Carnegie Mellon University
MELLON COLLEGE OF SCIENCE

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN THE FIELD OF PHYSICS

TITLE: "Innovative Techniques to Address the Non-linearity of the Large Scale Structure of the Universe."

PRESENTED BY: Siyu He

ACCEPTED BY THE DEPARTMENT OF PHYSICS

Shirley Ho	6/12/19
SHIRLEY HO, CHAIR PROFESSOR	DATE

Scott Dodelson	6/24/19
SCOTT DODELSON, DEPT HEAD	DATE

APPROVED BY THE COLLEGE COUNCIL

Rebecca Doerge	6/26/19
REBECCA DOERGE, DEAN	DATE

New Techniques to Address the Non-linearity of the Large Scale Structure of the Universe

Siyu He

Department of Physics

Carnegie Mellon University

Supervisor

Shirley Ho

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

May 28, 2019

Abstract

To fully understand the non-linear evolution of the large scale structure of the Universe and to extract useful information from the large scale structure are key subjects of modern cosmology. In this dissertation, I am going to address the non-linearity of the Universe from two new perspectives.

One way to study the non-linearity is to study filaments, which evolve non-linearly from the initial density fluctuations produced in the primordial Universe. In the first part of the dissertation, I am going to report the first detection of CMB (Cosmic Microwave Background) lensing by filaments. We propose a phenomenological model to interpret the detected signal, and we measure how filaments trace the matter distribution on large scales through filament bias.

In the second part of the dissertation, I will present the deep learning method as a practical and accurate alternative to learning the gravitational structure formation of the Universe. We build two deep neural networks, the D3M model and the multi-scale deep sets model, to predict the non-linear structure formation of the Universe. Our extensive results have shown our models outperform the second-order perturbation theory. I will also discuss our efforts in understanding the robustness of the trained deep learning model.

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Shirley Ho (Principle Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Rupert A. C. Croft

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Rachel Mandelbaum

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Martin White

Approved for the Carnegie Mellon University

Department of Physics:

Head of the Physics Department

(Principle Advisor)

To mom and dad

for everything ...

In memory of my grandfather

I miss you and love you every day ...

Acknowledgements

First of all, I would like to express my gratitude to my advisor, Professor Shirley Ho. Without her support, I couldn't become who I am today. She is very knowledgeable, intelligent, encouraging, and friendly. She is always providing insightful directions in my research. She has brought me from Carnegie Mellon University to the University of California, Berkeley, and then Flatiron Institute and sent me to different conferences around the world. She is always providing me with bigger stages full of valuable opportunities and resources. She is also very friendly and understanding. She will listen when I am down and provide me with caring suggestions.

I would also like to thank my committee members, Professor Rupert Croft, Professor Rachel Mandelbaum, and Professor Martin White, for their interest in my work, and many useful comments and suggestions.

I would like to thank Wei Chen. I couldn't imagine how I could survive the gloomy days without his supports. When I am down, he always listens to my complaints patiently, gives me suggestions, and tries every way to make me happy. When I am unconfident, he keeps saying to me how good I am. When I am making slight success, he shares my happiness and is truly happy for me. He helps and discusses a lot with me about my research projects. He kindly

becomes my "rubber duck" when I have problems debugging. He is the one I could share and talk about everything.

I would like to thank Professor Manfred Paulini, for all the supports while I am in CMU, and for helping me out about all the bureaucratic issues while I am away from CMU. Without his help, my stay in Berkeley and Flatiron Institute would never be this smooth.

I would like to thank my fellow graduate students at Carnegie Mellon University, Hung-Jin Huang, Shadab Alam, Even Tucker, Aklant Bhowmick, Bojun Feng, Sébastien fromenteau, Carolina Nunez, Franǫois Lanusse, Mao-Sheng Liu, Hongyu Zhu, Zhonghao Luo, Zongge Liu, Hsiu-Hsien Lin, Sukhdeep Singe, for making my life in CMU so happy and colorful, for giving me a sense of belonging in CMU, and for all the supports whenever I need their help.

I would like to thank Elena Massara and Elena Giusarma. I am grateful that we are together in the group for a long time. They are like my little sisters, giving me supports in both research and life.

I would like to thank Yin Li, Yu Feng, Ferraro Simone, for being my collaborators, for answering a significant number of questions patiently and for for making my life in Berkeley not lonely.

I would like to thank my fellow graduate students, postdocs, friends at the University of California, Berkeley, Alex Krolewski, Ellie Kitanidis, Chirag Modi, Ben Horowitz, ChangHoon Hahn, Emmanuel Schaan, Enea Di Dio, Vanessa Boehm, Siyao Jia, Teng Lei for providing the support I needed at critical times and for being such great and fun colleagues and officemates.

I would like to thank Daisy Leung, Avery Kim for sharing the office

with me in Flatiron Institute, for sharing both of my happiness and sadness. Both of them are very sympathetic, and I am grateful for that.

I would like to thank Gabriella Contardo, Laurence Levasseur, Francisco Antonio Villaescusa Navarro and Yashar Hezaveh, for being my collaborators and for all suggestions they make for my research. I enjoy a lot discussing with them about all the machine learning problems.

I would like to thank Yuanyuan Xiao, Kaibing Zhang, Shuangli Li, Tianyu Qin, Jingyao Wu for being my friends and bringing much joy to my everyday life.

I would like to take this opportunity to thank my parents, my grandparents, my uncles, my aunts, and my cousin, for giving me such a happy family and for their warm love and constant encouragement. Without their support, I couldn't have gone this far in my career. They are happy for however small achievements I make and proud of me about whatever I do.

Contents

1 Introduction **1**

1.1 Dynamics of structure formation 3

1.1.1 Analytical Theories 3

1.1.2 N-body dark matter only Simulations 6

1.1.3 Link between Mass Distribution and Galaxy Distribution 8

1.2 Measurements of the clustering of Large Scale Structure 9

1.2.1 Correlation Functions 9

1.2.2 Higher Order Statistics 11

1.2.2.1 Filaments 11

1.3 Deep Learning Methods 12

1.4 Organization of the Dissertation 14

2 Learning to Predict the Cosmological Structure Formation **20**

2.1 Abstract 20

2.2 Introduction 21

2.3 Setup 23

2.3.1 Displacement versus Density Field 24

2.4 Results and Analysis 26

2.4.1 Point-Wise Comparison 26

2.4.2 2-Point Correlation Comparison 27

2.4.3 3-Point Correlation Comparison 30
2.5 Generalizing to New Cosmological Parameters 33
2.5.1 Varying Primordial Amplitude of Scalar Perturbations $\mathbf{A_s}$ 33
2.5.2 Varying matter density parameter $\mathbf{\Omega_m}$ 34
2.6 Conclusions 36
2.7 Materials and Methods 37
2.7.1 Dataset 37
2.7.2 Model and Training 37
2.7.2.1 Details of the D^{3}M Architecture 38
2.7.2.2 Padding and Periodic Boundary 39
2.7.2.3 Loss Function 39
2.8 appendix 40
2.9 Comparison with One-point remapping method 43
2.10 Test Time Comparison 43

3 Beyond the D^{3}M Model **45**
3.1 Interrogation of the D^{3}M model with physics analysis 45
3.1.1 Zero Residual Test 46
3.1.2 Pancake Test 47
3.1.3 Two-Mode Coupling Test 48
3.1.4 Saliency Map 49
3.1.5 Conclusion 50
3.2 Multi-Scale Deep Sets to Learn Non-Linear Evolution of the Universe 55
3.2.1 Related Work 55
3.2.2 Data 56
3.2.3 Method 57
3.2.4 Result 58
3.2.5 Conclusion 59

4 The detection of the imprint of filaments on Cosmic Microwave Background (CMB) lensing **60**

4.1 Introduction 60

4.2 Main Result 61

4.3 Method 71

4.3.1 Filament finder 71

4.3.2 Filament length measurement. 75

4.3.3 Estimator. 75

4.3.4 Sky mock for filaments and dark matter. 76

4.4 Supplementary information 79

4.4.1 Jackknife regions 79

4.4.2 Error calculation for $C_l^{\kappa f}$ from simulation 79

5 Conclusions **85**

References **116**

List of Figures

1.1 This figure comes from [1]. The top and left slices illustrate the CfA2 'Great Wall' and Sloan Great Wall, respectively. The right and bottom slices of fig. 1.1 illustrate the mock galaxy catalog from Millennium simulation. The large scale structure could be clearly in both the galaxy surveys and the simulations. 2

2.1 (left) The displacement vector-field and (right) the resulting density field produced by D^3M. The vectors in the left figure are uniformly scaled down for better visualization. 22

2.2 The columns show 2-D slices of full particle distribution (top) and displacement vector (bottom) by various models, from left to right: **(a)** FastPM: the target ground truth, a recent approximate N-body simulation scheme that is based on a particle-mesh (PM) solver ; **(b)** Zel'dovich approximation (ZA): a simple linear model that evolves particle along the initial velocity vector; **(c)** Second order Lagrangian perturbation theory (2LPT): a commonly used analytical approximatation; **(d)** Deep learning model (D^3M) as presented in this work. While FastPM (a) served as our ground truth, the columns to its right includes color for the points or vectors. The color indicates the relative difference $(q_{model} - q_{target})/q_{target}$ between the target (a) location or displacement vector and predicted distributions by various methods (b-d). The error-bar shows denser regions have a higher error for all methods. which suggests that it is harder to predict highly non-linear region correctly for all models: D^3M, 2LPT and ZA. Our model D^3M has smallest differences between predictions and ground truth among the above models (b)-(d). 25

2.3 (**a**) From top to bottom: (top) displacement and density power-spectrum of FastPM (orange), 2LPT (blue), and D^3M (green); (middle) transfer function – i.e., the square root of the ratio of the predicted power-spectrum to the ground truth; (bottom) 1-r^2 where r is the correlation coefficient between the predicted fields and the true fields. Results are the averaged values of 1000 test simulations. The transfer function and correlation coefficient of the D^3M predictions is nearly perfect from large to intermediate scales and outperforms our benchmark 2LPT significantly. (**b**) The ratios of the multipole coefficients ($\zeta_l(r_1, r_2)$) (to the target) of the two 3-point correlation functions for several triangle configurations. The results are averaged over 10 test simulations. The error-bars (padded regions) are the standard deviations derived from 10 test simulations. The ratio shows the 3-point correlation function of D^3M is closer than 2LPT to our target FastPM with lower variance. 28

2.4 We show the differences of particle distributions and displacement fields when we change the cosmological parameters A_s and Ω_m. (**a**) The errorbar shows the difference of particle distribution (upper panel) and displacement fields (lower panel) between $A_s = A_0$ and the two extremes for $A_s = .2A_0$(left) and $A_s = 1.8A_0$ (right). (**b**) A similar comparison showing the difference of the particle distributions (upper panel) and displacement fields (lower panel) for smaller and larger values of $\Omega_m \in \{.1, .5\}$ with regard to $\Omega_m = 0.3089$, which was used for training. While the difference for smaller value of A_s (Ω_m) is larger, the displacement for larger A_s (Ω_m) is more non-linear. This non-linearity is due to concentration of mass and makes the prediction more difficult. 31

2.5 Similar plots as in Figure. 3(a), except we test the 2 point statistics when we vary the cosmological parameters without changing the training set (which has differnet cosmological parameters) nor the trained model. We show predictions from D^{3}M and 2LPT when tested on different (**a**) A_s and, (**b**) Ω_m. We show (top) the transfer function – i.e., the square root of the ratio of the predicted power-spectrum to the ground truth and (bottom) 1-r^2 where r is the correlation coefficient between the predicted fields and the true fields. D^{3}M prediction outperforms 2LPT prediction at all scales except in the largest scales as the perturbation theory works well in linear regime (large scales). 36

2.6 Same deep learning architecture with either density field or displacement field as input. The loss with displacement field as input is smaller, and this proves that displacement field contains information that is non-local in the density, which helps model to learn. 41

2.7 Each orange box corresponds to a multi-channel feature map. The number of channels and the output size of each channel are denoted on top of the box. Grey dashed boxes represent copied feature maps. The arrows denote the different operations. The black, red and blue arrays represent 3×3 convolution with stride one, 3×3 convolution with stride two, 3×3 transposed convolution with stride $1/2$ respectively with periodic padding followed by ReLU activation and batch normalization. The orange array represents 1×1 convolution with stride one. 42

2.8 From top to bottom: (top) density power-spectrum of FastPM (orange), 2LPT (blue), D^3M (green) and ZARM/2LPTRM (red); (middle) transfer function – i.e., the square root of the ratio of the predicted power-spectrum and the ground truth; (bottom) 1-r^2 where r is the correlation coefficient between the predicted fields and the true fields. Results are the average values of 1000 test simulations. It shows D^3M result outperforms the remapping method in both transfer function and correlation coefficient. 44

3.1 Residual Test. We feed the trained D^3M model with all zeros in the input displacement fields. Periodic patterns show up in the D^3M model prediction due to the $3\times3\times3$ filters. There is also a small bias in the x,y,z directions which might come from the noise of the D^3M model. But the D^3M model prediction is close to 0 and preserves homogeneity and isotropy with great significance as expected. 47

3.2 Pancake test. We decompose the input into single modes in Fourier space and feed each mode (planar perturbation as shown in eq. 3.1) to the trained D^3M model. The red triangles show the power spectrum for the input mode while the green crosses and orange stars show the power spectrum for FastPM and D^3M model predictions. The transfer function shows that the D^3M model captures quite well at the dominant scale of FastPM, which indicates the D^3M model is able to capture scale information. The D^3M model also captures the other modes of FastPM that are two orders smaller than the dominant mode and come from the numerical artifact of FastPM simulations. 52

3.3 Two-mode coupling check. We compare between FastPM, D^3M, and theory with 2 perpendicular fundamental modes as input. The agreement among the D^3M prediction, FastPM prediction and Lagrangian theory is pretty good on the first two modes. The remaining noisy modes are the artifacts of the FastPM simulations. And the D^3M predictions are also consistent with FastPM on these noisy modes. 53

3.4 2D slices of Saliency Map. The first row shows the input ZA which is fed into the D^3M model, the second row shows the D^3M prediction, and the third row shows the saliency map. The blue arrow points to where in the output map that the saliency map is calculated. The saliency map indicates the most crucial information to correct ZA for each particle comes from local regions. This is expected because ZA is correct on large scales, and most of the non-linear information should mostly come from small scale local regions. And the first plot also shows that the D^3M model correctly captures the periodic boundary condition of the simulations. 54

3.5 From top to bottom: (top) displacement and density power-spectrum of FastPM (orange), 2LPT (blue), and multi-Scale deep sets (green); (middle) transfer function – i.e., the square root of the ratio of the predicted power-spectrum to the ground truth; (bottom) 1-r^2 where r is the correlation coefficient between the predicted fields and the true fields. Results are the averaged values of 1000 test simulations. The results have shown that the multi-scale deep sets model outperforms the 2LPT. 59

4.1 The filament length as a function of redshift. The orange (red) crosses are the mean (median) of the filament length in each redshift bin, where the error bars come from the standard error of the mean (median). The large difference in the mean and the median values implies the filament length distribution is not Gaussian. The background mesh plot shows the 2d histogram of the number of filament length as a function of the redshift and the filament length. 69

4.2 Cross angular power spectrum. (***a***) shows the cross angular power spectrum of the filaments and the CMB convergence field. The blue crosses are measured with error bars from jackknife resampling of the sky into 77 equally weighted regions. The red and orange dashed lines are theoretical predictions based on different smoothing models (red: filament length and spacing smoothing, orange: filament length and statistical fit for perpendicular smoothing). The corresponding filament bias for the two models are 1.68 and 1.47. The green circles are from simulations. (***b***) a null test showing the cross angular power spectrum of the filament catalogue and the rotated CMB lensing convergence map. The cross signals fluctuate around 0. The $\chi^2_{NULL}/d.o.f.$ for the three scenarios are all ~ 1. 70

4.3 (**a**) The correlation coefficient of galaxy map and filament map and **b** the correlation of $C_l^{\kappa f}$ and $C_l^{\kappa g}$. Both plots show the filament and galaxy maps are not totally correlated with large deviations on small scales. 71

4.4 A demonstration of filament intensity overdensity and corresponding dark matter particle overdensity in simulation at redshift 0.57. (***a***) filament intensity overdensity at redshift 0.57. (***b***) dark matter overdensity at redshift 0.57. The color bars show the amplitude of the overdensity field in linear scale. 80

4.5 Visualization of jackknife regions. The jackknife regions are chosen so that each region has same effective observed area and is close to square. 81

4.6 Filament intensity distribution and galaxy redshift distribution as a function of redshift. The blue curve shows the filament intensity distribution as a function of redshift. The green curve shows the CMASS galaxy redshift distribution, defined as the normalized distribution of the number density of galaxies as a function of redshift. The decrease of the filament intensity distribution results from the decrease of CMASS galaxy redshift distribution, from where the filaments in each redshift slice are detected. 82

4.7 Quantification of the deviation between $C_l^{\kappa f}$ estimated from simulation and from theoretical $C_l^{\kappa f}$. Since the approximations to get $C_l^{\kappa f}$ from simulations are not perfect, the $C_l^{\kappa f}$ from simulation will slightly deviate from the true value of $C_l^{\kappa f}$. Γ quantifies the deviation. In theory, We use two models for the smoothing introduced by filaments. In model1, filament length is the overall smoothing scale. In model2, filament length is the smoothing scale along the filament and we fit α for the smoothing in the perpendicular direction, where $1/k_\perp(z) \sim \alpha \times 1/k_\parallel(z)$ and k is the wavenumber in Fourier space. As shown in the plot, the deviation of $C_l^{\kappa f}$ between simulation and theory is less than 5% from unity for both theoretical models. 83

4.8 Example of filament grouped in redshift bin 0.55. A line with the same color is considered as belonging to the same filament. 84

List of Tables

2.1 A summary of our analysis. 35

4.1 The final result for the bias fitting. Model 1 uses filament length as the overall smoothing scale. In model 2, the filament length is the smoothing along filaments; we fit α for the smoothing in the perpendicular direction, where $1/k_{\perp}(z) \sim \alpha \times 1/k_{\parallel}(z)$. We get α to be 0.65 as the best fit. The bias and the error of bias come from χ^2 fitting of theory model to data. 71

Chapter 1

Introduction

Observations have shown that galaxies are distributed in a web-like structure, consisting of over-dense galaxy clusters, elongated filaments, flattened sheets, and near-empty voids, which are also known as the cosmic web [2]. The first extensive 3D survey of galaxies — the CfA redshift survey [3], has provided us with the first glimpse of the large scale structure distribution of the galaxies, while subsequent large redshift galaxy surveys, such as 2dF [4], SDSS [5], 6df [6], GAMA [7], and VIPERS [8], have furnished critical information regarding the spatial distribution of galaxies. The top and left slices of Fig. 1.1 illustrate the CfA2 'Great Wall' and Sloan Great Wall, respectively. The galaxy surveys contain a wealth of information to learn about cosmology. Researchers are using these surveys to study the initial conditions in the early universe, the matter and energy contents of the cosmos, and the physics of galaxy formation.

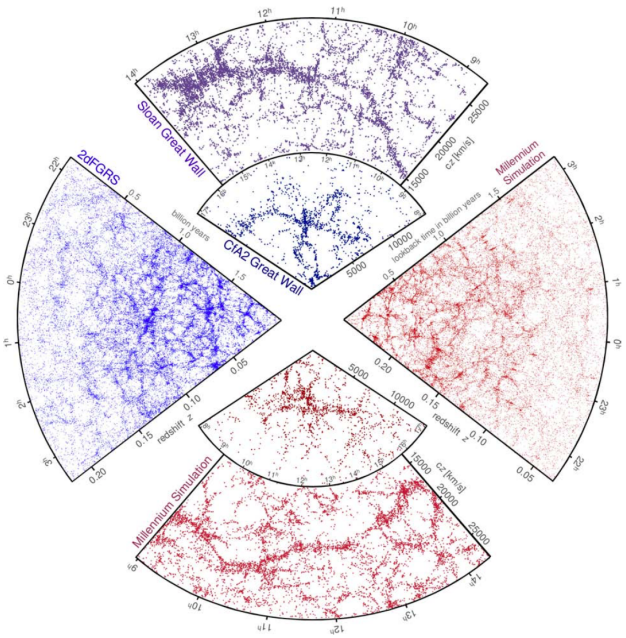


Figure 1.1: This figure comes from [1]. The top and left slices illustrate the CfA2 'Great Wall' and Sloan Great Wall, respectively. The right and bottom slices of fig. 1.1 illustrate the mock galaxy catalog from Millennium simulation. The large scale structure could be clearly in both the galaxy surveys and the simulations.

1.1 Dynamics of structure formation

The dynamics of structure formation is one of the key subjects of cosmology. In the ΛCDM model, the large scale structures in the Universe formed from a Gaussian initial condition. At the late time, gravity introduces the non-linear collapse of dark matter, generating non-Gaussian features such as the Sloan Great Wall[9]. Although the visible components of galaxies are made of baryonic matter (mostly hydrogen and helium), the gravitational forces that drive the growth of structure come mainly from dark matter[10]. The dark matter obeys the Vlasov equation – the governing equation to solve the distribution function in phase space, and the Vlasov equation is the master function in which all subsequent calculations of gravitational instability are derived[11]. The Vlasov equation is derived to be

$$\frac{df}{d\tau} = \frac{\partial f}{\partial \tau} + \frac{\boldsymbol{p}}{ma} \cdot \nabla f - am\nabla\Phi \cdot \frac{\partial f}{\partial \boldsymbol{p}} \tag{1.1}$$

where f is the particle number density in phase space, τ is conformal time, $\boldsymbol{x}$ is the particle position, $\boldsymbol{p}$ is the particle momenta, a is the cale factor, Φ is the gravitation potential and m is the particle mass. Needless to say, being a non-linear partial differential equation involving seven variables, this Vlasov equation is hard to solve.

1.1.1 Analytical Theories

To analytically understand the evolution of the density and velocity fields of matter fluctuation in the Universe, many approximate theories have been proposed. There are several branches of the approximate theories, including the Eulerian perturbation theory[11], Lagrangian perturbation Theory[11], effective field theory[12], etc. [13] has presented a direct comparison of several analytic predictions for the clustering of matter on quasi-linear scales. In general, on very

large scales, the matter distribution could be well modeled by the linear perturbation theory. On quasi-linear scales, the performances of the analytical theories begin to decrease, but there is still the possibility to model the matter distribution by extending the perturbation theories beyond the linear orders. But all the analytical theories would fail on small scales – highly non-linear scales, so far.

Among all the analytical theories, in this dissertation, we will focus on the Lagrangian perturbation Theory. The Lagrangian perturbation theory focuses on the trajectory of each particle. In the Lagrangian perturbation theory, the particles originally locate at coordinate $\boldsymbol{q}$, and the dynamic of the particles can be described by the Lagrangian displacement field $\boldsymbol{\Psi}$

$$\boldsymbol{x}(\tau) = \boldsymbol{q} + \boldsymbol{\Psi}(\boldsymbol{q}, \tau) \tag{1.2}$$

where τ is the comoving time and $\boldsymbol{\Psi}(\boldsymbol{q}, 0) = 0$.

The dynamics of the particles is governed by equation of emotion

$$\frac{d^2\boldsymbol{x}}{d\tau^2} + \mathcal{H}(\tau)\frac{d\boldsymbol{x}}{d\tau} = -\nabla_{\boldsymbol{x}}\Phi \tag{1.3}$$

where Φ is the gravitation potential governed by the Poisson equation

$$\nabla^2_{\boldsymbol{x}}\Phi = 4\pi G\rho = \frac{3}{2}\mathcal{H}^3\Omega_m\delta(\boldsymbol{x}) \tag{1.4}$$

where $\delta(\boldsymbol{x})$ is the density contrast, defined as $\delta(\boldsymbol{x}) = \rho(\boldsymbol{x})/\bar{\rho} - 1$. Due to the conservation of mass, we get the relation between Lagrangian and Eulerian coordinates

$$1 + \delta(\boldsymbol{x}, \tau) = |\frac{d^3q}{d^3x}| = \frac{1}{J(\boldsymbol{q}, \tau)} \tag{1.5}$$

where $J(\boldsymbol{q},\tau)$ is the Jacobian

$$J(\boldsymbol{q},\tau) = \det(\delta_{ij} + \Psi_{i,j}(\boldsymbol{q},\tau)) \tag{1.6}$$

By combinining eq. 1.3, eq. 1.4 and eq. 1.5, we get the master equation for Largaragian perturbation theory

$$J(\boldsymbol{q},\tau)\big[\delta_{ij}+\Psi_{i,j}(\boldsymbol{q},\tau)\big]^{-1}\big[\frac{d^2\Psi_{i,j}(\boldsymbol{q},\tau)}{d\tau^2}+\mathcal{H}(\tau)\frac{\Psi_{i,j}(\boldsymbol{q},\tau)}{d\tau}\big] = -\frac{3}{2}\mathcal{H}^2\Omega_m(\tau)[J(\boldsymbol{q},\tau)-1] \tag{1.7}$$

Lagrangian perturbation theory finds the perturbative solution for the displacement field Ψ

$$\boldsymbol{\Psi}(\boldsymbol{q},\tau) = \boldsymbol{\Psi}^{(1)}(\boldsymbol{q},\tau) + \boldsymbol{\Psi}^{(2)}(\boldsymbol{q},\tau) + \boldsymbol{\Psi}^{(3)}(\boldsymbol{q},\tau) + ... \tag{1.8}$$

By approximate eq. 1.7 to first order, we get the first order solution of the displacement field

$$\nabla_q \cdot \boldsymbol{\Psi}^{(1)} = -D_1(\tau)\delta(\boldsymbol{q}) \tag{1.9}$$

where $D_1(\tau)$ is the linear growth function

$$D_1^{''}(\tau) + \mathcal{H}(\tau)D_1^{'}(\tau) = \frac{3}{2}\mathcal{H}^2(\tau)\Omega_m(\tau)D_1(\tau) \tag{1.10}$$

The first order solution of perturbation theory is also called Zel'dovich approximation[14]. The solution of the local density of Zel'dovich approximation can be simplified as

$$1 + \delta(\boldsymbol{q},\tau) = \frac{1}{(1-\lambda_1 D_1(\tau))(1-\lambda_2 D_2(\tau))(1-\lambda_3 D_3(\tau))} \tag{1.11}$$

where λ_i are the eigenvalues of the $\Psi_{i,j}$. Let us assume $\lambda_1 > \lambda_2 > \lambda_3$. From this expression we could see when the linear perturbation grows as $D_1(\tau)$ reaches $1/\lambda_1$, the Zel'dovich solution leads to planar collapse and the density field becomes in-

finite. This stage is called Zel'dovich pancake. Indeed, Zel'dovich approximation breaks down before the density becomes infinite. Since for these small scales, we need to take into consideration the pressure effect and shell crossing.

By approximate eq. 1.7 to second order, we get solution of second order perturbation theory:

$$\nabla \cdot \mathbf{\Psi}^{(2)} = \frac{1}{2} D_2(\tau) \sum_{i \neq j} \left\{ \psi^{(1)}_{i,i}(\boldsymbol{q}, \tau) \psi^{(1)}_{j,j}(\boldsymbol{q}, \tau) - \psi^{(1)}_{i,j}(\boldsymbol{q}, \tau) \psi^{(1)}_{j,i}(\boldsymbol{q}, \tau) \right\} \tag{1.12}$$

where the time evolution is governed by

$$D_2''(\tau) + \mathcal{H}(\tau) D_2'(\tau) - \frac{3}{2} \mathcal{H}^2(\tau) \Omega_m(\tau) D_2(\tau) = \frac{3}{2} \mathcal{H}^2(\tau) \Omega_m(\tau) \left[D_1(\tau) \right]^2 \tag{1.13}$$

The second order perturbation theory adds a quadratic correction on top of Zel'dovich approximation to make a better prediction on quasi-linear regimes.

It is also possible to extend the second order solution to higher orders. However, it becomes more expensive to solve the equations as it goes to higher orders. In addition, going to higher orders does not guarantee to give us better solutions, leading one to question the convergence properties of such a series expansion of the perturbation theory[13].

1.1.2 N-body dark matter only Simulations

As mentioned before, it is intractable for analytical theories on small non-linear scales. To resolve the small scale clustering of the mass, we need to resort to numerical N-body simulations[15]. The N-body simulations solve the equation of motion (eq. 1.3) and Poisson equation (eq. 1.4) numerically. The idea of N-body

simulations is that the density-velocity field is represented by a set of particles. The N-body simulations evolve the trajectory of particles as a function of time with a numerical integration scheme. At every time step, the mass distribution is used to compute the gravitational potential field, which in turn yields the accelerations of particles.

One of the key aspects of N-body simulation is force evaluation. Because gravity is a long-range force, we cannot ignore the force from distant particles, and the force calculation can be time-consuming. Many different schemes have been proposed to address the force evaluation, such as particle-mesh method[16; 17], tree method[18; 19], P^3M method[20; 21], etc.

In this dissertation, we are going to use FastPM scheme [22]. FastPM scheme belongs to the big category of the particle-mesh method. For the particle-mesh method, the Poisson equation can be solved efficiently via Fast Fourier transforms (FFT), where the FFT mesh and the finite differentiation kernel acts as a force smoothing that regularizes the small scale structure formation near the resolution of the mesh. The acceleration is then fed into an integrator (e.g.Leap-frog) to update velocity and position. Additional care shall be taken when the time steps are sparse to conserve the growth of structures at large scales[23]. The FastPM enforce the linear theory growth to be correct with sparse time steps by modifying the traditional integrator in the particle-mesh scheme. FastPM scheme is shown to approach a full N-body simulation with high accuracy quickly.

Numerical N-body simulations have wide applications in cosmology. They are used to understand non-linear evolution of the Universe[24], to compute the covariance for different observables[25; 26], to identify the best observable to quantify the information content[27], etc.

Though N-body simulations are useful to study the structure formation, it is also essential to know the limitations of the N-body simulations[15; 28]. The limi-

tations include the miss long wave modes – wave modes larger than the simulation box, two body relaxation, poor accuracy in force calculation and discreteness, etc. One of the limitations we are going to talk more in this dissertation is that the simulations are computationally challenging even though the available computing power gets better.

1.1.3 Link between Mass Distribution and Galaxy Distribution

N-body simulations follow the gravitationally dominant dark matter component to simulate the large scale structure formation of the Universe. Another long-standing issue is how to compare simulated mass distributions with observed galaxy distributions. However, the relation between the galaxies and dark matter densities depends on the aspects of galaxy formation that are difficult to model theoretically[10]. One way to generate Mock catalogs of simulated galaxies are using friend-of-friends algorithms to find the dark matter halos from the dark matter only N-body simulations and then using halo occupation distributions to populate galaxies to halos based on halo mass[29]. Besides halo occupation distribution, halo abundance matching is another efficient method where more luminous galaxies are assigned to more massive halos[30]. An alternative way for generating Mock galaxy catalogs is to use Semi-analytic models[31], where a merger tree that represents the formation and growth of a dark matter halo is used. Besides the analytical models to connect observed galaxies to simulated dark matter halos, hydro-dynamical simulations can also be used. The hydro-dynamical simulations directly account for the baryonic component (gas, stars, supermassive black holes, etc.) in cosmological simulations as well as the dark matter. However, the hydrodynamics simulations are known to be computationally expensive. The right and bottom slices of fig. 1.1 illustrate the mock galaxy

catalog from Millennium simulation[24].

1.2 Measurements of the clustering of Large Scale Structure

With the observations from the surveys, Mock galaxy catalogs and simulations, quantitative measurements of the clustering of large scale structure are required to extract useful information. There are several ways to measure the clustering[10], including measurements of the correlation functions, angular clustering, galaxy bias, etc. In this dissertation, we are going to focus on the correlation functions, filaments as a way to extract higher order statistics, and deep learning methods.

1.2.1 Correlation Functions

The most widely used statistical method to study the galaxy clustering is the two-point correlation function[32]. The two-point correlation function ($\xi(r)$) measures the galaxy clustering as a function of scale. It is defined as the excess probability (dP) compared with a random distribution of finding a pair of galaxies with the separating distance r

$$dP = \bar{n}^2(1 + \xi(r_{12}))dV_1dV_2 \tag{1.14}$$

where $\bar{n}$ is the mean number density of galaxies.

The two-point correlation function could be measured through counting of galaxy pairs and comparing the counting with the Poisson distribution. Several

methods to measure the two-point correlation function have been proposed

$$\begin{aligned}\hat{\xi}_{DP} &= \frac{\langle DD\rangle}{\langle DR\rangle} - 1 \\ \hat{\xi}_{HAM} &= \frac{\langle DD\rangle\langle RR\rangle}{\langle DR\rangle^2} - 1 \\ \hat{\xi}_{LS} &= \frac{\langle (D-R)^2\rangle}{\langle RR\rangle} - 1\end{aligned} \tag{1.15}$$

where D and R represent data and random point set, respectively.

The Fourier transfer of two-point correlation function is the power spectrum (P(k)).

$$P(k) = \frac{1}{2\pi^2}\int dk k^2 P(k)\frac{sin(kr)}{kr} \tag{1.16}$$

where k is the wavenumber, defined as $k = 2\pi/\lambda$.

The power spectrum is predicted directly by theory of the formation of the large scale structure. The primordial power spectrum is assumed to follow a power law $P(k) \sim k^n$. The growth of density perturbation on different scales is determined by self-gravitation, pressure support and damping process, and so the shape of the primordial power spectrum would be modified. The modifications in primordial power spectrum are expressed by the tranfer function T(k)

$$P(k,z) = A(z)k^n T(k) \tag{1.17}$$

where A(z) is the normalization factor determined by observations.

If the described field is Gaussian, all the information would be included by the two-point statistics. But because the large scale structure of the Universe is highly non-Gaussian, we expect higher order statistics to provide us further information. Beyond the two-point statistics, researchers also measure the three-point correlation function [33; 34], which is the lowest order statistical tool to probe the non-Gaussianity. By studying the three-point statistics, we could put

strong constraints on models of structure formation [35].

The 3-point correlation function ($\zeta(r_1, r_2, \theta)$) expresses the correlation of the field among 3 locations in the configuration space, which is equivalently defined as bispectrum ($B(k_1, k_2, k_3)$) in Fourier space.

$$\begin{aligned} &\zeta(r_1, r_2, \theta) = \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x} + \boldsymbol{r_1})\delta(\boldsymbol{x} + \boldsymbol{r_2})\rangle. \\ &B(k_1, k_2, k_3)\delta_D(\boldsymbol{k_1} + \boldsymbol{k_2} + \boldsymbol{k_3}) = \langle \delta(\boldsymbol{k_1})\delta(\boldsymbol{k_2})\delta(\boldsymbol{k_3})\rangle \end{aligned} \tag{1.18}$$

1.2.2 Higher Order Statistics

By studying the summary statistics such as the two-point and three-point correlation function, we are still losing non-Gaussian information of the large scale structure encoded beyond. Additional methods are needed to extract non-Gaussian information from the large scale structure of the Universe.

1.2.2.1 Filaments

One of the methods to extract extra information is to study the properties filaments. The filaments are the most prominent and highly non-linear structure of the large scale structure of the Universe, which acts as the bridges between the galaxy clusters. It is shown in the simulations that filaments occupy 10% of the volume with 40% of mass at z=0[36].

Various methods have been proposed to identify and characterize the morphologies and properties of filaments[37; 38; 39; 40]. A comparison of different filament finders have been presented in [41; 42].

Filaments are of interest for several reasons. Detecting and characterizing filaments at a range of redshifts could help us understand the evolution of the structure formation of the Universe [2]. By studying the properties of filaments, one could have a more thorough understanding of the environmental influence

on the formation and evolution of galaxies[43; 44]. One active subject of the investigation is to find the alignment between galaxies and filaments[45; 46; 47]. Filaments are tracers of the underlying dark matter, and several studies have studied gravitational lensing of filaments to infer the mass distribution within the filaments[42; 48; 49]. By studying filament using thermal Sunyaev-Zeldovich effect might help reveal the properties of missing baryons, the vast majority of the gas which resides in the intergalactic medium and has so far evaded most observations[50].

1.3 Deep Learning Methods

Beyond the traditional methods, the fast-growing deep learning methods provide another direction to extract useful information from the large scale structure of the Universe. Deep learning is a fast-growing branch of machine learning. It is based on neural networks and is composed of multiple layers to learn the data representations. Deep learning has led many breakthroughs in multiple areas, such as image recognition[51; 52; 53], speech recognition[54; 55]. Deep learning is increasingly being adopted in cosmology researches too. It has applications in large scale structure[56; 57], gravitational lensing[58; 59], weak lensing[60], cosmic microwave background[61; 62; 63], cosmic reionization[64] supernova[65], etc.

Convolutional neural networks are a common class of deep learning methods[66], which have recently obtained remarkable experimental results. The convolutional neural networks are designed to process data in regular formats, such as 2D images or 3D volumetric data.

The typical architecture of the convolutional neural networks usually contains several convolution layers and pooling layers. The convolution layers consist of a set of learnable filters. These filters are connected to the local patches of

the output of the previous layers and a weighted sum is calculated. Next, the weighted sum is passed through non-linear activation, such as ReLU, tanh. All the local patches share the same weight filters, which is known as the weight sharing scheme. The idea of weight sharing is based on the assumption that local groups of values are often highly correlated and the local statistics of images and other signals are invariant to location. By using the weight sharing scheme, the number of learnable parameters is largely reduced. The role of convolution layers is to detect local conjunctions of features from the previous layer. The pooling layers are used to reduce the spatial size of the representation so that the number of parameters is reduced. A typical way to do the pooling is to calculate the maximum value or the average of the local patches of the output of the previous layer.

Other essential tricks widely used in building the convolution neural networks and are used in this dissertation are padding and batch normalization. Paddings are usually used before the convolution to control the spatial size of the output volumes and to maintain features at borders. Usually, zero-padding – pad around the borders with all 0, or reflective padding – pad around the borders with its reflection, is used. Batch normalization is known to improve the speed, performance, and stability of neural networks. The batch normalization is performed by subtracting the output of the activation layer with the batch mean and dividing by the batch standard deviation.

One of the limitations of the convolutional architectures is that they require highly regular input data formats. However, with data represented as point set – set of data points in space, the convolutional neural network is not an optimal choice anymore. Though the point set data can be converted into meshes or voxels representation, the information in scales smaller than the mesh or voxel size would be lost. In addition, if the data points are sparse, converting the data

to mesh or voxel representation would result in many empty regions, which makes it hard for the convolutional neural networks to extract features. To deal with the point set data, another recently developed class of deep learning methods, the point cloud model, can be used. Extensive experiments have shown the point cloud model performs well across several domains[67; 68]. In cosmology, the point cloud model is used to predict the redshift from the galaxy clustering[68].

So far, deep learning is shown to be a powerful method to learn complex functions and could have many applications in several domains in cosmology. But there are also limitations of the deep learning methods. Firstly, the deep learning model is a black box. Once a deep learning model is trained, it is not always clear how its making its decisions. Many efforts have been made to interpret the deep learning models. [69] gives an overview of techniques for interpreting complex machine learning models, with a focus on deep neural networks. However, as also mentioned in this paper, interpreting deep networks remains a young and emerging field of research. Secondly, Standard deep learning methods for regression and classification do not capture model uncertainty. However, error estimation is essential in cosmology; for example, to calculate the likelihood function. Though some work has been proposed to predict the model uncertainty[61; 70; 71], this field of research needs more investigations.

1.4 Organization of the Dissertation

The large scale structure of the Universe contains much information to constrain the cosmology. It is essential to understand the formation of large scale structure formation and to find statistical methods to extract useful information from large scale structure of the Universe. As is shown, one of the biggest challenges faced by studying large scale structure of the Universe is the non-linearity of small

scales – the small scale non-linearity causes analytical theories to fall; large computational power of simulations is needed to resolve the small scale clustering; other statistical methods are required to extract information which is ignored by the summary statistics. In this dissertation, we strive to find new techniques to address the non-linearity of the large scale structure of the Universe.

In the first part of this dissertation, we use the new technique, the deep learning methods, to address the non-linear evolution of large scale structure of the Universe. The main goal is to provide a fast and robust method to simulate the non-linear structure formation of the Universe.

In the second part of the dissertation, our goal is to find a new way to extract information from the non-linear structure of the Universe. We directly work on the highly non-linear structure of the large scale structure of the Universe – filaments, to provide us additional information.

The dissertation is organized as follows.

In Chapter 2, we build a deep neural network, the Deep Density Displacement Model, to predict the non-linear structure formation of the Universe from simple linear perturbation theory. Our extensive analysis demonstrates that our model outperforms the second order perturbation theory (hereafter 2LPT), the commonly used fast approximate simulation method, in point-wise comparison, 2-point correlation, and 3-point correlation. We also show that D^3M can accurately extrapolate far beyond its training data, and predict structure formation for significantly different cosmological parameters.

In chapter 3.1, we have interrogated the trained model in chapter 1 with physics analysis, showing the trained neural networks have expected behaviors as the physics prediction.

In chapter 3.2, we have developed a multi-scale deep set model, which directly uses sets of particles as input, to predict the non-linear evolution of the Universe.

Our objective is to show that our model is another efficient deep model that can be used in solving complex cosmological problems with cosmological data represented as point clouds.

In chapter 4, we report the first detection of CMB (Cosmic Microwave Background) lensing by filaments, and we apply a null test to confirm our detection. Furthermore, we propose a phenomenological model to interpret the detected signal, and we measure how filaments trace the matter distribution on large scales through filament bias. Our study provides a new scope to understand the environmental dependence of galaxy formation.

The following papers are discussed in this dissertation:

Learning to Predict the Cosmological Structure Formation
Siyu He, Yin Li, Yu Feng, Shirley Ho, Siamak Ravanbakhsh, Wei Chen, and Barnabs Pczos
Proceedings of the National Academy of Sciences of the United States of America in press
This paper is discussed in chaper 2.

Interrogation of the deep learning model with physics analysis
Siyu He, Yin Li, Shirley Ho, Yu Feng, Siamak Ravanbakhsh
In preparation.
This paper is discussed in chapter 3.1.

Multi-Scale Deep Sets to Learn Non-Linear Evolution of the Universe
Siyu He, Siamak Ravanbakhsh, Daniele Bertolini, Evan Cushing, Shirley Ho
In preparation.
This paper is discussed in chapter 3.2.

The detection of the imprint of filaments on Cosmic Microwave Background (CMB) lensing
Siyu He, Shadab Alam, Simone Ferraro, Yen-Chi Chen, and Shirley Ho
Nature Astronomy volume 2 pages 401-406 (2018)
This paper is discussed in chapter 4.

The following papers are not discussed in this dissertation:

HIGAN: Cosmic Neutral Hydrogen with Generative Adversarial Networks
Juan Zamudio-Fernandez, Atakan Okan, Francisco Villaescusa-Navarro, Seda Bilaloglu, Asena Derin Cengiz, **Siyu He**, Laurence Perreault Levasseur, Shirley Ho
preprint arXiv:1904.12846

From Dark Matter to Galaxies with Convolutional Networks
Xinyue Zhang, Yanfang Wang, Wei Zhang, Yueqiu Sun, **Siyu He**, Gabriella Contardo, Francisco Villaescusa-Navarro, Shirley Ho
preprint arXiv:1902.05965

Analysis of Cosmic Microwave Background with Deep Learning
Siyu He, Siamak Ravanbakhsh, Shirley Ho
Workshop at International Conference on Learning Representations 2018

CosmoFlow: using deep learning to learn the universe at scale
Amrita Mathuriya, Deborah Bard, Peter Mendygral, Lawrence Meadows, James Arnemann, Lei Shao, **Siyu He**, Tuomas Krn, Diana Moise, Simon J Pennycook, Kristyn Maschhoff, Jason Sewall, Nalini Kumar, Shirley Ho, Michael F Ringenburg, Prabhat Prabhat, Victor Lee
SC18: International Conference for High Performance Computing, Networking, Storage and Analysis

Detecting galaxyfilament alignments in the Sloan Digital Sky Survey III
Yen-Chi Chen, Shirley Ho, Jonathan Blazek, **Siyu He**, Rachel Mandelbaum, Peter Melchior, Sukhdeep Singh
Monthly Notices of the Royal Astronomical Volume 485 Pages 2492-2504 (2019)

Constraining gravity at the largest scales through CMB lensing and galaxy velocities

Anthony R Pullen, Shadab Alam, **Siyu He**, Shirley Ho

Monthly Notices of the Royal Astronomical Volume 460 Pages 4098-4108 (2016)

Chapter 2

Learning to Predict the Cosmological Structure Formation

2.1 Abstract

Matter evolved under influence of gravity from minuscule density fluctuations. Non-perturbative structure formed hierarchically over all scales, and developed non-Gaussian features in the Universe, known as the Cosmic Web. To fully understand the structure formation of the Universe is one of the holy grails of modern astrophysics. Astrophysicists survey large volumes of the Universe and employ a large ensemble of computer simulations to compare with the observed data in order to extract the full information of our own Universe. However, to evolve billions of particles over billions of years even with the simplest physics is a daunting task. We build a deep neural network, the Deep Density Displacement Model (hereafter D^3M), which learns from a set of pre-run numerical simulations, to predict the non-linear large scale structure of the Universe with Zel'dovich Ap-

proximation (hereafter ZA), an analytical approximation based on perturbation theory, as the input. Our extensive analysis, demonstrates that D^3M outperforms the second order perturbation theory (hereafter 2LPT), the commonly used fast approximate simulation method, in predicting cosmic structure in the non-linear regime. We also show that D^3M is able to accurately extrapolate far beyond its training data, and predict structure formation for significantly different cosmological parameters. Our study proves, for the first time, that deep learning is a practical and accurate alternative to approximate 3D simulations of the gravitational structure formation of the Universe.

2.2 Introduction

Astrophysicists require a large amount of simulations to extract the information from observations [4; 5; 6; 7; 8; 72; 73; 74]. At its core, modeling structure formation of the Universe is a computationally challenging task; it involves evolving billions of particles with the correct physical model over a large volume over billions of years [75; 76; 77]. To simplify this task, we either simulate a large volume with simpler physics or a smaller volume with more complex physics. In order to produce the cosmic web [78] in large volume, we select gravity, the most important component of the theory, to simulate at large scales. A gravity-only N-body simulation is the most popular; and effective numerical method to predict the full 6D phase space distribution of a large number of massive particles whose position and velocity evolve over time in the Universe [79]. Nonetheless, N-body simulations are relatively computationally expensive, thus making the comparison of the N-body simulated large-scale structure (of different underlying cosmological parameters) with the observed Universe a challenging task. We propose to use a deep model that predicts the structure formation as an alternative to N-body

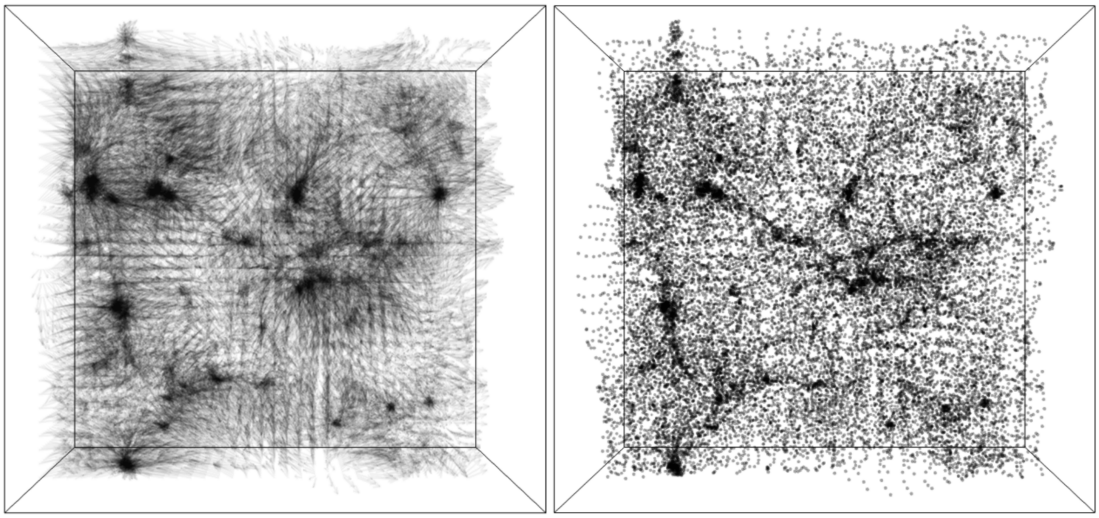


Figure 2.1: (left) The displacement vector-field and (right) the resulting density field produced by D^3M. The vectors in the left figure are uniformly scaled down for better visualization.

simulations.

Deep learning [80] is a fast growing branch of machine learning where recent advances have lead to models that reach and sometimes exceed human performance across diverse areas, from analysis and synthesis of images [51; 52; 53], sound [54; 55], text [81; 82] and videos [83; 84] to complex control and planning tasks as they appear in robotics and game-play [85; 86; 87]. This new paradigm is also significantly impacting a variety of domains in the sciences, from biology [88; 89] to chemistry [90; 91] and physics [92; 93]. In particular, in astronomy and cosmology, a growing number of recent studies are using deep learning for a variety of tasks, ranging from analysis of cosmic microwave background [61; 62; 63], large-scale structure [56; 57], and gravitational lensing effects [58; 59] to classification of different light sources [94; 95; 96].

The ability of these models to learn complex functions has motivated many to use them to understand the physics of interacting objects leveraging image, video and relational data [97; 98; 99; 100; 101; 102; 103; 104; 105; 106; 107].

However, modeling the dynamics of billions of particles in N-body simulations poses a distinct challenge.

In this paper we show that a variation on the architecture of a well-known deep learning model [108], can efficiently transform the first order approximations of the displacement field and approximate the exact solutions, thereby producing accurate estimates of the large-scale structure. Our key objective is to prove that this approach is an accurate and computationally efficient alternative to expensive cosmological simulations, and to this end we provide an extensive analysis of the results in the following section.

The outcome of a typical N-body simulation depends on both the initial conditions and on cosmological parameters which affect the evolution equations. A striking discovery is that D^3M, trained using a single set of cosmological parameters generalizes to new sets of significantly different parameters, minimizing the need for training data on diverse range of cosmological parameters.

2.3 Setup

We build a deep neural network, D^3M, with similar input and output of an N-body simulation. The input to our D^3M is the displacement field from ZA [109]. A displacement vector is the difference of a particle position at target redshift $z = 0$, i.e., the present time, and its Lagrangian position on a uniform grid. ZA evolves the particles on linear trajectories along their initial displacements. It is accurate when the displacement is small, therefore ZA is frequently used to construct the initial conditions of N-body simulations [14]. As for the ground truth, the target displacement field is produced using FastPM [22], a recent approximate N-body simulation scheme that is based on a particle-mesh (PM) solver. FastPM quickly approaches a full N-body simulation with high accuracy and provides a viable

alternative to direct N-body simulations for the purpose of our study.

A significantly faster approximation of N-body simulations is produced by second-order Lagrangian perturbation theory (hereafter 2LPT), which bends each particle's trajectory with a quadratic correction [110]. 2LPT is used in many cosmological analyses to generate a large number of cosmological simulations for comparison of astronomical dataset against the physical model [111; 112] or to compute the covariance of the dataset [113; 114; 115]. We regard 2LPT as an effective way to efficiently generate a relatively accurate description of the large-scale structure and therefore we select 2LPT as the reference model for comparison with $\mathrm{D}^3\mathrm{M}$.

We generate 10,000 pairs of ZA approximations as input and accurate FastPM approximations as target. We use simulations of 32^3 N-body particles in a volume of $128\,h^{-1}$Mpc (600 million light years, where $h = 0.7$ is the Hubble parameter). The particles have a mean separation of $4\,h^{-1}$Mpc per dimension.

2.3.1 Displacement versus Density Field

An important choice in our approach is training with displacement field rather than density field. Displacement field $\mathbf{\Psi}$ and density field ρ are two ways of describing the same distribution of particles. And an equivalent way to describe density field is the over-density field, defined as $\delta = \rho/\bar{\rho} - 1$, with $\bar{\rho}$ denoting the mean density. The displacement field and over-density field are related by eq. 2.1.

$$\begin{aligned} \mathbf{x} &= \mathbf{\Psi}(\mathbf{q}) + \mathbf{q} \\ \delta(\mathbf{x}) &= \int d^3q \delta_D(\mathbf{x} - \mathbf{q} - \mathbf{\Psi}(\mathbf{q})) - 1 \end{aligned} \tag{2.1}$$

When the displacement field is small and has zero curl, the choice of over-

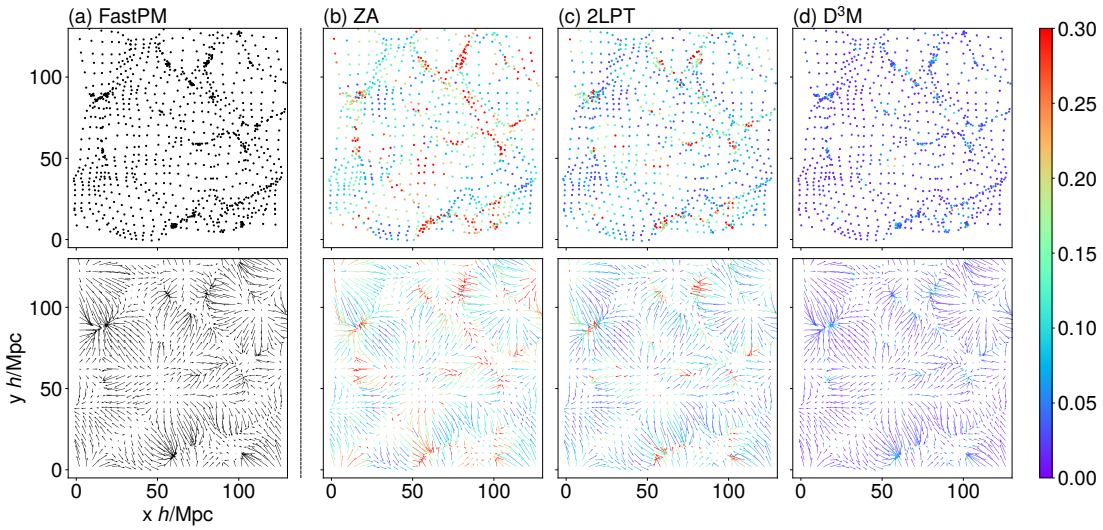


Figure 2.2: The columns show 2-D slices of full particle distribution (top) and displacement vector (bottom) by various models, from left to right:
(a) FastPM: the target ground truth, a recent approximate N-body simulation scheme that is based on a particle-mesh (PM) solver ;
(b) Zel'dovich approximation (ZA): a simple linear model that evolves particle along the initial velocity vector;
(c) Second order Lagrangian perturbation theory (2LPT): a commonly used analytical approximatation;
(d) Deep learning model (D^3M) as presented in this work.
While FastPM (a) served as our ground truth, the columns to its right includes color for the points or vectors. The color indicates the relative difference $(q_{model} - q_{target})/q_{target}$ between the target (a) location or displacement vector and predicted distributions by various methods (b-d). The error-bar shows denser regions have a higher error for all methods. which suggests that it is harder to predict highly non-linear region correctly for all models: D^3M, 2LPT and ZA. Our model D^3M has smallest differences between predictions and ground truth among the above models (b)-(d).

density vs displacement field for the output of the model is irrelevant, as there is a bijective map between these two representations, described by the equation:

$$\mathbf{\Psi} = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{q}} \frac{i\mathbf{k}}{k^2} \delta(\mathbf{k}) \tag{2.2}$$

However as the displacements grow into the non-linear regime of structure formation, different displacement fields can produce identical density fields [e.g. 116]. Therefore, providing the model with the target displacement field during the training eliminates the ambiguity associated with the density field. Our inability to produce comparable results when using the density field as our input and target attests that relevant information resides in the displacement field (See Appendix, Fig.2.6) .

2.4 Results and Analysis

Figure 2.1 shows the displacement vector field as predicted by D^3M (left) and the associated point-cloud representation of the structure formation (right). It is possible to identify structures such as clusters, filaments and voids in this point-cloud representation. We proceed to compare the accuracy of D^3M and 2LPT compared with ground truth.

2.4.1 Point-Wise Comparison

Let $\mathbf{\Psi} \in \mathbb{R}^{d\times d\times d\times 3}$ denote the displacement field, where d is the number of spatial resolution elements in each dimension ($d = 32$). A natural measure of error is the relative error $|\hat{\mathbf{\Psi}} - \mathbf{\Psi}_t|/|\mathbf{\Psi}_t|$, where $\mathbf{\Psi}_t$ is the true displacement field (FastPM) and $\hat{\mathbf{\Psi}}$ is the prediction from 2LPT or D^3M. Figure 2.2 compares this error for different approximations in a 2-D slice of a single simulation. We observe that

D^3M predictions are very close to the ground truth, with a maximum relative error of 1.10 over all 1000 simulations. For 2LPT this number is significantly higher at 4.23. In average, the result of D^3M comes with a 2.8% relative error while for 2LPT it equals 9.3%.

2.4.2 2-Point Correlation Comparison

As suggested by Figure 2.2 the denser regions seem to have a higher error for all methods – that is, more non-linearity in structure formation creates larger errors for both D^3M and 2LPT. The dependence of error on scale is computed with 2-point and 3-point correlation analysis.

Cosmologists often employ compressed summary statistics of the density field in their studies. The most widely used of these statistics are the 2-point correlation function (2PCF) $\xi(r)$ and its Fourier transform, the power spectrum $P_{\delta\delta}(k)$:

$$\begin{aligned} \xi(|\boldsymbol{r}|) &= \langle \delta_A(\boldsymbol{r}')\delta_B(\boldsymbol{r}'+\boldsymbol{r})\rangle, \\ P_{\delta\delta}(|\boldsymbol{k}|) &= \int \mathrm{d}^3\boldsymbol{r}\ \xi(r)e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \end{aligned} \tag{2.3}$$

where the ensemble average $\langle\,\rangle$ is taken over all possible realizations of the Universe. Our Universe is observed to be both homogeneous and isotropic on large scales, i.e. without any special location or direction. This allows one to drop the dependencies on $\boldsymbol{r}'$ and on the direction of $\boldsymbol{r}$, leaving only the amplitude $|\boldsymbol{r}|$ in the final definition of $\xi(r)$. In the second equation, $P_{\delta\delta}(k)$ is simply the Fourier transform of $\xi(r)$, and captures the dispersion of the plane wave amplitudes at different scales in the Fourier space. $\boldsymbol{k}$ is the 3D wavevector of the plane wave, and its amplitude k (the wavenumber) is related to the wavelength λ by $k = 2\pi/\lambda$. Due to isotropy of the Universe, we drop the vector form of $\boldsymbol{r}$ and $\boldsymbol{k}$.

Because FastPM, 2LPT and D^{3}M take the displacement field as input and output, we also study the two-point statistics for the displacement field. The displacement power spectrum is defined as:

$$P_{\mathbf{\Psi\Psi}}(k) = \langle \Psi_x(k)\Psi_x^*(k)\rangle + \langle \Psi_y(k)\Psi_y^*(k)\rangle + \langle \Psi_z(k)\Psi_z^*(k)\rangle \tag{2.4}$$

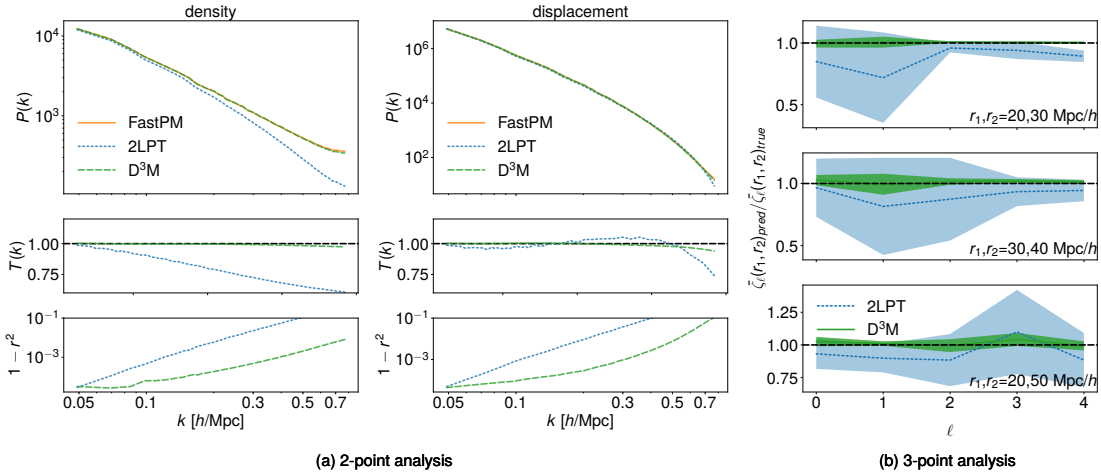


Figure 2.3: (**a**) From top to bottom: (top) displacement and density power-spectrum of FastPM (orange), 2LPT (blue), and D^{3}M (green); (middle) transfer function – i.e., the square root of the ratio of the predicted power-spectrum to the ground truth; (bottom) 1-r^2 where r is the correlation coefficient between the predicted fields and the true fields. Results are the averaged values of 1000 test simulations. The transfer function and correlation coefficient of the D^{3}M predictions is nearly perfect from large to intermediate scales and outperforms our benchmark 2LPT significantly.
(**b**) The ratios of the multipole coefficients ($\zeta_l(r_1, r_2)$) (to the target) of the two 3-point correlation functions for several triangle configurations. The results are averaged over 10 test simulations. The error-bars (padded regions) are the standard deviations derived from 10 test simulations. The ratio shows the 3-point correlation function of D^{3}M is closer than 2LPT to our target FastPM with lower variance.

We focus on the Fourier-space representation of the 2-point correlation. Because the matter and the displacement power spectrum take the same form, in

what follows we drop the subscript for matter and displacement field and use $P(k)$ to stand for both matter and displacement power spectrum. We employ the transfer function $T(k)$ and the correlation coefficient $r(k)$ as metrics to quantify the model performance against the ground truth (FastPM) in the 2-point correlation. We define the transfer function $T(k)$ as the square root of the ratio of two power spectra,

$$T(k) = \sqrt{\frac{P_{\rm pred}(k)}{P_{\rm true}(k)}}, \tag{2.5}$$

where $P_{\rm pred}(k)$ is the density or displacement power spectrum as predicted by 2LPT or $\mathrm{D}^3\mathrm{M}$, and analogously $P_{\rm true}(k)$ is the ground truth predicted by FastPM. The correlation coefficient r(k) is a form of normalized cross power spectrum,

$$r(k) = \frac{P_{\rm pred\times true}(k)}{\sqrt{P_{\rm pred}(k)P_{\rm true}(k)}}, \tag{2.6}$$

where $P_{\rm pred\times true}(k)$ is the cross power spectrum between 2LPT or $\mathrm{D}^3\mathrm{M}$ predictions and the ground truth (FastPM) simulation result. The transfer function captures the discrepancy between amplitudes, while the correlation coefficient can indicate the discrepancy between phases as functions of scales. For a perfectly accurate prediction, $T(k)$ and $r(k)$ are both 1. In particular, $1-r^2$ describes stochasticity, the fraction of the variance in the prediction that cannot be explained by the true model.

Figures 2.3(a) shows the average power spectrum, transfer function $T(k)$ and stochasticity $1 - r^2(k)$ of the displacement field and the density field over 1000 simulations. The transfer function of density from 2LPT predictions is 2% smaller than that of FastPM on large scales ($k \approx 0.05\,h\mathrm{Mpc}^{-1}$). This is expected since 2LPT performs accurately on very large scales ($k < 0.01\,h\mathrm{Mpc}^{-1}$). The displacement transfer function of 2LPT increases above 1 at $k \approx 0.35\,h\mathrm{Mpc}^{-1}$ and then drops sharply. The increase of 2LPT displacement transfer function is because

2LPT over-estimates the displacement power at small scales [see, e.g. 117]. There is a sharp drop of power near the voxel scale because smoothing over voxel scales in our predictions automatically erases power at scales smaller than the voxel size.

Now we turn to the D^{3}M predictions: both the density and displacement transfer functions of the D^{3}M differ from 1 by a mere 0.4% at scale $k \lesssim 0.4\,h\mathrm{Mpc}^{-1}$, and this discrepancy only increases to 2% and 4% for density field and displacement field respectively, as k increases to the Nyquist frequency around $0.7\,h\mathrm{Mpc}^{-1}$. The stochasticity hovers at approximately 10^{-3} and 10^{-2} for most scales. In other words, for both the density and displacement fields the correlation coefficient between the D^{3}M predictions and FastPM simulations, all the way down to small scales of $k = 0.7\,h\mathrm{Mpc}^{-1}$ is greater than 90%. The transfer function and correlation coefficient of the D^{3}M predictions shows that it can reproduce the structure formation of the Universe from large to semi-non-linear scales. D^{3}M significantly outperforms our benchmark model 2LPT in the 2 point function analysis. D^{3}M only starts to deviate from the ground truth at fairly small scales. This is not surprising as the deeply nonlinear evolution at these scales is more difficult to simulate accurately and appears to be intractable by current analytical theories[118].

2.4.3 3-Point Correlation Comparison

The 3-point correlation function (3PCF) expresses the correlation of the field of interest among 3 locations in the configuration space, which is equivalently defined as bispectrum in Fourier space. Here we concentrate on the 3PCF for computational convenience:

$$\zeta(r_1, r_2, \theta) = \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x} + \boldsymbol{r_1})\delta(\boldsymbol{x} + \boldsymbol{r_2})\rangle. \tag{2.7}$$

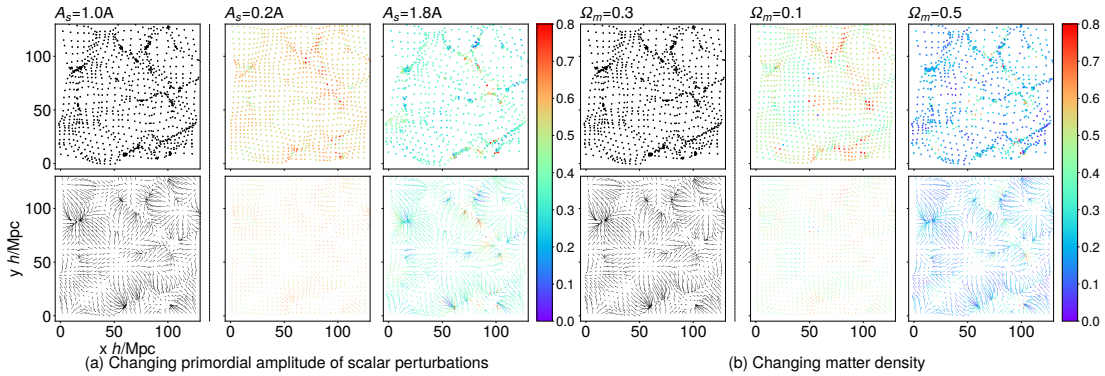


Figure 2.4: We show the differences of particle distributions and displacement fields when we change the cosmological parameters A_s and Ω_m.
(**a**) The errorbar shows the difference of particle distribution (upper panel) and displacement fields (lower panel) between $A_s = A_0$ and the two extremes for $A_s = .2A_0$(left) and $A_s = 1.8A_0$ (right).
(**b**) A similar comparison showing the difference of the particle distributions (upper panel) and displacement fields (lower panel) for smaller and larger values of $\Omega_m \in \{.1, .5\}$ with regard to $\Omega_m = 0.3089$, which was used for training.
While the difference for smaller value of A_s (Ω_m) is larger, the displacement for larger A_s (Ω_m) is more non-linear. This non-linearity is due to concentration of mass and makes the prediction more difficult.

where r_1=$|\boldsymbol{r_1}|$ and r_2=$|\boldsymbol{r_2}|$. Translation invariance guarantees that ζ is independent of $\boldsymbol{x}$. Rotational symmetry further eliminates all direction dependence except dependence on θ, the angle between $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. The multipole moments of $\zeta(r_1, r_2, \theta)$, $\zeta_\ell(r_1, r_2) = (2\ell + 1) \int d\theta P_\ell(\cos\theta)\zeta(r_1, r_2, \theta)$ where $P_\ell(\cos\theta)$ is the Legendre polynomial of degree ℓ, can be efficiently estimated with pair counting [34]. While the input (computed by ZA) do not contain significant correlations beyond the second order (power spectrum level), we expect D^{3}M to generate densities with a 3PCF that mimics that of ground truth.

We compare the 3PCF calculated from FastPM, 2LPT and D^{3}M by analyzing the 3PCF through its multipole moments $\zeta_\ell(r_1, r_2)$. Figure 2.3(b) shows the ratio of the binned multipole coefficients of the two 3PCF for several triangle configurations, $\bar{\xi}_\ell(r_1, r_2)_{pred}/\bar{\xi}_\ell(r_1, r_2)_{true}$, where $\bar{\xi}_\ell(r_1, r_2)_{pred}$ can be the 3PCF for D^{3}M or 2LPT and $\bar{\xi}_\ell(r_1, r_2)_{true}$ is the 3PCF for FastPM. We used 10 radial bins with $\Delta r = 5\,h^{-1}$Mpc. The results are averaged over 10 test simulations and the errorbars are the standard deviation. The ratio shows the 3PCF of D^{3}M is more close to FastPM than 2LPT with smaller errorbars. To further quantify our comparison, we calculate the relative 3PCF residual defined by

$$\begin{aligned} &\text{3PCF relative residual} \\ &= \frac{1}{9 \times N_r} \sum_{\ell=0}^{8} \sum_{r_1, r_2} \frac{|\zeta_\ell(r_1, r_2)_{\text{pred}} - \zeta_\ell(r_1, r_2)_{\text{true}}|}{|\zeta_\ell(r_1, r_2)_{\text{true}}|} \end{aligned} \tag{2.8}$$

where N_r is the number of (r_1,r_2) bins. The mean relative 3PCF residual of the D^{3}M and 2LPT predictions compared to FastPM are 0.79% and 7.82% respectively. The D^{3}M accuracy on 3PCF is also an order of magnitude better than 2LPT, which indicates that the D^{3}M is far better at capturing the non-Gaussian structure formation.

2.5 Generalizing to New Cosmological Parameters

So far, we train our model using a "single" choice of cosmological parameters $A_{\rm s} = 2.142 \times 10^{-9}$ (hereafter $A_0 = 2.142 \times 10^{-9}$) and $\Omega_m = 0.3089$ [119]. $A_{\rm s}$ is the primordial amplitude of the scalar perturbation from cosmic inflation, and Ω_m is the fraction of the total energy density that is matter at the present time, and we will call it matter density parameter for short. The true exact value of these parameters are unknown and different choices of these parameters change the large-scale structure of the Universe; see Figure 2.4.

Here, we report an interesting observation: the D^{3}M trained on a single set of parameters in conjunction with ZA (which depends on $A_{\rm s}$ and Ω_m) as input, can predict the structure formation for widely different choices of $A_{\rm s}$ and Ω_m. From a computational point of view this suggests a possibility of producing simulations for a diverse range of parameters, with minimal training data.

2.5.1 Varying Primordial Amplitude of Scalar Perturbations $\mathbf{A_{\rm s}}$

After training the D^{3}M using $A_{\rm s} = A_0$, we change $A_{\rm s}$ in the input of our test set by nearly one order of magnitude: $A_s = 1.8A_0$ and $A_{\rm s} = 0.2A_0$. Again, we use 1000 simulations for analysis of each test case. The average relative displacement error of D^{3}M remains less than 4% per voxel (compared to $< 3\%$ when train and test data have the same parameters). This is still well below the error for 2LPT, which has relative errors of 15.5% and 6.3% for larger and smaller values of A_s respectively.

Figure 2.5(a) shows the transfer function and correlation coefficient for both D^{3}M and 2LPT. The D^{3}M performs much better than 2LPT for $A_s = 1.8A_0$.

For small $A_s = 0.2A_0$, 2LPT does a better job than D^{3}M predicting the density transfer function and correlation coefficient at the largest scales, otherwise D^{3}M predictions are more accurate than 2LPT at scales larger than $k = 0.08\,h\text{Mpc}^{-1}$. We observe a similar trend with 3PCF analysis: the 3PCF of D^{3}M predictions are notably better than 2LPT ones for larger A_s, compared to smaller A_s where it is only slightly better. These results confirm our expectation that increasing A_s increases the non-linearity of the structure formation process. While 2LPT can predict fairly well in linear regimes, compared to D^{3}M its performance deteriorates with increased non-linearity. It is interesting to note that D^{3}M prediction maintains its advantage despite being trained on data from more linear regimes.

2.5.2 Varying matter density parameter $\Omega_\mathbf{m}$

We repeat the same experiments, this time changing Ω_m to 0.5 and 0.1, while the model is trained on $\Omega_m = 0.3089$, which is quite far from both of the test sets. For $\Omega_m = 0.5$ the relative residual displacement errors of the D^{3}M and 2LPT averaged over 1000 simulations are 3.8% and 15.2% and for $\Omega_m = 0.1$ they are 2.5% and 4.3%. Figures 2.5(c)(d) show the two-point statistics for density field predicted using different values of Ω_m. For $\Omega_m = 0.5$, the results show that the D^{3}M outperforms 2LPT at all scales, while for smaller $\Omega_m = 0.1$, D^{3}M outperforms 2LPT on smaller scales ($k > 0.1\,h\text{Mpc}^{-1}$). As for the 3PCF of simulations with different values of Ω_m, the mean relative 3PCF residual of the D^{3}M for $\Omega_m = 0.5$ and $\Omega_m = 0.1$ are 1.7% and 1.2% respectively and for 2LPT they are 7.6% and 1.7% respectively. The D^{3}M prediction performs better at $\Omega_m = 0.5$ than $\Omega_m = 0.1$. This is again because the Universe is much more non-linear at $\Omega_m = 0.5$ than $\Omega_m = 0.1$. The D^{3}M learns more non-linearity than is encoded in the formalism of 2LPT.

	point-wise	$T(k)$ $k = 0.11$ †	$r(k)$ $k = 0.11$	$T(k)$ $k = 0.50$	$r(k)$ $k = 0.50$	3PCF
test phase						
2 LPT Density	N/A	0.96	1.00	0.74	0.94	0.0782
D^3M Density	N/A	1.00	1.00	0.99	1.00	0.0079
2 LPT Displacement	0.093	0.96	1.00	1.04	0.90	N/A
D^3M Displacement	0.028	1.00	1.00	0.99	1.00	N/A
$\boldsymbol{A_s = 1.8A_0}$						
2LPT Density	N/A	0.93	1.00	0.49	0.78	0.243
D^3M Density	N/A	1.00	1.00	0.98	1.00	0.039
2LPT Displacement	0.155	0.97	1.00	1.07	0.73	N/A
D^3M Displacement	0.039	1.00	1.00	0.97	0.99	N/A
$\boldsymbol{A_s = 0.2A_0}$						
2LPT Density	N/A	0.99	1.00	0.98	0.99	0.024
D^3M Density	N/A	1.00	1.00	1.03	1.00	0.022
2LPT Displacement	0.063	0.99	1.00	0.95	0.98	N/A
D^3M Displacement	0.036	1.00	1.00	1.01	1.00	N/A
$\boldsymbol{\Omega_m = 0.5}$						
2LPT Density	N/A	0.94	1.00	0.58	0.87	0.076
D^3M Density	N/A	1.00	1.00	1.00	1.00	0.017
2LPT Displacement	0.152	0.97	1.00	1.10	0.80	N/A
D^3M Displacement	0.038	1.00	1.00	0.98	0.99	N/A
$\boldsymbol{\Omega_m = 0.1}$						
2LPT Density	N/A	0.97	1.00	0.96	0.99	0.017
D^3M Density	N/A	0.99	1.00	1.04	1.00	0.012
2LPT Displacement	0.043	0.97	1.00	0.97	0.98	N/A
D^3M Displacement	0.025	0.99	1.00	1.02	1.00	N/A

†The unit of k is $h\mathrm{Mpc}^{-1}$.

Table 2.1: A summary of our analysis.

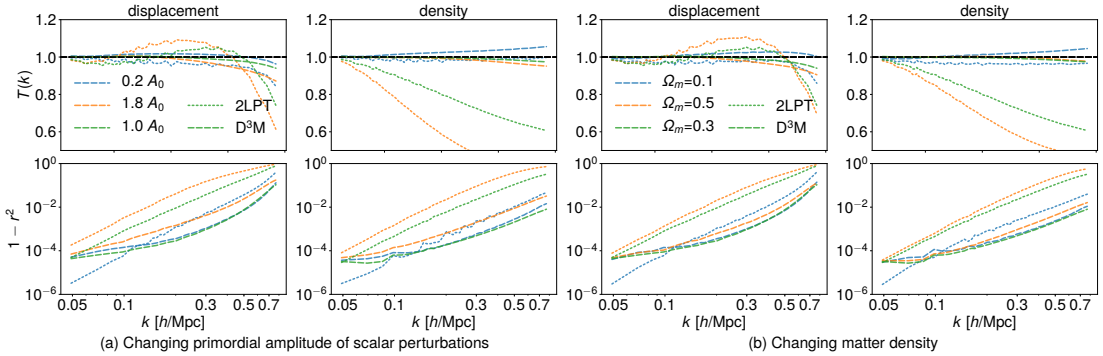


Figure 2.5: Similar plots as in Figure. 3(a), except we test the 2 point statistics when we vary the cosmological parameters without changing the training set (which has differnet cosmological parameters) nor the trained model. We show predictions from D^{3}M and 2LPT when tested on different (**a**) A_s and, (**b**) Ω_m. We show (top) the transfer function – i.e., the square root of the ratio of the predicted power-spectrum to the ground truth and (bottom) 1-r^2 where r is the correlation coefficient between the predicted fields and the true fields. D^{3}M prediction outperforms 2LPT prediction at all scales except in the largest scales as the perturbation theory works well in linear regime (large scales).

2.6 Conclusions

To summarize, our deep model D^{3}M can accurately predict the large-scale structure of the Universe as represented by FastPM simulations, at all scales as seen in the summary table in Table. 2.1. Furthermore, D^{3}M learns to predict cosmic structure in the non-linear regime more accurately than our benchmark model 2LPT. Finally, our model generalizes well to test simulations with cosmological parameters (A_s and Ω_m) significantly different from the training set. This suggests that our deep learning model can potentially be deployed for a ranges of simulations beyond the parameter space covered by the training data (Table 2.1). Our results demonstrate that the D^{3}M successfully learns the nonlinear mapping from first order perturbation theory to FastPM simulation beyond what higher order perturbation theories currently achieve.

Looking forward, we expect replacing FastPM with exact N-body simulations

would improve the performance of our method. As the complexity of our D^{3}M model is linear in the number of voxels, we expect to be able to further improve our results if we replace the FastPM simulations with higher resolution simulations. Our work suggests that deep learning is a practical and accurate alternative to the traditional way of generating approximate simulations of the structure formation of the Universe.

2.7 Materials and Methods

2.7.1 Dataset

The full simulation data consists of 10,000 simulations of boxes with ZA and FastPM as input-output pairs, with an effective volume of 20 $(\mathrm{Gpc/h})^3$ ($190 \times 10^9 \mathrm{ly}^3$), comparable to the volume of a large spectroscopic sky survey like DESI or EUCLID. We split the full simulation data set into 80%, 10% and 10% for training, validation and test, respectively. We also generated 1000 simulations for 2LPT for each set of tested cosmological parameters.

2.7.2 Model and Training

The D^{3}M adopts the U-Net architecture [108] with 15 convolution or deconvolution layers and approximately 8.4×10^6 trainable parameters. Our D^{3}M generalizes the standard U-Net architecture to work with three-dimensional data [120; 121; 122]. The details of the architecture are described in the following sections and a schematic figure of the architecture is shown in SI Appendix, Figure. 2.7. In the training phase, we employ the Adam Optimizer [123] with a learning rate of 0.0001, and first and second moment exponential decay rates equal to 0.9 and 0.999, respectively. We use the Mean-Squared Error as the loss

function (See Loss Function) and $L2$ regularization with regularization coefficient 0.0001.

2.7.2.1 Details of the D^{3}M Architecture

The contracting path follows the typical architecture of a convolution network. It consists of two blocks, each of which consists of two successive convolutions of stride 1 and a down-sampling convolution with stride 2. The convolution layers use 3×3×3 filters with a periodic padding of size 1 (see Padding and Periodic Boundary) on both sides of each dimension. Notice that at each of the two down-sampling steps, we double the number of feature channels. At the bottom of the D^{3}M, another two successive convolutions with stride 1 and the same periodic padding as above are applied. The expansive path of our D^{3}M is an inverted version of the contracting path of the network. (It includes two repeated applications of the expansion block, each of which consists of one up-sampling transposed convolution with stride 1/2 and two successive convolution of stride 1. The transposed convolution and the convolution are constructed with 3×3×3 filters.)

We take special care in the padding and cropping procedure to preserve the shifting and rotation symmetry in the up-sampling layer in expansive path. Before the transposed convolution we apply a periodic padding of length 1 on the right, down and back sides of the box (`padding=(0,1,0,1,0,1)` in pytorch), and after the transposed convolution, we discard one column on the left, up and front sides of the box and two columns on the right, down and back sides (`crop=(1,2,1,2,1,2)`).

A special feature of the D^{3}M is the concatenation procedure, where the up-sampling layer halves the feature channels and then concatenates them with the corresponding feature channels on the contracting path, doubling the number of

feature channels.

The expansive building block then follows a $1\times1\times1$ convolution without padding, which converts the 64 features to the the final 3-D displacement field. All convolutions in the network except the last one are followed by a rectified linear unit (ReLU) activation and batch normalization (BN).

2.7.2.2 Padding and Periodic Boundary

It is common to use constant or reflective padding in deep models for image processing. However, these approaches are not suitable for our setting. The physical model we are learning is constructed on a spatial volume with a periodic boundary condition. This is sometimes also referred to a torus geometry, where the boundaries of the simulation box are topologically connected – that is $x_{i+L} = x_i$ where i is the index of the spatial location, and L is the periodicity (size of box). Constant or reflective padding strategies break the connection between the physically nearby points separated across the box, which not only loses information but also introduces noise during the convolution, further aggravated with an increased number of layers.

We find that the periodic padding strategy significantly improves the performance and expedites the convergence of our model, comparing to the same network using a constant padding strategy. This is not surprising, as one expects it is easier to train a model that can explain the data than to train a model that does not.

2.7.2.3 Loss Function

We train the $\mathrm{D^3M}$ to minimize the mean square error on particle displacements

$$\mathcal{L} = \frac{1}{N}\sum_i (\hat{\mathbf{\Psi}}_i - \mathbf{\Psi}_{t,i})^2, \tag{2.9}$$

where i labels the particles and the N is the total number of particles. This loss function is proportional to the integrated squared error, and using a Fourier transform and Parsevals theorem it can be rewritten as

$$\int (\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_t)^2 \mathrm{d}^3\boldsymbol{q} = \int \left|\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_t\right|^2 \mathrm{d}^3\boldsymbol{k} = \int \mathrm{d}^3\boldsymbol{k} \left(\left|\boldsymbol{\Psi}_t\right|^2 (1-T)^2 + 2\left|\hat{\boldsymbol{\Psi}}\right|\left|\boldsymbol{\Psi}_t\right|(1-r) \right) \tag{2.10}$$

where $\boldsymbol{q}$ is the Lagrangian space position, and $\boldsymbol{k}$ its corresponding wavevector. T is the transfer function defined in Eq. 2.5, and r is the correlation coefficient defined in Eq. 2.6, which characterize the similarity between the predicted and the true fields, in amplitude and phase respectively. Eq. 2.10 shows that our simple loss function jointly captures both of these measures: as T and r approach 1, the loss function approaches 0.

2.8 appendix

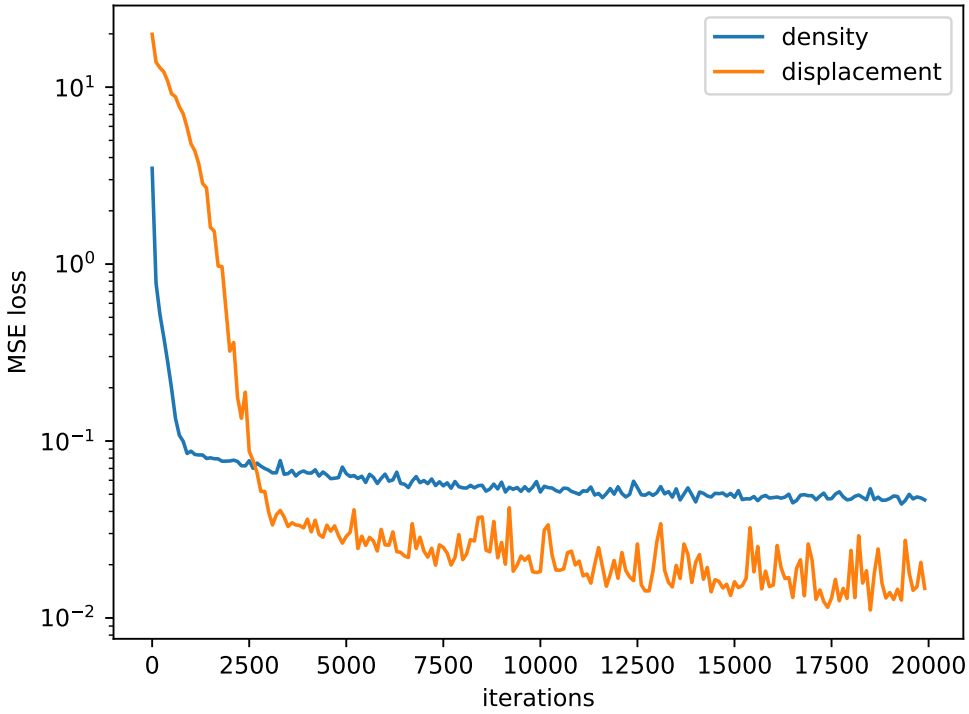


Figure 2.6: Same deep learning architecture with either density field or displacement field as input. The loss with displacement field as input is smaller, and this proves that displacement field contains information that is non-local in the density, which helps model to learn.

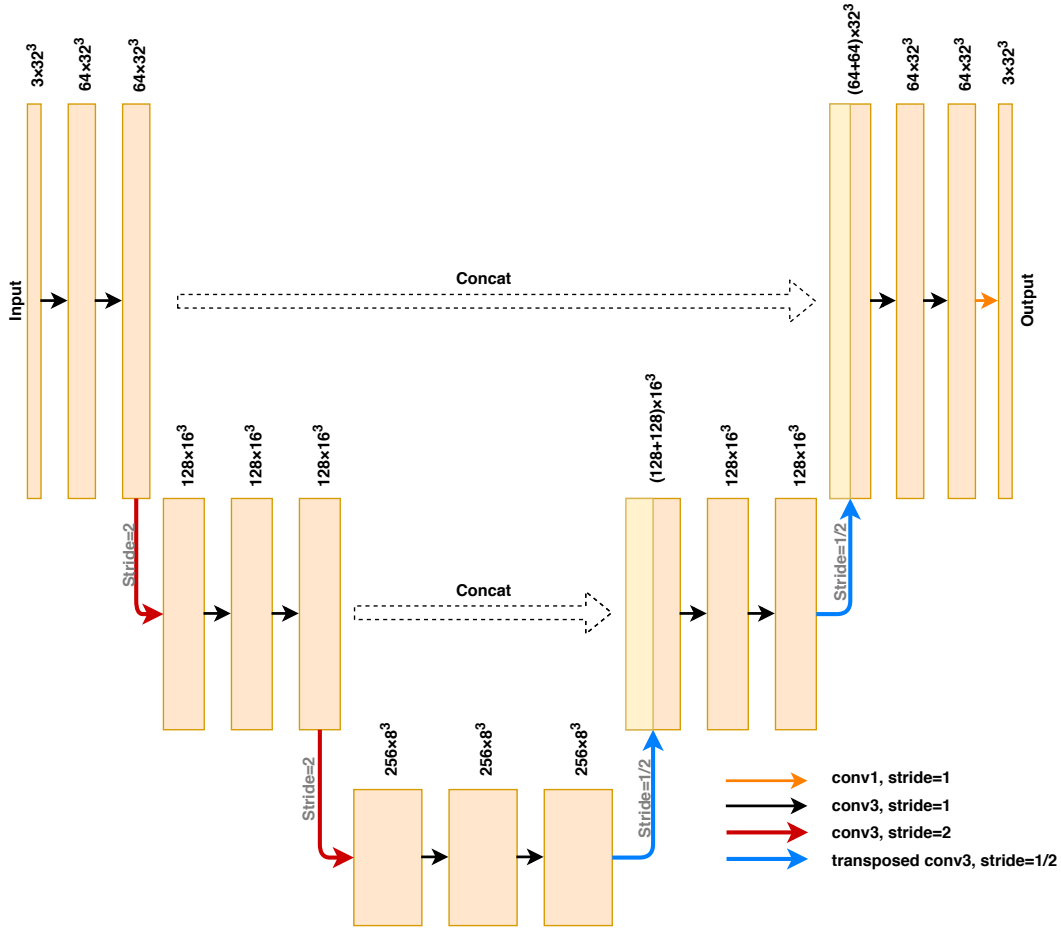


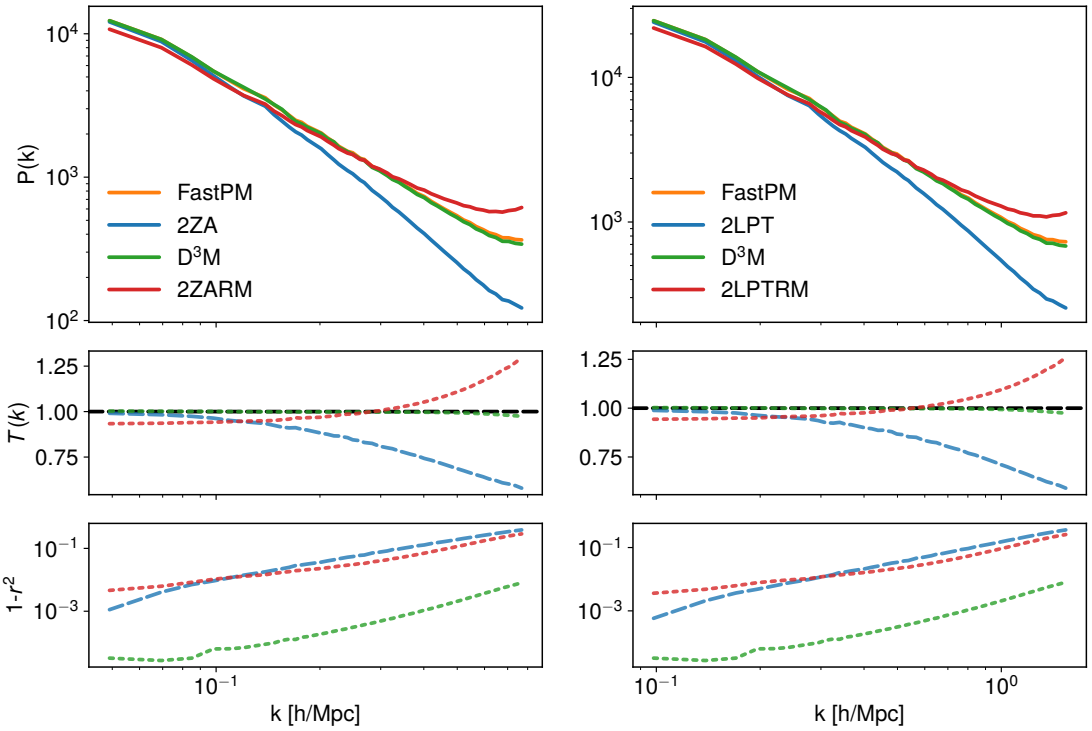
Figure 2.7: Each orange box corresponds to a multi-channel feature map. The number of channels and the output size of each channel are denoted on top of the box. Grey dashed boxes represent copied feature maps. The arrows denote the different operations. The black, red and blue arrays represent 3×3 convolution with stride one, 3×3 convolution with stride two, 3×3 transposed convolution with stride $1/2$ respectively with periodic padding followed by ReLU activation and batch normalization. The orange array represents 1×1 convolution with stride one.

2.9 Comparison with One-point remapping method

In addition, besides the benchmark method 2LPT, we have also compared our result with another state-of-the-art method: a non-linear remapping method [124]. The idea of this method is remapping the ZA/2LPT density to match the smoothed N-body density (N-body density smoothed by ZA/2LPT transfer function) probability density function, and then applying a reciprocal of the transfer function of the remapping field to arrive at the final estimation of the N-body simulation. All the computations for the one-point remapping method are done after binning the density fields with a Cloud in Cell (CiC) method with mesh size 4 Mpc/h. The transfer function and the remapping function are learned from 8000 simulations – the same simulations used for training of D^3M. And the remapping method is tested on 1000 simulations – the same simulations used for testing of D^3M. In Fig. 2.8, we use the power spectrum to compare our result and the remapping method. And it shows our result outperforms the remapping method in both transfer function and correlation coefficient. In addition, our method allows the prediction of displacement for each particle, so we predict final positions of all the particles.

2.10 Test Time Comparison

For the D^3M model, It takes 20s to generate 1000 simulations with a single gpu (Tesla V100-SXM2). And for FastPM, it takes 115s to generate 1000 simulations on a single Edison Haswell node, with 24 cores. The FastPM code is also fully optimized. This suggests the Deep Learning model is faster than the fast N-Body estimator.



(a) One-point remapping from ZA

(b) One-point remapping from 2LPT

Figure 2.8: From top to bottom: (top) density power-spectrum of FastPM (orange), 2LPT (blue), D^{3}M (green) and ZARM/2LPTRM (red); (middle) transfer function – i.e., the square root of the ratio of the predicted power-spectrum and the ground truth; (bottom) 1-r^2 where r is the correlation coefficient between the predicted fields and the true fields. Results are the average values of 1000 test simulations. It shows D^{3}M result outperforms the remapping method in both transfer function and correlation coefficient.

Chapter 3

Beyond the D^3M Model

In the last chapter, we have shown D^3M model to perform well in predicting the non-linear structure formation. However, there are limitations to the D^3M model. Firstly, the deep model is opaque – we don't know what does the black box learns. Secondly, the D^3M model is one kind of convolutional neural networks, which can only deal with the regular data format, such as images or 3D voxels. In this chapter, we strive to address these two challenges.

3.1 Interrogation of the D^3M model with physics analysis

The deep learning model is usually known as a black box. Specifically, in our problem, we don't know if the model is learning the right representations that are aligned with physics. However, the good part of the physics data is that it is interpretable, and we could use our knowledge to interrogate the model. In this section, we strive to probe the behavior of the trained D^3M with our known knowledge. The purpose of these tests is to gain some insights that our D^3M model is learning the underlying representations that are aligned with physics.

Firstly, we put the deep learning model under specific scenarios that are physically motivated and benchmark the model performance by comparing model prediction against the correct output data (from FastPM solutions). We consider three physically motivated scenarios and one machine learning motivated scenario: zero residual test, Zel'dovich Pancakes, and two-mode coupling test. We input the trained model with test data that is drawn from a very different distribution than the training data, and the test data is analytically interpretable. Secondly, we apply a popular machine learning visualization method to understand where does the most critical information comes from.

3.1.1 Zero Residual Test

We first check the residual output of the trained D^{3}M model by feeding the model with all zeros as the input displacement field. In doing so, we test if the D^{3}M model learns the physical symmetries of homogeneity and isotropy inherented in the training simulations. Given the uniform zero initial condition, we expect D^{3}M prediction to be homogeneous, isotropic, and close to zero in preserving the symmetries. Fig. 3.1 shows the output displacement field of the model Ψ_i with $i = \{\mathrm{x}, \mathrm{y}, \mathrm{z}\}$ averaged along the i-th axis. Periodic patterns show up as results of the 3×3×3 filters of the up-sampling steps in the expansive path. There are very small biases in x, y, and z directions: 0.07 ± 0.007, -0.03 ± 0.014 and 0.09 ± 0.009, respectively. Arising from the noise in the trained neural network, these biases are negligible compared to either the standard deviation of displacement in the simulation (7.7 Mpc/h) or the pixel size (4 Mpc/h), and thus have minimal impact on the tests below. Overall, the residual test agrees with our expectation.

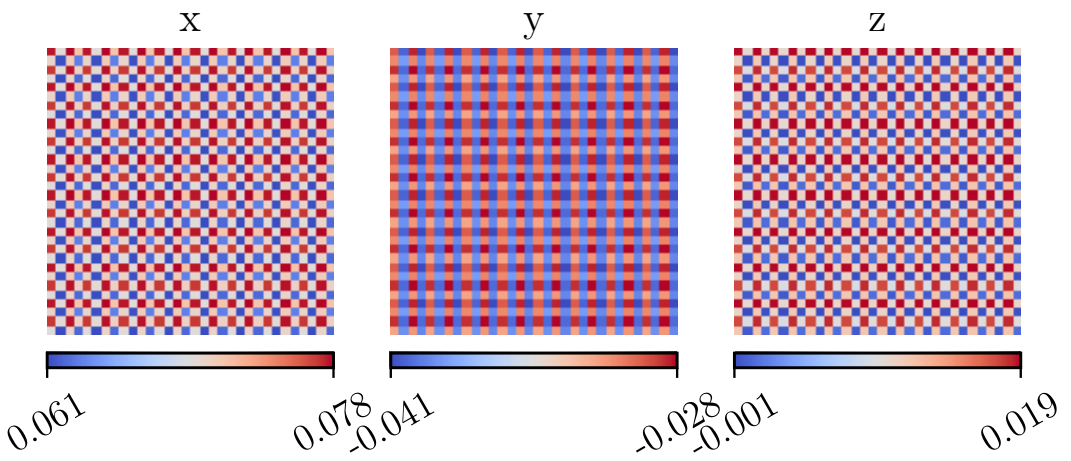


Figure 3.1: Residual Test. We feed the trained D^{3}M model with all zeros in the input displacement fields. Periodic patterns show up in the D^{3}M model prediction due to the 3×3×3 filters. There is also a small bias in the x,y,z directions which might come from the noise of the D^{3}M model. But the D^{3}M model prediction is close to 0 and preserves homogeneity and isotropy with great significance as expected.

3.1.2 Pancake Test

The training data have the same cosmological parameters with different random seeds, which means the power spectrum of training data is identical. From the perspective of Fourier space, the training data is a combination of the information from all scales. As a simple analysis, we want to check if the D^{3}M model is predicting well for information on a single scale, though the pattern of a single scale is not given in the training phase. To achieve this, we decompose the input into single modes (where the combination of all the orthogonal modes would form the full power spectrum) and feed each single mode to the trained D^{3}M model, that is, to test if the deep learning model is grasping the evolution of a planar perturbation. This test is well known in physics as Zel'dovich pancake test [125]. Mathematically, the input modes are of the following form

$$\vec{\Psi}(\hat{x}) = A_i \hat{k}_i cos(\vec{k}_i \cdot \vec{x} + \phi) \tag{3.1}$$

where $\vec{\Psi}(\hat{x})$ is the 3D displacement field, A_i is the amplitude of the single mode power which is of the same order of magnitude as the power spectrum at that scale, k is the scale factor, and ϕ is the phase term. Ideally, we would expect that if the D^{3}M model is learning the correct representation that is aligned with physics, it should predict precisely the evolution of a single mode.

Fig. 3.2 shows the result of pancake analysis. In all subplots, the red triangle and the dashed line shows the scale and the amplitude of the single input mode, which serves as one mode Zel'dovich approximation. The green crosses and orange stars are the results of FastPM simulation and D^{3}M model prediction, respectively. The transfer function shows that the D^{3}M model correctly captures the scales of the FastPM modes but with a slight wrong estimation of the amplitude of the modes. The D^{3}M model predicts perfectly at the dominant modes (where the input mode is) indicating that the D^{3}M model can capture information from different scales. The rest of the modes, whose amplitudes are at least two orders smaller than the dominant mode, are indeed the numerical artifacts of the FastPM (numerical noise from force resolution, cic window function, etc). But since FastPM simulations are the targets of the D^{3}M model so that it is expected that the D^{3}M model captures the numerical artifacts of the FastPM.

3.1.3 Two-Mode Coupling Test

Besides pancake test where a single mode is fed to the trained model, we also test the trained model with two perpendicular fundamental modes as input. During the evolution, the two waves will interact with each other and result in a coupling mode on smaller scales. This is to test if the D^{3}M model can also capture well the interaction between modes.

The two input perpendicular plane waves we use have the same amplitude of

wave number $k = 2\pi/128\ h/\mathrm{Mpc}$. Both of the modes are in linear regimes and can be well solved by the Lagrangian theory. We also rotate the two plane waves to test if the model could preserve the rotation symmetry.

Fig. 3.3 shows the result of two-mode coupling test, where the amplitude of the displacement in the Fourier space is calculated. The grey rhombus, blue cross, green plus, and orange dot show the initial input wave mode, FastPM prediction, D^{3}M prediction, and Lagrangian prediction respectively. The dominant mode is the input and the mode with the second largest amplitude is the result of mode coupling. The agreement among the D^{3}M prediction and Lagrangian theory is pretty good in the first two modes. The remaining noisy modes are the artifacts of the FastPM simulations. The noisy modes predicted by D^{3}M is also consistent with that of FastPM – the amplitudes of D^{3}M prediction are comparable to that of the FastPM. By rotating the two input perpendicular modes, the D^{3}M model predicts correctly at the dominate and coupling mode, whereas the noisy modes predicted vary slightly, which suggests the model has captured the rotation symmetry.

3.1.4 Saliency Map

The saliency map is first proposed as a visualization method for the computer vision classification task, and the purpose of the saliency map is to represent saliency at every location by a scalar quantity, highlighting the areas of the given image, discriminative with respect to the given class [126]. The idea is pretty simple. It computes the gradient of output to the input image:

$$\frac{\partial\ \text{output}}{\partial\ \text{input}} \tag{3.2}$$

And the result should tell us how output value changes with respect to a small change in input image pixels.

We adopt the idea of the saliency map and define the saliency map for our problem as

$$w_{i,j} = \sqrt{\left\| \frac{\partial \mathbf{\Psi}_{\mathrm{D^3M},i}}{\partial \mathbf{\Psi}_{\mathrm{ZA},j}} \right\|_2} \tag{3.3}$$

where i and j represents the i-th and j-th grids in the output and input respectively. The derivatives are calculated over the whole input map. By calculating the saliency map, we would know for each of the output grid, what are the areas that contribute the most to correct ZA to the right non-linear structure represented by FastPM.

Fig. 3.4 shows 2D slices of the saliency maps, with the first row showing the input ZA which is fed into the D^{3}M model, the second row showing the D^{3}M prediction and the third row showing the saliency map. The blue arrow points to where in the output map that the saliency map is calculated. Fig. 3.4 indicates most of the correcting effects for ZA come from local areas, which is expected because ZA is correct on large scales, and the corrections for ZA should mostly come from small scale local regions The first plot in Fig. 3.4 also shows that the D^{3}M model correctly captures the periodic boundary condition of the simulations.

3.1.5 Conclusion

In this chapter, we have interrogated the D^{3}M model with physics questions to check if the model is learning some representations that are aligned with physics. To achieve this, we do three physics analysis accordingly, which includes residual test, pancake test, and two-mode coupling test and we have proved that the D^{3}M model has learned the physical symmetries of homogeneity and isotropy,

the information of different scales, and the interaction between different modes, respectively. Besides the physics checks, we have also applied the saliency map method. The saliency map indicates that the most critical information to correct the ZA for each particle comes from local regions.

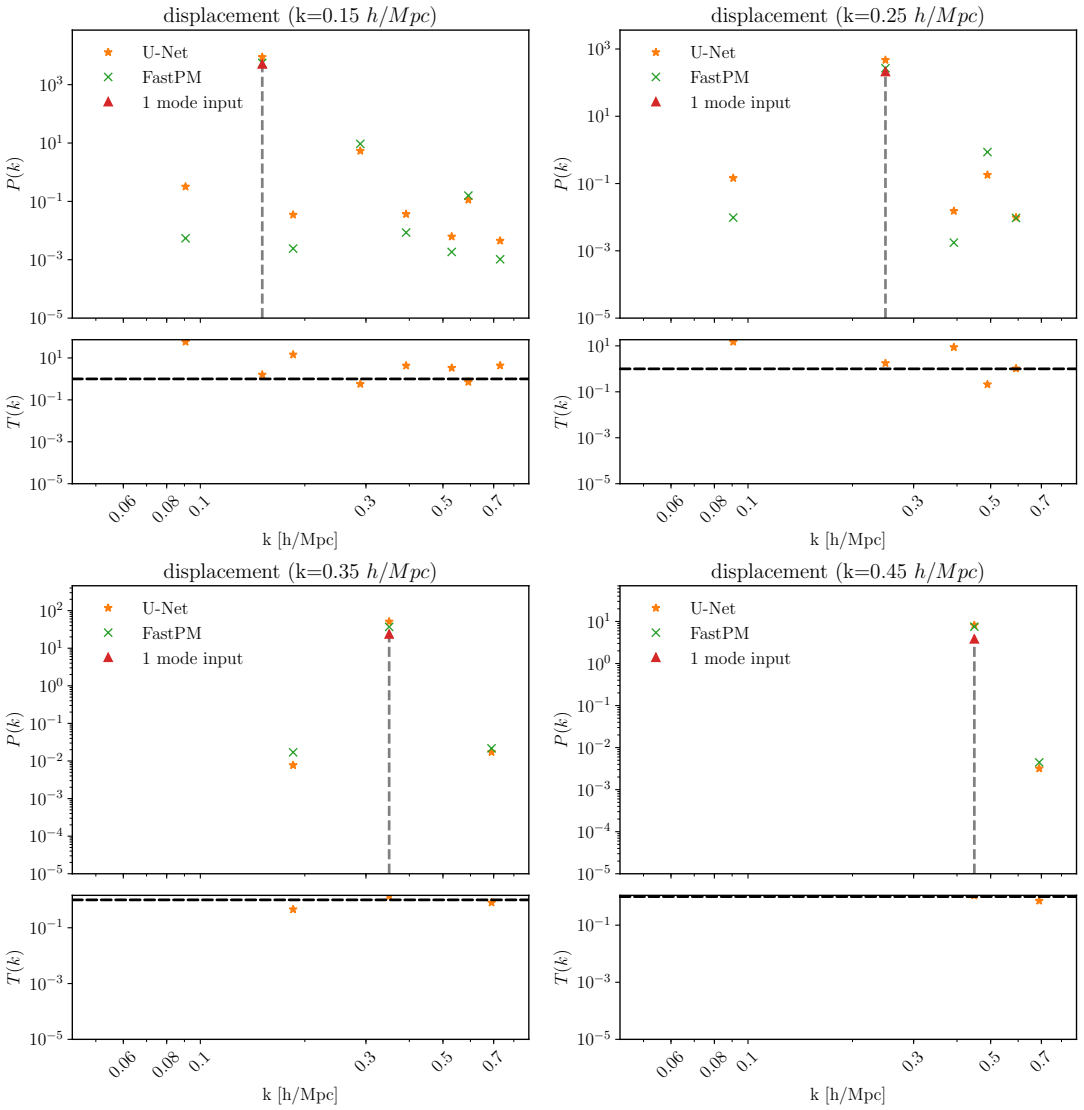


Figure 3.2: Pancake test. We decompose the input into single modes in Fourier space and feed each mode (planar perturbation as shown in eq. 3.1) to the trained D^{3}M model. The red triangles show the power spectrum for the input mode while the green crosses and orange stars show the power spectrum for FastPM and D^{3}M model predictions. The transfer function shows that the D^{3}M model captures quite well at the dominant scale of FastPM, which indicates the D^{3}M model is able to capture scale information. The D^{3}M model also captures the other modes of FastPM that are two orders smaller than the dominant mode and come from the numerical artifact of FastPM simulations.

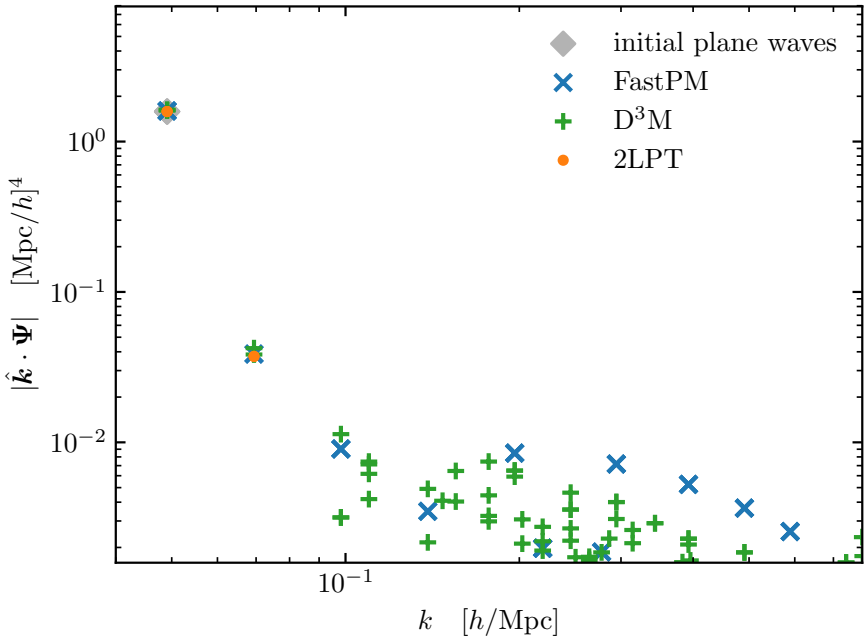


Figure 3.3: Two-mode coupling check. We compare between FastPM, D^{3}M, and theory with 2 perpendicular fundamental modes as input. The agreement among the D^{3}M prediction, FastPM prediction and Lagrangian theory is pretty good on the first two modes. The remaining noisy modes are the artifacts of the FastPM simulations. And the D^{3}M predictions are also consistent with FastPM on these noisy modes.

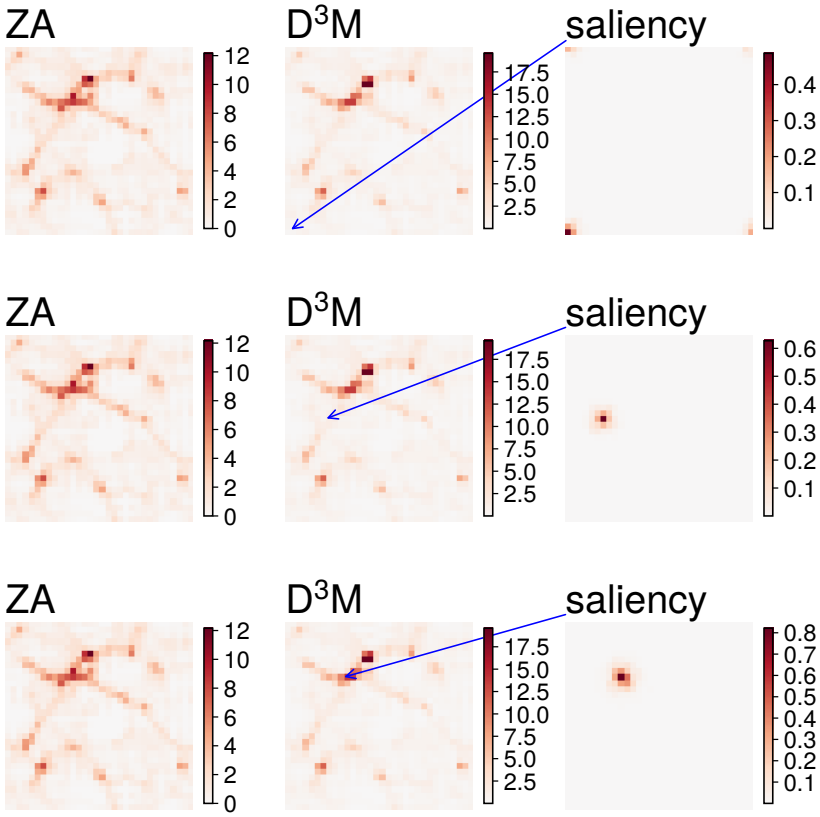


Figure 3.4: 2D slices of Saliency Map. The first row shows the input ZA which is fed into the D^{3}M model, the second row shows the D^{3}M prediction, and the third row shows the saliency map. The blue arrow points to where in the output map that the saliency map is calculated. The saliency map indicates the most crucial information to correct ZA for each particle comes from local regions. This is expected because ZA is correct on large scales, and most of the non-linear information should mostly come from small scale local regions. And the first plot also shows that the D^{3}M model correctly captures the periodic boundary condition of the simulations.

3.2 Multi-Scale Deep Sets to Learn Non-Linear Evolution of the Universe

In the previous chapter, the D^{3}M model, belonging to the big category of the convolutional neural networks, is shown to be a powerful tool that can efficiently learn complex functions. Typical convolutional neural networks require highly regular data format, which can be represented as fixed dimensional image grids or 3D voxels. However, many astronomical and cosmological data are represented as point clouds – set of data points in space. For instance, dark matter and galaxies are represented as a set of points in simulations. To apply convolutional neural networks on these datasets, one needs to convert the point-cloud data to voxel or mesh representation as a pre-processing step, resulting in smearing out information smaller than the scale of the voxel or mesh size [127; 128]. Besides, some point-cloud data are very sparse, making it hard for the convolutional neural network to extract useful information [129]. So the direct application of deep learning methods to point-cloud is highly desirable.

Recent progress in development of point-cloud models has made it possible to directly work on the point-cloud data [67; 68; 130; 131]. In this chapter, we design a multi-scale deep sets model – a point cloud model based on the deep sets model [68], to predict the non-linear structure formation of the Universe, the same task as in Chapter 2. Our objective is to show that the point-cloud model is another efficient deep model that can be used in solving complex cosmological problems with cosmological data represented as point clouds.

3.2.1 Related Work

The multi-scale deep sets model is built on top of deep sets [68]. Deep sets is a pioneering work in dealing with sets as inputs. It is developed based on powerful

permutation invariance and equivariance properties of sets, along with the theory to support its performance.

The permutation invariance deep sets model transforms each instance x_m into some representations $\phi(x_m)$. Then the representations are added up, and the output is processed by another ρ network ($\rho(\sum \phi(x_m))$), where ρ network can be any operation in the deep network, such as the fully connected layer, non-linear activation.

The requirement to preserve the permutation equivaraince in each layer $f_\Theta((\boldsymbol{x})) = \delta(\Theta(\boldsymbol{x}))$ is that the off-diagonal elements of Θ are tied together and all the diagonal elements are equal as well. Thus each layer of the permutation equivaraince deep sets model can be briefly summarized by eq. 3.4

$$f(x) = \sigma(\lambda \mathfrak{I}\boldsymbol{x} + \gamma \text{maxpool}(\boldsymbol{x})\mathbf{1}) \tag{3.4}$$

where σ is the non-linear activation, λ and γ are learnable weights.

Extensive experiments across several domains have shown that deep sets model has a strong generalization ability, and it outperforms other intuitive deep networks, which are not backed by theory.

3.2.2 Data

The data we use is the same as in chapter 2 (see Setup in Chapter 2). The only difference is that in chapter 2, the input and target data (ZA, FastPM respectively) is represented as displacement field $\boldsymbol{\Psi} \in \mathbb{R}^{d\times d\times d\times 3}$. But for our multi-scale deep set model, the data is represented as vectors $\boldsymbol{x} \in \mathbb{R}^{N\times C}$, where N is the number of particles and C is the number of features for each particle.

More specifically, we have 10,000 pairs of ZA approximations as input and accurate FastPM as the target. Each of the simulations has 32^3 N-body particles

in a volume of 128 Mpc/h. The input of our model are the positions and the ZA displacements of the particles (input $\in \mathbb{R}^{32^3 \times 6}$), and the target data are the FastPM displacements (target data$\in \mathbb{R}^{32^3 \times 3}$).

3.2.3 Method

By direct applying the deep sets model, we are unable to achieve a decent prediction comparable to the D^3M model in Chapter 2. To achieve better performance, we need a huge architecture because the deep sets model needs to learn the space – it needs to learn the interaction between particles based on distances, which is hard for the deep sets model especially with a large number of points.

However, we know that the interaction between particles depends on distance – particles far away from a particle has less influence on that particle than the particles near to that particle (scale as $1/r^2$ rule where r is the distance between two particles). Thus we develop the multi-scale deep sets model to help the model with more information on distances.

The basic idea multi-scale deep sets model is instead of calculating the mean-pooling features across the whole points, we calculate the mean-pooing feature for each point as a function of distance.

To achieve this, in each layer of the multi-scale deep sets model, we divide the simulation box into k^3 sub-boxes. Each particle is assigned a primary box. At each resolution level we are going to expand this box with neighboring boxes (e.g., 1 $\rightarrow$ 1+s $\rightarrow$ 1+2s $\rightarrow$ .. $\rightarrow$ k) where s is the step size. The features for each box are mean-pooling over all particle-features in that box. We do this for boxes at all resolution levels. The output feature for each point is calculated based on its input features, the features of the primary box and features for all boxes in which the primary box appears at the center.

However, implementation can be done more efficiently using convolution: In

order the calculate the features for each box at level h, we perform convolution with a choice of filter size that decides our number of levels. The convolution filters are fixed to identity to perform sum-pooling. Besides, we have an identity channel per particle that allows for counting of particles per level. Note that primary boxes will be double-counted at higher levels, but as long as we divide the features by the value we get for the identity channel the normalization is fine. The model can adjust the weights to account for double-count. In summary, we apply several convolution layers so that the final layer can see exactly all the input boxes. Then we calculate the output features for each particle by assigning a weight to each level as well as the features of the input particle.

3.2.4 Result

We use the same 2-Point Correlation statistics as in Chapter 2 (See 2-Point Correlation Comparison in Chapter 2) to evaluate our result. Fig. 3.5 shows the average power spectrum, transfer function T(k) and stochasticity $1 - r^2$ of the displacement field and the density field over 1000 simulations. For the Multi-Scale deep sets model, both the density and displacement transfer functions differ from 1 by a 1% at scale k $\sim$ 0.2 h/Mpc, and this discrepancy increases to 14% and 5% for density field and displacement field respectively, as k increases to the Nyquist frequency around 0.7 h/Mpc. The stochasticity hovers between 10^4 and 10^2 for most scales. But for the displacement field, the stochasticity goes up to 10^{-1} at scale k $>$ 0.62 h/Mpc. The performance of the multi-scale point cloud model is consistently better than the benchmark 2LPT. Though the multi-scale point cloud model is worse than the D^3M model, it is not surprising, because the problem is harder for the multi-scale point cloud model since it needs to learn the space information but for the D^3M model, while the same information is given as known information.

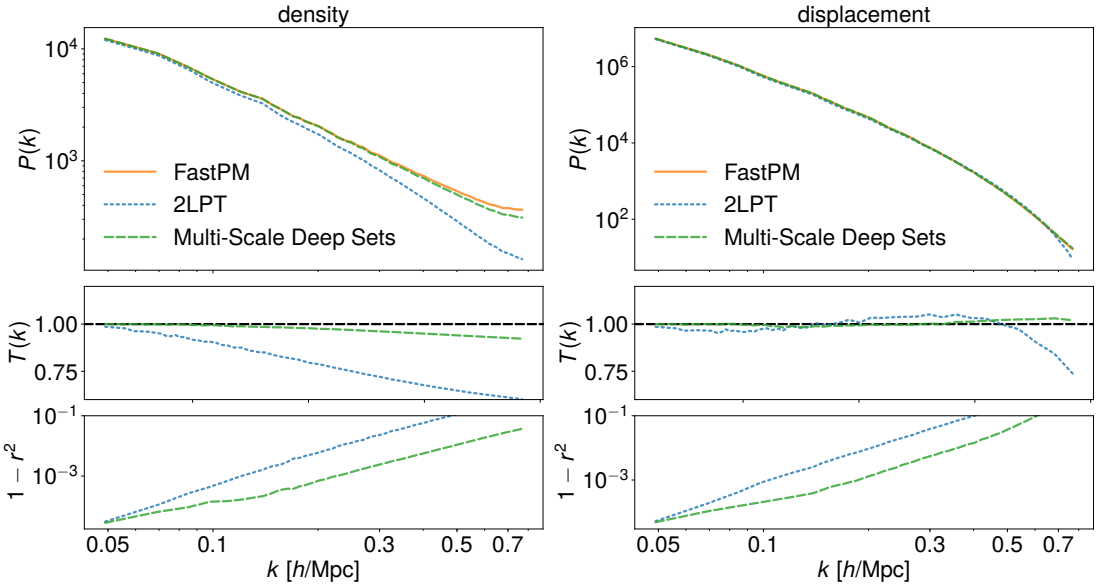


Figure 3.5: From top to bottom: (top) displacement and density power-spectrum of FastPM (orange), 2LPT (blue), and multi-Scale deep sets (green); (middle) transfer function – i.e., the square root of the ratio of the predicted power-spectrum to the ground truth; (bottom) 1-r^2 where r is the correlation coefficient between the predicted fields and the true fields. Results are the averaged values of 1000 test simulations. The results have shown that the multi-scale deep sets model outperforms the 2LPT.

3.2.5 Conclusion

In this chapter, we have developed a multi-scale deep sets model to learn the non-linear structure formation of the Universe. We have used the two-point statistics to evaluate our results. While the multi-scale deep sets model outperforms the benchmark 2LPT at all scales, its performance is worse than the D^{3}M model. This is within expectation because the multi-scale deep sets model needs to learn the extra space information from the positions of the particles, which is not trivial work. However, given the pretty good performance the multi-scale deep sets model achieves, we believe that the multi-scale deep sets model can be applied to other cosmological problems where standard convolution neural network is not feasible.

Chapter 4

The detection of the imprint of filaments on Cosmic Microwave Background (CMB) lensing

4.1 Introduction

Galaxy redshift surveys, such as 2dF [4], SDSS [5], 6df [6], GAMA [7] and VIPERS [8], have shown that the spatial distribution of matter forms a rich web, known as the cosmic web [2]. The majority of galaxy survey analyses measure the amplitude of galaxy clustering as a function of scale, ignoring information beyond a small number of summary statistics. Since the matter density field becomes highly non-Gaussian as structure evolves under gravity, we expect other statistical descriptions of the field to provide us with additional information. One way to study the non-Gaussianity is to study filaments, which evolve non-linearly from the initial density fluctuations produced in the primordial Universe. In our study, we report the first detection of CMB (Cosmic Microwave Background) lensing by filaments and we apply a null test to confirm our detection. Further-

more, we propose a phenomenological model to interpret the detected signal and we measure how filaments trace the matter distribution on large scales through filament bias, which we measure to be around 1.5. Our study provides a new scope to understand the environmental dependence of galaxy formation. In the future, the joint analysis of lensing and Sunyaev-Zel'dovich observations might reveal the properties of "missing baryons", the vast majority of the gas which resides in the intergalactic medium and has so far evaded most observations.

4.2 Main Result

The cross-correlations of CMB lensing with tracers of large-scale structure have been widely studied[132; 133; 134; 135; 136; 137; 138; 139; 140; 141; 142]. In our study, we detect the imprint of filaments on CMB lensing by cross-correlating filaments with the CMB lensing convergence (κ) map. We use the filament intensity map, which is derived from the Cosmic Web Reconstruction filament catalogue [143] (Public in `https://sites.google.com/site/yenchicr/catalogue`) from the Sloan Digital Sky Survey (SDSS) [5] Baryon Oscillations Spectroscopic Survey (BOSS) [113] Data Release 12 (DR 12) [144]. The filament finder (See Filament Finder in the Method section) partitions the universe from $z = 0.005$ to $z = 0.700$ into slices with $\Delta z = 0.005$. In our study, we use the filaments from $z = 0.450$ to z =0.700, which are detected from CMASS galaxy survey (a galaxy sample from SDSS which targets high redshift). Filaments are found in each redshift bin as the density ridge of the smoothed galaxy density field [145] and the filament uncertainty, which describes the uncertainty of filament position, is also calculated (see Uncertainty of Filaments in Method section). The filament

intensity, illustrated in Fig. 4.4, is defined as

$$I(\hat{n}, z) = \frac{1}{\sqrt{2\pi\rho_f(\hat{n}, z)^2}} \exp\left(-\frac{\|\hat{n} - \hat{\Pi}_f(\hat{n}, z)\|^2}{2\rho_f(\hat{n}, z)^2}\right) \tag{4.1}$$

where $\hat{n}$ is the angular position, $\hat{\Pi}_f(\hat{n}, z)$ is the angular position of the closest point to $\hat{n}$ on the nearest filament and $\rho_f(\hat{n}, z)$ is the uncertainty of the filament at the projected position $\hat{\Pi}_f(\hat{n}, z)$. Using the intensity map at each redshift bin, we construct the filament intensity overdensity map via

$$\delta_f(\hat{n}) = \frac{\int I(\hat{n}, z)dz - \bar{I}}{\bar{I}}, \quad \bar{I} = \frac{\int I(\hat{n}, z)\ d\Omega_{\hat{n}}dz}{\int d\Omega_{\hat{n}}} \tag{4.2}$$

In this work we use the CMB lensing convergence map (Public in) from the Planck [146] satellite experiment. The Planck mission has reconstructed the lensing potential of the CMB from a foreground-cleaned map synthesized from the Planck 2015 full-mission frequency maps using the SMICA code [147]. The lensing convergence κ is defined in terms of the lensing potential ϕ as

$$\kappa(\hat{n}) = \frac{1}{2}\nabla^2_{\hat{n}}\phi(\hat{n}) \tag{4.3}$$

We measure the cross angular power spectrum of CMB lensing convergence and filaments $C_l^{\kappa f}$ using standard techniques (See Estimator in Method section). We compute the error for each power spectrum by jackknife resampling the observed area into 77 equally weighted regions (see Supplementary Information and Fig. 4.5) that comprise the CMASS galaxy survey from where the filaments are detected.

We construct a phenomenological model to describe the cross-correlation of filaments and the CMB lensing convergence field. Instead of modeling the filament

profile on small scales[148; 149; 150], our model studies how filaments trace matter distribution on large scales through the use of the filament bias. We assume a ΛCDM cosmology with Planck parameters from the 2013 release [151], where $\Omega_m = 0.315, h = 0.673, \sigma_8 = 0.829, n_s = 0.9603$. In a spatially flat Friedmann-Robertson-Walker universe described by general relativity, the convergence field is

$$\kappa(\hat{n}) = 4\pi G_N \bar{\rho}_0 \int_0^{\chi_{CMB}} \frac{\chi(\chi_{CMB} - \chi)}{\chi_{CMB}} (1+z)\delta_m(\chi, \hat{n}) d\chi \tag{4.4}$$

where χ is the comoving radial distance, z is the redshift observed at radial distance, G_N is Newton's gravitational constant, $\bar{\rho}_0$ is the present-day mean density of the universe, and χ_{CMB} is the comoving distance to the CMB. On linear scale, we assume that filaments trace the matter as $\delta_f = b_f \delta_m$, where b_f is defined as the large-scale filament bias.

On large scale, we expect the filament overdensity δ_f to be related to the matter fluctuations through a linear filament bias b_f:

$$\delta_f(\hat{n}) = \int b_f f(z) \delta_m(\hat{n}, z) dz \tag{4.5}$$

where $f(z)$ is the mean filament intensity redshift distribution defined as

$$f(z) = \frac{F(z)}{\int F(z) dz}, \quad F(z) = \frac{1}{\Delta z} \int I(\hat{n}, z) d\Omega_{\hat{n}} \tag{4.6}$$

where $I(\hat{n}, z)$ is the filament intensity defined in eq. (4.1) and Δz is the width of redshift slice. In cross-correlation, on scales smaller than the typical filament length, using filaments introduces additional smoothing compared to the true matter density. We model the smoothing as follows: the filaments have typical length and we lose all small-scale information about fluctuations along the filament; therefore, we take the corresponding filament power spectrum to be

exponentially suppressed below the filament scale $k_{\parallel} \sim 1/(\text{filament length})$ in Fourier space. Similarly, any matter in between filaments is either assigned to a filament or eliminated from the catalog (in underdense regions) . For this reason we also introduce a suppression in power in the direction perpendicular to the filaments, with suppression scale $k_{\perp}$. We use two ways to model $k_{\perp}$. The detailed models are shown later in the paper. Using the Limber approximation [152] and the smoothing scale for small scales, the filament-convergence cross-correlation can be written as

$$C_l^{\kappa f} = \frac{3H_0^2\Omega_{m,0}}{2c^2} \int_{z_1}^{z_2} dz W(z) f(z) \chi^{-2}(z)(1+z) P_{mf}\left(\frac{l}{\chi(z)}, z\right) \tag{4.7}$$

where $W(z) = \chi(z)(1 - \frac{\chi(z)}{\chi_{CMB}})$ is the CMB lensing kernel, and P_{mf} is modeled as

$$P_{mf}(k,z) = \frac{1}{2\pi} \int d\phi \; b_f P_{mm}(k,z) e^{-(k\cos(\phi)/k_{\perp}(z))^2 - (k\sin(\phi)/k_{\parallel}(z))^2} \tag{4.8}$$

where P_{mm} is the matter power spectrum. We use CAMB (Code for Anisotropies in the Microwave Background) (`http://www.camb.info/`) to evaluate the theoretical matter power spectrum P_{mm}. The measurement of filament length is shown in Fig. 1. The mean and median length of the filaments increases as a function of redshift due to the combination of two factors. Firstly, the length of filaments, acting as the mass bridges between galaxy clusters, will decrease. Secondly, the number of filaments detected also depends on the number density of galaxies, which, in the CMASS sample, is low and decreases as a function of redshift (See Fig. 4.6). The large difference in the mean and median values of filament length indicates that the distribution of the filament length in each redshift bin is not Gaussian. We plot in the background the 2D histogram of filament length distribution as a function of redshift and filament length.

To check the validity of our model, we also compare the results to simula-

tions. The excellent agreement that we find in simulations provides an important consistency check. The theoretical prediction for $C_l^{\kappa f}$ is shown in eq. (4.7). The matter-filament correlation C_l^{mf} is defined as

$$C_l^{mf} = \int_{z_1}^{z_2} dz \frac{H(z)}{c} f(z)\chi^{-2}(z) P_{mf}\left(\frac{l}{\chi(z)}, z\right) \tag{4.9}$$

By taking the parameters that are slowly varying compared to $f^2(z)$, we get

$$C_l^{\kappa f} = \frac{3H_0^2\Omega_{m,0}W(z)(1+z)}{2cH(z)} C_l^{mf} \tag{4.10}$$

For the filament catalogue, the effective redshift, defined as the weighted mean redshift of filament intensity, is 0.56. This approximation is not perfect, leading to a systematic bias in the prediction for $C_l^{\kappa f}$. We propose an estimator for this systematic bias in Supplementary Information [139]. As shown in Fig. 4.7, the systematic bias is less than 5%. Thus, the approximation only causes a negligible bias. In our analysis, we measure C_l^{mf} using 10 realizations of sky mocks of dark matter and corresponding filaments (See sky mocks for dark matter and filament in Method section).

Fig. 2a shows the cross angular power spectrum of filaments and the CMB lensing convergence field. We bin our sample into 16 ℓ bins. Comparing simulation with data, we get $\chi^2/d.o.f. = 2.38$ with all 16 data points and $\chi^2/d.o.f. = 1.16$ without the first data point. The deviation of the first data point from the prediction is likely due to cosmic variance given the small sky area ($f_{sky} = 0.065$) covered by the simulations.

We fit the model of eq. 4.7 to the data with the filament bias b_f as the fitting parameter. We use two different smoothing methods to find $k_\perp$. The first method is to define the perpendicular smoothing scale as the filament spacing, since any scale smaller than the filament spacing is smoothed out. The filament spacing is

approximately the filament length. Thus, filament length is the overall smoothing scale for the effective power spectrum in eq. (4.8). The result is shown by the red line in Fig. 2. The best χ^2 fit gives $b_f = 1.68 \pm 0.334$. Since filaments also have width, filament spacing may be an overestimate of the smoothing perpendicular to filaments. In the second model, we also fit for smoothing scale in the perpendicular direction as a free parameter, where we assume $1/k_\perp(z) \sim \alpha \times 1/k_\parallel(z)$. We get $\alpha = 0.65$ and $b_f = 1.47 \pm 0.28$. The result is shown as the orange line in Fig. 2.

We measure the significance of the cross-correlation detection by measuring the signal-to-noise ratio (SNR). Our SNR is defined as follows

$$S/N = \sqrt{\chi^2_{NULL} - \chi^2_{fit}} \tag{4.11}$$

where

$$\chi^2_{NULL} = \sum_{ij} d_i^T \left(C_{ij}^{-1}\right) d_j \tag{4.12}$$

$$\chi^2_{fit} = \sum_{ij} (d_i - t_i)^T \left(C_{ij}^{-1}\right) (d_i - t_i) \tag{4.13}$$

where d_i is the cross angular signal in bin i, t_i is the best-fit theoretical prediction for the cross signal in bin i, and C is the covariance matrix estimated from jackknife resampling. The final result is shown in Table 4.1. The SNR values for both models show a significant detection of the cross-correlation. On large scales, we find that the filaments trace the matter with the filament bias around 1.5, which is somewhat smaller than galaxy bias from the same sample.

In order to validate the detection of our cross power spectrum, we perform a null test as follows. We rotate the CMB lensing convergence map by 90°, 135° and 180°, and then we cross correlate these rotated CMB convergence maps with

the filament intensity map. Fig. 2b shows that the cross signal with the rotated maps fluctuates around 0. $\chi^2_{NULL}/d.o.f.$ for the three cross angular power spectra is 0.79, 0.75 and 1.04, which means the cross-correlation between rotated CMB maps and the filament intensity map is consistent with 0. In addition, in order to test the impact of lensing generated by the clusters at the intersection of filaments on our signal, we mask out the redMaPPer clusters [153] in the CMB lensing map, finding a less than 4% difference in the cross angular power spectrum.

We define the cross-correlation coefficient between the filament and galaxy maps as $\rho = C_l^{fg}/\sqrt{C_l^{ff}C_l^{gg}}$, where C_l^{fg} is the cross angular power spectrum of filaments and galaxies, C_l^{ff} and C_l^{gg} are the auto angular power spectrum of filaments and galaxies. The result is shown in the left panel of Fig. 3a. The signal is highly correlated on large scales, since both galaxies and filaments trace the large-scale structure of the matter. However, the correlation decreases on small scales. Fig. 3b shows the cross-correlation of $C_l^{\kappa f}$ and $C_l^{\kappa g}$, where $C_l^{\kappa g}$ is the angular cross power spectrum of the CMB lensing convergence map and the CMASS galaxy catalogue. These two figures show that the maps are not totally correlated with large deviations at small scales. Establishing the amount of extra cosmological information present in the filaments field would require a joint analysis with galaxy clustering and lensing measurements; this is left to future work.

In our work, we have detected the effect of filaments lensing on the CMB by correlating filaments intensity map with CMB lensing convergence map. We measured filament bias, which is a quantitative description of how filaments trace the underlying matter, to be around 1.5. We perform null tests by rotating the CMB lensing map by more than its correlation length, obtaining results consistent with the null hypothesis. By comparing filaments with galaxies (both at the map and power spectrum level), we show an imperfect correlation, suggesting that there

might be additional information in the structure of the cosmic web, of which filaments provide an essential ingredient. In our study, the filament bias measured is significantly different from the mean bias of the CMASS galaxies used to create the filament catalog. This has important consequences for the environmental dependence of galaxy formation and can be key in generating accurate mocks for the next generations of surveys. In addition, the gas in filaments has been recently detected through the thermal Sunyaev-Zel'dovich (tSZ) effect derived from Planck maps [154] by measuring the gas pressure. A joint analysis of the mass profile and gas pressure can shed light on the majority of the gas in the intergalactic medium that resides outside of halos and hasn't been characterized so far.

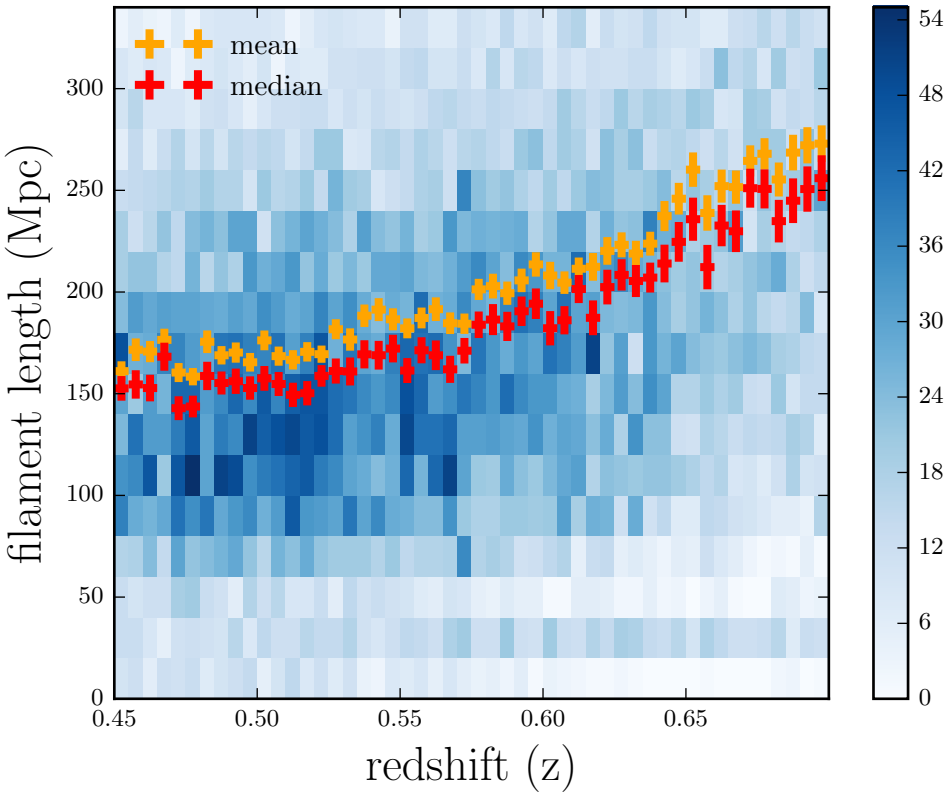


Figure 4.1: The filament length as a function of redshift. The orange (red) crosses are the mean (median) of the filament length in each redshift bin, where the error bars come from the standard error of the mean (median). The large difference in the mean and the median values implies the filament length distribution is not Gaussian. The background mesh plot shows the 2d histogram of the number of filament length as a function of the redshift and the filament length.

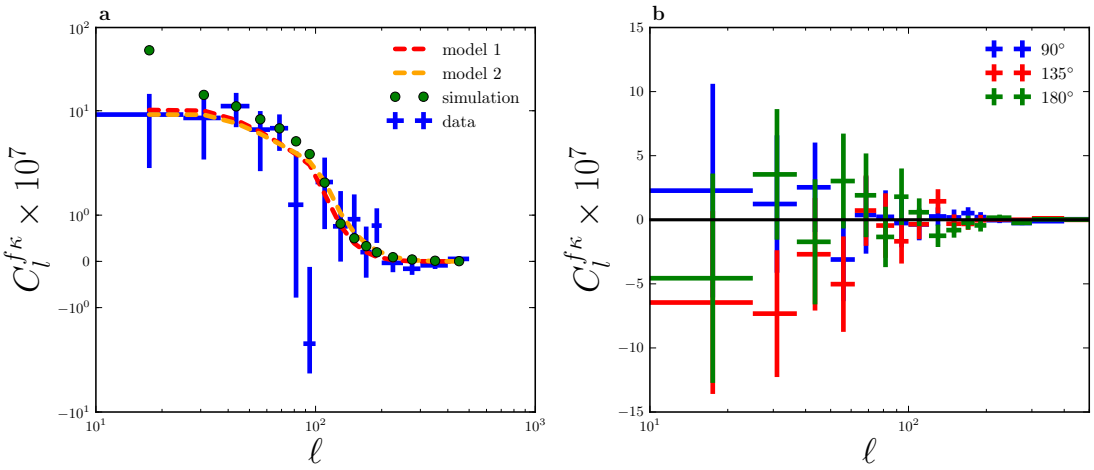


Figure 4.2: Cross angular power spectrum. (***a***) shows the cross angular power spectrum of the filaments and the CMB convergence field. The blue crosses are measured with error bars from jackknife resampling of the sky into 77 equally weighted regions. The red and orange dashed lines are theoretical predictions based on different smoothing models (red: filament length and spacing smoothing, orange: filament length and statistical fit for perpendicular smoothing). The corresponding filament bias for the two models are 1.68 and 1.47. The green circles are from simulations. (***b***) a null test showing the cross angular power spectrum of the filament catalogue and the rotated CMB lensing convergence map. The cross signals fluctuate around 0. The $\chi^2_{NULL}/d.o.f.$ for the three scenarios are all ~ 1.

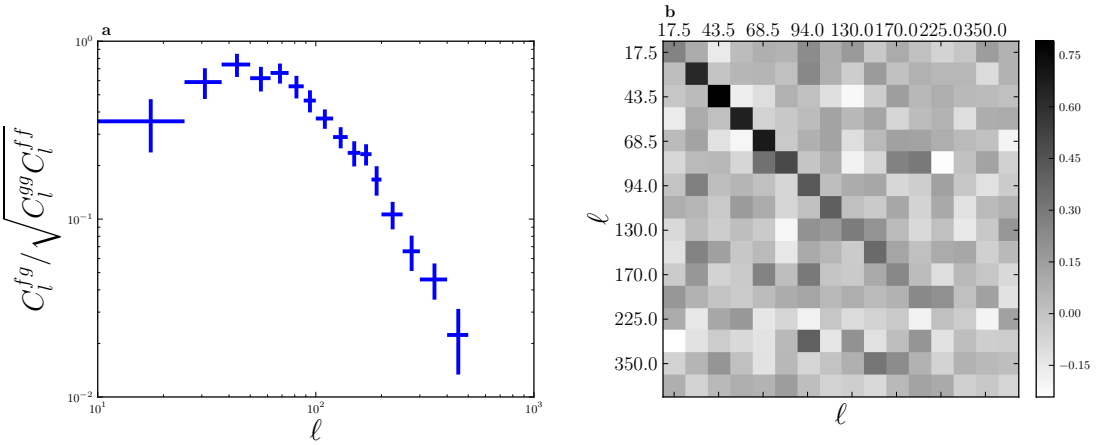


Figure 4.3: (**a**) The correlation coefficient of galaxy map and filament map and **b** the correlation of $C_l^{\kappa f}$ and $C_l^{\kappa g}$. Both plots show the filament and galaxy maps are not totally correlated with large deviations on small scales.

	model 1	model 2
b_f	1.68±0.334	1.47±0.28
S/N	5.0	5.2
χ^2_{fit}	25.77	24.39
$d.o.f.$	15	14

Table 4.1: The final result for the bias fitting. Model 1 uses filament length as the overall smoothing scale. In model 2, the filament length is the smoothing along filaments; we fit α for the smoothing in the perpendicular direction, where $1/k_\perp(z) \sim \alpha \times 1/k_\parallel(z)$. We get α to be 0.65 as the best fit. The bias and the error of bias come from χ^2 fitting of theory model to data.

4.3 Method

4.3.1 Filament finder

We obtain filaments from the publicly available Cosmic Web Reconstruction filament catalogue [143]. It finds filaments by applying the ridges finding algorithm

(filament detector) [145] to the galaxies in SDSS DR 12, ranging from $z = 0.050$ to $z = 0.700$. The spectroscopic galaxies are used since they give a reliable redshift estimate. Specifically, the catalogue is constructed using the following steps:

1. Partition the galaxies in redshift $z = 0.050 - 0.700$ into 130 redshift bins such that the bin width is $\Delta z = 0.005$. Galaxies within the same bin are projected onto a 2D space.

2. For each bin:

   (a) Reconstruct the (2D) galaxy probability density field by applying a kernel density estimator (KDE) with smoothing bandwidth chosen by the reference rule in [145].

   (b) Compute the root mean square (RMS) of the density field (ρ_{RMS}) within the area 150 deg $<$ RA $<$ 200 deg and 5 deg $<$ DEC $<$ 30 deg.

   (c) Remove galaxies in the regions where the probability density field is below a threshold density.

   (d) Apply the ridges finding algorithm [145] to the remaining galaxies.

   (e) Apply the galaxy mask to remove filaments outside the region of observations.

Here are some remarks on the construction of the catalogue.

- **The 2D projection.** The universe is sliced and galaxies are projected onto $2D$ space for several reasons. First, this enhances the stability of the filament detector. Second, this avoids the finger of god effect problem. Third, it's easy to compare filaments across different redshifts. More detailed discussion can be found in [143].

- **Choice of Δz.** The choice of $\Delta z = 0.005$ is to balance the estimation bias and the estimation random error. This is related to the so-called bias-variance tradeoff in statistics [155]. If Δz is very small, there will be a very limited number of galaxies, so the filament detector will be unstable. On the other hand, if Δz is large, the bin contains a very wide range of the universe so the filamentary structures may be washed away when projected onto $2D$ angular space. $\Delta z = 0.005$ is an empirical rule we discovered when applying to the SDSS data.

- **Area selection for calculating ρ_{RMS}.** The specific angular space ($[150, 200] \times [5, 30]\,\mathrm{deg}^2$) is selected to compute the RMS of the density field. The region is chosen because it is a wide region which is almost completely observed in SDSS samples. The range is large enough so the RMS calculation is stable.

- **Thresholding.** Before applying the ridges finding algorithm, galaxies are removed in the low density area. This thresholding stabilizes the ridges finding algorithm because the algorithm is very sensitive to density fluctuations in low density area.

The filament catalogue is shown to have strong agreement with the redMaPPer Catalogue, since most clusters in the redMaPPer Catalogue lie at the intersection of the filaments in the Cosmic Web Reconstruction filament catalogue, which is predicted by theory. The filament catalogue also has good consistency with the Voronoi model [145]. Furthermore, the effects of filaments on nearby galaxy properties (stellar mass, brightness, age, and orientation) are studied and it shows there is strong correlation of galaxy properties with filament environment, which satisfies theory prediction [156].

Uncertainty of filaments. The uncertainty of filaments is computed using the

bootstrap method [145; 157]. The filament detector returns a collection of points on filaments, which we call as filament points. For a given redshift bin, denote $F_1, \cdots, F_N$ as filament points. The uncertainty of filament points is computed as follows.

1. All galaxies (in one bin) are re-sampled with replacement to generate a new set of galaxies with the same total number of galaxies. This new set of galaxies is called a bootstrap sample.

2. Apply the same filament finding algorithm to the galaxies in the bootstrap sample. This yields a new set of filaments, which are called as the bootstrap filaments.

3. The distance of the filament point to the nearest filament point in the bootstrap filaments is calculated. Denote $\epsilon_1, \cdots, \epsilon_N$ as the distance for each filament point. This distance serves as an error measurement for $F_1, \cdots, F_N$.

4. Repeat the above 3 steps 1000 times (1000 : the number of bootstrap replicates). For each filament point, there will be 1000 error measurements. For instance, the ℓ-th filament point has 1000 error values:

$$\epsilon_\ell^{(1)}, \cdots, \epsilon_\ell^{(1000)}.$$

5. Compute the error (uncertainty) of each filament point by the RMS of the 1000 error measurements. Namely, for the ℓ-th filament point, the error is

$$\mathcal{E}_\ell = \sqrt{\frac{1}{1000}\sum_{j=1}^{1000}\left(\epsilon_\ell^{(j)}\right)^2}.$$

The bootstrap procedure measures the uncertainty due to the randomness of sampling [157].

4.3.2 Filament length measurement.

We get the filament intersections from Chen et al (Public in https://sites.google.com/site/yenchicr [143]. For each redshift bin, we use the hierarchical clustering method [145] to determine the number of branches at each intersection. The parameters in the hierarchical clustering are chosen to be the same as [145]:

$$r_{in} = \frac{2\omega}{3}, r_{out} = 2r_{in}, r_{sep} = (r_{in} + r_{out})/2. \tag{4.14}$$

where $\omega = \omega(z)$ is the smoothing bandwidth. At each intersection, we find the nearest point to the intersection point from each branch, and we group the nearest point as the filament point belonging to that filament (See Fig. 4.8). Then we keep finding the nearest point to the newly grouped filament to find the next filament point belonging to that branch. We stop the loop if the distance between filament points is less than r_{sep} and the distance between a filament point and the other intersection point is larger than $r_{sep}/2$.

4.3.3 Estimator.

We construct the filament map using the HEALPix pixelization with Nside=512. The CMB lensing convergence map is given directly by PLANCK using the HEALPix pixelization with Nside= 2048. We downgrade the lensing convergence map resolution to Nside=512 to cross-correlate with the filament map. The choice of resolution is consistent with the smoothing applied by the filament finder and

is large enough to fully resolve the scales relevant to our cross correlation.

We measure the cross angular power spectrum for the filament catalogue and the CMB lensing convergence field using a pseudo-C_l estimator:

$$\hat{C}_l^{\kappa f} = \frac{1}{(2l+1) f_{sky}^{\kappa f}} \sum_{m=-l}^{l} (\delta_f)_{lm} \kappa_{lm}^* \tag{4.15}$$

where $f_{sky}^{\kappa f}$ is the sky fraction observed by both the filament catalogue and the CMB lensing convergence field, κ_{lm} is the spherical harmonic transform of the lensing convergence field, and $(\delta_f)_{lm}$ is the spherical harmonic transform of the filament intensity overdensity. The spherical harmonic transform and C_l are computed using HEALPY.

4.3.4 Sky mock for filaments and dark matter.

We use N-body simulation runs using the TreePM method [158; 159; 160]. We use 10 realizations of this simulation based on the ΛCDM model with $\Omega_m = 0.292$ and $h = 0.69$. Although the parameters of the simulations are slightly different from the Planck cosmological parameters, if we compare the matter power spectrum with the cosmological parameters from the simulations and Planck, the difference is within 2%. Given the current noise in the data, we believe that this small difference is sub-dominant in our paper. These simulations are in a periodic box of side length $1380h^{-1}$Mpc and 2048^3 particles. A friend-of-friend halo catalogue is constructed at an effective redshift of $z = 0.55$. This is appropriate for our measurement since the galaxy sample used has effective redshift of 0.57. We use a Halo Occupation Distribution (HOD) [161; 162; 163; 164; 165; 166] to relate the observed clustering of galaxies with halos measured in the N-body simulation.

We have used the HOD model proposed in [167] to populate the halo catalogue with galaxies.

$$\begin{aligned} \langle N_{\rm cen}\rangle(M) &= \tfrac{1}{2}\left[1+{\rm erf}\left(\tfrac{\log M-\log M_{\rm min}}{\sigma_{\log M}}\right)\right] \\ \langle N_{\rm sat}\rangle(M) &= \langle N_{\rm cen}\rangle_M \left(\tfrac{M}{M_{\rm sat}}\right)^{\alpha} \exp\left(\tfrac{-M_{\rm cut}}{M}\right) \end{aligned} \tag{4.16}$$

where $\langle N_{\rm cen}\rangle(M)$ is the average number of central galaxies for a given halo mass M and $\langle N_{\rm sat}\rangle(M)$ is the average number of satellite galaxies. We use the HOD parameter set ($M_{\rm min} = 9.319\times10^{13}M_\odot/h, M_{\rm sat} = 6.729\times10^{13}M_\odot/h, \sigma_{\log M} = 0.2, \alpha = 1.1, M_{\rm cut} = 4.749\times10^{13}M_\odot/h$) from [167]. We have populated central galaxies at the center of our halo. The satellite galaxies are populated with radius (distance from central galaxy) distributed out to r_{200} as per the NFW profile; the direction is chosen randomly with a uniform distribution.

The sky mocks of dark matter and galaxy are obtained from the simulation box using the method described in [168]. We use publicly available "MAKE SURVEY" (https://github.com/mockFactory/make_survey) code to transform a periodic box into the pattern of survey. The first step of this transformation involves a volume remapping of the periodic box to sky coordinates preserving the structure in the simulation. This is achieved by using the publicly available package called "BoxRemap" (http://mwhite.berkeley.edu/BoxRemap) [169]. The BoxRemap defines an efficient volume-preserving, structure-preserving and one-to-one map to transform a periodic cubic box to non-cubical geometry. The non-cubical box is then translated and rotated to cover certain parts of the sky. We then convert the cartesian coordinate to the observed coordinate, which is right ascension, declination and redshift. We down-sample the galaxies with redshift to match the mock redshift with the redshift distribution observed in data. We request the reader refer to [168] for more details. We then apply the filament

detection algorithm to these simulated mocks using the method described in Filament Finder.

4.4 Supplementary information

4.4.1 Jackknife regions

To select the jackknife regions, we divide the observed sky in rectangular jackknife regions such that each region has same effective observed area by demanding equal number of randoms. We also tried to make a choice to keep the regions as close to square as possible so we dont introduce extra scales. We found that 11 × 7 (RA,DEC) jackknife regions satisfy all the constraints. We show an illustration of jackknife regions in Supplementary Figure 2.

4.4.2 Error calculation for $C_l^{\kappa f}$ from simulation

$C_l^{\kappa f}$ and C_l^{mf} are derived as

$$C_l^{\kappa f} = \frac{3H_0^2\Omega_{m,0}}{2c^2}\int_{z_1}^{z_2} dz W(z) f(z)\chi^{-2}(z)(1+z) \times P_{mf}\left(\frac{l}{\chi(z)}, z\right) \tag{4.17}$$

$$C_l^{mf} = \int_{z_1}^{z_2} dz \frac{H(z)}{c} f(z)\chi^{-2}(z) P_{mf}\left(\frac{l}{\chi(z)}, z\right) \tag{4.18}$$

By removing the appropriate functions from the integrands, which are slowly varying compared to $f(z)$, the correct expression between $C_l^{\kappa f}$ and C_l^{mf} is

$$C_l^{\kappa f} = \frac{3H_0^2\Omega_{m,0}W(z)(1+z)}{2cH(z)} C_l^{mf}$$

However, the approximations required to produce this expression are not perfect, causing the estimation of $C_l^{\kappa f}$ from simulation to slightly deviate from the true value of $C_l^{\kappa f}$. We estimate the deviation Γ by relating the theoretical prediction

for $C_l^{\kappa f}$ and C_l^{mf} by the following equation

$$\Gamma = \frac{2cH(z)C_l^{\kappa f}}{3H_0^2\Omega_{m,0}W(z)(1+z)C_l^{mf}} \tag{4.19}$$

The result is shown in Fig. 4.4. We see that the Γ is less than 5% from unity, which is much smaller than $\Delta(C_l^{\kappa f})/C_l^{\kappa f}$, where $\Delta(C_l^{\kappa f})$ is the error for $C_l^{\kappa f}$. Thus, the approximation for converting C_l^{mf} to $C_l^{\kappa f}$ only causes a negligible bias.

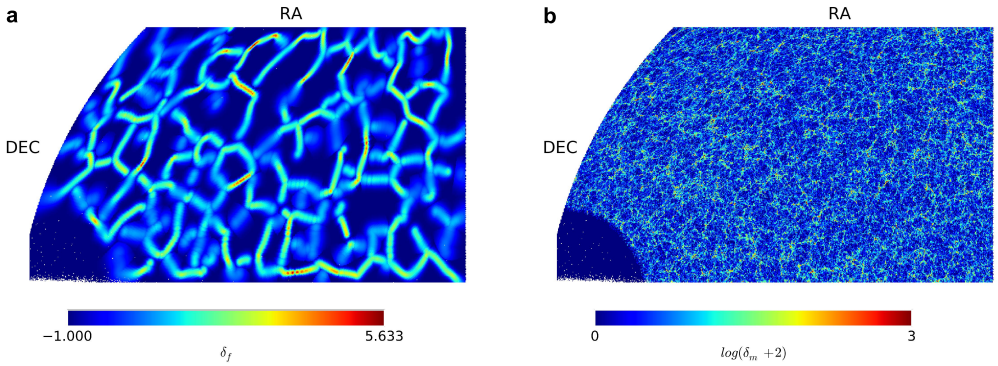


Figure 4.4: A demonstration of filament intensity overdensity and corresponding dark matter particle overdensity in simulation at redshift 0.57. (***a***) filament intensity overdensity at redshift 0.57. (***b***) dark matter overdensity at redshift 0.57. The color bars show the amplitude of the overdensity field in linear scale.

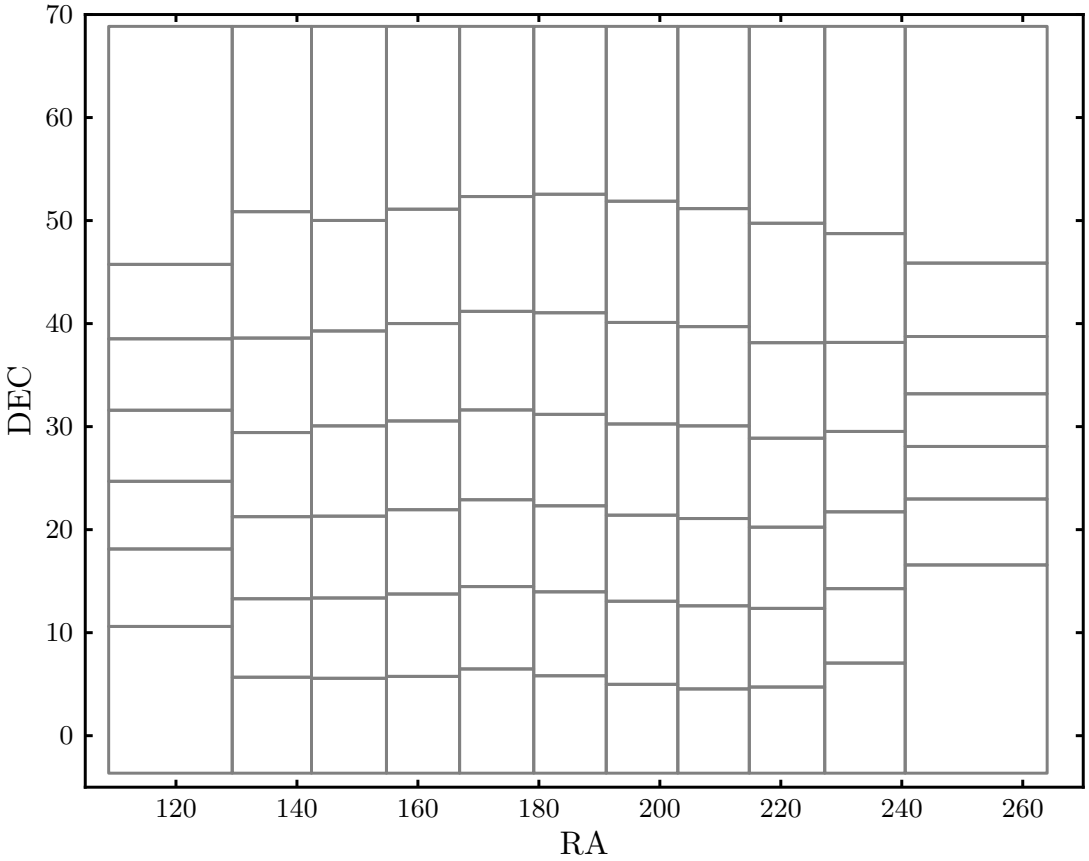


Figure 4.5: Visualization of jackknife regions. The jackknife regions are chosen so that each region has same effective observed area and is close to square.

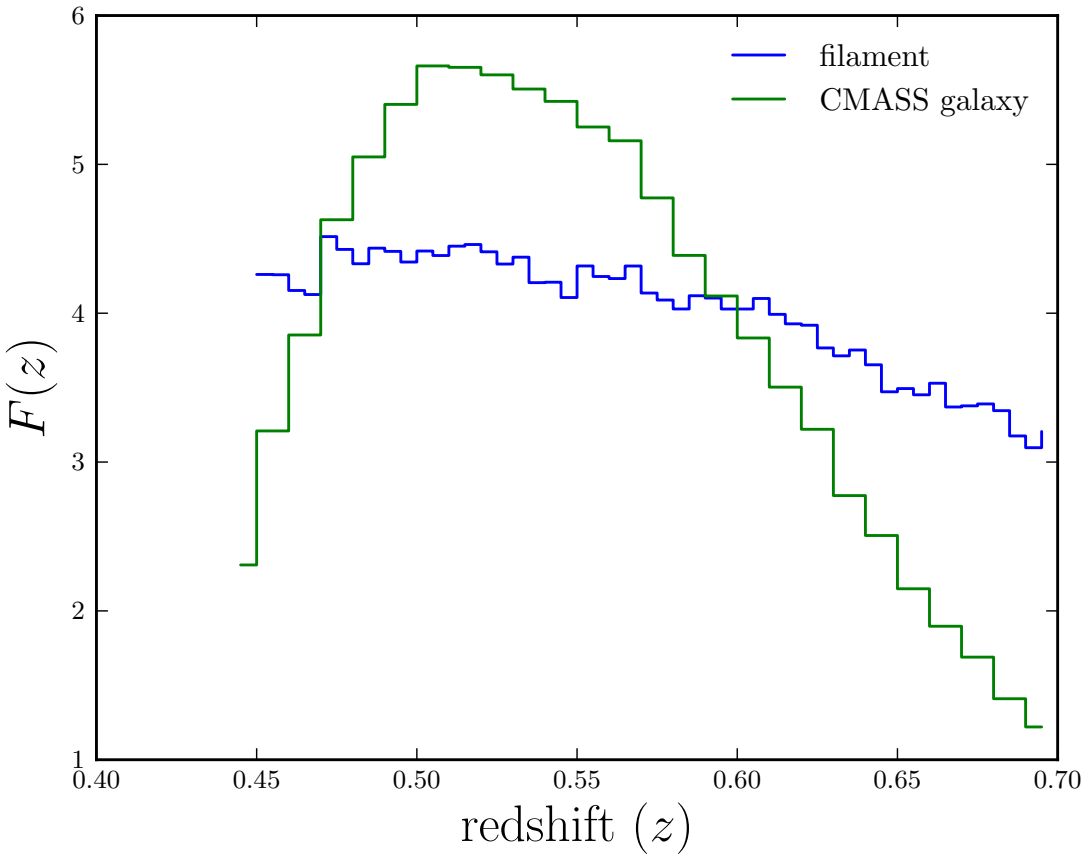


Figure 4.6: Filament intensity distribution and galaxy redshift distribution as a function of redshift. The blue curve shows the filament intensity distribution as a function of redshift. The green curve shows the CMASS galaxy redshift distribution, defined as the normalized distribution of the number density of galaxies as a function of redshift. The decrease of the filament intensity distribution results from the decrease of CMASS galaxy redshift distribution, from where the filaments in each redshift slice are detected.

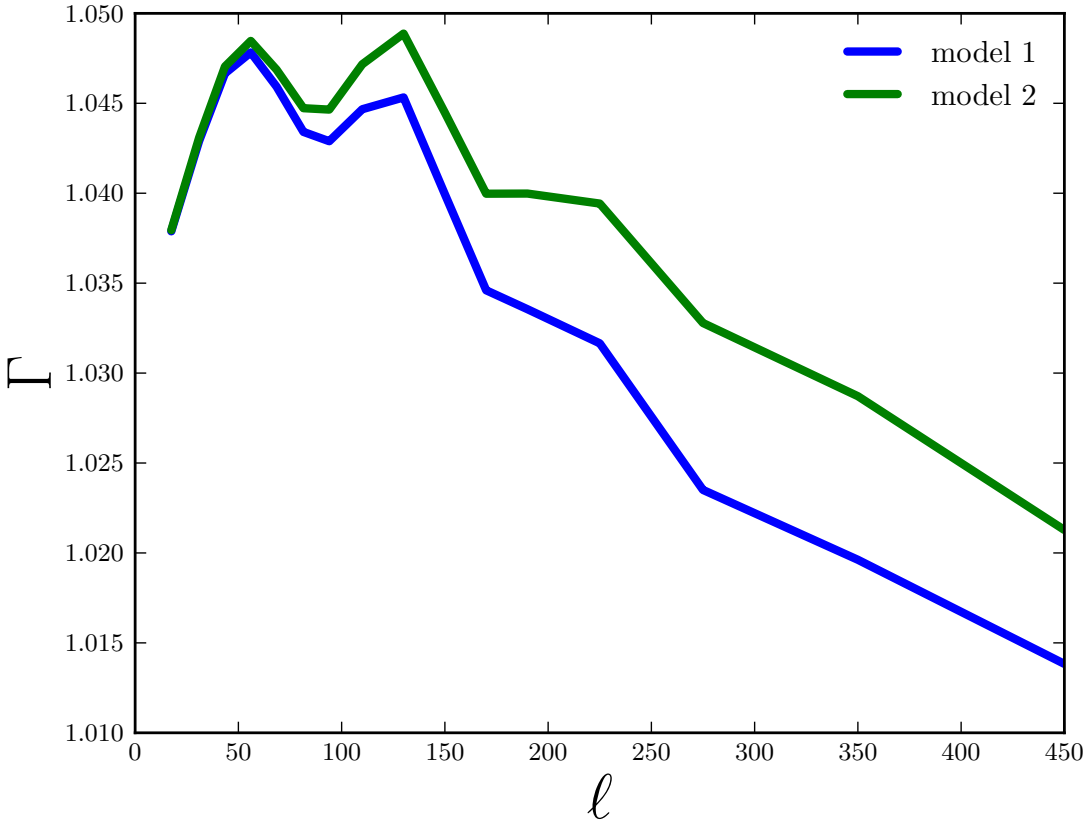


Figure 4.7: Quantification of the deviation between $C_l^{\kappa f}$ estimated from simulation and from theoretical $C_l^{\kappa f}$. Since the approximations to get $C_l^{\kappa f}$ from simulations are not perfect, the $C_l^{\kappa f}$ from simulation will slightly deviate from the true value of $C_l^{\kappa f}$. Γ quantifies the deviation. In theory, We use two models for the smoothing introduced by filaments. In model1, filament length is the overall smoothing scale. In model2, filament length is the smoothing scale along the filament and we fit α for the smoothing in the perpendicular direction, where $1/k_\perp(z) \sim \alpha \times 1/k_\parallel(z)$ and k is the wavenumber in Fourier space. As shown in the plot, the deviation of $C_l^{\kappa f}$ between simulation and theory is less than 5% from unity for both theoretical models.

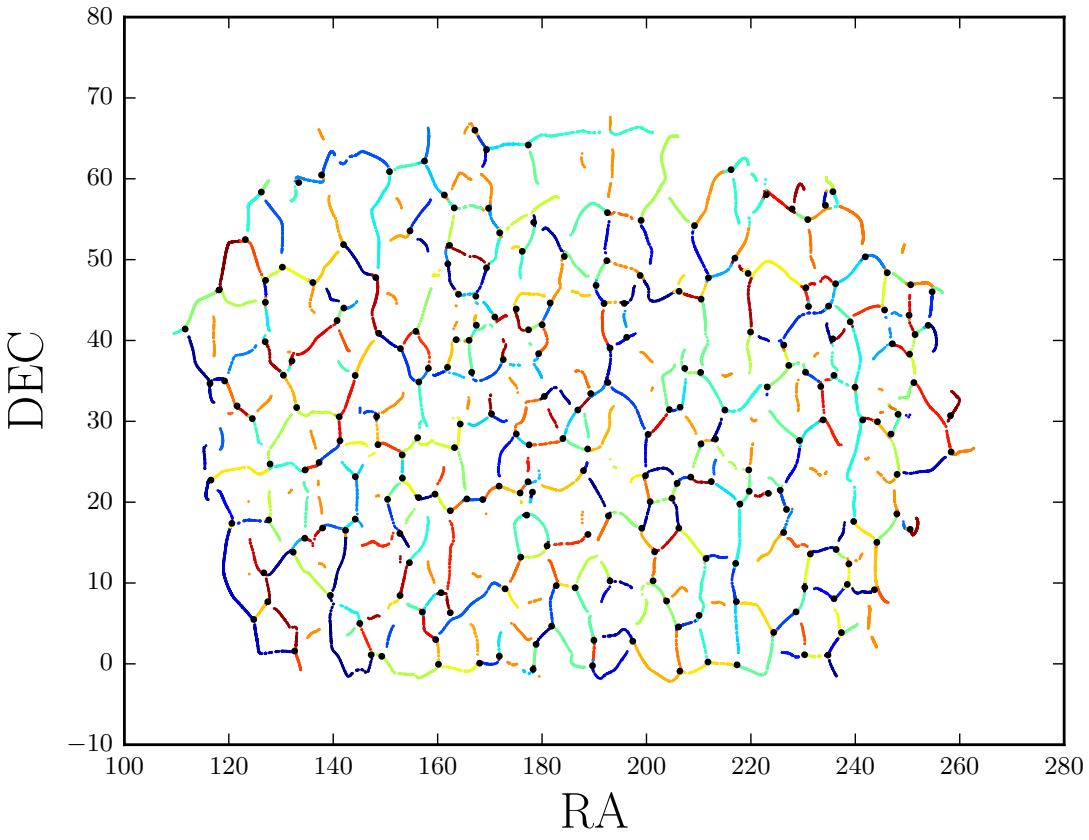


Figure 4.8: Example of filament grouped in redshift bin 0.55. A line with the same color is considered as belonging to the same filament.

Chapter 5

Conclusions

In this dissertation, we have addressed the non-linearity of the large scale structure from two new perspectives.

To understand the evolution of the Universe requires a concerted effort of accurate observation of the sky and fast prediction of structures in the Universe. N-body simulation is an effective approach to predicting structure formation of the Universe, though computationally expensive to resolve the small scale clustering. In the dissertation, we build a deep neural network (D^3M model) to predict the structure formation of the Universe. It outperforms the traditional fast analytical approximation (2LPT) in pair-wise, two-point and three-point statistics, and accurately extrapolates beyond its training data (Ω_m and A_s). Our study suggests that deep learning is an accurate alternative to the traditional way of generating approximate cosmological simulations. Our study is also the first to use deep learning to generate complex 3D simulations in cosmology, which suggests deep learning can provide a powerful alternative to traditional numerical simulations in cosmology. In the future, we could use the deep learning model to learn more complex simulations, such as the full N-body simulations and hydrodynamic simulations.

Though the deep learning method is shown to be powerful, one of the chal-

lenges of the deep model is that it's opaque – we don't know what the model learns. To address this challenge, we have used several simple analyses to interrogate the model. For the interrogations, the test data is drawn from a very different distribution than the training data and is analytically interpretable. Within the interrogations, the deep learning model is shown to have learned the underlying representations aligned with our expectations. However, our tests are minimal and in the future, more tests, such as the spherical collapse, should be applied to test the robustness of the model further.

The D^3M model falls into the big category of convolutional neural networks. However, one of the limitations of convolutional neural networks is that it requires the input data in a regular format, which is not suitable for data represented by point set. For the point set data, which is quite common in cosmology, we need the point cloud model. In this dissertation, we have presented the multi-scale deep sets model to learn the non-linear structure formation. The multi-scale deep sets model is shown to outperform the second order perturbation theory in two-point statistics. In the future, we expect the multi-scale deep sets could be useful for BAO reconstruction problems and other cosmology problems with data represented by a sparse point set.

Finally, in this dissertation, we have studied the cross correlation between filaments and CMB lensing convergence field to provide us additional information of our Universe. We report the first detection of the filaments lensing CMB. We propse a model to interpret the signal detected and we measure how filaments trace the matter distribution on large scales through filament bias. Our study provides a new scope to understand the environmental dependence of galaxy formation. In the future, the joint analysis of lensing and Sunyaev-Zeldovich observations might reveal the properties of missing baryons, the vast majority of the gas which resides in the intergalactic medium and has so far evaded most

observations.

References

[1] C. S. Frenk and S. D. M. White, "Dark matter and cosmic structure," *Annalen der Physik*, vol. 524, pp. 507–534, Oct 2012. vii, 2

[2] Bond, R.J., Kofman, L., and Pogosyan, D., "How filaments are woven into the cosmic web," *Nature*, vol. 380, pp. 603–606, 1996. 1, 11, 60

[3] M. Davis, J. Huchra, D. W. Latham, and J. Tonry, "A survey of galaxy redshifts. II - The large scale space distribution," *The Astrophysical Journal*, vol. 253, pp. 423–445, Feb. 1982. 1

[4] Colless, M. *et al.*, "The 2dF Galaxy Redshift Survey: Spectra and redshifts," *Mon. Not. Roy. Astron. Soc.*, vol. 328, p. 1039, 2001. 1, 21, 60

[5] Eisenstein, D.J. *et al.*, "SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems," *Astron. J.*, vol. 142, p. 72, 2011. 1, 21, 60, 61

[6] Jones, H.D. *et al.*, "The 6dF Galaxy Survey: Final Redshift Release (DR3) and Southern Large-Scale Structures," *Mon. Not. Roy. Astron. Soc.*, vol. 399, p. 683, 2009. 1, 21, 60

[7] Liske, J. *et al.*, "Galaxy and mass assembly (gama): end of survey report and data release 2," *Monthly Notices of the Royal Astronomical Society*, vol. 452, no. 2, p. 2087, 2015. 1, 21, 60

[8] Scodeggio, M. *et al.*, "The VIMOS Public Extragalactic Redshift Survey (VIPERS). Full spectroscopic data and auxiliary information release (PDR-2)," *ArXiv e-prints*, Nov. 2016. 1, 21, 60

[9] M. Einasto, L. J. Liivamägi, E. Tempel, E. Saar, E. Tago, P. Einasto, I. Enkvist, J. Einasto, V. J. Martínez, P. Heinämäki, and P. Nurmi, "The Sloan Great Wall. Morphology and Galaxy Content," *The Astrophysical Journal*, vol. 736, p. 51, July 2011. 3

[10] D. H. Weinberg, "Mapping the Large Scale Structure of the Universe," *arXiv e-prints*, pp. astro–ph/0510197, Oct 2005. 3, 8, 9

[11] Bernardeau, F., Colombi, S., Gaztaaga, E., and Scoccimarro, R., "Large-scale structure of the Universe and cosmological perturbation theory," *Physics Reports*, pp. 1–248, Feb. 2002. 3

[12] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, "The effective field theory of cosmological large scale structures," *Journal of High Energy Physics*, vol. 2012, p. 82, Sep 2012. 3

[13] J. Carlson, M. White, and N. Padmanabhan, "Critical look at cosmological perturbation theory techniques," *Physical Review D*, vol. 80, p. 043531, Aug 2009. 3, 6

[14] M. White, "The Zel'dovich approximation," *Monthly Notices of the Royal Astronomical Society*, vol. 439, pp. 3630–3640, Apr. 2014. 5, 23

[15] J. S. Bagla, "Cosmological N-Body simulation: Techniques, Scope and Status," *Current Science*, vol. 88, pp. 1088–1100, Apr 2005. 6, 7

[16] A. A. Klypin and S. F. Shandarin, "Three-dimensional numerical model of

the formation of large-scale structure in the Universe," *Monthly Notices of the Royal Astronomical Society*, vol. 204, pp. 891–907, Sept. 1983. 7

[17] F. R. Bouchet, J.-C. Adam, and R. Pellat, "On the way of understanding particle-in-cell simulations of gravitational clustering," *Astronomy and Astrophysics*, vol. 144, pp. 413–426, Mar. 1985. 7

[18] J. Barnes and P. Hut, "A hierarchical O(N log N) force-calculation algorithm," *Nature*, vol. 324, pp. 446–449, Dec. 1986. 7

[19] J. G. Jernigan and D. H. Porter, "A tree code with logarithmic reduction of force terms, hierarchical regularization of all variables, and explicit accuracy controls," *Astrophysical Journal Supplement Series*, vol. 71, pp. 871–893, Dec. 1989. 7

[20] G. Efstathiou, M. Davis, S. D. M. White, and C. S. Frenk, "Numerical techniques for large cosmological N-body simulations," *Astrophysical Journal Supplement Series*, vol. 57, pp. 241–260, Feb. 1985. 7

[21] H. M. P. Couchman, "Mesh-refined P3M - A fast adaptive N-body algorithm," *Astrophysical Journal*, vol. 368, pp. L23–L26, Feb. 1991. 7

[22] Y. Feng, M.-Y. Chu, U. Seljak, and P. McDonald, "FASTPM: a new scheme for fast simulations of dark matter and haloes," *Monthly Notices of the Royal Astronomical Society*, vol. 463, pp. 2273–2286, Dec. 2016. 7, 23

[23] A. Izard, M. Crocce, and P. Fosalba, "ICE-COLA: towards fast and accurate synthetic galaxy catalogues optimizing a quasi-N-body method," *Monthly Notices of the Royal Astronomical Society*, vol. 459, pp. 2327–2341, July 2016. 7

[24] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, and J. Helly, "Simulations of the formation, evolution and clustering of galaxies and quasars," *Nature*, vol. 435, pp. 629–636, Jun 2005. 7, 9

[25] C. B. Morrison and M. D. Schneider, "On estimating cosmology-dependent covariance matrices," *Journal of Cosmology and Astro-Particle Physics*, vol. 2013, p. 009, Nov 2013. 7

[26] A. Knebe, F. R. Pearce, P. A. Thomas, A. Benson, J. Blaizot, R. Bower, J. Carretero, F. J. Castander, A. Cattaneo, S. A. Cora, D. J. Croton, W. Cui, D. Cunnama, G. De Lucia, J. E. Devriendt, P. J. Elahi, A. Font, F. Fontanot, J. Garcia-Bellido, I. D. Gargiulo, V. Gonzalez-Perez, J. Helly, B. Henriques, M. Hirschmann, J. Lee, G. A. Mamon, P. Monaco, J. Onions, N. D. Padilla, C. Power, A. Pujol, R. A. Skibba, R. S. Somerville, C. Srisawat, C. A. Vega-Martínez, and S. K. Yi, "nIFTy cosmology: comparison of galaxy formation models," *Monthly Notices of the Royal Astronomical Society*, vol. 451, pp. 4029–4059, Aug. 2015. 7

[27] A. Cooray and R. Sheth, "Halo models of large scale structure," *Physics Reports*, vol. 372, pp. 1–129, Dec 2002. 7

[28] T. Boekholt and S. Portegies Zwart, "On the reliability of N-body simulations," *Computational Astrophysics and Cosmology*, vol. 2, p. 2, Mar 2015. 7

[29] A. A. Berlind and D. H. Weinberg, "The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass," *The Astrophysical Journal*, vol. 575, pp. 587–616, Aug 2002. 8

[30] S. Trujillo-Gomez, A. Klypin, J. Primack, and A. J. Romanowsky, "Galaxies in ΛCDM with Halo Abundance Matching: Luminosity-Velocity Relation, Baryonic Mass-Velocity Relation, Velocity Function, and Clustering," *The Astrophysical Journal*, vol. 742, p. 16, Nov. 2011. 8

[31] R. S. Somerville, G. Popping, and S. C. Trager, "Star formation in semi-analytic galaxy formation models with multiphase gas," *Monthly Notices of the Royal Astronomical Society*, vol. 453, pp. 4337–4367, Nov 2015. 8

[32] G. B. Dalton, R. A. C. Croft, G. Efstathiou, W. J. Sutherland , S. J. Maddox, and M. Davis, "The two-point correlation function of rich clusters of galaxies: results from an extended APM cluster redshift survey.," *Monthly Notices of the Royal Astronomical Society*, vol. 271, pp. L47–L51, Dec 1994. 9

[33] M. Takada and B. Jain, "The Three - point correlation function in cosmology," *Mon. Not. Roy. Astron. Soc.*, vol. 340, pp. 580–608, 2003. 10

[34] Z. Slepian and D. J. Eisenstein, "Computing the three-point correlation function of galaxies in $\mathcal{O}(N^2)$ time," *Mon. Not. Roy. Astron. Soc.*, vol. 454, no. 4, pp. 4142–4158, 2015. 10, 32

[35] F. Bernardeau, Y. Mellier, and L. van Waerbeke, "Detection of non-Gaussian signatures in the VIRMOS-DESCART lensing survey," *Astron. Astrophys.*, vol. 389, pp. L28–L32, 2002. 11

[36] M. A. Aragon-Calvo, R. van de Weygaert, and B. J. T. Jones, "Multiscale Phenomenology of the Cosmic Web," *Mon. Not. Roy. Astron. Soc.*, vol. 408, p. 2163, 2010. 11

[37] M. Alpaslan *et al.*, "Galaxy And Mass Assembly (GAMA): The large scale

structure of galaxies and comparison to mock universes," *Mon. Not. Roy. Astron. Soc.*, vol. 438, no. 1, pp. 177–194, 2014. 11

[38] R. E. Gonzalez and N. E. Padilla, "Automated detection of filaments in the large scale structure of the universe," *Mon. Not. Roy. Astron. Soc.*, vol. 407, pp. 1449–1463, 2010. 11

[39] J. E. Forero-Romero, Y. Hoffman, S. Gottloeber, A. Klypin, and G. Yepes, "A Dynamical Classification of the Cosmic Web," *Mon. Not. Roy. Astron. Soc.*, vol. 396, pp. 1815–1824, 2009. 11

[40] M. Cautun, R. van de Weygaert, and B. J. T. Jones, "NEXUS: Tracing the Cosmic Web Connection," *Mon. Not. Roy. Astron. Soc.*, vol. 429, p. 1286, 2013. 11

[41] N. I. Libeskind *et al.*, "Tracing the cosmic web," *Mon. Not. Roy. Astron. Soc.*, vol. 473, no. 1, pp. 1195–1217, 2018. 11

[42] M. Maturi and J. Merten, "Weak lensing detection of intra-cluster filaments with ground based data," *Astron. Astrophys.*, vol. 559, p. A112, 2013. 11, 12

[43] O. Hahn, C. M. Carollo, C. Porciani, and A. Dekel, "The Evolution of Dark Matter Halo Properties in Clusters, Filaments, Sheets and Voids," *Mon. Not. Roy. Astron. Soc.*, vol. 381, p. 41, 2007. 12

[44] M. Cautun, R. van de Weygaert, B. J. T. Jones, and C. S. Frenk, "Evolution of the cosmic web," *Mon. Not. Roy. Astron. Soc.*, vol. 441, no. 4, pp. 2923–2973, 2014. 12

[45] M. A. Aragon-Calvo, R. van de Weygaert, B. J. T. Jones, and J. M. T.

van der Hulst, "Spin alignment of dark matter haloes in filaments and walls," *Astrophys. J.*, vol. 655, pp. L5–L8, 2007. 12

[46] Y.-C. Chen, S. Ho, J. Blazek, S. He, R. Mandelbaum, P. Melchior, and S. Singh, "Detecting Galaxy-Filament Alignments in the Sloan Digital Sky Survey III," *Mon. Not. Roy. Astron. Soc.*, vol. 485, no. 2, pp. 2492–2504, 2019. 12

[47] J. Lee and U.-L. Pen, "Cosmic shear from galaxy spins," *Astrophys. J.*, vol. 532, p. L5, 2000. 12

[48] J. M. G. Mead, L. J. King, and I. G. McCarthy, "Probing the cosmic web: inter-cluster filament detection using gravitational lensing," *Mon. Not. Roy. Astron. Soc.*, vol. 401, p. 2257, 2010. 12

[49] J. Clampitt, B. Jain, M. Takada, and H. Miyatake, "Detection of Stacked Filament Lensing Between SDSS Luminous Red Galaxies," *Mon. Not. Roy. Astron. Soc.*, vol. 457, no. 3, pp. 2391–2400, 2016. 12

[50] A. de Graaff, Y.-C. Cai, C. Heymans, and J. A. Peacock, "Probing the missing baryons with the Sunyaev-Zel'dovich effect from filaments," *Astron. Astrophys.*, vol. 624, p. A48, 2019. 12

[51] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks.," in *CVPR*, vol. 1, p. 3, 2017. 12, 22

[52] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017. 12, 22

[53] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville,

"Improved training of wasserstein gans," in *Advances in Neural Information Processing Systems*, pp. 5767–5777, 2017. 12, 22

[54] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *CoRR abs/1609.03499*, 2016. 12, 22

[55] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, *et al.*, "Deep speech 2: End-to-end speech recognition in english and mandarin," in *International Conference on Machine Learning*, pp. 173–182, 2016. 12, 22

[56] Ravanbakhsh, S., Oliva, J., Fromenteau, S., Price, L. C., Ho, S., Schneider, J., and Poczos, B., "Estimating cosmological parameters from the dark matter distribution," *ArXiv e-prints*, Nov. 2017. 12, 22

[57] A. Mathuriya, D. Bard, P. Mendygral, L. Meadows, J. Arnemann, L. Shao, S. He, T. Karna, D. Moise, S. J. Pennycook, *et al.*, "Cosmoflow: using deep learning to learn the universe at scale," *arXiv preprint arXiv:1808.04728*, 2018. 12, 22

[58] Hezaveh, Y. D., Levasseur, L. P., and Marshall, P. J., "Fast automated analysis of strong gravitational lenses with convolutional neural networks," *Nature*, vol. 548, p. 555557, Aug. 2017. 12, 22

[59] Lanusse, F., Ma, Q., Li, N., Collett, T. E., Li, C., Ravanbakhsh, S., Mandelbaum, R., and Pczos, B., "Cmu deeplens: deep learning for automatic image-based galaxy-galaxy strong lens finding," *Monthly Notices of the Royal Astronomical Society*, pp. 3895–3906, Jan. 2018. 12, 22

[60] A. Peel, F. Lalande, J.-L. Starck, V. Pettorino, J. Merten, C. Giocoli,

M. Meneghetti, and M. Baldi, "Distinguishing standard and modified gravity cosmologies with machine learning," 2018. 12

[61] S. He, S. Ravanbakhsh, and S. Ho, "Analysis of cosmic microwave background with deep learning," 2018. 12, 14, 22

[62] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier, "Deepsphere: Efficient spherical convolutional neural network with healpix sampling for cosmological applications," *arXiv preprint arXiv:1810.12186*, 2018. 12, 22

[63] J. Caldeira, W. Wu, B. Nord, C. Avestruz, S. Trivedi, and K. Story, "Deepcmb: Lensing reconstruction of the cosmic microwave background with deep neural networks," *arXiv preprint arXiv:1810.01483*, 2018. 12, 22

[64] J. Chardin, G. Uhlrich, D. Aubert, N. Deparis, N. Gillet, P. Ocvirk, and J. Lewis, "A deep learning model to emulate simulations of cosmic reionization," 2019. 12

[65] V. A. Villar *et al.*, "Supernova Photometric Classification Pipelines Trained on Spectroscopically Classified Supernovae from the Pan-STARRS1 Medium-Deep Survey," 2019. 12

[66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012. 12

[67] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," *arXiv e-prints*, p. arXiv:1612.00593, Dec 2016. 14, 55

[68] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola, "Deep Sets," *arXiv e-prints*, p. arXiv:1703.06114, Mar 2017. 14, 55

[69] G. Montavon, W. Samek, and K.-R. Müller, "Methods for Interpreting and Understanding Deep Neural Networks," *arXiv e-prints*, p. arXiv:1706.07979, Jun 2017. 14

[70] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," *arXiv e-prints*, p. arXiv:1506.02142, Jun 2015. 14

[71] J. M. Hernández-Lobato and R. P. Adams, "Probabilistic Backpropagation for Scalable Learning of Bayesian Neural Networks," *arXiv e-prints*, p. arXiv:1502.05336, Feb 2015. 14

[72] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, D. Alonso, Y. AlSayyad, S. F. Anderson, J. Andrew, and et al., "LSST: from Science Drivers to Reference Design and Anticipated Data Products," *ArXiv e-prints*, May 2008. 21

[73] L. Amendola, S. Appleby, A. Avgoustidis, D. Bacon, T. Baker, M. Baldi, N. Bartolo, A. Blanchard, C. Bonvin, S. Borgani, E. Branchini, C. Burrage, S. Camera, C. Carbone, L. Casarini, M. Cropper, C. de Rham, J. P. Dietrich, C. Di Porto, R. Durrer, A. Ealet, P. G. Ferreira, F. Finelli, J. García-Bellido, T. Giannantonio, L. Guzzo, A. Heavens, L. Heisenberg, C. Heymans, H. Hoekstra, L. Hollenstein, R. Holmes, Z. Hwang, K. Jahnke, T. D. Kitching, T. Koivisto, M. Kunz, G. La Vacca, E. Linder, M. March, V. Marra, C. Martins, E. Majerotto, D. Markovic, D. Marsh, F. Marulli, R. Massey, Y. Mellier, F. Montanari, D. F. Mota, N. J. Nunes, W. Percival,

V. Pettorino, C. Porciani, C. Quercellini, J. Read, M. Rinaldi, D. Sapone, I. Sawicki, R. Scaramella, C. Skordis, F. Simpson, A. Taylor, S. Thomas, R. Trotta, L. Verde, F. Vernizzi, A. Vollmer, Y. Wang, J. Weller, and T. Zlosnik, "Cosmology and fundamental physics with the Euclid satellite," *Living Reviews in Relativity*, vol. 21, p. 2, Apr. 2018. 21

[74] D. Spergel, N. Gehrels, C. Baltay, D. Bennett, J. Breckinridge, M. Donahue, A. Dressler, B. S. Gaudi, T. Greene, O. Guyon, C. Hirata, J. Kalirai, N. J. Kasdin, B. Macintosh, W. Moos, S. Perlmutter, M. Postman, B. Rauscher, J. Rhodes, Y. Wang, D. Weinberg, D. Benford, M. Hudson, W.-S. Jeong, Y. Mellier, W. Traub, T. Yamada, P. Capak, J. Colbert, D. Masters, M. Penny, D. Savransky, D. Stern, N. Zimmerman, R. Barry, L. Bartusek, K. Carpenter, E. Cheng, D. Content, F. Dekens, R. Demers, K. Grady, C. Jackson, G. Kuan, J. Kruk, M. Melton, B. Nemati, B. Parvin, I. Poberezhskiy, C. Peddie, J. Ruffa, J. K. Wallace, A. Whipple, E. Wollack, and F. Zhao, "Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report," *ArXiv e-prints*, Mar. 2015. 21

[75] T. MacFarland, H. M. P. Couchman, F. R. Pearce, and J. Pichlmeier, "A new parallel P^3M code for very large-scale cosmological simulations," *New Astronomy*, vol. 3, pp. 687–705, Dec 1998. 21

[76] V. Springel, N. Yoshida, and S. D. M. White, "GADGET: a code for collisionless and gasdynamical cosmological simulations," *New Astronomy*, vol. 6, pp. 79–117, Apr 2001. 21

[77] J. S. Bagla, "TreePM: A Code for Cosmological N-Body Simulations," *Journal of Astrophysics and Astronomy*, vol. 23, pp. 185–196, Dec 2002. 21

[78] J. R. Bond, L. Kofman, and D. Pogosyan, "How filaments of galaxies are woven into the cosmic web," *Nature*, vol. 380, pp. 603–606, Apr. 1996. 21

[79] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, "The evolution of large-scale structure in a universe dominated by cold dark matter," *The Astrophysical Journal*, vol. 292, pp. 371–394, May 1985. 21

[80] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015. 22

[81] Z. Hu, Z. Yang, X. Liang, R. Salakhutdinov, and E. P. Xing, "Toward controlled generation of text," *arXiv preprint arXiv:1703.00955*, 2017. 22

[82] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017. 22

[83] E. Denton and R. Fergus, "Stochastic video generation with a learned prior," *arXiv preprint arXiv:1802.07687*, 2018. 22

[84] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2625–2634, 2015. 22

[85] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.*, "Mastering the game of go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, p. 484, 2016. 22

[86] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015. 22

[87] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1334–1373, 2016. 22

[88] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, *et al.*, "Opportunities and obstacles for deep learning in biology and medicine," *Journal of The Royal Society Interface*, vol. 15, no. 141, p. 20170387, 2018. 22

[89] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of dna-and rna-binding proteins by deep learning," *Nature biotechnology*, vol. 33, no. 8, p. 831, 2015. 22

[90] M. H. Segler, M. Preuss, and M. P. Waller, "Planning chemical syntheses with deep neural networks and symbolic ai," *Nature*, vol. 555, no. 7698, p. 604, 2018. 22

[91] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," *arXiv preprint arXiv:1704.01212*, 2017. 22

[92] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," *Science*, vol. 355, no. 6325, pp. 602–606, 2017. 22

[93] C. Adam-Bourdarios, G. Cowan, C. Germain, I. Guyon, B. Kégl, and D. Rousseau, "The higgs boson machine learning challenge," in *NIPS 2014 Workshop on High-energy Physics and Machine Learning*, pp. 19–55, 2015. 22

[94] N. Kennamer, D. Kirkby, A. Ihler, and F. J. Sanchez-Lopez, "ContextNet: Deep learning for star galaxy classification," in *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.), vol. 80 of *Proceedings of Machine Learning Research*, (Stockholmsmssan, Stockholm Sweden), pp. 2582–2590, PMLR, 10–15 Jul 2018. 22

[95] E. J. Kim and R. J. Brunner, "Star-galaxy classification using deep convolutional neural networks," *Monthly Notices of the Royal Astronomical Society*, p. stw2672, 2016. 22

[96] M. Lochner, J. D. McEwen, H. V. Peiris, O. Lahav, and M. K. Winter, "Photometric supernova classification with machine learning," *The Astrophysical Journal Supplement Series*, vol. 225, no. 2, p. 31, 2016. 22

[97] P. W. Battaglia, J. B. Hamrick, and J. B. Tenenbaum, "Simulation as an engine of physical scene understanding," *Proceedings of the National Academy of Sciences*, p. 201306572, 2013. 22

[98] P. Battaglia, R. Pascanu, M. Lai, D. J. Rezende, *et al.*, "Interaction networks for learning about objects, relations and physics," in *Advances in neural information processing systems*, pp. 4502–4510, 2016. 22

[99] R. Mottaghi, H. Bagherinezhad, M. Rastegari, and A. Farhadi, "Newtonian scene understanding: Unfolding the dynamics of objects in static images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3521–3529, 2016. 22

[100] M. B. Chang, T. Ullman, A. Torralba, and J. B. Tenenbaum, "A compositional object-based approach to learning physical dynamics," *arXiv preprint arXiv:1612.00341*, 2016. 22

[101] J. Wu, I. Yildirim, J. J. Lim, B. Freeman, and J. Tenenbaum, "Galileo: Perceiving physical object properties by integrating a physics engine with deep learning," in *Advances in neural information processing systems*, pp. 127–135, 2015. 22

[102] J. Wu, J. J. Lim, H. Zhang, J. B. Tenenbaum, and W. T. Freeman, "Physics 101: Learning physical object properties from unlabeled videos.," in *BMVC*, vol. 2, p. 7, 2016. 22

[103] N. Watters, D. Zoran, T. Weber, P. Battaglia, R. Pascanu, and A. Tacchetti, "Visual interaction networks: Learning a physics simulator from video," in *Advances in Neural Information Processing Systems*, pp. 4539–4547, 2017. 22

[104] A. Lerer, S. Gross, and R. Fergus, "Learning physical intuition of block towers by example," *arXiv preprint arXiv:1603.01312*, 2016. 22

[105] P. Agrawal, A. V. Nair, P. Abbeel, J. Malik, and S. Levine, "Learning to poke by poking: Experiential learning of intuitive physics," in *Advances in Neural Information Processing Systems*, pp. 5074–5082, 2016. 22

[106] K. Fragkiadaki, P. Agrawal, S. Levine, and J. Malik, "Learning visual predictive models of physics for playing billiards," *arXiv preprint arXiv:1511.07404*, 2015. 22

[107] J. Tompson, K. Schlachter, P. Sprechmann, and K. Perlin, "Accelerating eulerian fluid simulation with convolutional networks," *arXiv preprint arXiv:1607.03597*, 2016. 22

[108] P. Ronneberger, O.; Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *MICCAI 2015*, pp. 234–241, 2015. 23, 37

[109] Y. B. Zel'dovich, "Gravitational instability: An approximate theory for large density perturbations.," *Astronomy and Astrophysics*, vol. 5, pp. 84–89, Mar. 1970. 23

[110] T. Buchert, "Lagrangian Theory of Gravitational Instability of Friedman-Lemaitre Cosmologies - a Generic Third-Order Model for Nonlinear Clustering," *Monthly Notices of the Royal Astronomical Society*, vol. 267, p. 811, Apr. 1994. 24

[111] J. Jasche and B. D. Wandelt, "Bayesian physical reconstruction of initial conditions from large-scale structure surveys," *Monthly Notices of the Royal Astronomical Society*, vol. 432, pp. 894–913, June 2013. 24

[112] F.-S. Kitaura, "The initial conditions of the Universe from constrained simulations," *Monthly Notices of the Royal Astronomical Society*, vol. 429, pp. L84–L88, Feb. 2013. 24

[113] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouser, J. E. Bautista, A. Beifiori, A. A. Berlind, V. Bhardwaj, D. Bizyaev, C. H. Blake, M. R. Blanton, M. Blomqvist, A. S. Bolton, A. Borde, J. Bovy, W. N. Brandt, H. Brewington, J. Brinkmann, P. J. Brown, J. R. Brownstein, K. Bundy, N. G. Busca, W. Carithers, A. R. Carnero, M. A. Carr, Y. Chen, J. Comparat, N. Connolly, F. Cope, R. A. C. Croft, A. J. Cuesta, L. N. da Costa, J. R. A. Davenport, T. Delubac, R. de Putter, S. Dhital, A. Ealet, G. L. Ebelke, D. J. Eisenstein, S. Escoffier, X. Fan, N. Filiz Ak, H. Finley, A. Font-Ribera, R. Génova-Santos,

J. E. Gunn, H. Guo, D. Haggard, P. B. Hall, J.-C. Hamilton, B. Harris, D. W. Harris, S. Ho, D. W. Hogg, D. Holder, K. Honscheid, J. Huehnerhoff, B. Jordan, W. P. Jordan, G. Kauffmann, E. A. Kazin, D. Kirkby, M. A. Klaene, J.-P. Kneib, J.-M. Le Goff, K.-G. Lee, D. C. Long, C. P. Loomis, B. Lundgren, R. H. Lupton, M. A. G. Maia, M. Makler, E. Malanushenko, V. Malanushenko, R. Mandelbaum, M. Manera, C. Maraston, D. Margala, K. L. Masters, C. K. McBride, P. McDonald, I. D. McGreer, R. G. McMahon, O. Mena, J. Miralda-Escudé, A. D. Montero-Dorta, F. Montesano, D. Muna, A. D. Myers, T. Naugle, R. C. Nichol, P. Noterdaeme, S. E. Nuza, M. D. Olmstead, A. Oravetz, D. J. Oravetz, R. Owen, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. K. Parejko, I. Pâris, W. J. Percival, I. Pérez-Fournon, I. Pérez-Ràfols, P. Petitjean, R. Pfaffenberger, J. Pforr, M. M. Pieri, F. Prada, A. M. Price-Whelan, M. J. Raddick, R. Rebolo, J. Rich, G. T. Richards, C. M. Rockosi, N. A. Roe, A. J. Ross, N. P. Ross, G. Rossi, J. A. Rubiño-Martin, L. Samushia, A. G. Sánchez, C. Sayres, S. J. Schmidt, D. P. Schneider, C. G. Scóccola, H.-J. Seo, A. Shelden, E. Sheldon, Y. Shen, Y. Shu, A. Slosar, S. A. Smee, S. A. Snedden, F. Stauffer, O. Steele, M. A. Strauss, A. Streblyanska, N. Suzuki, M. E. C. Swanson, T. Tal, M. Tanaka, D. Thomas, J. L. Tinker, R. Tojeiro, C. A. Tremonti, M. Vargas Magaña, L. Verde, M. Viel, D. A. Wake, M. Watson, B. A. Weaver, D. H. Weinberg, B. J. Weiner, A. A. West, M. White, W. M. Wood-Vasey, C. Yeche, I. Zehavi, G.-B. Zhao, and Z. Zheng, "The Baryon Oscillation Spectroscopic Survey of SDSS-III," *The Astronomical Journal*, vol. 145, p. 10, Jan. 2013. 24, 61

[114] K. S. Dawson, J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, S. F. Anderson, E. Armengaud, É. Aubourg, S. Bailey, J. E. Bautista, A. A. Berlind, M. A. Bershady, F. Beutler, D. Bizyaev, M. R. Blanton, M. Blomqvist, A. S.

Bolton, J. Bovy, W. N. Brandt, J. Brinkmann, J. R. Brownstein, E. Burtin, N. G. Busca, Z. Cai, C.-H. Chuang, N. Clerc, J. Comparat, F. Cope, R. A. C. Croft, I. Cruz-Gonzalez, L. N. da Costa, M.-C. Cousinou, J. Darling, A. de la Macorra, S. de la Torre, T. Delubac, H. du Mas des Bourboux, T. Dwelly, A. Ealet, D. J. Eisenstein, M. Eracleous, S. Escoffier, X. Fan, A. Finoguenov, A. Font-Ribera, P. Frinchaboy, P. Gaulme, A. Georgakakis, P. Green, H. Guo, J. Guy, S. Ho, D. Holder, J. Huehnerhoff, T. Hutchinson, Y. Jing, E. Jullo, V. Kamble, K. Kinemuchi, D. Kirkby, F.-S. Kitaura, M. A. Klaene, R. R. Laher, D. Lang, P. Laurent, J.-M. Le Goff, C. Li, Y. Liang, M. Lima, Q. Lin, W. Lin, Y.-T. Lin, D. C. Long, B. Lundgren, N. MacDonald, M. A. Geimba Maia, E. Malanushenko, V. Malanushenko, V. Mariappan, C. K. McBride, I. D. McGreer, B. Ménard, A. Merloni, A. Meza, A. D. Montero-Dorta, D. Muna, A. D. Myers, K. Nandra, T. Naugle, J. A. Newman, P. Noterdaeme, P. Nugent, R. Ogando, M. D. Olmstead, A. Oravetz, D. J. Oravetz, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. K. Parejko, I. Pâris, J. A. Peacock, P. Petitjean, M. M. Pieri, A. Pisani, F. Prada, A. Prakash, A. Raichoor, B. Reid, J. Rich, J. Ridl, S. Rodriguez-Torres, A. Carnero Rosell, A. J. Ross, G. Rossi, J. Ruan, M. Salvato, C. Sayres, D. P. Schneider, D. J. Schlegel, U. Seljak, H.-J. Seo, B. Sesar, S. Shandera, Y. Shu, A. Slosar, F. Sobreira, A. Streblyanska, N. Suzuki, D. Taylor, C. Tao, J. L. Tinker, R. Tojeiro, M. Vargas-Magaña, Y. Wang, B. A. Weaver, D. H. Weinberg, M. White, W. M. Wood-Vasey, C. Yeche, Z. Zhai, C. Zhao, G.-b. Zhao, Z. Zheng, G. Ben Zhu, and H. Zou, "The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data," *The Astronomical Journal*, vol. 151, p. 44, Feb. 2016. 24

[115] DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E.

Allen, C. Allende Prieto, J. Annis, S. Bailey, C. Balland, and et al., "The DESI Experiment Part I: Science,Targeting, and Survey Design," *ArXiv e-prints*, Oct. 2016. 24

[116] Y. Feng, U. Seljak, and M. Zaldarriaga, "Exploring the posterior surface of the large scale structure reconstruction," *Journal of Cosmology and Astroparticle Physics*, vol. 7, p. 043, July 2018. 26

[117] K. C. Chan, "Helmholtz decomposition of the Lagrangian displacement," *Physical Review D*, vol. 89, p. 083515, Apr. 2014. 30

[118] A. Perko, L. Senatore, E. Jennings, and R. H. Wechsler, "Biased Tracers in Redshift Space in the EFT of Large-Scale Structure," *arXiv e-prints*, p. arXiv:1610.09321, Oct 2016. 30

[119] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al., "Planck 2015 results. XIII. Cosmological parameters," *Astronomy and Astrophysics*, vol. 594, p. A13, Sept. 2016. 33

[120] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," *ArXiv e-prints*, June 2016. 37

[121] P. Berger and G. Stein, "A volumetric deep Convolutional Neural Network for simulation of mock dark matter halo catalogues," *Monthly Notices of the Royal Astronomical Society*, vol. 482, pp. 2861–2871, Jan. 2019. 37

[122] M. A. Aragon-Calvo, "Classifying the Large Scale Structure of the Universe with Deep Neural Networks," *ArXiv e-prints*, Apr. 2018. 37

[123] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *eprint arXiv: 1412.6980*, 12 2014. 37

[124] F. Leclercq, J. Jasche, H. Gil-Marín, and B. Wandelt, "One-point remapping of Lagrangian perturbation theory in the mildly non-linear regime of cosmic structure formation," *Journal of Cosmology and Astro-Particle Physics*, vol. 2013, p. 048, Nov 2013. 43

[125] R. Teyssier, J. P. Chize, and J. M. Alimi, "Nonequilibrium Thermodynamics and Cosmological Pancake Formation," *The Astrophysical Journal*, vol. 509, pp. 62–73, 12 1998. 47

[126] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," *arXiv e-prints*, p. arXiv:1312.6034, Dec 2013. 49

[127] H.-W. Lin, T. Chiew-Lan, and G.-J. Wang, "A mesh reconstruction algorithm driven by an intrinsic property of a point cloud," *Computer-Aided Design*, vol. 36, pp. 1–9, Jan 2004. 55

[128] S. Ravanbakhsh, J. Oliva, S. Fromenteau, L. C. Price, S. Ho, J. Schneider, and B. Poczos, "Estimating Cosmological Parameters from the Dark Matter Distribution," *arXiv e-prints*, p. arXiv:1711.02033, Nov 2017. 55

[129] X. Zhang, Y. Wang, W. Zhang, Y. Sun, S. He, G. Contardo, F. Villaescusa-Navarro, and S. Ho, "From Dark Matter to Galaxies with Convolutional Networks," *arXiv e-prints*, p. arXiv:1902.05965, Feb 2019. 55

[130] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," *arXiv e-prints*, p. arXiv:1706.02413, Jun 2017. 55

[131] R. Klokov and V. Lempitsky, "Escape from Cells: Deep Kd-Networks for the Recognition of 3D Point Cloud Models," *arXiv e-prints*, p. arXiv:1704.01222, Apr 2017. 55

[132] K. M. Smith, O. Zahn, and O. Dore, "Detection of Gravitational Lensing in the Cosmic Microwave Background," *Phys. Rev.*, vol. D76, p. 043510, 2007. 61

[133] C. M. Hirata, S. Ho, N. Padmanabhan, U. Seljak, and N. A. Bahcall, "Correlation of CMB with large-scale structure: II. Weak lensing," *Phys. Rev.*, vol. D78, p. 043520, 2008. 61

[134] L. E. Bleem *et al.*, "A Measurement of the Correlation of Galaxy Surveys with CMB Lensing Convergence Maps from the South Pole Telescope," *Astrophys. J.*, vol. 753, p. L9, 2012. 61

[135] B. D. Sherwin *et al.*, "The Atacama Cosmology Telescope: Cross-Correlation of CMB Lensing and Quasars," *Phys. Rev.*, vol. D86, p. 083006, 2012. 61

[136] S. Ferraro, B. D. Sherwin, and D. N. Spergel, "WISE measurement of the integrated Sachs-Wolfe effect," *Phys. Rev.*, vol. D91, no. 8, p. 083533, 2015. 61

[137] R. Allison *et al.*, "The Atacama Cosmology Telescope: measuring radio galaxy bias through cross-correlation with lensing," *Mon. Not. Roy. Astron. Soc.*, vol. 451, no. 1, pp. 849–858, 2015. 61

[138] T. Giannantonio *et al.*, "CMB lensing tomography with the DES Science Verification galaxies," *Mon. Not. Roy. Astron. Soc.*, vol. 456, no. 3, pp. 3213–3244, 2016. 61

[139] Pullen, A.R., Alam, S., He, S., and Ho, S., "Constraining Gravity at the Largest Scales through CMB Lensing and Galaxy Velocities," *Mon. Not. Roy. Astron. Soc.*, vol. 460, no. 4, pp. 4098–4108, 2016. 61, 65

[140] C. Doux, E. Schaan, E. Aubourg, K. Ganga, K. Lee, D. N. Spergel, and J. Trguer, "First detection of cosmic microwave background lensing and Lyman- α forest bispectrum," *Phys. Rev.*, vol. D94, no. 10, p. 103506, 2016. 61

[141] S. Singh, R. Mandelbaum, and J. R. Brownstein, "Cross-correlating Planck CMB lensing with SDSS: Lensing-lensing and galaxy-lensing cross-correlations," *Mon. Not. Roy. Astron. Soc.*, vol. 464, no. 2, pp. 2120–2138, 2017. 61

[142] E. G. Geach and J. A. Peacock, "Cluster richnessmass calibration with cosmic microwave background lensing," *Nature Astronomy*, vol. 1, pp. 795–799, 2014. 61

[143] Chen, Y., Ho, S., Brinkmann, J., Freeman, P.E., Genovese, C.R., Schneider, D.P., and Wasserman, L., "Cosmic Web Reconstruction through Density Ridges: Catalogue," *Mon. Not. Roy. Astron. Soc.*, vol. 461, no. 4, pp. 3896–3909, 2016. 61, 71, 72, 75

[144] Alam, S. *et al.*, "The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III," *Astrophys. J. Suppl.*, vol. 219, no. 1, p. 12, 2015. 61

[145] Chen, Y., Ho, S., Freeman, P.E., Genovese, C.R., and Wasserman, L., "Cosmic Web Reconstruction through Density Ridges: Method and Algorithm," *Mon. Not. Roy. Astron. Soc.*, vol. 454, no. 1, pp. 1140–1156, 2015. 61, 72, 73, 74, 75

[146] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. I. R. Alves, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, H. Aussel, and et al., "Planck 2013 results. I. Overview of products and scientific results," *aap*, vol. 571, p. A1, Nov. 2014. 62

[147] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al., "Planck 2013 results. XVII. Gravitational lensing by large-scale structure," *aap*, vol. 571, p. A17, Nov. 2014. 62

[148] Clampitt, J., Jain, B., Takada, M., and Miyatake, H., "Detection of Stacked Filament Lensing Between SDSS Luminous Red Galaxies," *Mon. Not. Roy. Astron. Soc.*, vol. 457, no. 3, pp. 2391–2400, 2016. 63

[149] S. D. Epps and M. J. Hudson, "The Weak Lensing Masses of Filaments between Luminous Red Galaxies," *Mon. Not. Roy. Astron. Soc.*, vol. 468, pp. 2605–2613, Mar. 2017. 63

[150] Y. Higuchi, M. Oguri, and M. Shirasaki, "Statistical properties of filaments in weak gravitational lensing," *Mon. Not. Roy. Astron. Soc.*, vol. 441, no. 1, pp. 745–756, 2014. 63

[151] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al., "Planck 2013 results. XVI. Cosmological parameters," *aap*, vol. 571, p. A16, Nov. 2014. 63

[152] D. N. Limber, "The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field.," *The Astrophysical Journal*, vol. 117, p. 134, Jan. 1953. 64

[153] E. S. Rykoff *et al.*, "redMaPPer I: Algorithm and SDSS DR8 Catalog," *Astrophys. J.*, vol. 785, p. 104, 2014. 67

[154] A. de Graaff, Y.-C. Cai, C. Heymans, and J. A. Peacock, "Missing baryons in the cosmic web revealed by the Sunyaev-Zel'dovich effect," *ArXiv e-prints*, Sept. 2017. 68

[155] L. Wassermann, "All of nonparametric statistics," *Springer Science+ Business Media, New York*, 2017. 73

[156] Chen, Y., Ho, S., Mandelbaum, R., Bahcall, N.A., Brownstein, J.R., Freeman, P.E., Genovese, C.R., Schneider, D.P., and Wasserman, L., "Detecting Effects of Filaments on Galaxy Properties in the Sloan Digital Sky Survey III," *Mon. Not. Roy. Astron. Soc.*, vol. 466, p. 1880, 2017. 73

[157] Y.-C. Chen, C. R. Genovese, and L. Wasserman, "Asymptotic theory for density ridges," *The Annals of Statistics*, vol. 43, no. 5, pp. 1896–1928, 2015. 74

[158] Bagla, J.S., "A TreePM code for cosmological N-body simulations," *J. Astrophys. Astron.*, vol. 23, pp. 185–196, 2002. 76

[159] White, M.J., Hernquist, L., and Springel, V., "Simulating the Sunyaev-Zel'dovich effect(s): Including radiative cooling and energy injection by galactic winds," *Astrophys. J.*, vol. 579, p. 16, 2002. 76

[160] Reid, B.A., Seo, H., Leauthaud, A., Tinker, J.L., and White, M., "A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies," *Mon. Not. Roy. Astron. Soc.*, vol. 444, no. 1, pp. 476–502, 2014. 76

[161] Peacock, J.A. and Smith, R.E., "Halo occupation numbers and galaxy bias," *Mon. Not. Roy. Astron. Soc.*, vol. 318, p. 1144, 2000. 76

[162] Seljak, U., "Analytic model for galaxy and dark matter clustering," *Mon. Not. Roy. Astron. Soc.*, vol. 318, p. 203, 2000. 76

[163] Benson, A.J., Cole, S., Frenk, C.S., Baugh, C.M., and Lacey, C.G., "The Nature of galaxy bias and clustering," *Mon. Not. Roy. Astron. Soc.*, vol. 311, pp. 793–808, 2000. 76

[164] White, M.J., Hernquist, L., and Springel, V., "The Halo model and numerical simulations," *Astrophys. J.*, vol. 550, pp. L129–L132, 2001. 76

[165] Berlind, A.A. and Weinberg, D.H., "The Halo occupation distribution: Towards an empirical determination of the relation between galaxies and mass," *Astrophys. J.*, vol. 575, pp. 587–616, 2002. 76

[166] Cooray, A. and Sheth, R.K., "Halo models of large scale structure," *Phys. Rept.*, vol. 372, pp. 1–129, 2002. 76

[167] Beutler, F. *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Testing gravity with redshift-space distortions using the power spectrum multipoles," *Mon. Not. Roy. Astron. Soc.*, vol. 443, no. 2, pp. 1065–1089, 2014. 77

[168] White, M., Tinker, J.L., and McBride, C. K., "Mock galaxy catalogues using the quick particle mesh method," *Mon. Not. Roy. Astron. Soc.*, vol. 437, no. 3, pp. 2594–2606, 2014. 77

[169] Carlson, J. and White, M., "Embedding realistic surveys in simulations through volume remapping," *Astrophys. J. Suppl.*, vol. 190, pp. 311–314, 2010. 77

[170] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[171] T. Kipf, E. Fetaya, K.-C. Wang, M. Welling, and R. Zemel, "Neural relational inference for interacting systems," *arXiv preprint arXiv:1802.04687*, 2018.

[172] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*. 1981.

[173] J. S. Bagla, "Cosmological n-body simulation: Techniques, scope and status," *Current Science*, vol. 88, pp. 1088–1100, 04 2005.

[174] F. R. Bouchet, S. Colombi, E. Hivon, and R. Juszkiewicz, "Perturbative Lagrangian approach to gravitational instability.," *Astronomy and Astrophysics*, vol. 296, p. 575, Apr. 1995.

[175] P. Catelan, "Lagrangian dynamics in non-flat universes and non-linear gravitational evolution," *Monthly Notices of the Royal Astronomical Society*, vol. 276, pp. 115–124, Sept. 1995.

[176] N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak, and Z. Slepian, "nbodykit: an open-source, massively parallel toolkit for large-scale structure," *The Astronomical Journal*, vol. 156, no. 4, p. 160, 2018.

[177] R. Scoccimarro, "Transients from initial conditions: a perturbative analysis," *Monthly Notices of the Royal Astronomical Society*, vol. 299, pp. 1097–1118, Oct. 1998.

[178] A. Klypin and J. Holtzman, "Particle-Mesh code for cosmological simulations," *ArXiv e-prints*, 1997.

[179] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., and Berg, A. C., "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, pp. 211–252, 2015.

[180] D. Ribli, rmin Pataki, B., and Csabai, I., "Learning from deep learning: better cosmological parameter inference from weak lensing maps," *ArXiv e-prints*, 2018.

[181] S. Mallat, "Understanding deep convolutional networks," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, p. 20150203, Apr. 2016.

[182] van Engelen, A. *et al.*, "The Atacama Cosmology Telescope: Lensing of CMB Temperature and Polarization Derived from Cosmic Infrared Background Cross-Correlation," *Astrophys. J.*, vol. 808, no. 1, p. 7, 2015.

[183] Mead, J.M.G., King, L.J., and McCarthy, I.G., "Probing the cosmic web: inter-cluster filament detection using gravitational lensing," *Mon. Not. Roy. Astron. Soc.*, vol. 401, p. 2257, 2010.

[184] Maturi, M. and Julian, M. *A&A*, vol. 559, no. A112, 2013.

[185] Dietrich, J.P., Schneider, P., Clowe, D., Romano-Diaz, E., and Kerp, J., "Weak lensing evidence for a filament between the clusters A 222 and A 223 and its quantification," *Astron. Astrophys.*, vol. 440, pp. 453–471, 2005.

[186] Bond, N.A., Strauss, M.A., and Cen, R., "Crawling the Cosmic Network:

Identifying and Quantifying Filamentary Structure," *Mon. Not. Roy. Astron. Soc.*, vol. 409, p. 156, 2010.

[187] Sousbie, T., Pichon, C., Courtois, H. M., Colombi, S., and Novikov, D., "The 3D skeleton of the sdss," *Submitted to: Astrophys. J.*, 2006.

[188] Aragon-Calvo, M.A., Platen, E., van de Weygaert, R., and Szalay, A.S., "The Spine of the Cosmic Web," *Astrophys. J.*, vol. 723, pp. 364–382, 2010.

[189] Chen, Y., Ho, S., Tenneti, A., Mandelbaum, R., Croft, R., DiMatteo, T., Freeman, P.E., Genovese, C.R., and Wasserman, L., "Investigating Galaxy-Filament Alignments in Hydrodynamic Simulations using Density Ridges," *Mon. Not. Roy. Astron. Soc.*, vol. 454, no. 3, pp. 3341–3350, 2015.

[190] Ozertem, U. and Erdogmus, D., "Locally defined principal curves and surfaces," *J. Mach. Learn. Res.*, vol. 12, pp. 1249–1286, jul 2011.

[191] Gorski, K.M., Hivon, E., Banday, A.J., Wandelt, B.D., Hansen, F.K., Reinecke, M., and Bartelman, M., "HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere," *Astrophys. J.*, vol. 622, pp. 759–771, 2005.

[192] Ade, P. A. R. *et al.*, "Measurement of the Cosmic Microwave Background Polarization Lensing Power Spectrum with the POLARBEAR experiment," *Phys. Rev. Lett.*, vol. 113, p. 021301, 2014.

[193] Baxter, E.J. *et al.*, "A Measurement of Gravitational Lensing of the Cosmic Microwave Background by Galaxy Clusters Using Data from the South Pole Telescope," *Astrophys. J.*, vol. 806, no. 2, p. 247, 2015.

[194] Chen, Y., "A Tutorial on Kernel Density Estimation and Recent Advances," *ArXiv e-prints*, Apr. 2017.

[195] V. Springel, "The cosmological simulation code GADGET-2," *Monthly Notices of the Royal Astronomical Society*, vol. 364, pp. 1105–1134, Dec 2005.

[196] R. Teyssier, "Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES," *Astronomy and Astrophysics*, vol. 385, pp. 337–364, Apr. 2002.

[197] M. Takada and B. Jain, "The three-point correlation function in cosmology," *Monthly Notices of the Royal Astronomical Society*, vol. 340, pp. 580–608, Apr 2003.

[198] F. Governato, A. Babul, T. Quinn, P. Tozzi, C. M. Baugh, N. Katz, and G. Lake, "Properties of galaxy clusters: mass and correlation functions," *Monthly Notices of the Royal Astronomical Society*, vol. 307, pp. 949–966, Aug 1999.

[199] J. Schmelzle, A. Lucchi, T. Kacprzak, A. Amara, R. Sgier, A. Réfrégier, and T. Hofmann, "Cosmological model discrimination with Deep Learning," *arXiv e-prints*, p. arXiv:1707.05167, Jul 2017.

[200] T. Ciodaro, D. Deva, J. M. de Seixas, and D. Damazio, "Online particle detection with neural networks based on topological calorimetry information," *J. Phys. Conf. Ser.*, vol. 368, p. 012030, 2012.

[201] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, p. 436444, 2015.