NAGOYA UNIVERSITY

DOCTORAL THESIS

A Study on Enhancing Sensitivity to the Elements beyond the Standard Cosmological Model Using Machine Learning

(機械学習を用いた解析による標準宇宙モデルを超えた要素への感度向上)

Author:
Koya MURAKAMI

Supervisor:
Kiyotomo ICHIKI

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Science

in the

Cosmology Group
Department of Particle Physics and Astrophysics
Graduate School of Science
Nagoya University

March 25, 2024

Expected graduation: March 2024

Contents

Abstract

In this thesis, we investigate the elements beyond the ΛCDM model. The ΛCDM model consists of two assumptions; Λ is the cosmological constant and is related to dark energy, which causes the accelerated expansion of the universe, and cold dark matter (CDM), which is the dominant component of matter in the universe at present. However, the properties of them are still unknown, and there is room for modifications.

To investigate beyond ΛCDM, we consider two quantities: 1) the growth rate and 2) the mass of non-cold dark matter (NCDM), which are related to the accelerated expansion of the universe and the property of dark matter, respectively.

The growth rate corresponds to the time evolution of the density perturbation of matter. It is an important probe of gravity because the evolution of the matter distribution is driven by gravity. Source of modified gravity theories can explain the origin of the accelerated expansion. The growth rate is measured by the analysis of the large-scale structure of the universe and is sensitive to the modification of gravity.

Dark matter is one of the components of the universe. In the ΛCDM model, dark matter is non-relativistic matter when generated and is called CDM. The Λ CDM model is consistent with the observations at large scales. However, the ΛCDM model has problems at small scales, so-called small-scale crisis. The NCDM model is one of the solutions to these problems. NCDM is relativistic when generated and becomes non-relativistic matter as the universe expands and cools. An important property of NCDM is its velocity dispersion, which prevents structure evolution at small scales depending on its mass. Conversely, we can examine the mass of NCDM by the analysis of the large-scale structure of the universe.

The growth rate and the mass of NCDM can be measured by analyzing the large-scale structure. Traditionally, two-point statistics, such as the power spectrum of the large-scale structure, are used for the analysis. In the near future, we will obtain a large amount of high-sensitivity observation data. Therefore, we need to develop new methods to study these large amounts of data more effectively and accurately. In this thesis, we focus on the machine learning approach and conduct two works for the measurements of the growth rate and the mass of NCDM.

For the growth rate measurement, we use convolutional neural networks (CNN) trained by the Quijote N-body simulations to predict the growth rate for the simulation data. In general, we need a model for the non-linear part of the observed power spectrum to predict the growth rate. By using CNN, we aim to predict the growth rate directly from the matter distribution without non-linear modeling. As a result, we find that the predictions of the growth rate from our CNN are in good agreement with the fiducial values that the simulation assumes. The prediction errors are within a factor of order unity from those of the traditional optimistic Fisher approach. Therefore, we conclude that CNN can provide growth rate predictions without non-linear modeling on small scales.

For the mass of NCDM, we use CNN trained by the cosmological hydrodynamic simulation data to constrain the mass of NCDM. We focus on the HI distribution as a tracer of the dark matter distribution because the HI distribution can be observed as the 21cm signal using radio frequency observations. In this work, we conduct the hydrodynamic simulations for both the CDM model and NCDM models with different masses using `GADGET3-Osaka` and generate images of 21cm signals. We use CNN to classify the images of the 21cm maps in CDM and NCDM models and evaluate the classification results. As a result, we find that CNN outperforms the

power spectrum analysis in our study. In addition, we investigate the effects of different astrophysical models and the system noise assuming SKA-MID.

From these works, we demonstrate the potential of the machine learning approach for the analysis of the large-scale structure of the universe.

Acknowledgements

First, I would like to express my sincere thanks to my supervisor, Kiyotomo Ichiki, for his support, advice and discussion about my researches. I would like to thank Atsushi J. Nishizawa also. He always gave me the advice for all about my researches such as not only its plans, conducts, and sophistication but also the presentation of it. I would like to thank him for his collaboration for my researches also. I am grateful to Naoshi Sugiyama for his teaching me the basics of the cosmology and the fruitful discussion about my research.

I would like to thank my collaborators, Kentaro Nagamine and Ikko Shimizu, who are collaborators of my research for the constraints on the dark matter mass, They gave me the advice and the code for the hydrodynamic simulation, and helped me to interpret the results of the simulation and analysis. I am grateful to Indira Ocampo, Savvas Nesseris, and Sachiko Kuroyanagi, who are collaborators of my research about $f\sigma_8$ estimation. They gave me the idea of the research, and the essential analysis and discussion required to conduct my research. They helped also my stay for my research in Madrid.

I would like to express my thanks to past and present C-lab members for their advices and discussions in seminars and coffee breaks. They helped me to conduct my researches and I enjoyed the discussions with them.

I would like to take this opportunity to thank the "Nagoya University Interdisciplinary Frontier Fellowship" supported by Nagoya University and JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2120.

Part of the computation is performed on Cray XC50 and GPU cluster at the CfCA in NAOJ and the GPU workstation at Nagoya University. For introducing GPU workstation at Nagoya university, I would like to thank Atsushi J. Nishizawa and Hironao Miyatake.

Finally, I would like to express my sincere thanks to my family for supporting in all my life.

Chapter 1

Introduction

The ΛCDM model is widely accepted because it can account for the results of most observations. This model assumes two components in the universe: the cosmological constant Λ, which is related to dark energy and causes the accelerated expansion of the universe, and cold dark matter (CDM). However, the properties of these components are still unknown. To unveil them, we require more highly accurate observational data and methods for their analysis.

In this work, we focus on machine learning, especially the Neural Network (NN), as a method of analysis. NN is a type of machine learning used for the analysis of big data. This algorithm can be used to extract information from labeled data without specifying the features of data manually by a human. The traditional analysis techniques for the large-scale structure (LSS) of the universe such as the power spectrum, which is the two-point correlation of the matter distribution in Fourier space, can only obtain a part of the information in the observed data. In contrast, a machine learning algorithm can extract more complex information from the data and use various features beyond the two-point statistics. For example, the Convolutional Neural Network (CNN), which is an algorithm for image analysis, has been used to constrain cosmological parameters from the simulated weak lensing data (Ribli et al., 2019; Ribli, Pataki, and Csabai, 2019) or the simulated large-scale structure of dark matter distribution (Pan et al., 2020). Other examples use U-Net to detect the Sunyaev-Zel'dovich effect from Planck cosmic microwave background (CMB) data (Bonjean, 2020), distinguish modified gravity models from the standard model using CNN (Peel et al., 2019), and use NN to reconstruct the initial conditions of the universe from the galaxy distribution (Modi, Feng, and Seljak, 2018).

Here, we focus on the growth rate of the matter density perturbation related to dark energy and the mass of dark matter. These quantities are related to the statistics of the LSS of the universe. We use machine learning algorithms to investigate the LSS and compare the results with those from traditional analysis techniques. In the following subsection, we provide more details about the quantity we focus on.

1.1 Growth Rate

Recent cosmological observations indicate the expansion of the universe is accelerated and it is considered that this acceleration is typically caused by dark energy, specifically the cosmological constant in the ΛCDM model, which may cause the fine-tuning issues (Padmanabhan, 2003; Amendola, Marra, and Quartin, 2013). Various candidates for dark energy or alternative explanations for the accelerated expansion have been proposed beyond the cosmological constant. For example, quintessence (Fujii, 1982; Wetterich, 1988; Chiba, Sugiyama, and Nakamura, 1997) is one of the dark energy models that introduce a new scalar field beyond the Standard Model of particle physics. Another candidate is the modified gravity theory

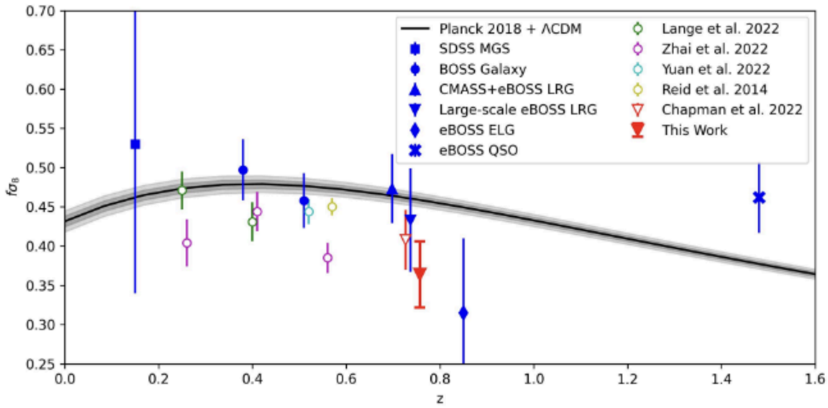

FIGURE 1.1: This figure shows the constraints on the growth rate quoted from a previous work (Chapman, Zhai, and Percival, 2023). The solid line represents the prediction of $f\sigma_8$ value for the ΛCDM model, and the plots show the constraints on $d\sigma_8$ by using the galaxy surveys.

(Tsujikawa, 2007; Joyce, Lombriser, and Schmidt, 2016), deviating from general relativity at scales corresponding to the size of galaxies or larger. One of the challenges in modern cosmology is understanding the properties of dark energy.

The LSS of the universe encodes important details for testing the cosmological model as it contains information about the late-time evolution of the Universe and the matter density distribution. The information is useful to constrain the fractional density parameters for CDM Ω_{cdm}, baryon Ω_{b}, dark energy Ω_Λ and other parameters such as the amplitude of the matter power spectrum σ_8. The LSS is also an important probe of dark energy and gravity theories because the dynamics of the LSS which is derived by gravity are strongly related to them.

One way of investigating the dynamics of the LSS is measuring the growth rate f of matter density perturbation, which is defined based on the time derivative of the matter density perturbation amplitude. In general, f is measured from galaxy survey data. However, galaxies are biased tracers of the matter distribution and f is sensitive to the bias parameter of galaxies b_g. So, we generally use $f\sigma_8$ as the bias-free parameter (see Section2.5), which can be measured, for example, through the two-point correlation of the galaxy distribution in redshift space (Song and Percival, 2009), the galaxy-clustering analysis using Effective Field Theory of the LSS (Carrasco et al., 2014; Carrilho, Moretti, and Pourtsidou, 2023), and the simulation-based emulator (Winther et al., 2019; Ramachandra et al., 2021; Brando et al., 2022).

Figure 1.1, which is quoted from Chapman, Zhai, and Percival, 2023, shows some constraints on the $f\sigma_8$ based on the LSS analysis ; the SDSS MGS (Howlett et al., 2015), BOSS galaxies (Alam et al., 2017), CMASS+eBOSS LRGs (Bautista et al., 2021), eBOSS ELGs (De Mattia et al., 2021), eBOSS quasars (Neveux et al., 2020). In addition, the constraints from some works (Chapman et al., 2022; Lange et al., 2022; Zhai et al., 2023; Yuan et al., 2022; Reid et al., 2014), which aim to analyze the small scales of the BOSS galaxy samples, are shown.

However, we need to make two assumptions to estimate $f\sigma_8$. Firstly, we need to specify a cosmological model to determine the relation between the redshift and the distance to a galaxy in order to compute the two-point correlation (Contreras et al., 2013). Secondly, we need to assume a model for the power spectrum including the non-linear part, typically done with phenomenological models (Scoccimarro and Frieman, 1996; Bernardeau, Rijt, and Vernizzi, 2012; Baldauf et al., 2015; Casas et al., 2017; Blanchard et al., 2020a). So, we want to develop a method to avoid these problems. For the first assumption, we can correct the dependence on the cosmology, for example, by Alcock-Paczynski type correction (Nesseris, Pantazis, and Perivolaropoulos, 2017). In this work, we try to develop a method to mitigate the second one.

In addition, we can expect more accurate data of the LSS from forthcoming cosmological observations such as *Euclid* (Laureijs et al., 2011), LSST (Abell et al., 2009), and DESI (Aghamousa et al., 2016). These observations data are rich in information in the LSS, so we need to develop statistical tools to use these data effectively.

To solve these problems, we focus on machine learning as the method to estimate $f\sigma_8$ without modeling the non-linear part of the power spectrum and the useful statistical tool for the rich LSS data. To demonstrate the performance of machine learning, we use the public N-body simulation data, Quijote simulation (Villaescusa-Navarro et al., 2020), and train and test our 3-dimensional CNN (3D-CNN) with this simulation data. The analysis with 3D-CNN has mainly two advantages. The first one is that the 3D-CNN can extract information from the matter density field directly. Here, we need assumptions only for the simulations. In this work, we use the data of Quijote simulation based on the ΛCDM model, so we need to assume the ΛCDM model. However, we do not need any assumptions for the non-linear part of the matter density field in estimating $f\sigma_8$. Secondly, we can estimate $f\sigma_8$ at high speed once we train CNN. CNN algorithm can simplify the computationally expensive procedures for data treatment (Lazanu, 2021a).

In this work, we aim to estimate the growth rate $f\sigma_8$ by 3D-CNN. Our 3D-CNNs are trained directly by the N-body simulations at different redshift bins. In addition, we use the power spectrum of these simulations, estimate $f\sigma_8$, and compare the estimated error of $f\sigma_8$ from 3D-CNN to that from the power spectrum analysis.

This work is published in Murakami et al., 2023b.

1.2 Mass of Dark Matter

Dark matter is one of the components of the universe and accounts for about 26% of the total energy density in the universe (Planck Collaboration et al., 2020). The existence of dark matter has been confirmed through various observations such as the motion of galaxies in a cluster (Zwicky, 1937) and the gravitational lensing effects (Natarajan et al., 2017). Dark matter interacts gravitationally with ordinary matter, while its interaction with the electromagnetic force is either absent or weak. Most of the properties of dark matter remain unknown and many dark matter models are suggested (Abdalla et al., 2019). In this work, we focus on the dark matter model in which dark matter is an elementary particle. One of the important properties of dark matter is its mass. In the ΛCDM model, dark matter is characterized as a heavy and non-relativistic, i.e., cold particle at its freeze-out. This type of dark matter is called Cold Dark Matter (CDM), and the CDM model is consistent with many observation results. However, the ΛCDM model does not specify any particular dark matter model and makes no assumptions about the mass of a dark matter particle.

Various dark matter models are suggested. Figure 1.2, which is quoted from Tuominen, 2021 shows the examples of dark matter models. Axion dark matter is produced by the vacuum realignment and is the ultralight particle (Abbott and Sikivie, 1983; Dine and Fischler, 1983; Preskill, Wise, and Wilczek, 1983). On the other hand, the weakly interacting massive particle (WIMP) dark matter is a heavy particle and its mass range is from 10 GeV to 100 TeV (Alvarez et al., 2020). In our

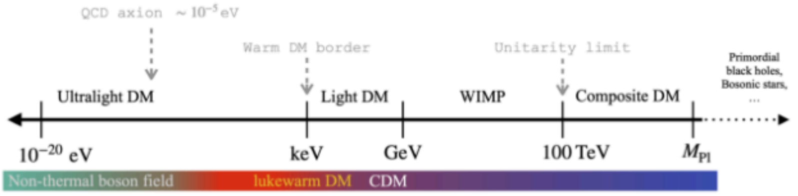

FIGURE 1.2: Examples of dark matter models. This figure is quoted from Figure 1 of the previous work (Tuominen, 2021).

work, we focus on the sterile neutrino, which is a fundamental right-handed particle added to the standard model and distinct from active neutrinos (a left-handed particle), as a warm dark matter model (Dodelson and Widrow, 1994).

The sterile neutrino dark matter model is one of the dark matter models with particle masses in the keV scale. In these models, dark matter is a relativistic particle at the time of freeze-out. And then, these dark matter particles have a velocity dispersion that depends on their mass. In this work, we refer to this type of dark matter as Non-Cold Dark Matter (NCDM) or Warm Dark Matter (WDM). The velocity dispersion of dark matter has an effect on the dark matter distribution at scales corresponding to the free-streaming length of dark matter. This effect is one of the solutions of the small-scale crises in the ΛCDM model such as the core-cusp and too-big-to-fail problems (Weinberg et al., 2015). These problems represent inconsistencies between observations and simulations (ΛCDM model) regarding the density profile of the halo and the halo abundance, respectively.

As mentioned earlier, the mass of dark matter affects its spatial distribution, allowing us to estimate the dark matter mass based on the dark matter distribution. For example, the power spectrum of Lyman-α forest, which traces the dark matter distribution, has been used to constrain the dark matter particle mass. Previous studies have shown that the lower limit of the dark matter mass is on the order of 1 keV (Viel et al., 2013; Garzilli et al., 2021; Garzilli et al., 2019; Villasenor et al., 2023). However, a more stringent constraint on the dark matter mass is required to explore various dark matter models. Therefore, it is important to develop new methods that can extract more complex information from the LSS than the power spectrum. In this work, we focus on the machine learning techniques to access the information including three- or higher-order statistics.

As we cannot see the dark matter directly, we need to see the tracer of the dark matter distribution in practical observations. In this work, we use the 21cm radiation emitted from neutral hydrogen (HI) due to the hyperfine splitting. The HI is one of the tracers of the dark matter distribution and the 21cm radiation can be observed by radio telescopes. A lot of ongoing or planned observations of the HI such as the Murchison Wide-field Array (MWA) (Tingay et al., 2013), Canadian Hydrogen Intensity Mapping Experiment (CHIME) (Bandura et al., 2014), Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX) (Newburgh et al., 2016), and Square Kilometer Array (SKA) (Santos et al., 2015), contribute to our understanding of the HI distribution. These surveys will provide us with the information of the HI distribution and we can use it as a tool to constrain the dark matter mass. For example, a previous study (Carucci et al., 2015) showed the forecasts of the constraint on the dark matter mass by the SKA observation.

In this work, our goal is to demonstrate the potential of CNN in constraining the dark matter mass. Our focus is on the HI distribution at redshift $z = 3$ during the post-reionization epoch, as will be observed by SKA. During the reionization epoch, the HI distribution is significantly influenced by ionization processes, so we focus on the post-reionization. The previous study (Bauer et al., 2021) discussed the HI power spectrum for the 21cm intensity mapping, which is modeled by using the N-body simulation and assuming the properties of the HI halo and forecasted the improvement of the constraints on the axion dark matter. As an example of applying machine learning to the cosmological analysis, the previous work (**2024MNRAS.527..739R**) investigated the constraints on the warm dark matter mass by using CNN based on the N-body simulation and showed that CNN analysis can outperform the analysis by the two-point statistics. In our approach, we consider a map-based analysis with CNN to extract additional information to the power spectrum included in the HI map generated from the cosmological hydrodynamic simulation including models such as the star formation, supernova feedback, UV radiation background, and radiative cooling and heating.

To test our CNN analysis, we run hydrodynamic simulations using the CDM and NCDM models with various masses, generating corresponding HI maps. Subsequently, we use CNN and the power spectrum to classify the data assuming different dark matter models, and we compare the results.

Furthermore, we consider two effects that exist in practical observations. Firstly, we consider the different astrophysical assumptions such as the self-shielding effect of HI gas, the star formation effect, and the UV background model. In this thesis, we refer to these assumptions as the astrophysical model. These models have an effect on the ionization of hydrogen, so it is probable that they affect the results of our analysis. For example, in the work by (Villanueva-Domingo and Villaescusa-Navarro, 2021), they removed the astrophysical effect from the map of the 21cm signal and created the map of the matter distribution. On the other hand, our work uses the map with the astrophysical effects directly to constrain the dark matter mass. Secondly, we consider the system noise that exists in the SKA observation. This noise contaminates the map of the 21cm signals, affecting our analysis. We generate the map data contaminated by the mock noised map assuming the SKA observation, and investigate the effect of the noise on our analysis.

This work is published in Murakami et al., 2023a.

This thesis is organized as follows. In Chapter 2, we present the theory for the evolution of the matter distribution, the statistics representing the properties of the matter distribution, the redshift space distortion related to the growth rate, NCDM models, and the 21cm signal. Chapter 3 introduces the algorithm of machine learning we focus on. Chapter 4 and 5 present our works for estimating $f\sigma_8$ and constraining dark matter mass, respectively. Finally, in Chapter 6 we summarize this thesis. Unless otherwise described, we use natural units and consider a flat universe in this thesis.

Chapter 2

Theoretical Framework

In this chapter, we introduce the theoretical frameworks for our work. Section 2.1-2.5 introduces the standard cosmology, the evolution and statistics of the LSS, and the effects of the practical observations. In the discussion of these sections refer to (Matsubara, 2014). Section 2.6 and 2.7 provide a brief introduction to the modified gravity theory and NCDM models for the preparation of our work. Section 2.8 explains the observable we focus on in Chapter 5.

2.1 FLRW Metric

Firstly, we consider the homogeneous and isotropic spacetime. The homogenous and isotropic metric is described by the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric, and its line element ds^2 is expressed as,

$$ds^2 = -dt^2 + a^2(t)\gamma_{ij}dx^i dx^j, \tag{2.1}$$

where a is the scale factor that represents the size of the universe, γ_{ij} is the 3-dimensional, homogenous, and isotropic metric, where i and j take values of 1, 2, or 3, representing the components of 3-dimensional space. In our work, we normalize the scale factor such that $a(t_0) = 1$, where t_0 represents the present time. When we consider the spherical coordinate for 3-dimensional space, the line element is written as

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{\sqrt{1-Kr^2}} + r^2(d\theta^2 + \sin\theta d\phi^2)\right], \tag{2.2}$$

where r is the radial coordinate scaled with the cosmic expansion, θ and ϕ are the angular coordinates, and K describes the spatial curvature; $K = 1, 0,$ and -1 correspond to a closed, flat, and open universe, respectively. In the following, we set $d\theta = d\phi = 0$ since the spacetime is homogenous and isotropic, and $K = 0$, i.e., the universe is flat. The trajectory of a photon in the universe is described as $ds = 0$, so we can derive the following relation for the photon trajectory from Eq. (2.2),

$$\frac{dt}{a(t)} = -d\chi, \tag{2.3}$$

where $d\chi$ is defined as $d\chi = dr/\sqrt{1-Kr^2}$, and χ is the comoving distance representing the distance measured at $t = t_0$. And then, in the case that the photon is emitted from $\chi = \chi$ at $t = t$ and reaches us at $\chi = 0$ at the present time t_0, we derive the relation as

$$\int_t^{t_0} \frac{dt}{a(t)} = \int_0^{\chi} d\chi. \tag{2.4}$$

By integrating, we can calculate the comoving distance χ to the source of the photon as

$$\chi = \int_t^{t_0} \frac{dt}{a(t)} = \int_a^1 \frac{da}{a^2 H} = \int_0^z \frac{dz}{H} \ , \tag{2.5}$$

where $H = \dot{a}/a$ is the Hubble parameter representing the expansion rate of the universe, and z is a redshift representing an increase in the wavelength caused by the expansion of the universe, which satisfies $1 + z = 1/a$. Note that we express the time derivative of a quantity A as $\dot{A}$.

2.2 Friedmann Equation

In general relativity, the equation for the gravity is expressed by the Einstein equation

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{2.6}$$

where $G_{\mu\nu}$ is the Einstein tensor, μ and ν take values of 0, 1, 2, or 3 and represent the components of the 4-dimensional spacetime, Λ is a cosmological constant, $g_{\mu\nu}$ is a metric, G is the gravitational constant, and $T_{\mu\nu}$ is the energy-momentum tensor. For the FLRW metric, we can derive two equations from the Einstein equation, and they are written as

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} \ , \tag{2.7}$$

$$\left(\frac{\ddot{a}}{a}\right) = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3} \ , \tag{2.8}$$

where $\ddot{A}$ represents the second time derivative of a quantity A, ρ is the energy density of the matter and radiation components of the universe, and P is the pressure of the radiation component. Eq. (2.7) is called as Friedmann equation. In addition, we can derive the equation from the above two equations,

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0. \tag{2.9}$$

Based on Eq. (2.7), we define the following parameters;

$$H_0 = \dot{a}(t_0), \quad \Omega_0 = \frac{8\pi G}{3H_0^2}, \quad \Omega_{\Lambda 0} = \frac{\Lambda}{3H_0^2} \ , \tag{2.10}$$

where H_0 is the Hubble parameter at the present time referred to as the Hubble constant, Ω_0 is the density parameter representing the rate of the energy density of the matter and radiation components to the total energy density in the universe, and Ω_Λ is the cosmological constant parameter, which is the nondimensionalized value of Λ. We can rewrite the Friedmann equation (2.7) at the present time by using Ω_0 and Ω_Λ as

$$\Omega_0 + \Omega_{\Lambda 0} = 1. \tag{2.11}$$

The dimensionless Hubble constant,

$$h = \frac{H_0}{100 \text{ km s}^{-1}\text{Mpc}^{-1}} \ , \tag{2.12}$$

is often used to express the value of the Hubble constant or the density parameter in the form of Ωh^2. The parameters h (or H_0), Ω_0, and Ω_Λ are part of the cosmological parameters which specify the cosmological model.

When matter is the dominant component in the universe, we can derive the evolution of matter density from Eq. (2.8) by setting $P = 0$ as $\rho \propto a^{-3}$. And then, the Friedmann equation (2.7) during the matter-dominated epoch is rewritten as

$$H^2 = H_0^2 \left(\frac{\Omega_{m0}}{a^3} + \Omega_\Lambda \right), \tag{2.13}$$

where the Ω_{m0} is the density parameter of the matter at the present time defined as

$$\Omega_{m0} = \frac{8\pi G \rho_{m0}}{3H_0^2}, \tag{2.14}$$

where ρ_{m0} is the energy density of the matter at the present time. From Eq. (2.13), we can describe the time dependence of the Hubble parameter by the function of the redshift z as,

$$H(z) = H_0 \sqrt{\Omega_{m0}(1+z)^3 + \Omega_{\Lambda 0}}. \tag{2.15}$$

From this equation and Eq. (2.5), we can calculate the comoving distance depending on the value of the cosmological parameters.

When we define two quantities as

$$\rho_\Lambda = \frac{\Lambda}{8\pi G} \ , \quad P_\Lambda = -\frac{\Lambda}{8\pi G} \ , \tag{2.16}$$

ρ_Λ and P_Λ correspond to the energy density and the pressure of the cosmological constant in Eq. (2.8). Then, we can deal with the cosmological parameter as the energy component which has the constant density and the negative pressure for $\Lambda > 0$. In Eq. (2.8), $\ddot{a}$ is positive when $P < -\rho/3$ and $\Lambda > 0$. Therefore, the cosmological constant causes the accelerated expansion of the universe by its negative pressure. The energy component that causes the accelerated expansion, such as the cosmological constant, is referred to as dark energy. By defining the equation of state parameter of dark energy as $w = P_{\rm DE}/\rho_{\rm DE}$, where $P_{\rm DE}$ and $\rho_{\rm DE}$ represent the pressure and energy density of dark energy, respectively, w should be $< -1/3$ for the accelerated expansion, e.g., $w = -1$ for the cosmological constant.

2.3 Evolution of Density Perturbation

In this section, we explore the evolution of matter density perturbations. In the early universe, the matter distribution is nearly homogeneous. However, slight inhomogeneities in the matter distribution lead to density perturbations around the mean density, giving rise to the structures observed in the present universe. As matter moves from lower-density to higher-density regions due to gravitational instability, these density perturbations grow. A plausible theory explaining the origin of these initial perturbations is inflation theory, which posits the exponential expansion of the universe in its early stages, offering a coherent explanation for various observations.

In the following, we approximate the matter distribution by the non-relativistic Newtonian fluid. For spatial coordinates $\boldsymbol{r}$ defined by a physical distance, the fluid follows the equations below;

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial \boldsymbol{r}} \cdot (\rho \boldsymbol{u}) = 0, \tag{2.17}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + \left(\boldsymbol{u} \cdot \frac{\partial}{\partial \boldsymbol{r}} \right) \boldsymbol{u} = -\frac{1}{\rho} \frac{\partial P}{\partial \boldsymbol{r}} - \frac{\partial \phi}{\partial \boldsymbol{r}}, \tag{2.18}$$

where $\rho(\boldsymbol{r}, t)$ is the mass density, $P(\boldsymbol{r}, t)$ is the pressure, $\boldsymbol{u}(\boldsymbol{r}, t)$ is the velocity vector, and $\phi(\boldsymbol{r}, t)$ is the gravitational potential. Eq. (2.17) is the continuity equation representing the mass conservation, and Eq. (2.18) is the Euler equation which is the equation of motion of the fluid element.

These equations are described in the physical rest-frame coordinates $\boldsymbol{r}$. Next, in order to describe the motion of a fluid in the expanding universe, we rewrite Eq. (2.17) and Eq. (2.18) by the comoving coordinates $\boldsymbol{x}$, which expand with cosmic expansion, defined as $\boldsymbol{r} = a(t)\boldsymbol{x}$ with a scale factor a. When the position of a fluid element is expressed by $\boldsymbol{r}(t)$, its velocity is described by $\boldsymbol{u} = \dot{\boldsymbol{r}} = a\dot{\boldsymbol{x}} + \dot{a}\boldsymbol{x}$. To disentangle the effects of cosmic expansion from the velocity caused by local motions, we define the velocity in comoving coordinates as

$$\boldsymbol{v} \equiv \dot{\boldsymbol{r}} - \dot{a}\boldsymbol{x} = a\dot{\boldsymbol{x}}, \tag{2.19}$$

which means the relative velocity to the comoving rest-frame and $\boldsymbol{v}$ is called a peculiar velocity to discriminate the velocity caused by the cosmic expansion. The relation between $\boldsymbol{v}$ and $\boldsymbol{u}$ is described as $\boldsymbol{u} = \dot{a}\boldsymbol{x} + \boldsymbol{v}$. Here, we can rewrite the spatial derivative for the coordinates transformation $(\boldsymbol{r}, t) \to (\boldsymbol{x}, \mathrm{t})$;

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} - \left(\frac{\dot{a}}{a} \right) \boldsymbol{x} \cdot \boldsymbol{\nabla}, \quad \frac{\partial}{\partial \boldsymbol{r}} \to \frac{1}{a} \boldsymbol{\nabla}, \tag{2.20}$$

where $\boldsymbol{\nabla} = \partial / \partial \boldsymbol{x}$. Therefore, we rewrite Eq. (2.17) and Eq. (2.18) in the comoving coordinates as

$$\frac{\partial \rho}{\partial t} + 3\frac{\dot{a}}{a}\rho + \frac{1}{a}\boldsymbol{\nabla} \cdot (\rho \boldsymbol{v}) = 0, \tag{2.21}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + \frac{\dot{a}}{a}\boldsymbol{v} + \frac{1}{a}(\boldsymbol{v} \cdot \boldsymbol{\nabla})\boldsymbol{v} = -\frac{1}{a\rho}\boldsymbol{\nabla}P - \frac{1}{a}\boldsymbol{\nabla}\Phi, \tag{2.22}$$

where Φ defined as

$$\Phi = \phi + \frac{1}{2} a\ddot{a} |\boldsymbol{x}|^2, \tag{2.23}$$

which is the gravitational potential in the comoving frame. These equations help isolate the effects of cosmic expansion and provide a clearer understanding of the fluid dynamics in an expanding universe.

In a physical rest frame, the gravitational potential ϕ follows the Poisson equation derived from Einstein equation for Newtonian limit;

$$\triangle \phi = 4\pi G \rho - \Lambda, \tag{2.24}$$

where $\triangle \phi$ represents Laplacian. The mean matter density $\bar{\rho}$ in the universe satisfies the below equation from Eq. (2.8) during the matter-dominated epoch;

$$\left(\frac{\ddot{a}}{a} \right) = -\frac{4\pi G}{3}\bar{\rho} + \frac{\Lambda}{3}, \tag{2.25}$$

where the mean pressure $\bar{p}$ is much less than $\bar{\rho}$ during the matter dominated epoch. And then, in the comoving frme, Eq. (2.24) is written as

$$\triangle \Phi = 4\pi G a^2 (\rho - \bar{\rho}). \tag{2.26}$$

The mass density perturbation δ and pressure perturbation δP, which are the small deviations of the matter density and pressure from the homogenous background, respectively, are defined as

$$\delta(x, t) \equiv \frac{\rho(x, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \quad \delta P(x, t) \equiv P(x, t) - \bar{P}(t). \tag{2.27}$$

Finally, we can rewrite Eq. (2.21), Eq. (2.22), and Eq. (2.26) as follows:

$$\dot{\delta} + \frac{1}{a} \boldsymbol{\nabla} \cdot [(1 + \delta) v] = 0, \tag{2.28}$$

$$\dot{v} + \frac{\dot{a}}{a} v + \frac{1}{a} (v \cdot \boldsymbol{\nabla}) v = \frac{1}{a \bar{\rho} (1 + \delta)} \boldsymbol{\nabla} \delta P - \frac{1}{a} \boldsymbol{\nabla} \Phi, \tag{2.29}$$

$$\triangle \Phi = 4\pi G a^2 \bar{\rho} \delta.. \tag{2.30}$$

From these equations, we can calculate the time evolution of the matter density perturbation in the expanding universe. In addition, the following equation derived from Eq. (2.28)-(2.30) is often used;

$$\begin{aligned} &\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - 4\pi G \bar{\rho} (1 + \delta) \delta = \\ &\qquad \frac{\triangle \delta P}{\bar{\rho} a^2} + \frac{1}{a^2} \boldsymbol{\nabla} \delta \cdot \boldsymbol{\nabla} \Phi + \frac{1}{a^2} \sum_i \sum_j \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^j} \left[(1 + \delta) v^i v^j \right], \end{aligned} \tag{2.31}$$

where x^i and v^i are the i-th component of the comoving coordinates and the peculiar velocity, respectively.

Next, we consider the linear evolution of the density perturbation of dark matter. In the late universe, which we focus on in our work, the non-relativistic matter (dark matter and baryon) is the dominant energy component relative to radiation. In the standard cosmology, dark matter is a pressureless particle, which is cold when it is generated. This kind of dark matter is called as cold dark matter (CDM). We can approximate CDM as a pressureless and non-relativistic fluid. Therefore, the evolution of the density perturbation of CDM is written as

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - 4\pi G \bar{\rho} \delta = 0, \tag{2.32}$$

which derived from Eq. (2.31), where the righthand-side is zero since we ignore the second order of the perturbation and dark matter is pressureless. When dark energy is modeled as a cosmological constant, we can derive the time evolution of the Hubble parameter $H = \dot{a}/a$ from the time derivative of Eq. (2.25) and $\rho \propto a^{-3}$;

$$\ddot{H} + 2H\dot{H} - 4\pi G \bar{\rho} H = 0. \tag{2.33}$$

By comparing Eq. (2.32) with Eq. (2.33), we can see $\delta \propto H$ is a particular solution of Eq. (2.32). And then, we put $\delta = H y(t)$, where $y(t)$ is an unknown function depending on time, to find the other solution. From Eq. (2.32), $y(t)$ satisfies the below differential equation;

$$\dot{a}\ddot{y} + 2\ddot{a}\dot{y} = 0. \tag{2.34}$$

Therefore, we can derive $\dot{y} \propto \dot{a}^{-2}$ by integrating Eq. (2.34). Subsequently, we find

$$y \propto \int_0^t \frac{dt}{\dot{a}^2} = \int_0^a \frac{da}{\dot{a}^3} = \int_0^a \frac{da}{a^3 H^3}. \tag{2.35}$$

From the above, we find that two independent solutions D_- and D_+ of Eq. (2.32) are

$$D_- \propto H, \quad D_+ \propto H \int_0^a \frac{da}{a^3 H^3}, \tag{2.36}$$

where the Hubble parameter is described as the function of the scale factor $a = 1/(1+z)$ by

$$H(a) = H_0 \sqrt{\frac{\Omega_{m0}}{a^3} + \Omega_{\Lambda 0}}. \tag{2.37}$$

When the matter-dominated epoch, the term of Ω_{m0}/a^3 is dominant in Eq. (2.37). Therefore, D_- is $\propto a^{-3/2}$ and vanishes with the expansion of the universe. On the other hand, D_+ is $\propto a$ and grows with the expansion.

After enough time in the matter-dominated epoch, the term proportional to D_- vanishes, and we can describe the time evolution of the matter density perturbation δ by D_+. The linear growth factor, D_+, quantifies the growth of density perturbations in the linear regime. In the case that the linear growth factor is normalized by its present value $D_+(t_0)$ and expressed as $D(t) \equiv D_+(t)/D_+(t_0)$, the time evolution of the matter density perturbation δ in the linear regime is described as

$$\delta(x, t) = D(t)\delta_0(x), \tag{2.38}$$

where δ_0 is the value of the matter density perturbation in the linear regime at the present time.

For Eq. (2.28) and Eq. (2.29), we consider the peculiar velocity v. The rotation of Euler equation (2.29) in the linear regime is

$$\frac{\partial}{\partial t}(a \boldsymbol{\nabla} \times v) = 0. \tag{2.39}$$

Therefore, the rotation of the peculiar velocity vector $\boldsymbol{\nabla} \times v$ is proportional to the inverse of the scale factor a^{-1}. After enough time, the rotation of v decays and we can describe the peculiar velocity as

$$v = -\frac{\boldsymbol{\nabla}\psi}{aH}, \tag{2.40}$$

where $\psi(x, t)$ is a scalar function. From Eq. (2.40) and the continuity equation (2.28) in the linear regime, we find

$$\dot{\delta} = \frac{\triangle\psi}{a^2 H}. \tag{2.41}$$

By combining this equation with the time evolution of the density perturbation (2.38), the below equation is derived;

$$\triangle\psi = a^2 H^2 f \delta, \tag{2.42}$$

where f is the linear growth rate representing the time evolution of the linear growth factor $D(t)$, and it is defined as

$$f(t) = \frac{d\ln D}{d\ln a} = \frac{\dot{D}}{HD} = \frac{d\ln\delta}{d\ln a}. \tag{2.43}$$

The linear growth rate is an important quantity to discuss the redshift space distortion since it is related to the peculiar velocity (see Section 2.5). In addition, it is one of the probes of gravity theory since f describes the evolution of the matter distribution depending on the gravity theory.

2.4 Statistics of the Large-Scale Structure

Cosmological models predict the properties of the entire universe, not those of individual objects such as gas, stars, and galaxies. The actual distribution of galaxies is determined by the initial matter distribution in the universe, but the value of the density perturbation at an individual location is randomly determined and we cannot know its specific value from a cosmological model alone. Therefore, the statistics of the matter distribution are generally used to compare theoretical predictions by cosmological models and observations. In this section, we explore the power spectrum, a commonly used statistic to investigate the Large-Scale Structure (LSS).

Firstly, we define the Fourier transform of the matter density perturbation as

$$\tilde{\delta}(\boldsymbol{k}) = \int d^3x e^{-\boldsymbol{k}\cdot\boldsymbol{x}}\delta(\boldsymbol{x}), \tag{2.44}$$

where $\tilde{\delta}(\boldsymbol{k})$ is the Fourier counterpart of the matter density perturbation and $\boldsymbol{k}$ is a wave vector. Here, we define the two-point correlation function ξ as

$$\xi(x) = \langle\delta(\boldsymbol{x})\delta(\boldsymbol{x'})\rangle, \tag{2.45}$$

where $\langle\cdots\rangle$ represents the ensemble average, and we assume the cosmological principle, which means the universe is globally homogenous and isotropic, and ξ depends only on the distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ described as $x = |\boldsymbol{x} - \boldsymbol{x}'|$. ξ represents the correlation between the values of δ at $\boldsymbol{x}$ and $\boldsymbol{x}'$. And then, we consider the Fourier counterpart of $\langle\delta(\boldsymbol{x})\delta(\boldsymbol{x'})\rangle$;

$$\langle\tilde{\delta}(\boldsymbol{k})\tilde{\delta}(\boldsymbol{k}')\rangle = \int\int d^3x d^3x' e^{-\boldsymbol{k}\cdot\boldsymbol{x}} e^{-\boldsymbol{k}'\cdot\boldsymbol{x}'}\langle\delta(\boldsymbol{x})\delta(\boldsymbol{x'})\rangle \tag{2.46}$$

$$= \int\int d^3x d^3x' e^{-\boldsymbol{k}\cdot\boldsymbol{x}} e^{-\boldsymbol{k}'\cdot\boldsymbol{x}'}\xi(x). \tag{2.47}$$

we can do the integration in the above equation by converting $\boldsymbol{x} \to \boldsymbol{x}'' = \boldsymbol{x} - \boldsymbol{x}'$ and find

$$\langle\tilde{\delta}(\boldsymbol{k})\tilde{\delta}(\boldsymbol{k}')\rangle = (2\pi)^3\delta_D^3(\boldsymbol{k} + \boldsymbol{k}')P(k), \tag{2.48}$$

where $\delta_D^3(\boldsymbol{k})$ is the 3-dimensional Kronecker delta function, and $P(k)$ is power spectrum defined as

$$P(k) = \int d^3x e^{-\boldsymbol{k}\cdot\boldsymbol{x}}\xi(x) = 4\pi\int x^2 dx \frac{\sin kx}{kx}\xi(x), \tag{2.49}$$

where k is $=$ $|\boldsymbol{k}|$. Note that the power spectrum depends on the direction of the $\boldsymbol{k}$ vector when the matter distribution is anisotropic, e.g., the power spectrum in the redshift space (see Section 2.5). In the case that the spatial distribution of δ follows Gaussian, the two-point correlation function or the power spectrum represents all statistical information of the matter distribution. For non-Gaussian distribution, three- or higher-order statistics such as bispectrum (three-point correlation) and trispectrum (four-point correlation) include additional information to the power spectrum.

In the early universe, the matter distribution follows nearly Gaussian, and the density perturbation at each location grows linearly like Eq. (2.38). Furthermore, if $\delta(t')$ follows Gaussian, $\delta(t) \propto D(t)\delta(t')$ also follows Gaussian due to the property of random valuables following Gaussian. Therefore, from Eq. (2.38) and Eq. (2.48) the power spectrum at a time $t = t$ in the linear regime is described by

$$P(k,t) = D^2(t)P(k,t_0), \tag{2.50}$$

where $D(t)$ is the growth factor, and t_0 represents the present time. Following the matter-dominated epoch, the functional form of the power spectrum undergoes little change. On the other hand, during the radiation-dominated epoch, the growth of δ depends on the wave number k as follows; Firstly, we consider the perturbations for the scale that k is smaller than the horizon scale, which is the maximum scale that particles can interact each other. The growth of these perturbations is suppressed by the pressure of the radiation component. Smaller scales (larger k) are surpassed more strongly, as they remain within the horizon scale for a longer duration.

In a linear regime, the strength of the above suppression is determined independently for each scale. Therefore, we can describe the density perturbation evolves in the linear regime as

$$\tilde{\delta}(\boldsymbol{k}) = \frac{D(t)}{D(t_{\rm in})}T(k)\tilde{\delta}_{\rm in}(\boldsymbol{k}), \tag{2.51}$$

where $T(k)$ is a transfer function representing the suppression for each k, $t_{\rm in}$ is a time during the radiation-dominated epoch, and $\delta_{\rm in}$ is the value of the density perturbation in Fourier space at $t = t_{\rm in}$. Subsequently, the power spectrum is described as

$$P(k) = \frac{D^2(t)}{D^2(t_{\rm in})}T^2(k)P_{\rm in}(k), \tag{2.52}$$

where $P_{\rm in}(k)$ in the power spectrum at $t = t_{\rm in}$. The specific form of the initial power spectrum $P_{\rm in}(k)$ depends on the theoretical model. In general, we assume the initial power spectrum is proportional to the power function of k;

$$P_{\rm in} \propto k^{n_s}, \tag{2.53}$$

where n_s is the spectral index, representing the scale dependence of the initial power spectrum.

In summary, in the linear regime, the power spectrum for the initial power spectrum given by Eq. (2.53) is described as

$$P(k) = AD^2(t)T^2(k)k^{n_s}, \tag{2.54}$$

where A is the amplitude of the initial power spectrum, determined through observations. The parameters A and n_s are part of the cosmological parameters, and the functional forms of $D(t)$ and $T(k)$ can be calculated when assuming the cosmological model and the values of the cosmological parameters.

As another expression of the amplitude of the power spectrum, σ_8 is often used.

It is the density perturbation averaged within a sphere with a radius $R = 8\ \mathrm{Mpc}/h$, and defined as

$$\sigma_8^2 = \int \frac{d^3k}{(2\pi)^3} \tilde{W}(kR) P(k)|_{R=8\ \mathrm{Mpc}/h}, \tag{2.55}$$

where $\tilde{W}(kR)$ is the Fourier counterpart of the top-hat window function and is written as

$$\tilde{W}(kR) = \frac{3\sin(kR) - 3kR\cos(kR)}{(kR)^3}. \tag{2.56}$$

2.5 Observational effects

In this section, we discuss the observational effects when we use practical observables.

Firstly, in Section 2.4, we discuss the statistics of the matter density distribution in the universe. However, in practical observation, we cannot see dark matter, which is the dominant component of matter. Therefore, we need to observe a tracer of the dark matter density distribution. For example, we consider galaxies as observables. In general, we can consider that more galaxies exist in the high-density region. Therefore, we can describe the perturbation of the galaxy number density as

$$\delta_g = b\delta, \tag{2.57}$$

where δ_g is the galaxy number density perturbation, δ is the matter density perturbation, and b is the bias parameter representing the relation between the distribution of the matter density and galaxies. This bias relation (Eq. (2.57)) is not limited to the galaxy distribution; it can be adapted by considering the bias corresponding to the specific observable under consideration.

Next, in practical observations of the LSS, the distance to an observable such as a galaxy is measured through its redshift because photons from more distant objects are redshifted more. However, the redshift does not always correspond directly to the distance to the observable, as some observables have their peculiar velocity. We observe the redshift to an observable, which is derived from the cosmic expansion and the peculiar velocity of the observable. Therefore, when the distance to an observable is determined by its redshift, the observed distribution of the observables is distorted from the actual distribution along the line of sight. This distortion is referred to as Redshift Space Distortion (RSD).

In a homogenous and isotropic space-time, when λ_1' is the wavelength observed by the rest observer at the location of the source of light and λ_0 is the wavelength observed by us, the redshift derived from the cosmic expansion is

$$1 + z = \frac{\lambda_0}{\lambda_1'}, \tag{2.58}$$

where we consider the rest frame of us, i.e., our peculiar velocity is zero. And then, the observed redshift z_{obs} deviates from Eq. (2.58) due to the peculiar velocity of the source;

$$1 + z_{\mathrm{obs}} = \frac{\lambda_0}{\lambda_1}, \tag{2.59}$$

where λ_1 is the wavelength observed by the observer moving with the source. The relation between λ_1' and λ_1 is derived from the formula of the relativistic Doppler effect as

$$\lambda_1' = \frac{1+v_\parallel}{\sqrt{1+v_\parallel^2}}\lambda_1, \tag{2.60}$$

where $v_\parallel$ is the peculiar velocity of the source relative to us. In the case that the motion of the source is non-relativistic, i.e, $v_\parallel \ll 1$ (in natural units), we ignore the terms of the second and higher order of $v_\parallel$ and derive the relation from Eq. (2.58)-(2.60)

$$1 + z_{\rm obs} = (1+z)(1+v_\parallel), \tag{2.61}$$

$$z_{\rm obs} = z + \frac{v_\parallel}{a}, \tag{2.62}$$

where a is the scale factor at the time of the photon emission.

From Eq. (2.5), the observed distance s to the source is

$$s = \int_0^{z_{\rm obs}} \frac{dz}{H}, \tag{2.63}$$

$$s = \chi + \frac{v_\parallel}{aH}, \tag{2.64}$$

where χ is the actual comoving distance to the source, and we consider that the Hubble parameter H is nearly constant from z to $z_{\rm obs}$ by assuming the duration is enough small in deriving the second equation. Here, we can describe the relation between the actual comoving coordinates $\boldsymbol{x}$ (referred to as real space) and the coordinates $\boldsymbol{s}$ measured by the redshift (referred to as redshift space) as

$$(s_1, s_2, s_3) = \left(x_1, x_2, x_3 + \frac{v_\parallel}{aH}\right), \tag{2.65}$$

where x_3 and s_3 correspond to the direction along the line of sight. The number of observables is conserved in converting from the real space to the redshift space, so we find

$$n_r(\boldsymbol{x})d^3x = n_s(\boldsymbol{s})d^3s, \tag{2.66}$$

where n_r and n_s are the number density of the observables in the real space and redshift space, respectively. From Eq. (2.65)-(2.66), the relation between n_r and n_s is written as

$$n_s(\boldsymbol{s}) = \frac{n_r(\boldsymbol{x})}{1 + \frac{\partial}{\partial x_3}\left(\frac{v_\parallel}{aH}\right)}. \tag{2.67}$$

From the above, in the linear regime for the perturbation, the relation between the number density perturbation in the real space and the redshift space is described as

$$\delta_s(\boldsymbol{s}) = \delta_r(\boldsymbol{x}) - \frac{\partial}{\partial x_3}\left(\frac{v_\parallel}{aH}\right), \tag{2.68}$$

where $\delta_s = n_s/\bar{n}_s - 1$ and $\delta_r = n_r/\bar{n}_r$ is the number density perturbation in the real and redshift space, respectively, and we use the relation $\bar{n}_s = \bar{n}_r$ derived from Eq. (2.66). We can consider $\delta_s(\boldsymbol{s}) = \delta_s(\boldsymbol{x})$ in the linear regime from Eq. (2.65), and we assume the scales we focus on is enough small along the line of sight to approximate aH is constant. In this case, Eq. (2.67) is written as

$$\delta_s(\boldsymbol{x}) = \delta_r(\boldsymbol{x}) - \frac{1}{aH}\left(\frac{\partial v_\parallel}{\partial x_3}\right). \tag{2.69}$$

By Fourier transform, we find

$$\tilde{\delta}_s(\boldsymbol{k}) = \tilde{\delta}_r(\boldsymbol{k}) - \frac{ik_3}{aH}\tilde{v}_\parallel(\boldsymbol{k}), \tag{2.70}$$

where $\tilde{A}$ represents the Fourier counterpart of the quantity A, and k_3 is the component of the wave vector along the line of sight. From Eq. (2.40) and Eq. (2.42), $\tilde{v}_\parallel$ satisfy

$$\tilde{v}_\parallel = aHf\frac{ik_3}{k^2}\delta(\boldsymbol{k}), \tag{2.71}$$

where f is the growth rate, and δ is the mass density perturbation of matter. Therefore, we find the below relation;

$$\tilde{\delta}_s(\boldsymbol{k}) = \tilde{\delta}_r(\boldsymbol{k}) + f\mu^2\delta(\boldsymbol{k}), \tag{2.72}$$

where μ is defined as $\mu \equiv k_3/k$ representing the direction of the wave vector $\boldsymbol{k}$.

From Eq. (2.72), the power spectrum in the redshift space $P_s(\boldsymbol{k})$ is defined as

$$\langle\tilde{\delta}_s(\boldsymbol{k})\tilde{\delta}_s(\boldsymbol{k}')\rangle = (2\pi)^3\delta_D^3(\boldsymbol{k}+\boldsymbol{k}')P_s(\boldsymbol{k}), \tag{2.73}$$

and described by using the power spectrum in the real space as

$$P_s(k,\mu) = b^2(1+\beta\mu^2)P(k), \tag{2.74}$$

where b is the bias parameter for the observable we consider, and β is $= f/b$. The redshift space is anisotropic, so P_s depends on not only the absolute value but also the direction of $\boldsymbol{k}$ via μ. P_s is often described by its Legendre multipole P_l, i.e.,

$$P_l(k) = \frac{2l+1}{2}\int_{-1}^{1} P_s(k,\mu)L_l(\mu)d\mu, \tag{2.75}$$

where $L_l(\mu)$ is the Legendre polynomial of order l. The advantage of this expression is that we can express the power spectrum in the redshift space by using k without μ. For $l = 0,\ 2,$ and 4, the Legendre polynomial is

$$L_0(\mu) = 1, \tag{2.76}$$

$$L_2(\mu) = \frac{1}{2}(3\mu^2-1), \tag{2.77}$$

$$L_4(\mu) = \frac{1}{8}(35\mu^4-30\mu^2+3). \tag{2.78}$$

By comparing the observed power spectrum $P_s(k,\mu)$ with the theoretical predicted power spectrum $P(k)$, we can estimate the value of f. In practice, the value of $f\sigma_8$, which is independent of the bias parameter, is estimated(Song and Percival, 2009).

2.6 Modified Gravity

In this section, we briefly introduce modified gravity. While General relativity is consistent with observations at scales of the solar system, there is room for modification of the gravity theory at cosmological scales, such as the scale of galaxy clusters. Modified gravity is one of the theories that can explain the accelerated expansion of the universe, and it has been actively studied recently.

Firstly, we consider the perturbed FLRW metric. In the conformal Newtonian gauge, the perturbed FLRW metric is written as (Ma and Bertschinger, 1995; Bernardeau et al., 2002)

$$ds^2 = a(\tau)^2 \left[-(1 + 2\Psi(x,\tau))d\tau^2 + (1 - 2\Phi(x,\tau))dx^2 \right], \tag{2.79}$$

where Φ and Ψ are the scalar metric perturbation and correspond to the Newtonian potential, and τ is the conformal time defined as $\tau = \int^t 1/a(t')dt'$. In the case that the matter component is a perfect fluid, the energy-momentum tensor of the matter described as

$$T^\mu_\nu = p\delta^\mu_\nu + (\rho + P)U^\mu U_\nu, \tag{2.80}$$

where U^μ is the 4-dimensional velocity defined as $U^\mu = \frac{dx^\mu}{\sqrt{-ds^2}}$, δ^μ_ν is the Kronecker delta, ρ is the total energy density, and P is the total pressure. ρ and P can be written as $\bar{\rho}(1+\delta)$ and $\bar{P} + \delta P$ like Eq. (2.27). Therefore, the components of the energy-momentum tensor are (Nesseris, 2022)

$$T^0_0 = -\bar{\rho}(1+\delta), \tag{2.81}$$

$$T^0_i = (\bar{\rho} + \bar{P})u_i, \tag{2.82}$$

$$T^i_j = (\bar{P} + \delta P)\delta^i_j + \Sigma^i_j, \tag{2.83}$$

where Σ is the anisotropic shear perturbation, and u is the conformal time derivative of x, i.e., $u = \partial x/\partial\tau$. In the following in this section, the conformal time derivative of a quantity A is denoted by $A' = \partial A/\partial\tau$.

By using this energy-momentum tensor, we can write the perturbed Einstein equations in Fourier space as (Ma and Bertschinger, 1995; Nesseris, 2022)

$$k^2\Phi + 3\frac{a'}{a}\left(\Phi' + \frac{a'}{a}\Psi\right) = 4\pi G a^2 \delta T^0_0, \tag{2.84}$$

$$k^2\left(\Phi' + \frac{a'}{a}\Psi\right) = 4\pi G a^2 (\bar{\rho} + \bar{P})\theta, \tag{2.85}$$

$$\Phi'' + \frac{a'}{a}(\Psi' + 2\Phi') + \left(2\frac{a''}{a} - \frac{a'^2}{a^2}\right)\Psi + \frac{k^2}{3}(\Phi - \Psi) = \frac{4\pi}{3} G a^2 \delta T^i_i, \tag{2.86}$$

$$k^2(\Phi - \Psi) = 12\pi G a^2 (\bar{\rho} + \bar{P})\sigma, \tag{2.87}$$

where G is the gravitational constant, δT^0_0 and δT^i_i is the term of the perturbation in T^0_0 and T^i_i, respectively, $\theta = ik_j u^j$ is the fluid velocity, and σ is the rewritten anisotropic shear perturbation; $(\bar{\rho} + \bar{P})\sigma \equiv -(-\hat{k}_i\hat{k}_j - \frac{1}{3}\delta_{ij})\Sigma^{ij}$, where $\hat{k}_i$ is the unit vector of k_i. Here, we use the Einstein summation convention. From the above equations, the Newtonian potentials Φ and Ψ are expressed as (Nesseris, 2022)

$$\Psi(k,a) = -4\pi G \frac{a^2}{k^2} \mu(k,a) \bar{\rho}\delta, \tag{2.88}$$

$$\Phi(k,a) = -4\pi G \frac{a^2}{k^2} Q_{\rm eff}(k,a) \bar{\rho}\delta, \tag{2.89}$$

where μ and $Q_{\rm eff}$ are the phenomenological parameters representing the deviation from General relativity and equal to unity in General relativity. In general, μ and $Q_{\rm eff}$ are dependent on time and scale (Nesseris, 2022).

For the most modified gravity theories, the time evolution of the matter density perturbation δ is described as (Tsujikawa, 2007)

$$\delta''(a) + \left(\frac{3}{a} + \frac{H'(a)}{H(a)}\right)\delta'(a) - \frac{3\Omega_{\mathrm{m},0}\,\mu(k,a)}{2a^5\,H(a)^2/H_0^2}\delta(a) = 0, \tag{2.90}$$

where we assume the initial condition of $\delta(a \ll 1)$ is $\sim a$ and the one of $\delta'(a \ll 1)$ is ~ 1 at the initial time during the matter-dominated era.

By introducing the modified gravity, the expansion rate of the universe deviates from the one in General relativity and accelerated expansion can be caused. The modified gravity changes the time evolution of the matter density perturbation as Eq. (2.90), and then the growth rate f is different between General relativity and the modified gravity. Therefore, the growth rate f (multiplied by σ_8) is used as a probe of the gravity theory. However, we have some problems in estimating the value of $f\sigma_8$ such as the difficulty of the modeling of the theoretical power spectrum and the computational cost to explore the large amount of data. As the solution to these problems, we focus on machine learning analysis (see Chapter 4).

2.7 NCDM Models

In this section, we introduce the Non-Cold Dark Matter (NCDM) or Warm Dark Matter (WDM). Note that we focus on dark matter models where dark matter is a fundamental particle.

Dark matter is the dominant matter component in the universe and interacts with the gravitational force while it does not (or weakly) interact with the electromagnetic force. In the ΛCDM model, we assume dark matter is Cold Dark Matter (CDM), which is the non-relativistic particle when it is generated. However, ΛCDM does not specify the particular dark matter model, and dark matter is still mysterious matter. In addition, there is the small-scale crisis, referred to in Chapter 1, and this problem motivates the NCDM (or WDM) model.

The difference between CDM and NCDM is described as its velocity dispersion. CDM is non-relativistic just after it is generated and does not have velocity dispersion. On the other hand, NCDM is relativistic when it is generated and has some velocity dispersion. In general, the phase space distribution of NCDM is given by (Colombi, Dodelson, and Widrow, 1996)

$$f(v) = \frac{\beta}{\exp\left(\frac{p}{\alpha T_\gamma}\right) + 1}, \tag{2.91}$$

where $v = p/\sqrt{p^2 + m^2}$, T_γ is the cosmic microwave background photon temperature at the present time, α and β are the free parameters and m is the mass of NCDM. From this, we can also describe the energy density of NCDM relative to the total energy density in the universe as

$$\Omega_{\mathrm{NCDM}}h^2 = \beta\frac{\alpha^3}{(4/11)}\frac{m}{94\ \mathrm{eV}}, \tag{2.92}$$

where h is the dimensionless Hubble constant. The velocity dispersion depends on the ratio m/T, where $T = \alpha T_\gamma$ (Colombi, Dodelson, and Widrow, 1996), and the

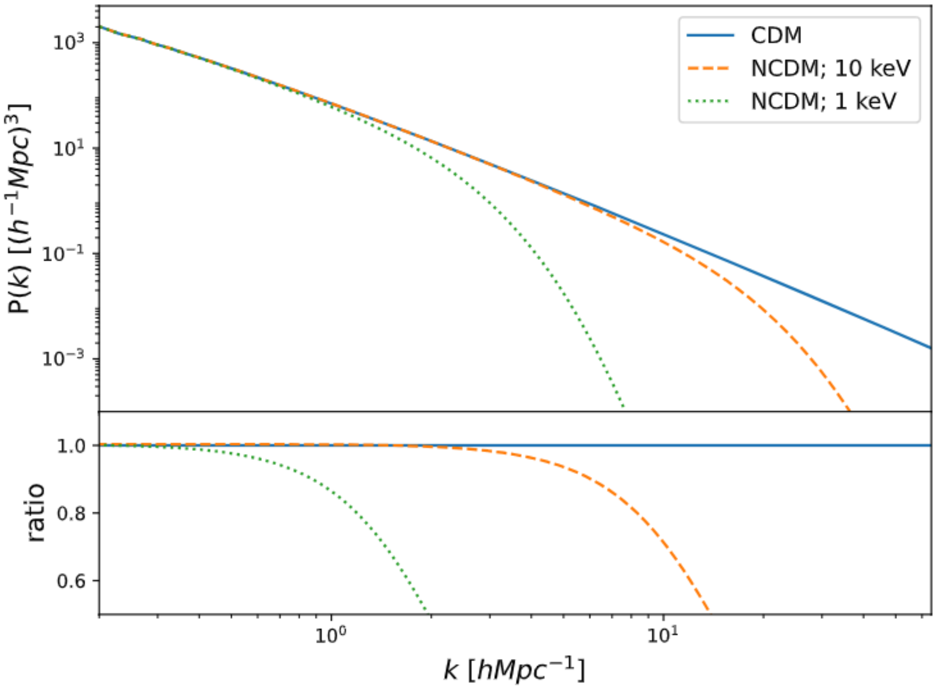

FIGURE 2.1: The linear matter power spectrum for CDM model and NCDM models with $m = 1$ keV and $m = 10$ keV at $z = 0$. The horizontal axis represents the wave number, The upper panel shows the amplitude of the power spectrum and the lower panel shows the ratio of the power spectrum of the NCDM to the one of CDM.

free streaming length of dark matter is $\sim T/m$. The free streaming of dark matter prevents the clustering of matter, and the amplitude of the power spectrum is suppressed within the scale of the free streaming length.

In this work, we consider the Dodelson-Widrow sterile neutrino (Dodelson and Widrow, 1994) as the NCDM model. A sterile neutrino is an undetected fundamental particle beyond the standard model of particle physics and is the right-handed particle while the active neutrino is the left-handed particle. It is expected that a sterile neutrino interacts with gravity while it does not have electromagnetic and weak interaction, unlike the active neutrino. Therefore, a sterile neutrino is one of the candidates for dark matter. The space space distribution of Dodelson-Widrow sterile neutrino is given by setting $\alpha = (4/11)^{1/3}$, which corresponds to the temperature of the active neutrino, in Eq. (2.91). Once we decide the abundance of NCDM, Ω_{NCDM}, and the mass of NCDM, we can decide the value of β from Eq. (2.92). In this work, we do not consider the mixture of CDM and NCDM, and we set $\Omega_{\mathrm{NCDM}} = 0.2621$, which is the value of the abundance of CDM from the Planck observation(Planck Collaboration et al., 2020).

Based on this phase space distribution, we can calculate the linear matter power spectrum. To calculate them in this paper, we use the Boltzmann solver code Cosmic Linear Anisotropy Solving System (CLASS) (Lesgourgues, 2011). This code solves the Boltzmann equation, which represents the time evolution of the phase space distribution function, and computes the power spectrum of the density perturbation numerically. Fig. 2.1 shows the linear power spectra for the CDM model and NCDM models with $m = 1$ keV and $m = 10$ keV at redshift $z = 0$. The upper panel shows the amplitude of the power spectrum and the lower panel shows the ratio of the amplitude of the NCDM power spectrum to the CDM power spectrum. We can see the suppression of the amplitude by the free streaming of NCDM and about ten times difference between the suppressed scale of NCDM models with $m = 1$ keV ($k \sim 0.3\, h\mathrm{Mpc}^{-1}$) and $m = 10$ keV ($k \sim 3\, h\mathrm{Mpc}^{-1}$).

When we know the matter distribution, we can extract the information about NCDM mass, e.g., by studying the power spectrum. However, dark matter has no or very weak interaction with the electromagnetic force and we cannot observe the dark matter distribution directly by telescopes. Therefore, we focus on the distribution of neutral hydrogen as a tracer of the one of dark matter (see Section 2.8). In addition, the effect of the NCDM mass appears at the small scale, where the non-linear effects arise and the matter distribution does not follow Gaussian. The power spectrum cannot all information included in the non-Gaussian distribution, so we want a method to access the information in the three- or higher-order statistics. In this work, we focus on the method using machine learning (Chapter 5).

2.8 21cm Signal

21cm signal is the radio waves emitted from neutral hydrogen (HI). It is emitted from the hyperfine structure of HI, and its wavelength is 21 cm corresponding to 1420 MHz in frequency. In recent years, there has been a lot of ongoing and planned radio observation such as the Murchison Wide-field Array (MWA) (Tingay et al., 2013), Canadian Hydrogen Intensity Mapping Experiment (CHIME) (Bandura et al., 2014), Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX) (Newburgh et al., 2016), and Square Kilometer Array (SKA) (Santos et al., 2015). These observations will provide us with the HI distribution, which we can use as a tracer of the dark matter distribution, through the 21cm signals.

In this work, we consider the post-reionization epoch ($z < 6$), when almost HI is ionized and HI resides in high-density regions such as haloes and galaxies. We can observe 21cm signals as the differential brightness temperature δT_b, which is the difference between the brightness temperature of the 21cm radiation and the cosmic microwave background photon temperature and is described as (Field, 1958)

$$\delta T_b = \frac{T_\mathrm{S} - T_\gamma}{1+z}\left(1 - e^{-\tau_{\nu_0}}\right), \tag{2.93}$$

where T_S is the spin temperature of HI, T_γ is the cosmic microwave background photon temperature, $\nu_0 = 1420$ MHz is the frequency of the 21cm radiation at the rest frame, and τ_{ν_0} is the optical depth of HI.

The spin temperature represents the temperature, corresponding to the ratio of the number of HI in the excited states of the hyperfine structure to that of the ground. It is defined by using the functional form of the Boltzmann distribution as

$$\frac{n_1}{n_0} = 3\exp\left(-\frac{h_p \nu_0}{k_B T_\mathrm{S}}\right), \tag{2.94}$$

where h_p and k_B are the Planck constant and the Boltzmann constant, respectively, n_1 and n_0 are the number density of the excited- and ground-state HI, respectively, and the coefficient of the exponential corresponds to the ratio of the statistical weight of the excited-state to that of the ground. Note that we use the MKS system of units in this section. The spin temperature is also described as (Field, 1958)

$$T_{\mathrm{S}}^{-1} = \frac{T_{\gamma}^{-1} + x_{\alpha} T_{\alpha}^{-1} + x_c T_{\mathrm{K}}^{-1}}{1 + x_{\alpha} + x_c}, \tag{2.95}$$

where T_{α} and x_{α} is the temperature of Ly-α and its coupling coefficient, and T_{K} and x_c is the kinetic gas temperature and its coefficient. In Chapter 5, we compute the spin temperature following (Furlanetto, Oh, and Briggs, 2006; Endo, Tashiro, and Nishizawa, 2020).

The optical depth is given by (Furlanetto, Oh, and Briggs, 2006)

$$\tau_{\nu_0} = \frac{3}{32\pi} \frac{h_p c^3 A_{10}}{k_B T_S \nu_0^2} \frac{n_{HI}}{(1+z)(dv_{\parallel}/dr_{\parallel}))}, \tag{2.96}$$

where c is the speed of light, $A_{10} = 2.85 \times 10^{-15}\ [s^{-1}]$ is the Einstein coefficient, corresponding to the possibility of the emission of the 21cm radiation from a HI per second, n_{HI} is the number density of HI, z is the redshift of the source of 21cm radiation, and $dv_{\parallel}/dr_{\parallel}$ is the gradient of the velocity of the source along the line of sight. In the following, we replace the $dv_{\parallel}/dr_{\parallel}$ with the expansion rate (or Hubble parameter) $H(z)$ because the peculiar velocity of HI is significantly smaller than the receding speed caused by the expansion (Ando et al., 2021).

We can assume $\tau_{\nu_0} \ll 1$ during the post-reionization epoch, when almost HI is ionized, and rewrite Eq. (2.93) as

$$\delta T_b \sim \frac{3}{32\pi} \frac{h_p c^3 A_{10}}{k_{\mathrm{B}} \nu_0^2} \left(1 - \frac{T_{\gamma}(z)}{T_{\mathrm{S}}}\right) \frac{n_{\mathrm{HI}}}{(1+z)H(z)}. \tag{2.97}$$

As we can see in Eq. (2.97), δT_b is proportional to the HI number density n_{HI}. Thus, we can infer the HI distribution by observing the differential brightness temperature.

Chapter 3

Convolutional Neural Networks

In this chapter, we introduce the basics of convolutional neural networks (CNN). In Section 3.1, we introduce the single-layer and multi-layer Perceptrons, which form the basis of neural networks. Section 3.2 shows the basic structure and the algorithms used to train the neural network. In Section 3.3, we illustrate the structure of the convolutional neural network, which is employed in our work.

3.1 Perceptron

In this section, we consider the regression problem as an example. Our aim is the prediction of a value of y by using the data vector x.

The single-layer Perceptron (Rosenblatt, 1958) models the function of our brain and is one of the oldest and simplest structures of neural networks. It describes the output by the linear combination of the input as

$$y = \sum_{i=0}^{n} w_i x_i = w \cdot x, \tag{3.1}$$

where x_i is the i-th component of the input data vector $x = (x_1, x_2, \cdots, x_n)$, w is the vector of the weight parameters, n is the number of components of x, and we set $x_0 = 1$ to represent the constant bias with w_0. And then, $y = w \cdot x$ is converted to $f(w \cdot x)$, where f is the activation function for the non-linear transformation and (Rosenblatt, 1958) suggests the sigmoid function $f(t) = 1/(1 + e^{-t})$ as the activation function. The other activation functions are introduced in Section 3.1.1. In training, w is optimized to fit the training data, where we know the correct value of y corresponding to the input data vector x. The single-layer Perceptron can also be illustrated as shown in Figure 3.1. Such illustrations are convenient for understanding more complex structures of neural networks. Fitting the single-layer Perceptron to a single training data is straightforward by choosing w that minimizes the squared error $E = (y - w \cdot x)^2/2$.

To express more complex function, the multi-layer Perceptron (Cybenko, 1989) is suggested to express the more complex functions. The multi-layer Perceptron is composed of multiple single-layer Perceptrons and can describe the non-linear function of x. Figure 3.2 illustrates an example of the multi-layer Perceptron, which consists of the input and output layer, and a hidden layer. The hidden layer is the layer between the input and output layer and consists of nodes (shown by circles in Figure 3.2), where the input to each node is weighed by the different weight parameters. In addition, the hidden layer is not necessarily a single layer. Applying multiple hidden layers allows the multi-layer Perceptron to express more complex functions, although the training becomes more challenging.

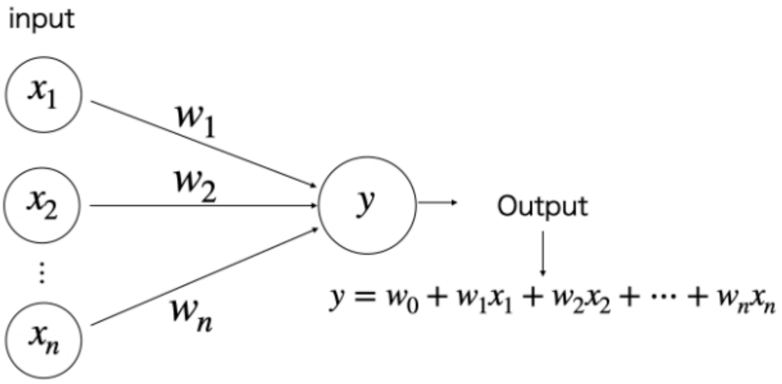

FIGURE 3.1: The illustration of the single-layer Perceptron.

For the multi-layer Perceptron in Figure 3.2, the input to the hidden layer is transformed as

$$u_l = \sum_k w_{kl}^{(1)} x_k, \tag{3.2}$$

where l and k represent the corresponding node of the hidden layer and component of the input, respectively, and the superscript of w indicates the number of the previous layer. And then, the activation function f is applied to u_l in a node for the non-linear transformation of u_l. Finally, the output value y is described by the linear combination of $f(u_l)$ as

$$y = \sum_i w_i^{(2)} f(u_i), \tag{3.3}$$

where i represents the i-th input from the node in the previous layer. In general, the output value is also transformed by the activation function.

The multi-layer Perceptron includes non-linear transformations and can describe the complex conversion of input data through multiple hidden layers and their nodes. However, optimizing the multi-layer Perceptron is challenging due to its complex architecture. To optimize the parameter, we use gradient descent (Section 3.2.2) and backpropagation (Section 3.2.3). Neural Network is an algorithm that consists of the multi-layer Perceptron and its training processes, such as gradient descent and backpropagation.

3.1.1 Activation function

The activation function is used to express the non-linear transformation. In this work, we use the following activation functions:

- ReLU (Rectified Linear Unit)

$$f(x) = \left\{ \begin{array}{ll} 0 & (x \leq 0) \\ x & (x > 0) \end{array} \right. . \tag{3.4}$$

It is commonly in hidden layers. Its derivative is defined as

$$\frac{df(x)}{dx} = \left\{ \begin{array}{ll} 0 & (x \leq 0) \\ 1 & (x > 0) \end{array} \right. . \tag{3.5}$$

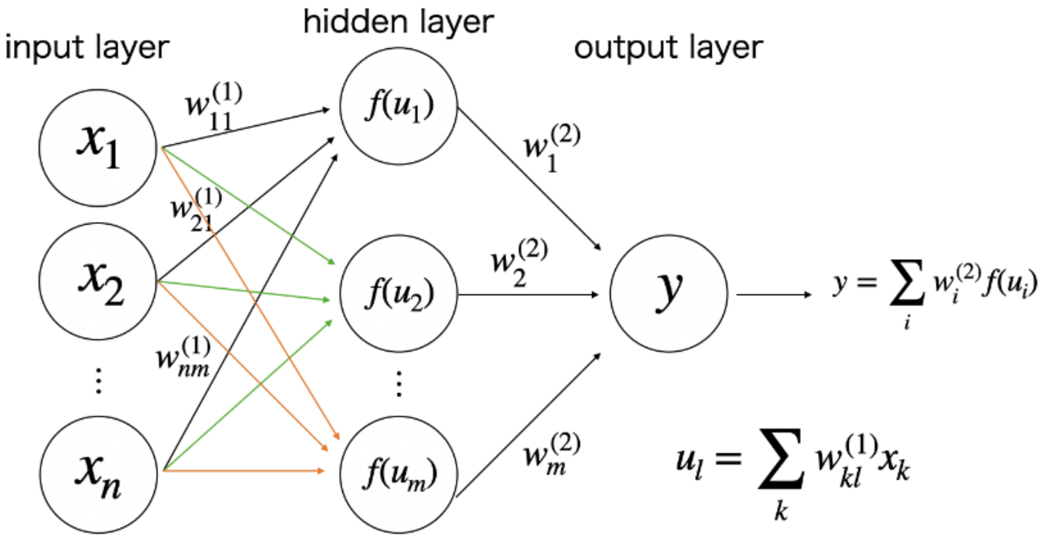

FIGURE 3.2: The illustration of the multi-layer Perceptron.

In training, we need to calculate the derivative of the activation function (see Section 3.2.3). The derivative of ReLU is unity for positive input, and it reduces the risk of the vanishing gradient problem (He et al., 2016), where the parameters are not updated due to a very small gradient.

- identity function

  It is used in the output layer of neural networks for regression. The output value is directly used as the predicted value.

- softmax function

  $$y = \frac{\exp(x)}{\sum_i \exp(x_i)}. \tag{3.6}$$

  It is used in the output layer of neural networks for data classification. In neural networks for classification, the output layer has multiple nodes. Each node corresponds to a target category, e.g., cat and dog for image recognition. i represents the label of the node in the output layer, and the value of the softmax function represents the probability that the input data belongs to the category corresponding to that node.

3.2 Training Neural Network

This section shows the loss function, optimization of the weight parameter, backpropagation method, and some learning techniques for the training of neural networks.

3.2.1 Loss function

In training neural networks, we need to evaluate the difference between the target value (true value for regression or true category for classification) and the predicted value from the neural network. Various loss functions are suggested to do it.

- Mean squared error (MSE)

  $$E = \frac{1}{2}\sum_k (y_k - t_k)^2, \tag{3.7}$$

  where y_k and t_k are the predicted and true values for the k-th node in the output layer, respectively. This loss function is used in neural networks for regression.

- Mean Absolute Error (MAE)

  $$E = \frac{1}{2}\sum_k |y_k - t_k|, \tag{3.8}$$

  where y_k and t_k are the predicted and true values for the k-th node in the output layer, respectively. This loss function is also used in neural networks for regression. This Loss function is more robust against outliers than MSE because MSE is the squared value of the difference between y_k and t_k and tends to be larger than MAE. However, MSE is more sensitive to larger differences y_k and t_k.

- Huber loss (Huber, 1964)

  $$E = \left\{ \begin{array}{ll} \frac{1}{2}a^2 & (|a| \leq \delta) \\ \delta(|a| - \frac{1}{2}\delta) & (|a| > \delta) \end{array} \right. , \tag{3.9}$$

  where a corresponds to the difference between the true and predicted values, and δ is the hyperparameter. This loss function is used in neural networks for regression and has similar properties of MAE, but we need to choose the value of δ.

- Likelihood Free Inference (LFI) loss (Villaescusa-Navarro et al., 2022; Jeffrey and Wandelt, 2020)

  $$E = \log\left(\sum_j (\theta_j - \mu)^2\right) + \log\left(\sum_j \left((\theta_j - \mu)^2 - \sigma^2\right)^2\right), \tag{3.10}$$

  where we need the subgroup of the training data, j represents the label of data in the subgroup, θ is the predicted value, and μ and σ are the mean and standard deviation of the true values of the data in the subgroup, respectively. This loss function is used in neural networks for regression. It can use the information of the mean and standard deviation of the training data.

- cross entropy loss

  $$E = -\sum_k t_k \ln(y_k), \tag{3.11}$$

  where y_k are the predicted probability of the k-th node in the output layer ranging from 0 to 1, and t_k is unity for the node corresponding to the true category and is zero for the node corresponding to the false category. This loss function is used in neural networks for classification. When the predicted probability y_k is 1 for the node corresponding to the false category, this loss function tends to ∞.

In our work, we use MSE and cross entropy loss mainly.

3.2.2 Optimization

In algorithms of neural networks, the weight parameters w are optimized to minimize the value of the loss function. In general. the loss function is minimized when its gradient with respect to w is zero. However, the architecture of neural networks is complex and includes many weight parameters. It is impractical to solve the equation $\partial E/\partial w = 0$ analytically.

In general, the optimization of the weight parameters in neural networks is conducted by using gradient descent. In the gradient descent, the weight parameters are optimized by the following process:

$$w \leftarrow w - \eta \frac{\partial E}{\partial w}, \tag{3.12}$$

where η is the hyperparameter and referred as to the learning rate. By updating the weight parameters using the gradient of the loss function E with Eq. (3.12), the loss function becomes close to its minimum value. For multiple weight parameters, w in Eq. (3.12) is replaced with the vector of weight parameters $\boldsymbol{w}$.

Eq. (3.12) is the basic form of weight parameter optimization. In practice, various optimization methods are employed to effectively update the parameters."

- Stochastic Gradient Descent (SGD)

  $$w \leftarrow w - \eta \frac{\partial E}{\partial w}. \tag{3.13}$$

  The form of the formula to update the parameters is the same as Eq. (3.12). In SDG, the data that is used for the optimization are randomly selected in each update of parameters. The shape of the loss function is different in each update, making it less likely for the parameters to get stuck at a local minimum of the loss function.

- Adam (Kingma and Ba, 2014)

  $$m_0 = v_0 = 0, \tag{3.14}$$
  $$m_t = \beta_1 m_{t-1} + (1-\beta_1)\frac{\partial E}{\partial w}, \tag{3.15}$$
  $$v_t = \beta_2 v_{t-1} + (1-\beta_2)\left(\frac{\partial E}{\partial w}\right)^2, \tag{3.16}$$
  $$\hat{m}_t = \frac{v_t}{1-\beta_1^t}, \tag{3.17}$$
  $$\hat{v}_t = \frac{v_t}{1-\beta_2^t}, \tag{3.18}$$
  $$w \leftarrow w - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \tag{3.19}$$

  where t is the number of times of the update, β_1 and β_2 are hyperparameters, ϵ is the small value to avoid the divergence to infinity of the fraction in the last formula. This optimization adjusts the amount of the update by using the previous update. As the initial value of the hyperparameters, the following values are recommended; $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\eta = 0.001$, $\epsilon = 10^{-8}$

- AMSGRAD with Adam (Reddi, Kale, and Kumar, 2019)

$$
\begin{aligned}
m_0 &= v_0 = 0, & (3.20)\\
m_t &= \beta_1 m_{t-1} + (1-\beta_1)\frac{\partial E}{\partial w}, & (3.21)\\
v_t &= \beta_2 v_{t-1} + (1-\beta_2)\left(\frac{\partial E}{\partial w}\right)^2, & (3.22)\\
\hat{m}_t &= \frac{v_t}{1-\beta_1^t}, & (3.23)\\
\hat{v}_t &= \frac{v_t}{1-\beta_2^t}, & (3.24)\\
\hat{v}^{\max} &\leftarrow max(\hat{v}, \hat{v}^{\max}), & (3.25)\\
w &\leftarrow w - \eta\frac{\hat{m}_t}{\sqrt{v_t^{\hat{\max}}} + \epsilon}, & (3.26)
\end{aligned}
$$

This is an improved optimization method compared to Adam. The optimization process is more complex, and the AMSGRAD is not always superior to Adam.

3.2.3 Backpropagation

For the optimization process, we need to calculate the gradient of the loss function. However, neural networks have many weight parameters and the computation of the gradient is challenging. In backpropagation(Rumelhart, Hinton, and Williams, 1986), the gradient is calculated step by step from the output to the input layer using the chain rule.

In the following, we show an example for the same architecture of the neural network shown in Figure 3.2.

Firstly, the gradient of the loss function E with respect to $w^{(2)}$ is calculated using the chain rule as

$$\frac{\partial E}{\partial w_i^{(2)}} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial w_i^{(2)}}, \tag{3.27}$$

where i is the label of the parameter corresponding to the node of the previous layer. In this equation, $\partial E/\partial y$ can be computed by specifying the loss function and $\partial y/\partial w_i^{(2)}$ is expressed as

$$
\begin{aligned}
\frac{\partial y}{\partial w_i^{(2)}} &= \frac{\partial}{\partial w_i^{(2)}}\left[f\left(\sum_j w_j^{(2)} f(u_j)\right)\right], & (3.28)\\
&= \frac{\partial}{\partial w_i^{(2)}}\left(\sum_j w_j^{(2)} f(u_j)\right) \times \frac{\partial f(X)}{\partial X}, & (3.29)\\
&= f(u_i)\frac{\partial f(X)}{\partial X}, & (3.30)
\end{aligned}
$$

where X is $= \left(\sum_j w_j^{(2)} f(u_j)\right)$, and $\partial f(X)/\partial X$ can be computed by specifying the activation function f. Therefore, we can write the gradient of the loss function E with respect to $w^{(2)}$ as

$$\frac{\partial E}{\partial w_i^{(2)}} = \frac{\partial E}{\partial y} f(u_i)\frac{\partial f(X)}{\partial X}. \tag{3.31}$$

Next, the gradient of the loss function E with respect to $w^{(1)}$ is calculated using the chain rule as

$$\frac{\partial E}{\partial w_{kl}^{(1)}} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial f(u_l)}\frac{\partial f(u_l)}{\partial w_{kl}^{(1)}}, \tag{3.32}$$

$$= \frac{\partial E}{\partial y}\frac{\partial y}{\partial f(u_l)}\frac{\partial f(u_l)}{\partial u_l}\frac{\partial u_l}{\partial w_{kl}^{(1)}}. \tag{3.33}$$

In this equation, the value of $\partial E/\partial y$ is already calculated, and we can determine the value of and $\partial f(u_l)/\partial u_l$ by choosing the activation function. And then, $\partial y/\partial f(u_l)$ and $\partial u_l/\partial w_{kl}^{(1)}$ are described as

$$\frac{\partial y}{\partial f(u_l)} = \frac{\partial}{\partial f(u_l)}\left(\sum_j w_j^{(2)} f(u_j)\right), \tag{3.34}$$

$$= w_l^{(2)}, \tag{3.35}$$

and

$$\frac{\partial u_l}{\partial w_{kl}^{(1)}} = \frac{\partial}{\partial w_{kl}^{(1)}}\left(\sum_p w_{pl}^{(1)} x_p\right), \tag{3.36}$$

$$= x_k, \tag{3.37}$$

respectively. Therefore, $\partial E/\partial w_{kl}^{(1)}$ is given by

$$\frac{\partial E}{\partial w_{kl}^{(1)}} = \frac{\partial E}{\partial y} w_l^{(2)} \frac{\partial f(u_l)}{\partial u_l} x_k. \tag{3.38}$$

As above, we can write the gradient of E with respect to w. Even when considering neural networks with multiple nodes in the output layer, different activation functions in the output layer, or multiple hidden layers, we can easily extend the above discussion. By using the gradient of E and the optimization method introduced in Section 3.2.2, we can train the neural network.

3.2.4 Learning Techniques

In this section, we introduce some techniques for training of neural networks

- mini-batch learning

  In our work, the term "one epoch" refers to one complete path through the entire training dataset during the training of neural networks. During each epoch, the neural networks learn from the entire training dataset once. In an epoch, the training dataset is divided into some subgroups, and the weight parameters are updated by each mean value of the loss function E for the data in the subgroup. This learning process is referred to as mini-batch learning. The subgroup and the number of data in a subgroup are called the mini-batch and the batch size, respectively. Applying mini-batch learning makes it less likely for the weight parameters to be trapped at the local minimum of the loss function because the shape of the loss function is different for each mini-batch, and it takes less time to train the neural network compared to the case where the parameters are updated by individual data.

- L2 regularization

  The L2 regularization is used for neural networks to reduce overfitting the training data. The following value is added to the loss function;

$$E_{\mathrm{L2}} = \frac{\lambda}{2}\sum_i |w_i|^2, \tag{3.39}$$

  where λ is the hyperparameter, and the summation is performed over all weight parameters in the neural network. By doing this, the extremely large values of the weight parameters are suppressed, and reduce the overfitting of the neural network to the training data.

- dropout

  In training a neural network, some weight parameters are temporarily set to zero to reduce the overfitting of the neural network to the training dataset. This method is referred to as dropout. By using dropout, the complexity of the neural network's architecture is suppressed, making it less likely to overfit the training dataset. In the testing phase, the dropout is not applied, so the neural network retains its original configuration.

- batch normalization (Ioffe and Szegedy, 2015)

  Batch normalization is a method used to prevent the vanishing and exploding gradient problems, which are that the gradient of the loss function becomes extremely small or large, respectively, and stabilize the learning of neural networks. This method suppresses the internal covariate shift, which is the change in the distribution of the input to layers of a neural network during training. The deeper the neural network we consider, the more internal covariate shifts emerge because the input data is transformed more times. To apply batch normalization, mini-batch learning is used. We consider a mini-batch that includes the data $x = (x_1, x_2, \cdots, x_n)$. The mini-batch normalization is described as

$$\hat{x}_k = \frac{x_k - \mu}{\sqrt{\sigma^2}}, \tag{3.40}$$

  where μ and σ are the mean and standard deviation of the components of x, respectively. By using batch normalization, the inputs to the layers of the neural networks are standardized in a mini-batch. As a result, the internal covariate shift is suppressed. Batch normalization not only stabilizes the learning but also generalizes the neural network, i.e., the neural network shows high performance on the test dataset.

3.3 Convolutional Neural Network

A Convolutional Neural Network (CNN) is one of the neural network algorithms used for image analysis. In CNN, the convolution layer and pooling layer are added to the usual multi-layer Perceptron for image processing.

3.3.1 convolution layer

The convolution layer extracts information from images by convolving them with kernels (or filters). Figure 3.3 shows an example of the calculation in the convolution layer, where the size of the input image and the kernel are 3×3 and 2×2,

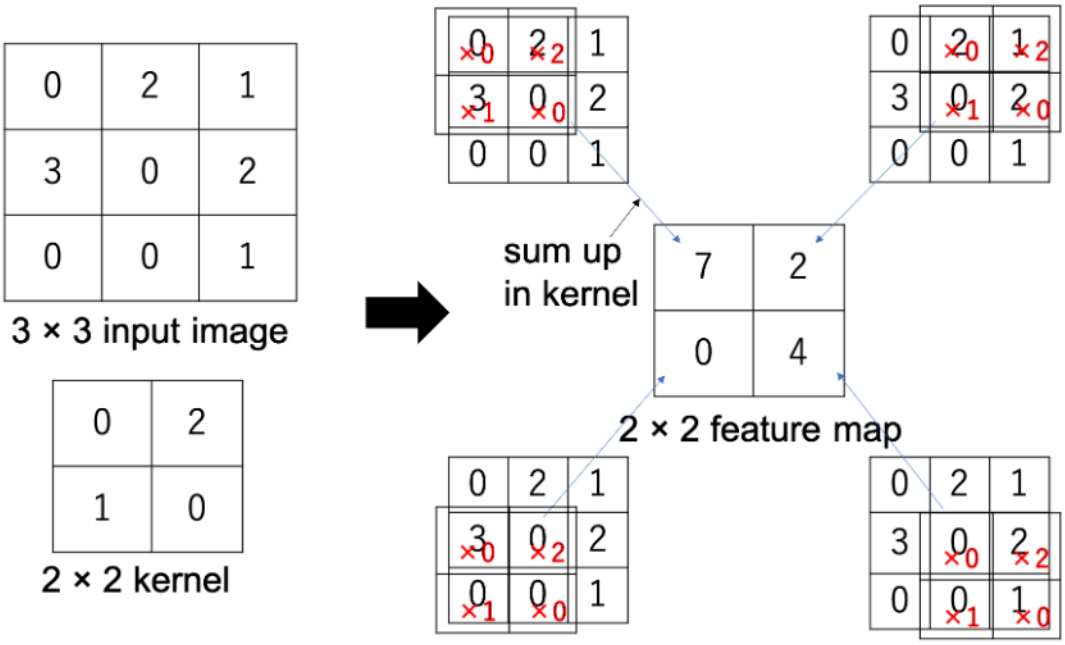

FIGURE 3.3: The illustration of the calculation of the convolution of the 3×3 input image with the 2×2 kernel. In this figure, the stride s is $= 1$.

respectively. In the convolution layer, the kernel is overlaid on the input image. The value in the input image is multiplied by the corresponding value in the kernel at each position, and the results are summed up to generate the feature map. Subsequently, the kernel is moved spatially across the input image and the above process is repeated. The feature map, representing the extracted information from the input image, can be described as

$$F_{ij} = \sum_{k=0}^{n_x-1} \sum_{l=0}^{n_y-1} I_{i+k,j+l} \times K_{k+1,l+1} \, , \tag{3.41}$$

where F_{ij}, $I_{i,j}$, and $K_{i,j}$ represent the $i \times j$ element of the $[(N_x - (n_x - 1))/s] \times [(N_y - (n_y - 1))/s]$ feature map, the $N_x \times N_y$ input image, and the $(n_x \times n_y)$ kernel, respectively. Here, s is the stride, which is the step size when the kernel is moved spatially across the input. The values of the elements of K are updated by the learning process introduced in Section 3.2. This calculation can be easily extended to 3D input images by using 3D kernels.

The number of convolution layers and the size of the kernel are the hyperparameters. We need to choose the proper values of these parameters. It is generally recommended to have a dataset size of at least one-tenth of the total number of weight parameters (Han et al., 2015).

3.3.2 pooling layer

The pooling layer compresses the input image and suppresses the computational cost. In addition, it makes the CNN less likely to be affected by the position of the feature in the input image. In the pooling layer, the average- or max-pooling is often used.

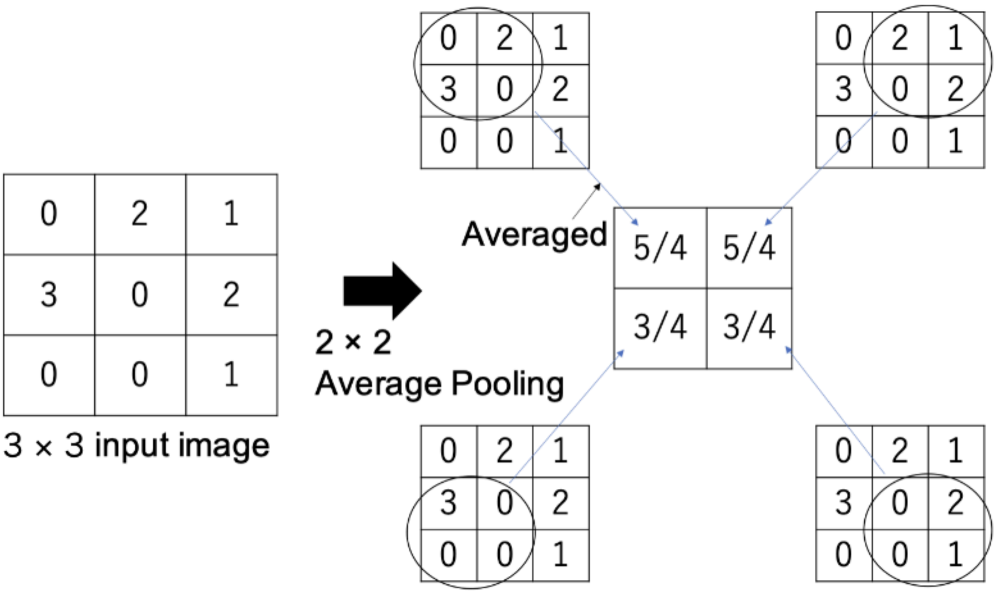

FIGURE 3.4: The illustration of the pooling layer calculation for the 3×3 input image with the 2×2 average pooling.

Figure 3.4 shows the calculation of the 2×2 average pooling of the 3×3 input image. In the 2×2 average pooling, the 2×2 patch is trimmed out the values in the patch, and the feature map is generated by averaging the values in the patch. The patch is moved spatially across the input image. The feature map from the average $n_x \times n_y$ pooling layer can be described as

$$F_{ij} = \frac{1}{n_x \times n_y} \sum_{k=0}^{n_x-1} \sum_{l=0}^{n_y-1} I_{(N_x/n_x)\times i-k,(N_y/n_y)\times j-l}, \tag{3.42}$$

where F_{ij} and $I_{i,j}$ represent the $i \times j$ element of the $(N_x/n_x) \times (N_y/n_y)$ feature map and the $N_x \times N_y$ input image, respectively. For max-pooling, we just pick up the maximum value in the patch instead of averaging.

Related to the average pooling layer, the global average pooling layer (Lin, Chen, and Yan, 2013) is often used to convert the extracted information to the output of the CNN instead of the fully connected layer, where the inputs are transformed by the equation such as Eq. (3.2) or Eq. (3.3). The calculation in global average pooling layer correspond to Eq. (3.42) with $N_x = n_x$ and $N_y = n_y$

Considering the $N \times (N_x \times N_y)$ input, where N represents the number of the input, the fully connected layer requires $N \times N_x \times N_y$ weight parameters to convert the input to the output. On the other hand, the global average pooling layer needs only N weight parameters. The global average pooling layer often shows higher performances than the fully connected layer (Lin, Chen, and Yan, 2013).

Chapter 4

Estimation of the Growth Rate

This chapter presents the data, methods, results, and discussion related to predicting $f\sigma_8$ directly from the 3D matter distribution using machine learning. We aim to investigate the machine learning approaches, to avoid the modeling of the observed power spectrum in the non-linear regime. This work uses the Quijote simulations. We analyze the dark matter and halo distribution of the simulations by the image-based analysis with CNN and the power spectrum analysis with the neural network. In addition, we compare these results with the errors estimated by the Fisher matrix approach.

The contents of this chapter follow Murakami et al., 2023b.

4.1 Data

4.1.1 Quijote simulation

In this work, we use the publicly available N-body simulation data, the Quijote simulations (Villaescusa-Navarro et al., 2020). The Quijote simulations are a set of 44100 N-body dark matter simulations with ΛCDM model. These simulations are created by running the GADGET-3 (described in Springel, 2005), which is a TreePM code. The size of each simulation box is 1 h^{-1}Gpc on a side, and the number of dark matter particles is 512^3 in a simulation. We use these simulations to train and test our CNN in this work.

This work focuses on 2000 realizations out of 44100 simulations, which assume the different combinations of cosmological parameters $(\Omega_{m0}, \Omega_{b0}, h, n_s, \sigma_{8,0})$. Here, Ω_{b0} represents the present abundance of baryons defined as

$$\Omega_{b0} = \frac{8\pi G\rho_{b0}}{3H_0^2}, \tag{4.1}$$

where ρ_{b0} is the energy density of baryons at the present time. Ω_{m0}, h, n_s, and $\sigma_{8,0}$ are the abundance of matter, dimensionless Hubble constant, spectral index, and the amplitude of the power spectrum, respectively, and are already introduced in Chapter 2. The set of cosmological parameters is chosen randomly by Latin-hypercube sampling within the range of

$$\begin{aligned} \Omega_m &\in [0.1, 0.5], \\ \Omega_b &\in [0.03, 0.07], \\ h &\in [0.5, 0.9], \\ n_s &\in [0.8, 1.2], \\ \sigma_8 &\in [0.6, 1.0]. \end{aligned} \tag{4.2}$$

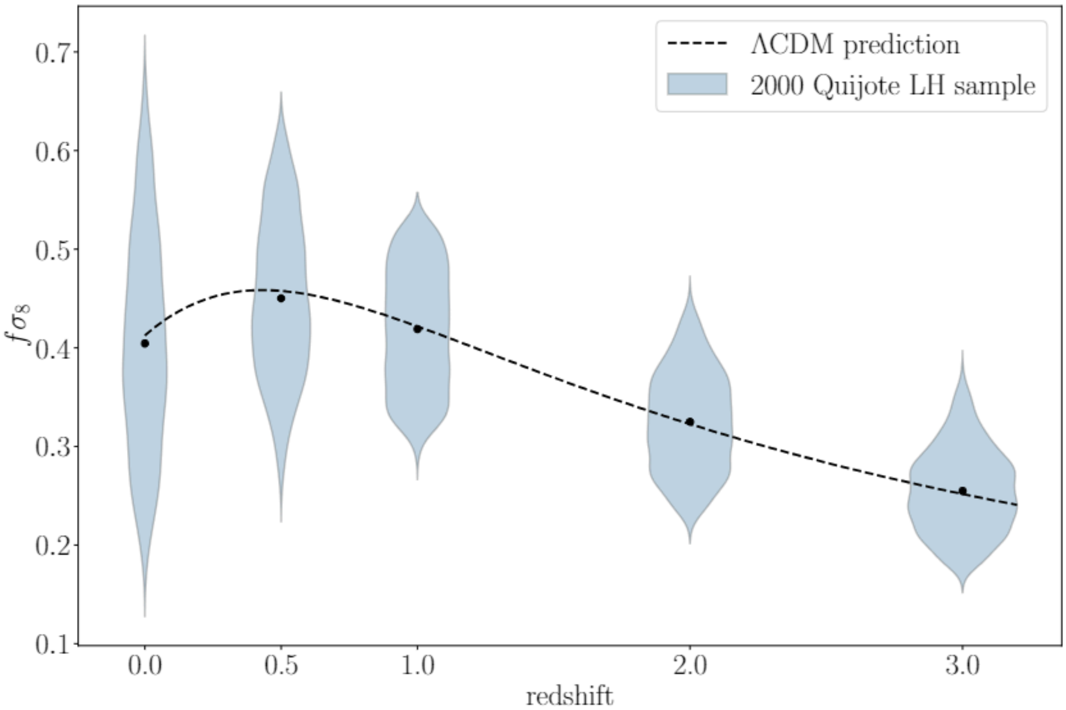

FIGURE 4.1: The distribution of $f\sigma_8$ for 2000 realizations at their corresponding redshifts. The black dots and black dashed line represent the mean of $f\sigma_8$ and the ΛCDM prediction based on the mean value of each cosmological parameter at each redshift, respectively.

We can compute the $f\sigma_8$ from Ω_m and σ_8 based on the ΛCDM model as (Padmanabhan, 2003; Silveira and Waga, 1994; Percival, 2005; Bueno belloso, Garcia-Bellido, and Sapone, 2011)

$$f(a) = \frac{\mathrm{d}\log\delta_\mathrm{m}}{\mathrm{d}\log a}, \tag{4.3}$$

$$\delta_\mathrm{m}(a) = a \cdot {}_2F_1\left(\frac{1}{3}, 1; \frac{11}{6}; a^3\left(1 - \frac{1}{\Omega_{\mathrm{m},0}}\right)\right), \tag{4.4}$$

where ${}_2F_1$ is the Gaussian hypergeometric function given by

$${}_2F_1(a,b,c,z) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n}\frac{z^n}{n!}, \tag{4.5}$$

$$(x)_n = x(x+1)(x+2)\cdots(x+n-1). \tag{4.6}$$

Here, we assume the quasi-static and sub-horizon (Tsujikawa, 2007).

All simulations are generated from different seeds of the initial conditions. Each realization has 5 snapshots at redshifts $z = [3, 2, 1, 0.5, 0]$, and Figure4.1 shows the distribution of the values of $f\sigma_8$ for 2000 realizations at each corresponding redshift.

As a rough approximation, we use the snapshots at $z = [3, 2, 1, 0.5]$ to represent four redshift bins. The mean redshift in each bin is denoted as $z = [2.5, 1.75, 1.25, 0.75]$ in the following.

z range	z mean	relative number	n_{halo} $[(\mathrm{Gpc}h^{-1})^{-3}]$	comoving distance $[h^{-1}\mathrm{Mpc}]$	box size [deg]
$[0.5, 1.0]$	0.75	2	8.0×10^3	1.9×10^3	30.2
$[1.0, 1.5]$	1.25	4	1.6×10^4	2.7×10^3	20.9
$[1.5, 2.0]$	1.75	3	1.2×10^4	3.4×10^3	16.9
$[2.0, 3.0]$	2.5	1	8.0×10^3	4.1×10^3	14.0

TABLE 4.1: Specifications for the simulations. The relative number is our assumed ratio of the number of halos in each redshift bin, which roughly corresponds to the *Euclid* survey and n_{halo} is the halo number density included in the simulation box after the random draw. We also provide observation-related quantities, such as the comoving distance to the mean redshift and the subtended angle of the simulation box located at the comoving distance.

4.1.2 the Halo Catalog

The Quijote simulations provide the halo catalog made by the Friends-Of-Friends (FOF) algorithm, which identifies halos in a N-body simulation by considering particles as belonging to the same halo if they are close. In the catalog from the Quijote simulation, a halo is defined as having at least 20 dark matter particles. Therefore, the lower limit of a halo mass is $\mathcal{O}(10^{12})$ - $\mathcal{O}(10^{13})$ $[M_\odot/h]$, which varies with the set of cosmological parameters because a particle mass in the N-body simulation is determined by the number of dark matter particles and the cosmological parameters related to the total mass in the box.

To reproduce the redshift distribution of the observables of the practical observation such as *Euclid* (Laureijs et al., 2011), we assume the ratio of the halo number in each redshift bin, and we draw halos randomly for the halo number to match our assumption. Table 4.1 shows our assumption of the relative number and the halo number density after the random draw. This table also shows the specifications of the simulations such as the comoving distance to the mean redshift of each redshift bin, and the subtended angle of the simulation box at the mean redshift. Here, we calculate the apparent angular size as

$$\theta\ [\mathrm{rad}] = 2 \arctan\left(\frac{l/a}{2\chi(z)}\right), \tag{4.7}$$

where l is the physical size of the simulation box, a is the scale factor, and $\chi(z)$ is the comoving distance to the redshift z.

4.1.3 Data for Machine Learning

In this work, we use the CNN to analyze the 3D images of the dark matter or halo distribution, which are generated from the simulations as

1. We define the 40^3 grids in a simulation box. The grid size corresponds to $k = 0.25\ h\mathrm{Mpc}^{-1}$ in Fourier space, which is the scale we focus on and is the approximately linear scale.

2. The dark matter particles or halos are redistributed to the cells by the Nearest Gridding Point (NGP).

Through these processes, we obtain a 3D image of the dark matter or halo distribution with a size of 40^3 from a simulation box. The voxel value is the density of the dark matter or halo. This size of the grid is larger than the typical scale of the redshift space distortion in the simulations. Therefore, we do not consider its effect in generating the images. To validate this assumption, we test our CNN by the images including the RSD effects, and we confirm the results are not changed from the case that we use the images without the RSD effects.

In addition, we also use the Legendre multipoles of the power spectrum for the dark matter distribution in the 2000 simulations. We then proceed to train and test the neural network using these multipoles. The Legendre multipoles are defined as Eq. (2.75). Here, we use the coefficient $P_0(k)$ (monopole), $P_2(k)$ (quadrupole), and $P_4(k)$ (hexadecapole), and these coefficients are publicly available for each Quijote simulation we use. We use a wave number range of $0.089 < k\ [h\mathrm{Mpc}^{-1}] < 0.25$, divided into 39 linearly separated bins to match the grid size of the 3D images for our CNN.

In this work, we divide the 2000 simulations into the subgroups of the 1500, 100, and 400 simulations. The first one is used as the training data, which is the data for the update of the weight parameter, the second is used as the validation data, which is the data for the monitoring of the progress of the training, and the last one is the test data, which is the data for the evaluation of our CNN and neural network.

4.2 Methods

4.2.1 Error Estimation from Fisher matrix

For comparison with our machine learning analysis, we forecast the measurement errors of the cosmological parameters with the Fisher matrix from the forthcoming galaxy surveys (Tegmark, 1997; Seo and Eisenstein, 2003; Yahia-Cherif et al., 2021; Blanchard et al., 2020a). We can obtain an optimistic estimation of the errors of the cosmological parameters by assuming some specifications for a survey through the Fisher matrix approach.

Firstly, we assume a fiducial galaxy survey split in N_{bins} redshift bins. The Fisher matrix in each redshift bin can be written as (Tegmark, 1997; Seo and Eisenstein, 2003; Yahia-Cherif et al., 2021; Blanchard et al., 2020a)

$$F^{\mathrm{bin}}_{\alpha\beta}(z_i) = \frac{1}{8\pi^2}\int_{-1}^{1} d\mu \int_{k_{\min}}^{k_{\max}} k^2\, dk \left[\frac{\partial \ln P_{\mathrm{obs}}}{\partial \theta_\alpha}\frac{\partial \ln P_{\mathrm{obs}}}{\partial \theta_\beta}\right] V_{\mathrm{eff}}(z_i, k, \mu), \tag{4.8}$$

where μ is the cosine of the angle between the line of sight and the wave vector, θ_α is the set of parameters varied in this analysis (specified below), P_{obs} is the observed power spectrum, and V_{eff} is the effective survey volume described as

$$V_{\mathrm{eff}}(k, \mu; z) = V_s(z)\left[\frac{n(z)\, P_{\mathrm{obs}}(k, \mu; z)}{n(z)\, P_{\mathrm{obs}}(k, \mu; z) + 1}\right]^2, \tag{4.9}$$

where V_s is the comoving survey volume, and $n(z)$ is the number density of galaxies. For the range of the integration in Eq. (4.8), $k_{\min}$ is determined by the simulation box size, and $k_{\max}$ is $= 0.25\ h\mathrm{Mpc}^{-1}$ corresponding to the grid size of the images for our CNN analysis. The scale of $k_{\max}$ corresponds to the scale where the non-linearities become important for $z = 0$, based on the criterion $\Delta^2(k) \equiv k^3 P(k)/(2\pi^2) \sim 1$ (Dodelson and Schmidt, 2020).

For the Fisher analysis by Eq. (4.8), we need to model the non-linear effects in the galaxy clustering for the observed power spectrum P_{obs}. The model is typically a phenomenological model and is employed in this work and in some works

(Tegmark, 1997; Seo and Eisenstein, 2003; Yahia-Cherif et al., 2021; Blanchard et al., 2020a). In this approach, the observed power spectrum an be written as

$$P_{\text{obs}}(k,\mu;z) = \frac{1}{q_\perp^2\, q_\parallel} \frac{\left[b\sigma_8(z) + f\sigma_8(z)\mu^2\right]^2}{1 + [f(z)k\mu\sigma_{\text{p}}]^2} \frac{P_{\text{dw}}(k,\mu;z)}{\sigma_8^2(z)}\, F_z(k,\mu;z) + P_s(z), \tag{4.10}$$

where $q_\perp$ and $q_\parallel$ are the Alcock-Paczynski projection coefficients, b is the bias of the observable, σ_p is a non-linear parameter representing the deviation from the linear prediction and calculated from the linear power spectrum, P_{dw} is the de-wiggled power spectrum, which is the linear power spectrum without the feature of the baryon acoustic oscillation (BAO), F_z describes the smearing of the galaxy density field along the line of sight due to the redshift error, and $P_s(z)$ is the scale-independent offset due to the imperfect removal of the shot noise (Yahia-Cherif et al., 2021; Blanchard et al., 2020b). Here, we set $q_\perp = q_\parallel = 1$ because the Alcock-Paczynski effect, which is the geometrical distortion of the 3D galaxy distribution due to the error of the measurement of the distance to the observable, is irrelevant to our work. P_{dw} describes the effect of the non-linear spreading caused of the matter distribution caused by BAO during the structure formation. F_z is $= \mathrm{e}^{-k^2\mu^2\sigma_r^2(z)}$, where σ_r is the error of the measurement of the comoving distance and is written as

$$\sigma_r(z) = \frac{\partial r}{\partial z}\sigma_z(z) = \frac{c}{H(z)}(1+z)\,\sigma_{0,z}, \tag{4.11}$$

where $\sigma_{0,z} = 0.001$ represents the redshift error Wang, Chuang, and Hirata, 2013; Blanchard et al., 2020b.

In our work, we calculate the linear power spectrum numerically using the Boltzmann code, the Cosmic Linear Anisotropy Solving System (`CLASS`) (Lesgourgues, 2011).

To compute the Fisher matrix Eq. (4.8), we need the derivative of P_{obs} with respect to a parameter. We can only evaluate its derivative numerically. Therefore, we use a central difference scheme for the derivative. We confirm the convergence of the Fisher matrix calculation by varying the parameter ϵ representing the small difference to compute the derivative. We set $\epsilon = 0.01$ as the default value, and we find the robustness of the Fisher analysis for the different values of ϵ.

For P_{dw}, we use a Savitzky-Golay filter with a window size of 201 and polynomial order equal to 3 and obtain a de-wiggled power spectrum from the power spectrum computed by the CLASS including the BAO feature. We implement the Savitzky-Golay filter by `scipy.signal.savgol_filter`(data, window_length, poly_order) in the `scipy` (Virtanen et al., 2020) python module.

As the bias parameter b depends on the redshift, we use the value calculated for the simulations to compare the Fisher analysis with the machine learning fairly. To calculate b, we compute the power spectrum for both halo and dark matter for the simulations and fit the ratio of the power spectrum of halo to the one of dark matter at linear scale $k \sim 0.01\, h\text{Mpc}^{-1}$ by using the form of

$$b_{\text{halo}}(z) = A + B\, e^{C\,z} \tag{4.12}$$

The obtained values are $A = 0.648$, $B = 0.689$, and $C = 0.792$.

Subsequently, the total Fisher matrix is calculated by summing over all the redshift bins, that is,

$$F_{\alpha\beta} = \sum_{i=1}^{N_{\mathrm{bins}}} F^{\mathrm{bin}}_{\alpha\beta}(z_i), \tag{4.13}$$

where $F^{\mathrm{bin}}_{\alpha\beta}(z_i)$ is a Fisher matrix for a redshift bin. In our analysis, we choose the cosmological parameter set $\theta_{\mathrm{cosmo}} = \{\Omega_{\mathrm{b},0}, \Omega_{\mathrm{m},0}, h, n_s, \sigma_8\}$ and the non-cosmological parameters $\theta_{\mathrm{non-cosmo}} = \{\sigma_p, P_s\}$ (Blanchard et al., 2020a). The fiducial values of the cosmological parameters are the same as the fiducial values of the Quijote simulations, $\theta_{\mathrm{cosmo,fid}} = \{\Omega_{\mathrm{b},0} = 0.05, \Omega_{\mathrm{m},0} = 0.32, h = 0.67, n_s = 0.96, \sigma_8 = 0.816\}$.

The total vector of the parameters for the Fisher matrix is $\theta_{\mathrm{total}} = \theta_{\mathrm{cosmo}} + \theta_{\mathrm{non-cosmo}}$. In general, other parameters such as the angular diameter distance and the Hubble parameter are considered in the Fisher analysis. However, our analysis assumes the ΛCDM model and these parameters depend on the cosmological parameters. It is enough to consider θ_{total} in our analysis.

Finally, by determining the total Fisher matrix, we consider the error propagation from the covariance matrix $C_{\alpha\beta} \equiv F^{-1}_{\alpha\beta}$ of θ_{total} to the covariance of the growth rate $f\sigma_8(z_i)$ in each redshift bin. This process can be described as

$$C_{ij} = \frac{\partial f\sigma_{8,i}}{\partial\theta_\alpha}\frac{\partial f\sigma_{8,j}}{\partial\theta_\beta} C_{\alpha\beta}, \tag{4.14}$$

where $f\sigma_{8,i}$ represents $f\sigma_8(z_i)$ and we can calculate the $\partial f\sigma_{8,i}/\partial\theta_\alpha$ numerically. And then, the error of $f\sigma_8$ is the diagonal elements of C_{ij} and can be written as

$$\sigma^2_{f\sigma_8(z_i)} = C_{ii}. \tag{4.15}$$

We perform the extensive comparison with other analyses such as (Yahia-Cherif et al., 2021; Blanchard et al., 2020a) and we confirm that the results of our analysis are in agreement with other analyses.

4.2.2 Error Estimation by Maximum Likelihood analysis

For comparison, we estimate the $f\sigma_8$ value by the maximum likelihood analysis. In this work, we model the Legendre multipoles P_0, P_2, P_4 of the dark matter power spectrum defined by Eq. (2.75) following Noriega et al., 2022 (please see section 3.3 of (Noriega et al., 2022)). This model of the Legendre multipole includes the 1-loop correction of the perturbation theory and the modeling of the non-linear part. We use `CLASS`(Lesgourgues, 2011) and `FOLPS-nu` for the non-linear power spectra[1].

In our work, we focus on the analysis of the dark matter density field for the maximum likelihood analysis, and the bias parameter b_1 and b_2 are unity and zero, respectively. Here, we estimate 5 cosmological parameters [Ω_m, Ω_b, h, n_s, σ_8] and 4 nuisance parameters [α_0, α_2, α_0^{shot}, α_2^{shot}], where the first two parameters model the non-linear part of the power spectrum and the latter two parameters model the shot noise.

We fit the modeled power spectrum $P_l^{\mathrm{model}}(k)$ to the power spectra of the Quijote simulation data $P_l^{\mathrm{data}}(k)$ by maximizing the likelihood for $k < 0.25$ $[h\mathrm{Mpc}^{-1}]$ defined as

$$L = [P^{\mathrm{model}}(k_i) - P^{\mathrm{data}}(k_i)]\mathrm{Cov}^{-1}(k_i, k_j)[P^{\mathrm{model}}(k_j) - P^{\mathrm{data}}(k_j)], \tag{4.16}$$

---

[1] https://github.com/henoriega/FOLPS-nu?tab=readme-ov-file

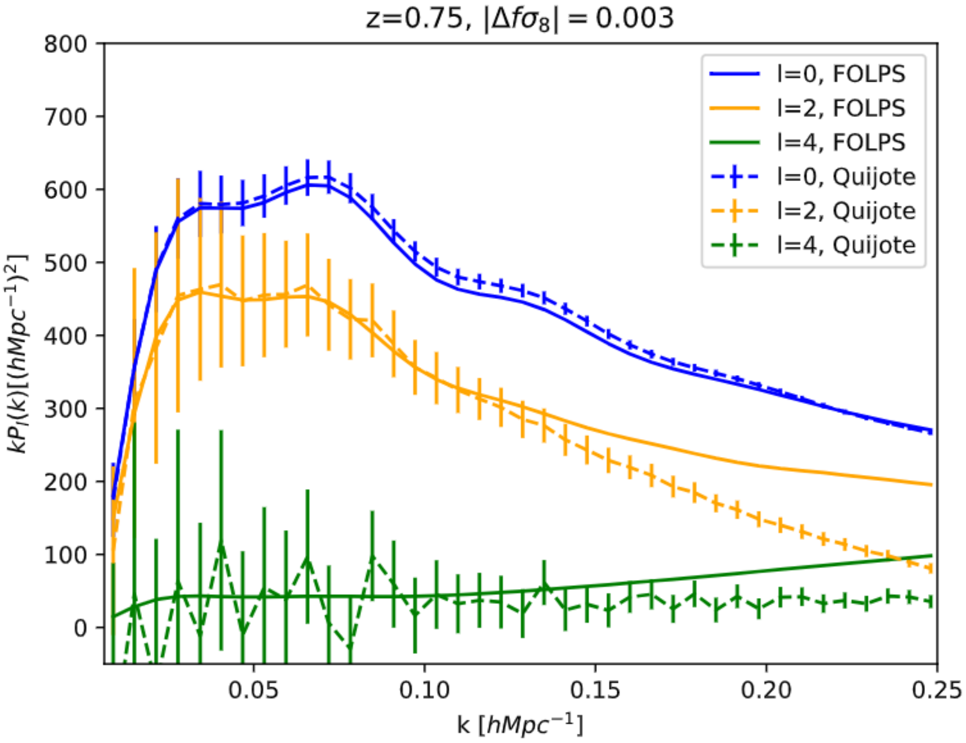

FIGURE 4.2: The modeled Legendre multipoles of the power spectra by using `FOLPS-nu` (dashed), where the parameters are fit to the fiducial simulation, and the ones calculated directly for the fiducial simulation data (solid). These power spectra are computed for $z = 0.75$ and the error bar is estimated from the 2000 realization for the fiducial cosmology.

where k_i is the wavenumber of the center of the i-th bin, $P^{\mathrm{model}}(k_i)$ and $P^{\mathrm{data}}(k_i)$ are the concatenated Legendre multipoles for $l = 0, 2$, and 4 for the power spectra of the model and the simulation data, respectively, and $\mathrm{Cov}^{-1}(k_i, k_j)$ is the covariance matrix of $P^{\mathrm{data}}(k)$. In our work, we calculate the covariance matrix by using the 2000 realizations of the fiducial cosmology in the Quijote simulations, where the random seeds of the initial condition are different but the assumed cosmological parameters are the same; $[\Omega_m, \Omega_b, h, n_s, \sigma_8] = [0.3175, 0.049, 0.6711, 0.9624, 0.834]$.

And then, we picked up 400 realizations randomly, where the parameters are determined by the Latin-Hypercube described in Section 4.1.1, and compute the $f\sigma_8$ following Eq. (4.3) for the parameters maximizing the likelihood for each 400 realization. To maximize the likelihood (equivalent to the minimizing of the $-\log L$), we use the `optimize.minimize()` in `scipy`, which is the Python module. Figure 4.2 shows the Legendre multipoles calculated from the fiducial simulation and the model of the power spectra fitted to the model as an example.

Finally, we evaluate the error of the $f\sigma_8$ estimation by the standard deviation of $\Delta f\sigma_8$, which is the difference between the prediction from the maximum likelihood and the true value for the realization.

4.2.3 Machine Learning architecture

	Layer	Output map size
1	Input	$40 \times 40 \times 40 \times 1$
2	$3 \times 3 \times 3$ convolution	$38 \times 38 \times 38 \times 2$
3	BatchNorm3d	$38 \times 38 \times 38 \times 2$
4	$2 \times 2 \times 2$ MaxPool	$19 \times 19 \times 19 \times 2$
5	$2 \times 2 \times 2$ convolution	$18 \times 18 \times 18 \times 64$
6	BatchNorm3D	$18 \times 18 \times 18 \times 64$
7	$2 \times 2 \times 2$ MaxPool	$9 \times 9 \times 9 \times 64$
8	$3 \times 3 \times 3$ convolution	$7 \times 7 \times 7 \times 64$
9	$3 \times 3 \times 3$ convolution	$5 \times 5 \times 5 \times 64$
10	$2 \times 2 \times 2$ convolution	$4 \times 4 \times 4 \times 128$
11	BatchNorm3d	$4 \times 4 \times 4 \times 128$
12	Flatten	$8192 (= 4^3 \times 128)$
13	FullyConnected	512
14	FullyConnected	256
15	FullyConnected	1

TABLE 4.2: Our CNN architecture. In all convolutional layers, stride $s = 1$. Output map size corresponds to (height, width, depth, and channel), where the channel represents the number of input images to the layer. After each convolution layer and FullyConnected layer, we apply the ReLU as the activation function except for the last layer. The total number of trainable parameters is 5,345,341.

In this work, we use the 3D-CNN for the image analysis and the neural network for the power spectrum analysis. To implement the CNN and neural network, we use the PyTorch (Paszke et al., 2019), which is one of the Python modules for deep learning, and these architectures are based on (Lazanu, 2021b). Note that some hyperparameters are different from this previous work because the size of the input image is different. Table 4.2 and Table 4.3 show the architectures of our 3D-CNN and neural network.

	Layer	Output size
1	Input	3×39
2	FullyConnected	512
3	FullyConnected	512
4	FullyConnected	512
5	FullyConnected	512
6	FullyConnected	1

TABLE 4.3: Our neural network architecture. After each FullyConnected layer, we apply the dropout layer with rates of 0.1 except for the last layer. The total number of trainable parameters is 848897.

In our CNN, we use ReLU as the activation function in the convolution and FullyConnected layers except for the last layer. Also, in our neural network, ReLU is used in the FullyConnected layers except for the last layer. In both our CNN and neural network, the activation function is the identity function in the last layer, and the output is the predicted value of the growth rate $f\sigma_8$. The batch normalization and dropout are used in our CNN and neural network, respectively. Please see the caption of Table 4.2 and Table 4.3 for the details.

We apply the mini-batch learning for both our CNN and neural network with the batch size $N_b = 16$. As the loss function, we use the MSE loss;

$$E = \frac{1}{N_b} \sum_{i=1}^{N_b} (y_i - \hat{y}_i)^2. \tag{4.17}$$

where y_i is the predicted value of $f\sigma_8$ from our CNN and neural network, and $\hat{y}_i$ is the correct value of the corresponding input data. We discuss other loss functions in Appendix. A.

In updating the weight parameters in our CNN and neural network, we use the Adam optimizer (Kingma and Ba, 2014). We implement this optimizer by using `torch.optim.Adam()` in PyTorch. For our CNN, we set $\mathrm{lr} = 5 \times 10^{-7}$ and $\mathrm{weight\_decay} = 0.1$, which are the argument of `torch.optim.Adam()`, and correspond to the learning rate and the hyperparameter of the L2 regularizatio, respectively. For our neural network, we set $\mathrm{lr} = 5 \times 10^{-6}$ and $\mathrm{weight\_decay} = 0$. We use default values for other parameters in `torch.optim.Adam()` for both our CNN and neural networks.

For our CNN, the input is the $40 \times 40 \times 40$ image of the dark matter or halo distribution. For our Neural network, The input is the Legendre multipoles $P_0(k)$, $P_2(k)$, and $P_4(k)$. The size of the input is 3 [$P_0(k)$, $P_2(k)$, and $P_4(k)$] $\times$ 39 [the number of k-bins].

Our CNN and neural network are trained during 100 and 1000 epochs, respectively, and we use the set of weight parameters that minimizes the loss function for the validation data. The errors of the prediction by the machine learning are estimated by the standard deviation of the vector $\Delta f\sigma_{8,i}$, which is the difference between the predicted and true $f\sigma_8$ and defined as $\Delta f\sigma_8 \equiv (y_{\mathrm{pred},i} - y_{\mathrm{true},i})$, where i is the label of the input data, and$y_{\mathrm{pred},i}$ and $y_{\mathrm{true},i})$ are the predicted and true value of the $f\sigma_{8,i}$, respectively. We validate this estimation by the following test:

- We make the 30 sets of the cosmological parameters for $\Omega_m \in [0.1, 0.5]$ and $\sigma_8 \in [0.6, 1.0]$.

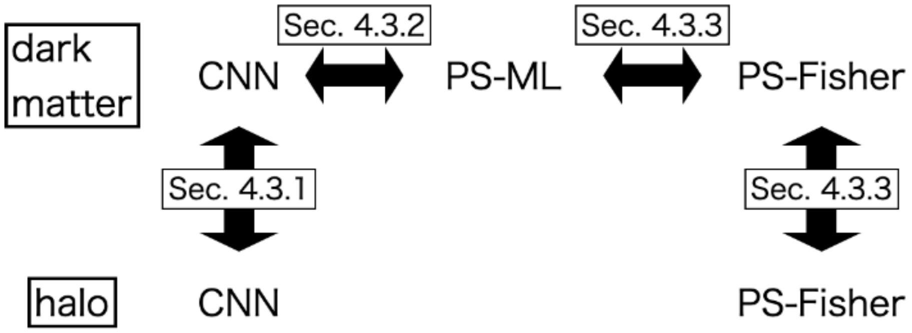

FIGURE 4.3: This is an overview of our work, highlighting the various comparisons between the different reconstruction methods. PS-ML represents the Power Spectrum and the Machine Learning, and CNN stands for Convolutional Neural Networks.

- We evaluate the $f\sigma_8$ value for the above parameter sets, and pick the 400 sets (same as the number of the test data).
- We draw the predictions for $f\sigma_8$ from the Gaussian distribution with mean given by the value of each parameter set and with the standard deviation σ. The assumption that the predictions follow Gaussian is justified by a large number of the data and the central limit theorem.
- We compare the predicted error from the standard deviation of the $\Delta f\sigma_8$ of the previous step with the fiducial error δ assumed in the previous step.

For this test, the predicted errors from the $f\sigma_8$ agree with the fiducial error σ.

4.3 Results

method	CNN		PS-ML	Likelihood	Fisher	
tracer	DM	halo	DM	DM	DM	halo
$z = 0.75$	3.8	4.5 (3.9)	2.3	7.9	1.4	0.34
1.25	2.3	2.5 (2.2)	2.0	5.7	1.6	0.46
1.75	1.2	1.7 (1.1)	2.7	3.9	1.6	0.49
2.5	0.74	1.2 (1.2)	2.9	3.3	1.4	0.46

$(\times 10^{-2})$

TABLE 4.4: The predicted errors on the $f\sigma_8$ as derived by different methods and tracers, where the combination of methods and tracers are summarized in Figure 4.3. The values for "halo" are obtained from the halo data, drawn randomly from the parent halo to match the redshift distribution shown in Table. 4.1) for CNN, while those in parenthesis are from the parent halo sample as a reference.

Here, we show and compare the results of the inference of the growth rate $f\sigma_8$ from the dark matter or halo distribution in the simulations by the image analysis with the CNN and the power spectrum analysis with the neural network and Fisher matrix. Firstly, in Section 4.3.1, we show the results of the image-based analysis with the CNN for both the images of the dark matter and halo. Additionally, we discuss the effects of the random drawing of the halo. Next, in Section 4.3.2, we show the results of the power spectrum analysis with the neural network, denoted as PS-ML in the following. And then, we show the results of the Fisher analysis for the power spectrum of the dark matter and halo distribution in Section 4.3.3. In addition, we discuss the effect of the random seed for the initial condition of the simulations to investigate the origin of the errors in Section 4.3.4. Figure 4.3 shows the overview of the comparison between the different methods of the parameter inference, and Table 4.4 summarizes the error of the estimation from the different methods (CNN, PS-ML, and Fisher) and tracers (dark matter and halo).

4.3.1 CNN results for dark matter and halo

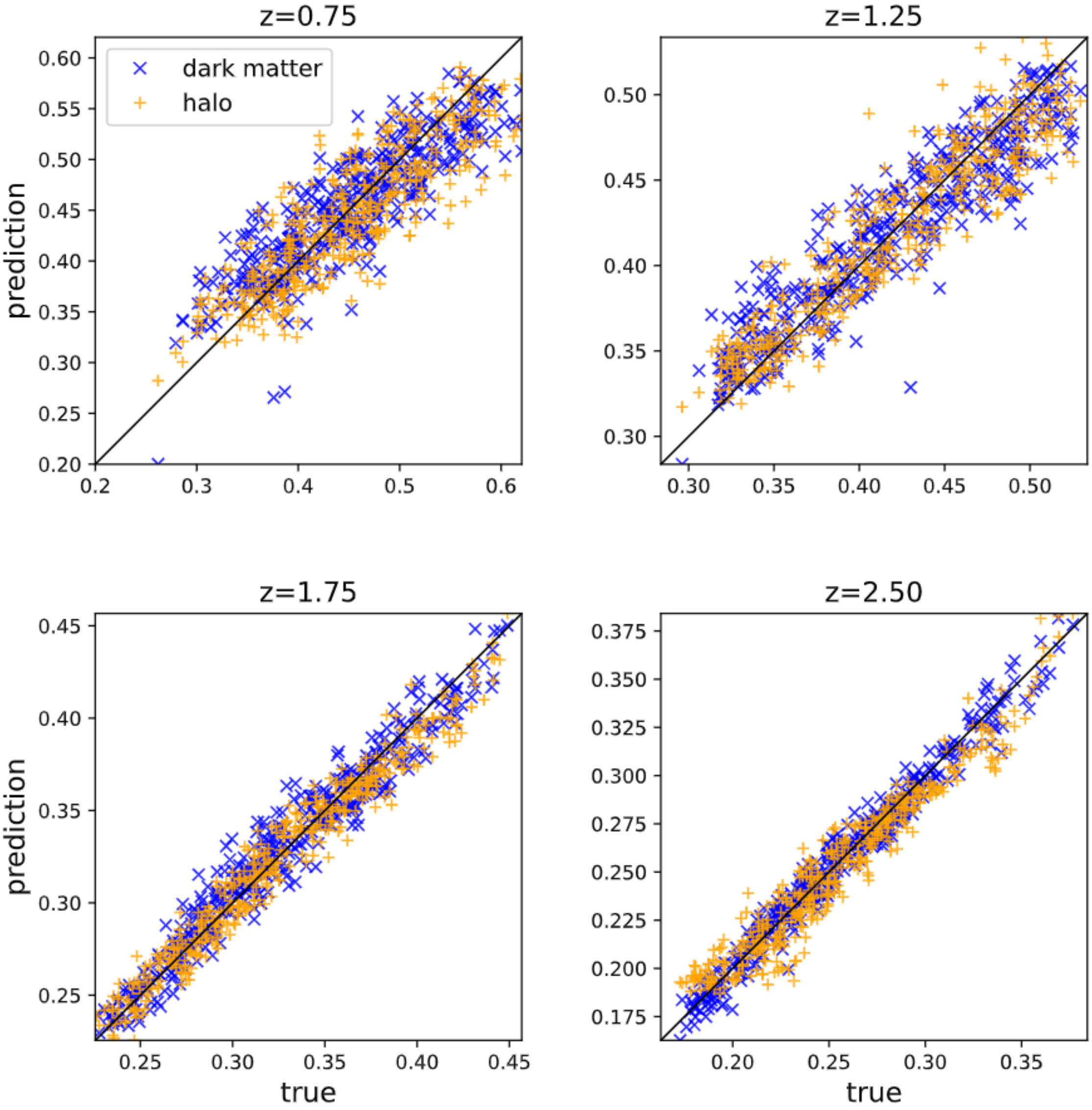

FIGURE 4.4: The scatter plot of the prediction of $f\sigma_8$ from our CNN for dark matter images (blue) and halo images, including all halo (orange) for each redshift. The horizontal and vertical axis shows the true and predicted value of $f\sigma_8$, respectively.

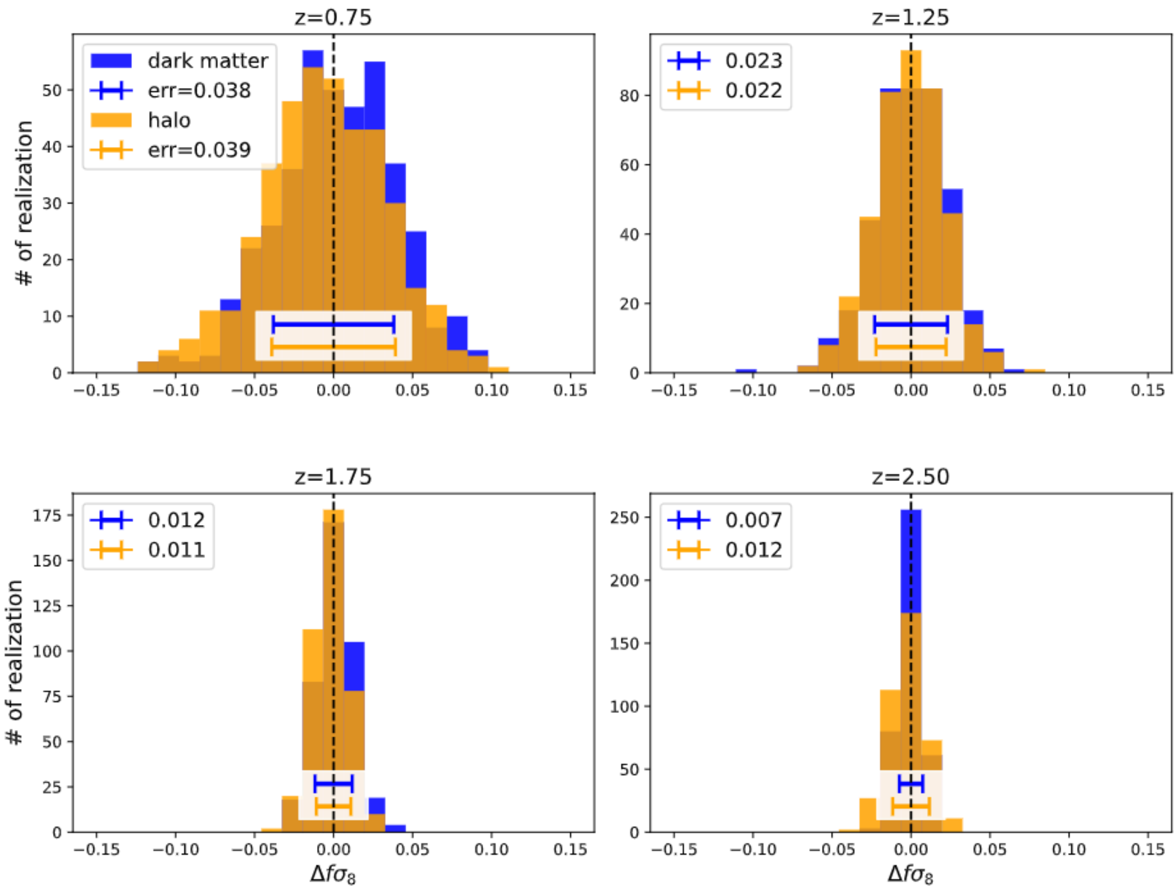

FIGURE 4.5: The histograms of the CNN prediction for dark matter images (blue) and halo images (orange) for each redshift bin. The horizontal axis corresponds to the value of $\Delta f\sigma_8$ predicted by our CNN. The error bars show the standard deviation of $\Delta f\sigma_8$; its value is shown in the legend. As seen, all histograms are centered around zero, indicating that our CNN successfully constrains $f\sigma_8$.

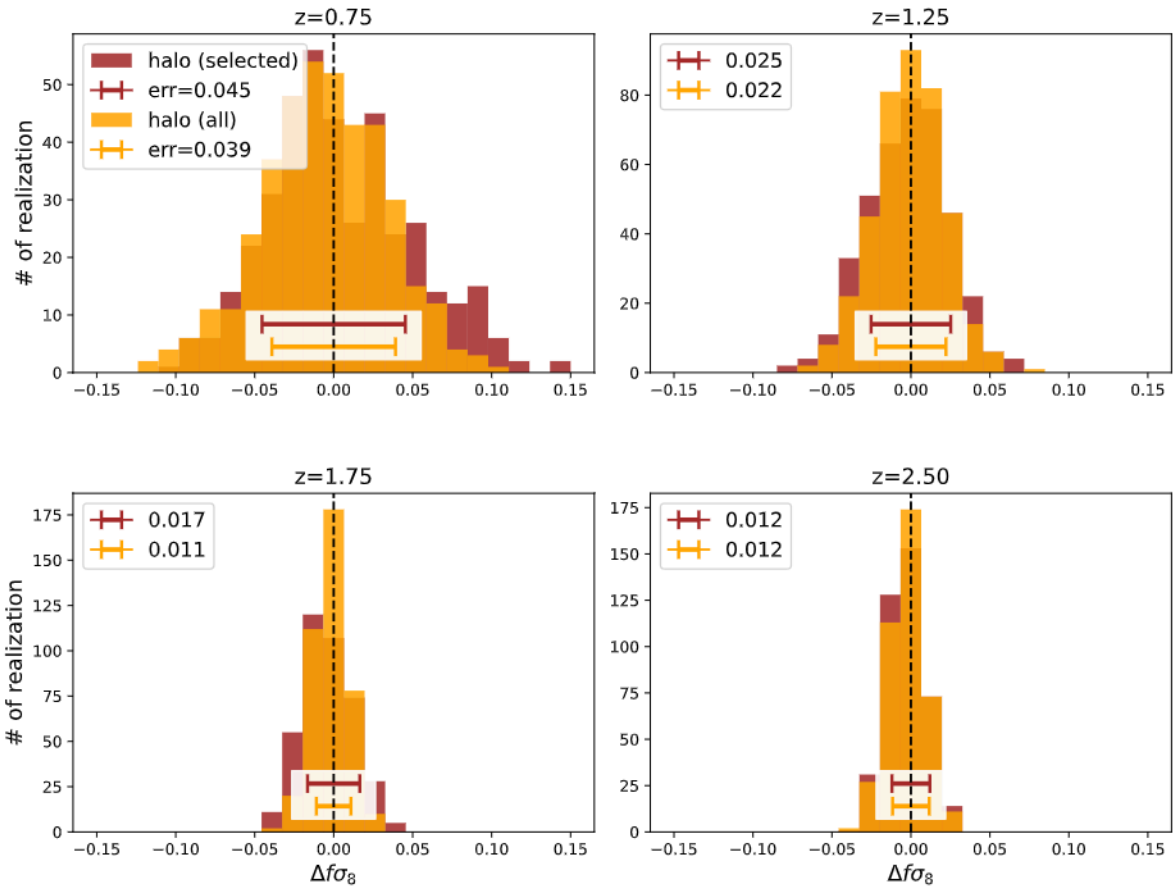

FIGURE 4.6: The results of our CNN for images include all halos (orange, same as Figure 4.5) and randomly selected halos (brown) for each redshift bin. Each axis, the error bar, and the legend represent the same as Figure 4.5

Here, we show the results of the prediction of the $f\sigma_8$ with our CNN shown in Table 4.2 for the images of the dark matter and halo distribution in the simulations. The process of the training and the error evaluation is mentioned in Section 4.2.3.

Figure 4.4 shows the scatter plot of the predicted and true value of the $f\sigma_8$ for each test image in the vertical and horizontal axis, respectively, and each panel corresponds to the different redshift bin. The blue and orange plots represent the prediction of our CNN for the dark matter and halo images, respectively. Figure 4.5 shows the histogram of $\Delta f\sigma_8$ predicted by our CNN. The horizontal axis represents the $\Delta f\sigma_8$, the vertical axis shows the frequency, and each panel corresponds to the different redshift bin. In addition, the error bars, which is predicted by our CNN and the standard deviation of $\Delta f\sigma_{8,i}$, are shown in this Figure, and the values of the errors are provided in the legend. As we can see in Figure 4.5, all histograms are centered around $\Delta f\sigma_8 = 0$, and we find our CNN successfully constrains the value of the $f\sigma_8$.

In these figures, we use the halo images including all of the halos (without drawing the halo) to compare the results for the dark matter and halo. As a result, we find the errors are mostly comparable for each redshift bin, while the error for the dark matter images is a bit smaller than the one for the halo images at the redshift bin of $z = 2.5$. And then, we can see a decrease in the error for both the dark matter and halo image. One of the possibilities is that the non-linearity makes it more challenging to extract the information included in the matter distribution at the lower redshift.

The error for the halo images is larger than the one for the dark matter at $z = 2.5$, while these errors are comparable at $z = 1.75$. We can consider that the one reason is the shot noise because the number of the halo in the simulation at $z = 2.5$ is less than the one-tenth of the simulations at the other redshift bins in using all of the halos. The shot noise is proportional to the inverse of the number density of the halo $1/n$ and has a larger impact on the analysis at the higher redshift.

Next, we show the effects of the random drawing of the halos in Figure 4.6. This figure also shows the histogram of the $\Delta f\sigma_8$ from our CNN, like Figure 4.5, for all halos (orange, same as Figure 4.5) and the randomly drawn halos (brown), where the halos in the simulation are drawn randomly to follow the redshift distribution shown in Table 4.1. Even in the case of the randomly selected halos, we find the same dependence of the error on the redshift in the case of all halos.

The error for the selected halo is slightly larger than the one for all halos. This is reasonable because the images of the selected halo distribution lose the information compared to the ones for all halos.

In addition, we test some loss functions other than the MSE loss function, such as the MAE loss, Hubber loss, and LFI loss, which are introduced in Section 3.2.1. We find that the error is slightly improved at low redshift with the LFI loss. However, the loss curve, which represents the change of the value of the loss function over the training, for the validation data, is more stable with the MSE loss. Therefore, we choose the MSE loss for the final results. In Appendix A, we show the error with other loss functions.

4.3.2 Power Spectrum analysis on Dark Matter based on Machine Learning

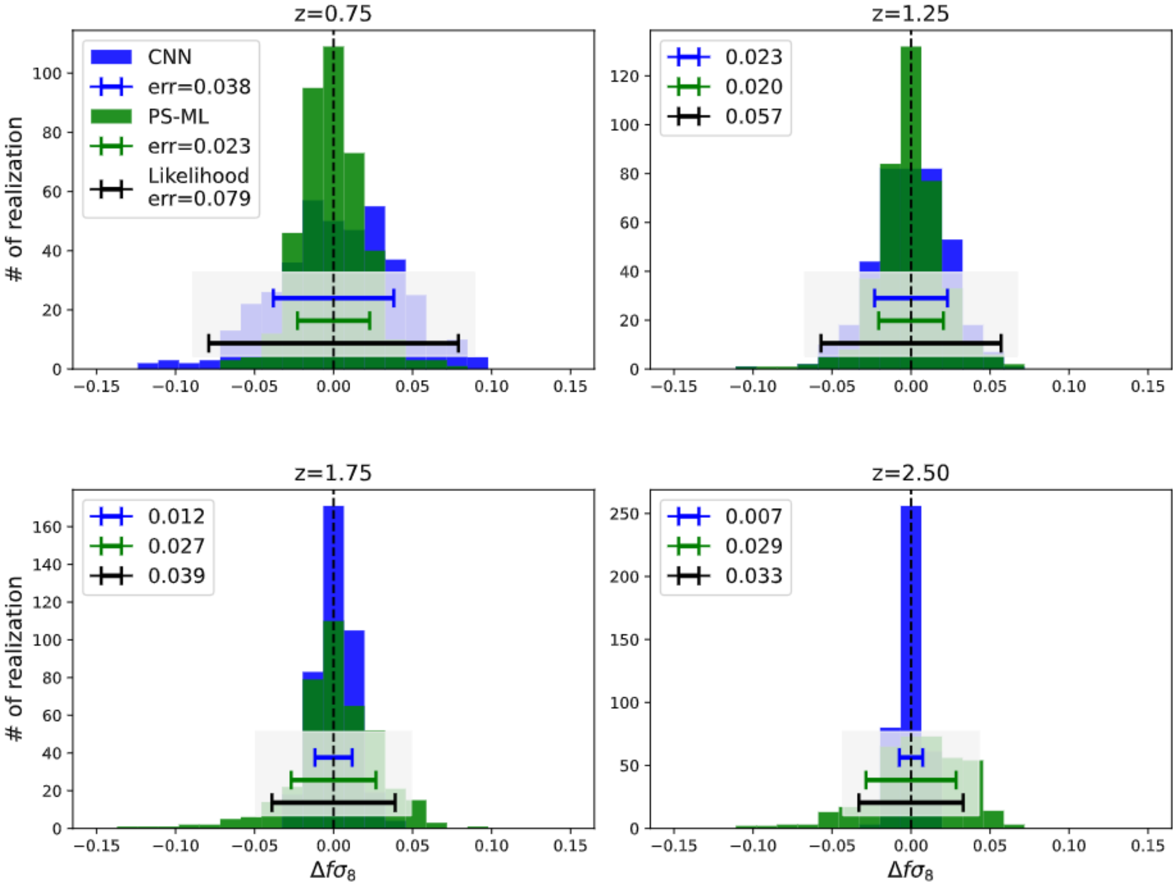

FIGURE 4.7: The histograms of our CNN for dark matter images (blue, same as Figure 4.5) and PS-ML (green) for each redshift bin. The horizontal axis corresponds to $\Delta f\sigma_8$, and the vertical axis shows the frequency of $\Delta f\sigma_8$ for each test image. The error bars show the standard deviation of $\Delta f\sigma_8$, and its value is shown in the legend. In addition, the black error bar shows the standard deviation of $\Delta f\sigma_8$ for the likelihood analysis.

In this subsection, we show the results of PS-ML, where the architecture of the neural network is defined in Table 4.3. The process of the training and the error evaluation is mentioned in Section 4.2.3.

Figure 4.7 shows the histograms of $\Delta f\sigma_8$ for our CNN for the dark matter images (blue) and PS-ML (green). Both results are for the dark matter images and we compare them. As we can see in this figure, the constraints on $f\sigma_8$ from PS-ML are weaker than our CNN at higher redshift $z > 1.75$ (lower two panels), while PS-ML provides tighter constraints than our CNN as lower redshift $x < 1.25$ (upper two panels). The observed variations in constraints between PS-ML and our CNN raise interesting questions. One possibility is that the image recognition by our CNN might struggle to extract the features of the data in a highly non-linear regime corresponding to lower redshift. Another possibility is the architecture of our CNN may not be perfectly optimized for the lower redshift data. Further investigation is needed to explore these potential explanations.

Additionally, we compare the error from the PS-ML and the maximum likelihood analysis for the dark matter density field. The errors from the likelihood show the redshift dependence similar to the CNN analysis. We can consider that this redshift dependence is caused by the difficulty of the modeling of the non-linear part of the power spectrum. On the other hand, the PS-ML does not show the redshift dependence. The PS-ML learns the power spectra directly including the non-linear part and can probably estimate the $f\sigma_8$ more effectively than the maximum likelihood analysis. At the higher redshift ($z = 2.5$), the model of the Legendre multipoles can express the power spectra including the non-linear part better, and the PS-ML and the likelihood analysis show a comparable result.

4.3.3 Comparison with the Fisher analysis

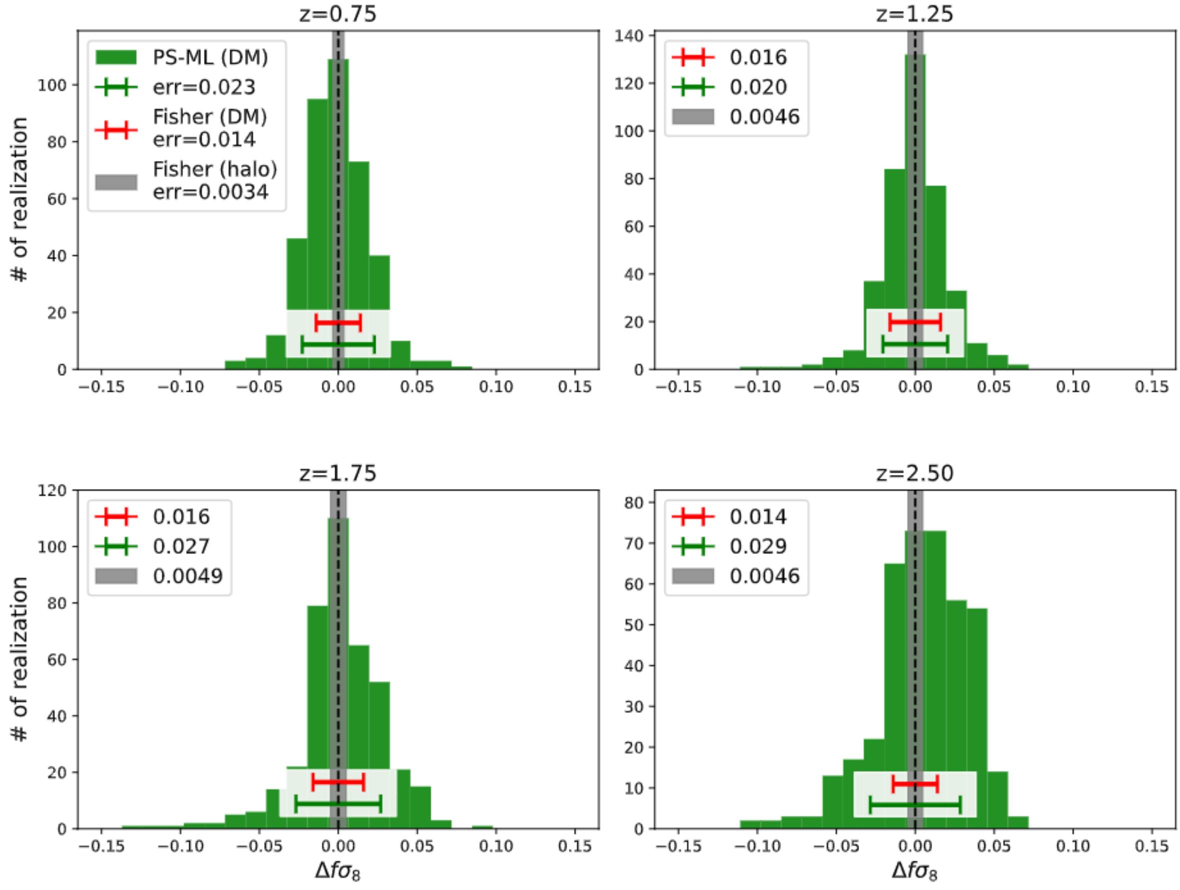

FIGURE 4.8: The histograms of $\Delta f\sigma_8$ from PS-ML (green, same as Figure 4.7), plotted together with the errors from the Fisher analysis for the power spectrum of dark matter (red) and the one of halo (grey-shaded). The horizontal axis corresponds to $\Delta f\sigma_8$, and the vertical axis shows the frequency of $\Delta f\sigma_8$ for each test image. The value of the error is shown in the legend.

In this subsection, we show the comparison of the error from PS-ML with the one from the Fisher matrix. Both analyses use the power spectrum of the dark matter distribution, so we compare the results of them.

The results are shown in Figure 4.8. The errors for the predicted value of $f\sigma_8$ of the halo distribution from the Fisher analysis are calculated by using the bias b of Eq. (4.12), and are shown in Figure 4.8 by the gray-shaded region. The errors for the dark matter are computed by setting $b = 1$, and shown in Figure 4.8 by the red error bar.

For the analysis of the dark matter distribution, the errors $\Delta f\sigma_8$ from PS-ML and the Fisher analysis are not largely dependent on the redshift. Additionally, the errors from the Fisher analysis are smaller than those from PS-ML. This suggests a potential limitation in our PS-ML's ability to access all information in the power spectra of the dark matter distribution, highlighting a potential area for improvement in our neural network architecture. Additionally, we can see the CNN shows comparable errors with the Fisher analysis in Table. 4.4 especially at the lower redshifts and CNN can estimate the $f\sigma_8$ without the modeling of the non-linear part.

As we can see in Figure 4.8, the error from the Fisher analysis for the halos (gray-shaded region) is smaller than the one for the dark matter (red error bar). This result can be explained by the shot noise. Firstly, the number density of the dark matter particles is larger than the one of the halos. In addition, the amplitude of the power spectrum of the halos is larger than the one of the dark matter because the halo power spectrum is amplified by the bias given by Eq. (4.12). Therefore, the shot noise for the halo has less effect on the analysis than the one for the dark matter.

4.3.4 Effect of the random seed

Each of the 2000 simulations, where the cosmological parameters are sampled by the Latin hypercube, has a different random seed of its initial condition. Therefore, for the parameter inference in the previous sections, the error of the predicted value includes the effects of both the variation of the cosmological parameters and the variance of the initial conditions. In this subsection, we discuss the effect of the random seed of the initial conditions on the errors $\Delta f\sigma_8$. To discuss the effect of the variation of the random seed, we use another dataset of the Quijote simulations. The Quijote simulations have 15000 realizations, which have the same cosmological parameters $(\Omega_{m0}, \Omega_{b0}, h, n_s, \sigma_{8,0}) = (0.3175, 0.049, 0.6711, 0.9624, 0.824)$ while Each realizations has the different random seed. In (Villaescusa-Navarro et al., 2020), these realizations are denoted as 'Fid' simulations.

Firstly, we pick up 7000 realizations from 'Fid' simulations. And then, from these simulations, we generate the 3D images of the randomly selected halo following the process in Section 4.1.3. Next, by using these images from 'Fid' simulations, we test our CNN already trained by the randomly selected halo in Section 4.3.1. From these processes, we can evaluate the error of the predicted value of $f\sigma_8$ from the variation of the random seed.

As a result, we find the error for 'Fid' simulations is around $\mathcal{O}(10^{-3})$. This value is one-fifth of the total error shown in Figure 4.5. Therefore, we conclude that the errors of the predicted values in our CNN analysis are dominated by the errors from the variation of the cosmological parameters. On the other hand, this value of the error for 'Fid' simulations is close to the one for the Fisher analysis of the halo power spectrum shown in Table 4.4. Therefore, the error from the Fisher analysis is affected by the variation of the random seed.

4.4 Discussion and Conclusion

In this work, we consider the CNN approach to extract information and to constrain the value of $f\sigma_8$ without modeling the power spectrum. We use the Quijote simulation to train and evaluate our CNN and compare the results from our CNN with those from the power spectrum analysis of the neural network and Fisher analysis.

We overview our analysis in our work in Figure 4.3. The figures 4.4, 4.5, 4.6, 4.7, and 4.8 show the results of the different tracers of the matter distribution and the different methods of the analysis, and Table 4.4 shows the summary of the estimated error for the results shown in these figures.

As we can see in these figures and Table 4.4, the machine learning approach can estimate the value of $f\sigma_8$ accurately. The histograms of $\Delta f\sigma_8$ shown in the figures are centered at around zero. And then, the predicted errors from our CNN for each redshift are comparable within a factor of order unity for both the CNN and neural network based on both of the images of the dark matter and halo distribution.

In comparing the results from the machine learning approaches with those from the Fisher analysis, we find that CNN shows comparable results at higher redshifts to the Fisher analysis. On the other hand, the Fisher analysis shows an optimistic error compared to the machine learning approach at lower redshifts, and the errors from the machine learning approaches show the redshift dependence. Specifically, the errors estimated by our CNN show a factor of reduction of two for low redshifts compared to the dark matter, and a factor of reduction of two to ten for our CNN for the halo in Figure 4.5. On the other hand, the errors from the Fisher analysis do not show the redshift dependence. These errors seem to track the redshift distribution of the number density of the ideal survey sample. To probe for the cause of this redshift dependence of the predicted errors from our machine learning approaches, we test several possibilities.

Firstly, one possibility of the reason that our CNN results have the redshift dependence and deviate from the Fisher results may be the effects that emerged at the low redshifts. The enhanced non-linearity at low redshifts might introduce more scatter in the predicted $f\sigma_8$ by our CNN. The Fisher analysis with the phenomenological model of the observed power spectrum might not correctly capture the effects of the non-linearity. The Fisher analysis seems to follow the redshift distribution of the observables in the ideal survey and ignores part of the complicated effects of the non-linearity at low redshifts.

Next, another possible reason for the redshift dependence of the errors from our CNN might be because $f\sigma_8$ shown in Figure 4.1 is not well sampled while the cosmological parameters $(\Omega_{m0}, \Omega_{b0}, h, n_s, \sigma_{8,0})$ are well sampled by the Latin-hypercube in the Quijote simulation. As we can see in Figure 4.1, the range of $f\sigma_8$ is decreased with redshift. This might introduce the reduction of the errors from our CNN because the range of $f\sigma_8$ expected by our CNN might be smaller at higher redshifts. However, we should note that the range of $f\sigma_8$ in the training sample is generally greater than the errors from our CNN. To test this effect, we use new training samples, where the range of $f\sigma_8$ is narrower, but we find no significant effect on the error estimations.

This possibility might be tested by either the larger number of simulations, which have a wider range of the cosmological parameters at higher redshifts, used to train our CNN or the direct sampling in $f\sigma_8$ space. We think that the main scope of our work is to highlight our methodology and these tests are beyond the scope.

Third, the processes of the training and test of our CNN have some stochasticity and randomness depending on the choices of the training and test dataset. To investigate this possibility, we test our CNN by the different test datasets with the different random seeds and calculate the mean and standard deviation of the errors from our CNN, As a result, we find that the mean of the errors is consistent with the values in Table. 4.4 and the standard deviation of the errors, i.e., the error of the error is much smaller. Therefore, this reason can only have a small contribution to the redshift dependence of our CNN results.

Additionally, we have more tests. First, we investigate the effects of the random selection of halos by using the different halos, but there is no significant difference in the results and we also see the redshift dependence in the case of all halos. Second, we examine the different loss functions as we mentioned in Section 4.2.3, and we see the same redshift dependence as the case of the MSE loss. Third, we test the different random seeds of the initial condition in Section 4.3.4, and we find the errors from the difference of the random seed in our CNN analysis are only one-fifth of the total error. Therefore, we conclude the dominant component of the error of our CNN analysis is the one from the variation of the cosmological parameters.

For future works, we leave the extension of our analysis of models beyond ΛCDM. To eliminate the assumption that we need to specify the cosmological model in the current $f\sigma_8$ measurements, we require the N-body simulations assuming another model such as the model based on the modified gravity. Also, we leave a more detailed comparison between our CNN and the Fisher analysis, as it would require significant modifications of the theoretical model on the non-linear regime for the Fisher analysis and more optimization of the CNN architecture to study the effects at low redshifts.

To conclude, the forthcoming galaxy surveys will provide a lot of high-quality observation data of the large-scale structure of the universe. To deal with these data effectively, we need new ways of analysis to minimize the theoretical errors from assumptions such as the modeling of the non-linear part of the power spectrum as small scales. In this work, we provide the first step in this direction, but more investigation will be required to bridge the gap between the theory and the observation.

Chapter 5

Constraint on Non-CDM mass

In this chapter, we show constraints on NCDM mass with machine learning. NCDM mass affects the large-scale structure of the universe at small scales, but the structure at small scales follows a non-Gaussian distribution due to the non-linear evolution. Instead of the power spectrum which cannot extract all information included in the higher-order statistics, we focus on the CNN analysis.

In this work, we assume neutral hydrogen (HI) as the observable because we can obtain its distribution from the observation of the 21cm signals (Section 2.8). We use the Gadget3-OSAKA simulation, which is the cosmological hydrodynamic simulation, and obtain the distribution of the dark matter and HI at redshift $z = 3$ in the post-reionization epoch. We classify the images from CDM and NCDM simulations using both CNN and the power spectrum and compare the results from each method.

In addition, we investigate the effects of some astrophysical assumptions and the thermal noise of the SKA-MID observation.

The contents of this chapter follow Murakami et al., 2023a

5.1 Data

5.1.1 GADGET3-Osaka simulation

In this work, a series of cosmological hydrodynamic simulations are conducted for both the Cold Dark Matter (CDM) model and Non-Cold Dark Matter (NCDM) models with varying particle masses. Specifically, the Dodelson-Widrow model is considered the NCDM model, as introduced in Section 2.7. In the range of the NCDM mass denoted by $m_{\rm DM}$ in our analysis, dark matter particles have gravitational interactions only with each other after their initial conditions at $z = 99$. The features of the initial dark matter distribution for each dark matter model are described by the matter power spectrum, assuming a Gaussian distribution for the initial conditions. The cosmological parameters are based on Planck satellite (Planck Collaboration et al., 2020): $\Omega_m = 0.311$, $\Omega_\Lambda = 0.689$, $\Omega_b = 0.049$, $h = 0.677$, and $\ln 10^{10} A_s = 3.047$, with $\ln 10^{10} A_s$ representing the amplitude of the primordial matter power spectrum.

We examine NCDM models with six distinct particle masses, logarithmically sampled from 10^3 to $10^{4.66}$ eV. A simulation is prepared for an NCDM model with a particle mass of 10^5 eV; however, throughout our analyses, this particular NCDM model cannot be distinguished from the CDM model. Consequently, we exclude it from further consideration.

Our focus in this work is solely on a single dark matter component. As illustrated in Figure 5.1, we present examples of the linear 2D matter power spectrum at $z = 3$ for the CDM model, as well as for NCDM models with particle masses of 10 keV and

1 keV. It is important to note that we concentrate on the projected 2D matter distribution of dark matter and HI in this work. Our focus is the 21 cm intensity mapping. Additionally, we want to avoid the effects of redshift space distortion and the light-cone effect and use the projected matter distribution with enough projection length. The power spectra are calculated for the 2D matter distribution, wherein the 3D matter distribution is projected onto a 2D plane over 50 h^{-1}Mpc along the line of sight. The 3D matter power spectra are computed using `CLASS` (Lesgourgues, 2011). As shown in the figure, the power spectrum's amplitude is suppressed for NCDM models, with suppression on larger scales for the lighter NCDM model.

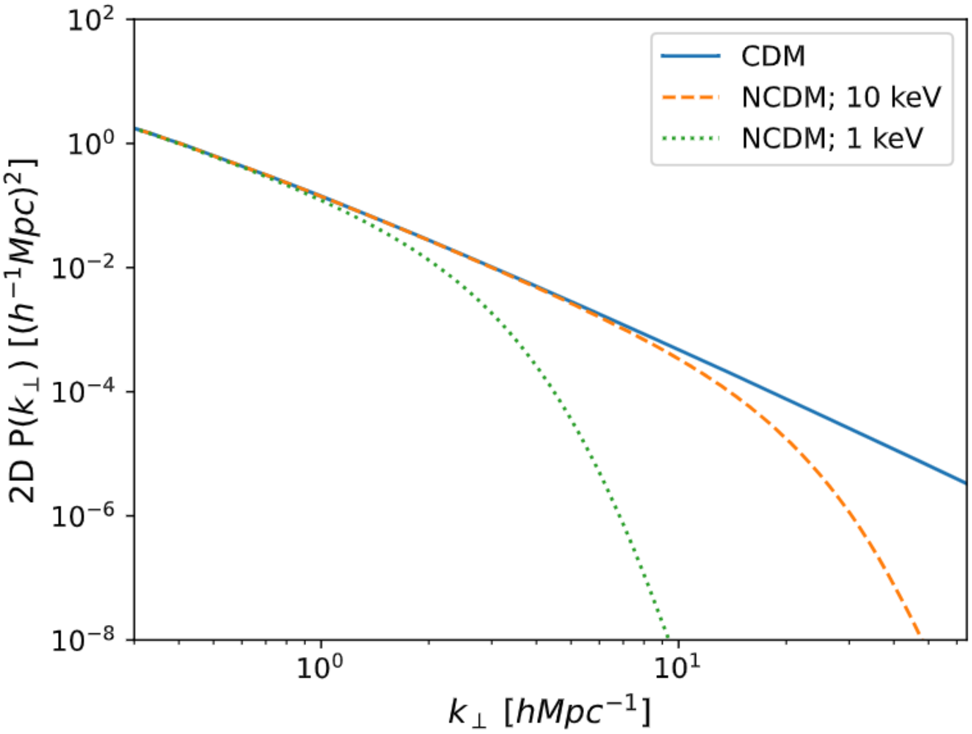

FIGURE 5.1: Examples of the 2D matter power spectra at $z = 3$, derived by projecting the 3D linear power spectrum along the line of sight with a width of 50 h^{-1}Mpc. The 3D power spectra are computed using `CLASS` (Lesgourgues, 2011)

We use the linear matter power spectrum calculated by `CLASS` as the initial power spectrum for our simulation. First, we calculate the matter power spectrum at redshift $z = 0$, and its amplitude is adjusted to the one at $z = 99$, which is the initial redshift of our simulations, by using the linear growth factor D_+ in Eq. (2.36). Finally, we generate the initial matter distribution by using `2LPTic` (Crocce, Pueblas, and Scoccimarro, 2006), where the distribution of the matter density perturbation is calculated based on the second-order Lagrangian perturbation theory (Scoccimarro, 1998). In generating the initial conditions, we apply the glass-like (White, 1994) realization to make a uniform and isotropic distribution. In general, the initial condition is generated from the grid-based realization, but previous works (Götz and Sommer-Larsen, 2002; Götz and Sommer-Larsen, 2003) show that initial conditions from grid-based realizations can introduce unrealistic features in the matter distribution. In our work, we test the effect of the choice of grid or glass and find AUC (see

Section 5.2.3) is slightly increased by $\mathcal{O}(10^{-2})$ for the case of the glass-like realization compared to the grid-based realization.

We use the cosmological hydrodynamic simulation code `GADGET3-Osaka` (Aoyama et al., 2016; Shimizu et al., 2019) to simulate the time evolution of the matter distribution. `GADGET3-Osaka` is a Smoothed Particle Hydrodynamics (SPH) code designed for cosmological simulations and is based on `GADGET-3` described in (Springel, 2005).

The simulation box has a size of 100 h^{-1}Mpc on each side, with 512^3 dark matter particles and 512^3 gas particles. Our simulations are started at $z = 99$ and terminated at $z = 3$, corresponding to the post-reionization epoch. In our analyses, we focus on the simulation data at $z = 3$, where the ionization processes of neutral hydrogen (HI) are mostly completed, and the HI distribution is minimally affected by these processes. Our simulation setup closely follows that of Nagamine et al., 2021, except for using different initial conditions to account for the NCDM models.

`GADGET3-Osaka` incorporates models for star formation, supernova feedback, UV radiation background, and radiative cooling and heating, which are related to HI ionization. The code also considers the self-shielding effect of optically thick HI gas, mitigating the impact of UV radiation that ionizes HI. Cooling in our simulations is solved using the Grackle chemistry and cooling library (Smith et al., 2016). In this work, we use the HI distribution from the simulations without modeling the relation between the distributions of dark matter and HI.

In this work, we explore the different astrophysical assumptions such as the self-shielding of the HI gas, star formation, and the UV background model related to the reionization of the universe. These assumptions are referred to as the astrophysical model in our work. To investigate the effects of the varying astrophysical models, we consider the following models:

- *Fiducial*

  We apply the star formation model in the AGORA project (Kim et al., 2014; Kim et al., 2016) and the supernova feedback model described in Shimizu et al., 2019. The self-shielding of the HI gas is not considered in this model. The reionization is driven based on the UV background model described in Haardt and Madau, 2012 at $z = 6$. For *Fiducial* model, we conduct the simulations for a CDM model and 6 NCDM models.

- *Shield*

  The configurations are almost the same as those of *Fiducial* model, but this model considers the self-shielding of the HI gas.

- *NoSF*

  This model does not consider star formation. The other configurations are the same as those of *Fiducial*.

- *FG09*

  This model applies the different UV background model (Faucher-Giguère et al., 2009) from the case of *Fiducial* model. The other configurations are the same as those of *Fiducial*.

The simulations for the *Shield*, *NoSF*, and *FG09* models are only conducted for the CDM model. The details of *Fiducial*, *Shield*, and *FG09* are discussed in Nagamine et al., 2021.

Basically, this work focuses on the classification of the models of CDM and NCDM based on *Fiducial* model. In Section 5.3.2, we investigate the effect of the varying astrophysical models.

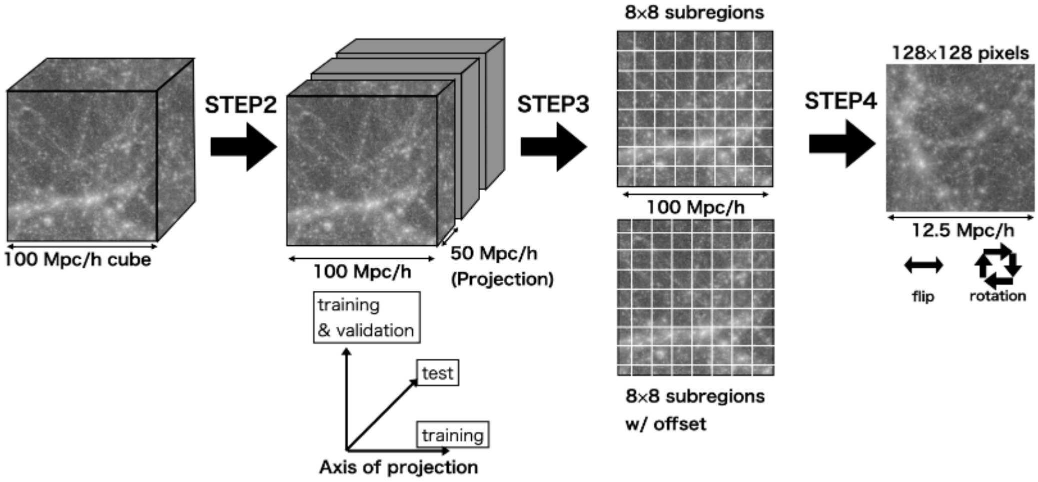

FIGURE 5.2: Illustration of the procedure for generating images from simulations.

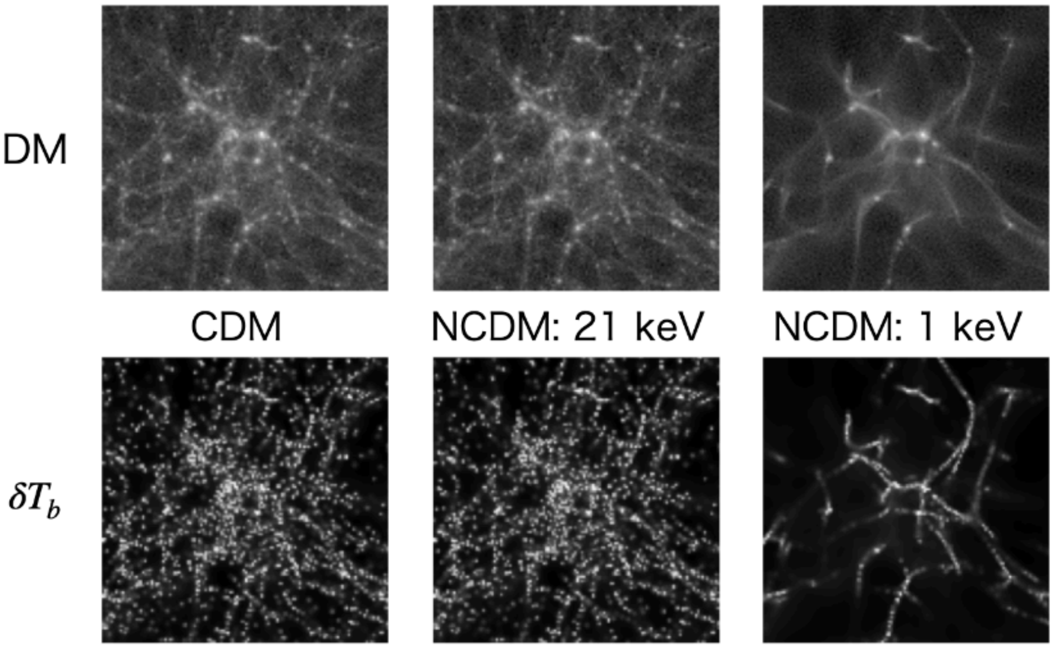

FIGURE 5.3: Example images of the CDM, 21 keV NCDM, and 1 keV NCDM models, arranged from left to right. The upper and lower panels show the images of dark matter and differential brightness temperature δT_b , respectively. For visibility, the δT_b images represent the logarithm of the actual images.

5.1.2 the Procedure of generating Images

In this subsection, we outline the process of generating images for training, validation, and testing of our CNN using data from the hydrodynamic simulations. The scale of matter power spectrum suppression resulting from the free streaming of dark matter is $k_\perp \sim 1\ h\mathrm{Mpc}^{-1}$ for the 1 keV NCDM and $k_\perp \sim 50\ h\mathrm{Mpc}^{-1}$ for the 46 keV (approximately $10^{4.66}$ eV) NCDM in the 2D Fourier space, where $k_\perp$ represents the wave number in the direction perpendicular to the line of sight.

Therefore, it is necessary for the size of the image to encompass the mode $k_\perp \sim 1\ h\mathrm{Mpc}^{-1}$ and for the pixel size to be sufficient to resolve the mode $k_\perp \sim 50\ h\mathrm{Mpc}^{-1}$. For the configurations of our simulation, the box size and the number of particles, which determine the spatial resolution of our simulations, meet these requirements.

We generate images following the steps below (summarized in Fig 5.2):

- [STEP 1]

  Firstly, a 1024^3 grid is defined in the simulation box. And then, in the case of dark matter particles, they are relocated to their nearest grid point. For HI gas, the HI number density n_{HI} in a grid, with its center located at $\mathbf{x}_i$, is computed by summing over all particles contributing to this grid:

$$n_{\mathrm{HI}}(x_i) = \sum_j W_{\mathrm{SPH}}(x_i - x_j | h_j) n_{\mathrm{HI},j} \tag{5.1}$$

  where $n_{\mathrm{HI},j}$ is the HI number density represented by the j-th particle located at $\mathbf{x}_j$, and W_{SPH} is the SPH kernel following the cubic spline kernel (Monaghan and Lattanzio, 1985). The SPH kernel is computed as

$$\begin{aligned} &W_{\mathrm{SPH}}(r,h) \\ &= A \begin{cases} 1 - \frac{2}{3}\left(\frac{r}{h/2}\right)^2 + \frac{3}{4}\left(\frac{r}{h/2}\right)^3 & (0 < r < \frac{h}{2}) \\ \frac{1}{4}\left(2 - \frac{r}{h/2}\right)^3 & (\frac{h}{2} < r < h) \\ 0 & (h < r)\,, \end{cases} \end{aligned} \tag{5.2}$$

  where h denotes the smoothing length for each particle, and r represents the distance between the particle and the center of the cell. The constant A is determined such that the sum of W_{SPH} over all grids equals unity for each particle. This normalization ensures the proper weighting of contributions from neighboring particles in the SPH calculations. The HI number density n_{HI} is converted to the differential brightness temperature δT_b following Eq. (2.97).

- [STEP 2]

  We cut out a part of the simulation box with a size of $100 \times 100\ (h^{-1}\mathrm{Mpc})^2$ (transverse) $\times 50\ h^{-1}$Mpc (line of sight) , and obtain three slices. These slices correspond to regions from 0 to 50 h^{-1}Mpc, 25 to 75 h^{-1}Mpc, and 50 to 100 h^{-1}Mpc along the line of sight, respectively. However, for one direction of the line of sight, which is used to generate test images, only two slices from 0 to 50 h^{-1}Mpc and 50 to 100 h^{-1}Mpc are used to prevent overlap in the regions when generating test images.

  To determine the optimal length of the slice along the line of sight, we investigate the length from 50 h^{-1}Mpc (limited by the sufficient number of images for CNN training) to 0.1 h^{-1}Mpc (the cell size in STEP 1). The performance, evaluated by the AUC of the classification between the CDM and 10 keV NCDM models, is found to be maximized when the projection depth is set to 50 h^{-1}Mpc.

Given that there are three degrees of freedom for the direction of the line of sight and these can be treated as independent realizations, a total of (2 line-of-sight directions) $\times$ (3 slices) $=$ 6 slices are used to generate training and validation images. For test images, (1 line-of-sight direction) $\times$ (2 slices) $=$ 2 slices are used. Specifically, images from five slices are used as the training data, and those from the remaining slice are used as validation data for the two line-of-sight directions. Test images are generated using the two slices from the other line-of-sight direction.

- [STEP 3]

  Within each slice in STEP 2, we perform integration along the line of sight for the mass density of dark matter $\rho_{\rm DM}(\boldsymbol{x})$, where $\boldsymbol{x}$ is the 3D coordinate of the center of cell. For dark matter, this integration yields a 2D map of the projected dark matter density $\rho_{\rm DM}(\boldsymbol{n})$, where $\boldsymbol{n}$ is the 2D coordinate of the center of the cell. And then, we compute the 2D mass density fluctuation of dark matter $\delta_{\rm DM}(\boldsymbol{n})$, defined as $\delta_{\rm DM}(\boldsymbol{n}) = (\rho_{\rm DM}(\boldsymbol{n}) - \bar{\rho}_{\rm DM})/\bar{\rho}_{\rm DM}$, where $\bar{\rho}_{\rm DM}$ is the average projected density of dark matter. For 21cm signals, δT_b is accumulated along the line of sight, described as $\delta T_b(\boldsymbol{n}) = \sum_{\rm los} \delta T_b(\boldsymbol{x})$.

- [STEP 4]

  From each slice, we crop 8$\times$8 subregions with each size of 12.5×12.5 (h^{-1}Mpc)^2. Each subregion corresponds to a single image with 128^2 pixels, where the pixel size corresponds to the scale of $\sim 0.1 h^{-1}$Mpc.

  To augment the training data, when dividing the slices into 8×8 regions to create training and validation data, we apply multiple offsets for the location of the edge for cropping. These offsets are defined as $\Delta = 12.5i/16$ h^{-1}Mpc, where $i = 0, 1, \cdots, 15$, in both the directions parallel and perpendicular to a side. When the edge for cropping is out of the slice, we apply the periodic boundary condition. While the shifted images may not be entirely independent, this approach significantly increases the number of available images and helps our training process.

Consequently, for the training, validation, and test images, we obtain (8×8 cut out in STEP 4) $\times$ (5, 1, and 3 slices in STEP 2) $\times$ (16^2, 16^2, and 1 offsets) $=$ 81920, 16384, and 128 images from one realization of the simulation, respectively. Additionally, in the training process of CNN, we apply the horizontal flip and rotation every 90 degrees randomly to the images. Finally, the effective number of the training images is 81920×2 (flip) $\times 4$ (rotation) $= 655360$. The images for validation and test of our CNN are not flipped and rotated. The validation images are only used to monitor the progress of the training and do not contribute to the updating of the weight parameters.

The test images are not completely independent of the training images because these images are generated from the same simulation box. This may affect the results of CNN. To examine that the test of CNN by using the test images from the same realization as the training images is valid, we prepare another simulation of the CDM and 10 keV NCDM model with the different random seeds of the initial conditions. And then, we generate 128 test images from each of these new realizations following the above procedure. Finally, we use these new test images to test our CNN trained by the original dataset. As a result, the AUC for the new test images is 0.80 and consistent with AUC$= 0.78$ for the original dataset.

The images of the dark matter density fluctuation δ and the differential brightness temperature δT_b have large dynamic ranges caused by the evolution of the large-scale structure of the universe in the non-linear regime. It is not easy for our CNN to extract information from images with large dynamic ranges. Therefore, we transform the pixel values in the images as

$$m_\delta(x) = \sinh^{-1}\left[\frac{\delta(x)}{b}\right], \tag{5.3}$$

$$m_T(x) = \sinh^{-1}\left[\frac{\delta T_b(x)}{b}\right], \tag{5.4}$$

where b is a softening parameter that controls the smooth transition scale from linear at $x \ll 1$ to logarithm at $x \gg 1$. We apply $b = 1$ for the dark matter images and $b = 1$ [nK] for the δT_b images. Figure 5.3 shows examples of the images of the dark matter and δT_b.

This transformation is originally designed for the magnitude system, known as *Luptitude*, which is introduced by the Sloan Digital Sky Survey (Lupton, Gunn, and Szalay, 1999). It is useful for reducing the dynamic range and can deal with negative pixel values, unlike a simple logarithmic scale. We see the results of our CNN for $b = 1$ [nK] and $b = 1$ [mK] for the δT_b images of the CDM and 10 keV NCDM, and the performance of our CNN for nK becomes better; the AUC is 0.95 for nK while it is 0.78 for mK.

5.2 Methods

5.2.1 Power Spectrum

For comparison with the CNN approach, we use the power spectrum analysis. Note that here we consider the classification problem between CDM and NCDM with mass $m_{\rm DM}$, unlike the usual parameter inferences. To do this, we calculate the 2D power spectrum of the test images generated in Section 5.1.2. In this work, we define the 2D Fourier counterpart of the quantity $A(\boldsymbol{n})$ as

$$\tilde{A}(\boldsymbol{k}_\perp) = \int \exp\{(-i\boldsymbol{k}_\perp \cdot \boldsymbol{n})\} A(\boldsymbol{n}) d^2 n. \tag{5.5}$$

In this work, we calculate the discrete 2D Fourier transform of images, and the discrete wave number of the i-th bin is represented as $\boldsymbol{k}_{\perp,i}$.

For the dark matter test images, we calculate the 2D power spectrum as

$$P^{\rm DM}(k_{\perp,i})\ [(h^{-1}{\rm Mpc})^2] = \frac{1}{L^2}\frac{1}{N_{k_{\perp,i}}}\sum_j \tilde{\delta}_{\rm DM}(\boldsymbol{k}_{\perp,j})\tilde{\delta}^*_{\rm DM}(\boldsymbol{k}_{\perp,j}), \tag{5.6}$$

and for the δT_b test images, the 2D power spectrum is written as

$$P^{\delta T_b}(k_{\perp,i})\ [{\rm mK}^2(h^{-1}{\rm Mpc})^2] = \frac{1}{L^2}\frac{1}{N_{k_{\perp,i}}}\sum_j \delta\tilde{T}_b(\boldsymbol{k}_{\perp,j})\delta\tilde{T}_b{}^*(\boldsymbol{k}_{\perp,j}), \tag{5.7}$$

where $\tilde{\delta}_{\rm DM}$ and $\delta\tilde{T}_b$ are the 2D Fourier counterparts of $\delta_{\rm DM}$ and δT_b respectively, $k_{\perp,i}$ is $= |\boldsymbol{k}_{\perp,i}|$, $\boldsymbol{k}_{\perp,j}$ and $N_{k_{\perp,i}}$ are the vectors of the wave number and the number of

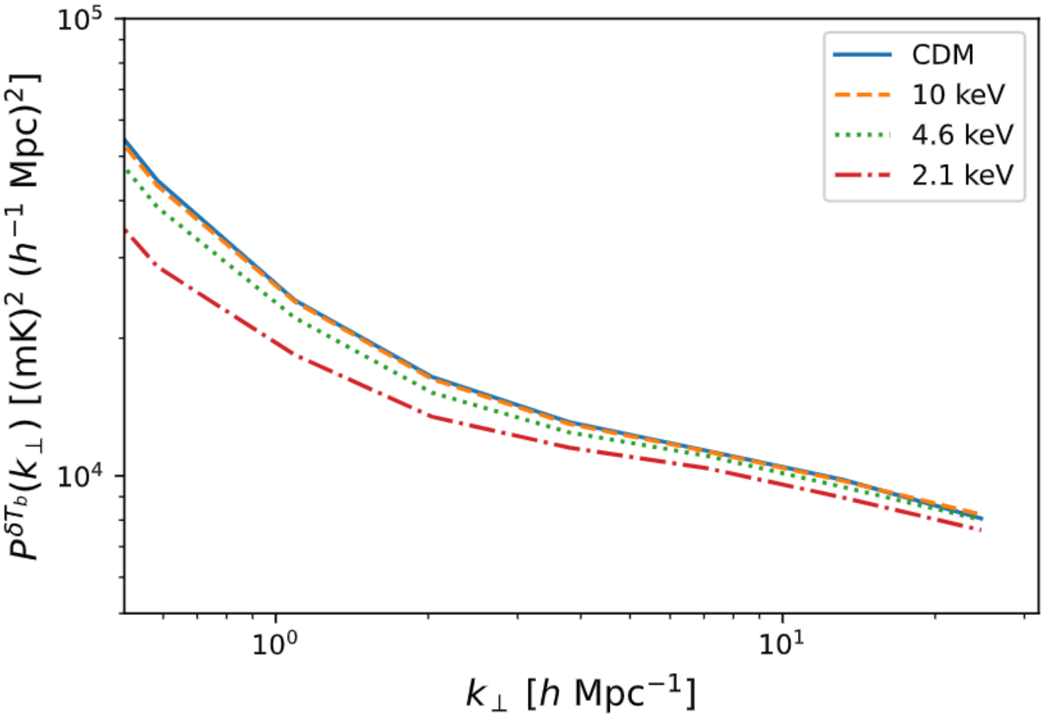

FIGURE 5.4: The 2D δT_b power spectra measured from the entire box of the simulation for CDM (blue-solid), 10 keV (orange-dashed), 4.6 keV (green-dotted), and 2.1 keV (red-dash-dotted) models, respectively.

the modes satisfying $k_{\perp,i} \leq |\mathbf{k}_{\perp,j}| < k_{\perp,i+1}$, respectively, and L is $=$12.5 h^{-1}Mpc, corresponding to the images size. The factor $1/L^2$ is required for the 2D Fourier transform for the finite interval and L becomes $= 2\pi$ if the image size is infinite. The minimum and maximum of $k_{\perp,i}$ correspond to the inverse multiplied by 2π of the size and resolution of the image, multiplied by 2π, respectively. We evaluate the classification results for the number of $k_\perp$-bins ranging from 1 to 20 and find that 4 bins are optimal.

Figure 5.4 shows the 2D power spectra of δT_b. Unlike the dark matter power spectrum, we see the overall suppression of the amplitude for NCDM models. This can be explained as follows. The NCDM model affects the HI ionization through the star formation efficiency due to the different amplitude of clustering of dark matter. It drastically changes the abundance of HI with respect to the dark matter halo. This modifies the HI bias, and affects the overall amplitude of the power spectrum.

The covariance matrix of the power spectra for the test images from the CDM simulation can be measured as

$$C_{mn} = \frac{1}{N_{\text{img}}} \sum_l \left(P_l(k_{\perp,m}) - \bar{P}(k_{\perp,m})\right)\left(P_l(k_{\perp,n}) - \bar{P}(k_{\perp,n})\right), \tag{5.8}$$

where P_l is the 2D power spectrum for the l-th CDM test image, of which the number is 128 and is denoted as N_{img} and $\bar{P}(k_\perp)$ is the averaged power spectra over the power spectra for the CDM test images.

	Layer	Output
1	Input	$128 \times 128 \times 1$
2	3×3 convolution	$126 \times 126 \times 32$
3	3×3 convolution	$124 \times 124 \times 32$
4	3×3 convolution	$122 \times 122 \times 64$
5	3×3 convolution	$120 \times 120 \times 64$
6	3×3 convolution	$118 \times 118 \times 128$
7	1×1 convolution	$118 \times 118 \times 64$
8	3×3 convolution	$116 \times 116 \times 128$
9	2×2 AvgPool	$58 \times 58 \times 128$
10	3×3 convolution	$56 \times 56 \times 256$
11	1×1 convolution	$56 \times 56 \times 128$
12	3×3 convolution	$54 \times 54 \times 256$
13	2×2 AvgPool	$27 \times 27 \times 256$
14	3×3 convolution	$25 \times 25 \times 512$
15	1×1 convolution	$25 \times 25 \times 256$
16	3×3 convolution	$23 \times 23 \times 512$
17	2×2 AvgPool	$12 \times 12 \times 512$
18	3×3 convolution	$10 \times 10 \times 512$
19	1×1 convolution	$10 \times 10 \times 256$
20	3×3 convolution	$8 \times 8 \times 512$
21	1×1 convolution	$8 \times 8 \times 256$
22	3×3 convolution	$6 \times 6 \times 512$
23	GlobalAvgPool	$1 \times 1 \times 512$
24	FullyConnected	2

TABLE 5.1: The architecture of our CNN. In the second column, we show the layer type, and in the third column, we show the output size from the layer (height × width × channel), where the channel denotes the number of the feature maps. This CNN has the 8328610 weight parameters.

5.2.2 CNN architecture

Here, we describe the architecture of our CNN. Table. 5.1 shows the architecture, which follows the previous work (Ribli et al., 2019) but we remove the first two average pooling layers in (Ribli et al., 2019) to adjust to our input image with 128^2 pixels. Here, in the second column, "convolution", "AvgPool", "GlobalAvgPool", and "FullyConnected" represent the convolution, average-pooling, global average pooling, and fully connected layer. To construct this CNN, we use PyTorch, which is the Python module for the implementation of neural networks. Our CNN has the $\sim 8 \times 10^6$ weight parameters. Han et al., 2015 suggests that the number of the training data should be more than one-tenth of the number of its weight parameters to avoid both over- and under-fitting to the training data. We have sufficient training images, of which the number is $\sim 10^5$ effectively for a realization.

Here, we explore the optimal number of layers. In the case that we reduce the number of layers by removing the 4th, 5th, 6th, 12th, 16th, and 22nd layers in Table 5.1, the cross entropy loss Eq. (3.11) for this smaller architecture becomes 10 times larger compared to the one for the original architecture, and our CNN classifies the images at random and cannot distinguish the dark matter models. These results mean that this smaller architecture is too simple to classify the dark matter models. On the contrary, we consider the case that we add convolution layers after each original convolution layer with zero-padding, where the size and number of the kernel are the same as the one in the previous layer. Zero-padding means the adding of extra pixels with the value of zero, surrounding the input image, and is used to keep the size of the input image, In this case, we cannot observe the reduction of the loss value in training. This is because this architecture is too complex and its parameters cannot be optimized properly. It is probable that too large number of the weight parameters cause the vanishing gradients (He et al., 2016). This larger architecture also classifies images randomly.

In training our CNN, the initial value of the weight parameters is given randomly, and they are updated through the training. For efficient training, we apply a batch normalization introduced in Section 3.2.4 after each convolution layer in Table 5.1. As an activation function in each convolution layer, we use ReLU described as Eq. 3.5.

Here, we can describe the outputs as

$$p_i(\mathrm{M}) = \{p_i(\mathrm{CDM}|\mathrm{M}), p_i(\mathrm{NCDM}|\mathrm{M})\}, \tag{5.9}$$

where M represents the true dark matter model for the i-th input image, and $p_i(k|\mathrm{M})$ is the predicted probability that the i-th input image from model M is from k model. Our CNN is trained to enlarge the element of $P_i(\mathrm{M}|\mathrm{M})$.

We obtain the output value written as Eq. (5.9) by converting of the raw output from the last FullyConnected layer, described as $y(\mathrm{M}) = \{y_i(k|\mathrm{M}), y_i(k|\mathrm{M})\}$, using the softmax function:

$$p_i(k|\mathrm{M}) = \frac{\exp\{(y_i(k|\mathrm{M}))\}}{\exp\{(y_i(\mathrm{CDM}|\mathrm{M}))\} + \exp\{(y_i(\mathrm{NCDM}|\mathrm{M}))\}}. \tag{5.10}$$

For the loss function, we apply a cross-entropy loss using the above expression of the outputs:

$$E_i(w) = -\sum_k \tilde{p}_i(k|\mathrm{M}) \ln\left(p_i(k|\mathrm{M})\right). \tag{5.11}$$

Here, $\tilde{p}_i$ corresponds to the ground truth, and takes 1 for the correct class ($k = \mathrm{M}$)

and 0 for the other case ($k \neq \mathrm{M}$). p_i corresponds to our CNN prediction takes from 0 to 1 continuously.

To optimize the weight parameters using the loss function E, we apply the AMSGRAD(Reddi, Kale, and Kumar, 2019), described in Section 3.2.2. In our work, we choose $\eta = 10^{-5}$ as the learning rate. We apply the mini-batch learning and set 8 as the batch size in training our CNN. The validation images are used to monitor the training progress, and we define the completion of the training is that the loss for the validation images averaged over the latest 5 epochs converges to 1%.

5.2.3 Evaluation of Classification

In this study, our focus is to distinguish between the images of the CDM model and NCDM models. We employ the Kolmogorov-Smirnov test and Area Under the Curve (AUC) as metrics to assess the effectiveness of our classification. The former metric is used to evaluate the classification performance of our CNN and the power spectrum, while the latter provides a quantitative measure of the CNN classification results.

Kolmogolov-Smirnov Test

The Kolmogorov-Smirnov (KS) test is a method to determine if two different finite samples have the same probability distribution (Kolmogorov, 1933; Smirnov, 1939). We apply the KS test to classify the images of the CDM and NCDM models using either our Convolutional Neural Network (CNN) or the power spectra.

For both the power spectrum and our CNN outputs, we evaluate the classification performance using the distributions of χ^2 values defined in the following. For the i-th test image from the dark matter model M, the χ^2 value of the power spectrum is computed as follows:

$$\chi^2_{\mathrm{PS},i}(\mathrm{M}) = \mathbf{\Delta} P_i(k_\perp|\mathrm{M})\mathbf{C}^{-1}\mathbf{\Delta} P_i(k_\perp|\mathrm{M}). \tag{5.12}$$

Here, $\mathbf{\Delta} P_i(k_\perp|\mathrm{M})$ is $= P_i(k_\perp|\mathrm{M}) - \bar{P}(k_\perp|\mathrm{CDM})$, where $P_i(k_\perp|\mathrm{M})$ is the power spectrum of the i-th input image from the dark matter model M calculated by Eq. (5.6) or Eq. (5.7), and $\bar{P}(k_\perp)$ is the averaged power spectrum over the CDM images. C^{-1} is the inverse covariance matrix for $P(k_\perp|\mathrm{CDM})$ calculated by Eq. (5.8). Concerning our CNN classification, the χ^2 value is determined by evaluating the difference between the outputs for two dark matter models:

$$\chi^2_{\mathrm{CNN},i}(\mathrm{M}) = \frac{(y_i(\mathrm{M}) - \bar{y}(\mathrm{CDM}))^2}{\frac{1}{N}\sum_j \left(y_j(\mathrm{CDM}) - \bar{y}(\mathrm{CDM})\right)^2}. \tag{5.13}$$

where $y_i(M)$ represents the output from CNN for the i-th input image of dark matter model M. $\bar{y}(\mathrm{CDM})$ is the mean of our CNN outputs over the CDM test images and the denominator of the right-hand side is the variance of $y_i(\mathrm{CDM})$.

Following that, we apply the Kolmogorov-Smirnov (KS) test to compare the distributions of $\chi^2_i(\mathrm{CDM})$ and $\chi^2_i(\mathrm{NCDM})$ using the `stats.ks_2samp` method available in the `SciPy` library (Virtanen et al., 2020), a Python module.

Our null hypothesis is that the distribution of $\chi^2_i(\mathrm{NCDM})$ follows the same distribution as the one of $\chi^2_i(\mathrm{CDM})$. In this analysis, we set the significance level to a p-value of 0.01 ($\sim 2.6\sigma$).

Note that the KS test only informs us about the presence or absence of a significant difference between the images of the two dark matter models. It does not provide a quantification of whether the output is correct or not. We evaluate the correctness of the classification by the AUC introduced in the next section.

AUC

Firstly, we introduce the Receiver Operating Characteristic (ROC) curve. Our CNN provides the probability that the input image is of the NCDM model denoted. We denote the predicted probability that the i-th input image is the NCDM model as p_i. In binary classification, a specific threshold t is required. The CNN recognizes the input image as the NCDM model if $p_i > t$. Therefore, we can consider the four different cases

1. True Positive (TP): $p_i(\mathrm{NCDM}|\mathrm{NCDM}) \geq t$,
2. True Negative (TN): $p_i(\mathrm{NCDM}|\mathrm{CDM}) < t$,
3. False Positive (FP): $p_i(\mathrm{NCDM}|\mathrm{CDM}) \geq t$,
4. False Negative (FN): $p_i(\mathrm{NCDM}|\mathrm{NCDM}) < t$.

Each quantity is a function of t.

And then, we can define the following quantities:

$$\mathrm{ROC} : x(t) = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}, \quad y(t) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{5.14}$$

where x represents the ratio of the number of misclassified images as the NCDM model to the one of the CDM test images, and y represents the ratio of the number of correctly classified images as the NCDM model to the one of the NCDM test images. The trajectory of the point (x, y) with varying threshold t continuously from 0 to 1 is referred to as the ROC curve.

Subsequently, we can consider the area under the ROC curve (AUC). The AUC approaches unity when the classification is efficient and complete, while it approaches 0.5 when the classification is insufficient and incomplete.

5.3 Results

In this section, we show the results of the binary classification between the images of the CDM and NCDM models. The classifications are conducted by the image analysis with CNN and the power spectrum analysis, and the results of the classification by these analyses are compared. Section 5.3.1 shows the results of the classifications by the CNN and the power spectrum for both the image of the dark matter and the δT_b. We compare the results. In Section 5.3.2 and 5.3.3, we discuss the effects of the choice of the astrophysical models and those of the system noise existing in the observations.

For convenience, we first define the acronyms X-Y, where X represents the method (either CNN or PS for power spectrum) and Y represents the observable (either DM for dark matter or δT_b). For example, CNN-δT_b corresponds to the classification by the CNN using the δT_b map.

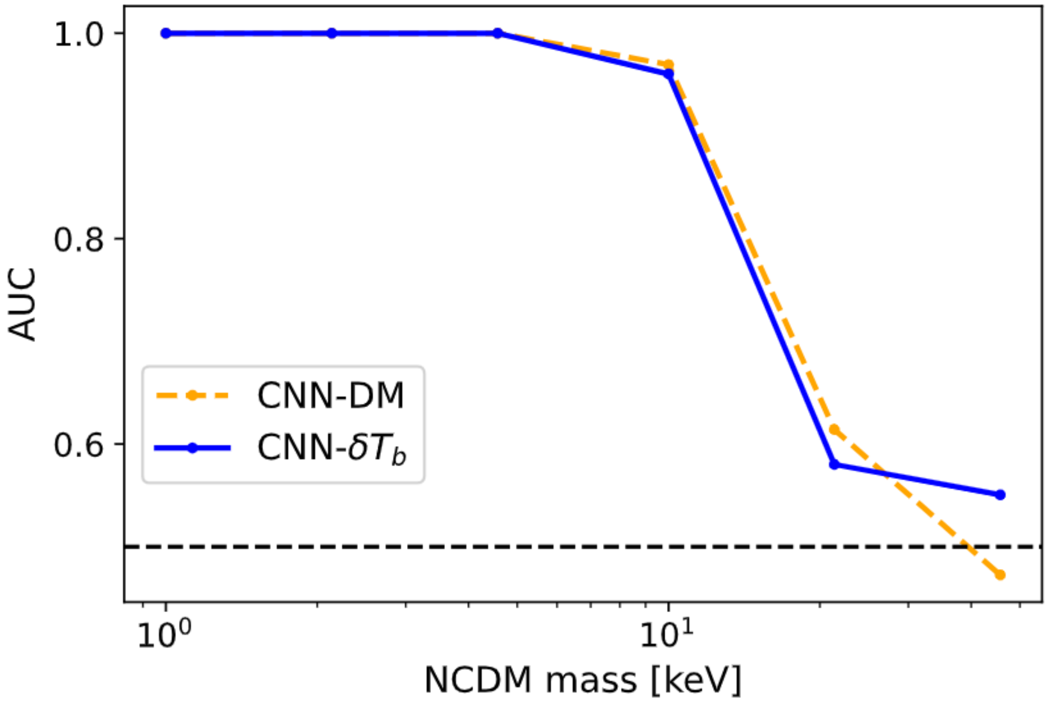

FIGURE 5.5: The AUC for the classification between CDM and NCDM, of which the mass is represented in the horizontal axis. The horizontal dashed line corresponds to the case of the random classification. The blue solid and orange dashed line represent the results of CNN-DM and CNN-δT_b, respectively, and both methods show comparable results.

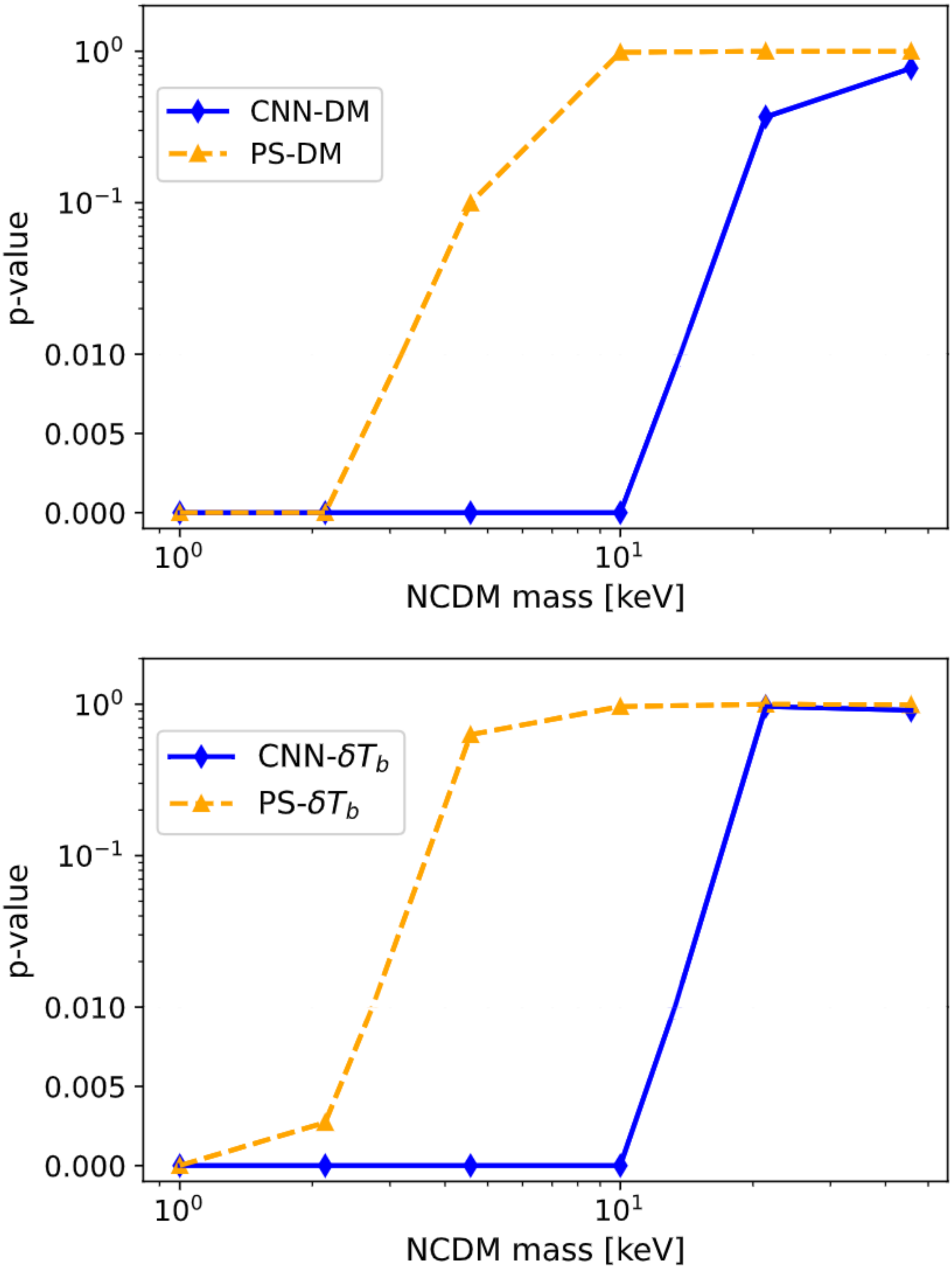

FIGURE 5.6: p-values of our KS test represented as a function of the dark matter mass m_{DM}. We compare CNN (blue solid) with the power spectrum (orange). The upper and lower panel show the results of the classification of the dark matter and δT_b images, respectively. CNN shows a better performance than the power spectrum for both cases of the dark matter and δT_b images.

5.3.1 the Classifications of dark matter and δT_b images

In this subsection, we show the classification results of the dark matter model and compare the results for images of dark matter and δT_b using both CNN and power spectrum analyses. These classifications are evaluated using the AUC and the KS test.

First, we compare the results of CNN-δT_b with those of CNN-DM. Figure 5.5 shows the AUC values of the classification results. As shown in this figure, for both of the cases of dark matter and δT_b images, AUC values are greater than 0.95 for $m_{\rm DM} \leq 10$ keV. AUC values of CNN-δT_b are comparable to those of CNN-DM for across the entire mass range of NCDM we consider. Thus, we can conclude that δT_b is a valid tracer of the dark matter distribution for our CNN.

Next, we compare the p-value of the KS test for CNN-DM and PS-DM, as well as CNN-δT_b and PS-δT_b. Figure 5.6 shows the results of CNN-DM and PS-DM (upper panel) and of CNN-δT_b and PS-δT_b (lower panel). The results of CNN and the power spectrum are described by the blue solid and orange dashed line, respectively. Our CNN outperforms the power spectrum for both of the cases of the dark matter and δT_b images. The p-value of CNN-DM and CNN-δT_b is less than 0.001 at $m_{\rm DM} = 4.6$ keV, and these methods can reject the null-hypothesis with high significance. On the other hand, the p-values of PS-DM and PS-δT_b are larger than 0.1.

Finally, we compare CNN-DM with CNN-δT_b. These methods provide a similar performance for the KS test. The p-values of the classifications by both of them are less than 0.001 for the $m_{\rm DM} \leq 10$ keV NCDM and these methods can distinguish the images of CDM and NCDM with high significance. Their performances diminish for more massive NCDM, for example, the p-values of CNN-DM and CNN-δT_b are $= 0.37$ and > 0.99 for $m_{\rm DM} = 21$ keV, respectively.

5.3.2 Effects of the Astrophysical Models

Here, we examine the effects of the different astrophysical models from *Fiducial* in our classifications. To assess their effects on the classifications, we replace the CDM test images of *Fiducial* model with those of the different astrophysical models such as *Shield*, *NoSF*, and *FG09*. This implies that our CNN is trained using images from the incorrect astrophysical model. In the following, we only consider PS-δT_b and CNN-δT_b.

Figure 5.7 shows the results of the KS test for PS-δT_b (upper panel) and CNN-δT_b (lower panel). The p-values for PS-δT_b in the upper panel of Figure 5.7 are less than 0.001 at $m_{\rm DM} \leq 4.6$ keV, regardless of the astrophysical model. They exceed 0.1 at more massive NCDM, where the power spectrum cannot classify the dark matter models. Hence, the differences in the astrophysical models do not have significant impacts on our analysis using the power spectrum.

Next, the p-values for CNN-δT_b are also shown in Figure 5.7 (lower panel). Our CNN can distinguish between CDM and NCDM with $m_{\rm DM} \leq 4.6$ keV independent of the astrophysical model, except for *FG09*. The p-values for *NoSF* (orange dashed) *Shield* (green dotted) models are comparable to those of *Fiducial* (blue solid), while *FG09* (red plots) exhibit different behavior. Table 5.2 shows AUC values for the different astrophysical model for $m_{\rm DM} = 2.1, 4.6, 10, 21$ keV. As we can see in this table, our CNN can distinguish between images of CDM and NCDM for $m_{\rm DM} \leq 10$ keV for *Fiducial*, *Shield*, and *NoSF*. However, our CNN cannot correctly classify the images of *FG09* CDM model and NCDM models with $m_{\rm DM} \geq 4.6$ keV, despite the p-values in Figure 5.7 being close to zero for these classifications. Therefore, the difference

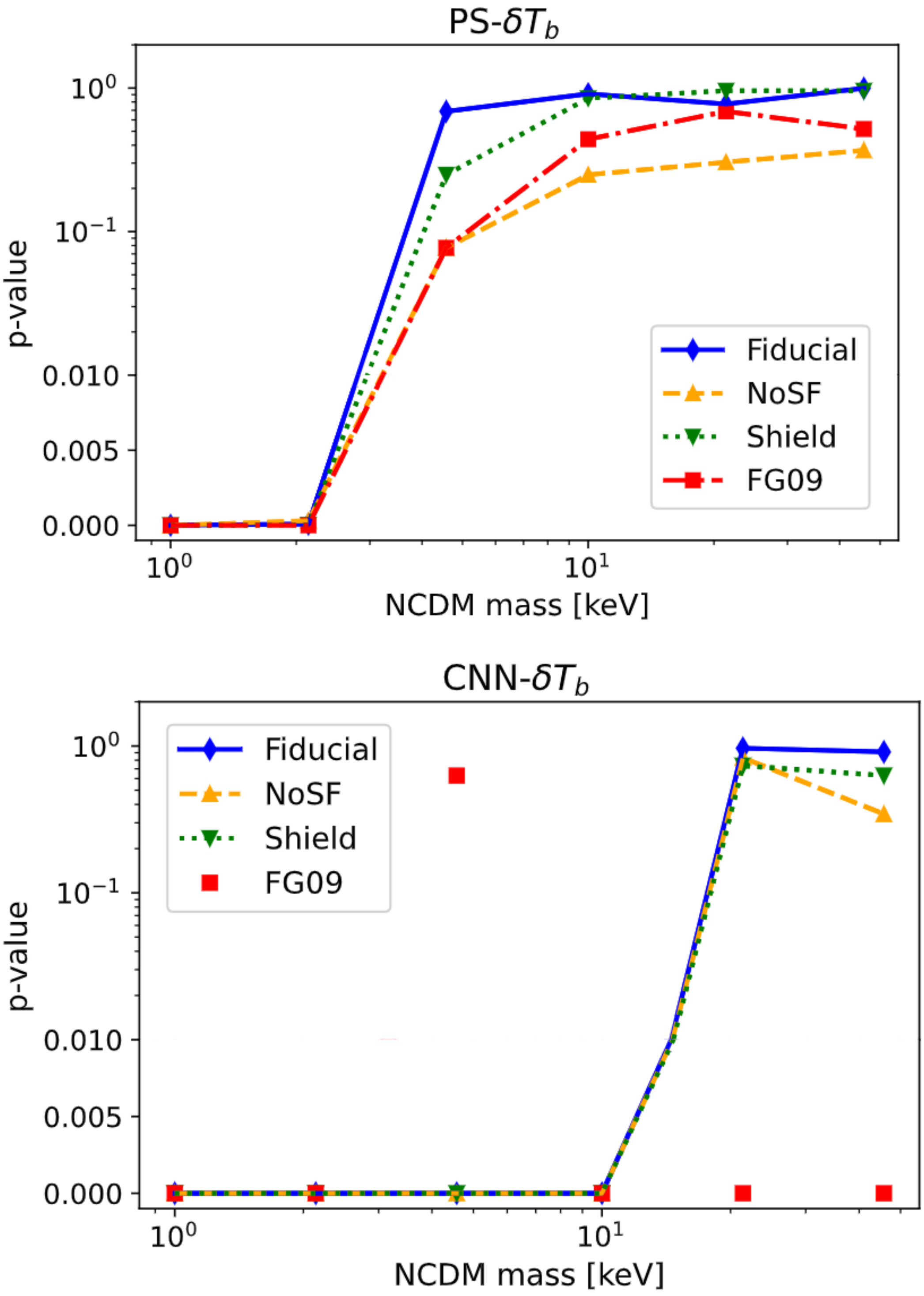

FIGURE 5.7: The p-values of the KS test. In the vertical axis, we show the p-value for PS-δT_b (upper panel) or CNN-δT_b (lower panel). Here, the *Fiducial* (blue solid), *NoSF* (orange dashed), *Shield* (green dotted), and *FG09* (red square plot) models are assumed. For CNN-δT_b, p-value for *FG09* model is significantly different from that of *Fiducial* model. CNN-δT_b provide small p-values for *FG09* model, but CNN cannot classify the images correctly, as we can see in Figure 5.8 and Table 5.2.

Model	2.1 keV	4.6 keV	10 keV	21 keV
Fiducial	1.00	1.00	0.96	0.58
Shield	1.00	1.00	0.96	0.58
NoSF	1.00	1.00	0.95	0.51
FG09	1.00	0.5	0.28	0.02

TABLE 5.2: The AUC values for different astrophysical models and dark matter mass models. The *Shield* and *NoSF* models do not have significant effects on the classification, but the *FG09* model drastically spoils the classification.

in the HI gas density map between CDM and NCDM might be partly mimicked by the effect of the different UV background models. However, for $m_{\rm DM} < 2.1$ keV, we do not observe the effects of the astrophysical models in the classifications. Thus, we can conclude that our CNN classification for this mass range is robust against differences in the astrophysical model, at least within the astrophysical models we consider in this work.

For further discussion, we evaluate the classification results using the confusion matrix, defined as

$$\begin{pmatrix} \frac{\rm TP}{\rm TP+FN} & \frac{\rm FN}{\rm TP+FN} \\ \\ \frac{\rm FP}{\rm TN+FP} & \frac{\rm TN}{\rm TN+FP} \end{pmatrix}, \tag{5.15}$$

where TP, FN, TN, and FP are calculated at threshold $t = 0.5$. The diagonal components represent the results for the correct classification, where the upper left and lower right components correspond to the fraction of correctly classified test images of NCDM and CDM, respectively. On the other hand, the non-diagonals represent the results for the misclassified images, where the upper right and lower left components correspond to the fraction of misclassified test images of NCDM and CDM, respectively.

Figure 5.8 shows the confusion matrix of the results of CNN-δT_b. The astrophysical models, including *Fiducial*, *Shield*, *NoSF*, and *FG09*, are considered from top to bottom, and three NCDM models with $m_{\rm DM} = 4.6$, 10, and 21 keV are arranged from left to right. There are little differences among *Fiducial*, *Shield*, *NoSF* in the confusion matrix, similar to AUC values shown in Table 5.2. However, all CDM test images from the *FG09* model are consistently misclassified as NCDM, regardless of the NCDM mass, and AUC values also decrease drastically. These results imply that the number of correctly classified images of CDM decreases when we assume the incorrect astrophysical model. The underlying reason might be the presence of more HI gas in the *FG09* model compared to the UV background model used in *Fiducial* model (Haardt and Madau, 2012).

We also investigate the effects of the astrophysical model on the δT_b power spectra. Fig 5.9 shows the fractional difference of $P^{\delta T_b}(k_\perp)$ for different astrophysical models and NCDM modes with respect to the $P^{\delta T_b}(k_\perp)$ for *Fiducial* CDM model. The errors are depicted by the dark-gray shaded region, where the error corresponds to the 1-σ error from cosmic variance $1/\sqrt{N_{k_\perp}}$. The power spectra of NCDM models are suppressed compared to the *Fiducial* CDM model. On the other hand, the power spectra of the *NoSF* and *FG09* models are amplified, while there is little difference between the *Shield* and *Fiducial* models. However, *NoSF* model does not have a large effect on the classifications of CNN-δT_b, and the images from *FG09* model are

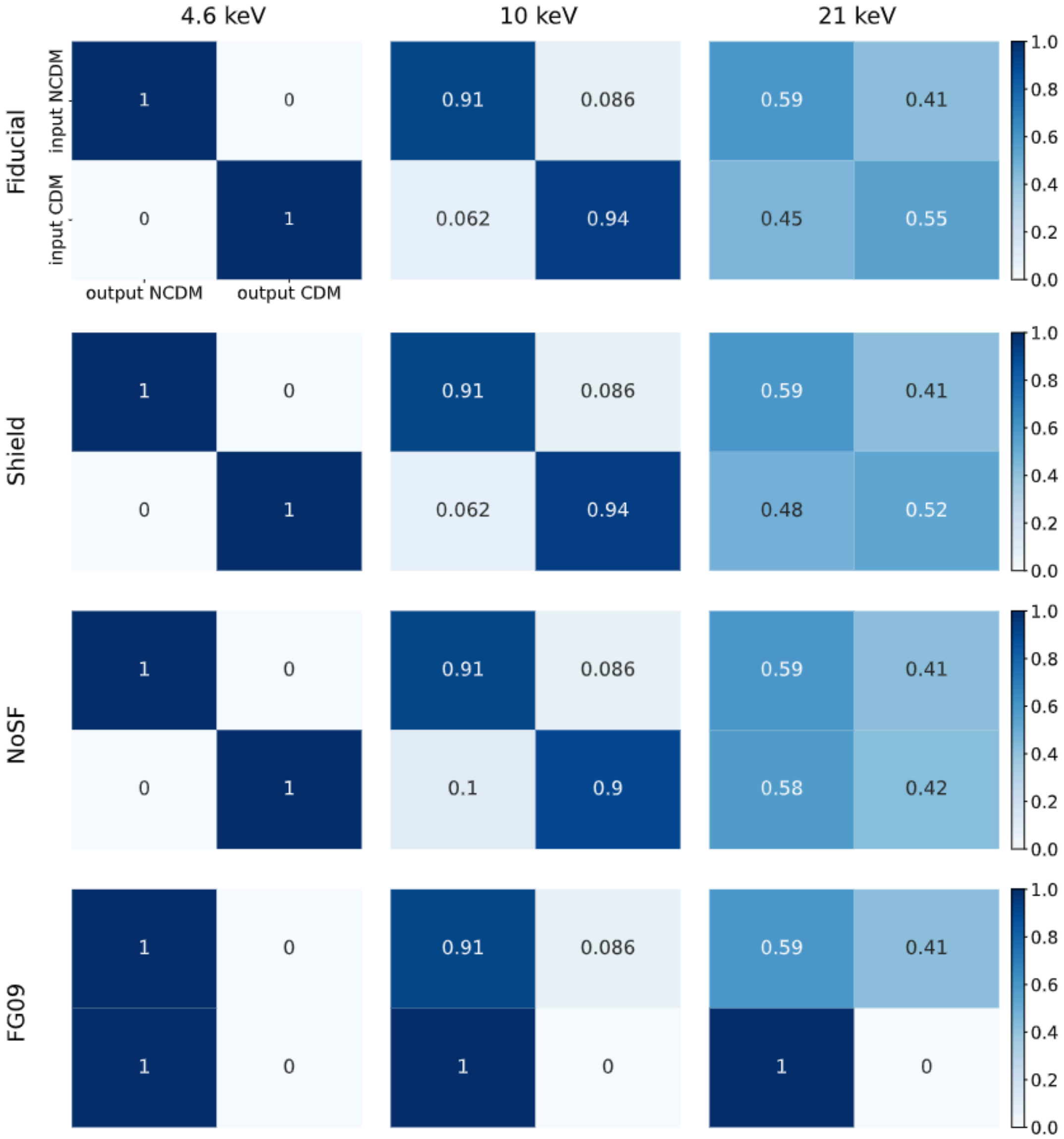

FIGURE 5.8: The confusion matrices for CNN-δT_b.
Column: the NCDM masses, 4.6, 10, and 21 keV from left to right
Row: the astrophysical models, *Fiducial*, *Shield*, *NoSF*, and *FG09* from top to bottom.
The confusion matrices of *Shield* and *NoSF* are almost identical to that of the *Fiducial* model. On the other hand, for the *FG09* model, the CDM test images are classified into NCDM incorrectly, regardless of the mass of NCDM.

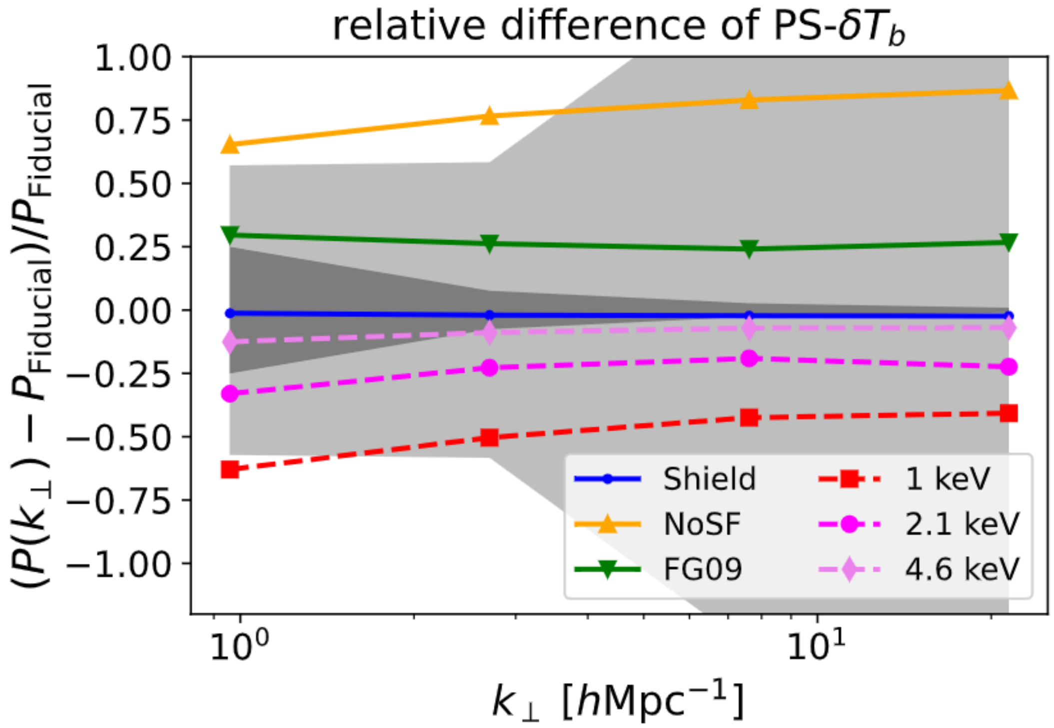

FIGURE 5.9: The ratio of the δT_b power spectrum for the *Fiducial* CDM model to *Shield* (blue solid), *NoSF* (orange solid), *FG09* (green solid), 4.6 keV (violet dashed) or 2.1 keV (red dashed) model. The shaded region shows $1-\sigma$ error of the cosmic variance (dark gray) and 1-σ error of the cosmic variance + the system noise introduced in Section 5.3.3 with $t_0 = 1,000$ hours (light gray).

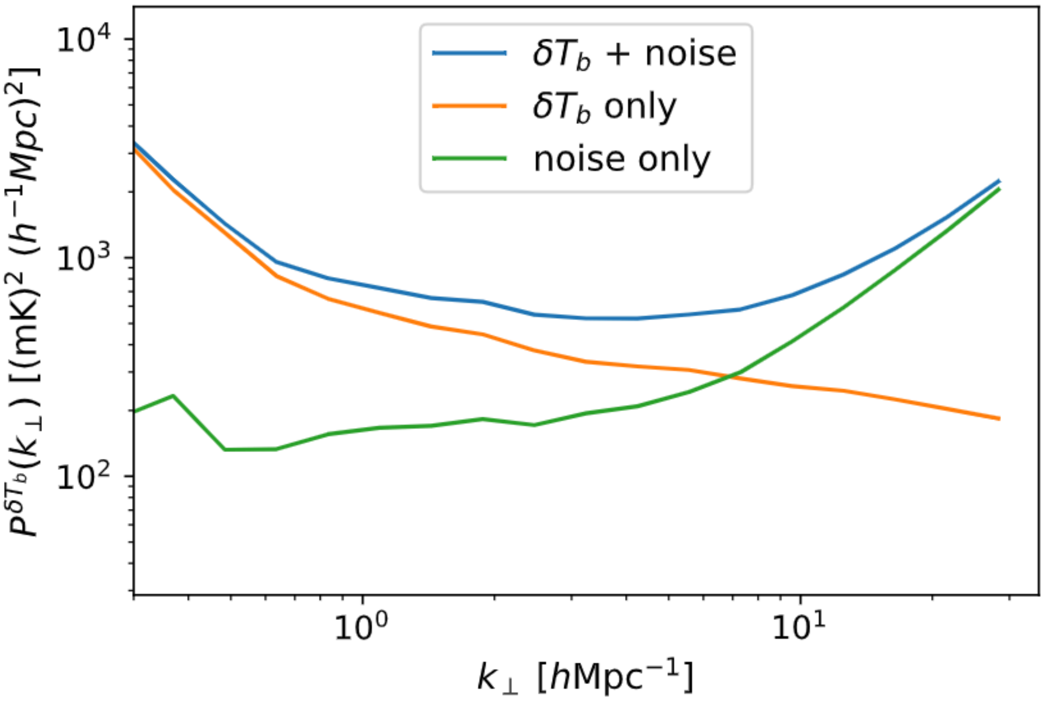

FIGURE 5.10: The 2D power spectrum of $\delta T_b +$ noise (blue), δT_b only (orange), and noise only (green) for the *Fiducial* CDM simulation. For the noise power spectrum here, the integration time t_0 is 1,000 hours.

misclassified, as shown in Figure 5.7 and 5.8. Therefore, the effects of the different astrophysical models on the classifications of CNN-δT_b cannot be explained solely by the effects on the power spectrum. It is probable that the reason for the misclassifications in the *FG09* model is that the classification with our CNN depends on whether the image is of the CDM model or not, rather than the difference between the two dark matter models.

In Appendix B, we explore the properties of the HI halo and the effects of the astrophysical models on the halo properties.

5.3.3 Effect of System Noise

Here, we consider the effect of the system noise for the SKA-MID that can observe 21cm signals at $0 < z < 3$. Indeed, we should consider the effects of foreground contamination, which is the contamination at radio frequencies by Galactic synchrotron emission, Galactic free-free emission, and extragalactic sources. The signals from these contaminants are much brighter than the signals from the HI gas. To remove these contaminations, various methods of foreground removal, such as the principal component analysis (PCA) (Spinelli et al., 2022), generalized morphological component analysis (GMCA) (Carucci, Irfan, and Bobin, 2020)m and Gaussian process regression (Soares et al., 2022; Chen et al., 2023), have been suggested. In this work, we assume the most optimistic case, where the foreground contaminations are completely removed from our data. This is because our aim in this work is to demonstrate the potential of CNN for the image analysis of 21cm intensity mapping using the simulation data. In fact, foreground removal is challenging, but the investigation of the effects of foreground contamination will be part of our future work.

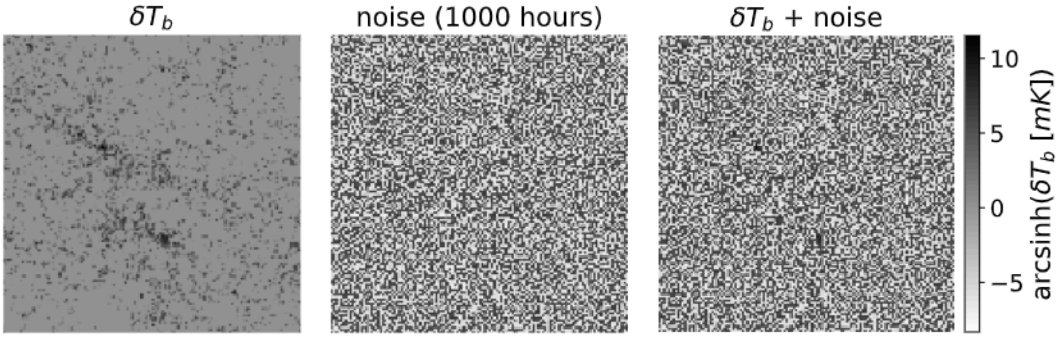

FIGURE 5.11: Each panel show the image of 12.5×12.5 $(\mathrm{Mpc}/h)^2$ region of δT_b (left), only noise with $t_0 = 1{,}000$ hours (middle), and the map of the sum of δT_b and noise (right). These images show the same region in a *Fiducial* simulation box. The pixel value is $\mathrm{arcsinh}(\delta T_b\ [\mathrm{mK}])$. Note that the darker pixel represents the higher δT_b for visibility.

To generate the images including the system noise, we first generate the 3D noise map with a size of 100 h^{-1}Mpc on each side and 1024^3 grids, consistent with the configuration used in STEP 1 in Section 5.1.2. We assume the noise follows a Gaussian distribution with a mean of 0 and a variance of P_{noise} as a simplistic assumption, where P_{noise} is the noise power spectrum introduced in the following part. Next, we combine this noise map with the simulation data after completing STEP 1 in Section 5.1.2. Finally, we generate the images following the same procedure as described in STEP 2-4 in Section 5.1.2. However, we apply the transformation of Eq. (5.4) with a different b from previous sections. For the δT_b images including system noise, we apply the transformation with $b = 1$ mK instead of $b = 1$ nK, i.e.,

$$m_T^{\mathrm{obs}} = \sinh^{-1}\left[\frac{\delta T_b(x) + \mathrm{noise}}{(1\mathrm{mK})}\right]. \tag{5.16}$$

The transformation of $\sinh^{-1}(x/b)$ is a linear transformation at scales of $x/b \ll 1$ while it is a logarithmic one at scales of $x/b \gg 1$, where the softening parameter b controls the transition scale. We find that the structure of the HI distribution in low-density regions contains significant information for our classifications and $b = 1$ nK works well in the images without the noise. However, low-density regions are highly contaminated by the system noise, with an amplitude typically on the order of 1 mK in our assumption. In the case of the images including noise, we focus on high-density regions of HI, where δT_b is > 1 mK, by setting $b = 1$ mK. In practice, for the images including noise, the AUC value of the classification between the CDM and 1 keV NCDM for $b = 1$ mK is 0.78 while it for $b = 1$ nK is 0.67. b is the hyperparameter in our analysis and needs to be optimized, but we use $b = 1$ mK in this work.

The power spectrum of Gaussian noise is described as (Geil, Gaensler, and Wyithe, 2011)

$$P_{\mathrm{noise}}(\boldsymbol{k}_\perp) = \frac{T_{\mathrm{sys}}^2}{Bt_0}\frac{\chi^2\Delta\chi}{n_{\mathrm{b}}(k_\perp D/2\pi, \nu)}\left(\frac{\lambda^2}{A_{\mathrm{e}}}\right)^2, \tag{5.17}$$

where $T_{\mathrm{sys}}/A_{\mathrm{e}} \sim 0.2$ represents the ratio of the system temperature to the effective area, B is the frequency band, t_0 represents the total integration time of the assumed observation, χ $[h^{-1}\mathrm{Mpc}]$ is the comoving distance from us to the source at $z = 3.0$,

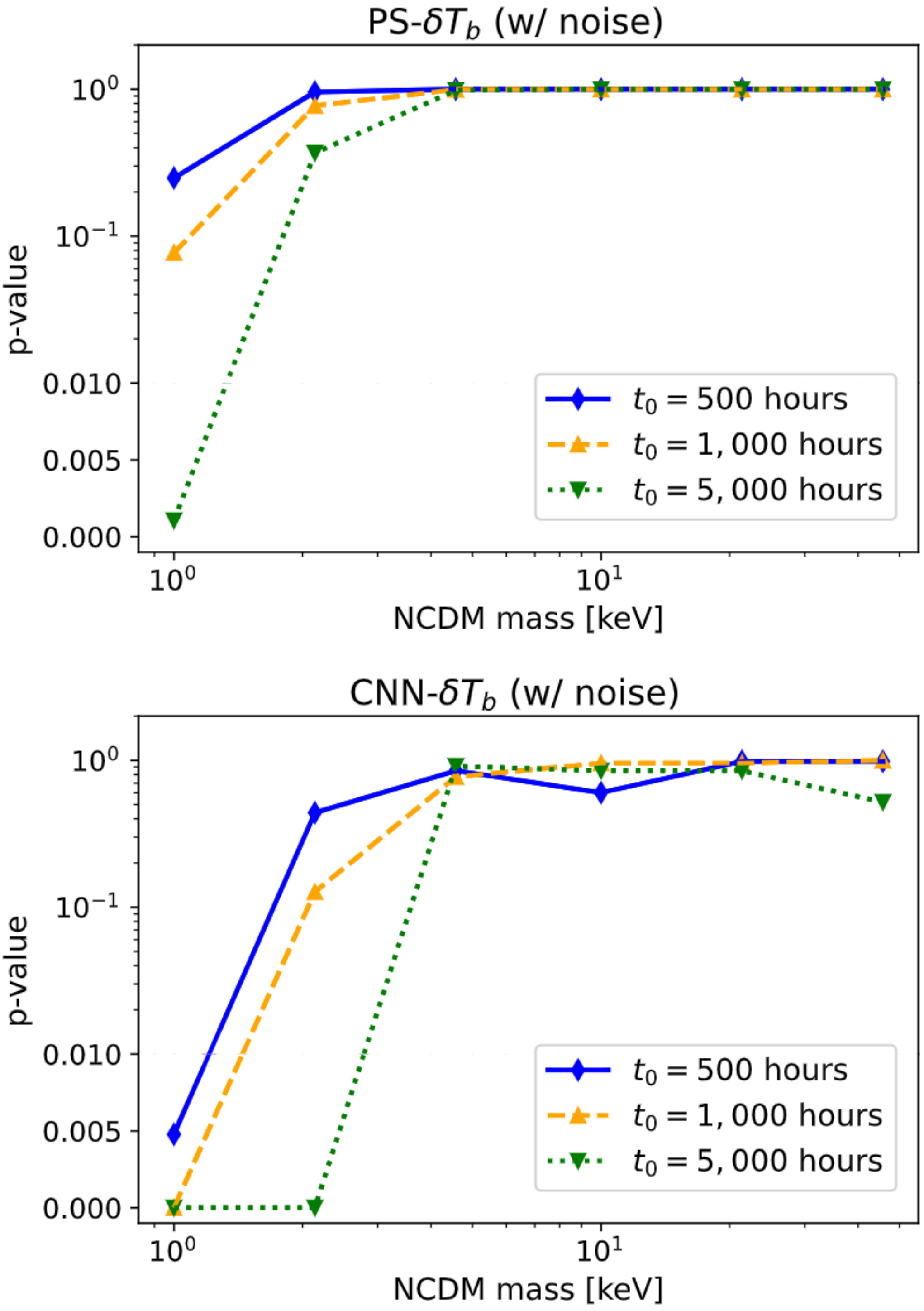

FIGURE 5.12: The p-values for the classification of the noised images with the observation time t_0 =500 (blue solid), 1,000 (orange dashed), and 5,000 (green dotted) hours. The upper and lower panels corresponds to the results of the power spectrum (upper panel) and our CNN (lower panel), respectively.

$\Delta\chi$ is the depth of the survey volume, λ is the observed wavelength of the 21cm signals, and c and ν_0 are the speed of light and the rest frequency of a 21cm signal. In our analysis, $\Delta\chi$ is 50 $[h^{-1}Mpc]$, which is the projection length in making images and we can approximate $\Delta\chi \sim \frac{c(1+z)^2}{\nu_0 H(z)} B$. Here, B is $\sim$ 4.5 MHz and the SKA-Mid can resolve this frequency band. $n_b(U, \nu)$ represents the baseline number density of the interferometer, given by

$$\begin{aligned} n_b(k_\perp \chi/2\pi, \nu) =& C_b \int_0^{r_{\max}} dr\, 2\pi r n_a(r) \\ & \times \int_0^{2\pi} d\phi\, n_a(|\mathbf{r} - \lambda \mathbf{k}_\perp \chi/2\pi|), \end{aligned} \tag{5.18}$$

where ν is the observed frequency, r [km] is the distance between the center of the region where the antennas distribute and another antenna, $r_{\max}$ is the maximum value of r and is $= 150$ km, n_a is the number density of the antenna described as

$$n_a(r) = \begin{cases} n_c & (r \le r_c) \\ n_c\,(r_c/r)^2 & (r_c < r < r_{\max}) \\ 0 & (r \ge r_{\max})\ , \end{cases} \tag{5.19}$$

where n_c represents the number density of the antenna in its core region and is written as $n_c = N_a/(\pi r_c^2 + \ln(r_{\max}/r_c))$, r_c [km] is the core radius and is 2 km, $Na = 197$ is the total number of antennas. C_b in Eq. (5.18) is the normalization constant and is determined so that n_b satisfies

$$(1/2\pi) \int k_\perp \chi^2 n_{\mathrm{b}}(k_\perp \chi/2\pi) dk_\perp = N_{\mathrm{a}}(N_{\mathrm{a}} - 1)/2. \tag{5.20}$$

The specifications for the observation used in this calculation are taken from the pages[1][2][3]. Note that the field of view of the SKA-Mid is $\sim$ 10 $[\mathrm{deg}^2]$ while the box size ($= 100$ $[h^{-1}\mathrm{Mpc}]$) corresponds to 1.6 $[\mathrm{deg}^2]$. Therefore, the practical observation by the SKA-Mid can provide us with a larger survey area and this difference can affect the results.

In this work, we assume the integration times of $t_0 = 500$ hours, 1000 hours following Villaescusa-Navarro et al., 2015, 5000 hours to investigate the effects of the integration time. The value of $t_0 = 1000$ hours is often used such as (Villaescusa-Navarro et al., 2015; Crocce, Pueblas, and Scoccimarro, 2006; Pritchard et al., 2015), for example.

Figure 5.10 shows the 2D power spectra of δT_b only (orange), noise only (green), and their combination (blue). We can see that the noise power spectrum exceeds the one of the δT_b signals around $k_\perp > 5\ h\mathrm{Mpc}^{-1}$. Figure 5.11 shows examples of the images of δT_b only (left), noise only (middle), and $\delta T_b +$ noise (right). The bright regions are barely visible in the right image. Figure 5.9 also shows the fractional difference of $P^{\delta T_b}(k_\perp)$ for the NCDM models with respect to the one for the CDM model. It compares these difference with the errors, including the system noise depicted in the light-gray shaded region. We can see the dark matter model classification is challenging for the images including the system noise.

---

[1] https://www.astron.nl/telescopes/square-kilometre-array/

[2] https://www.skao.int/en/science-users/118/ska-telescope-specifications

[3] https://indico.skatelescope.org/event/940/contributions/8511/attachments/7800/12784/MA-SUM_Dish%20pdf.pdf

For the images including the system noise, we train our CNN following the procedure described in Section 5.2.2. Figure 5.12 shows the classification results described by the p-values of the KS test for PS-δT_b (upper panel) and CNN-δT_b (lower panel). Each line shows the p-value as a function of $m_{\rm DM}$ for the classification of the noised images with $t_0 = 500$ (blue solid), 1000 (orange dashed), and 5000 (green dotted) hours, respectively. PS-δT_b can classify the noised images of the CDM and 1 keV NCDM with $t_0 = 5000$ hours with high significance ($\sim$ 3-σ) but cannot distinguish the 1 keV NCDM from CDM with $t_0 = 1000$ hours. On the other hand, CNN-δT_b can distinguish the images between the CDM and the1 keV NCDM with $t_0 = 500$ hours as well as the 2.1 keV NCDM with $t_0 = 5000$ hours. The system noise prevents the classifications with both the power spectrum and CNN, but CNN still outperforms the power spectrum. The feature of 1 keV NCDM mass is obscured by the system noise of SKA-MID with $t_0 = 1000$ hours observation in the power spectrum analysis, but our CNN succeeds in capturing this feature.

Finally, we discuss the comparison between the effects of the astrophysical models and the system noise. The classifications with CNN-δT_b without considering the system noise are not largely affected by the *Shield* and *NoSF* models and are affected by *FG09* model for $m_{\rm DM} > 4.6$ keV as shown in Table. 5.2. However, the classifications of CNN-δT_b are affected by the system noise with $t_0 = 1000$ hours for $m_{\rm DM} > 2.1$ keV as shown in Figure 5.12. Therefore, the effects of the system noise exceed those of the difference of the astrophysical models, but the effects of the astrophysical models become important in future observations with higher sensitivity.

5.4 Conclusion

In this work, we investigate the image-based CNN analysis of 21cm intensity mapping to constrain the mass of dark matter particle. We conduct the cosmological hydrodynamic simulations with the different dark matter models, such as the CDM model and NCDM models with varying the mass of dark matter particle. We generate the images of the mass density distribution of dark matter, and those of the 21cm signals described by the differential brightness temperature δT_b. And then, we compare the performance of the classification using CNN and the power spectrum and demonstrate the potential of the CNN analysis.

Firstly, we compare the results of the classifications for the four cases, such as PS-DM, CNN-DM, PS-δT_b, and CNN-δT_b. Here, the AUC and KS test are used to evaluate the classifications for our CNN, and both of our CNN and the power spectrum, respectively. In Figure 5.5, we can see that the results described by the AUC for both images of dark matter and δT_b are comparable. We can conclude that δT_b is a valid tracer of the dark matter distribution in our CNN analysis.

In Figure 5.6, we compare the performance of the classification described by the p-values of the KS test between our CNN and the power spectrum. For both the images of dark matter and δT_b, our CNN outperforms the power spectrum. Specifically, our CNN can classify the CDM model and the 10 keV NCDM model with high significance, while the power spectrum can distinguish between the CDM and 2.1 keV NCDM.

Next, we examine the effects of the different astrophysical models. To do this, we conduct the additional simulations of the CDM model with varying astrophysical models, such as *Shield*, *NoSF*, and *FG09*. Subsequently, we generate images from these additional simulations and replace the test images of the *Fiducial* CDM with those of the different astrophysical models. From Figure 5.7, we obtain the following results; For PS-δT_b, we do not find the significant impacts of the different astrophysical models on the classification. However, for CNN-δT_b, we can see *FG09* model has a large effect on our classifications. From these results, we conclude that our CNN is robust against the difference of the astrophysical models we consider, for $m_{\mathrm{DM}} < 2.1$ keV.

Table 5.2 also shows the effect of the *FG09* model, and our CNN cannot correctly classify the images of the *FG09* model from the $m_{\mathrm{DM}} > 4.6$ keV NCDM. In Figure 5.8, which shows the confusion matrix, we see the *FG09* CDM images are misclassified as NCDM. We explore how different astrophysical models have effects on the power spectrum in Figure 5.9. However, the behaviors of the power spectra for the different astrophysical models cannot explain the reason for the misclassifications of CNN for *FG09* model. There is a possibility that the results are affected by the presence of more HI gas in the *FG09* model compared to the UV background model used in *Fiducial* model (Haardt and Madau, 2012).

Finally, we study the effects of the system noise. In this work, we assume the SKA-MID survey. As a simplistic assumption, we add the Gaussian noise to the images obtained from the simulations. The power spectrum and image considering this noise are shown in Figure 5.10 and Figure 5.11. Figure 5.9 compares the power spectrum differences between the CDM model and the NCDM models, considering errors including system noise depicted by the light gray-shaded region As we can see in this figure, the classifications of dark matter models become challenging at the presence of the system noise. In fact, Figure 5.12 shows the degradation of the performance of PS-δT_b and CNN-δT_b. However, even if the system noise is included in images, the performance of our CNN exceeds than the one of the power spectrum. As an example, CNN-δT_b can distinguish the 1 keV NCDM model with the noise from the $t_0 = 500$ hours observation, and the 2.1 keV NCDM model with the noise for $t_0 = 5000$ hours, while PS-δT_b can only distinguish the 1 keV NCDM model with $t_0 = 5000$ hours.

In addition, we compare the impacts of the different astrophysical models and those of the system noise. For the astrophysical model, *FG09* model has a significant effect on the classification of CNN-δT_b for $m_{\mathrm{DM}} > 4.6$ keV as shown in Table 5.2. However, the system noise with $t_0 = 1000$ hours has a large effect on the classification for $m > 2.1$ keV. Therefore, the effects of the different astrophysical models are hidden behind those of the system noise in our assumption.

This work demonstrates the potential of the CNN to constrain the particle mass of dark matter more effectively than the power spectrum, where We assume the δT_b signals as an observable, which can be observed by radio observations such as SKA instead of the dark matter distribution. However, some effects, which we ignore but should consider in practical observation, such as foreground contamination, the selection of the target redshift, and redshift space distortion, are not considered yet. We will revisit these challenges in future work.

Chapter 6

Summary

In this thesis, we focus on machine learning approaches, to investigate the elements beyond the ΛCDM model. The ΛCDM model consists of two assumptions: Λ as the cosmological constant, related to dark energy causing the accelerated expansion of the universe, and cold dark matter (CDM), which is the dominant component of matter in the universe and derive the growth of the large-scale structure of the universe. However, their properties are still unknown, and there is a room for modifications.

To investigate models beyond ΛCDM, we consider two quantities; the growth rate and mass of non-cold dark matter (NCDM), which are related to the accelerated expansion of the universe and dark matter, respectively. However, we encounter some challenges, such as the non-linear modeling of the power spectrum for measuring the growth rate and the analysis of the non-Gaussian distribution of matter caused by non-linear evolution for constraining the mass of NCDM. To solve these problems, we utilize machine learning approaches. The image-based analysis with machine learning enables us to analyze the large-scale structure of the universe directly from its image.

This thesis demonstrates the potential of machine learning approaches to predict the growth rate through observational data, such as galaxy clustering, without a theoretical model of the statistics of observables. It also explores the constraint on the mass of NCDM using the HI distribution, a tracer of the dark matter distribution observable through radio frequency observations. This paper consists of two works with machine learning approaches: (1) estimation of the growth rate and (2) constraint on the mass of NCDM.

(1) This work is in Chapter 4 and Murakami et al., 2023b.

The growth rate $f\sigma_8$ represents the time evolution of the density perturbation of matter. Structure formation is mainly caused by gravity, and therefore the growth rate reflects the effects of modifications to the standard gravity theory. Therefore, it is often used as a probe of modified gravity theories. Modified gravity theories alter the equations describing the evolution of the universe, and may explain the origin of the accelerated expansion of the universe.

In general, we measure the growth rate through the observations of the large-scale structure such as galaxy clustering, and two-point statistics such as the power spectrum are used to analyze the data. However, estimating $f\sigma_8$ requires a theoretical model of the statistics of observables, and modeling at non-linear scales is challenging.

To avoid this modeling difficulty, we use convolutional neural networks (CNN) and estimate $f\sigma_8$ directly from the matter distribution in the N-body simulations without theoretical modelings. In this work, we use the Quijote simulations, which are the publicly available dataset of the N-body simulations. We generate 3-dimensional images of the distribution of dark matter particles and dark matter halo from the simulations with varying cosmological parameters and train our CNN.

First, we show the results of the prediction of $f\sigma_8$ for images of dark matter and halo in Figure 4.4, 4.5, 4.6. We find that $\Delta f\sigma_8$ is centered around zero for each analysis of dark matter, halo, and randomly selected halo. Therefore, we conclude that our CNN successfully constrains $f\sigma_8$ without a significant bias.

In addition, this work compares the errors of the predicted value of $f\sigma_8$ from our CNN with those of the power spectrum analysis with machine learning as shown in Figure 4.7. As a result, our CNN shows the redshift dependence of the results and provides smaller errors at high redshift while it shows larger errors at low redshift. One possible reason for these results is that our CNN might struggle to obtain the information in the data at scales of the non-linear regime at low redshift. Another possibility is that there is room for optimizing our CNN architecture for low redshift data.

We also investigate the reason for the redshift dependence of our CNN results, which the Fisher analysis does not show. We show some possibilities: non-linearity, parameter sampling, choice of the training, validation, and test dataset, random selection of halo, and random seed of the initial condition of the simulations. We have examined these possibilities, but we do not completely explain the reason for the redshift dependence of our CNN results. We leave the extension of our analysis for future works.

(2) This work is in Chapter 5 and Murakami et al., 2023a.

Since the mass of NCDM affects the matter distribution at small scales due to their free streaming, we can obtain information of the NCDM mass from the analysis of the matter distribution in the universe. However, we cannot observe the dark matter distribution directly, and the distribution is highly non-Gaussian at small scales including the information of the NCDM mass.

In this work, we focus on the HI distribution, which can be observed as 21cm signals δT_b, as a tracer of dark matter distribution, and image-based analysis with CNN to extract features from the non-Gaussian distribution of δT_b.

We use the cosmological hydrodynamic simulation `GADGET3-Osaka` in the CDM model and NCDM models with different masses, and generate images of the distribution of dark matter and δT_b to train and test our CNN. In this work, to demonstrate the potential of CNN to constrain the mass of NCDM, we classify the images of CDM and NCDM by CNN and the power spectra and evaluate the classification results by comparing these methods.

Figure 5.5 shows the difference in the results of our CNN between the cases where we use the images of dark matter and δT_b. These results are comparable, and we conclude that δT_b is a good tracer of the dark matter distribution in our CNN analysis. Then, we compare the results of CNN with those of the power spectrum in Figure 5.6. As a result, we can see that our CNN exhibits better performance in the classification of CDM and NCDM for both the analyses of dark matter and δT_b images.

Next, we investigate the effects of the different astrophysical models such as *Shield*, *NoSF*, and *FG09* by conducting additional simulations with the CDM model for these astrophysical models. We replace test images of the *Fiducial* CDM with those of the different astrophysical models to test our CNN. Figure 5.7 shows the classification results for the different astrophysical models. There is little effect of *Shield* and *NoSF* models on our CNN analysis, but *FG09* model has a large effect. From the confusion matrix shown in Figure 5.8 and the AUC values shown in Table 5.2, we can see that the images of the FG CDM model are misclassified as NCDM.

Therefore, the features of NCDM might be partly mimicked by the *FG09* model. We do not see the effect of the *FG09* model on the classification for $m_{\rm DM} < 2.1$ keV, so we conclude our CNN is robust against the different astrophysical models if we consider $m_{\rm DM} < 2.1$ keV.

Finally, we investigate the effects of the system noise, assuming SKA-MID observations. Gaussian noise, whose power spectrum is shown in Figure 5.10 as an example, is added to the simulation data like Figure 5.11 and the noised images are generated. Figure 5.12 shows the results of the classification of the noised images of CDM and NCDM by our CNN and the power spectrum. The system noise degrades the performance of both CNN and the power spectrum analysis, but our CNN outperforms the power spectrum for the classification of the noised images. In addition, we compare the effects of the system noise with those of the different astrophysical models and find that the effects of the different astrophysical models are hidden behind the system noise with $t_0 = 1000$ hours or less.

Over the next decade, we will obtain a large amount of data with high accuracy from the forthcoming surveys.

For the constraints of $f\sigma_8$ by the analysis of the galaxy clustering, we need more galaxy samples and the three or higher statistics. The forthcoming galaxy surveys *Euclid* Laureijs et al., 2011, LSST Abell et al., 2009, and DESI Aghamousa et al., 2016 provide us with a large number of the galaxy samples for the large survey area and the machine learning approaches enable us to investigate the galaxy distribution in more detail and at higher speed. Our work in Chapter4 demonstrates the potential of CNN. However, we must consider some effects of the practical observations such as the masked region in the survey area, the redshift distribution of the galaxies, and the selection of galaxies. we will investigate these effects in our future work. In addition, we leave the question for the comparison of the results between the machine learning and the Fisher analysis. We will also revisit this problem.

For the dark matter mass constraints, we will obtain the HI distribution in the universe by the radio interferometer, SKA with a high resolution. To constrain the mass of the dark matter, we need information of the matter distribution at small scales. At small scales, the matter distribution does not follow the Gaussian due to the non-linear evolution of the matter density and the higher-order statistics have the important information for the dark matter mass. The image-based analysis by machine learning allows us to access this information than the power spectrum analysis as we demonstrate in Chapter 5 in the noise-level of the SKA observations. To apply our method of analysis to the data from the practical observations, we need further examination of the effects of the foreground contamination and the survey area. Additionally, we can consider other redshifts for the observation or joint analysis of the multiple redshifts to improve the constraints on the dark matter mass in future work.

In our works, we use the CNN for the image-based analysis. In recent years, many machine learning algorithms for image analysis have been developed. Recurrent Neural Network (RNN) is one of the developed neural network algorithms. In RNN, the data is inputted to the layers in the middle of the network. By doing this, RNN can avoid the vanishing of information of the input data and can have a deeper network than a usual neural network. Another example is the Bayesian Neural network (BNN). In BNN, the weight parameters are given by the probability distribution function. BNN can avoid overfitting training data and evaluate the uncertainty of the output explicitly. These algorithms are applied to some cosmological analyses such as Escamilla-Rivera, Carvajal Quintero, and Capozziello, 2020. As another method for image analysis, a Vision Transformer (ViT) is suggested. This is not a kind of neural network. ViT applies the algorithm for a Large Language Model to the image-based analysis. A previous work demonstrates the ViT performance for the cosmological image-based analysis. This work does not show the superiority of the ViT to CNN. However, in the other case, ViT outperforms CNN (Raghu et al., 2021) and there is probably room for improvement of the ViT analysis in cosmology. Additionally, we usually use the simulation data to train our machine-learning architecture and need a large amount and various cosmological simulations. In recent years, there has been a lot of public simulation data such as Quijote simulations (Villaescusa-Navarro et al., 2020) and CAMELS project (Villaescusa-Navarro et al., 2021). Therefore, we can expect the improvement of the analysis by the new algorithm and training data.

Our works demonstrate the potential of image-based analysis by using CNN for the constraints of the growth rate and the mass of dark matter and the possibility that the image-based analyses outperform the conventional method such as power spectrum. To apply our method to the practical data and improve the constraints, we should consider the effects of the practical observations, another algorithm of machine learning, and more simulation data in future work.

Appendix A

Loss Functions tests

redshift	MSE Loss	MAE Loss	Huber Loss	LFI Loss
0.75	3.8	4.1	3.9	3.3
1.25	2.3	3.1	2.9	2.3
1.75	1.2	1.5	1.8	1.7
2.5	0.74	1.0	1.1	1.3

$(\times 10^{-2})$

TABLE A.1: Predicted errors from our CNN for each redshift using Mean Squared Error (MSE), Mean Absolute Error (MAE), Huber Loss and the moments networks Loss (LFI) Functions.

This appendix is related to Chapter 4.

We introduce several loss functions in Section 3.2.1. In the work of Chapter 4, we use the MSE loss in the calculations for the main results. Here, we show the effects of the choice of the loss function on the errors, which is defined as the standard deviation of $\Delta f\sigma_8$, from our CNN analysis. In our work, the LFI loss is written as

$$\mathcal{E}_{\rm LFI} = \log\left(\sum_{j\in\ \rm batch} (f\sigma_{8,j} - \mu_j)^2\right) + \log\left(\sum_{j\in\ \rm batch} \left((f\sigma_{8,j} - \mu_j)^2 - \sigma_j^2\right)^2\right), \quad \text{(A.1)}$$

where the summation is done over the data in a mini-batch, μj and σ_j is the mean and standard deviation of the true value of $f\sigma_8$ in the mini-batch.

Table A.1 shows the errors from our CNN analysis for each redshift for the MSE loss, MAE loss, Huber loss, and LFI loss function. We find that our CNN with each loss function provides a similar error. The LFI loss shows a slightly better result at the low redshift $z = 0.75$. However, the loss curve, which is the change of the loss function with epochs, for the MSE loss shows a smaller variance than the one for the LFI loss, and the training with the MSE loss is more stable. Therefore, we use the MSE loss to obtain the final result.

Appendix B

the Property of HI halo

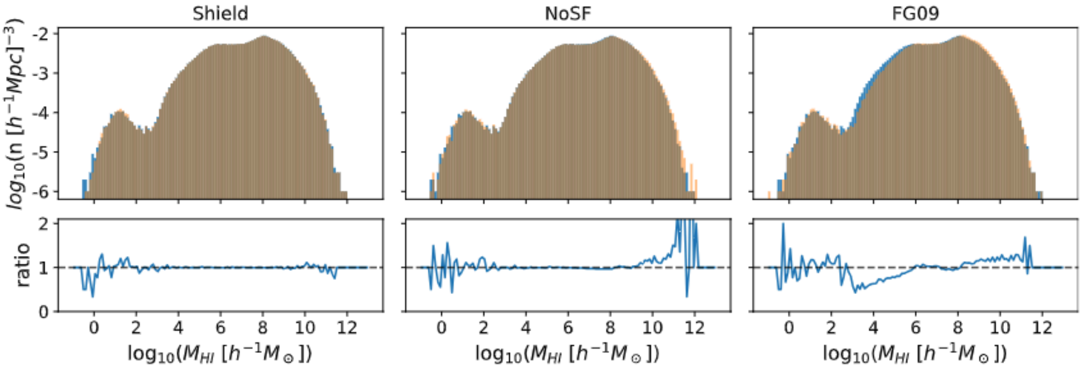

FIGURE B.1: The mass function of HI in halos. The horizontal axis represents the HI mass in a halo, and the vertical axis represents the comoving number density of the halos, where the halo includes HI mass corresponding to the horizontal axis. The upper and lower panel show the histograms of *Fiducial* (blue) and other astrophysical models (orange), and their ratios, respectively.

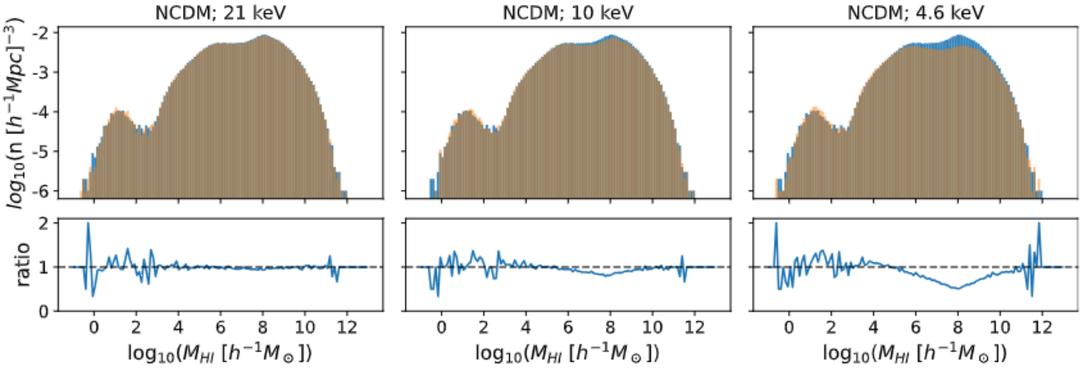

FIGURE B.2: This figure shows the same as Fig. B.1, but the orange histograms correspond to the NCDM models.

In this appendix, we investigate the halo properties for further discussion related to Section 5.3.2.

As we discussed in Fig. 5.7, the *FG09* CDM test images are misclassified as the NCDM model. In this appendix, we explore the effects of the astrophysical models on the halo properties such as its size, number, mass function, and density profile. We identify the dark matter halo using the ROCKSTAR code (Behroozi, Wechsler, and Wu, 2013), and the size of the dark matter halo is defined by its virial radius.

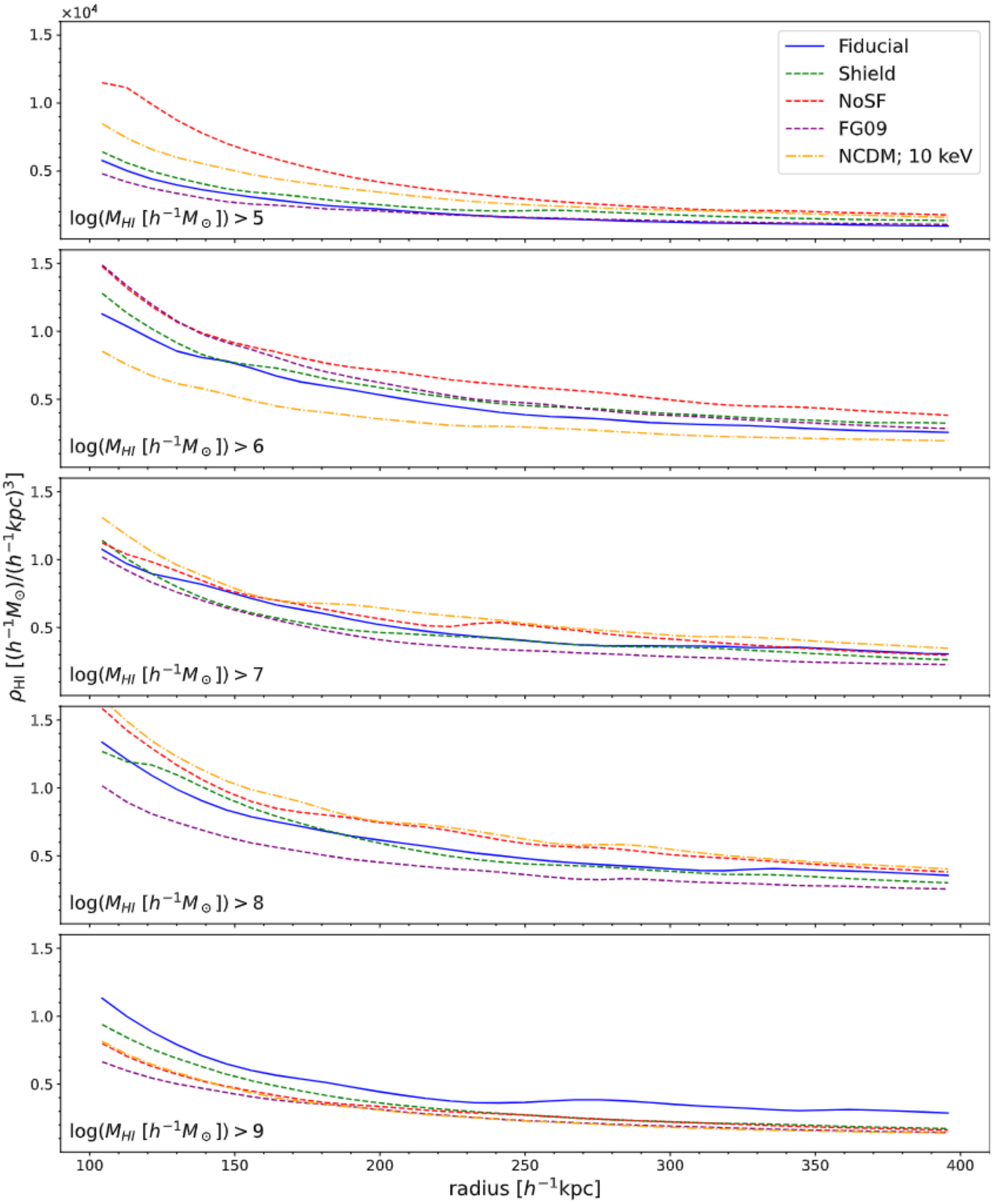

FIGURE B.3: The HI mass density profile of halos. The horizontal axis represents the halo radius. To calculate the density profile, we pick up and stack the 3000 halos, where the lowest mass of the halos is $\log\left(M_{\rm HI}\ [h^{-1}M_\odot]\right) \geq 5, 6, 7, 8$ and 9, corresponding to each panel from top to bottom. The stacked density profiles for each model are shown, such as *Fiducial* (blue solid), *Shield* (green dashed), *NoSF* (red dashed), *FG09* (purple dashed), and 10 keV NCDM (orange dash-dot). Note that our images cannot resolve the scales $\sim$ 100 h^{-1}kpc

Model	μ [h^{-1}kpc]	σ [h^{-1}kpc]	Total Number
Fiducial	66.2	28.2	270,931
Shield	66.2	28.2	270,930
NoSF	66.2	28.2	271,092
FG09	66.2	28.2	270,936
46 keV	66.2	28.2	269,283
21 keV	66.2	28.5	264,444
10 keV	66.2	28.5	245,598
4.6 keV	65.2	31.9	196,314
2.1 keV	61.9	34.6	126,987
1 keV	59.9	36.7	54,755

TABLE B.1: The properties of dark matter halos, which are its mean size (μ) and its standard deviation σ, and its number, in a simulation box for each astrophysical model and NCDM model. Each row corresponds to the difference between the astrophysical models and NCDM models.

And then, the HI halo is identified as the collection of the HI gas particles included the dark matter halo.

In Table B.1, we show the mean size, the standard deviation of the halo size, and the number of dark matter halos in a simulation box. Each row corresponds to the difference of the assumed model for the astrophysics or NCDM. It is observed that there is no significant relation between the halo number and the assumed astrophysical models. However, the halo number decreases when the NCDM mass is smaller, especially for NCDM models with $m_{\mathrm{DM}} \leq 4.6$ keV. This reduction is attributed to the prevention of clustering in light dark matter due to the free-streaming of NCDM (refer to Section 2.7), resulting in the absence of halo formation. In contrast, the halo size remains unaffected by both astrophysical models and NCDM models.

Fig. B.1 and Fig. B.2 show the comparison of HI mass function between the *Fiducial* CDM model and the astrophysical models or the NCDM models, respectively. The count of halos with $M_{\mathrm{HI}} > 10^9 M_\odot/h$ shows an increases in the *NoSF* and *FG09* models compared to the *Fiducial* model. Simultaneously, halos with $M_{\mathrm{HI}} \sim 10^8\ h^{-1}M_\odot$ decrease for $m_{\mathrm{DM}} \leq 10$ keV NCDM models. Nevertheless, these observed trends fail to elucidate the resemblance between the *FG09* model and the NCDM model in terms of classification. We do not find the similarities in the HI halo mass functions between the NCDM models and *FG09* model.

Subsequently, we will examine the similarity of the halo density profile between the NCDM models and *FG09* model. Fig. B.3 shows the mass density profiles of the HI halo. This figure shows the averaged HI density profiles over the 3000 halos, of which the lowest HI halo mass are 10^5, 10^6, 10^7, 10^8, and $10^9\ h^{-1}M_\odot$ in each panel from top to bottom. The halos in the *Fiducial* (blue solid) and *Shield* (green dashed) models have relatively comparable profiles. For *NoSF* (red dashed) and 10 keV NCDM (orange dash-dot) models, the halo profiles are similar, except for the second panel corresponding to $\log\left(M_{\mathrm{HI}}\ [h^{-1}M_\odot]\right) \geq 6$. However, the results in Section 5.3.2 do not show the misclassification between the *NoSF* CDM model and the 10 keV NCDM model. Concerning the *FG09* model, its profiles (purple dashed) are different from *Fiducial*'s ones, particularly for massive halos ($M_{\mathrm{HI}} > 10^8\ h^{-1}M_\odot$), and they also differ from the NCDM profile. As indicated in Section 5.3.2, our CNN misclassification is likely instigated not only by features resembling NCDM but also by features that diverge from CDM-like characteristics.

Bibliography

Abbott, Laurence F and P Sikivie (1983). “A cosmological bound on the invisible axion”. In: *Physics Letters B* 120.1-3, pp. 133–136.

Abdalla, E. et al. (Dec. 2019). “Brazilian Community Report on Dark Matter”. In: *arXiv e-prints*, arXiv:1912.10076, arXiv:1912.10076. DOI: 10.48550/arXiv.1912.10076. arXiv: 1912.10076 [hep-ph].

Abell, Paul A. et al. (Dec. 2009). “LSST Science Book, Version 2.0”. In: arXiv: 0912.0201 [astro-ph.IM].

Aghamousa, Amir et al. (Oct. 2016). “The DESI Experiment Part I: Science,Targeting, and Survey Design”. In: arXiv: 1611.00036 [astro-ph.IM].

Alam, Shadab et al. (2017). “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample”. In: *Monthly Notices of the Royal Astronomical Society* 470.3, pp. 2617–2652.

Alvarez, Alexandre et al. (Sept. 2020). “Dark matter constraints from dwarf galaxies with data-driven J-factors”. In: *J. Cosmology Astropart. Phys.* 2020.9, 004, p. 004. DOI: 10.1088/1475-7516/2020/09/004. arXiv: 2002.01229 [astro-ph.HE].

Amendola, Luca, Valerio Marra, and Miguel Quartin (2013). “Internal robustness: systematic search for systematic bias in SN Ia data”. In: *Monthly Notices of the Royal Astronomical Society* 430.3, pp. 1867–1879.

Ando, Rika et al. (Oct. 2021). “Reconstructing H I power spectrum with minimal parameters using the dark matter distribution beyond haloes”. In: *MNRAS* 507.2, pp. 2937–2948. DOI: 10.1093/mnras/stab2284. arXiv: 2011.13165 [astro-ph.CO].

Aoyama, Shohei et al. (Nov. 2016). “Galaxy simulation with dust formation and destruction”. In: *Monthly Notices of the Royal Astronomical Society* 466.1, pp. 105–121. ISSN: 0035-8711. DOI: 10.1093/mnras/stw3061. eprint: https://academic.oup.com/mnras/article-pdf/466/1/105/10865208/stw3061.pdf. URL: https://doi.org/10.1093/mnras/stw3061.

Baldauf, Tobias et al. (2015). “Equivalence Principle and the Baryon Acoustic Peak”. In: *Phys. Rev. D* 92.4, p. 043514. DOI: 10.1103/PhysRevD.92.043514. arXiv: 1504.04366 [astro-ph.CO].

Bandura, Kevin et al. (July 2014). “Canadian Hydrogen Intensity Mapping Experiment (CHIME) pathfinder”. In: *Ground-based and Airborne Telescopes V*. Ed. by Larry M. Stepp, Roberto Gilmozzi, and Helen J. Hall. Vol. 9145. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 914522. DOI: 10.1117/12.2054950. arXiv: 1406.2288 [astro-ph.IM].

Bauer, Jurek B. et al. (Jan. 2021). “Intensity mapping as a probe of axion dark matter”. In: *MNRAS* 500.3, pp. 3162–3177. DOI: 10.1093/mnras/staa3300. arXiv: 2003.09655 [astro-ph.CO].

Bautista, Julian E et al. (2021). “The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1”. In: *Monthly Notices of the Royal Astronomical Society* 500.1, pp. 736–762.

Behroozi, Peter S., Risa H. Wechsler, and Hao-Yi Wu (Jan. 2013). “The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores”. In: *ApJ* 762.2, 109, p. 109. DOI: `10.1088/0004-637X/762/2/109`. arXiv: `1110.4372 [astro-ph.CO]`.

Bernardeau, Francis, Nicolas Van de Rijt, and Filippo Vernizzi (2012). “Resummed propagators in multi-component cosmic fluids with the eikonal approximation”. In: *Phys. Rev. D* 85, p. 063509. DOI: `10.1103/PhysRevD.85.063509`. arXiv: `1109.3400 [astro-ph.CO]`.

Bernardeau, Francis et al. (2002). “Large-scale structure of the Universe and cosmological perturbation theory”. In: *Physics reports* 367.1-3, pp. 1–248.

Blanchard, A. et al. (2020a). “Euclid preparation: VII. Forecast validation for Euclid cosmological probes”. In: *Astron. Astrophys.* 642, A191. DOI: `10.1051/0004-6361/202038071`. arXiv: `1910.09273 [astro-ph.CO]`.

Blanchard, Alain et al. (2020b). “Euclid preparation-VII. Forecast validation for Euclid cosmological probes”. In: *Astronomy & Astrophysics* 642, A191.

Bonjean, V. (Feb. 2020). “Deep learning for Sunyaev-Zel’dovich detection in Planck”. In: *A&A* 634, A81, A81. DOI: `10.1051/0004-6361/201936919`. arXiv: `1911.10778 [astro-ph.CO]`.

Brando, Guilherme et al. (2022). “Enabling matter power spectrum emulation in beyond-ΛCDM cosmologies with COLA”. In: *JCAP* 09, p. 051. DOI: `10.1088/1475-7516/2022/09/051`. arXiv: `2203.11120 [astro-ph.CO]`.

Bueno belloso, Alicia, Juan Garcia-Bellido, and Domenico Sapone (2011). “A parametrization of the growth index of matter perturbations in various Dark Energy models and observational prospects using a Euclid-like survey”. In: *JCAP* 10, p. 010. DOI: `10.1088/1475-7516/2011/10/010`. arXiv: `1105.4825 [astro-ph.CO]`.

Carrasco, John Joseph M. et al. (2014). “The Effective Field Theory of Large Scale Structures at Two Loops”. In: *JCAP* 07, p. 057. DOI: `10.1088/1475-7516/2014/07/057`. arXiv: `1310.0464 [astro-ph.CO]`.

Carrilho, Pedro, Chiara Moretti, and Alkistis Pourtsidou (2023). “Cosmology with the EFTofLSS and BOSS: dark energy constraints and a note on priors”. In: *JCAP* 01, p. 028. DOI: `10.1088/1475-7516/2023/01/028`. arXiv: `2207.14784 [astro-ph.CO]`.

Carucci, Isabella P., Melis O. Irfan, and Jérôme Bobin (Nov. 2020). “Recovery of 21-cm intensity maps with sparse component separation”. In: *MNRAS* 499.1, pp. 304–319. DOI: `10.1093/mnras/staa2854`. arXiv: `2006.05996 [astro-ph.CO]`.

Carucci, Isabella P. et al. (July 2015). “Warm dark matter signatures on the 21cm power spectrum: intensity mapping forecasts for SKA”. In: *J. Cosmology Astropart. Phys.* 2015.7, 047, p. 047. DOI: `10.1088/1475-7516/2015/07/047`. arXiv: `1502.06961 [astro-ph.CO]`.

Casas, Santiago et al. (2017). “Linear and non-linear Modified Gravity forecasts with future surveys”. In: *Phys. Dark Univ.* 18, pp. 73–104. DOI: `10.1016/j.dark.2017.09.009`. arXiv: `1703.01271 [astro-ph.CO]`.

Chapman, Michael J., Zhongxu Zhai, and Will J. Percival (Oct. 2023). “Isolating the linear signal when making redshift space distortion measurements”. In: *MNRAS* 525.2, pp. 2135–2153. DOI: `10.1093/mnras/stad2351`. arXiv: `2302.11621 [astro-ph.CO]`.

Chapman, Michael J et al. (2022). “The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the growth rate of structure from the small-scale clustering of the luminous red galaxy sample”. In: *Monthly Notices of the Royal Astronomical Society* 516.1, pp. 617–635.

Chen, Zhaoting et al. (Sept. 2023). "Detecting the H I power spectrum in the post-reionization Universe with SKA-Low". In: *MNRAS* 524.3, pp. 3724–3740. DOI: `10.1093/mnras/stad2102`. arXiv: `2302.11504 [astro-ph.CO]`.

Chiba, Takeshi, Naoshi Sugiyama, and Takashi Nakamura (Aug. 1997). "Cosmology with x-matter". In: *MNRAS* 289.2, pp. L5–L9. DOI: `10.1093/mnras/289.2.L5`. arXiv: `astro-ph/9704199 [astro-ph]`.

Colombi, Stephane, Scott Dodelson, and Lawrence M. Widrow (Feb. 1996). "Large-Scale Structure Tests of Warm Dark Matter". In: *ApJ* 458, p. 1. DOI: `10.1086/176788`. arXiv: `astro-ph/9505029 [astro-ph]`.

Contreras, Carlos et al. (2013). "The WiggleZ Dark Energy Survey: measuring the cosmic growth rate with the two-point galaxy correlation function". In: *Mon. Not. Roy. Astron. Soc.* 430, p. 924. DOI: `10.1093/mnras/sts608`. arXiv: `1302.5178 [astro-ph.CO]`.

Crocce, Martín, Sebastián Pueblas, and Román Scoccimarro (Nov. 2006). "Transients from initial conditions in cosmological simulations". In: *MNRAS* 373.1, pp. 369–381. DOI: `10.1111/j.1365-2966.2006.11040.x`. arXiv: `astro-ph/0606505 [astro-ph]`.

Cybenko, George (1989). "Approximation by superpositions of a sigmoidal function". In: *Mathematics of control, signals and systems* 2.4, pp. 303–314.

De Mattia, Arnaud et al. (2021). "The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the emission line galaxy sample from the anisotropic power spectrum between redshift 0.6 and 1.1". In: *Monthly Notices of the Royal Astronomical Society* 501.4, pp. 5616–5645.

Dine, Michael and Willy Fischler (1983). "The not-so-harmless axion". In: *Physics Letters B* 120.1-3, pp. 137–141.

Dodelson, S. and F. Schmidt (2020). *Modern Cosmology*. Elsevier Science. ISBN: 9780128159484. URL: `https://books.google.es/books?id=GGjfywEACAAJ`.

Dodelson, Scott and Lawrence M. Widrow (Jan. 1994). "Sterile neutrinos as dark matter". In: *Phys. Rev. Lett.* 72.1, pp. 17–20. DOI: `10.1103/PhysRevLett.72.17`. arXiv: `hep-ph/9303287 [hep-ph]`.

Endo, Takao, Hiroyuki Tashiro, and Atsushi J. Nishizawa (Nov. 2020). "The Alcock Paczynski test with voids in 21 cm intensity field". In: *MNRAS* 499.1, pp. 587–596. DOI: `10.1093/mnras/staa2822`. arXiv: `2002.00348 [astro-ph.CO]`.

Escamilla-Rivera, Celia, Maryi A. Carvajal Quintero, and Salvatore Capozziello (Mar. 2020). "A deep learning approach to cosmological dark energy models". In: *J. Cosmology Astropart. Phys.* 2020.3, 008, p. 008. DOI: `10.1088/1475-7516/2020/03/008`. arXiv: `1910.02788 [astro-ph.CO]`.

Faucher-Giguère, Claude-André et al. (2009). "A NEW CALCULATION OF THE IONIZING BACKGROUND SPECTRUM AND THE EFFECTS OF He II REIONIZATION". In: *The Astrophysical Journal* 703.2, pp. 1416–1443. DOI: `10.1088/0004-637x/703/2/1416`. URL: `https://doi.org/10.1088/0004-637x/703/2/1416`.

Field, George B. (Jan. 1958). "Excitation of the Hydrogen 21-CM Line". In: *Proceedings of the IRE* 46, pp. 240–250. DOI: `10.1109/JRPROC.1958.286741`.

Fujii, Yasunori (1982). "Origin of the gravitational constant and particle masses in a scale-invariant scalar-tensor theory". In: *Phys. Rev. D* 26 (10), pp. 2580–2588. DOI: `10.1103/PhysRevD.26.2580`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.26.2580`.

Furlanetto, Steven R., S. Peng Oh, and Frank H. Briggs (Oct. 2006). "Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe". In: *physrep*

433.4-6, pp. 181–301. DOI: 10.1016/j.physrep.2006.08.002. arXiv: astro-ph/0608032 [astro-ph].

Garzilli, Antonella et al. (Nov. 2019). "The Lyman-α forest as a diagnostic of the nature of the dark matter". In: *MNRAS* 489.3, pp. 3456–3471. DOI: 10.1093/mnras/stz2188. arXiv: 1809.06585 [astro-ph.CO].

Garzilli, Antonella et al. (Apr. 2021). "How to constrain warm dark matter with the Lyman-α forest". In: *MNRAS* 502.2, pp. 2356–2363. DOI: 10.1093/mnras/stab192. arXiv: 1912.09397 [astro-ph.CO].

Geil, Paul M., B. M. Gaensler, and J. Stuart B. Wyithe (Nov. 2011). "Polarized foreground removal at low radio frequencies using rotation measure synthesis: uncovering the signature of hydrogen reionization". In: *MNRAS* 418.1, pp. 516–535. DOI: 10.1111/j.1365-2966.2011.19509.x. arXiv: 1011.2321 [astro-ph.CO].

Götz, Martin and Jesper Sommer-Larsen (July 2002). "Warm Dark Matter and the Missing Satellites Problem". In: *Ap&SS* 281.1, pp. 415–416. DOI: 10.1023/A:1019543230202.

— (Apr. 2003). "Galaxy formation: Warm dark matter, missing satellites, and the angular momentum problem". In: *Ap&SS* 284.2, pp. 341–344. DOI: 10.1023/A:1024073909753. arXiv: astro-ph/0210599 [astro-ph].

Haardt, Francesco and Piero Madau (2012). "RADIATIVE TRANSFER IN A CLUMPY UNIVERSE. IV. NEW SYNTHESIS MODELS OF THE COSMIC UV/X-RAY BACKGROUND". In: *The Astrophysical Journal* 746.2, p. 125. DOI: 10.1088/0004-637x/746/2/125. URL: https://doi.org/10.1088/0004-637x/746/2/125.

Han, Song et al. (2015). "Learning both weights and connections for efficient neural network". In: *Advances in neural information processing systems* 28.

He, Kaiming et al. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Howlett, Cullan et al. (2015). "The clustering of the SDSS main galaxy sample–II. Mock galaxy catalogues and a measurement of the growth of structure from redshift space distortions at z= 0.15". In: *Monthly Notices of the Royal Astronomical Society* 449.1, pp. 848–866.

Huber, P. J. (1964). "Robust estimation of a location parameter". English. In: *Ann. Math. Stat.* 35, pp. 73–101. ISSN: 0003-4851. DOI: 10.1214/aoms/1177703732.

Ioffe, Sergey and Christian Szegedy (Feb. 2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *arXiv e-prints*, arXiv:1502.03167, arXiv:1502.03167. DOI: 10.48550/arXiv.1502.03167. arXiv: 1502.03167 [cs.LG].

Jeffrey, Niall and Benjamin D Wandelt (2020). "Solving high-dimensional parameter inference: marginal posterior densities & Moment Networks". In: *arXiv preprint arXiv:2011.05991*.

Joyce, Austin, Lucas Lombriser, and Fabian Schmidt (2016). "Dark Energy Versus Modified Gravity". In: *Ann. Rev. Nucl. Part. Sci.* 66, pp. 95–122. DOI: 10.1146/annurev-nucl-102115-044553. arXiv: 1601.06133 [astro-ph.CO].

Kim, Ji hoon et al. (2014). "THE AGORA HIGH-RESOLUTION GALAXY SIMULATIONS COMPARISON PROJECT". In: *The Astrophysical Journal Supplement Series* 210.1, p. 14. DOI: 10.1088/0067-0049/210/1/14. URL: https://doi.org/10.1088/0067-0049/210/1/14.

Kim, Ji hoon et al. (2016). "THE

AGORA

/*i*

HIGH-RESOLUTION GALAXY SIMULATIONS COMPARISON PROJECT. II. ISOLATED DISK TEST". In: *The Astrophysical Journal* 833.2, p. 202. DOI: `10.3847/1538-4357/833/2/202`. URL: `https://doi.org/10.3847/1538-4357/833/2/202`.

Kingma, Diederik P. and Jimmy Ba (Dec. 2014). "Adam: A Method for Stochastic Optimization". In: *arXiv e-prints*, arXiv:1412.6980, arXiv:1412.6980. DOI: `10.48550/arXiv.1412.6980`. arXiv: `1412.6980 [cs.LG]`.

Kolmogorov, A. L. (1933). "Sulla determinazione empirica di una legge di distribuzione". In: *G. Ist. Ital. Attuari* 4, pp. 83–91. URL: `https://cir.nii.ac.jp/crid/1571135650766370304`.

Lange, Johannes U et al. (2022). "Five per cent measurements of the growth rate from simulation-based modelling of redshift-space clustering in BOSS LOWZ". In: *Monthly Notices of the Royal Astronomical Society* 509.2, pp. 1779–1804.

Laureijs, R. et al. (Oct. 2011). "Euclid Definition Study Report". In: arXiv: `1110.3193 [astro-ph.CO]`.

Lazanu, Andrei (2021b). "Extracting cosmological parameters from N-body simulations using machine learning techniques". In: *JCAP* 09, p. 039. DOI: `10.1088/1475-7516/2021/09/039`. arXiv: `2106.11061 [astro-ph.CO]`.

— (2021a). "Extracting cosmological parameters from N-body simulations using machine learning techniques". In: *Journal of Cosmology and Astroparticle Physics* 2021.09, p. 039.

Lesgourgues, Julien (Apr. 2011). "The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview". In: *arXiv e-prints*, arXiv:1104.2932, arXiv:1104.2932. arXiv: `1104.2932 [astro-ph.IM]`.

Lin, Min, Qiang Chen, and Shuicheng Yan (Dec. 2013). "Network In Network". In: *arXiv e-prints*, arXiv:1312.4400, arXiv:1312.4400. arXiv: `1312.4400 [cs.NE]`.

Lupton, Robert H., James E. Gunn, and Alexander S. Szalay (Sept. 1999). "A Modified Magnitude System that Produces Well-Behaved Magnitudes, Colors, and Errors Even for Low Signal-to-Noise Ratio Measurements". In: *AJ* 118.3, pp. 1406–1410. DOI: `10.1086/301004`. arXiv: `astro-ph/9903081 [astro-ph]`.

Ma, Chung-Pei and Edmund Bertschinger (1995). "Cosmological perturbation theory in the synchronous and conformal Newtonian gauges". In: *arXiv preprint astro-ph/9506072*.

Matsubara, Takahiko (2014). **大規模構造の宇宙論 宇宙に生まれた絶妙な多様性**. 共立出版.

Modi, Chirag, Yu Feng, and Uroš Seljak (Oct. 2018). "Cosmological reconstruction from galaxy light: neural network based light-matter connection". In: *J. Cosmology Astropart. Phys.* 2018.10, 028, p. 028. DOI: `10.1088/1475-7516/2018/10/028`. arXiv: `1805.02247 [astro-ph.CO]`.

Monaghan, J. J. and J. C. Lattanzio (Aug. 1985). "A refined particle method for astrophysical problems". In: *A&A* 149.1, pp. 135–143.

Murakami, Koya et al. (2023a). *Impact of astrophysical effects on the dark matter mass constraint with 21cm intensity mapping*. arXiv: `2305.01256 [astro-ph.CO]`.

Murakami, Koya et al. (2023b). *Non-Linearity-Free prediction of the growth-rate* $f\sigma_8$ *using Convolutional Neural Networks*. arXiv: `2305.12812 [astro-ph.CO]`.

Nagamine, Kentaro et al. (June 2021). "Probing Feedback via IGM tomography and the Lyα Forest with Subaru PFS, TMT/ELT, and JWST". In: *ApJ* 914.1, 66, p. 66. DOI: `10.3847/1538-4357/abfa16`. arXiv: `2007.14253 [astro-ph.GA]`.

Natarajan, Priyamvada et al. (Feb. 2017). "Mapping substructure in the HST Frontier Fields cluster lenses and in cosmological simulations". In: *Monthly Notices of the Royal Astronomical Society* 468.2, pp. 1962–1980. ISSN: 0035-8711. DOI: `10.1093/mnras/stw3385`. eprint: `https://academic.oup.com/mnras/article-pdf/468/2/1962/11210742/stw3385.pdf`. URL: `https://doi.org/10.1093/mnras/stw3385`.

Nesseris, Savvas (2022). "The Effective Fluid approach for Modified Gravity and its applications". In: *Universe* 9.1, p. 13.

Nesseris, Savvas, George Pantazis, and Leandros Perivolaropoulos (2017). "Tension and constraints on modified gravity parametrizations of $G_{\rm eff}(z)$ from growth rate and Planck data". In: *Phys. Rev. D* 96.2, p. 023542. DOI: `10.1103/PhysRevD.96.023542`. arXiv: `1703.10538 [astro-ph.CO]`.

Neveux, Richard et al. (2020). "The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from the anisotropic power spectrum of the quasar sample between redshift 0.8 and 2.2". In: *Monthly Notices of the Royal Astronomical Society* 499.1, pp. 210–229.

Newburgh, L. B. et al. (Aug. 2016). "HIRAX: a probe of dark energy and radio transients". In: *Ground-based and Airborne Telescopes VI*. Ed. by Helen J. Hall, Roberto Gilmozzi, and Heather K. Marshall. Vol. 9906. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 99065X. DOI: `10.1117/12.2234286`. arXiv: `1607.02059 [astro-ph.IM]`.

Noriega, Hernán E. et al. (Nov. 2022). "Fast computation of non-linear power spectrum in cosmologies with massive neutrinos". In: *J. Cosmology Astropart. Phys.* 2022.11, 038, p. 038. DOI: `10.1088/1475-7516/2022/11/038`. arXiv: `2208.02791 [astro-ph.CO]`.

Padmanabhan, T. (July 2003). "Cosmological constant-the weight of the vacuum". In: *Phys. Rep.* 380.5-6, pp. 235–320. DOI: `10.1016/S0370-1573(03)00120-0`. arXiv: `hep-th/0212290 [hep-th]`.

Padmanabhan, Thanu (2003). "Cosmological constant—the weight of the vacuum". In: *Physics Reports* 380.5-6, pp. 235–320.

Pan, ShuYang et al. (Sept. 2020). "Cosmological parameter estimation from large-scale structure deep learning". In: *Science China Physics, Mechanics, and Astronomy* 63.11, 110412, p. 110412. DOI: `10.1007/s11433-020-1586-3`. arXiv: `1908.10590 [astro-ph.CO]`.

Paszke, Adam et al. (2019). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., pp. 8024–8035. URL: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

Peel, Austin et al. (July 2019). "Distinguishing standard and modified gravity cosmologies with machine learning". In: *Phys. Rev. D* 100.2, 023508, p. 023508. DOI: `10.1103/PhysRevD.100.023508`. arXiv: `1810.11030 [astro-ph.CO]`.

Percival, Will J. (2005). "Cosmological structure formation in a homogeneous dark energy background". In: *Astron. Astrophys.* 443, p. 819. DOI: `10.1051/0004-6361:20053637`. arXiv: `astro-ph/0508156`.

Planck Collaboration et al. (Sept. 2020). "Planck 2018 results. VI. Cosmological parameters". In: *A&A* 641, A6, A6. DOI: `10.1051/0004-6361/201833910`. arXiv: `1807.06209 [astro-ph.CO]`.

Preskill, John, Mark B Wise, and Frank Wilczek (1983). "Cosmology of the invisible axion". In: *Physics Letters B* 120.1-3, pp. 127–132.

Pritchard, J. et al. (Apr. 2015). "Cosmology from EoR/Cosmic Dawn with the SKA". In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, p. 12. arXiv: `1501.04291 [astro-ph.CO]`.

Raghu, Maithra et al. (Aug. 2021). "Do Vision Transformers See Like Convolutional Neural Networks?" In: *arXiv e-prints*, arXiv:2108.08810, arXiv:2108.08810. DOI: `10.48550/arXiv.2108.08810`. arXiv: `2108.08810 [cs.CV]`.

Ramachandra, Nesar et al. (2021). "Matter Power Spectrum Emulator for f(R) Modified Gravity Cosmologies". In: *Phys. Rev. D* 103.12, p. 123525. DOI: `10.1103/PhysRevD.103.123525`. arXiv: `2010.00596 [astro-ph.CO]`.

Reddi, Sashank J., Satyen Kale, and Sanjiv Kumar (Apr. 2019). "On the Convergence of Adam and Beyond". In: *arXiv e-prints*, arXiv:1904.09237, arXiv:1904.09237. arXiv: `1904.09237 [cs.LG]`.

Reid, Beth A. et al. (Oct. 2014). "A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies". In: *MNRAS* 444.1, pp. 476–502. DOI: `10.1093/mnras/stu1391`. arXiv: `1404.3742 [astro-ph.CO]`.

Ribli, Dezső, Bálint Ármin Pataki, and István Csabai (Jan. 2019). "An improved cosmological parameter inference scheme motivated by deep learning". In: *Nature Astronomy* 3, pp. 93–98. DOI: `10.1038/s41550-018-0596-8`. arXiv: `1806.05995 [astro-ph.CO]`.

Ribli, Dezső et al. (Dec. 2019). "Weak lensing cosmology with convolutional neural networks on noisy data". In: *MNRAS* 490.2, pp. 1843–1860. DOI: `10.1093/mnras/stz2610`. arXiv: `1902.03663 [astro-ph.CO]`.

Rosenblatt, Frank (1958). "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6, p. 386.

Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams (1986). "Learning representations by back-propagating errors". In: *nature* 323.6088, pp. 533–536.

Santos, Mario et al. (2015). "Cosmology from a SKA HI intensity mapping survey". In: *PoS* AASKA14, p. 019. DOI: `10.22323/1.215.0019`.

Scoccimarro, Roman (Oct. 1998). "Transients from initial conditions: a perturbative analysis". In: *MNRAS* 299.4, pp. 1097–1118. DOI: `10.1046/j.1365-8711.1998.01845.x`. arXiv: `astro-ph/9711187 [astro-ph]`.

Scoccimarro, Roman and Joshua Frieman (1996). "Loop corrections in nonlinear cosmological perturbation theory". In: *Astrophys. J. Suppl.* 105, p. 37. DOI: `10.1086/192306`. arXiv: `astro-ph/9509047`.

Seo, Hee-Jong and Daniel J. Eisenstein (2003). "Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys". In: *Astrophys. J.* 598, pp. 720–740. DOI: `10.1086/379122`. arXiv: `astro-ph/0307460`.

Shimizu, Ikkoh et al. (Jan. 2019). "Osaka feedback model: isolated disc galaxy simulations". In: *Monthly Notices of the Royal Astronomical Society* 484.2, pp. 2632–2655. ISSN: 0035-8711. DOI: `10.1093/mnras/stz098`. eprint: `https://academic.oup.com/mnras/article-pdf/484/2/2632/27662451/stz098.pdf`. URL: `https://doi.org/10.1093/mnras/stz098`.

Silveira, V. and I. Waga (1994). "Decaying Lambda cosmologies and power spectrum". In: *Phys. Rev. D* 50, pp. 4890–4894. DOI: `10.1103/PhysRevD.50.4890`.

Smirnov, Nikolai V (1939). "Estimate of deviation between empirical distribution functions in two independent samples". In: *Bulletin Moscow University* 2.2, pp. 3–16.

Smith, Britton D. et al. (Dec. 2016). *Grackle: Chemistry and radiative cooling library for astrophysical simulations*. Astrophysics Source Code Library, record ascl:1612.020. ascl: `1612.020`.

Soares, Paula S. et al. (Mar. 2022). “Gaussian Process Regression for foreground removal in H I Intensity Mapping experiments”. In: *MNRAS* 510.4, pp. 5872–5890. DOI: `10.1093/mnras/stab2594`. arXiv: `2105.12665 [astro-ph.CO]`.

Song, Yong-Seon and Will J Percival (2009). “Reconstructing the history of structure formation using Redshift Distortions”. In: *Journal of Cosmology and Astroparticle Physics* 2009.10, p. 004.

Spinelli, Marta et al. (Jan. 2022). “SKAO H I intensity mapping: blind foreground subtraction challenge”. In: *MNRAS* 509.2, pp. 2048–2074. DOI: `10.1093/mnras/stab3064`. arXiv: `2107.10814 [astro-ph.CO]`.

Springel, Volker (Dec. 2005). “The cosmological simulation code GADGET-2”. In: *MNRAS* 364.4, pp. 1105–1134. DOI: `10.1111/j.1365-2966.2005.09655.x`. arXiv: `astro-ph/0505010 [astro-ph]`.

Tegmark, Max (1997). “Measuring cosmological parameters with galaxy surveys”. In: *Phys. Rev. Lett.* 79, pp. 3806–3809. DOI: `10.1103/PhysRevLett.79.3806`. arXiv: `astro-ph/9706198`.

Tingay, S. J. et al. (2013). “The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies”. In: *Publications of the Astronomical Society of Australia* 30, e007. DOI: `10.1017/pasa.2012.007`.

Tsujikawa, Shinji (2007). “Matter density perturbations and effective gravitational constant in modified gravity models of dark energy”. In: *Physical Review D* 76.2, p. 023514.

Tuominen, Kimmo (2021). “Cold Particle Dark Matter”. In: *Symmetry* 13.10, p. 1945.

Viel, Matteo et al. (Aug. 2013). “Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data”. In: *Phys. Rev. D* 88.4, 043502, p. 043502. DOI: `10.1103/PhysRevD.88.043502`. arXiv: `1306.2314 [astro-ph.CO]`.

Villaescusa-Navarro, Francisco et al. (Mar. 2015). “Cross-correlating 21cm intensity maps with Lyman Break Galaxies in the post-reionization era”. In: *J. Cosmology Astropart. Phys.* 2015.3, 034, p. 034. DOI: `10.1088/1475-7516/2015/03/034`. arXiv: `1410.7393 [astro-ph.CO]`.

Villaescusa-Navarro, Francisco et al. (2020). “The quijote simulations”. In: *The Astrophysical Journal Supplement Series* 250.1, p. 2.

Villaescusa-Navarro, Francisco et al. (July 2021). “The CAMELS Project: Cosmology and Astrophysics with Machine-learning Simulations”. In: *ApJ* 915.1, 71, p. 71. DOI: `10.3847/1538-4357/abf7ba`. arXiv: `2010.00619 [astro-ph.CO]`.

Villaescusa-Navarro, Francisco et al. (2022). “The camels multifield data set: Learning the universe’ s fundamental parameters with artificial intelligence”. In: *The Astrophysical Journal Supplement Series* 259.2, p. 61.

Villanueva-Domingo, Pablo and Francisco Villaescusa-Navarro (2021). “Removing Astrophysics in 21 cm Maps with Neural Networks”. In: *The Astrophysical Journal* 907.1, p. 44. DOI: `10.3847/1538-4357/abd245`. URL: `https://dx.doi.org/10.3847/1538-4357/abd245`.

Villasenor, Bruno et al. (July 2023). “New constraints on warm dark matter from the Lyman-α forest power spectrum”. In: *Phys. Rev. D* 108.2, 023502, p. 023502. DOI: `10.1103/PhysRevD.108.023502`. arXiv: `2209.14220 [astro-ph.CO]`.

Virtanen, Pauli et al. (2020). “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17, pp. 261–272. DOI: `10.1038/s41592-019-0686-2`.

Wang, Yun, Chia-Hsun Chuang, and Christopher M Hirata (2013). “Towards more realistic forecasting of dark energy constraints from galaxy redshift surveys”. In: *Monthly Notices of the Royal Astronomical Society* 430.3, pp. 2446–2453.

Weinberg, David H. et al. (Oct. 2015). “Cold dark matter: Controversies on small scales”. In: *Proceedings of the National Academy of Science* 112.40, pp. 12249–12255. DOI: `10.1073/pnas.1308716112`. arXiv: `1306.0913 [astro-ph.CO]`.

Wetterich, C. (1988). “Cosmology and the Fate of Dilatation Symmetry”. In: *Nucl. Phys. B* 302, pp. 668–696. DOI: `10.1016/0550-3213(88)90193-9`. arXiv: `1711.03844 [hep-th]`.

White, Simon D. M. (Oct. 1994). “Formation and Evolution of Galaxies: Les Houches Lectures”. In: *arXiv e-prints*, astro-ph/9410043, astro–ph/9410043. DOI: `10.48550/arXiv.astro-ph/9410043`. arXiv: `astro-ph/9410043 [astro-ph]`.

Winther, Hans et al. (2019). “Emulators for the nonlinear matter power spectrum beyond ΛCDM”. In: *Phys. Rev. D* 100.12, p. 123540. DOI: `10.1103/PhysRevD.100.123540`. arXiv: `1903.08798 [astro-ph.CO]`.

Yahia-Cherif, S. et al. (2021). “Validating the Fisher approach for stage IV spectroscopic surveys”. In: *Astron. Astrophys.* 649, A52. DOI: `10.1051/0004-6361/201937312`. arXiv: `2007.01812 [astro-ph.CO]`.

Yuan, Sihan et al. (2022). “Stringent $\sigma 8$ constraints from small-scale galaxy clustering using a hybrid MCMC+ emulator framework”. In: *Monthly Notices of the Royal Astronomical Society* 515.1, pp. 871–896.

Zhai, Zhongxu et al. (May 2023). “The Aemulus Project. V. Cosmological Constraint from Small-scale Clustering of BOSS Galaxies”. In: *ApJ* 948.2, 99, p. 99. DOI: `10.3847/1538-4357/acc65b`. arXiv: `2203.08999 [astro-ph.CO]`.

Zwicky, F. (Oct. 1937). “On the Masses of Nebulae and of Clusters of Nebulae”. In: *ApJ* 86, p. 217. DOI: `10.1086/143864`.