

# Deep Learning for Cosmological Parameter Inference from Dark Matter Halo Density Field

Zhiwei Min,<sup>1,2</sup> Xu Xiao,<sup>1,2</sup> Jiacheng Ding,<sup>1,2</sup> Liang Xiao,<sup>1,2</sup> Jie Jiang,<sup>3</sup> Donglin Wu,<sup>3</sup> Qiufan Lin,<sup>4</sup> Yin Li,<sup>4</sup> Yang Wang,<sup>4</sup> Shuai Liu,<sup>5</sup> Zhixin Chen,<sup>5</sup> Xiangru Li,<sup>3</sup> Jinqu Zhang,<sup>3\*</sup> Le Zhang,<sup>1,2,4\*\*</sup> Xiao-Dong Li<sup>1,2\*†</sup>

<sup>1</sup>*School of Physics and Astronomy, Sun Yat-Sen University, Zhuhai 519082, China*

<sup>2</sup>*CSST Science Center for the Guangdong–Hong Kong–Macau Greater Bay Area, SYSU, Zhuhai 519082, China*

<sup>3</sup>*School of Computer Science, South China Normal University, Guangzhou 510631, China*

<sup>4</sup>*Peng Cheng Laboratory, No. 2, Xingke 1st Street, Shenzhen 518000, China*

<sup>5</sup>*School of Educational Science, Hunan Normal University, Changsha 410081, China*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

We propose a lightweight deep convolutional neural network (ICNN) to estimate cosmological parameters from simulated three-dimensional DM halo distributions and associated statistics. The training dataset comprises 2000 realizations of a cubic box with a side length of  $1000 h^{-1}$  Mpc, and interpolated over a cubic grid of  $300^3$  voxels, with each simulation produced using  $512^3$  DM particles and  $512^3$  neutrinos. Under the flat  $\Lambda$ CDM model, simulations vary standard six cosmological parameters including  $\Omega_m$ ,  $\Omega_b$ ,  $h$ ,  $n_s$ ,  $\sigma_8$ ,  $w$ , along with the neutrino mass sum,  $M_\nu$ . We find that: 1) within the framework of ICNN, extracting large-scale structure information is more efficient from the halo density field compared to relying on the statistical quantities including the power spectrum, the two-point correlation function, and the coefficients from wavelet scattering transform; 2) combining the halo density field with its Fourier transformed counterpart enhances predictions, while augmenting the training dataset with measured statistics further improves performance; 3) achieving high accuracy in inferring  $\Omega_m$ ,  $h$ ,  $n_s$ , and  $\sigma_8$  by the neural network model, while being inefficient in predicting  $\Omega_b$ ,  $M_\nu$  and  $w$ ; 4) compared to the simple random forest network trained with three statistical quantities, ICNN yields unbiased estimations with reduced statistical errors: approximately 33.3% for  $\Omega_m$ , 20.0% for  $h$ , 8.3% for  $n_s$ , and 40.0% for  $\sigma_8$ . Our study emphasizes this ICNN-based novel approach in extracting large-scale structure information and estimating cosmological parameters.

**Key words:** cosmological parameters – deep learning – large-scale structure of Universe

## 1 INTRODUCTION

One of the compelling challenges in modern cosmology is the precise estimation of cosmological parameters. With the continuous development of observational techniques, our understanding of the Universe is progressively deepening. However, to comprehensively and accurately understand the evolution and nature of the Universe, key parameters such as the expansion rate and dark energy density need more sophisticated measurement and analysis. This is crucial for validating cosmological models and unlocking the puzzles of the Universe such as and Hubble and  $S_8$  tensions. High-precision parameter estimates will validate or challenge existing theories, e.g., the  $\Lambda$ CDM model (Weinberg 1989; Peebles & Ratra 2003; Li et al. 2011), leading to greater progress in understanding the nature of the Universe.

The large-scale structure (LSS) of the Universe holds significant cosmological information. These vast and intricate formations depict the distribution, accumulation, and evolution of matter in the

Universe, serving as crucial observables for comprehending cosmic origins and evolution (Bardeen et al. 1986; De Lapparent et al. 1986; Huchra et al. 2012; Tegmark et al. 2004; Guzzo et al. 2014). Through the observation and analysis of LSS, we can track the evolution of the Universe, comprehend its expansion history across various redshifts, explore the formation mechanisms of galaxy clusters and superclusters, and investigate the impacts of DM and dark energy on the evolution of LSS.

At present, the two-point correlation function (2PCF) and its Fourier counterpart, the power spectrum, are the most commonly used statistical tools for analyzing LSS (Zhong et al. 2024), due to the fact that their sensitivity to both the geometry and the cosmic evolution (Kaiser 1987; Ballinger et al. 1996; Eisenstein et al. 1998; Blake & Glazebrook 2003; Seo & Eisenstein 2003), allowing for the effective extraction of information regarding Gaussian perturbations. These methods have been successfully applied in analyzing galaxy redshift surveys such as the 2dFGRS (Colless et al. 2003), 6dFGS (Beutler et al. 2011), the WiggleZ Survey (Riemer-Sørensen et al. 2012), and the SDSS Survey (York et al. 2000; Eisenstein et al. 2005; Percival et al. 2007; Anderson et al. 2014; Samushia et al. 2014; Ross et al. 2015; Beutler et al. 2017; Sánchez et al. 2017; Alam et al. 2017;

\* E-mail: zhangle7@mail.sysu.edu.cn

† E-mail: lixiaod25@mail.sysu.edu.cn

(Chuang et al. 2017; Neveux et al. 2020). However, they encounter difficulties in extracting small-scale information, e.g.,  $\lesssim 40 h^{-1}$  Mpc, from the LSS due to the pronounced influence of nonlinear structure evolution caused by gravitational collapse on such scales. Consequently, direct comparisons between observations and theories on the nonlinear scales become challenging.

Alternative statistical measures have been turned to in probing the small-scale properties of the Universe beyond 2PCF. The three-point correlation function (Sabiu et al. 2016; Slepian et al. 2017) has been utilized to improve cosmological constraints, while the more complicated four-point correlation function (Sabiu et al. 2019), has demonstrated more stringent constraints. Furthermore, more recently, Lavaux & Wandelt (2012) employed cosmic voids as a means to probe the cosmic geometry. Additionally, Li et al. (2017) explored the redshift dependence of the 2PCF along the line-of-sight as a probe for cosmological parameters. The symmetry of galaxy pairs was tested (Marinoni & Buzzi 2010) and the redshift dependence of the Alcock-Paczynski effect (AP effect) can be exploited to mitigate redshift distortions (RSDs) (Li et al. 2015). Additionally, Li et al. (2018) utilized the tomographic AP method on SDSS galaxy data to obtain a strong constraint on dark energy.

More recently, the mark weighted correlation function (MCF) (White 2016) has proposed as an alternative approach. It assigns density weights to various galaxy features to extract non-Gaussian information on LSS. Demonstrated effectiveness in capturing detailed clustering information has led to significantly enhanced constraints on cosmological parameters such as  $\Omega_m$  and  $w$  (Yang et al. 2020; Lai et al. 2023; Yin et al. 2024). Moreover, Fang et al. (2019); Yin et al. (2024) utilized the  $\beta$ -skeleton statistics to constrain cosmological parameters.

Although the methods mentioned above can extract rich information from LSS, they also exhibit certain drawbacks. Some methods are overly complex and demand substantial computational resources. Moreover, no single statistical method can completely extract all the information embedded within LSS. In recent years, the rapid development and application of machine learning have introduced new and powerful technical tools for astronomical data analysis, offering innovative solutions to the challenges encountered in survey data analysis (Way et al. 2012; Chen & Zhang 2014; Jordan & Mitchell 2015; Rodríguez-Mazahua et al. 2016; Ball et al. 2017; El Boucheffy & de Souza 2020; Sen et al. 2022). Machine learning-based data analysis methods offer significant advantages over traditional approaches in terms of efficiency, accuracy, and feature extraction capabilities. For instance, Wu et al. (2021, 2023) developed a deep learning technique to infer the non-linear velocity field from the DM density field. In addition, (Wang et al. 2023) present a deep-learning technique for reconstructing the dark-matter density field from the redshift-space distribution of dark-matter halos.

Ravanbakhsh et al. (2016); Pan et al. (2020) utilized convolutional neural networks (CNNs) to extract information from 3D DM distribution and accurately estimate cosmological parameters. Meanwhile, (Lazanu 2021) employed the Quijote simulation (Villaescusa-Navarro et al. 2020) to estimate cosmological parameters from 3D DM distribution using CNNs, comparing the constraints with those obtained from power-spectrum-based methods. Additionally, Hortua (2021) utilized Quijote simulation data to estimate cosmological parameters from a Bayesian neural network, resulting in a posterior distribution of parameters. Recently, Hwang et al. (2023) applied the Vision Transformer, known for its advantages in natural language processing, to the estimation of cosmological parameters, and compares its performance with traditional CNNs and 2PCF.

In this study, we explore a deep-learning-based approach to ex-

tract cosmological information from the halo number density field. In contrast to previous studies (Lazanu 2021; Ravanbakhsh et al. 2016; Pan et al. 2020), we utilize the halo number density field instead of the DM particle density field. Additionally, we incorporate redshift-space distortion (RSD) effects into our mock samples based on a fiducial cosmology, aiming to more realistically reflect real observations. Using the halo catalog of the Quijote's LH $\nu$ w simulation (Villaescusa-Navarro et al. 2020), our proposed ICNN framework demonstrates the ability to provide reliable constraints on cosmological parameters. Furthermore, we observe that by combining various statistics as input to the ICNN, the performance of the neural network can be noticeably enhanced.

This paper is part of the "Dark-AI" project<sup>1</sup>, a project aims to apply state-of-the-art machine learning algorithms to address frontier problems in cosmology. The structure of this paper is as follows. In Sect. 2, we introduce the samples utilized for training and testing, whereas in Sect. 3, we outline the architecture of our neural network. Sect. 4 is dedicated to presenting the results. Finally, we conclude in Sect. 5 by discussing the results.

## 2 DATA

To estimate cosmological parameters, training and test samples are constructed using the DM halo catalogues from LH $\nu$ w simulations, a subset of 2000 simulations within the Quijote simulations (Villaescusa-Navarro et al. 2020)—an ensemble of publicly available  $N$ -body simulations. These simulations utilize the TreePM code Gadget-III (Springel 2005) and are conducted in boxes with side length  $1 h^{-1}$  Gpc. The LH $\nu$ w simulations offer various cosmological results, evolving  $512^3$  DM particles together with  $512^3$  neutrino particles. For this study, we focus on the snapshot at  $z = 0.5$ . Beginning from  $z = 127$ , the simulations evolve over time, with matter power spectra and transfer functions obtained from CAMB (Lewis et al. 2000), appropriately adjusted. These quantities are used to determine displacements and peculiar velocities via second-order perturbation theory, which are then employed to assign initial particle positions on a regular grid using the 2LPT<sup>2</sup>. The simulations are executed by employing Latin-hypercube sampling, a statistical technique for generating a quasi-random sample of parameter values from a multidimensional distribution, with 7 cosmological parameters. The parameter ranges are as follows:  $\Omega_m \in [0.1, 0.5]$ ,  $\Omega_b \in [0.03, 0.07]$ ,  $h \in [0.5, 0.9]$ ,  $n_s \in [0.8, 1.2]$ ,  $\sigma_8 \in [0.6, 1.0]$ ,  $M_\nu \in [0, 1]$ , and  $w \in [-1.3, -0.7]$ .

In this study, we performed the following preprocessing steps on the halo catalogs in the Quijote LH $\nu$ w simulations to make it available as data for use by the neural network: 1) The RSD effect was incorporated along the line of sight (LoS) to more accurately reproduce real observational conditions, as expressed by:

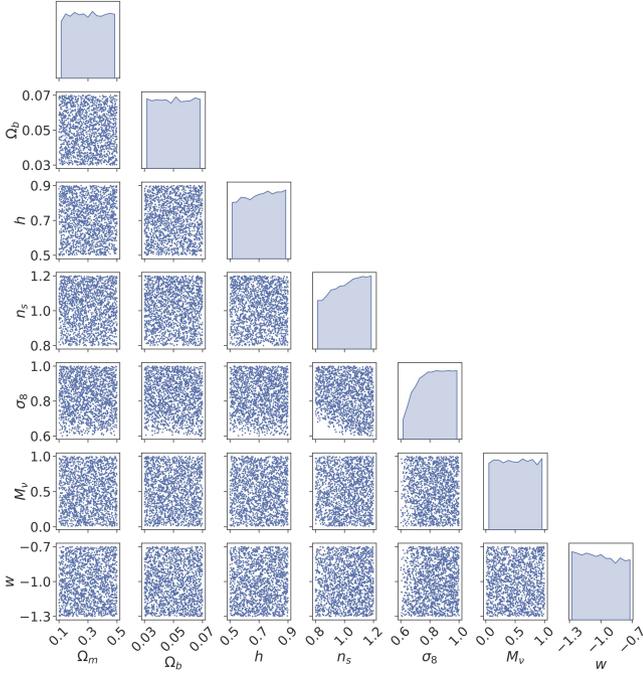
$$s = s^0 + \frac{\mathbf{v} \cdot \hat{\mathbf{z}}}{aH(a)} \quad (1)$$

where  $s^0$  is the original position of halos, while  $s$  is the new positions after the RSD effect is considered.

2) According to the different cosmological parameters of each simulation, we converted the halo positions into a fiducial cosmology background. The fiducial cosmology is derived from Planck 2018 measurements (Aghanim et al. 2020), where  $\Omega_m = 0.3071$ ,  $w =$

<sup>1</sup> <https://dark-ai.top/>

<sup>2</sup> <https://cosmo.nyu.edu/roman/2LPT/>



**Figure 1.** Distribution of cosmological parameters across the simulation boxes following the preprocessing steps on the halo catalogs in the Quijote LHVW simulations. As observed, by discarding simulations with halo number densities that are too low, the resulting distributions for each parameter exhibit slight deviations from a uniform distribution.

–1. The relation between the Quijote cosmologies and the fiducial cosmology is expressed by:

$$s_{\perp} = s_{\perp}^0 \frac{d_A^f(z)}{d_A(z)}, \quad s_{\parallel} = s_{\parallel}^0 \frac{H(z)}{H^f(z)} \quad (2)$$

where  $d_A(z)$  and  $H(z)$  represent the angular diameter distance and the Hubble parameter at redshift  $z$ , respectively. The superscript  $f$  denotes the fiducial cosmology, while  $s_{\perp}$  and  $s_{\parallel}$  represent components perpendicular and parallel to LoS, respectively.

3) After conversion from the fiducial cosmology, the box sizes are no longer the same in all three dimensions. Therefore, to conveniently feed the data cubes into the neural network, we cut the converted boxes into sides of equal length, specifically  $744 h^{-1}\text{Mpc}$ . Consequently, only the halos within such box in each simulation are considered.

4) Considering that DM halos with very low mass contribute significant noise, we implemented a cutoff for small mass halos. This cutoff was chosen appropriately such that the number of DM halos has a density equal to  $2 \times 10^{-4} h^3\text{Mpc}^{-3}$  in each box to be compatible with current spectral observations. Furthermore, if the halo number density in a simulation box is lower than that value, the box is discarded, resulting in 1710 data cubes remaining. Of these, 1500 are used for training and 210 for testing. Note that, the parameter distributions deviate from a uniform distribution due to the discarding of some simulation boxes corresponding to different cosmological models, as illustrated in Fig. 1.

5) The halo number density field is discretized into mesh cells by assigning the haloes to a  $300^3$  mesh using the Cloud-in-Cell (CIC) scheme, with a cell resolution of  $(2.48 h^{-1}\text{Mpc})^3$ .

## 2.1 Training and Test Samples

After preprocessing the simulation data, as mentioned previously, we obtained halo catalogs at the redshift of 0.5 for 1710 cosmological models. We utilized the spatial distribution information of halos together with various associated statistics as both training and test sets for the neural network. This study utilized three datasets for training and testing, as described below.

**Dataset A:** we utilized the three-dimensional distribution of the DM halo number density field,  $n(\mathbf{x})$ , which is interpolated onto a  $300^3$  mesh with a resolution of  $(2.48 h^{-1}\text{Mpc})$  along each side, to extract the input cosmological parameters. The first and second rows of Fig. 2 display the projected halo number density fields in three different cosmological models, along with their corresponding zoomed-in plots.

**Dataset B:** We utilized the Fourier-transformed halo density field with  $300^3$  grids. Letting  $\tilde{\delta}(\mathbf{k})$  denote the Fourier transform of the overdensity  $\delta(\mathbf{x})$ , defined by

$$\tilde{\delta}(\mathbf{k}) = \int \frac{d^3x}{(2\pi)^{3/2}} \delta(\mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}), \quad (3)$$

where  $\delta(\mathbf{r}) = n(\mathbf{r})/\bar{n} - 1$  is the density contrast, a dimensionless measure of overdensity at each point.

In practice, to complement  $n(\mathbf{x})$ , we retain only the low-frequency (i.e. large-scale) modes in the Fourier space field, which are not captured by the configuration space field. Specifically, we filter  $\tilde{\delta}(\mathbf{k})$  with  $|\mathbf{k}| < 0.5h\text{Mpc}^{-1}$ , resulting in a database of Fourier modes on a  $60^3$  grid. The third and fourth rows of Fig. 2 show the amplitudes of the Fourier fields for the three different cosmological models, along with their corresponding zoomed-in plots. Here, the zero-frequency mode is located at the center of each plot. Note that both amplitude and phase are input into the neural network, where each Fourier mode can be expressed as  $\tilde{\delta} = A e^{i\phi}$ , with  $A$  representing amplitude and  $\phi$  representing phase.

**Dataset C:** in addition to the density field information, we have integrated various statistics into our training samples. These statistics comprise the two-point correlation function of halos,  $\xi(r)$ , and the corresponding power spectrum,  $P(k)$ , and the wavelet scattering transform (WST) coefficients, labeled as  $S_n$ , where  $n$  denotes the order of WST coefficients.

The power spectrum is given by the following average over Fourier space:

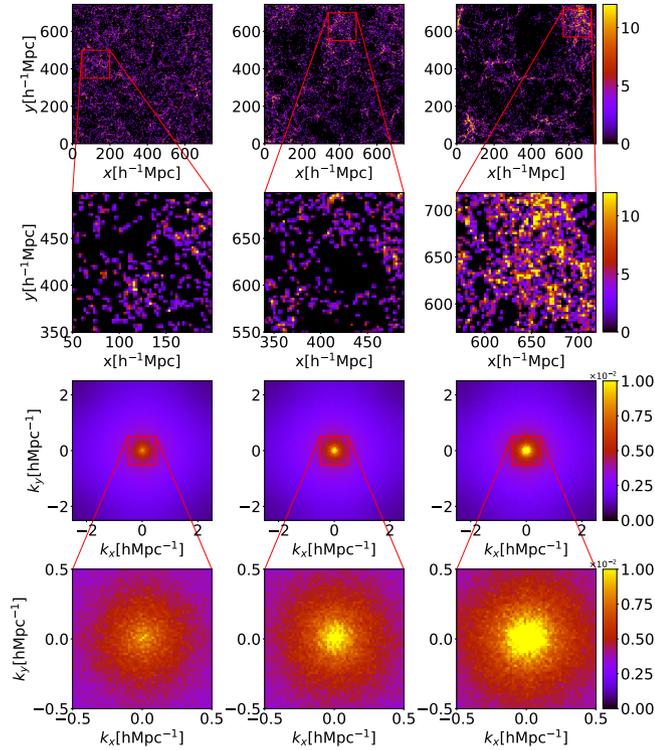
$$\langle \tilde{\delta}(\mathbf{k}) \tilde{\delta}^*(\mathbf{k}') \rangle = (2\pi)^3 P(k) \delta^3(\mathbf{k} - \mathbf{k}'). \quad (4)$$

The relationship between  $p(k)$  and  $\xi(r)$  is a Fourier transform, which can be mathematically expressed as follows,

$$\xi(r) \equiv \langle \delta(\mathbf{x}) \delta(\mathbf{x} - \mathbf{r}) \rangle = \int \frac{d^3k}{(2\pi)^3} P(k) e^{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{x}')}. \quad (5)$$

Considering the relatively large uncertainty of statistics at small scales due to noise,  $\xi(r)$  and  $P(k)$  are normalized by their mean value, namely we only utilized their shapes and focused on specific ranges:  $r \in [20, 200] h^{-1}\text{Mpc}$  for  $\xi(r)$  and  $k \in [0.05, 0.6] h\text{Mpc}^{-1}$  for  $P(k)$ . In other words, we discarded magnitude information, keeping only shape information.

The wavelet scattering transform (WST) was originally introduced in the context of signal processing in computer vision, as discussed by Bruna & Mallat (2013); Mallat (2012). This method serves the purpose of capturing the statistical properties inherent in an input field. In the WST framework, an input field  $I(\mathbf{x})$  undergoes two primary nonlinear operations: wavelet convolutions and modulus calculations. Essentially, when  $\Psi_{j_1, l_1}(\mathbf{x})$  denotes an oriented wavelet



**Figure 2.** Projected DM halo number density fields of a region of  $744 \times 744 \times 124$  ( $h^{-1}\text{Mpc}$ )<sup>3</sup> and their corresponding Fourier counterparts, selected from the training set. Three different cosmological models are presented from left to right, with the parameters as follows:  $(\Omega_m, \Omega_b, h, n_s, \sigma_8, M_\nu, w) = \{0.32, 0.045, 0.75, 0.93, 0.74, 0.07, -1\}$  (left),  $\{0.10, 0.049, 0.88, 1.08, 0.93, 0.11, -0.94\}$  (middle), and  $\{0.13, 0.054, 0.62, 0.90, 0.94, 1.00, -1.18\}$  (right). The first and second rows display the spatial distributions of the halo number density field and their zoomed-in plots, where we show the projected field with a thin slice depth of  $124 h^{-1}\text{Mpc}$ . The third and last rows correspond to the amplitudes of the corresponding Fourier modes of the density fields and their zoomed-in versions, where the depth along LoS is within  $\Delta k \in [-0.5, 0.5] h\text{Mpc}^{-1}$ .

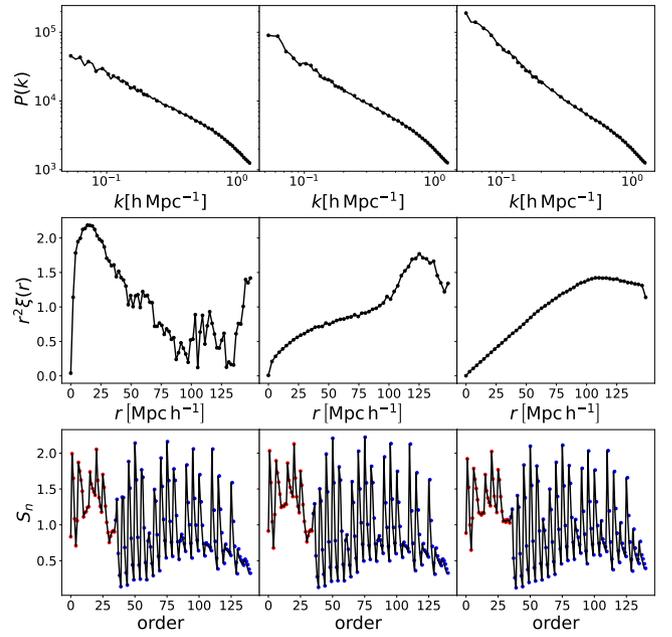
probing a scale  $j_1$  and angle  $l_1$ , the WST operation transforms  $I(\mathbf{x})$  as follows:

$$I'(\mathbf{x}) = |I(\mathbf{x}) \otimes \Psi_{j_1, l_1}(\mathbf{x})|. \quad (6)$$

Here,  $\otimes$  represents convolution. The averaging of this operation produces a WST coefficient  $S_n$ , essentially a real number describing the characteristics of the field. Through the utilization of a set of localized wavelets  $\Psi_{j_1, l_1}(\mathbf{x})$ , exploring different scales  $j_1$  and angles  $l_1$ , repeated iterations of this process generate a scattering network. The WST coefficients,  $S_n$ , up to order  $n = 2$ , are determined by the following relationships:

$$\begin{aligned} S_0 &= \langle |I(\mathbf{x})| \rangle, \\ S_1(j_1, l_1) &= \langle |I(\mathbf{x}) \otimes \Psi_{j_1, l_1}(\mathbf{x})| \rangle, \\ S_2(j_2, l_2, j_1, l_1) &= \langle |(|I(\mathbf{x}) \otimes \Psi_{j_1, l_1}(\mathbf{x})|) \otimes \Psi_{j_2, l_2}(\mathbf{x})| \rangle. \end{aligned} \quad (7)$$

Here,  $\langle \cdot \rangle$  denotes averaging over samples. Generally, a family of wavelets  $\Psi_{j_1, l_1}(\mathbf{x})$  can be generated by applying dilations and rotations to a mother wavelet. In our study, the mother wavelet is a solid harmonic multiplied by a Gaussian envelope, taking the form



**Figure 3.** Three statistics are utilized as training data for cosmological parameter inference. They are derived from the same density fields representing three distinct cosmologies, as depicted in Fig. 2. According to Eqs. 4, 5 and 6, the power spectra, 2PCFs, and WST coefficients for the halo number density fields are presented from top to bottom, respectively. In the third row, the red and blue dots represent the WST coefficients of  $S_1$  and  $S_2$ , respectively, each of which is normalized by its mean value.

of

$$\Psi_l^m(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}} e^{-|\mathbf{x}|^2/2\sigma^2} |\mathbf{x}|^l Y_l^m\left(\frac{\mathbf{x}}{|\mathbf{x}|}\right), \quad (8)$$

where,  $Y_l^m$  represents the Laplacian spherical harmonics, and  $\sigma$  denotes the Gaussian width measured in field pixels. In this study, we set  $\sigma = 0.25$ . Given a 3D input field, along with a total number of spatial dyadic scales  $J$  and total orientations  $L$ , WST coefficients can be calculated to any order. Here, the coefficient order is defined as a function of  $(j, l)$ , where  $j \in [0, 1, \dots, J-1, J]$  and  $l \in [0, 1, \dots, L-1, L]$ . Detailed information on the coefficients can be found in Valogiannis et al. (2023).

In our analysis, we set  $J = 6$  and  $L = 4$ , resulting in a total of 140 WST coefficients, excluding  $S_0$ . In summary, the WST coefficients in our work are

$$\begin{aligned} S_0 &= \langle |I(\mathbf{x})^q| \rangle, \\ S_1(j_1, l_1) &= \left\langle \left( \sum_{m=-l_1}^{m=l_1} |I(\mathbf{x}) \otimes \Psi_{j_1, l_1}^m(\mathbf{x})|^2 \right)^{q/2} \right\rangle, \\ S_2(j_2, l_2, j_1, l_1) &= \left\langle \left( \sum_{m=-l_1}^{m=l_1} |U_1(j_1, l_1)(\mathbf{x}) \otimes \Psi_{j_2, l_2}^m(\mathbf{x})|^2 \right)^{q/2} \right\rangle, \end{aligned} \quad (9)$$

with

$$U_1(j_1, l_1)(\mathbf{x}) = \sum_{m=-l_1}^{m=l_1} |I(\mathbf{x}) \otimes \Psi_{j_1, l_1}^m(\mathbf{x})|^2, \quad (10)$$

where  $q$  is a specified power governing operations on a target field. Choosing  $q > 1$  or  $q < 1$  highlights overdense or underdense regions, respectively, while  $q = 1$  represents the basic WST scenario. In our

analysis, we consider all three cases:  $q = 0.5$ ,  $q = 1$ , and  $q = 2$ . Fig. 3 presents these three statistics, derived from the same three simulation boxes as depicted in Fig. 2. For WST coefficients, only  $q = 0.5$  is displayed.

Finally, employing the principal component analysis (PCA) technique, an efficient compression scheme is utilized to retain most of the signal information encoded in the data while projecting out the noise-dominated modes. The original  $\xi(r)$  has 266 bins,  $P(k)$  has 243 bins, and  $S_1, S_2$  totally consist of 200 bins. Through PCA, each measurement statistic is compressed into a one-dimensional vector of 20 dimensions.

### 3 METHOD

To fully exploit the three-dimensional nature of the data, we employed a deep 3D convolutional network. After investigating several architectures, we propose a lightweight deep convolutional neural network (ICNN) that is efficient and present high performance in parameter estimation. Fig. 4 schematically depicts ICNN designed for determining cosmological parameters.

The network comprises three types of layers: 3D convolutions which is followed by batch normalization, max pooling layers, and fully connected layers. It begins with a  $60^3$ -voxel input layer representing the density field. When incorporating the Fourier transform of the density, two extra channels are introduced to accommodate the amplitude and phase of the Fourier modes. Thus, we represent the dimension of the input data cube as  $60^3 \times n$ , where  $n$  is for the number of channels utilized. Following this 4 convolutional layers are applied, and each is accompanied by batch normalization and a max-pooling layer with a kernel (2, 2, 2) for dimensionality reduction. The size of kernel in convolutional layers is (3, 3, 3) except the third, which is (4, 4, 4). After four 3D convolutions, the input data information are encoded by  $128 \times 2^3$  voxels, which are transited into a standard deep neural network after the flatten operation. Here, we introduce two new hidden layers, with 1324 and 128 neurons respectively, before concluding with a seven-neuron output layer. This output layer corresponds to the seven parameters that have been varied in the simulations. Additionally, when employing statistical measurements such as power spectra, 2PCF, and WST coefficients, each measurement originally has a dimension of 20. We then construct two fully connected layers to transform the dimension of each statistic to 100, which are concatenated with the output features of ICNN before passing them through the fully connected layers.

Throughout the network, rectified linear unit (ReLU) activation functions are employed. For optimization, the Adam optimizer (Kingma & Ba 2017) is utilized with a learning rate of  $5 \times 10^{-5}$ . For our machine learning task, we opted for the widely-used Mean Squared Error (MSE) loss function. This metric quantifies the average squared difference between predicted and true values, defined as

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i^{\text{pred}} - y_i^{\text{true}})^2 \quad (11)$$

In the training and testing process, we employed a density field that has been interpolated into a  $300^3$  grid using the CIC scheme, as previously mentioned. For training purposes, we divided a single  $300^3$  data cube into  $5^3$  sub-boxes, each with dimensions of  $60^3$ . When combining the density field in Fourier space, which has dimensions of  $60^3 \times 2$ , the density and its Fourier transform were concatenated into a sub-box with dimensions of  $60^3 \times n$ , where  $n = 3$ . When

utilizing only statistical measurements without incorporating density fields, parameter estimation is exclusively performed using the random forest network.

During training, for each epoch, we randomly selected a sub-box to feed into the neural network. For testing, all  $5^3$  sub-boxes are fed into the neural network, and the predictions are averaged to estimate the cosmological model. The training and testing processes are schematically depicted in Fig. 5.

### 3.1 Evaluation Metrics

Once the data is divided into training and test sets, we proceed to estimate the cosmological parameters for inputs outlined in subsequent sections. We evaluate the performance of each model through four approaches on the test set to quantify the results: 1) plotting the predicted values against the ground truth for the test set, quantified by the coefficient of determination  $R^2$ ,  $R^2$  ranges from 0 to 1, where 1 represents perfect inference.; 2) calculating the averaged bias (Bias) for each parameter; 3) calculating the relative squared error (RSE) for each parameter; 4) calculating the Root Mean Square (RMSE) for each parameter. These quantities are defined as follows:

$$R^2 = 1 - \frac{\sum_i (y_i^{\text{pred}} - y_i^{\text{true}})^2}{\sum_i (y_i^{\text{true}} - \bar{y}^{\text{true}})^2}, \quad (12)$$

$$\text{Bias} = \frac{1}{N} \sum_i (y_i^{\text{pred}} - y_i^{\text{true}}), \quad (13)$$

$$\text{RSE} = \frac{\sum_i (y_i^{\text{pred}} - y_i^{\text{true}})^2}{\sum_i (y_i^{\text{true}} - \bar{y}^{\text{true}})^2}, \quad (14)$$

and

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_i (y_i^{\text{pred}} - y_i^{\text{true}})^2}, \quad (15)$$

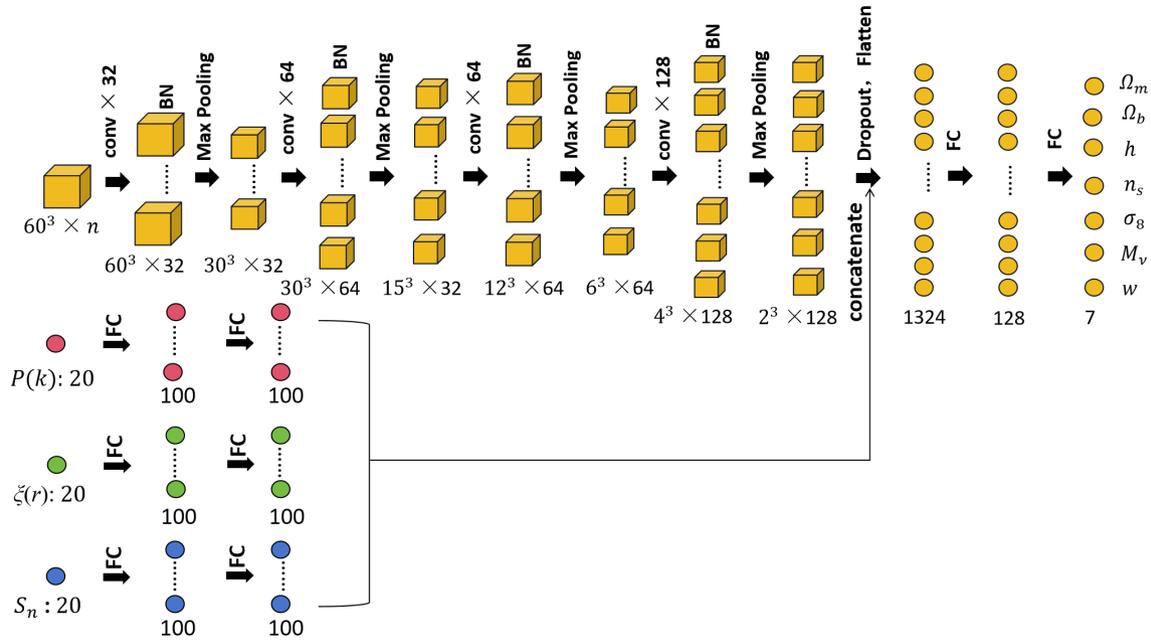
where the summation runs through the entire  $N$  test samples, with the bar indicating the average. The  $R^2$  quantifies the fraction by which the error variance is less than the true variance, while the RMSE provides an overall measure of the model's prediction accuracy, with lower values indicating better performance. Similarly, RSE measures the relative error between predicted and true values by comparing the squared difference between them. On other hand, Bias denotes the systematic error of predictions from true values. A Bias close to zero indicates that, on average, the model is making predictions that are unbiased.

## 4 RESULTS

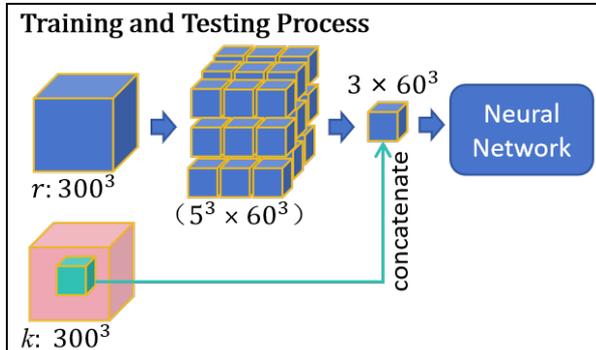
In this section, we present the results obtained from various models using different inputs, including the density field, its Fourier modes, the three statistical measurements, and their combinations. Additionally, we compare the performance of predictions made by the random forest model using statistical measurements alone.

Five distinct models were devised to evaluate the optimal choice among different datasets as inputs, including:

- (i) Model ‘‘CNN( $r$ )’’: utilizing solely the density field.
- (ii) Model ‘‘CNN( $r + k$ )’’: incorporating the density field along with its Fourier modes.



**Figure 4.** Architecture of the proposed lightweight deep convolutional neural network (ICNN) for parameter estimation. Starting with a cube of size  $60^3 \times n$ , where  $n = 1$  for the density field and  $n = 3$  for the combination of density and Fourier modes (2 for amplitude and phase), the network comprises several convolutional layers. Each convolutional layer is followed by Batch Normalisation (BN) layers to enhance convergence, and Max Pooling layers to reduce dimensionality. The dropout technique is also used for preventing the network from overfitting. Subsequently, the network is flattened and includes two fully connected layers consisting of 1024 and 128 neurons. The output layer with seven neurons corresponds to the original input parameters  $\{\Omega_m, \Omega_b, h, n_s, \sigma_8, M_v, w\}$ . For feature extraction, four convolutional layers with filters 32, 64, 64, and 128 are applied. If the statistical measurements, including the power spectrum, 2PCF, and WST coefficients, each having a dimension of 20, are utilized, we then construct two fully connected layers to transform the dimension of each statistic to 100. These layers are then concatenated with the density field network.



**Figure 5.** Training and testing processes. Data cubes are divided into  $5^3$  sub-boxes of dimensions  $60^3$ . When incorporating density fields in Fourier space, with dimensions  $60^3 \times 2$ , they are concatenated into sub-boxes of size  $60^3 \times 3$ . We combine the density fields in configuration and Fourier spaces to utilize clustering information from both small-scale and large-scale structures. During training, sub-boxes are randomly selected for each epoch, while testing involves feeding all  $5^3$  sub-boxes into the neural network, with predictions averaged to estimate the cosmological model.

- (iii) Model “CNN( $r$ )+statistics”: employing the density field together with three statistics including  $P(k)$ ,  $\xi(r)$ , and  $S_n$ .
- (iv) Model “CNN( $r+k$ )+statistics”: combining the case “CNN( $r+k$ )” with three statistics.
- (v) Model “statistics”: using only the three measured statistics.

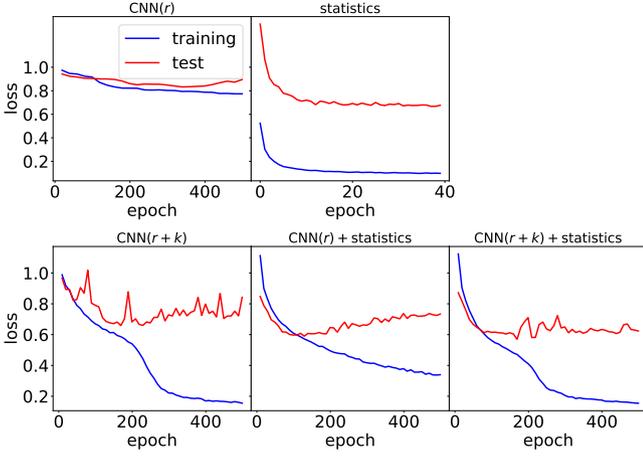
Fig. 6 displays loss curves against the number of epochs for the five different training sets. The blue and red curves represent the loss for the training and testing data sets, respectively, corresponding to 87.7% and 12.3% of the full dataset.

It can be observed that when using “CNN( $r$ )” alone, the loss function on the training set decreases slower compared to other cases, gradually converging to 0.77. However, incorporating “CNN( $k$ )” or statistics significantly reduces the loss function value, indicating that either the Fourier modes of the density field or the statistics can greatly enhance network performance. Moreover, as the dimension of each statistic is only 20, the random forest network exhibits rapid decrease and convergence at about 20 epochs.

For the case of “CNN( $r+k$ )+statistics”, the loss function drops to the lowest value of 0.57 at epoch about 200, outperforming all other cases. Importantly, comparing with the loss functions for the testing dataset, combining both fields and statistics results in the lowest loss values. Thus, as expected, this case demonstrates the best performance for parameter estimation, as we will demonstrate later. Additionally, we trained ICNN until achieving the lowest loss values for the test set.

In Fig. 7, we present the actual predictions from our designed ICNN using the entire test dataset. The corresponding  $R^2$  value for each case is also listed in each panel. For comparison, black curves are drawn to represent perfect parameter recovery ( $R^2 = 1$ ), where the correlation between the predicted and true values is 100%.

The five panels in each row correspond to the five input models for fixed cosmological parameters. Results for the seven cosmological parameters are presented from top to bottom, respectively. The prediction accuracies for  $\Omega_m$ ,  $\sigma_8$  and  $h$  from ICNN



**Figure 6.** Loss functions, as defined in Eq. 11, for five different cases for both the training and testing datasets across epochs. The blue and red curves represent training and testing losses, respectively, covering 87.7% and 12.3% of the full dataset, respectively. When using “CNN( $r$ )” alone, the training loss decreases gradually, converging to 0.77. Incorporating “CNN( $k$ )” or statistics markedly reduces the loss, indicating significant performance enhancement. The random forest network achieves rapid convergence due to the low dimensionality of each statistic. For “CNN( $r+k$ )+statistics”, the loss reaches its minimum approximately at epoch 200, outperforming all other cases. Such case also yields the lowest testing loss values, demonstrating best performance for parameter estimation.

are significantly higher compared to other parameters. Especially, the predicted  $\Omega_m$  values closely match the ground truth (black lines), with relatively small scatters. As expected, overall, the model “CNN( $r+k$ )+statistics” demonstrates the best performance for parameter predictions among other models, evident from its highest average  $R^2$  value. However, none of the five models perform well for predicting  $\Omega_b$ ,  $M_\nu$ , and  $w$ , with the highest  $R^2$  values only reaching 0.152, 0.037, and 0.196, respectively. This is because these parameters do not visibly imprint unique features in LSS in our simulation mocks and also degenerate with other parameters. For example, there is a percent-level suppression effect in the density fluctuations at small scales from  $M_\nu$ . Additionally, LSS is sensitive to the total density of  $\Omega_m$ , rather than the relatively small quantity of  $\Omega_b$ , as no baryonic feedback is considered in the cold DM simulations. Moreover, one snapshot at  $z = 0.5$  cannot effectively distinguish the different dark energy equations of state  $w$ . Since both ICNN and the random forest fail to provide effective predictions for  $\Omega_b$ ,  $M_\nu$  and  $w$ , we do not display their results in the following.

In Tab. 1, the Bias, RMSE, and RSE metrics are presented for detailed comparison across the four cosmological parameters among the five models. The results agree with the  $R^2$  values depicted in Fig. 7. The “CNN( $r+k$ )+statistics” model demonstrates the smallest RMSE and RSE values across almost all of these parameters, indicating high accuracy and small uncertainty estimation compared to other models. Notably, Bias values for all models closely match the true values within a  $2\sigma$  level compared to RMSE, highlighting the robustness of the networks and negligible systematic errors.

To emphasize the MSE values for different models, in Fig. 8, we display the relative MSE values compared to the model “statistics”. This is represented as the ratio of MSE for each model to that from the random forest network. Since our trained ICNN models are ineffective for  $\Omega_b$ ,  $M_\nu$ , and  $w$ , due to the much low  $R^2$  values, we only compare MSE values of the four parameters ( $\Omega_m$ ,  $h$ ,  $n_s$ ,  $\sigma_8$ )

relative to that of model “statistics”. From this comparison, we observe that, except for the parameter  $n_s$  with the model “CNN( $r$ )”, ICNN performs significantly better than with “statistics”. Additionally, the combination of the density field and the Fourier modes, i.e., the model “CNN( $r+k$ )”, performs better than using the density field alone, corresponding to the model “CNN( $r$ )”, except for the parameter  $h$ . Moreover, feeding the three measured statistics to ICNN further enhances the accuracy of prediction, effectively lowering the MSE values. The best performance is achieved for the model “CNN( $r+k$ )+statistics” (red line), reducing MSE by about 15–47% when compared with “statistics” model across the parameters.

To further investigate the performance of ICNN in prediction, we illustrate the joint distribution of each parameter pair and the histogram of each parameter in Fig. 9. Two models are presented for comparison: “CNN( $r+k$ )+statistics” and “statistics”. For clarity, the plots display the distributions of the errors of cosmological parameters, centered around the mean of the parameter space, i.e.

$$p_i = p_i^{\text{pred}} - p_i^{\text{true}} + \bar{p}^{\text{true}}, \quad (16)$$

where  $\bar{p}^{\text{true}}$  denotes the averaged true value over all 210 test samples with varied cosmological parameters. The two-dimensional contour plots illustrate the joint probability distribution at 68% and 95% levels, respectively, providing information about the correlation between each pair of parameters. Meanwhile, the one-dimensional distribution displays the marginalized probability of each parameter.

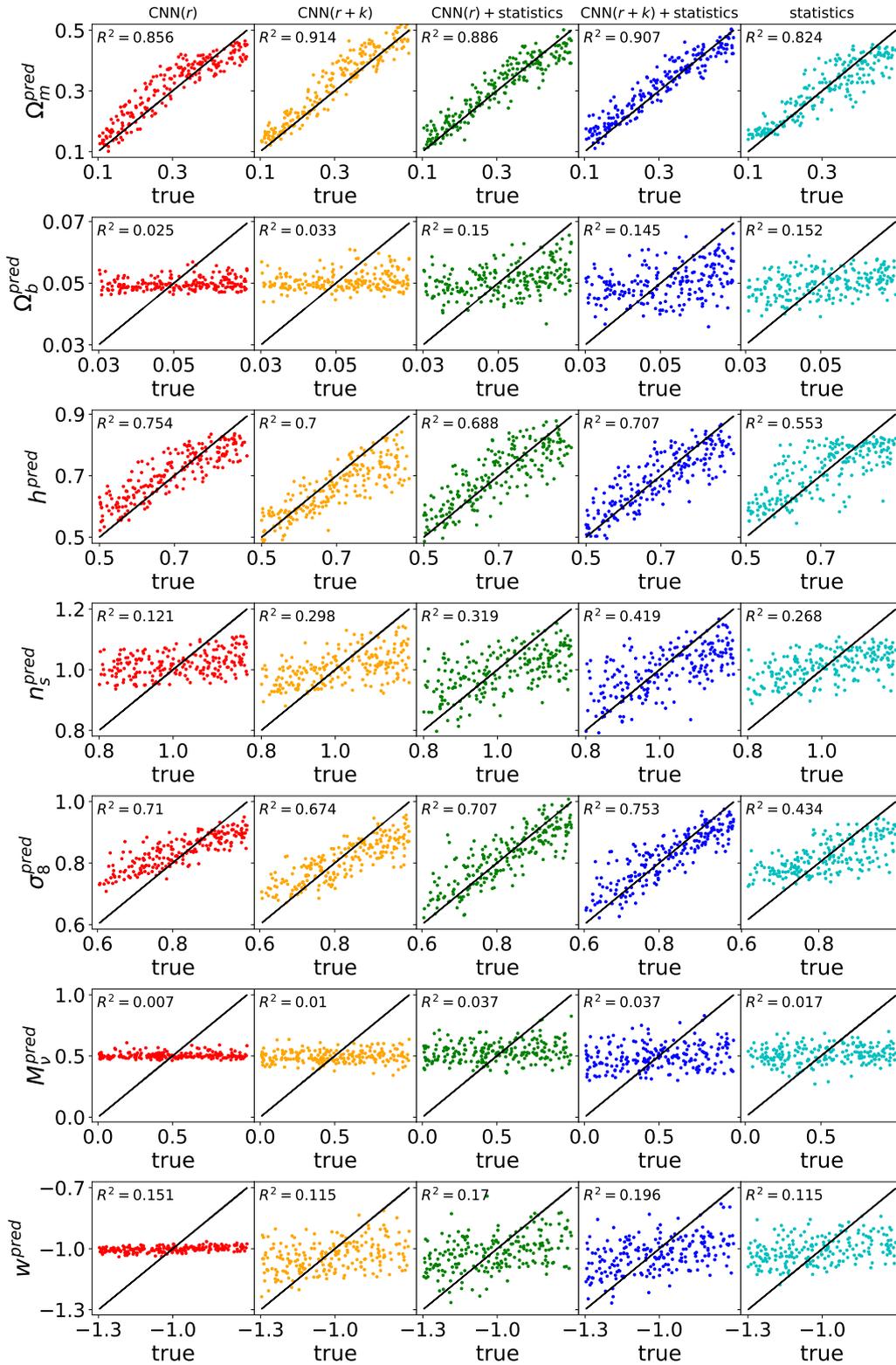
As observed, for each parameter, the model “CNN( $r+k$ )+statistics” offers a more sharply marginalized probability distribution than that for “statistics”. In other words, the former provides more accurate predictions on the cosmological parameters. This finding is further confirmed by the contour plots.

The contour area corresponds to the statistical uncertainty. As observed, the scatters of the difference between the predicted values and the true ones for all test datasets, derived from ICNN are considerably smaller than those from the random forest. Both the centers of the two-dimensional contour and the one-dimensional distribution for each parameter are close to the averaged true value (black dashed), and the deviation is significantly smaller than the statistical uncertainty, indicating an unbiased estimate. In particular, the mean and the associated standard deviation,  $\sigma_p$ , for the marginalized distribution are listed at the top of each one-dimensional plot. As observed, all the mean values closely agree with the true ones. The deviation for “CNN( $r+k$ )+statistics” from the averaged true value is 2.8% for  $\Omega_m$ , 2.4% for  $h$ , 2.5% for  $n_s$ , and 0.6% for  $\sigma_8$ , respectively. In comparison with using the model “statistics”, ICNN yields smaller  $\sigma_p$  values. Specifically, it is reduced by 33.3% for  $\Omega_m$ , 20.0% for  $h$ , 8.3% for  $n_s$  and 40.0% for  $\sigma_8$ , respectively.

We also observe that there is almost no correlation between the predicted parameters overall, although weak correlations exist between certain parameter pairs, such as  $n_s$  and  $h$ . Since the parameters in the test sample were randomly generated and uncorrelated, there should be no significant correlation between the values of the parameters predicted by ICNN. The results in the test sample meet our expectations, indicating that the estimates of the parameters by ICNN are unbiased.

## 5 CONCLUDING REMARKS

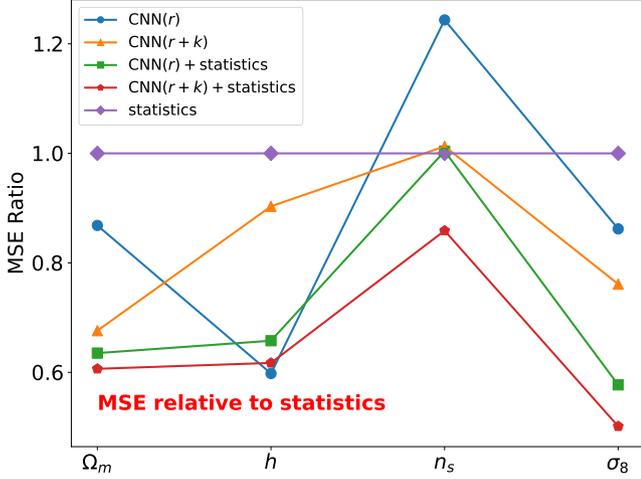
In this study, we have designed a lightweight deep convolutional neural network, ICNN, aimed at estimating cosmological parameters from simulated three-dimensional DM halo number density field and associated statistics. Our training dataset consists of 2000 realizations



**Figure 7.** Comparison between the true values of ( $\Omega_m$ ,  $\Omega_b$ ,  $h$ ,  $n_s$ ,  $\sigma_8$ ,  $M_v$ ,  $w$ ) (from top to bottom) from the test sample and their predictions from the five models (from left to right) of ICNN. The black lines indicate perfect prediction matching the truth. The resulting  $R^2$  values for each case are also shown. On average, the model “CNN( $r+k$ )+statistics” exhibits better performance for prediction than other models. ICNN performs well for predicting parameters  $\Omega_m$ ,  $h$ , and  $\sigma_8$ , but the prediction ability becomes weaker for  $\Omega_b$ ,  $M_v$ , and  $w$ . For comparison, predictions from the random forest (the fifth column) trained with the three statistics only, including the power spectrum, the 2PCF, and the WST coefficients, are shown.

Bias/RMSE/RSE \ Model	CNN( $r$ )	CNN( $r+k$ )	CNN( $r$ )+statistics	CNN( $r+k$ )+statistics	statistics
$\Omega_m$	0.013/0.046/0.161	0.022/0.041/0.284	0.005/0.040/0.881	0.014/0.039/0.440	-0.005/0.050/0.950
$h$	0.011/0.059/0.125	-0.038/0.073/0.429	0.004/0.062/0.717	-0.004/0.060/0.388	0.008/0.077/0.915
$n_s$	0.004/0.110/0.118	-0.009/0.099/0.314	-0.020/0.099/0.713	-0.020/0.091/0.293	-0.003/0.099/0.841
$\sigma_8$	0.017/0.072/0.112	-0.014/0.067/0.295	-0.0007/0.059/0.610	-0.001/0.055/0.255	-0.0003/0.078/0.817

**Table 1.** Summary of the measured evaluation metrics of Bias, RMSE, and RSE for the five models across four cosmological parameters. A small value for Bias, RMSE, or RSE indicates that the predictions made by the model are close to the true values, suggesting that the performance is relatively good.

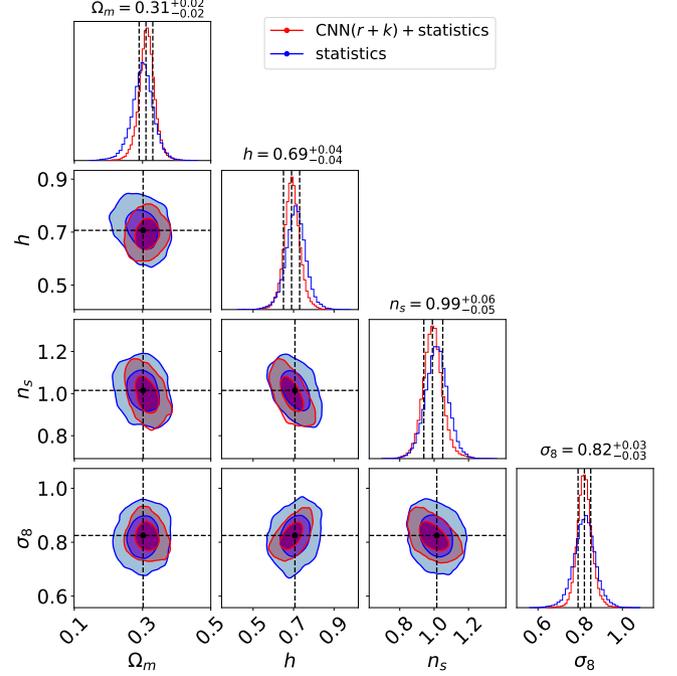


**Figure 8.** RMS values for different models relative to those of the model “statistics” using the simple random forest network across cosmological parameters, including  $\Omega_m$ ,  $h$ ,  $n_s$ , and  $\sigma_8$ . The smaller the value of MSE indicates better performance in parameter prediction. Therefore, the model “CNN( $r+k$ )+statistics” (red) exhibits the best performance among other models. As the predictions for  $\Omega_b$ ,  $M_\nu$ , and  $w$  are ineffective by ICNN, we do not show their results.

of a cubic box with a side length of  $1000 h^{-1}$  Mpc, each sampled with  $512^3$  DM particles and  $512^3$  neutrinos interpolated over a cubic grid of  $300^3$  voxels. Under the flat  $\Lambda$ CDM model, simulations vary the standard six cosmological parameters, including  $\Omega_m$ ,  $\Omega_b$ ,  $h$ ,  $n_s$ ,  $\sigma_8$ ,  $w$ , along with the neutrino mass sum,  $M_\nu$ .

Five distinct models have been considered to assess the optimal input datasets, including: “CNN( $r$ )”, which utilizes solely the density field; “CNN( $r+k$ )”, incorporating both density field and its Fourier modes; “CNN( $r$ )+statistics”, employing the density field along with three statistics (i.e., the halo density power spectrum, the 2PCF, and the WST coefficients); “CNN( $r+k$ )+statistics”, combining “CNN( $r+k$ )” with three statistics; and “statistics”, utilizing the random forest neural network trained solely with the three measured statistics, for comparison with the ICNN.

Our findings reveal several key insights: 1) within the framework of ICNN, extracting LSS information is more efficient from the halo density field compared to relying on statistical quantities including the power spectrum, 2PCF, and WST coefficients; 2) combining the halo density field with its Fourier-transformed counterpart enhances predictions, and augmenting the training dataset with measured statistics further improves performance; 3) the neural network model achieves high accuracy in inferring  $\Omega_m$ ,  $h$ ,  $n_s$ , and  $\sigma_8$ , while showing inefficiency in predicting  $\Omega_b$ ,  $M_\nu$ , and  $w$ ; 4) moreover, compared to the simple random forest network trained with three statistical quantities, our proposed ICNN model yields essentially



**Figure 9.** Probability distributions of the cosmological parameters predicted from the model “CNN( $r+k$ )+statistics” in ICNN (red) and “statistics” from the random forest (blue). The averaged value (black dashed) of each parameter over all test datasets is displayed for comparison.

unbiased estimations and provides smaller statistical errors, reducing the errors by about 33.3% for  $\Omega_m$ , 20.0% for  $h$ , 8.3% for  $n_s$ , and 40.0% for  $\sigma_8$ , respectively.

Machine learning is highly effective at analyzing complex features in complicated datasets. From this perspective, a limitation of our study is that our training samples are composed of sparse halo fields with a low number density of  $2 \times 10^{-4}$ . Consequently, many small-scale structures and clustering details are not captured in such sparse fields. A promising direction for future investigation would be to increase the number density by one to two orders of magnitude to better mimic the observational data from stage-IV surveys. In such scenarios, we expect that machine learning could significantly enhance performance and offer substantial advantages over traditional statistical methods.

In future work, we intend to evaluate the ability of the network to predict cosmological parameters from light-cone simulations, and finally, apply it to real observational data.

## ACKNOWLEDGEMENTS

We thank Francisco Villaescusa-Navarro for helpful discussions. This work is supported by National SKA Program of China (2020SKA0110401, 2020SKA0110402, 2020SKA0110100), the National Key R&D Program of China (2020YFC2201600, 2018YFA0404504, 2018YFA0404601), the National Science Foundation of China (11890691, 12203107, 12073088, 12373005), the China Manned Space Project with No. CMS-CSST-2021 (A02, A03, B01), the Guangdong Basic and Applied Basic Research Foundation (2019A1515111098), and the 111 project of the Ministry of Education No. B20019. We also wish to acknowledge the Beijing Super Cloud Center (BSCC) and Beijing Beilong Super Cloud Computing Co., Ltd (<http://www.blsc.cn/>) for providing HPC resources that have significantly contributed to the research results presented in this paper.

## REFERENCES

- Aghanim N., et al., 2020, *Astronomy & Astrophysics*, 641, A6
- Alam S., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 470, 2617
- Anderson L., et al., 2014, *Monthly Notices of the Royal Astronomical Society*, 441, 24
- Ball J. E., Anderson D. T., Chan C. S., 2017, *Journal of applied remote sensing*, 11, 042609
- Ballinger W., Peacock J., Heavens A., 1996, *Monthly Notices of the Royal Astronomical Society*, 282, 877
- Bardeen J. M., Bond J., Kaiser N., Szalay A., 1986, *Astrophysical Journal*, Part 1 (ISSN 0004-637X), vol. 304, May 1, 1986, p. 15-61. SERC-supported research., 304, 15
- Beutler F., et al., 2011, *Monthly Notices of the Royal Astronomical Society*, 416, 3017–3032
- Beutler F., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 466, 2242
- Blake C., Glazebrook K., 2003, *The Astrophysical Journal*, 594, 665
- Bruna J., Mallat S., 2013, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 1872
- Chen C. P., Zhang C.-Y., 2014, *Information sciences*, 275, 314
- Chuang C.-H., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 471, 2370
- Colless M., et al., 2003, *arXiv e-prints*, pp astro-ph/0306581
- De Lapparent V., Geller M. J., Huchra J. P., 1986, *Astrophysical Journal*, Part 2-Letters to the Editor (ISSN 0004-637X), vol. 302, March 1, 1986, p. L1-L5. Research supported by the Smithsonian Institution., 302, L1
- Eisenstein D. J., Hu W., Tegmark M., 1998, *The Astrophysical Journal*, 504, L57
- Eisenstein D. J., et al., 2005, *The Astrophysical Journal*, 633, 560
- El Boucheffry K., de Souza R. S., 2020, in , *Knowledge discovery in big data from astronomy and earth observation*. Elsevier, pp 225–249
- Fang F., Forero-Romero J., Rossi G., Li X.-D., Feng L.-L., 2019, *Monthly Notices of the Royal Astronomical Society*, 485, 5276
- Guzzo L., et al., 2014, *Astronomy & Astrophysics*, 566, A108
- Hortua H. J., 2021, *arXiv preprint arXiv:2112.11865*
- Huchra J. P., et al., 2012, *The Astrophysical Journal Supplement Series*, 199, 26
- Hwang S. Y., Sabiu C. G., Park I., Hong S. E., 2023, *Journal of Cosmology and Astroparticle Physics*, 2023, 075
- Jordan M. I., Mitchell T. M., 2015, *Science*, 349, 255
- Kaiser N., 1987, *Monthly Notices of the Royal Astronomical Society*, 227, 1
- Kingma D. P., Ba J., 2017, *Adam: A Method for Stochastic Optimization* ([arXiv:1412.6980](https://arxiv.org/abs/1412.6980))
- Lai L. M., et al., 2023, *Improving Constraint on  $\Omega_m$  from SDSS Using Marked Correlation Functions* ([arXiv:2312.03244](https://arxiv.org/abs/2312.03244))
- Lavaux G., Wandelt B. D., 2012, *The Astrophysical Journal*, 754, 109
- Lazanu A., 2021, *Journal of Cosmology and Astroparticle Physics*, 2021, 039
- Lewis A., Challinor A., Lasenby A., 2000, *The Astrophysical Journal*, 538, 473–476
- Li M., Li X.-D., Wang S., Wang Y., 2011, *Communications in theoretical physics*, 56, 525
- Li X.-D., Park C., Sabiu C. G., Kim J., 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 807
- Li X.-D., Park C., Sabiu C. G., Park H., Cheng C., Kim J., Hong S. E., 2017, *The Astrophysical Journal*, 844, 91
- Li Y., Schmittfull M., Seljak U., 2018, *Journal of Cosmology and Astroparticle Physics*, 2018, 022
- Mallat S., 2012, *Group Invariant Scattering* ([arXiv:1101.2286](https://arxiv.org/abs/1101.2286))
- Marinoni C., Buzzi A., 2010, *Nature*, 468, 539
- Neveux R., et al., 2020, *Monthly Notices of the Royal Astronomical Society*, 499, 210–229
- Pan S., Liu M., Forero-Romero J., Sabiu C. G., Li Z., Miao H., Li X.-D., 2020, *Science China Physics, Mechanics & Astronomy*, 63, 110412
- Peebles P. J. E., Ratra B., 2003, *Reviews of modern physics*, 75, 559
- Percival W. J., Cole S., Eisenstein D. J., Nichol R. C., Peacock J. A., Pope A. C., Szalay A. S., 2007, *Monthly Notices of the Royal Astronomical Society*, 381, 1053
- Ravanbakhsh S., Oliva J., Fromenteau S., Price L., Ho S., Schneider J., Póczos B., 2016, in *International conference on machine learning*. pp 2407–2416
- Riemer-Sørensen S., et al., 2012, *Physical Review D*, 85
- Rodríguez-Mazahua L., Rodríguez-Enríquez C.-A., Sánchez-Cervantes J. L., Cervantes J., García-Alcaraz J. L., Alor-Hernández G., 2016, *The Journal of Supercomputing*, 72, 3073
- Ross A. J., Samushia L., Howlett C., Percival W. J., Burden A., Manera M., 2015, *Monthly Notices of the Royal Astronomical Society*, 449, 835
- Sabiu C. G., Mota D. F., Llinares C., Park C., 2016, *Astronomy & Astrophysics*, 592, A38
- Sabiu C. G., Hoyle B., Kim J., Li X.-D., 2019, *The Astrophysical Journal Supplement Series*, 242, 29
- Samushia L., et al., 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 3504
- Sánchez A. G., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 464, 1640
- Sen S., Agarwal S., Chakraborty P., Singh K. P., 2022, *Experimental Astronomy*, 53, 1
- Seo H.-J., Eisenstein D. J., 2003, *The Astrophysical Journal*, 598, 720
- Slepian Z., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 468, 1070
- Springel V., 2005, *Monthly Notices of the Royal Astronomical Society*, 364, 1105–1134
- Tegmark M., et al., 2004, *The Astrophysical Journal*, 606, 702
- Valogiannis G., Yuan S., Dvorkin C., 2023, *arXiv preprint arXiv:2310.16116*
- Villaescusa-Navarro F., et al., 2020, *The Astrophysical Journal Supplement Series*, 250, 2
- Wang Z., Shi F., Yang X., Li Q., Liu Y., Li X., 2023, *Science China Physics, Mechanics & Astronomy*, 67
- Way M. J., Scargle J. D., Ali K. M., Srivastava A. N., 2012, *Advances in machine learning and data mining for astronomy*. CRC Press Boca Raton
- Weinberg S., 1989, *Reviews of modern physics*, 61, 1
- White M., 2016, *Journal of Cosmology and Astroparticle Physics*, 2016, 057–057
- Wu Z., et al., 2021, *The Astrophysical Journal*, 913, 2
- Wu Z., Xiao L., Xiao X., Wang J., Kang X., Wang Y., Wang X., Li X.-D., 2023, *Monthly Notices of the Royal Astronomical Society*, 522, 4748
- Yang Y., et al., 2020, *The Astrophysical Journal*, 900, 6
- Yin F., et al., 2024, *arXiv preprint arXiv:2403.14165*
- York D. G., et al., 2000, *The Astronomical Journal*, 120, 1579
- Zhong K., Gatti M., Jain B., 2024, *Improving Convolutional Neural Networks for Cosmological Fields with Random Permutation* ([arXiv:2403.01368](https://arxiv.org/abs/2403.01368))

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.