

## COSMOLOGICAL PARAMETER ESTIMATION FROM LARGE-SCALE STRUCTURE MACHINE LEARNING

SHUYANG PAN, MIAOXIN LIU, ZIYONG WU, HAITAO MIAO, XIAOLIN LUO, XIAO-DONG LI,<sup>1</sup> CRISTIANO G. SABIU,<sup>2</sup> AND JAIME FORERO-ROMERO<sup>3</sup>

<sup>1</sup>*School of Physics and Astronomy, Sun Yat-Sen University, Guangzhou 510297, P.R.China*

<sup>2</sup>*Department of Astronomy, Yonsei University, Seoul, Korea*

<sup>3</sup>*Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10 Edificio Ip, CP 111711, Bogotá, Colombia*

(Received November 6, 2017; Revised November 27, 2017; Accepted June 24, 2019)

Submitted to ApJ

### ABSTRACT

*Keywords:* large-scale structure of Universe — dark energy — cosmological parameters

## 1. INTRODUCTION

Machine learning is useful ...

LSS is non-linear... To build the connection between the LSS and the underlying cosmological parameters, machine learning should be useful...

Pioneering works have been done in (Ravanbakhsh et al. 2017). ...

We tested ...

## 2. DATA

Training sample:

$0.16 \leq \Omega_m \leq 0.46$ , step size 0.01, 31 numbers;

$2.0 \leq 10^9 A_s \leq 2.3$ , step size 0.02, 15 numbers;

For reference, Planck 2015 (TT,TE,EE+lowP+lensing) gives  $\Omega_m = 0.3121 \pm 0.0087$ ,  $10^9 A_s = 2.13 \pm 0.053$ ,  $\sigma_8 = 0.8150 \pm 0.0087$

In total  $31*15 = 465$  cosmologies

Each cosmology: one COLA simulation,  $128^3$  particles,  $256^3$  box, COLA simulation, 40 timesteps

Redshift  $z = 0$  sample used in machine learning.

Test sample:

sing-cosmology test: 500 BigMD cosmologies ( $\Omega_m, \sigma_8 = (0.3072, 0.8228)$ ). multi-cosmology test: all settings the same as training, except that using different initial condition

## 3. METHODOLOGY

We use the following architecture...

The three convolution layers and three dense layers have (896, 55,360, 221,312) and (1,049,600, 262,400, 514) parameters.

We varied the options...

Learning curve...

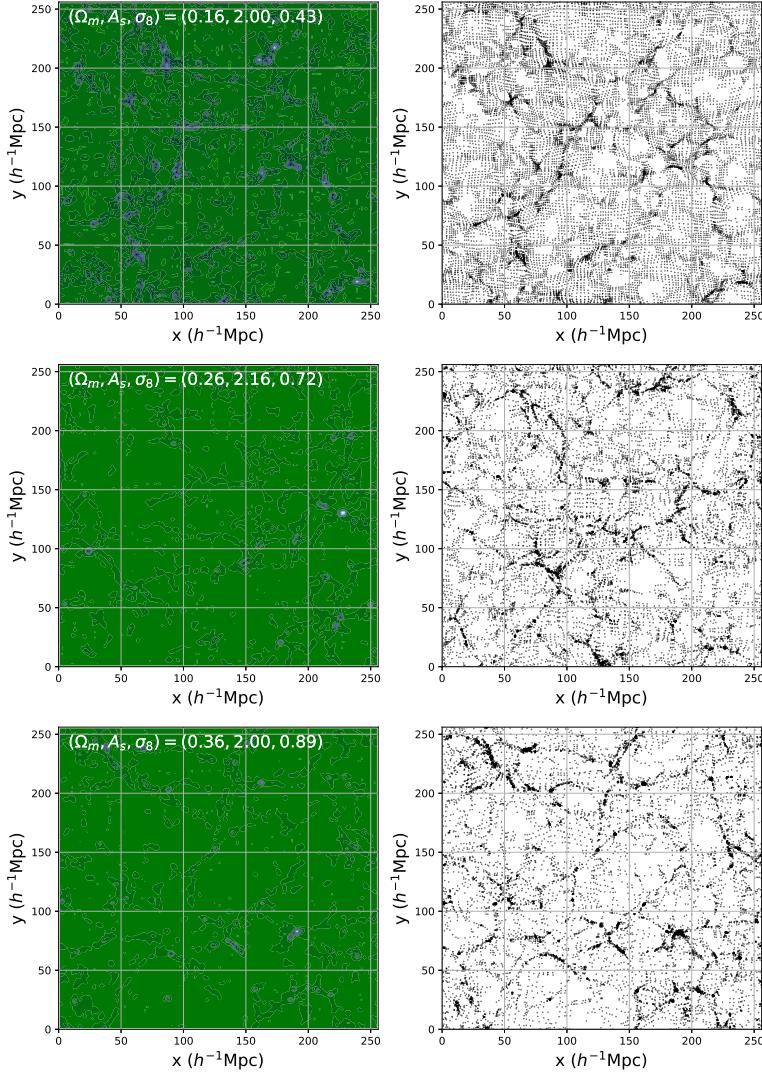
## 4. RESULTS

## 5. CONCLUDING REMARKS

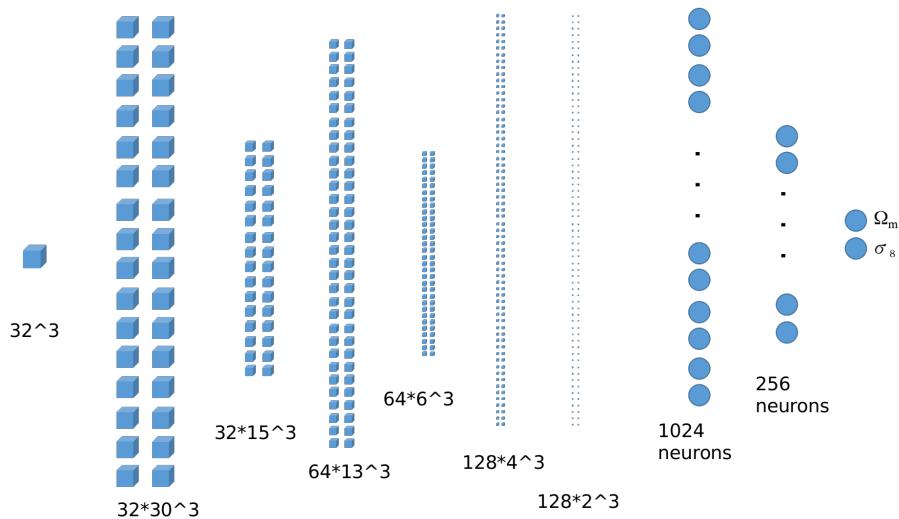
We thank Cheng Cheng, Xin Wang and Kwan-Chuen Chan for helpful discussions. XDL acknowledges the supported from NSFC grant (No. 11803094).

## REFERENCES

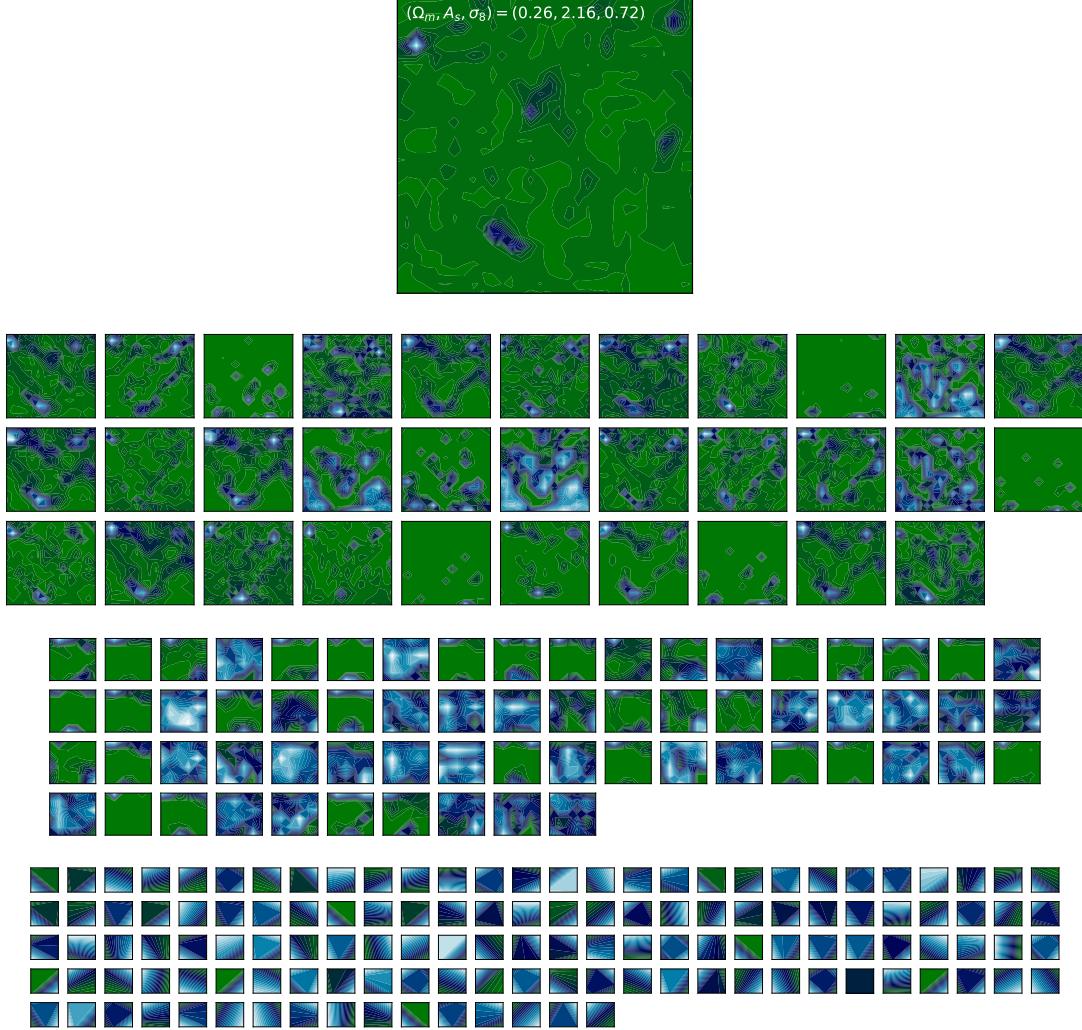
- Ade, P.A.R., Aghanim, N., & Arnaud, M., et al. arXiv:1502.01589
- Aghamousa, A., 2016, arXiv:1611.00036
- Alam, S., Ata, M., & Bailey, S., et al. 2016, submitted to MNRAS (arXiv:1607.03155)
- Alcock, C., & Paczynski, B. 1979, Nature, 281, 358 We did not discuss the systematics of the AP method in details.
- Anderson, L., Aubourg, É., & Bailey, S. et al. 2014, MNRAS, 441, 24
- Ballinger, W.E., Peacock, J.A., & Heavens, A.F. 1996, MNRAS, 282, 877
- Betoule, M., Kessler, R., & Guy, J., et al. 2014, A&A, 568, 32
- Beutler, F., Blake, C., & Colless, M., et al. 2011, MNRAS, 416, 3017
- Bond, J. R., Efstathiou, G., & Tegmark M. 1997, MNRAS, 291, L33
- Blake, C., Glazebrook, K., & Davis, T. M., 2011, MNRAS, 418, 1725
- Crittenden, R.G. et al. 2012, JCAP, 02, 048
- Crittenden, R. G., Pogosian, Li, & Zhao, G.-B., 2009. JCAP, 0912, 025
- Crittenden, R. G., Zhao, G.-B., & Pogosian, Li, et al. 2012. JCAP, 1202, 048
- Delubac, T. et al. 2015, Astron. & Astrophys. 574, A59
- Kodi Ramanah, D., Lavaux, G, Jasche, J., & Wandelt, B.D., et al. 2019, A&A, 621, A69
- Efstathiou, G. 2014, MNRAS, 440, 1138
- Feng, B., Wang., X. L., & Zhang, X. M., 2005. Phys. Lett. B., 607, 35
- Heymans, C. et al. 2013, Mon. Not. R. Astron. Soc. 432, 24332453. 1303.1808.
- Kim, J., Park, C., L'Huillier, B., & Hong, S. E. 2015, JKAS, 48, 213
- Kitaura, F.S., Rodriguez-Torres, S., Chuang, C.-H., et al. arXiv:1509.06400
- Laureijs, R., Amiaux, J., & Arduini, S., et al. 2011, arXiv:1110.3193
- Lavaux, G., & Wandelt, B.D. 2012, ApJ, 754, 109
- Li, M., Li, X.-D., Wang, S., & Wang, Y. 2011, Commun. Theor. Phys., 56, 525
- Li, X.-D., Park, C., Forero-Romero, J., & Kim, J. 2014, ApJ, 796, 137
- Li, X.-D., Park, C., Sabiu, C.G., & Kim, J. 2015, MNRAS, 450, 807
- Li, X.-D., Park, C., & Sabiu, C.G., et al. 2016, ApJ, 832, 103
- Li, X.-D., Sabiu, C.G., & Park, C., et al. 2018, ApJ, 856, 88
- Li, X.-D., Miao, H, & Wang, X., et al. 2019, submitted to ApJ, arXiv:1903.04757
- Mao, Q., Berlind, A.A., Scherrer, R.J., et al. 2016, submitted to ApJ
- Marco R., 2019, arXiv:1902.01366



**Figure 1.** The density field (left) and particle distribution (right) of three cosmologies  $(\Omega_m, A_s, \sigma_8) = (0.16, 2.0, 0.43)$ ,  $(0.26, 2.16, 0.72)$ ,  $(0.36, 2.0, 0.89)$ , selected within the training sample. Thin slices with  $0h^{-1}\text{Mpc} < z < 2h^{-1}\text{Mpc}$  are plotted. The clustering strength increases when increasing  $\Omega_m$  or  $A_s$ , making the structures more compact. We train the neuron network to build up the connection between the density field and underlying cosmological parameters.



**Figure 2.** Architecture of our CNN (default settings). We split the  $128^3$  voxels box into 64 sets of  $32^3$  voxels sub-boxes, and then feed them to the CNN. After three convolution+pooling – with 32, 64, 128 filters in each convolution, respectively, we got  $128 \times 2^3$  voxels as the extracted features of the initial field. These voxels are then connected with three dense layers with 1028, 24 and 2 neurons, to predict  $\Omega_m$  and  $\sigma_8$ .



**Figure 3.** Feature extraction in the CNN, taking the cosmology  $(\Omega_m, A_s, \sigma_8) = (0.26, 2.16, 0.72)$  as the example. The CNN was feeded with a  $32^3$ -box, whose  $z < 2h^{-1}\text{Mpc}$  slice is plotted in the uppermost panel. It was firstly convolved with 32  $3^3$ -filters, and then pooled to generate 32  $15^3$ -boxes. They capturing different types of features in the original box. Then, the 32 boxes are further convolved with 64  $3^3$ -filters, which are also pooled to generate 64  $6^3$ -boxes, containing more compressed features. Finally, the last convolution+pooling outputs 128  $2^3$ -boxes, containing the most compressed information. These information are passed to the dense layers to produce values of  $\Omega_m$  and  $\sigma_8$ . The number of free parameters in the three convolutions are 896, 55,360, 221,312, respectively. These free parameters are tuned in the training process so that the convolutions can extract the features most sensitive to the  $\Omega_m$  and  $\sigma_8$  values.

Marshall, Phil, Anguita, Timo, & Bianco, F. B., et al. 2017,  
arXiv:1708.04058

Matsubara T., & Suto, Y. 1996, ApJ, 470, L1

Moresco, M. et al. 2016, J. Cosmol. Astropart. Phys. 5, 014.  
1601. 01701.

Outram, P.J., Shanks, T., Boyle, B.J., Croom, S.M., Hoyle,  
F., Loaring, N.S., Miller, L., & Smith, R.J. 2004,  
MNRAS, 348, 745

Park, C., & Kim, Y.-R. 2010, ApJL, 715, L185

Parkinson, D. et al. 2012, Phys. Rev. D 86, 103518.  
1210.2130.

Perlmutter, S., Aldering, G., & Goldhaber, G., et al. 1999,  
ApJ, 517, 565

Ravvanbakhsh, S., Oliva, J., Fromenteau, S., et al. 2017,  
arXiv:1711.02033

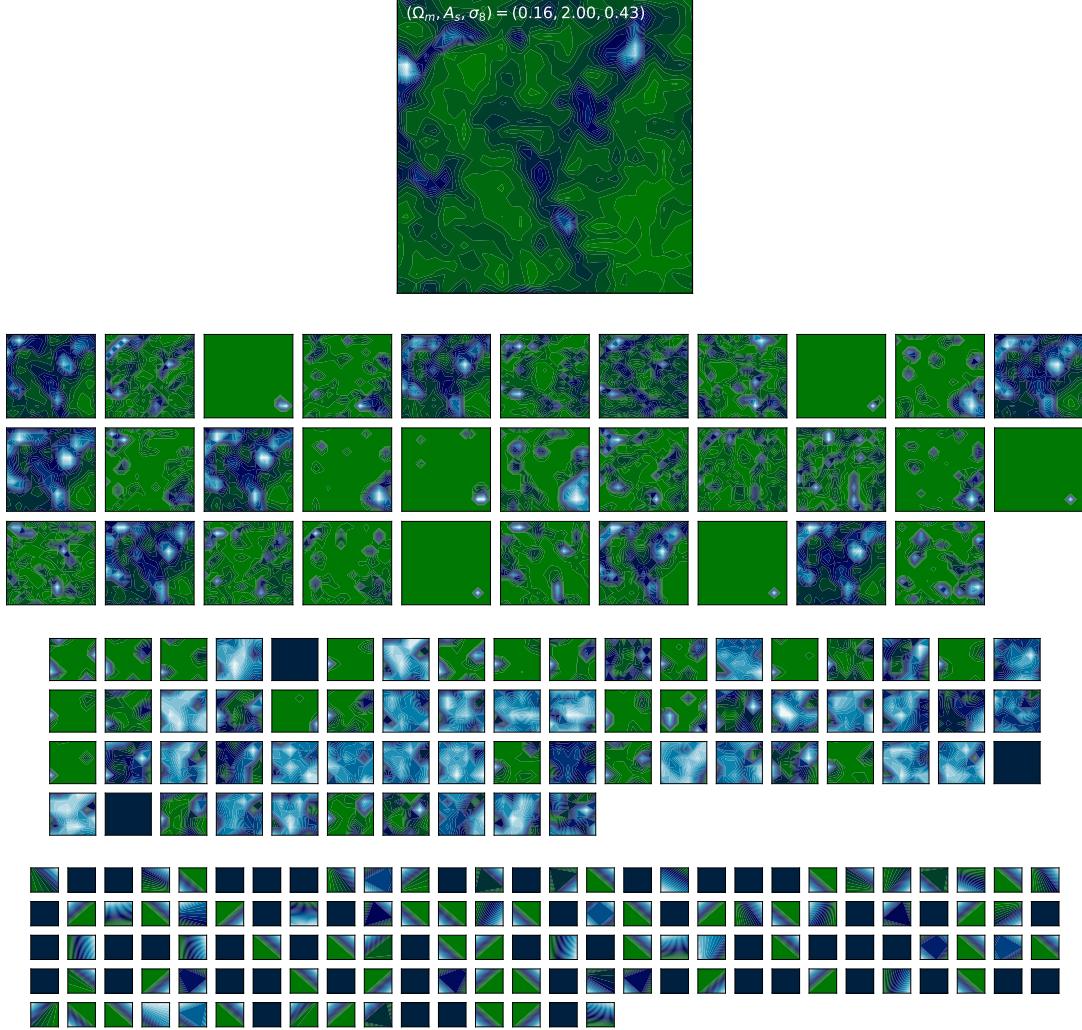
Riess, A.G., Filippenko, A.V., & Challis, P., et al. 1998,  
AJ, 116, 1009

Riess, A.G., Macri, L., & Casertano, S., et al. 2011, ApJ,  
730, 119

Ross, A.J., Samushia, L., & Howlett, C., et al. 2015,  
MNRAS, 449, 835

Ryden, B.S. 1995, ApJ, 452, 25

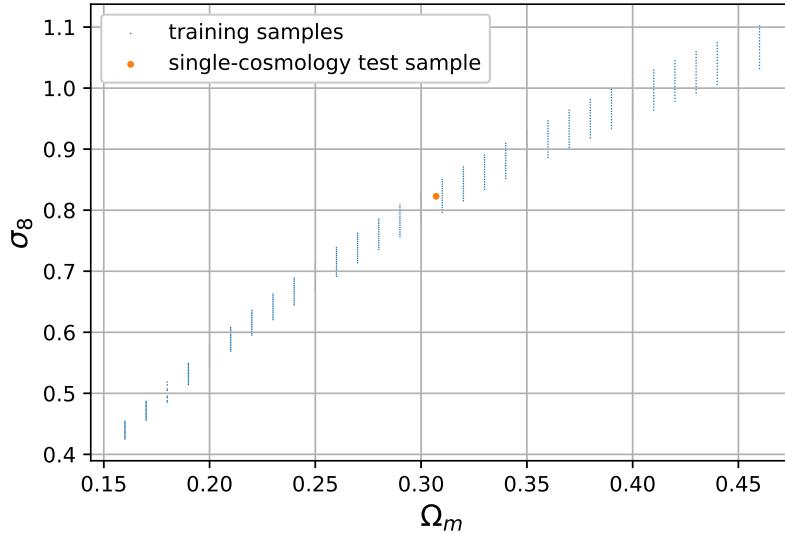
Weinberg, S. 1989, Reviews of Modern Physics, 61, 1



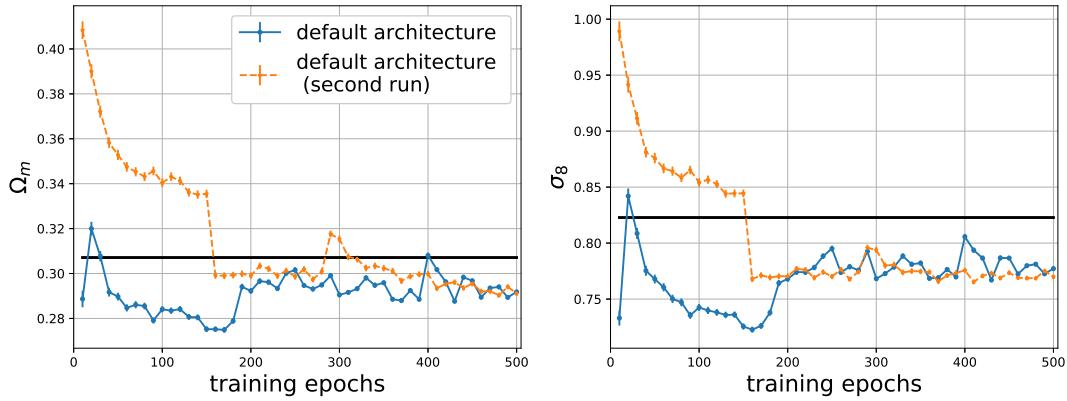
**Figure 4.** The same as Figure 3, except that the feature extraction was acted on the cosmology  $(\Omega_m, A_s, \sigma_8) = (0.26, 2.00, 0.43)$ . The features extracted by the convolutions are different from those in the cosmology  $(\Omega_m, A_s, \sigma_8) = (0.26, 2.16, 0.72)$ . In this way, the CNN was able to tell that the input cosmology is different from the previous one, and outputs different  $\Omega_m$  and  $\sigma_8$  values for different cosmologies.

- Weinberg, D.H., Mortonson, M.J., Eisenstein, D.J., et al.  
2013, Physics Reports, 530, 87  
Wang, Y., Pogosian, L., Zhao, G.-B., & Zucca, A. 2018.  
accepted by ApJL  
Yoo, J., & Watanabe, Y. 2012, International Journal of  
Modern Physics D, 21, 1230002  
Zhang, X., Huang, Q.-G., & Li, X.-D. 2018. Mon. Not. R.  
Astron. Soc., 483, 1655

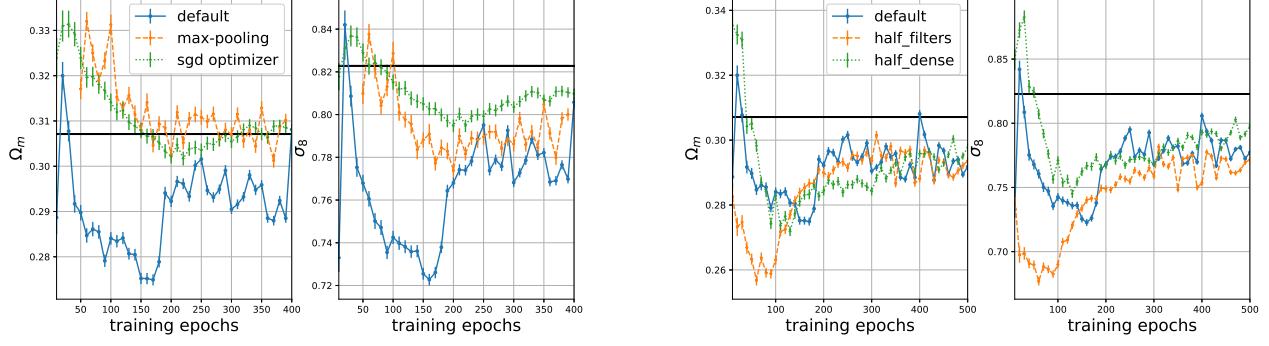
- Zhao, G.-B., Crittenden, R.-G., Pogosian, L., & Zhang, X,  
2012. Phys. Rev. Lett., 109, 171301  
Zhao, G.-B., Raveri, M., & Pogosian, L., et al. 2017, Nat.  
Astron., 1, 627  
Zhao, G.-B. et al. 2017, Mon. Not. R. Astron. Soc. 466, 762  
Zhao, G.-B. et al. 2019, MNRAS, 482, 3497



**Figure 5.** Cosmological parameters of the 465 training samples (little blue dots) and single-cosmology test samples (big yellow dot). The multi-cosmology test samples have exactly the same parameters of the training samples, created using different initial conditions.



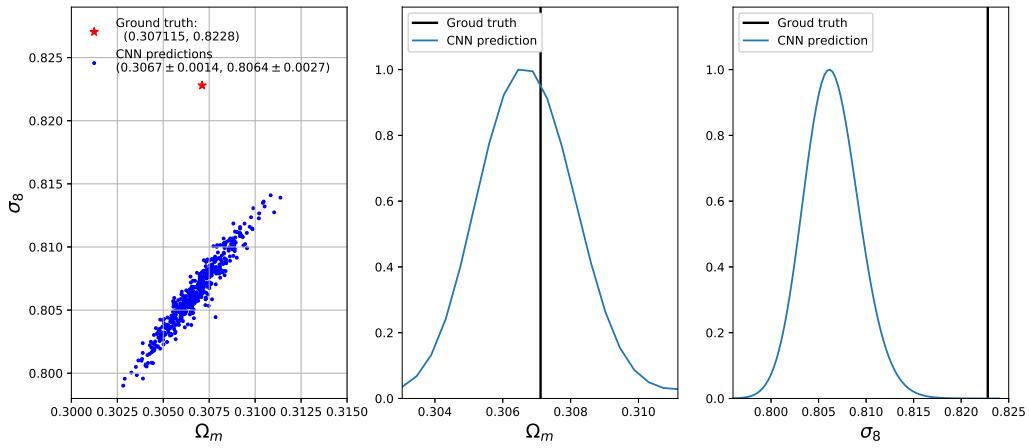
**Figure 6.** Learning curve using the default architecture (separately trained by two times), tested on the single-cosmology samples. The underlying true cosmology parameters were marked by the thick black line. The outputs of CNN as a function of epochs are plotted in the blue solid and yellow dashed curves. The statistical errors (not quite visible after epochs  $\gtrsim 200$ ) are estimated using the variance of the outputs of the 500 realizations. At the beginning of the training, the behavior was bad and random; the performance then converges and becomes stable after 200 epochs training. The two separate training yield very similar results. Very roughly, the CNN has systematical errors of 0.02 and 0.05 in the estimation of  $\Omega_m$  and  $\sigma_8$ .



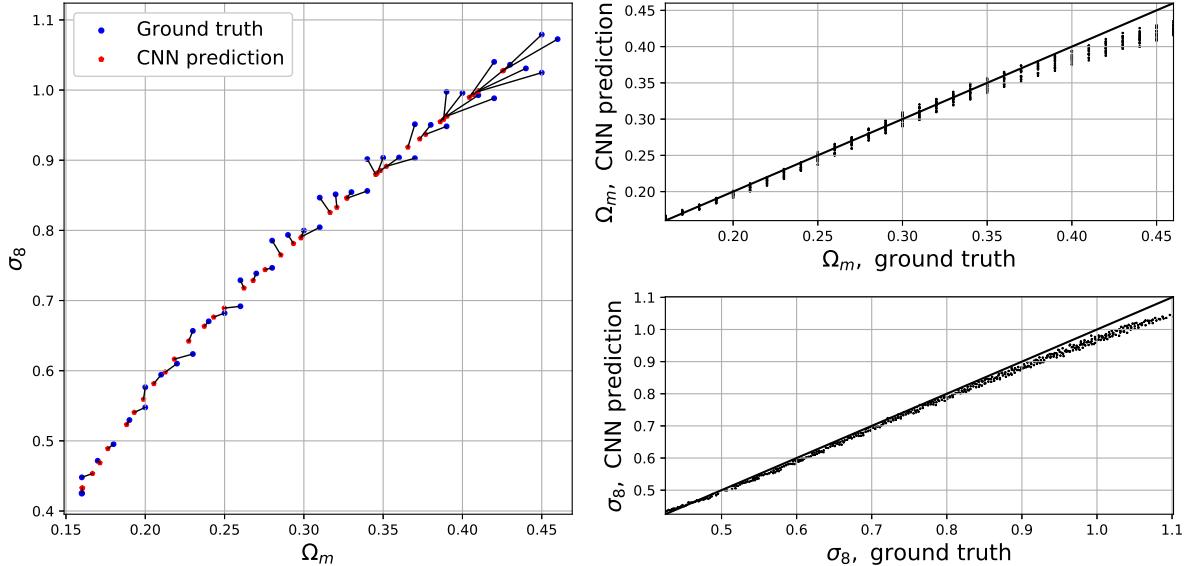
**Figure 7.** Learning curve using different settings of the CNN architecture and the optimization process. For comparison, the default settings are also plotted in the panels. *Left panel:* Using max-pooling rather than average-pooling in the pooling layer, the parameter estimation becomes better. This is different from the conclusion of (Ravanbakhsh et al. 2017), where the authors got good estimations using average pooling, but failed to get reasonable parameter estimation using max-pooling. This should be due to the difference in the CNN architectures used in their and our paper <sup>a</sup>. Also we find sgd (stochastic gradient descent) optimizer is helpful in improving the parameter estimation. Possibly, the default Adam optimizer enters local minimum and was trapped there. *Right panel:* Tests on the CNN architecture. After decreasing the number of filters in the convolution layers (half\_filters), or the number of neurons in the dense layers (half\_dense), the parameter estimation is, basically, as good as the default architecture.

<sup>a</sup>They use much less filters and more convolution layers.

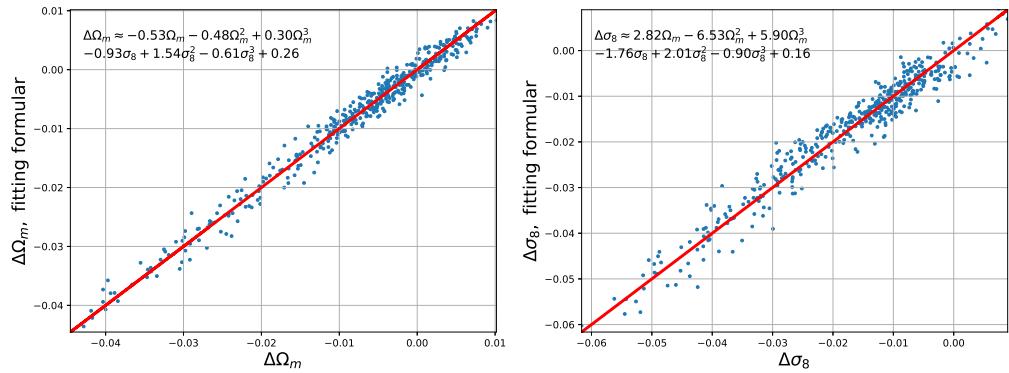
<sup>b</sup>They use much less filters and more convolution layers.



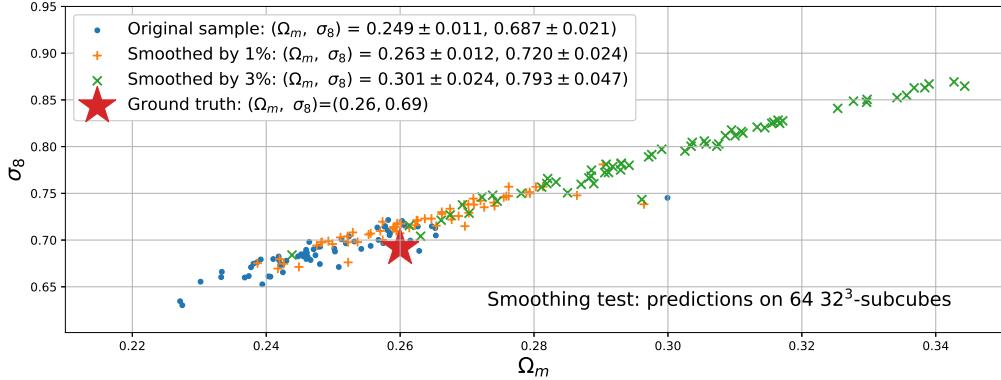
**Figure 8.** Test of a CNN architecture (sgd) on the single-cosmology samples. *Left panel:* Ground truth (red star) and CNN predictions (blue dots) of  $\Omega_m$  and  $\sigma_8$ , in the 2-d parameter space. The CNN well predicts the values of  $\Omega_m$ , but has a bias in estimating  $\sigma_8$ . *Middle and Right panels:* Likelihood distribution of  $\Omega_m$ ,  $\sigma_8$  from the CNN predictions.



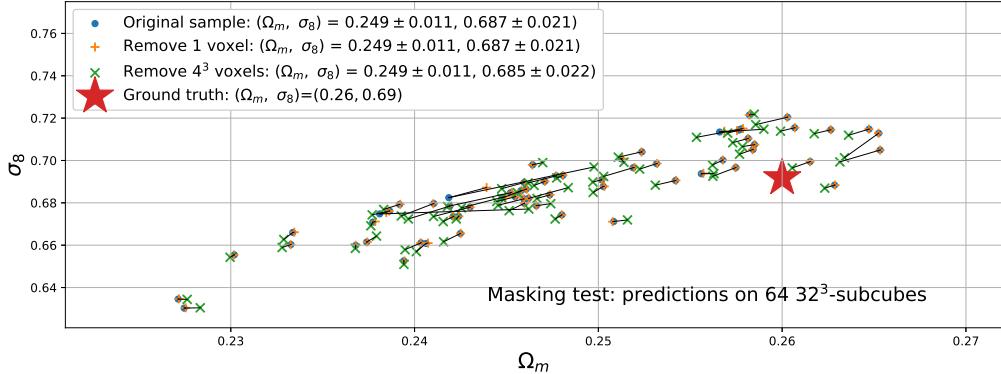
**Figure 9.** Test of a CNN architecture (sgd) on a multi-cosmology grid. There is a strong degeneracy between  $\Omega_m$  and  $\sigma_8$ . *Left panel:* Ground truth and CNN predictions of  $\Omega_m$  and  $\sigma_8$ , in the 2-d parameter space. The black lines show the difference between them. The error bar is larger at the upper-right corner of the parameter space. *Right panels:* Ground truth and CNN predictions for  $\Omega_m$  and  $\sigma_8$  panels, respectively. The CNN has slight under-estimation of the parameters at the large value tail.



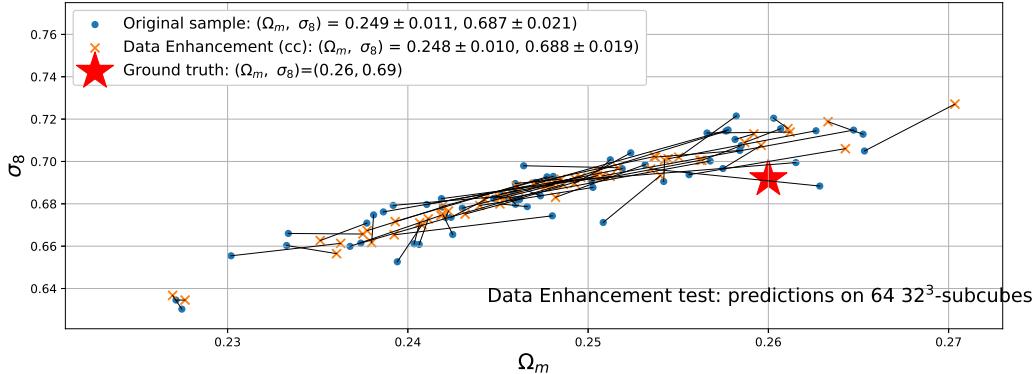
**Figure 10.** Distribution of the systematic bias in the CNN predicted  $\Omega_m$  and  $\sigma_8$  (denoted as  $\Delta\Omega_m$  and  $\Delta\sigma_8$ ). Very roughly, in the parameter space we studied, there is  $|\Delta\Omega_m| \lesssim 0.03$  and  $|\Delta\sigma_8| \lesssim 0.05$ , with mean value of  $|\Delta\Omega_m| = 0.01$  and  $|\Delta\sigma_8| = 0.018$ . In practice one can calibrate the results by subtract the systematic bias in the CNN predictions (e.g., using the fitting formula shown in the panels), making the final estimation unbiased.



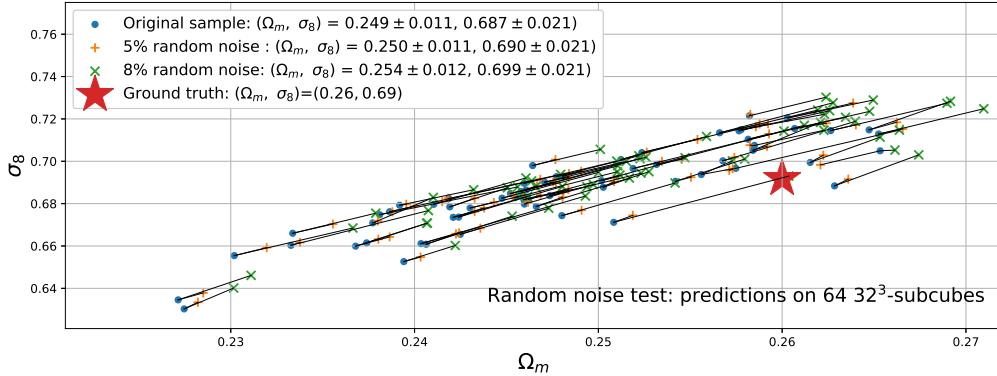
**Figure 11.** A test about error-tolerance of the CNN: sensitivity to smoothing. The input sample is a  $128^3$ -grid with cosmology  $(\Omega_m, \sigma_8) = (0.26, 0.69)$ . Predictions on the  $64 32^3$ -subgrids are plotted. The CNN can correctly predict the input cosmologies using the input sample. If we artificially smooth the grid by, say, 1% and 3% (replacing the density at each voxel point by adding a fraction of the density in 6 nearby voxels to it), we find  $2\sigma$  error in the predictions. Also, the statistical error is amplified by 2 times in the 3%-smoothing case.



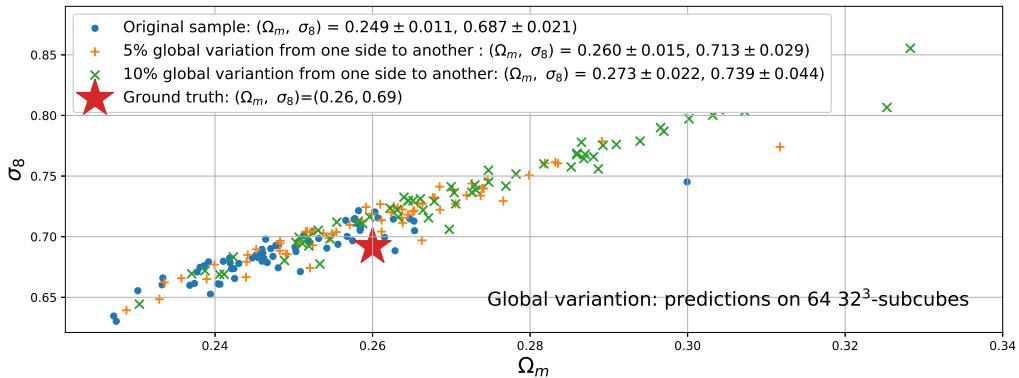
**Figure 12.** A test about error-tolerance of the CNN: sensitivity to masking. The input sample is a  $128^3$ -grid with cosmology  $(\Omega_m, \sigma_8) = (0.26, 0.69)$ . Predictions on the  $64 32^3$ -subgrids are plotted. In case we mask 1 or  $4^3$  voxels in each of the  $32^3$ -subgrid (set the values to 0), the predicted results are almost unchanged. This error-tolerance ability is very helpful when we apply the CNN to real observational data where there are many masked regions.



**Figure 13.** A test about using data enhancement (DE) in prediction. The number of test samples, after DE, is increased by as much as 48 times, yet no significant improvement in the prediction is detected.



**Figure 14.** An error-tolerance test about noise. In this test, all pixels are multiplied by a Gaussian random variable with mean value  $\mu=1$  and  $\sigma=0.05$  or  $0.1$ . In both cases we do not find significant change in the prediction (well below  $1\sigma$ ).



**Figure 15.** An error-tolerance test about global change. A global variation in the density field can create notable change in the predicted results. Cases of 5% or 10% change (linearly increase from 0% at  $x = 0$  to 5% or 10% at  $x = 256h^{-1}\text{Mpc}$ ) are plotted. In case of a 5% change, the mean values are overestimated by  $(0.024, 0.052)$ , and the scattering is increased by 100%. In case of a 10% change, the mean values are overestimated by  $(0.056, 0.11)$ , while the scattering being enlarged by 4 times.