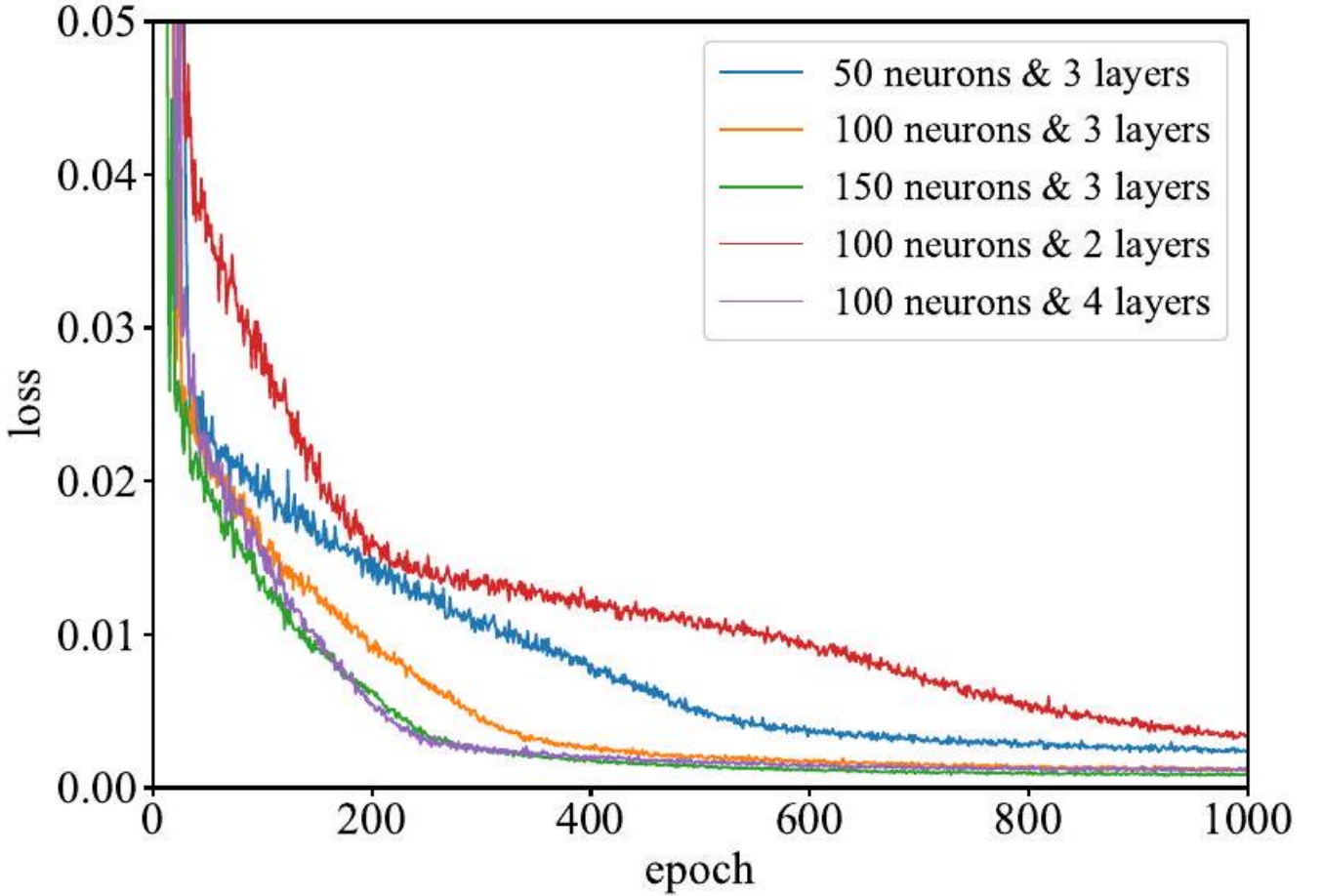
FIG. 4. - The training loss in the cases of different numbers of neurons and layers with the activation function fixed to tanh.

where $F_{\text{beam}} \equiv 1 - \cos\theta_{\text{jet}}$ is the beaming factor, θ_{jet} is the jet opening angle. The uncertainty of E_γ propagates from the uncertainties of $E_{\text{iso and }} F_{\text{beam}}$.

Our GRBs sample is taken from Wang et al. (2011), which consists of 116 long GRBs in redshift range $z \in [0.17, 8.2]$. Following Lin et al. (2016), we divide GRBs into two subsamples, i.e. the low-z sample ($z \leq 1.4$) which consists of 50GRBs, and the high-z sample ($z > 1.4$) which consists of 66 GRBs. We investigate the redshift dependence of luminosity correlations for this two subsamples, as well as for the full GRBs sample. To fit the six luminosity correlations, we apply the D'Agostini's liklihood (D'Agostini 2005)

$$\mathcal{L}(\sigma_{\text{int}}, a, b) \propto \prod_i \frac{1}{\sqrt{\sigma_{\text{int}}^2 + \sigma_{yi}^2 + b^2\sigma_{xi}^2}} \times \exp\left[-\frac{(y_i - a - bx_i)^2}{2\left(\sigma_{\text{int}}^2 + \sigma_{yi}^2 + b^2\sigma_{xi}^2\right)}\right]$$

By maximizing this joint likelihood function, we can derive the best-fitting parameters $(a, b, \sigma_{\text{int}})$, where the intrinsic scatter σ_{int} denotes any other unknown errors except for the measurement errors.

We perform a Markov Chain Monte Carlo analysis to calculate the posterior probability density function (PDF) of parameter space. We assume a flat prior on all the free parameters and limit $\sigma_{\text{int}} > 0$. Note that not all GRBs can be used to analyze each luminosity correlation, because not all the necessary quantities are measurable for some GRBs. For example, GRBs without measurement of the spectrum lag can not used in the $\tau_{\text{lag}} - L$ analysis. Hence, we present the best-fitting parameters, together with the number of available GRBs in each fitting in Table 1 In Figure 5 we plot all the six luminosity correlations in logarithmic coordinates. Low-z and high-z GRBs are represented by blue and red dots with the error bars denoting 1σ uncertainties. The blue line, red line and black line stand for the best-fitting results for low-z GRBs, high-z GRBs and all-z GRBs, respectively. The 1σ and 2σ contours and the PDFs for parameter space are plotted in Figure 6

As shown in Table 1 low-z GRBs have a smaller intercept, but a sharper slope than high-z GRBs for all the six luminosity correlations. All-z GRBs have the parameter values between that of low-z and high-z subsamples. For the intrinsic scatter, low-z GRBs have larger value than high-z GRBs, and the $E_p - E_\gamma$ relation has the smallest intrinsic scatter hence we can only obtain its upper limit. The $V - L$ relation has the largest intrinsic scatter, thus it can not be fitted well with a simple line, which is legible in Figure [5 In Figure 6 the contours in the (a, b) plane indicate that the $E_p - E_\gamma$ relation of low-z GRBs is consistent with that of high-z GRBs at 1σ confidence level. For the rest luminosity correlations, however, the intercepts and slopes for low-z GRBs differ from that of high-z GRBs at more than 2σ confidence level.

Having luminosity correlations calibrated, we can conversely using these correlations to calibrate the distance of GRBs, and further use GRBs to constrain cosmological models. Since our calibration of luminosity correlations is independent of cosmological model, the circularity problem is avoided. As we have seen, the $E_p - E_\gamma$ relation is not significantly evolving with redshift, so we use this relation to calibrate the distance of GRBs. Due to that the

TABLE 1

THE BEST-FITtING PARAMETERS OF GRB LUMINOSITY CORRELATIONS. N IS THE NUMBER OF GRBS IN EACH SUBSAMPLE.

Correlation	Sample	N	a	b	σ_{int}
$\tau_{\text{lag}} - L$	low-z	37	52.076 ± 0.106	-0.789 ± 0.159	0.495 ± 0.0888
	high-z	32	52.610 ± 0.069	-0.633 ± 0.115	0.144 ± 0.099
	All-z	69	52.313 ± 0.073	-0.757 ± 0.114	0.462 ± 0.059
$V - L$	low-z	47	51.524 ± 0.202	0.440 ± 0.301	0.863 ± 0.084
	high-z	57	52.355 ± 0.131	0.264 ± 0.125	0.480 ± 0.073
	All-z	104	51.784 ± 0.140	0.558 ± 0.160	0.747 ± 0.070
$E_p - L$	low-z	50	51.855 ± 0.092	1.481 ± 0.185	0.576 ± 0.075
	high-z	66	52.495 ± 0.053	1.151 ± 0.139	0.192 ± 0.088
	All-z	116	52.167 ± 0.059	1.454 ± 0.136	0.530 ± 0.049
$E_p - E_\gamma$	low-z	12	50.602 ± 0.088	1.538 ± 0.192	< 0.176
	high-z	12	50.768 ± 0.126	1.162 ± 0.377	< 0.238
	All-z	24	50.650 ± 0.067	1.486 ± 0.157	< 0.137
$\tau_{\text{RT}} - L$	low-z	39	52.671 ± 0.132	-1.354 ± 0.194	0.458 ± 0.074
	high-z	40	52.880 ± 0.080	-0.816 ± 0.161	0.249 ± 0.097
	All-z	79	52.766 ± 0.078	-1.250 ± 0.132	0.424 ± 0.054
$E_p - E_{\text{iso}}$	low-z	40	52.536 ± 0.099	1.600 ± 0.201	0.557 ± 0.085
	high-z	61	53.013 ± 0.055	1.293 ± 0.135	0.192 ± 0.088
	All-z	101	52.778 ± 0.058	1.546 ± 0.129	0.461 ± 0.051

parameters of $E_p - E_\gamma$ relation in three samples (low-z, high-z and full-z samples) are consistent with each other, we directly apply the parameters obtained with all-z sample. The distance of 24 GRBs calibrated using $E_p - E_\gamma$ are shown in the Hubble diagram in Figure 7 With the Pantheon dataset, the matter density of the flat Λ CDM model is constrained to be $\Omega_{\text{M}} = 0.278^{+0.007}_{-0.008}$. With 24 long GRBs alone, the matter density is constrained to be $\Omega_{\text{M}} = 0.307^{+0.065}_{-0.073}$. It indicates that the Hubble diagram in high redshift is consistent with the ΛCDM model.

5 DISCUSSIONS AND CONCLUSION

We have investigated the redshift dependence of six luminosity correlations in long GRBs using deep learning. We first reconstructed the distance-redshift relation to high redshift from the Pantheon compilation using RNN, and derived the uncertainty using BNN. Then the luminosity distance of GRBs is obtained without any assumption about the cosmological model. To test the possible redshift dependence of luminosity correlations, we divided GRBs into low-z and high-z subsamples and investigate the correlations for each subsample separately. It is found that, for all six luminosity correlations, low-z subsample has a smaller intercept, but a sharper slope than high-z subsample. In four out of six correlations ($\tau_{\text{lag}} - L$, $E_p - L$, $\tau_{\text{RT}} - L$ and $E_p - E_{\text{iso}}$), the intercept and slope for low-z subsample and high-z subsample differ at more than 2σ. For the $V - L$ relation, the intrinsic scatter is too large to make a reliable

conclusion. For the $E_p - E_\gamma$ relation, there is no evidence for redshift evolution, and the intrinsic scatter is the smallest amono the six correlations. However, the number of available GRBs for the $E_p - E_{\text{relation}}$ is small, because most long GRBs lack the measurement of jet opening angle which is necessary to calculate the beaming factor. The constraint on flat A CDM model from 24 GRBs calibrated 1 ing $E_p - E_{\text{relation oives}}$ $\Omega_M = 0.307 + 0.065$ which is The constraint on flat Λ CDM model from 24 GRBs calibrated using $E_p - E_\gamma$ relation gives $\Omega\Omega_M = 0.307 + 0.065$, which is

well consistent with the result constrained from Pantheon. Our network is trained using Pantheon sample, so GRBs calibrated using the network should be consistent with Pantheon sample.

In the RNN+BNN network, the network is trained using the Pantheon data to extend the Hubble diagram to high redshift range without any assumption about the cosmological model. In the deep learning method, the kernel is achieving the most efficient expression to denote the relationship between the features and the targets of Pantheon data, where the validity is reflected in the architecture of network. In the cooperation of the Monte Carlo dropout and the activation function, our network not only predicts the central value, but also the corresponding confidence region. Because the number of SNe Ia whose redshift beyond 1.4 is small, the uncertainty in the poor observational data range would be large. Therefore, we choose a small dropout rate 0.2 with the specific activation function $A_{f_{\text{Tanh}}}$ to acquire a relatively small uncertainty.

The method proposed here has some advantages compared with other methods already existing in literatures. Lin et al. (2016) have tested the redshift-dependence of six luminosity correlations by dividing GRBs into low- z and high-z subsamples. In their work, the slope and intercept parameters (a, b) are calculated based on a specific cosmological model. In contrast, our method proposed is completely independent on cosmological model. Several works have been devoted to calibrating the distance of GRBs in a model-independent way (Firmani et al. 2005: Liang et al. 2005: Capozziello & Izzo 2010 . Wei 2010 . Liu & Wei 2015 . All of these methods rely on the assumption that the luminosity correlations have no redshift evolution, which couldn't be tested by the method itself. To calibrate high- z GRBs, one must first obtain the luminosity correlation coefficients (a, b) from the low- z GRB, then directly extrapolate the correlation to high- z GRBs. Wang et al. (2011) proposed to calibrate GRBs by fitting the coefficients (a, b) and the correlation to high-z GRBs. Wang et al. (2011) proposed to calibrate GRBs by fitting the coefficients (a,b) and the This calibrating method is still model-dependent, and the GRB calibrated in one cosmological model couldn't be used
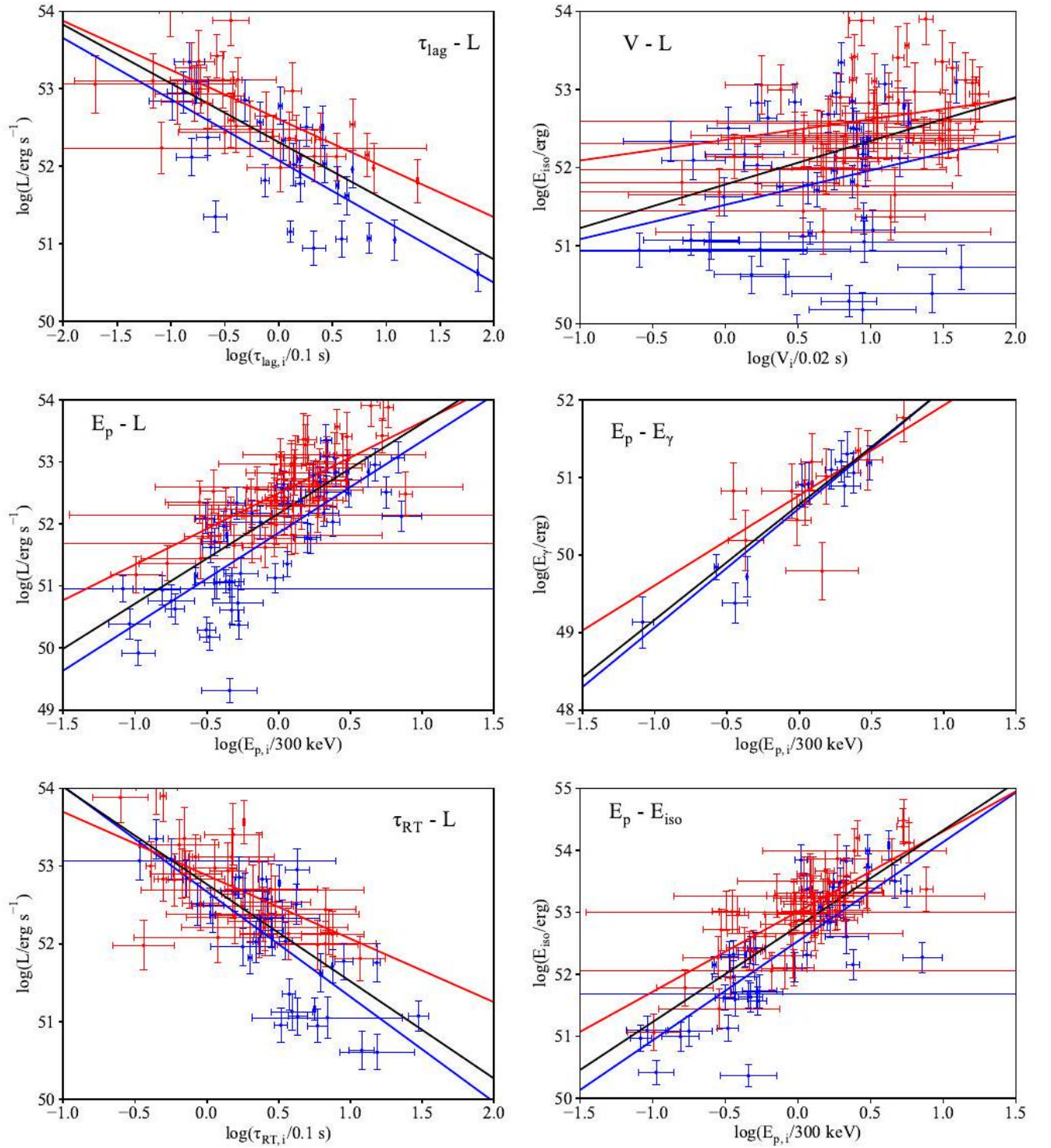
FIG. 5. - The luminosity correlations for low- z (blue dots) and high- z (red dots) GRBs. Error bars denote the 1σ uncertainties. The lines are the best-fitting results, blue line for low- z GRBs, red line for high- z GRBs and black line for all- z GRBs.

to constrain other models. To constrain other cosmological models, the coefficients (a, b) should be refitted again. In comparison, our method can be used to test the redshift evolution of luminosity correlations model-independently. If a correlation is redshift-independent, low- z and high- z GRBs can be calibrated simultaneously using this correlation. GRBs calibrated in this way can be directly used to constrain cosmological models.

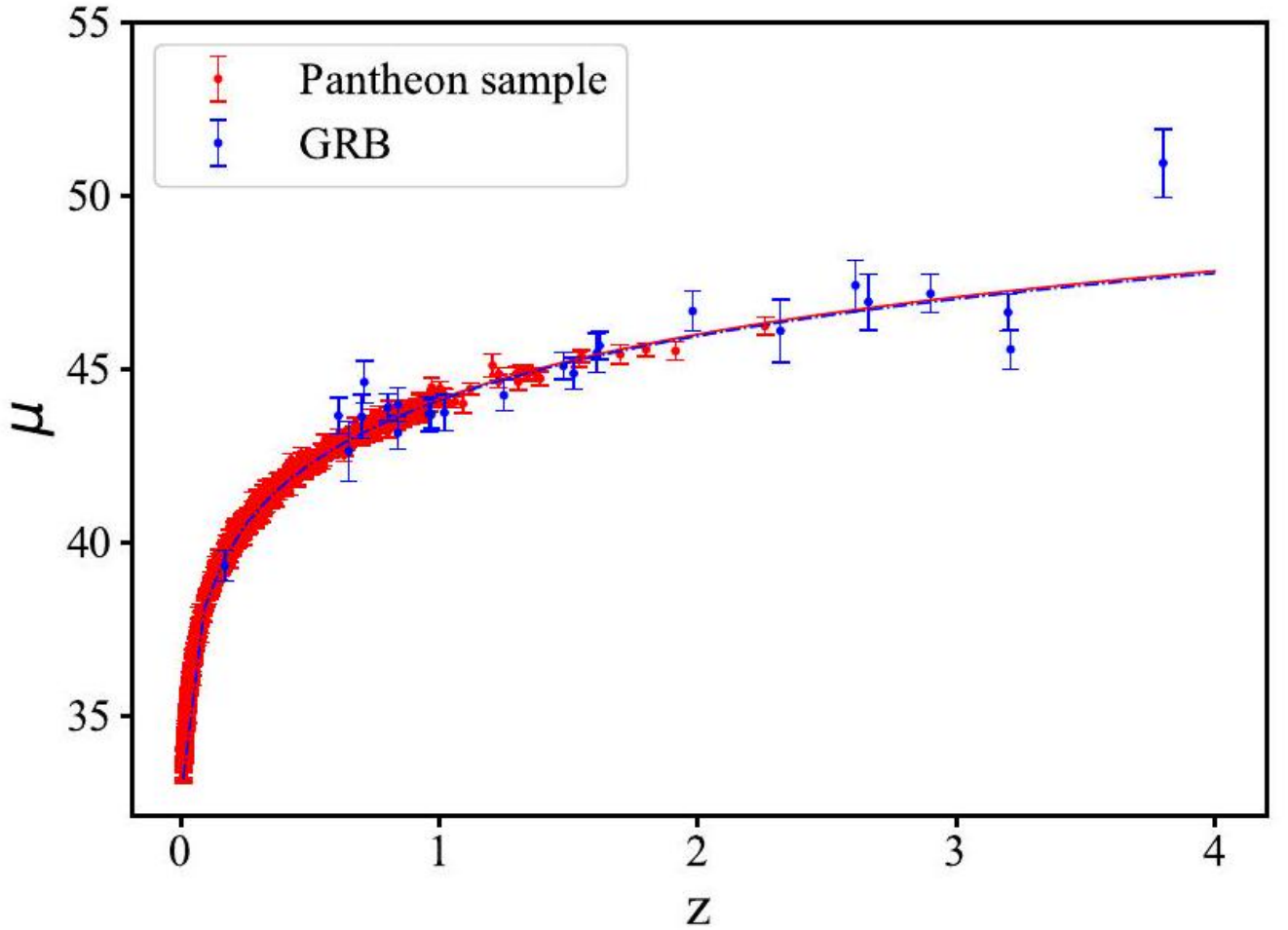FIG. 6. - The confidence contours and marginalized PDFs for the parameters.

FIG. 7. - The Hubble diagram of GRBs. The red dots with 1σ error bars are the Pantheon data points. The blue dots with with 1σ error bars are 24 long GRBs calibrated using the $E_p - E_\gamma$ relation. The lines are the best-fitting results to Λ CDM model with Pantheon (red line) and GRBs (blue dot-dashed line), respectively. Note: the two best-fitting lines seem to overlap with each other.

6 REFERENCES

Amati L. et al., 2002, A&A, 390, 81

Aurelien G., Hands-On Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build

Intelligent Systems, O'Reilly Media (2017)

[ISBN-10:1491962291]

Basilakos S., Perivolaropoulos L., 2008, MNRAS, 391, 411

Betoule M., Kessler R., Guy J. et al., 2014, Astron. Astrophys, 568, A22

Bloom J. S., Frail D. A., Kulkarni S. R., 2003, ApJ, 594, 674

Bonjean V., 2020, Astron. Astrophys, 634, A81

Cai R. G., Ma Y.Z., Tang B., Tuo Z. L., 2013, Phys. Rev. D, 87, 123522

Capozziello S., Polarski, and Izzo L., 2010, Astron. Astrophys. 519, A73

Chang Z., Li X., Lin H.-N., Wang S., 2014, Mod. Phys. Lett. A, 29,1450067

Chevallier M., Polarski, 2001, Int. J. Mod. Phys. D, 10, 213

Clevert D.-A., Unterthiner T., Hochreiter S., 2015,

arXiv:1511.07289

Cucchiara A. et al., 2011, ApJ, 736, 7

D'Agostini G., 2005, preprint (arXiv:physics/0511182

Dai Z. G., Liang E. W., Xu D., 2004, Ap.J., 612, L101

Demianski M., Demianski E., 2011, MNRAS, 415, 3580

Dieleman S., Willett K. W., and Dambre J., 2015, MNR.AS, 450 , 1441

Dolgov A., Halenka V., Tkachev I., 2014, J. Cosmol. Astropart. Phys. 1410,047

Escamilla-Rivera C.. Quintero M.A.C., and Capozziello S., 2020 , JCAP, 2020, 008

Fenimore E.E., Ramirez-Ruiz E., 2000, preprint astro-ph /0004176

Firmani *C.*, *Ghisellini G.*, *Ghirlanda G.*, Avila-Reese V., 2005 , MNRAS, 360, L1

Friedman A.S., Bloom J. S., 2005, ApJ, 627, 1

Gal Y. & Ghahramani Z., 2016, preprint $\sqrt{\text{arXiv:1506.02142}}$

Gal Y. & Ghahramani Z., 2016, preprint arXiv:1506.02157

Gal Y. & Ghahramani Z., 2016, preprint (arXiv:1512.05287)

Ghirlanda G., Ghisellini G., Lazzati D., Firmani C., 2004, Ap.J. 613, *L*13

Ghirlanda G., Ghisellini G., Lazzati D., 2004, ApJ, 616, 331

Hornik K., 1991 , Neural Networks, 4, 251

Izzo L., Capozziello S., Covone G., Capaccidi M., 2009, A&A, 508,63

Klambauer, G.. Unterthiner T., Mayr A., & Hochreiter S., 2017, Self-Normalizing Neural Networks. In Advances in Neural Information Processing Systems (NIPS) (arXiv:1706.02515 v5)

Li L.-X., 2007, MNRAS, 379, L55

Liang E. W., Zhang B. 2005, Ap. 633,611

Liang N., Xiao W. K., Liu Y., Zhang S. N., 2008, ApJ, 685, 354

Liang N., Wu P. X., Zhu Z. H., 2010, Res. Astron. Astrophys., 11, 1019

Lin H.-N., Li X., Wang S., Chang Z., 2015, MNRAS, 453. 128

Lin H.-N., Li X., Chang Z., 2016, MNRAS, 455, 2131

Lin H.-N., Li M.-H., Li X., 2018, MNRAS, 480, 3117

Liu, J., Wei, H., Gen Relativ Gravit, 2015, 47, 141

Louizos C. and Welling M., 2016, arXiv:1603.04733

Mangena T., Hassan S., Santos M G., 2020, MNRAS, 494, 600

Mathuriya A. et al., 2018, preprint (arXiv:1808.04728)

Norris J. P., Marani G.F., Bonnell J. T., 2000, ApJ, 534, 248

Pan S., Weiqiang Yang W.-Q. and Paliathanasis A., 2020, Eur. Phys. J. C, 80, 274

Perlmutter S. et al., 1999, ApJ, 517, 565

Riess A. G. et al., 1998, AJ, 116, 1009 Schaefer B. E., 2007, ApJ, 660, 16

Scolnic D. M. et al., Astrophys. J., 2018, 859, 101

Suzuki N., Rubin D., Lidman C.et al., 2012, ApJ, 746, 85

Tutusaus I., Lamine B., Blanchard A., et al., 2016, Phys. Rev. D, 94,103511

Wang F.Y., Dai Z.G., 2006. MNRAS,368, 371

Wang F. Y., Qi S., Dai Z. G., 2011, MNRAS, 415, 3423

Wang S., Wang Y., Li M., 2017, Phys. Rep. 696, 1

Wei H. & Zhang S. N., 2009. Eur. Phys. J. C, 63. 139

Wei J. J.. WuX. F.. Melia F.,2016. Mon. Not. Roy. Astron. Soc. 463,1144 Wei H., 2010, J. Cosmol. Astropart. Phys., 08, 020

Xu D., Dai Z. G., Liang E.W., 2005, ApJ, 633, 603

Yonetoku D., Murakami T., Nakamura T., Yamazaki R., Inoue A. K., Ioka K., 2004, ApJ, 609, 935

Zhu Z. H., Hu M., Alcaniz J. S., Liu Y. X.. 2008, Astron. Astrophys. 483,15