Doctoral Dissertation

博士論文

Analysis of the Large-Scale Structure of the Universe Using Cosmological Simulations and Machine Learning

（宇宙論的シミュレーションと機械学習を用いた宇宙大規模構造の解析）

A Dissertation Submitted for the Degree of Doctor of Philosophy

December 2021

令和 3 年 12 月 博士（理学）申請

Department of Physics, Graduate School of Science,
The University of Tokyo
東京大学大学院 理学系研究科物理学専攻

Kana Moriwaki
森脇 可奈

Abstract

Various astronomical observations have revealed the non-uniform distributions of matter and radiation of the universe. Such distribution is called the large-scale structure (LSS) of the universe. The LSS is formed via gravitational evolution from tiny fluctuations that were generated in the early universe 13.8 billion years ago. It contains information on the initial state, expansion history, and contents of the universe. Galaxy surveys have played an important role in revealing the LSS in the present universe. Current observational results are explained by a model where we assume the cosmological constant (Λ) and cold dark matter (CDM), the so-called ΛCDM model.

Several unresolved issues in cosmology can be studied by probing the LSSs in more detail than ever before. The issues include the properties of dark energy and dark matter, the mechanism that generates the initial state of the universe, and the properties of cosmic neutrinos. Observations of the LSS of the distant universe will also allow us to study galaxy formation and evolution and other astronomical phenomena, such as the cosmic reionization of the intergalactic medium (IGM) that occurred when the universe was a few billion years old.

An emerging observational method, line intensity mapping (LIM) will allow us to probe large-scale distributions of galaxies and IGM in greater volumes and greater distances than has been done before. In LIM observations, spectroscopic observations are conducted over large fields of the sky with lower angular/spectral resolutions. Instead of detecting individual emission sources, we aim at detecting the intensity fluctuations of spectral line emissions, which trace three-dimensional distributions of the line emitters. The LIM technique will allow us to survey large volumes efficiently and provide precise constraints on cosmology and astrophysics. A variety of future LIM observations are currently planned and proposed.

In LIM observations, there are serious problems of large observational noises and contaminantion including line interlopers that cannot be easily removed. While several methods (e.g., cross-correlation analysis) have been proposed to statistically extract line emission signals at a specific redshift, these methods abandon phase information of line intensity signals, which are valuable for studying cosmology and astrophysics.

In this thesis, we propose, for the first time, to use machine learning to perform pixel-by-pixel extraction of signals at specific redshifts from LIM observational data. We perform a demonstration of our methods assuming future near-infrared LIM observations such as SPHEREx and CDIM. These observations will probe the LSSs from intermediate redshifts to the epoch of reionization traced by rest-frame optical and ultra-violet emission lines including Hα and [OIII]5007Å. We use convolutional neural networks (CNNs) to deal with the line intensity maps. We build conditional generative adversarial networks (cGANs), which are known to be one of the best image-to-image translation models, and train them on signal extraction tasks in several conditions.

We first devise a method to remove the observational noises from LIM observational data. We generate mock observational data using a cosmological simulation and train cGANs to

extract line emission signals from noisy observational maps. The trained networks reconstruct the line emission signals properly although the observed data is noise-dominated. We find that the reproducibility depends on the size and resolution of the training data. When we adopt the original angular resolution of SPHEREx of 0.1 arcmin, the locations of bright point sources are accurately reconstructed, but the faint extended signals are not reproduced well. When we train the network with maps of a coarser resolution of 2 arcmin, the faint signal distributions and large-scale power spectrum are reproduced.

Next, we consider the line confusion problem. Assume that there are two emission lines with rest-frame wavelengths λ_1 and λ_2. Emissions of line 1 from redshift z_1 are observed at the wavelength $\lambda_{\rm obs} = \lambda_1(1 + z_1)$. The observed signals at this wavelength are also contaminated by line 2 from redshift $z_2 = \lambda_{\rm obs}/\lambda_2 - 1$. The contributions of individual lines cannot be easily separated because of low resolutions of LIM observations. The line confusion problem is one of the most serious problems in the LIM observations. We construct a cGAN to separate two emission signals originating from different elements, for example, the Hα line from $z = 1.3$ and the [OIII] line from $z = 2.0$. We find the LSSs traced by two signals are clearly separated by our cGAN. The peak positions of the intensities are detected with 76 and 32 percent precision for Hα and [OIII]. Statistics such as the power spectra, probability distribution functions, and mean intensities are also properly reproduced. By looking at the convolutional filters, we find that our cGAN extracts characteristic large-scale features from the observed images. This indicates that our cGAN learns the differences in the LSSs at each redshift to perform the separation.

In the above two experiments, we only deal with two-dimensional LIM data at a single observed wavelength. In actual LIM observations, spectral information is also available. We develop a cGAN that can be applied to the three-dimensional observational data cubes. We construct a network architecture that provides the machine with physical information on the wavelengths of individual emission lines. We find that our physics-informed cGAN can extract the bright peaks of individual emission lines with 84 and 68 percent precision. We investigate how the machine extracts the signals by performing various methods and find that the cGAN may focus on the co-existence of multiple emission line signals in the spectral direction.

The reconstructed intensity maps trace the LSS of the universe and also provide information on the galaxy population at high redshifts. They can be used in various studies on cosmology and astrophysics. As one of these applications, we investigate using the line intensity maps to study the epoch of reionization. In particular, we predict the cross-correlation signals between the 21-cm line emissions from neutral hydrogen in the IGM and [OIII] line emissions from galaxies at the epoch of reionization. We find that the large-scale cross-power spectrum can be detected with future LIM observations if we can properly extract foreground interlopers with methods like ours. The observed signals depend on the reionization process and can be used to reveal the cosmic reionization history.

The machine learning methods we devised in this thesis enable pixel-by-pixel extraction of the LIM signals at a particular redshift. Such methods are crucial to fully utilize the huge amount of observational data obtained in future observations. When our methods are applied to future observational data over unprecedented scales, we can robustly detect faint LSS signals and solve various issues in cosmology and astrophysics.

Contents

Chapter 1

Introduction

Observations of the cosmic microwave background (CMB) have revealed the existence of tiny fluctuations in the early universe. These fluctuations grow with time via gravitational interaction and form the large-scale structure (LSS) observed in the present universe. The LSS contains information on the initial state and evolutionary history of the universe. Current observations can be explained by the cold dark matter (CDM) model with a cosmological constant (Λ), the so-called ΛCDM model.

The LSS of the universe can be measured by observing the distributions of galaxies as well as the CMB, weak lensing signals, and Lyα absorption in the spectra of distant quasars. Future galaxy surveys will provide LSSs of the universe much farther and wider than we have reached so far. They will allow us to study several important issues of cosmology and astrophysics. It is crucial to predict the observational signals and to prepare methods to analyze future observational data. For this purpose, cosmological simulations will absolutely play an important role, and we can also use machine learning techniques.

Scientific Goals

Future LSS observations allow us to study the following:

- **Dark energy** The cosmic expansion is considered to be caused by dark energy. The energy conservation law of the spatially homogeneous and isotropic universe is given by

  $$\dot{\rho} + 3\frac{\dot{a}}{a}\left[\rho + \frac{p}{c^2}\right] = 0, \tag{1.1}$$

  where a is the scale factor of the universe, ρ is the density, p is the pressure, and c is the speed of light. The equation of state of the dark energy is represented by

  $$p_d = w\rho_d c^2. \tag{1.2}$$

  It explains the accelerated expansion of the universe (i.e., $\ddot{a} > 0$) when $w < -1/3$ and corresponds to the cosmological constant Λ when $w = -1$. One of the major goals of modern cosmology is to determine the evolution of the equation of state $w(z)$. We can explore it by observing the expansion history of the universe. The expansion of the universe is denoted by the Hubble parameter

$$H(z) \equiv \frac{\dot{a}}{a} = H_0 \sqrt{\frac{\Omega_r}{a^4} + \frac{\Omega_m}{a^3} - \frac{K}{H_0^2 a^2} + \Omega_d \exp\left[3\int_a^{a_0} (1+w)\frac{da}{a}\right]}, \tag{1.3}$$

where H_0 is the Hubble parameter at the present universe, Ω_r, Ω_m, and Ω_d are the energy ratios of radiation, matter, and dark energy to the total energy in the present universe, and K is the curvature. Current observational results are consistent with $w = -1$ (Planck Collaboration XIV 2016), but various models suggest that deviation from it may be occurring at $z \gtrsim 2$, where we have yet to fully explore (e.g., Raveri et al. 2017).

- **Gravity theory** The cosmic expansion may also be explained by a modified gravity theory instead of assuming mysterious dark energy. Since general relativity has been tested very well on small scales by the observations of the Solar system and pulsars, we generally consider theories in which there is (some) deviation only at large scales. Large-scale observations are thus essential to test these theories.

- **Mechanism that generates initial state** One of the most plausible models of the early universe is the inflation model, in which space expands exponentially during a certain period in the early universe. While the simplest inflation models predict initial density fluctuations that are described by a Gaussian random field, some models predict non-Gaussian distribution. To test these models, it is important to constrain the primordial non-Gaussianity $f_{\rm NL}$. Currently, the best constraints are given by the bispectrum of the CMB (e.g., $f_{\rm NL} = -0.9 \pm 5.1$ for local non-Gaussianity; Planck Collaboration IX 2020). The primordial non-Gaussianity can also be probed by observing the present LSS on ultra-large scales that are not subject to nonlinear effects.

- **Neutrino mass** Observations of neutrino oscillations have shown that neutrinos are massive. Astronomical observations put stronger upper limits on the total mass than particle experiments. The massive neutrinos suppress the structure formation on small scales and thus can be studied with the LSS of the unvierse.

- **Cosmic reionization and heating** Various observations suggest that the intergalactic medium (IGM) extending over the entire universe becomes ionized at $z \sim 8$. This epoch is referred to as the epoch of reionization (EoR). The IGM heating is also considered to precede the cosmic reionization. Details of these phenomena are still unclear, and they are among the frontiers in cosmology and astrophysics.

Cosmological Simulation

For a given cosmological model, we can numerically simulate the structure formation of the universe. Cosmological simulation is an important method to connect the observations and theories on the LSS. As the cosmological parameters have been determined with good accuracy, cosmological simulations can be used for predicting the observational data.

In cosmological simulations, we simulate the distribution and physical properties of matter by numerically solving physical processes from the initial state of the universe. It allows us to follow nonlinear evolution, which is difficult to handle analytically. Since it is practically impossible to resolve all matter into individual elementary particles in the simulations, we replace them with particles of certain masses or divide them into grids of finite size. In the simplest case (e.g., N-body simulations), we only consider the dark matter and solve its gravitational growth. We can simulate the matter distribution of a very large volume with such simulations.

To follow the galaxy formation, we need to include baryonic physics. In cosmological hydrodynamics simulations, we follow the evolution of the dark matter and baryons by solving both gravity and hydrodynamics. Hydrodynamics simulations are costly compared to the dark-matter-only simulations. Thus we adopt smaller spatial/mass resolutions and smaller volumes. A typical mass resolution of the recent cosmological hydrodynamics simulations is $10^4 - 10^6$ $\mathrm{M}_\odot$, which is not enough to resolve the molecular cloud where stars form. For this reason, we use subgrid recipes to model the physical processes such as gas cooling and heating, star formation, feedback from the supernova explosions and active galactic nuclei (AGN), metal yield from stars, and so on. The free parameters of the subgrid models are chosen so that the simulation results fit various observational results including statistical observables (e.g., star formation history, stellar mass function, stellar-to-halo mass ratio, mass-metallicity relation, galaxy size) and other individual observables (e.g., Kennicutt-Schmidt law, morphology and radial profile of individual galaxies, phase-space distribution of gas).

So far, it has been difficult to resolve the internal structures of galaxies in a simulation of cosmological volume. Recently, large cosmological hydrodynamics simulation projects such as EAGLE (Schaye et al. 2015), Illustris (Vogelsberger et al. 2014), and IllustrisTNG (Pillepich et al. 2018) have made it possible. These simulations resolve individual galaxies even for large box sizes ~ 100 Mpc. The simulation outputs obtained from these projects are open access in most cases, and anyone can use them for various kinds of studies.

Machine Learning Techniques

Data analysis methods using machine learning have been attracting attention in various fields. The machines learn features of training data (e.g., complex images) and perform automated processes (e.g., classification). With the advent of graphics processing units (GPUs), complex machine learning models such as convolutional neural networks have been under intense investigation. Machine learning models are used for many kinds of tasks such as classification, clustering analysis, regression, and generation. In astrophysics and cosmology, machine learning methods have been widely used for various tasks as reviewed by, e.g., Ball & Brunner (2010), Ntampaka et al. (2019), and Fluke & Jacobs (2020).

One of the biggest advantages of using machine learning is that it can process large amounts of data at high speed. Machine learning can be used to automate the detection and classification

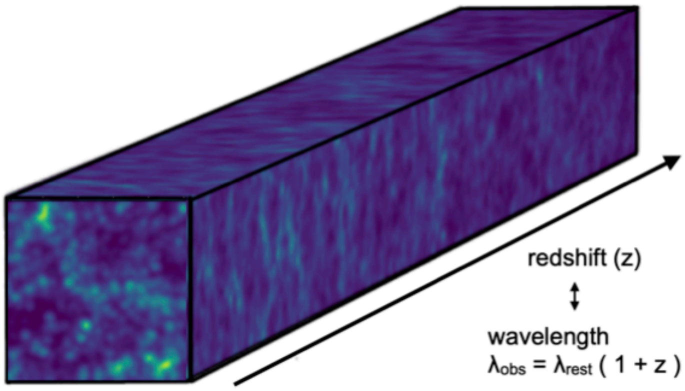


Figure 1.1: Line intensity mapping measures the fluctuations of the line emission with low spatial and spectral resolution. If only one emission line with rest-frame wavelength $\lambda_{\rm rest}$ contributes to the fluctuations, we can reconstruct the three-dimensional distribution of line emitters by converting the observed wavelengths to redshifts as $\lambda_{\rm obs} = \lambda_{\rm rest}(1+z)$.

of galaxies, stars, and transients obtained from surveys (e.g., Ackermann et al. 2018; Cui et al. 2021; du Buisson et al. 2015; Kojima et al. 2020; Ntampaka et al. 2015; Pearson et al. 2019). Machines can also discover unexpected rare sources such as galaxy-galaxy gravitational lenses and extreme emission/absorption line sources (Baron & Poznanski 2017). Such usage of machine learning will be more crucial to analyze the huge amount of data from future observations.

Another advantage is that machines can capture complex structures, such as the LSS in the present universe. Gupta et al. (2018) have used a machine learning model to derive cosmological information from the LSS observed in weak lensing observations and shown that it could outperform the conventional analytical methods. Other studies also demonstrate that we can estimate cosmological and astrophysical parameters from LSSs (e.g. Hong et al. 2021; Ravanbakhsh et al. 2017), CMB (e.g., Caldeira et al. 2019) and 21-cm line intensity maps at the EoR (e.g., Hassan et al. 2019; Shimabukuro & Semelin 2017), and clean noisy observational data (e.g., Shirasaki et al. 2021) using machine learning.

A general issue of machine learning is that it is like a black box. We cannot easily understand how a machine deals with the data. Thus even if errors and biases exist in its output, we cannot easily recognize them. Recently, several methods have been proposed to understand the machine learning processes (e.g., Erhan et al. 2009; Kendall & Gal 2017; Papernot & McDaniel 2018; Simonyan et al. 2013), and they are being adopted in astrophysics problems as well (e.g., Acquaviva et al. 2020; Matilla et al. 2020; Petroff et al. 2020; Villanueva-Domingo & Villaescusa-Navarro 2021).

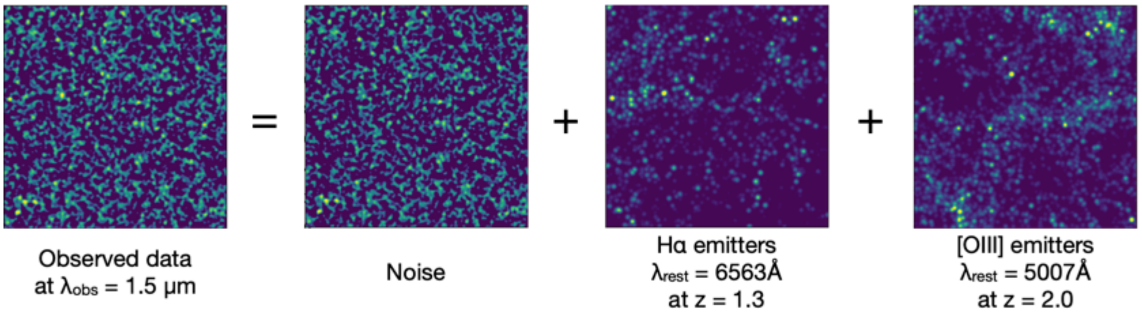


Figure 1.2: The line intensity mapping suffers from large observational noise and line confusion. Several emission lines from different redshifts contribute to the same observed wavelength. For example, both Hα 6563Å from $z = 1.3$ and [OIII] 5007Å from $z = 2.0$ are observed at 1.5 μm.

Aims of This Thesis

In this study, we use cosmological simulations to predict the observational signals in future observations and devise analysis methods using machine learning techniques. We review the observations of the LSS of the universe in Chapter 2 and describe our line emission model used in this thesis in Chapter 3. In particular, we consider two types of emission lines with rest-frame optical wavelengths: Hα and [OIII].

We focus on an emerging method of observing LSSs, line intensity mapping (LIM), in which we perform a spectroscopic observation to measure the fluctuations of the line emissions over a large area with low spatial/spectral resolutions. From the measured fluctuation, three-dimensional distribution of line emitters can be reconstructed by converting the observed wavelengths to the redshifts using $\lambda_{\rm obs} = \lambda_{\rm rest}(1 + z)$ (Fig.1.1). While LIM is a promising method to survey large volumes very efficiently, there are several problems. If there are several bright emission lines, distributions at different redshifts contaminate (Fig.1.2). They are difficult to separate because of the low sensitivity and resolution. Large observational noises are also problematic in the LIM observations.

In Chapters 4 - 6, we develop machine learning networks to extract LSS signals from LIM observation data. The experiments are carried out in stages (Fig.1.3). First, (i) we extract signals from noisy observational data. Then, (ii) we separate multiple emission lines from noiseless data. Finally, (iii) we investigate the separation of multiple emission lines using three-dimensional data including spectral information. For the first two experiments, we use two-dimensional data at a single wavelength bin.

(i) Chapter 4 (2D): signal (Hα) + noise
(ii) Chapter 5 (2D): signals (Hα, [OIII])
(iii) Chapter 6 (3D): signals (Hα, [OIII]) + noise

We build our machine learning models using Tensorflow (Abadi et al. 2016). All the experiments are performed on a single NVIDIA Titan RTX GPU.

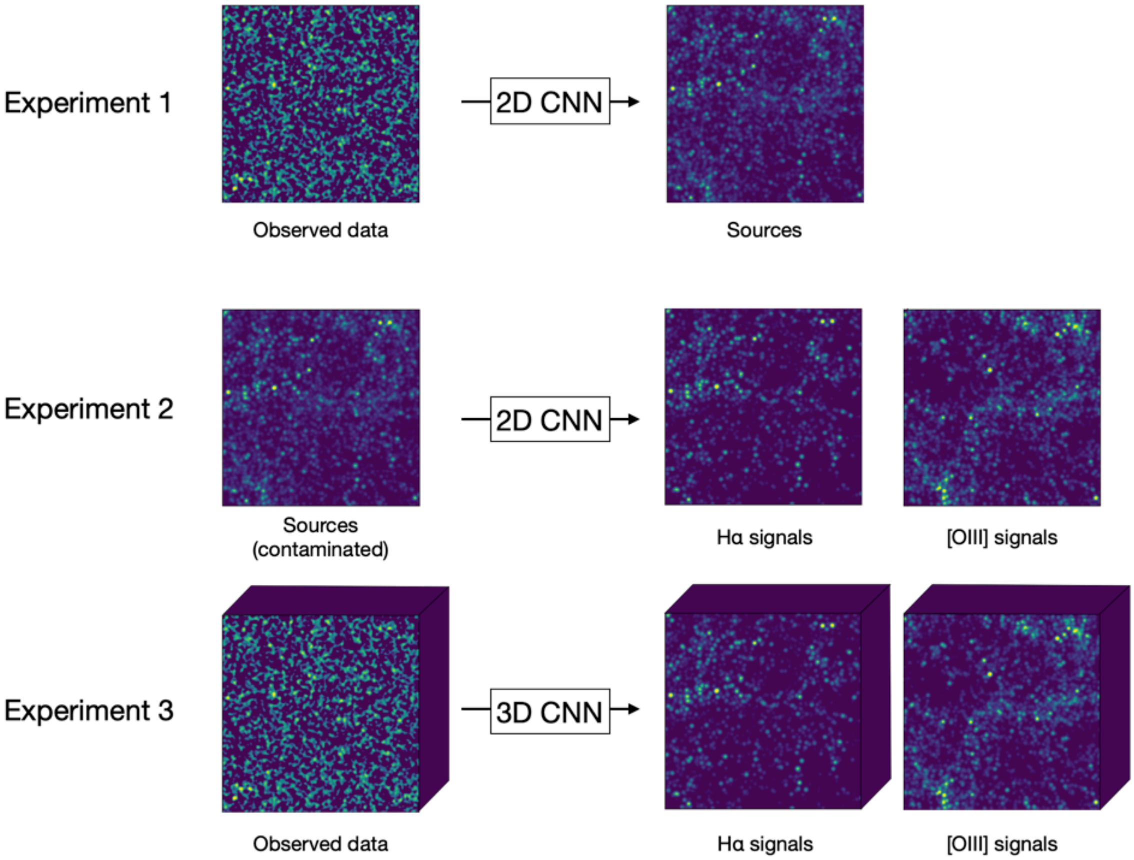


Figure 1.3: The experiments are carried out in stages. In experiment 1, we extract signals from noisy observational data. In experiment 2, we separate multiple emission lines from noiseless data. In experiment 3, we separate multiple emission lines in the presence of noise. In experiments 1 and 2, we use two-dimensional data observed at a single wavelength slice, and in experiment 3, we use three-dimensional data including spectral information.

The extracted LSSs can be used for various studies in cosmology and astrophysics. As one of the applications, we consider using the line intensity maps to study the EoR in Chapter 7. We predict the signals obtained from the combination of the 21-cm and [O III] LIM observations, and investigate what we can learn from future observations. In this thesis, we adopt the ΛCDM model with $\Omega_m = 0.316$, $\Omega_\Lambda = 0.684$, $\Omega_b = 0.044$, $h = 0.673$, and $\sigma_8 = 0.8$ unless otherwise noted.

Chapter 2

Observations of the Large-Scale Structure of the Universe

In this chapter, we review the observations of the large-scale structure. In Section 2.1, we discuss the characteristics of the observed LSSs, and overview the past and future observations of the LSS in Section 2.2. Cosmic reionization can also be studied by observing the LSS in the distant universe. In Section 2.3, we summarize the observations of cosmic reionization.

2.1 Large-Scale Structure of the Universe

We consider a matter density $\rho(\boldsymbol{x})$ and its fluctuation

$$\delta(\boldsymbol{x}) = \rho(\boldsymbol{x})/\overline{\rho} - 1, \tag{2.1}$$

where $\overline{\rho}$ is the mean matter density. In the inflation model, the density fluctuations are generated by quantum fluctuations and the actual value of the fluctuation at each point is not predictable. Therefore, statistics such as the power spectrum are used to compare the observed data with the theory. The power spectrum $P(k)$ is defined by

$$\langle \tilde{\delta}(\boldsymbol{k})\tilde{\delta}(\boldsymbol{k}')\rangle = (2\pi)^3 \delta_{\mathrm{D}}^3(\boldsymbol{k}+\boldsymbol{k}')P(k), \tag{2.2}$$

where

$$\tilde{\delta}(\boldsymbol{k}) = \int d^3x \delta(\boldsymbol{x}) \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{2.3}$$

is the Fourier transform of the density fluctuations, and δ_{D}^3 is the Dirac delta function in three-dimensional space. In galaxy surveys, the correlation function is obtained by counting the number of galaxy pairs that are separated by a certain distance. The correlation function is the Fourier transform of the power spectrum

$$\xi(r) = \int \frac{d^3k}{(2\pi)^3}\, \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}} P(k) \tag{2.4}$$

$$P(k) = \int d^3x\, \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \xi(r) \tag{2.5}$$

and provides the same information as the power spectrum.

In the early universe, the fluctuation grows linearly, and the power spectrum on a scale k at time t is given by

$$P(k,t) \propto D^2(t)T^2(k,t)P_{\rm in}(k), \tag{2.6}$$

where $P_{\rm in}(k)$ is the power spectrum in the initial state, $D(t)$ is the growth function, and $T(k,t)$ is the transfer function. In simple inflation models, the initial power spectrum follows the power law $P_{\rm in} \propto k^n$. Its amplitude is not determined theoretically and has to be determined observationally, so we adopt the variance of the matter fluctuations in the present universe averaged over a sphere of radius $8h^{-1}$ Mpc, σ_8^2, as a free parameter. The growth function $D(t)$ depends on the energy budget of the universe. In a matter-dominated universe, it is proportional to the scale factor, $D(t) = a$, and it is suppressed in the presence of dark energy. The transfer function $T(k,t)$ describes how differently the fluctuation on each scale grows. For example, small-scale fluctuations that experience horizon crossing earlier than the matter-radiation equality become smaller than the other scales because their growth is suppressed during the radiation-dominated phase. Therefore, the observed power spectrum today is bent around the wavelength corresponding to the Hubble radius at the time of equality, $k_{\rm eq} = a_{\rm eq}H_{\rm eq}/c$.

As the fluctuations become large, they grow nonlinearly. At small scales, the fluctuations grow through gravitational instability, forming a gravitationally bound dark matter halo. The dark matter is then virialized and further contraction is prevented. Baryons, on the other hand, can be cooled via the interaction with photons and contract further. When the gas density becomes sufficiently high, stars are formed. The cloud of the bright stars are observed as galaxies. Galaxy surveys provide the distributions of the galaxies that satisfy certain criteria on the brightness, distance, etc. The galaxy distributions reflect the background matter distribution. On large scales, the fluctuation of the number density of galaxies δ_g is considered to be roughly proportional to the matter density fluctuation[1]

$$\tilde{\delta}_g(\boldsymbol{k}) = b\tilde{\delta}(\boldsymbol{k}), \quad P_g(\boldsymbol{k}) = b^2 P(\boldsymbol{k}). \tag{2.7}$$

The coefficient b is called the bias parameter, which is generally scale-independent on large scales. The bias parameter depends on the choice of galaxy sample. Since two scaling factors of the galaxy power spectrum, the bias b and σ_8, are essentially degenerate, the scale dependence of the power spectrum is more important than its amplitude.

Redshift Space Distortion

The distances of galaxies from us are measured with redshift. In the Friedmann-Robertson-Walker metric, there is a one-to-one correspondence between the redshift, $z = 1/a - 1$, and the

---

[1] Eq.2.7 is called 2-halo term. We also observe a clustering term between galaxies within a halo (1-halo term) and a shot noise due to the discrete distribution of the galaxies on small scales.

comoving distance

$$x(z) = \int_0^z \frac{c\,\mathrm{d}z'}{H(z')}. \tag{2.8}$$

In general, the redshift is measured from the deviation of the observed wavelength from the rest-frame wavelength

$$\lambda_{\mathrm{obs}} = \lambda_{\mathrm{rest}}(1 + z_{\mathrm{s}}). \tag{2.9}$$

If the galaxy has a proper velocity v along the line-of-sight, the Doppler shift affects the observed redshift as $z_{\mathrm{s}} = z + v/ca$ and the derived comoving distance as $x_s = x + v/aH$. The index s denotes redshift space.

In a non-uniform universe, galaxies move toward overdense regions on average. Making a linear approximation, the relationship between the number density of galaxies in real space, $n_g(\boldsymbol{x})$, and that in redshift space, $n_s(\boldsymbol{x_s})$, is given by

$$n_s(\boldsymbol{x_s}) = \left[1 - \frac{\partial}{\partial x_3}\left(\frac{v}{aH}\right)\right] n_g(\boldsymbol{x}). \tag{2.10}$$

The density and velocity fields are related as $ikv \sim -\dot{\delta}$ in linear theory, and in the later universe, the density scales with the growth function D. We thus obtain Kaiser's formula (Kaiser 1987)

$$\tilde{\delta}_s(\boldsymbol{k}) = (b + f\mu^2)\tilde{\delta}(\boldsymbol{k}), \tag{2.11}$$

where $\mu = k_3/k$ is the cosine of the angle between $\boldsymbol{k}$ and the line of sight, and

$$f \equiv \frac{d \log D}{d \log a} \tag{2.12}$$

is the growth rate. We call $\beta \equiv f/b$ the redshift-space distortion (RSD) parameter. At $z \sim 0$, $f \sim \Omega_m^{0.6}$. The observed large-scale power spectrum becomes anisotropic

$$P_s(\boldsymbol{k}) = (b + f\mu^2)^2 P(k). \tag{2.13}$$

By determining β from the observations of the anisotropy and the bias from the other observations, we can put constraints on the cosmological parameters.

Baryonic Acoustic Ocsillation

Before photon decoupling, the baryon couples with the photon and behaves differently from dark matter. The fluid mixture of baryons and photons has a high pressure that pushes back the gravitational contraction force from the dark matter potential, resulting in acoustic oscillations. The oscillation starts when the fluctuation wavelength enters the horizon and continues until decoupling. The phase of oscillation at decoupling varies with wavelength because of the difference in the start time and duration of the oscillation. While the baryons catch up with the gravitational growth of dark matter after decoupling, the distribution of baryons in the present universe still shows a characteristic oscillation pattern. The scales at which the galaxy BAO signals are seen with respect to those of the CMB depend on the expansion history of the universe, and thus can be used to investigate the nature of dark energy.

Primordial Non-Gaussianity

The primordial non-Gaussianity can be traced by the LSS on ultra-large scales ($k < k_{\mathrm{eq}}$). It does not appear in the matter power spectrum but appears in other statistics such as the one-point probability distribution function and bispectrum. it also appears in the galaxy power spectrum as a scale-dependent bias. Dalal et al. (2008) show that the local non-Gaussianity skews the probability distribution of density fluctuations toward larger values, and produces more galaxies. In this case, the deviation of the bias from that of the Gaussian distribution is given by

$$\Delta b \propto \frac{f_{\mathrm{NL}}}{k^2}. \tag{2.14}$$

While it may be more difficult to obtain stronger constraints on f_{NL} from a single galaxy survey than from the CMB observations, we may be able to obtain competitive or more constraints by using multiple tracers (Seljak 2009).

Suppression due to Free Streaming

The presence of light elements such as cosmic neutrinos modifies the power spectrum amplitude. The temperature of the cosmic neutrino background decreases as the universe expands,

$$T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma = 195\ (1+z)\ \mathrm{K}, \tag{2.15}$$

where T_γ is the temperature of the CMB.[2] The neutrinos become non-relativistic when $3T_\nu \sim m_\nu$ $(1+z \sim 2000\ (m_\nu/1\ \mathrm{eV}))$. Non-relativistic neutrinos moving very fast with a thermal velocity $v \sim 3T_\nu/m_\nu$ stream out of the dense regions. Therefore, the matter fluctuations on scales smaller than the free-streaming scale, i.e.,

$$k > k_{\mathrm{fs}} = \sqrt{\frac{3}{2}} \frac{H(z)}{v(1+z)} \propto \frac{m_\nu H(z)}{(1+z)^2}, \tag{2.16}$$

are suppressed. Detecting such scale dependent suppression provides information on neutrinos. At $m_\nu \gg T_\nu$, energy density of neutrinos in the universe scales with the mass of the neutrinos, and the amplitude of the suppression depends on the neutrino mass.

2.2 Observations of Large-Scale Distribution of the Galaxies

2.2.1 Galaxy Surveys

In galaxy surveys, the positions of individual galaxies are measured. The simplest method is the imaging survey, which measures their locations on the celestial sphere and obtains the projected distribution along the line of sight. To determine the depth position of a galaxy, we

---

[2] The neutrinos decouple from the cosmic plasma before the electrons and positrons annihilate. Photons decouple after obtaining energy from the annihilation. Thus the temperature of the cosmic neutrino background is smaller than that of the CMB.

need to measure its spectral energy distribution (spectrum). In redshift surveys, we apply a spectrometer to the catalog galaxies and detect line emissions to measure their redshifts one by one (Eq.2.9). The first detection of a filamentary structure between clusters has been done by Gregory & Thompson (1978), who have done a redshift survey over a few hundred galaxies around the Coma cluster. A number of surveys have then followed. One of the largest redshift surveys in the early days is the CfA redshift survey (de Lapparent et al. 1986), which have targeted $\sim$ 10,000 galaxies at $z \lesssim 0.05$ and have provided their distribution over $\sim$ 100 Mpc. They have revealed the existence of the LSS in the universe, the bubble-like structure composed of voids and filaments. Later, the two-degree Field Galaxy Survey (2dFGRS; Colless et al. 2001) observed $\sim$ 220,000 galaxies at $z < 0.3$ and the Sloan Digital Sky Survey (SDSS; Park et al. 2005) observed $\sim$ 930,000 galaxies at $z < 0.5$. In these observations, BAO signals at around $100h^{-1}$ Mpc were detected for the first time in the galaxy power spectrum (Cole et al. 2005; Eisenstein et al. 2005). These surveys have also made an important contribution to the study of galaxy formation and evolution. For example, elliptical galaxies, in which star formation is already quenched, tend to be in filaments and clusters, while actively star-forming spiral galaxies tend to be in low-density regions. This strongly suggests that the environment has an effect on galaxy formation. Such environmental effects can also be important in constraining the cosmology as they would affect the bias parameter of a particular galaxy catalog.

High-redshift observations are generally difficult because of the smaller apparent brightness and apparent size of the distant galaxies. Several redshift surveys were conducted at $z \lesssim 1$ (Ellis et al. 1996; Lilly et al. 1995). In the Deep2 Redshift Survey (Davis et al. 2003), 50,000 galaxies over $\sim 1\mathrm{deg}^2$ at $z \sim 1$ were observed. These surveys revealed that the typical luminosity of the galaxy changes between $z = 1$ and 0. More recently, survey projects of SDSS, the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013) and Extended BOSS (eBOSS; Prakash et al. 2016; Raichoor et al. 2017), have provided the distribution of 1.3 million galaxies at $z < 0.6$ and 0.5 million galaxies at $z < 1.1$. The total effective volume of BOSS/eBOSS observations is as large as $\sim$ 10 Gpc^3. They are the largest redshift surveys currently available.

Future and ongoing galaxy surveys will allow us to study the LSSs at greater distances and/or over a larger area. Recently launched ground-based experiment Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016),[3] will observe $\sim$ 40 million galaxies at redshifts up to $z = 1.7$ covering $14,000$ deg^2 area. The European Space Agency survey mission Euclid (Laureijs et al. 2011)[4] and the Nancy Grace Roman Space Telescope (Spergel et al. 2015),[5] will conduct deeper observations and will provide the distributions of galaxies up to $z \sim 2$ including more faint galaxies than ever. They would be complementary to a large-area observation by DESI. At higher redshift, the Subaru Prime Focus Spectrograph (PFS; Takada et al. 2014)[6] will map galaxies up to $z = 2.4$, covering a larger comoving volume than the current largest survey. These surveys will constrain the cosmological parameters with greater

---

[3] `https://www.desi.lbl.gov`
[4] `https://www.euclid-ec.org`
[5] `https://roman.gsfc.nasa.gov`
[6] `https://pfs.ipmu.jp`

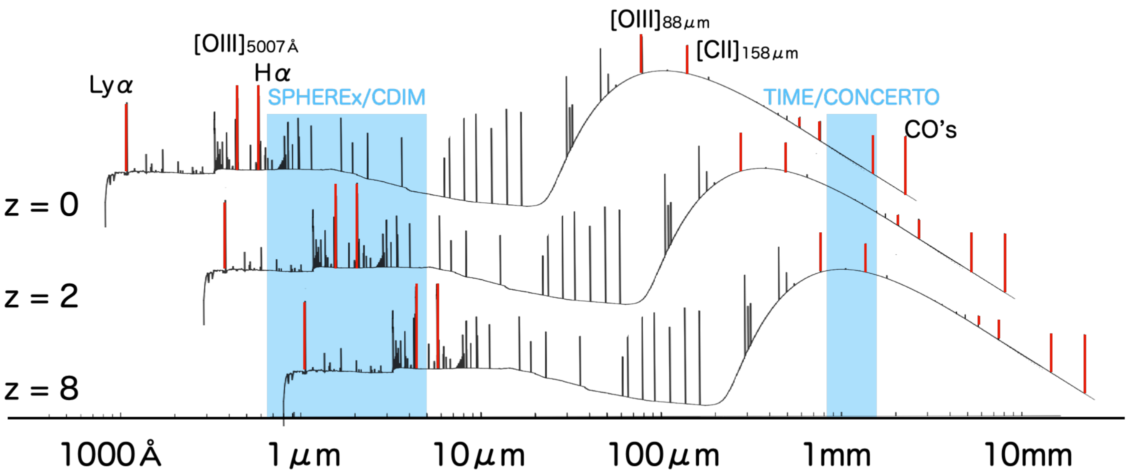


Figure 2.1: Examples of the spectra at $z = 0$, 2, 8. The vertical axis is log scale. Several bright emission lines are indicated by red. The blue shaded regions indicate the approximate coverage of the LIM observations by SPHEREx, CDIM, TIME, and CONCERTO. Intermediate redshifts can be probed by optical [OIII] and Hα lines and CO lines, and high redshifts can be probed by Lyα, optical [OIII], FIR [OIII] and [CII] lines.

precision than ever before.

The LSSs of an even more distant universe have also been partly revealed by, for example, the Subaru Telescope (Ouchi et al. 2005). They have performed narrowband imaging surveys to probe ~ 10 Mpc scale distributions of Lyα emitters at $z \sim 3, 5, 6$. Such survey data can be used for studying the EoR as well as galaxy formation and evolution.

2.2.2 Line Intensity Mapping

An emerging observational technique, line intensity mapping (LIM), plays a complementary role to the conventional galaxy redshift surveys. It performs a spectroscopic observation over wide area and provides the LSS traced by complete samples. In LIM observations, we do not resolve individual emission sources but measure the integrated emissions from galaxies and IGM. Both continuum and line emissions contribute to the observed intensities, but we can solely extract the fluctuations of line emissions (i.e., the three-dimensional galaxy distribution) by removing the smooth component from the observed spectrum. The low spatial and spectral resolutions allow us to observe large volumes efficiently. This method would be more important especially in observing LSSs in the distant universe, where individual galaxy signals are weaker.

In the following, we will focus on the LIM observations targeting emission lines from galaxies. Those from IGM at the EoR will be discussed in Section 2.3. Fig.2.1 shows an example of the galaxy spectrum. Several bright emission lines are highlighted: hydrogen recombination lines (Lyα, Hα) and forbidden lines of oxygen and carbon ions ([OIII], [CII]) from ionized regions around young massive stars and rotational transition lines of carbon-monoxide (CO) form molecular clouds. These lines trace the star forming activity and the properties of the interstellar medium (ISM) of the galaxy. They are commonly used to study individual high-redshift galaxies. We can also consider the 21-cm hyperfine transition line from neutral hydrogen in galaxies. The 21-cm line is intrinsically very weak, but the integrated signals can be observed with the LIM technique.

In LIM observations, we observe the integrated line intensity within each observational voxel

$$I_{\rm line} = \frac{\sum_i L^i_{\rm line}}{4\pi D_{\rm L}^2}\frac{1}{d\theta^2 d\nu}, \tag{2.17}$$

where $\sum_i L^i_{\rm line}$ is the total luminosity within the voxel, $D_{\rm L}$ is the luminosity distance, and $d\theta$ and $d\nu$ are the angular and spectral resolutions. As in the galaxy surveys, we compute statistics such as the power spectrum to study cosmology and mean properties of the galaxy populations. The power spectrum of the line intensity map is given by

$$P_I(\boldsymbol{k}) = \overline{I}^2_{\rm line}\left[(b + f\mu^2)^2 P(k) + \frac{U_{1-{\rm halo}}(k)}{\overline{n}^{\rm eff}_{\rm halo}} + \frac{1}{\overline{n}^{\rm eff}_{\rm g}}\right], \tag{2.18}$$

where $\overline{I}_{\rm line}$ is the mean line intensity.[7] As the line emitters are biased tracers of the large-scale matter distribution, the power spectrum is proportional to the matter power spectrum $P(k)$ on large scales (the first term of Eq.2.18). The second and third terms are called the 1-halo term and the shot noise, and

$$\overline{n}^{\rm eff}_{\rm halo} = \frac{\overline{L}^2_{\rm halo}}{\overline{L^2}_{\rm halo}}, \quad \overline{n}^{\rm eff}_{\rm g} = \frac{\overline{L}^2_{\rm g}}{\overline{L^2}_{\rm g}}. \tag{2.19}$$

are the effective number densities of halos and galaxies, where $L_{\rm halo}$ and $L_{\rm g}$ represent the total line luminosities of halos and galaxies. The shape of the 1-halo term depends on the distribution of galaxies within a halo, which is denoted by $U_{1-{\rm halo}}(k)$.

Much of the earliest work on the LIM observations comprised of detecting the 21-cm line from neutral hydrogen. The first detection of the LIM signal was made by Chang et al. (2010), who used the Green Bank Telescope (GBT) and detected the cross-correlation signal between 21-cm intensity and galaxies from the DEEP2 survey at $z \sim 0.8$. The 21-cm auto-power spectra have also been detected (e.g., Switzer et al. 2013). In recent years, other emission lines in different wavelength regimes are also becoming the target of the LIM observations. In millimeter wavelengths, the CO and [CII] lines are good targets. Keating et al. (2016) have reported a 3-σ detection of the CO(1-0) intensity power spectrum at $z \sim 3$ in the CO Power Spectrum Survey (COPSS), where they used the Sunyaev-Zel'dovich Array to survey a total area of 0.7 deg^2. The intensity power spectra of the other CO lines have also been detected at $z = 1.3$, 2.5, and 3.6 in Millimeter-wave Intensity Mapping Experiment (mmIME; Keating et al. 2020), which have used the Atacama Large Millimeter/submillimeter Array (ALMA) and Atacama Compact Array (ACA) to observe 22.4 arcmin^2 in total. Pullen et al. (2018)

---

[7] We do not use the fluctuation δ for computing the power spectrum as it is generally difficult to obtain the mean intensity because of a large observational noise.

have used the intensity map obtained by Planck and given an upper limit on the cross-power spectrum between the [CII] intensity maps and quasar samples from BOSS at $z \sim 2.5$.

Future observations will probe line intensity maps over larger volumes. In radio observations, the 21-cm intensity maps at $z = 0.2 - 2.5$ will be observed by the Canadian Hydrogen Intensity Mapping Experiment (CHIME)[8] currently in operation and by the SKA (Carilli 2015) Hydrogen Intensity and Real-time Analysis eXperiment (HIRAX; Newburgh et al. 2016),[9] Baryon acoustic oscillations In Neutral Gas Observations (BINGO; Battye et al. 2016),[10] and Five-hundred-meter Aperture Spherical radio Telescope (FAST; Bigot-Sazy et al. 2016)[11] in the future. The SKA surveys over as large as $\sim 30,000$ deg^2. As for CO and [CII] lines, the CO Mapping Array Pathfinder (COMAP; Li et al. 2016)[12] and CarbON CII line in post-rEionisation and ReionisaTiOn epoch (CONCERTO; Concerto Collaboration et al. 2020)[13] have started LIM observations over an area of ~ 1 deg^2 at $z \sim 3$. They will observe signals up to $z = 7$. Recently, the Tomographic Ionized-carbon Mapping Experiment (TIME; Crites et al. 2014) has also begun observations of [CII] at the epoch of reionization. Future LIM plans for such wavelength regimes include the Terahertz Intensity Mapper (TIM; Vieira et al. 2020), South Pole Telescope Summertime Line Intensity Mapper (SPT-SLIM; Karkare et al. 2021), CCAT-prime (Stacey et al. 2018)[14], and Experiment for Cryogenic Large-aperture Intensity Mapping (EXCLAIM; Ade et al. 2020).

LIM observations in optical to near-infrared regimes will also be conducted. For Lyα intensity mapping, we can use the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX Hill et al. 2008).[15] This was originally designed to detect individual bright Lyα emitters in the 400 deg^2 at $z = 1.9 - 3.5$, but the data obtained by its integral field unit also provides the line intensity maps. Rest-frame optical lines such as Hα, [OIII], and [OII] will also be observed by NASA's all-sky survey instrument Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (SPHEREx; Doré et al. 2018).[16] Data in its deep regions of 200 deg^2 can be used for LIM study. In the future, the Cosmic Dawn Intensity Mapper (CDIM; Cooray et al. 2019) is planned to observe them with higher sensitivity. We show the wavelength ranges covered by SPHEREx-like and TIME-like LIM observations in Fig.2.1.

Most of the ongoing LIM observations aim at detecting small-scale signals. The observed small-scale signals can be used to study galaxy population at the specific redshifts. We can estimate the average properties of ISM such as neutral hydrogen abundance (e.g., Keating et al. 2020) as well as mean star formation rate density. Line luminosity functions can also be estimated from the probability distribution function (PDF) (Breysse et al. 2017). As the LIM

---

[8] `https://chime-experiment.ca/en`
[9] `https://hirax.ukzn.ac.za`
[10] `https://bingotelescope.org`
[11] `https://fast.bao.ac.cn`
[12] `https://comap.caltech.edu`
[13] `https://mission.lam.fr/concerto/`
[14] `https://www.ccatobservatory.org`
[15] `https://hetdex.org`
[16] `https://spherex.caltech.edu`

measures integrated signals including contributions from faint galaxies, weak emission lines could also be detected. Visbal et al. (2015) have proposed to estimate the influence of Pop III stars by statistically detecting intrinsically weak HeII line emissions.

Future observations are expected to detect large-scale signals and provide cosmological constraints. Fonseca et al. (2017) have shown that HETDEX, SPHEREx, and TIME-like[17] observations can detect auto-power spectra at $z = 1 - 2$ and $k \sim 0.01 - 0.1\ \mathrm{Mpc}^{-1}$ (corresponding to $\sim 50 - 500$ Mpc) with signal-to-noise ratio $\gtrsim 100$ and can detect the BAO signals. Detection of the large-scale signals would allow us to investigate the modified gravity theories as well as the equation of state of the dark energy (e.g., Bull 2016). Many of the LIM projects will cover regions that have already been observed by previous galaxy surveys. Combination of the LIM data and the galaxy catalogs would allow more robust constraints. Villaescusa-Navarro et al. (2015) have shown that the neutrino mass can be constrained within an error competitive to the current constraints by combining the SKA LIM data and other galaxy surveys. Future LIM observations will provide maps tracing the ultra-large scale, which can be used for multi-tracer analysis to study the primordial non-Gaussianity (Fonseca et al. 2015). We may also be able to study the properties of dark matter by detecting radiation from decaying dark matter. Creque-Sarbinowski & Kamionkowski (2018) have shown that the lifetime of 10^{-6} - 10 eV axion-like particles that decay into 0.1 GHz - 1000 THz can be constrained with ongoing/future experiments such as CHIME, COMAP, and SPHEREx. Moreover, several planned or proposed line intensity mappers such as CDIM and TIME-NG (Sun et al. 2021) can perform LIM observations at the EoR. Such LIM data can be used to study the cosmic reionization in combination with the high-redshift 21-cm line intensity maps (e.g., Dumitru et al. 2019).

2.3 Observations of the Cosmic Reionization

2.3.1 Current Observational Constraints on the Reionization

One of the observables to study the reionization is the Lyα absorption signal in distant quasar spectra. Neutral hydrogen at redshift z in the line of sight of a distant quasar absorbs and scatters Lyα photons, which are observed as absorption signals at wavelength $1216(1+z)$ Å. By determining the optical depth τ from the observed spectra assuming intrinsic intensity of the quasars, we can obtain the neutral hydrogen density as

$$n_{\mathrm{HI}}(z) = \frac{a(z)H(z)\nu_{\mathrm{Ly}\alpha}}{(1+z)c\alpha_{\mathrm{Ly}\alpha}}\tau(z), \tag{2.20}$$

where $\alpha_{\mathrm{Ly}\alpha} = \int \sigma_{\mathrm{Ly}\alpha} d\nu$ is the integrated optical depth. These observations, so-called Gunn-Peterson tests, suggest that the reionization is complete at $z \sim 6$ (e.g., Fan et al. 2006; McGreer et al. 2015). This test can only be used to study the very late stages of reionization because the cross-section of Lyα is so large that even a small amount of neutral hydrogen would cause a saturated absorption.

---

[17] Fonseca et al. (2017) consider observations with larger survey area than that of actual TIME.

Another constraint comes from the CMB observations. The presence of free electrons in the IGM suppresses the temperature fluctuation and generates the polarization of the CMB via inverse Compton scattering (Thomson scattering). From these observations, the Thomson optical depth, i.e., the integrated number of free electrons along the line-of-sight direction,

$$\tau = \sigma_{\mathrm{T}} \int_0^z n_e(z') \frac{dl(z')}{dz'}\, dz' \tag{2.21}$$

is estimated. Recent Planck observation suggests that the reionization occurred at $z \sim 8.8$, assuming instantaneous reionization (Planck Collaboration XIII 2016).

The reionization can also be probed by the observation of Lyα emitters. If a galaxy is surrounded by neutral hydrogen, it scatters the Lyα photons emitted from the galaxy, making it more difficult to detect. Thus we can estimate the neutral hydrogen abundance from the difference between the Lyα emitters and other galaxy samples in their number densities or in the degree of their clustering. Results from the recent Lyα observations suggest a consistent reionization history with the other observations (Konno et al. 2018; Ouchi et al. 2018).

The reionization is considered to be caused by the photoionization by photons from high-redshift galaxies. Recent high-redshift observations suggest that there are a sufficient number of galaxies to induce the cosmic reionization if we assume a somewhat larger escape fraction, $f_{\mathrm{esc}} \gtrsim 0.2$, than local galaxies (Ouchi et al. 2009).

2.3.2 Observations of the 21-cm Lines at the EoR

The above observations provide only indirect and/or limited information on cosmic reionization. Direct and more detailed studies can be done using the 21-cm spin-flip transition emission from neutral hydrogen (Field 1959). The 21-cm line is observed as an emission or absorption line depending on the spin temperature relative to the temperature of the background radiation (CMB). The 21-cm intensity scales with the amount of neutral hydrogen and also depends on the IGM temperature. More details are summarized in Appendix B.1.

One of the observational methods for the 21-cm line is detecting global signals. Various observations of the global 21-cm signals have been made (Price et al. 2018; Voytek et al. 2014). Recently, Bowman et al. (2018) reported the detection of the 21-cm absorption at $z \sim 18$ by the Experiment to Detect the Global Epoch of Reionization Signature (EDGES). The detected signal is considered to correspond to the decrease of the spin temperature by Lyα coupling and the increase of it by IGM heating. The amplitude of the absorption signal is much larger than predicted by any theoretical model at that time, suggesting several possibilities including the existence of mechanisms other than adiabatic cooling that are responsible for the IGM cooling and existence of additional background radiation other than the CMB.

In addition to global signal observations, we can observe the 21-cm line intensity fluctuations by LIM observations using interferometers. So far, observations by the LOw-Frequency ARray (LOFAR; van Haarlem et al. 2013), Murchison Widefield Array (MWA; Bowman et al. 2013), Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010), Giant Metrewave Radio Telescope (GMRT; Paciga et al. 2011), and Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017) have already been conducted, in which the upper limits on the 21-cm power spectrum at the EoR are obtained. From these observations, a few reionization/heating models in which the IGM heating occurs at a later epoch have been disfavored (e.g., Ali et al. 2015; The HERA Collaboration et al. 2021). The SKA will provide more sensitive observation soon.

The greatest difficulty in observing the 21-cm signals is the large foreground, including the Galactic synchrotron emissions and extra-galactic radio sources. The mean amplitude of the foregrounds of the 21-cm line is estimated to be about four orders of magnitude larger than those of the distant 21-cm signals. These foregrounds have smooth spectra, and they can partly be removed or avoided by several methods (e.g., Liu et al. 2014; Mertens et al. 2018). However, it is quite difficult to remove all the contributions, and they could still cause systematic errors on the EoR 21-cm observations.

To statistically remove the foreground contribution, we can take a cross-correlation with other tracers such as galaxies at the same redshift. Many works discuss the information obtained from the cross-correlation analysis and predict the detectability of the cross-correlation signals at the EoR. In particular, the combination of the 21-cm observations and Lyα emitter surveys has been considered in many studies (e.g., Heneka & Mesinger 2020; Lidz et al. 2009; Vrbanec et al. 2016; Wiersma et al. 2013). While the Lyα emitters are good tracers in the late stages of reionization, recent observations show a tendency that the Lyα emissions weaken rapidly at $z > 7$ (Konno et al. 2014).

Metal lines can also be good high-redshift tracer candidate. We can use single-dish telescopes such as the Large Submillimeter Telescope (LST; Kohno et al. 2020) to perform large-scale surveys of far-infrared [Oiii]/[Cii] line emitters at high redshifts in the future. Another possibility is to use [Cii] line intensity maps (e.g., Dumitru et al. 2019), which can be measured by future TIME-like observations (e.g., TIME-NG; Sun et al. 2021). The [Oiii] intensity maps observed by SPHEREx-like telescopes can also be a good tracer of the LSS at the high-redshift universe (e.g., Kannan et al. 2021).

Chapter 3

Modeling Emission Line Galaxies

Predicting the emission line signals from galaxies and IGM observed in future experiments is crucial for their interpretation as well as for forecasting their detectability. Previous studies have predicted the LIM signals using simple models, such as adopting a simple one-to-one relation for the halo mass-to-luminosity ratio (e.g., Silva et al. 2018; Yue & Ferrara 2019). More accurate predictions require more realistic modeling.

In this chapter, we build a model of emission line galaxies taking into account the ISM properties. We first summarize the basics of emission lines from ionized regions (HII regions) of galaxies in Section 3.1, and then we describe our emission line model in Section 3.2. In Section 3.3, we describe our method based on Moriwaki et al. (2018) to generate a large number of mock LIM catalogs used to train and test the neural networks in this study.

3.1 Line Emissions from HII Regions

Photons with $E > 13.6$ eV emitted from stars in galaxies ionize the surrounding hydrogen and form ionized regions (HII regions). We first compute the size of the HII regions. In a steady-state, ionization and recombination are balanced. Assuming that the photons emitted by recombination to the ground state immediately ionize nearby neutral hydrogen (case-B assumption), this balance in a pure hydrogen gas can be written as

$$n_{\rm H} J_* \overline{\sigma}_{\rm H} = n_{\rm e} n_{\rm p} \alpha_B, \tag{3.1}$$

where $n_{\rm H}$, $n_{\rm e}$, and $n_{\rm p}$ are the number densities of hydrogen atoms, electrons, and protons, the flux J_* is the number of ionizing photons per unit area and time, $\overline{\sigma}_{\rm H}$ is the photoionization cross-section, and $\alpha_{\rm B}$ is the effective recombination coefficient under the case-B assumption.

Let us assume that a young stellar cloud emits ionizing photons in a uniform gas. If the radiation is emitted isotropically, the number of ionizing photons passing through a spherical shell at radius r per unit time, $Q(r)$, decreases with radius as

$$\begin{aligned} \frac{{\rm d}Q(r)}{{\rm d}r} &= -4\pi r^2 n_{\rm H} J_* \overline{\sigma}_{\rm H} && (3.2) \\ &= -4\pi r^2 n_{\rm e}^2 \alpha_{\rm B}. && (3.3) \end{aligned}$$

Note we have used Eq. 3.1 with $n_{\rm e} = n_{\rm p}$. As we assume constant gas density, this can be solved as

$$Q(r) = Q_0 - \frac{4\pi}{3} r^3 n_{\rm e}^2 \alpha_{\rm B}, \tag{3.4}$$

where Q_0 is the total number of photons per unit time emitted from the stellar cloud at the center. The boundary of the HII region $R = R_{\rm S}$ is given by the radius where $Q(r) = 0$

$$R_{\rm S} = \left(\frac{3Q_0}{4\pi n^2 \alpha_{\rm B}} \right)^{1/3}. \tag{3.5}$$

This radius is called the Strömgren radius. The mean free path of an ionizing photon in a neutral medium is quite small and thus a photon cannot travel far enough beyond the boundary of the HII region. The above results do not significantly change when we include a small amount of metal.

We then formulate the line emissivity from atoms and ions in HII regions. We consider metal lines as well as hydrogen lines. The emissivity of a line from a transition i → j (i > j) is written as

$$\epsilon_{\rm ij} = E_{\rm ij} n_{\rm i} A_{\rm ij}, \tag{3.6}$$

where $E_{\rm ij}$ is the energy gap between the two levels, $n_{\rm i}$ is the number density of hydrogen atoms in level i, and $A_{\rm ij}$ is the probabilities of radiative transition from level i to j. For hydrogen recombination lines (i.e., lines emitted in a cascade after an atom captures a free electron), the collisional transitions are negligible, and the excitation and deexcitation for level i are balanced as

$$n_{\rm p} n_{\rm e} \alpha_{\rm i} + \sum_{\rm j>i} n_{\rm j} A_{\rm ji} = n_{\rm i} \sum_{\rm j<i} A_{\rm ij} \tag{3.7}$$

under the statistical equilibrium, where $\alpha_{\rm i}$ is the recombination coefficient into level i. In this case, the emissivity is proportional to the density squared, and the effective coefficient $\alpha_{\rm ij}^{\rm eff}$ is defined as

$$\epsilon_{\rm ij} \equiv E_{\rm ij} n_{\rm p} n_{\rm e} \alpha_{\rm ij}^{\rm eff}. \tag{3.8}$$

There are other important types of emission lines that are mainly emitted after collisional excitation. These emission lines are mostly the forbidden lines (i.e., lines from transitions forbidden by the selection rule), so the recombination term can be ignored. The excitation and deexciation balance is written as

$$\sum_{\rm j>i} n_{\rm j} A_{\rm ji} + n_{\rm e} \sum_{\rm j\neq i} n_{\rm j} q_{\rm ji} = n_{\rm i} \sum_{\rm j<i} A_{\rm ij} + n_{\rm e} n_{\rm i} \sum_{\rm j\neq i} q_{\rm ij}, \tag{3.9}$$

where n_i is the number density of atoms or ions in level i, and $q_{\rm ij}$ is the probabilities of collisional transition from level i to j. If we consider a two-level system for simplicity, Eq. 3.9 is reduced to

$$q_{12} n_{\rm e} n_1 = n_2 A_{21} + q_{21} n_{\rm e} n_2. \tag{3.10}$$

As the forbidden transition is a rare event, most of the bound electrons reside in the lower level. Solving Eq. 3.10 for the density in the upper level, we get

$$n_2 = \frac{q_{12} n_\mathrm{e}}{A_{21} + q_{21} n_\mathrm{e}} n_1. \tag{3.11}$$

The collisional transition probabilities are balanced in equilibrium as

$$\frac{q_{12}}{q_{21}} = \frac{g_2}{g_1} \mathrm{e}^{-E_{21}/k_\mathrm{B}T}, \tag{3.12}$$

where g_1, g_2 are statistical weights. The temperature of the HII region is balanced by the photoionization heating and the radiative cooling by metal gas. When the temperature increases, the number density of the upper-level n_i and therefore the emissivity increase.

When the gas density is low, radiative deexcitation occurs more frequently than collisional deexcitation and the emissivity scales with the square of the density

$$\epsilon_{21} \sim E_{21} n_\mathrm{e} n_1 q_{12} \propto n_\mathrm{e}^2 \quad (n_\mathrm{e} \ll A_{21}/q_{21}), \tag{3.13}$$

while at high density, the emissivity scales with the density

$$\epsilon_{21} \sim E_{21} n_1 A_{21} \frac{q_{12}}{q_{21}} \propto n_\mathrm{e} \quad (n_\mathrm{e} \gg A_{21}/q_{21}). \tag{3.14}$$

The density at which the radiative and collisional deexcitations are balanced

$$n_\mathrm{crit} \equiv \frac{A_{21}}{q_{21}} \tag{3.15}$$

is called the critical density. The emissivity of the forbidden lines also scales with the abundance of the atoms or ions, which depend on the strength of the ionization radiation. If the number of ionizing photons per gas particle is large, each atom is ionized to a higher level. The relative strength of the radiation is expressed in terms of the ionization parameter, which is defined as the ratio between the ionizing photon density n_LyC and the electron density:

$$U \equiv \frac{n_\mathrm{LyC}}{n_\mathrm{e}}. \tag{3.16}$$

The luminosity of the emission line is computed by multiplying the emissivity by the volume of the emission region:

$$L_\mathrm{ij} = \epsilon_\mathrm{ij} V_\mathrm{HII}. \tag{3.17}$$

The volume of the HII region scales as $V_\mathrm{HII} \propto R_\mathrm{S}^3 \propto Q_0/n^2$ (Eq.3.5). For the recombination lines and forbidden lines in a low-density medium, the emissivity scales with density as $\epsilon_\mathrm{ij} \propto n_\mathrm{e}^2$, and thus the luminosity is independent of the density

$$L_\mathrm{ij} \propto n^2 \frac{Q_0}{n^2} = Q_0. \tag{3.18}$$

Table 3.1: Bright optical emission lines. From left to right, the line name, rest-frame wavelength, ion, ionization potential $E_{\rm ion}$ of the ion, critical density $n^{\rm e}_{\rm crit}$, and excitation temperature $T_{\rm exc} = E_{\rm ul}/k_{\rm B}$. The critical densities are shown for $T = 10^4$ K.

	$\lambda_{\rm rest}$[Å]	ion	$E_{\rm ion}$ [eV]	$n^{\rm e}_{\rm crit}$ [cm^{-3}]	$T_{\rm exc}$ [K]
Hα	6563	$\rm H^+$	13.6		2.2×10^4
[OIII]5007	5007	$\rm O^{++}$	35.1	6.8×10^5	2.9×10^4
[OII]3727	3727	$\rm O^+$	13.6	3.4×10^3	3.9×10^4

If there is dust in the ISM, they absorb the photons emitted from the stars and the ISM, and re-radiate them in the far-infrared. In general, the dust extinction is expressed by the optical depth, τ, as

$$I = I_0 {\rm e}^{-\tau}, \tag{3.19}$$

or the dust extinction magnitude, A, as

$$I = I_0 \times 10^{-A/2.5} \tag{3.20}$$

where I_0 and I are the intrinsic and observed intensities. The dust absorption spectrum is known to have a peak in the ultra-violet (UV) band.

3.2 Emission Line Model

In this thesis, we consider several emission lines with rest-frame optical wavelengths observed by future surveys such as SPHEREx. There are three bright emission lines: a recombination line Hα and two forbidden lines from doubly-ionized (OIII) and singly-ionized (OII) oxygen. Their properties are summarized in Table 3.1. In the following, these emission lines are abbreviated as Hα, [OIII], and [OII]. These lines are all emitted from HII regions of galaxies. As we have seen in Section 3.1, the luminosities of the hydrogen recombination lines including Hα are almost independent of the ISM properties, except for the effect of absorption by dust. On the other hand, the luminosities of lines from oxygen ions, [OIII] and [OII], depend on the oxygen abundance and the ionization parameter. We build an emission line model that takes into account these dependencies. We note that both the [OIII] and [OII] lines have higher critical densities than the typical gas density of the HII regions (Osterbrock & Ferland 2006), and the gas density has little effect on their luminosity (Eq.3.18).

In Section 3.1, we have seen that the luminosity of the emission line from the HII region roughly scales with the emission rate of ionizing photons. As young and massive stars emit more ionizing photons, the total emission rate of ionizing photons is proportional to the instantaneous star formation rate (SFR). Therefore, we compute the line luminosities of galaxies as

$$L_{\rm line} = 10^{-A_{\rm line}/2.5}(1 - f_{\rm esc})\, C_{\rm line}\, {\rm SFR}, \tag{3.21}$$

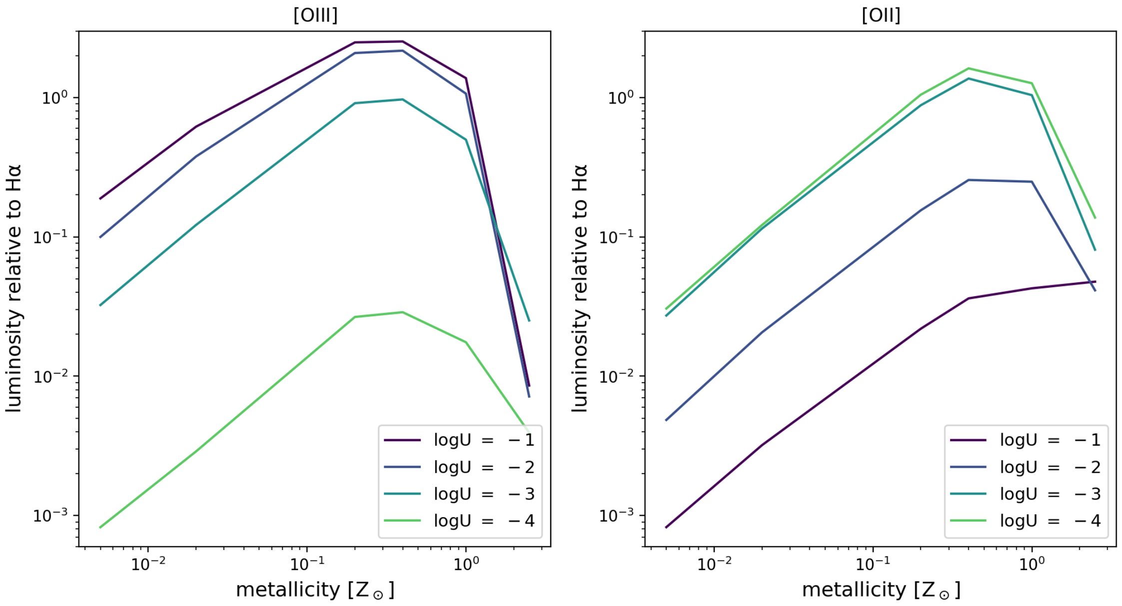


Figure 3.1: The luminosity of [O III]5007Å (solid) and [O II]3727Å (dashed) relative to Hα as a function of the metallicity computed with CLOUDY. The blue, orange, and green lines are those computed with $\log U = -1, -2$, and -3, respectively. The luminosity is uncorrected for the dust extinction. The solar metallicity is set to be $= 0.02$ for this plot. We adopt the solar abundance.

where $A_{\rm line}$ is the dust attenuation magnitude, and $f_{\rm esc}$ is the escape fraction of ionizing photons from the galaxy. For the dust extinction, we adopt the typical values inferred from the galaxy observations (Khostovan et al. 2015; Sobral et al. 2016): $A_{\rm H\alpha} = 1.0$ mag, $A_{\rm [OIII]} = 1.35$ mag, and $A_{\rm [OII]} = 1.0$ mag. The escape fraction of intermediate-redshift galaxies are estimated as small as $f_{\rm esc} < 0.1$ (e.g., Rutkowski et al. 2016). We thus assume $f_{\rm esc} = 0$ for simplicity, except in Chapter 7, where we assume a non-zero escape fraction to be consistent with the reionization simulation.

The coefficient $C_{\rm line}$ in Eq.3.21 depends on the spectrum of the ionizing source and ISM properties such as the temperature and element abundance. To compute the coefficients, we use the plasma simulation code CLOUDY (Ferland et al. 2017). In CLOUDY, physical conditions (i.e., thermal and ionization states and chemical compositions) of a gas cloud illuminated by an external field are computed. We run CLOUDY with the gas metallicity Z and the ionization parameter U (Eq. 3.16) as variable parameters to compute the coefficient:

$$C_{\rm line} = C_{\rm line}(Z, U). \tag{3.22}$$

In practical terms, we compute the coefficients for parameters listed in Table 3.2 to generate a lookup table. We obtain the coefficient for each simulated galaxy by interpolating the table values.

In the CLOUDY calculation, we set up a plane-parallel cloud with a constant gas density that is exposed to an ionizing source. We adopt the typical gas density observed in nearby

Table 3.2: The metallicity Z and the ionization parameter U used in the CLOUDY simulation. The ionization parameter is defined at the irradiated surface of the gas cloud.

metallicity Z	0.0001, 0.0004, 0.004, 0.008, 0.02($\sim Z_\odot$), 0.05
ionization parameter U	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}

HII regions, $n = 100$ cm^{-3} (Osterbrock & Ferland 2006). We consider a stellar cloud as the ionizing source and use the spectral energy distribution (SED) synthesis code BC03 (Bruzual & Charlot 2003) to compute the SEDs of the stellar cloud.[1] We adopt Salpeter IMF (Salpeter 1955) and assume that the metallicity of the stars is the same as that of the gas. We adopt the solar abundance ratios of heavy elements. That is, the abundance of the heavy element y scales with the metallicity Z as

$$y = y_\odot \times \left(\frac{Z}{Z_\odot}\right), \tag{3.23}$$

where $y_\odot$ is the solar abundance, and $Z_\odot$ is the solar metallicity. The calculation is terminated at the region where the ionization fraction falls below 10^{-3}.

Fig.3.1 shows how the luminosity ratios of [OIII] and [OII] to Hα change with respect to the metallicity and the ionization parameter. We assume solar abundance and thus the oxygen abundance is proportional to the metallicity. The luminosities scale with the oxygen abundance at the low-Z limit. When the metallicity is high, the gas is cooled down and the luminosities decrease because of the lower collisional excitation rate. For [OIII] line, the luminosity relative to the Hα line luminosity thus has a peak at around $Z \sim 0.3$Z$_\odot$. When the ionization parameter is high, more OII ions are converted to OIII ions resulting in a large luminosity ratio of [OIII] to [OII].

In LIM observations, we observe the integrated line emission signals. Here, we compare the mean contributions of the individual emission lines observed at far-infrared wavelengths. To compute the mean line intensities that depend on the galaxy population, we use cosmological hydrodynamics simulation IllustrisTNG (Nelson et al. 2019, see Section 3.3 for more detail). We use the metallicity averaged within the stellar half mass radius. We adopt $U = 10^{-2}$ based on recent observations of [OIII]/[OII] lines at intermediate redshifts (Nakajima et al. 2013). Fig.3.2 shows the mean intensities of emission lines including subdominant lines from sulfur and nitrogen ions ([SII] 6731/6717Å, [NII] 6584Å/6548Å) and a hydrogen recombination line (Hβ). We find that the Hα and [OIII] are two dominant emission lines. The [OII] is subdominant in this wavelength regime. We should note that the [OII] line could be brighter if the ionization parameter is much smaller or there is less dust extinction than we expect. In such cases, the [OII] emissions from $z \sim 2$ could dominate at $\lambda_{\rm obs} \sim 1.0$ μm. Another possible interloper is hydrogen Lyα emission from the EoR, which we do not include in Fig.3.2. The Lyα line from $z = 10$ is observed at 1.2 μm, for example. While the emissivity of the Lyα line at such high

---

[1] The SED depends on the star formation history. We generate SED assuming constant star formation history with a duration of 10 Myr. The choice of the star formation duration does not significantly change the results when it is longer than 10 Myr (Inoue 2011).

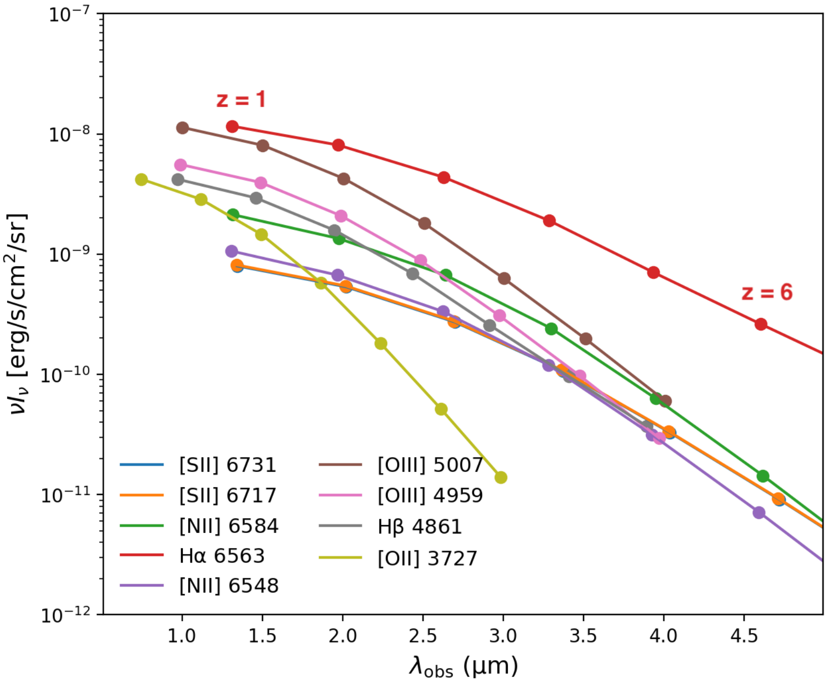


Figure 3.2: Mean intensity of line emissions as a function of observed wavelength. The dots correspond to z = 1, 2, 3, 4, 5, 6, 7 from left to right, respectively.

Table 3.3: Observational parameters of SPHEREx Deep (Doré et al. 2014).

Field of view	200 deg^2
Angular resolution (l)	6".2
Wavelength ($\lambda_{\rm obs}$)	0.75-4.18 μm / 4.18-5.0 μm
5σ point source sensitivity ($m_{\rm AB}$)	$\sim$ 22 / 21-20
Spectral resolution (R)	41 / 135
Hα redshift ($z_{\rm H\alpha}$)	0.14-5.4 / 5.4-6.6
[OIII] redshift ($z_{\rm [OIII]}$)	0.5-7.4 / 7.4-9.0

redshifts is highly uncertain, its mean contribution is much smaller than the other lines because of the low star formation rate density at such high redshifts. In the following, we only consider the Hα and [OIII] emission lines for simpllicity.

3.3 Mock Observational Line Intensity Maps

In this section, we describe how we create mock observational data used for training and testing of the neural networks in Chapter 4 - 6. We consider the LIM observation by SPHEREx planned to be launched in 2024 (Doré et al. 2014). It will perform spectroscopic observation at 0.75 - 4.18 μm with $R \sim 41$ and 4.18 - 5.0 μm with $R \sim 135$. The deep observation will cover 200 deg^2. The details of the deep observation are summarized in Table 3.3. The instrumental noise level is $\sigma_n = 2.6 \times 10^{-6}$ erg/s/cm^2/sr at 1.5 μm.

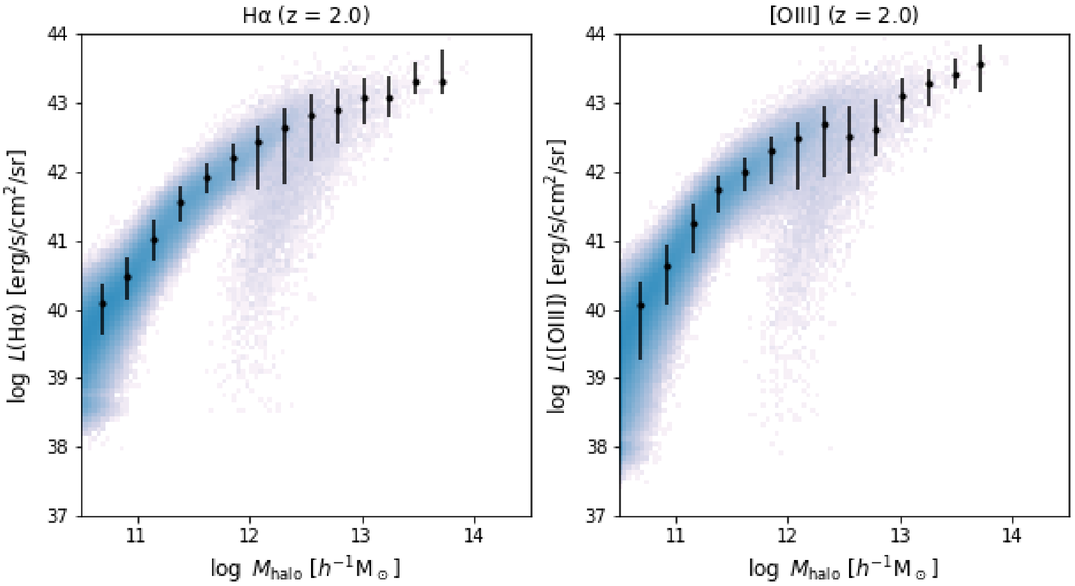


Figure 3.3: The halo mass-to-luminosity relation for Hα. The color map shows the distributions of galaxies at $z = 2$ in IllustrisTNG. The error bars show the mean and variance for individual halo-mass bins.

To generate a large number of mock observational data, we use the fast halo catalog generation code PINOCCHIO (Monaco et al. 2013). In PINOCCHIO, the density field is grown based on Lagrangian perturbation theory (LPT; Moutarde et al. 1991) and dark-matter halos are generated based on a fragmentation algorithm. We first generate a halo catalog with box size $L_{\rm box}$.

To assign the line luminosities to the catalog halos, we use the halo mass-to-line luminosity relation computed with IllustrisTNG (Nelson et al. 2019). IllustrisTNG follows the formation and evolution of galaxies from $z = 127$ to $z = 0$ using moving-mesh code AREPO (Springel 2010). The simulation includes star formation and evolution, chemical enrichment, radiative cooling of gas, black hole formation and evolution, and feedback from supernovae and black holes (Pillepich et al. 2018; Weinberger et al. 2017), and is calibrated to reproduce observational results such as the galaxy stellar mass function and stellar-to-halo mass relation. The initial mass function (IMF) of Chabrier (2003) is adopted in the star formation. We use a dataset TNG300-1 that has a box size of $(205h^{-1}\ \mathrm{Mpc})^3$ and dark matter particle resolution of $m_{\rm DM} = 7.6 \times 10^6 h^{-1}\ \mathrm{M}_\odot$. The basic results of IllustrisTNG are summarized in a series of papers (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018; Springel et al. 2018).

We compute the luminosities of galaxies in IllustrisTNG using Eq.3.21 and then sum up all the luminosities of the galaxies within a halo to compute the total luminosity of each halo. Note that we do not consider the distribution of the galaxies within a halo because the spatial extent of halos is mostly smaller than the angular resolution of SPHEREx (0.1 arcmin). The

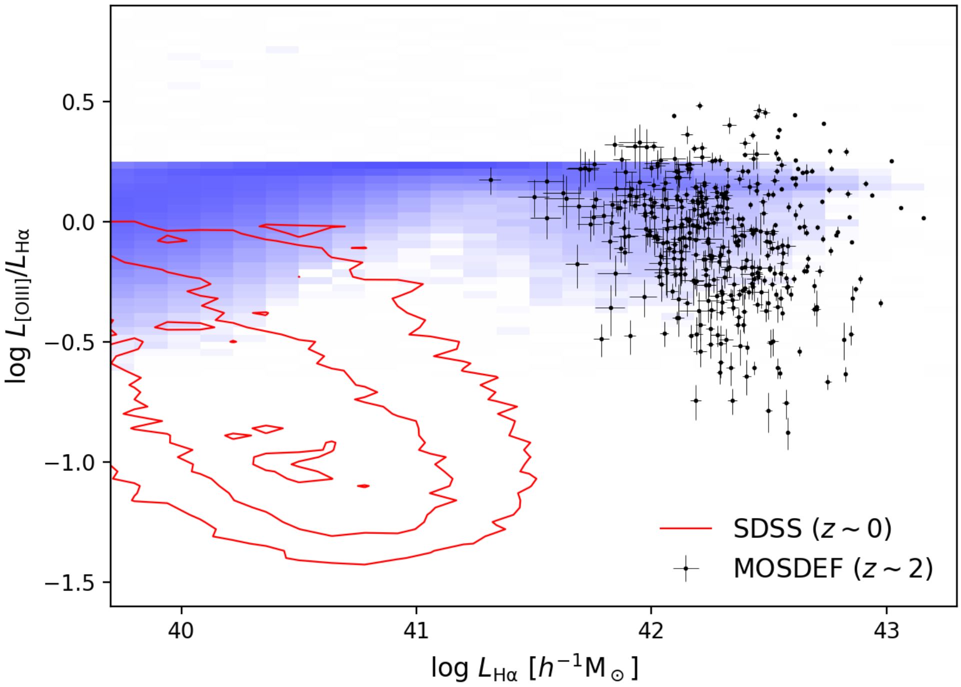


Figure 3.4: The distribution of the line luminosity ratios [OIII]/Hα. The color map shows the distribution of galaxies at $z = 2$ in IllustrisTNG. The error bars and contours show the distributions of galaxies at $z \sim 2.3$ from the MOSDEF survey (Kriek et al. 2015) and line emitters at $z \sim 0$ from SDSS. For SDSS, we only show the line emitters with S/N > 5.

obtained halo mass-to-line luminosity relations at $z = 2$ are shown in Fig.3.3. The luminosity scales with the halo mass at the low mass end and is suppressed at $M_{\mathrm{halo}} \gtrsim 10^{12}h^{-1}$ M$_\odot$. This is because of the suppression of the star formation by AGN feedback. We find that the total line luminosity is mostly contributed by $M_{\mathrm{halo}} \gtrsim 2 \times 10^{11}h^{-1}$ M$_\odot$ at $z = 2$. To compare our line emission model with observations, we show the distribution of the line luminosity ratios [OIII]/Hα in Fig.3.4. The color map shows the distribution of the galaxies at $z = 2$ in our model. Since we fix the ionization parameter, we observe a sharp edge at log [OIII]/Hα ~ 0.2. The contours and the error bars in Fig.3.4 show the distributions of the line emitters at $z \sim 0$ from SDSS[2] and those at $z \sim 2.3$ from the MOSFIRE Deep Evolution Field (MOSDEF) survey (Kriek et al. 2015)[3], respectively. At the bright end, our model is in good agreement with the $z \sim 2.3$ galaxies. This is because we adopt a higher ionization parameter ($U = 10^{-2}$) than a typical value of local galaxies ($U \sim 10^{-3}$) based on recent observations at $z = 2 - 3$ (e.g., Nakajima et al. 2013). Varying the ionization parameter would reproduce the observed distribution at $z \sim 2.3$ more properly, but this is beyond the scope of this study.

We calculate the mean logarithmic luminosity $\overline{\log L_i}$ and variance σ_i^2 for each halo mass bin which are shown as the errors in Fig.3.3. The line luminosity of a catalog halo in mass bin M_i

---

[2] http://www.mpa-garching.mpg.de/SDSS/DR7/

[3] https://mosdef.astro.berkeley.edu

Table 3.4: Properties of mock observational data used for our experiments: the box size ($L_{\rm box}$) and minimum halo mass ($M_{\rm min}$) of PINOCCHIO simulation, number of pixels ($n_{\rm pix}$), angular size of pixels ($l_{\rm pix}$), total area ($A_{\rm total} = n_{\rm pix} l_{\rm pix}^2$), dimension of the data, and number of realizations. We prepare two datasets with different total area for Chapter 4.

	$L_{\rm box}$[Mpc/h]	$M_{\rm min}$ [M$_\odot$/h]	$n_{\rm pix}$	$l_{\rm pix}$ [']	$A_{\rm total}$ [deg^2]	dim.	$n_{\rm rea}$
Chapter 4 (small)	280	3×10^{10}	512^2	0.1	0.85^2	2	300
Chapter 4 (large)	700	3×10^{11}	256^2	2.0	8.53^2	2	300
Chapter 5	280	3×10^{10}	256^2	0.4	1.71^2	2	300
Chapter 6	600	2×10^{11}	64^2	0.1	0.11^2	3	500

is calculated so that the logarithmic luminosity follows a normal distribution with mean $\overline{\log L_i}$ and variance σ_i^2

$$\log L(M_i) = \overline{\log L_i} + \sigma_i \mathcal{N}(0,1), \tag{3.24}$$

The luminosities are randomly sampled from this normal distribution. When considering multiple emission lines from a galaxy, we assign the luminosities using the same random value sampled from $\mathcal{N}(0,1)$. By computing the luminosities in this way, we can reproduce the rough distribution of the line luminosity ratio [OIII]/Hα of IllustrisTNG.

After assigning the luminosities to halo catalogs, we generate the line intensity maps using Eq.2.17 with the spectral resolution of SPHEREx ($R \sim 41$). For angular resolution ($l_{\rm pix}$), we adopt those listed in Table 3.4. When we include the observational noise, we consider Gaussian noise with variance

$$\sigma_n' = \sigma_n \frac{l}{l_{\rm pix}}, \tag{3.25}$$

where σ_n and $l = 6.2$" is the noise level and original angular resolution of SPHEREx.

For training, we generate $n_{\rm rea}$ halo catalogs with different realizations (initial conditions). We then generate 100 line intensity maps from each catalog by randomly extracting 100 regions of area $A_{\rm total} = n_{\rm pix} l_{\rm pix}^2$, obtaining $100 n_{\rm rea}$ training data in total. After training, we test the performance of the machines using datasets that are independent of the training data. For the test dataset, we use 1000 data generated with 1000 realizations that are different from those of training dataset. Table 3.4 shows the parameters used for generating mock data. In Chapter 4, we prepare two training datasets with different resolutions and areas.

Chapter 4

Signal Extraction from Noisy LIM Data

LIM is a very powerful method to efficiently survey a large volume. On the other hand, its sensitivity is generally not as good as that of conventional galaxy surveys. It is thus important to properly extract the signals from noisy observational maps. Recently, machine learning methods have been applied to remove noises from various types of observational data. (e.g., Li et al. 2019; Shirasaki et al. 2021). The machine learning may also be used for denoising the observed line intensity maps.

In this chapter, we develop conditional generative adversarial networks (Isola et al. 2016) that denoise the observed line intensity maps. In Section 4.1 we will look at the basic idea of deep neural networks, convolutional neural networks, and generative adversarial networks. We then describe the choice of training data and network architecture in Section 4.2 and the results in 4.3. We further discuss the results in Section 4.4. The contents in this chapter are based on Moriwaki et al. (2021).

4.1 Machine Learning Algorithms

4.1.1 Basics of Neural Networks

There are two types of machine learning methods: supervised learning, in which training data is given with labels, and unsupervised learning, in which the machine is trained without labels. Supervised learning can be used for classification and regression tasks, and unsupervised learning can be used for tasks such as clustering analysis. In this thesis, we consider supervised learning with neural networks.

We consider solving a regression problem by training a deep neural network that gives a number $f(x)$ for an input vector x. As for the training, a set of answers $\{p_i\}$ to the input data $\{x_i\}$ is given, and the network is trained so that $\{f(x_i)\}$ gets as close as possible to $\{p_i\}$. A deep neural network is a network that repeats similar operations (layers) multiple times

$$x \to h_1(x) \to h_2(h_1(x)) \to ... \to f(x), \tag{4.1}$$

and is known to be capable of extracting more complex information than a single-layer network. A typical operation is a dense layer (or linear layer, feed-forward layer), in which the operation $x \to y$ at each layer is expressed as

$$y = a(wx + b), \tag{4.2}$$

where a is the nonlinear activation function, w is the weight matrix, and b is the bias matrix. The weight and bias matrices are the trainable parameters of the network. The dimension of the layer output can be chosen arbitrarily by changing the dimension of the weight matrix.

There are many kinds of **activation functions** that allow a network to perform nonlinear and complex operations. Commonly used activation functions include the Rectified Linear Unit (ReLU)

$$a(x) = \begin{cases} 0 & (x < 0) \\ x & (x \geq 0), \end{cases} \tag{4.3}$$

Leaky ReLU

$$a(x) = \begin{cases} -\alpha x & (x < 0) \\ x & (x \geq 0), \end{cases} \tag{4.4}$$

where α is a small constant, tanh

$$a(x) = \tanh(x), \tag{4.5}$$

and sigmoid

$$a(x) = \frac{1}{1 + \mathrm{e}^{-x}}. \tag{4.6}$$

Leaky ReLU is known to be more stable than the others when it is used between layers in a network. Other functions are also used especially in the output layer to introduce upper and lower bounds on the output. For example, the sigmoid function is used for binary (0 or 1) classification networks.

During the training, a set of the trainable parameters of the network including the w's and b's in Eq.4.2 are updated by evaluating its output with a **loss function**. Let n be the number of data used for the evaluation. The objective of the network is to return values as close as possible to the pre-given answers $\{p_i\}_{i=1}^{n}$. We can use, for example, the mean squared error

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{n} \sum_{i=1}^{n} (f(x_i) - p_i)^2 \tag{4.7}$$

as a loss function. Another important loss function is binary cross-entropy

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{n} \sum_{i=1}^{n} (p_i \log f(x_i) + (1 - p_i) \log(1 - f(x_i))), \tag{4.8}$$

which is commonly used for classification tasks.

An ideal way of training may be to update the network by evaluating the loss function on the entire training data every time. Actually, this is not a realistic strategy as it requires a large amount of memory and training time. In general, only a part of the training data (**mini-batch**) is used in each training. In (mini-)batch training, a batch is randomly selected for each time from the entire training data. This process is repeated until all the training data are used at least once, and the series of these training steps is called one epoch. Several epochs of training are often performed. The total number of training is (number of epochs) × (number of training data)/(batch size).

The trainable parameters, $\{\theta\}$, are updated toward the direction of the gradient of the loss function

$$g = \nabla_\theta \mathcal{L}[f_\theta]. \tag{4.9}$$

To perform stable training, we use an **optimizer** for the update. In a commonly used Adam Optimizer (Kingma & Ba 2014), parameters (weight, bias, etc.) are updated as

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \tag{4.10}$$

at t-th training, where we call the free parameter α the learning rate, and

$$\hat{m}_t = m_t/(1 - \beta_1^t) \tag{4.11}$$

$$\hat{v}_t = v_t/(1 - \beta_2^t) \tag{4.12}$$

are moments corrected with free parameters β_1^t and β_2^t. The moments m_t and v_t are updated as

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \tag{4.13}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2. \tag{4.14}$$

Adam optimizer suppresses sudden changes of parameters that can occur especially during the mini-batch training and prevents learning delays. In all of our experiments below, we adopt the same hyperparameters as in Isola et al. (2016) ($\alpha = 0.0002$, $\epsilon = 10^{-8}$, $\beta_1 = 0.5$, and $\beta_2 = 0.999$).

There are many other methods to improve the accuracy and stability of machine learning. One of them is **batch normalization** (Ioffe & Szegedy 2015), which is known to allow the training to be stable. In the batch normalization, the output of a layer, $\{y_i\}_{i=1}^n$, is normalized before input into the following layer as

$$\hat{y}_i = \frac{y_i - \mu}{\sqrt{\sigma^2 + \delta}}, \tag{4.15}$$

where μ and σ^2 are the mini-batch mean and the mini-batch variance,

$$\mu = \frac{1}{n}\sum_{i=1}^{m} y_i, \tag{4.16}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{m}(y_i - \mu)^2, \tag{4.17}$$

where n is the batch size, and δ is a small constant. We adopt $\delta = 10^{-5}$ in our experiments. For validation and test, moving mean $\overline{\mu}$ and variance $\overline{\sigma^2}$ computed during the training as

$$\overline{\mu}_t = m\,\overline{\mu}_{t-1} + (1-m)\,\mu_t \tag{4.18}$$

$$\overline{\sigma^2}_t = m\,\overline{\sigma^2}_{t-1} + (1-m)\,\sigma_t^2 \tag{4.19}$$

are used for batch normalization. We adopt $m = 0.9$ in the following.

Another method is **dropout** (Srivastava et al. 2014), in which some nodes are randomly dropped out and the others are scaled up by $1/(1-r)$, where r is the dropout rate. The operation $x \to y$ at a layer including dropout is expressed as

$$y = a[d(wx + b)], \tag{4.20}$$

where w and b are the weight and bias, and d is a diagonal matrix whose diagonal terms take either 0 or $1/(1-r)$. Dropout encourages every node to be useful and is known to suppress overfitting problems.

The output of machine learning varies depending on the random initialization of learnable parameters, the training data, and the hyperparameters. This non-deterministic nature of machine learning is suitable for some tasks but becomes noise for others, such as regression tasks, where we want to reproduce ground truth. To reduce such a generalization error, a technique called **bagging** is often used (Breiman 1996). This is a method of training several networks with datasets obtained by bootstrap sampling from the entire training dataset and of taking an average of their outputs, and is one of the ensemble methods. The generalization error can be reduced with this method when the errors of the networks are independent of each other (Goodfellow et al. 2016).

4.1.2 Convolutional Neural Networks

Next, we consider a network that can be used for analyzing images. In image processing, the relative positional relationship between pixels is important, but it cannot be handled by dense layers (Eq. 4.2). The most common way to deal with it is **convolution**. Convolution is an operation that multiplies a portion of an image by a filter (kernel) pixel-wise. The filter is shifted gradually and applied over the whole image. The number of pixels s to be shifted in one multiplication is called the stride. When the input image size is $N_{\rm pix} \times N_{\rm pix}$, the output size is $N_{\rm pix}/s \times N_{\rm pix}/s$. By using stride 2 or more, the data size in the hidden layer is reduced.

The convolution extracts features in the image that are spatially correlated with the filter. If we adopt a filter size of $k \times k$, only spatial features of the neighboring $k \times k$ pixels are extracted in a single operation of convolution, but in principle, larger-scale features can also be incorporated by repeating convolution multiple times. The number of layers required to share the information of all pixels from one end of the image to the other is of order $\log_k N_{\text{pix}}$.

One of the characteristics of convolution is its translational symmetry. As convolution performs the same operation for every part of the image, the output does not change except near the edge of the image.

The neural networks that consists of convolution layers is called a convolutional neural network (CNN). The operation performed in each CNN layer is written as

$$y = a(W * x + b), \tag{4.21}$$

where $W*$ represents the operation of the convolution with a filter W. The filter W and the bias b are trainable parameters.

4.1.3 Generative Adversarial Networks

A generative model generates new data that has similar properties to those in a training dataset. Generally, it learns the probability density that the training dataset follows, and generates new data by sampling random points from it. Several models such as autoencoder (AE) and variational autoencoder (VAE; Kingma & Welling 2013) have been proposed. Among them, generative adversarial networks (GAN; Goodfellow et al. 2014) attracted particular attention. GANs have the advantage of being able to model more complex probability densities than other models.

Let $\{y_i\}_{i=1}^N$ be the training dataset from which the GAN should learn the probability distribution. In a GAN, two CNNs called the generator and discriminator are trained adversarially. The objective of the generator is to generate new data that follows the same distribution as the training dataset, while that of the discriminator is to distinguish the training data and those generated by the generator, i.e., to return 1 for the training data and 0 for the generated data. The generator takes a random vector z as input and generate various data, $G : z \to G(z)$. The discriminator consists of a series of convolution layers, and the generator consists of a series of deconvolution layers. Here, deconvolution is defined as an operator that increases the number of pixels by a factor of $s \times s$, where s (> 1) is the stride. Note that the definition of stride for deconvolution is different from that for convolution. One of the operations that works as a deconvolution is the convolutional transpose which generates a larger image by convolving the input and then padding zero pixels between the pixels.

We use a sum of binary cross-entropy (Eq. 4.8) for the training data (y_i, $p_i = 1$) and the generated data ($G(z_i)$, $p_i = 0$) as the loss function (see Appendix A.1 for more details):

$$\begin{aligned} \mathcal{L}_{\text{GAN}} &= -[\mathcal{L}_{\text{BCE}}(y) + \mathcal{L}_{\text{BCE}}(G(z))] \\ &= \sum_{i=1}^{n}[\log D(y_i) + \log(1 - D(G(z_i)))]. \end{aligned} \tag{4.22}$$

Note that we prepare the same number of training and generated data for each training. In general, a sigmoid function is placed in the output layer of the discriminator to keep its output within $[0, 1]$. In this case, the loss function is called sigmoid binary cross-entropy. The objectives of the discriminator and generator are to maximize and minimize it, respectively. Thanks to adversarial training, the generator of the GAN can mimic more complex probability densities than other methods.

In the vanilla GAN described above, new data is generated from a random input. If we want to solve problems such as image segmentation or coloring with the CNN, network is required to generate an image corresponding to a particular input image. To solve these image-to-image problems, Isola et al. (2016) have proposed the conditional GAN (cGAN), in which the generator generates an image from an input image as $G : x \to G(x)$[1]. The objective of the generator is to generate an image $G(x)$ as close as possible to the ground truth, y. The discriminator of cGAN takes a pair of input, either (x, y) or $(x, G(x))$, and distinguishes the true pairs and fake (generated) pairs.

Like all the other CNNs, in cGAN, we reduce the size of the images in hidden layers by adopting stride greater than 1 to save computational costs. Fig.4.1 shows an example architecture of the generator of cGAN. If we want to get an output with the same size as the input, we connect a series of convolutional layers (encoder) followed by the same number of deconvolutional layers (decoder) with the same strides. While the high-resolution features extracted in the early layers could often be important in image transformation, they are not directly transferred to later layers when we use convolution alone. **Skip connection** is what makes this possible. Skip connection is an operation that concatenates the features generated by encoder layers with the features in the mirrored layer of the decoder. An encoder-decoder with skip connections is called U-Net.

To train the cGAN, we prepare a set of pairs of an input image and its expected output (ground truth), $\{x_i, y_i\}$. The loss function of the cGAN is

$$\mathcal{L}_{\mathrm{cGAN}} = \sum_{i=1}^{n} [\log D(x_i, y_i) + \log(1 - D(x_i, G(x_i)))]. \tag{4.23}$$

Isola et al. (2016) show that adding an L1 norm between the ground truth and the generated data

$$\mathcal{L}_{\mathrm{L1}} = \sum_{i=1}^{n} |y_i - G(x_i)| \tag{4.24}$$

makes learning more efficient. Note that the L1 norm encourages less blurring of the output than the mean squared error. The loss function is the sum of the cGAN loss and the L1 norm weighted by a hyperparameter λ:

$$\mathcal{L} = \mathcal{L}_{\mathrm{cGAN}} + \lambda \mathcal{L}_{\mathrm{L1}}. \tag{4.25}$$

---

[1] In general, random noise z is also input to the generator, but we omit it in our notation for simplicity.

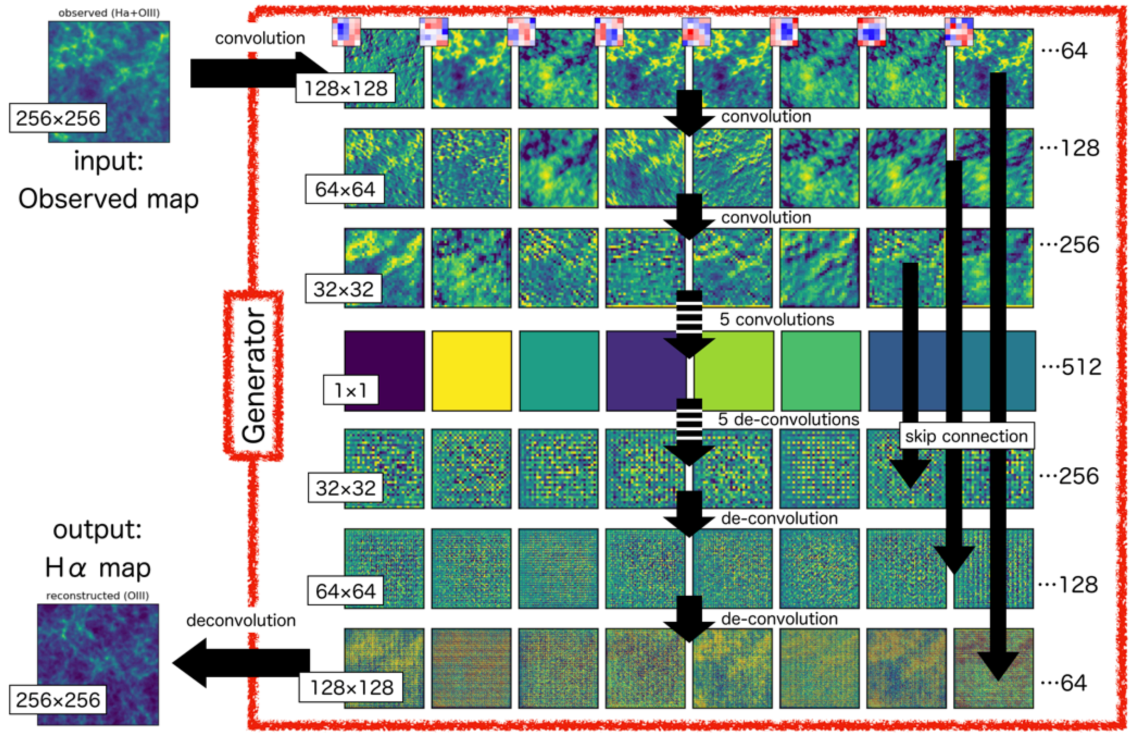


Figure 4.1: An example of the generator architecture that takes an observed map as input and reconstructs a Hα map. The encoder part consists of eight convolutional layers and the decoder part consists of eight deconvolutional layers (five among each of them are omitted in the figure). The size of the images in hidden layers are reduced by adopting stride 2 to save computational costs. Skip connections bring the outputs of the earlier layers to the corresponding later layers.

The objective of the generator (discriminator) is to minimize (maximize) this loss function.

One of the important extensions of the vanilla GAN is Wasserstein GAN (WGAN; Arjovsky et al. 2017). WGAN consists of two CNNs, generator (G) and critic (D), and is known to mitigate the training instability (see Appendix A.1). In the WGAN, we do not place a sigmoid function in the output layer of the critic (unlike the discriminator of the vanilla GAN) and thus it returns any value in $[-\infty, \infty]$. The loss function of WGAN is

$$\mathcal{L}_{\mathrm{WGAN}} = \frac{1}{n}\sum_{i=1}^{n}[D(y_i) - D(G(x_i))]. \tag{4.26}$$

The objective of the generator (critic) is to increase (decrease) the loss function. Arjovsky et al. (2017) suggest that the loss function (Eq.4.26) plays an approximately equivalent role to the Earth-Mover distance, which is continuous and differentiable everywhere, by allowing the weights of the critic to vary only in a small parameter space – most simply, by clipping all the weights to lie within $[-0.01, 0.01]$ after every gradient update. The critic developed in this way does not experience vanishing gradients unlike the discriminator of vanilla GAN and thus allows more stable training of the generator.

4.2 Methods: Training Data and Network Architecture

In this study, we develop CNNs to extract signals from noisy LIM data. We use a single spectral bin at $\lambda_{\rm obs} = 1.5\ \mu$m observed by SPHEREx. In near-infrared wavelength, there are several foregrounds such as Zodiacal light and diffuse Galactic light from interstellar dust, but they are spatially smooth and can be removed with several techniques (Doré et al. 2014). The biggest problem is remaining thermal noises.

The choice of the size and resolution of training data is important. Due to limited computational resources, there is a limit to the number of pixels that a machine can deal with, and it is difficult to achieve both high resolution and large area. With SPHEREx resolution of 0.1 arcmin, most galaxies are observed as point sources. To detect individual point sources, it is better to adopt as fine resolution as possible.[2] On the other hand, to extract extended faint signals, it may be better to degrade the spatial resolution and adopt a larger area instead. We thus prepare two different datasets with different angular resolutions and areas. We adopt the original resolution (0.1 arcmin) with an area $(0.85\ \rm deg)^2$ for the first training dataset and coarser resolution (2.0 arcmin) with an area $(8.5\ \rm deg)^2$ for the second training dataset. We refer to these datasets as “small” and “large” datasets, respectively.

We use cGANs to extract signals.[3] For simplicity, we consider a single emission line Hα from $z = 1.3$, which is observed at $\lambda_{\rm obs} = 1.5\ \mu$m. Our task is to reconstruct Hα intensity from the observed map (Hα + noise). There are two ways to make a machine learn this: to let it output Hα itself, or to let it output noise and reconstruct Hα intensities by subtracting the reconstructed noise from the observed image. We find that the signals are more accurately reconstructed when we let the network learn to reconstruct the noise. The same result is obtained in previous studies of denoising weak lensing maps using a cGAN (Shirasaki et al. 2019).

We build a cGAN based on Isola et al. (2016). The generator consists of eight convolutional layers and eight deconvolutional layers, and the discriminator consists of four convolutional layers. We include skip connection, batch normalization, and dropout. The hyperparameters of the generator and discriminator are listed in Table 4.1. The cGAN is trained with the loss function Eq.4.25. The hyperparameters λ, batch size, and epoch are determined with 10 validation data (see Appendix A.2 for more detail). We adopt $\lambda = 1000$ and 200 for “small” and “large” datasets, respectively, and batch size 4, and epoch 8 for both of them. The images (intensities) are normalized with 1.0×10^{-4} erg/s/cm^2/sr (“small”) and 1.0×10^{-6} erg/s/cm^2/sr (“large”) before input into the networks.

---

[2] The intensity of a point source per pixel is inversely proportional to the pixel area $l_{\rm pix}^2$, but the noise intensity per pixel is inversely proportional to $l_{\rm pix}$ as in Eq.3.25.

[3] We find that a simpler machine learning method using only the L1 norm as a loss function does not work well because it encourages the network to reconstruct smooth intensity maps. See Appendix A.2.

Table 4.1: The architecture of the generator and discriminator. For each convolutional (Conv.) or deconvolutional (Deconv.) layer, the number of filters N_{filter}, filter size, stride, presence of the batch normalization, dropout rate, and activation function are listed. The shape of filters is in a form of (width x, height y, features).

Generator

Layer	N_{filter}	Filter size	Stride	BN	Dropout	Activation
Conv. 1	64	(5, 5, 1)	(2, 2)	×	×	Leaky ReLU
Conv. 2	128	(5, 5, 64)	(2, 2)	○	×	Leaky ReLU
Conv. 3	256	(5, 5, 128)	(2, 2)	○	×	Leaky ReLU
Conv. 4	512	(5, 5, 256)	(2, 2)	○	×	Leaky ReLU
Conv. 5	512	(5, 5, 512)	(2, 2)	○	×	Leaky ReLU
Conv. 6	512	(5, 5, 512)	(2, 2)	○	×	Leaky ReLU
Conv. 7	512	(5, 5, 512)	(2, 2)	○	×	Leaky ReLU
Conv. 8	512	(5, 5, 512)	(2, 2)	○	×	Leaky ReLU
Deconv. 1	512	(5, 5, 512)	(2, 2)	○	0.5	Leaky ReLU
Deconv. 2	512	(5, 5, 512)	(2, 2)	○	0.5	Leaky ReLU
Deconv. 3	512	(5, 5, 512)	(2, 2)	○	0.5	Leaky ReLU
Deconv. 4	512	(5, 5, 512)	(2, 2)	○	×	Leaky ReLU
Deconv. 5	256	(5, 5, 512)	(2, 2)	○	×	Leaky ReLU
Deconv. 6	128	(5, 5, 256)	(2, 2)	○	×	Leaky ReLU
Deconv. 7	64	(5, 5, 128)	(2, 2)	○	×	Leaky ReLU
Deconv. 8	1	(5, 5, 64)	(2, 2)	×	×	tanh

Discriminator

Layer	N_{filter}	Filter size	Stride	BN	Dropout	Activation
Conv. 1	64	(5, 5, 2)	(2, 2)	×	×	Leaky ReLU
Conv. 2	128	(5, 5, 64)	(2, 2)	○	×	Leaky ReLU
Conv. 3	256	(5, 5, 128)	(2, 2)	○	×	Leaky ReLU
Conv. 4	512	(5, 5, 256)	(2, 2)	○	×	Leaky ReLU
Dense	1[a]			×	×	sigmoid

[a] Dimension of the output.

4.3 Extracted Signals from Noisy Maps

We first visually compare the true and reconstructed maps. Fig.4.2 shows an example of the reconstructed maps for the "small" dataset. From left to right, the observed map (Hα + noise), true Hα map, and reconstructed Hα map are shown. The images are smoothed for visibility. The locations and intensities of individual bright sources are well reproduced though the observed data is noise-dominated. For instance, when we pick up pixels above $3.5\sigma_n$ ($= 9.1 \times 10^{-6}$ erg/s/cm^2/sr) from observed maps, only $\sim$ 20 percent of the extracted peaks are true $3.5\sigma_n$ peaks. We find that when we choose pixels with $> 3.5\sigma_n$ in the reconstructed maps by cGAN, $\sim$ 70 percent of them are true $3.5\sigma_n$ peaks. This indicates that the machine selectively extracts more plausible signals.

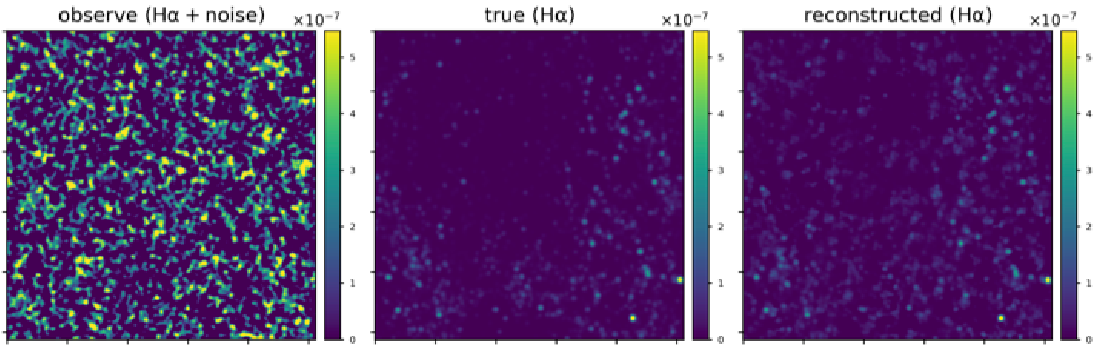


Figure 4.2: The observed map (Hα + noise, left), the true Hα map (middle), and the reconstructed Hα map (right) for the "small" dataset. The side length is 0.85 deg. The maps are smoothed for visibility. The intensities are in units of erg/s/cm^2/sr.

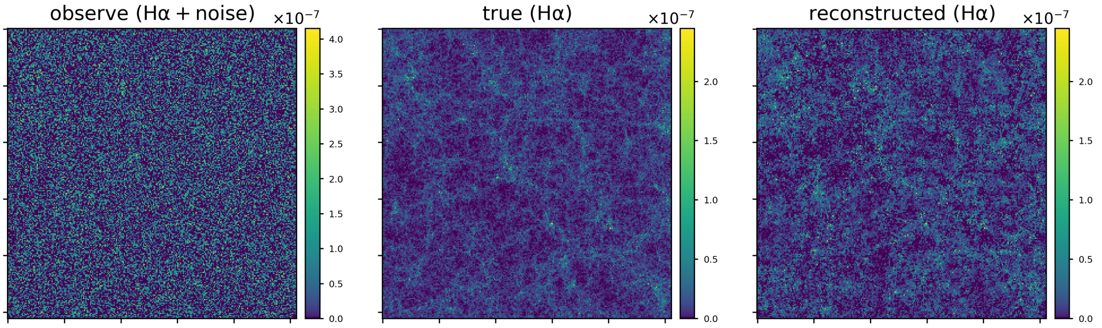


Figure 4.3: Same as Fig. 4.2, but for the "large" dataset. The side length is 8.5 deg.

Fig.4.3 shows an example for the "large" dataset. We find that the detectability of individual emission sources is degraded compared to the "small" dataset. On the other hand, the positions of voids and large-scale filamentary structures are relatively well reproduced. Such differences of datasets can also be seen in the pixel-by-pixel comparison between the true and reconstructed maps shown in Fig.4.4. In the case of the "large" dataset, the network properly reproduces the intensities of truly bright signals but predicts bright intensities even for some pixels where there are only faint true emissions. This is because the large noise makes the pixel-by-pixel reconstruction difficult.

Summary Statistics

The statistics including the power spectrum and PDF are important for studying cosmology as well as galaxy formation and evolution. We examine the reconstruction of them using 1000 test data. To compute the statistical values, we use the bagging method (see Section 4.1). In particular, we prepare five different datasets generated with different realizations in PINOCCHIO and train five cGANs using each dataset. We compute the statistics from the maps obtained from each network and take the average of them.

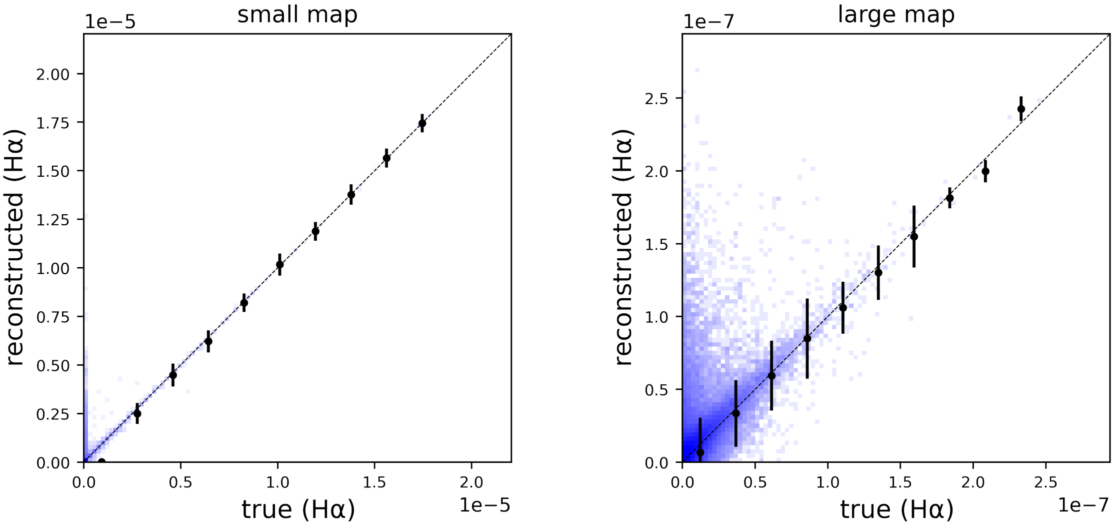


Figure 4.4: The pixel-by-pixel comparisons between true and reconstructed images that are shown in Fig.4.2 and 4.3 for “small” (left) and “large” (right) datasets. The error bars indicate the variance of the reconstructed intensity in each true intensity bin.

Fig.4.5 shows the variance of the observed (light shade), true (dark shade), and reconstructed (error bars) PDFs over 1000 test data, for “small” (left) and “large” (right) datasets. In both cases, the PDFs are reconstructed reasonably well even in domains where the observed data is dominated by noise. As we have seen above, the values of the individual pixels are not always properly reconstructed for both cases. Nevertheless, the networks successfully reconstruct the PDFs. This is because they learn the typical properties (including the statistical properties) of the training data. The bagging process also contributes to the reconstruction of the statistics. It is remarkable that the cGANs are able to *reproduce* the PDFs well even though we do not directly train them to learn the statistics. Here we note that we use data that was not used in the training, and thus the network does not just memorize the training data. We will discuss how it behaves when different data from the training data is input in Section 4.4.1.

We also compute the power spectrum. In Fig.4.6, the gray (orange) shade regions and black (red) error bars indicate the variance of the two-dimensional power spectrum of true and reconstructed Hα intensity maps over 1000 test data for the “small” (“large”) dataset. The gray dashed line is the noise power spectrum

$$P_n = A_{\rm pix}\sigma_n^2, \tag{4.27}$$

where $A_{\rm pix} = l_{\rm pix}^2$ is the pixel size and σ_n is the noise level per pixel. The power spectra are reproduced well although these scales are dominated by noise. The power spectrum of the line intensity map scales with the squared mean intensity $\overline{I}^2$ on large scales. On small scales, the

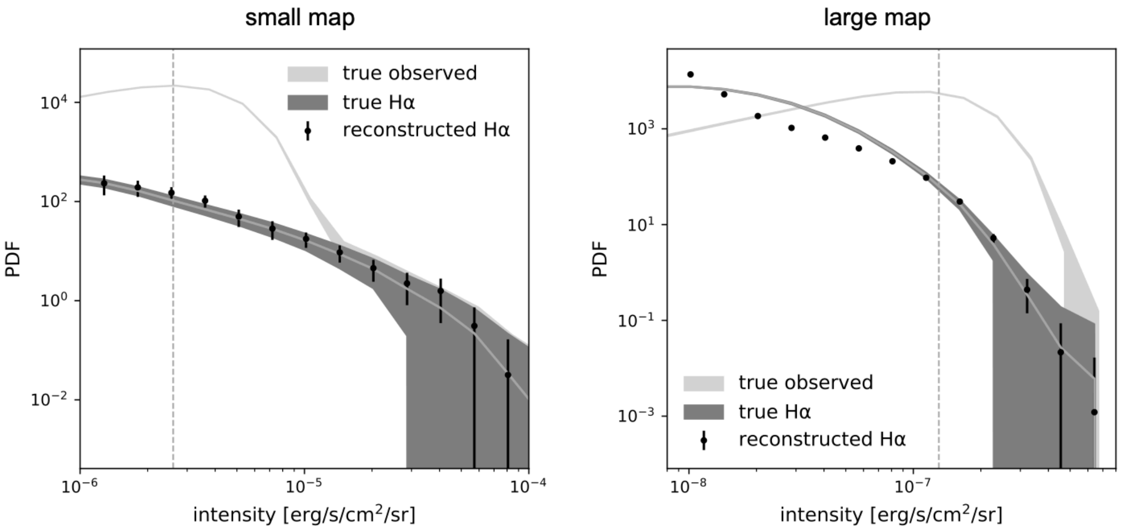


Figure 4.5: The point distribution function (PDF) for the "small" (left) and "large" (right) datasets. The light and dark shaded regions and error bars show the variance of the PDF of the observed maps, true Hα maps, and reconstructed Hα maps over 1000 test data. The vertical dashed line indicates the 1-σ_n.

shot noise that is proportional to the mean of the squared luminosity $\overline{L^2_{\rm halo}}$

$$\text{shot noise} = \bar{I}^2 \frac{\overline{L^2_{\rm halo}}}{\bar{L}^2_{\rm halo}} \propto \overline{L^2_{\rm halo}} \tag{4.28}$$

is observed (Eq.2.18). Therefore, the amplitude of the small-scale power is mainly determined by the bright galaxy population.

In the "small" case, the galaxy shot noise is well reproduced. This is consistent with the fact that the bright galaxies are properly reconstructed as seen above. The large-scale power at $k \lesssim 0.5$ arcmin^{-1}, however, is underestimated. This regime corresponds to the 2-halo term (the clustering of neighbor halos), which is determined by the overall intensity distribution including the faint sources. The underestimation may be because of the small boxsize of the training data. In the case of the "large" dataset, the shape of the large-scale power spectrum is properly reconstructed. We should note that the reconstructed maps of the "small" dataset still contain large-scale clustering information. If one uses the galaxy number density derived from the reconstructed Hα maps over a large volume, 2-halo terms (Eq.2.13) would be detected as in the usual galaxy surveys. Such reconstructed data can be used to detect cosmological large-scale signatures.

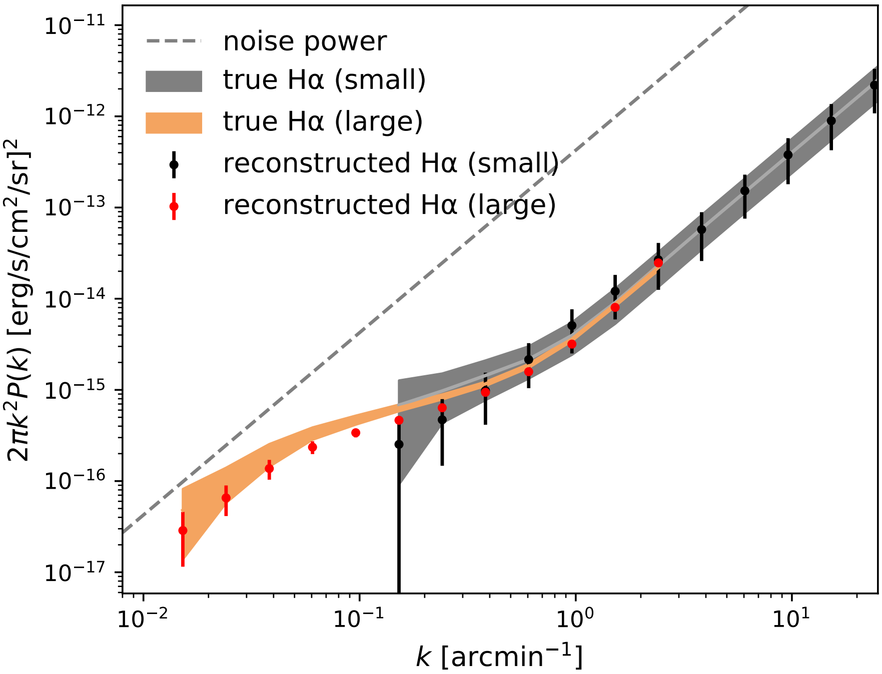


Figure 4.6: The power spectra for the "small" (gray/black) and "large" (orange/red) datasets. The shaded regions and error bars show the variance of the power spectra of the true and reconstructed maps over 1000 test data. The dashed line is the noise power spectrum (Eq.4.27).

4.4 Discussions

4.4.1 Different Emission Line Models

In machine learning, it is important to investigate how well the signal extraction works for data that is different from the training data. The properties of the emission-line galaxies at $z \lesssim 3$ has been studied in galaxy surveys (e.g., Khostovan et al. 2015; Sobral et al. 2013), but there are still some uncertainties. In this section, we test the network using datasets generated with three different line models from the training dataset.

The first two datasets are generated by uniformly amplifying or reducing the overall line intensity. We generate Hα intensity maps by multiplying a constant value $c_{\mathrm{H}\alpha}$ and leave the noise intensity unchanged

$$I_{\mathrm{obs}} = c_{\mathrm{H}\alpha} I_{\mathrm{H}\alpha} + I_{\mathrm{noise}}. \tag{4.29}$$

We adopt $c_{\mathrm{H}\alpha} = 0.5$ and 2. In addition, we adopt an emission line model by Silva et al. (2018), where they assume the SFR-halo mass relation derived by a semi-analytical model (Guo et al. 2013). Silva et al. (2018) use a fitting function

$$\mathrm{SFR}(M) = 10^a \left(\frac{M}{M_1}\right)^b \left(1 + \frac{M}{M_2}\right)^c, \tag{4.30}$$

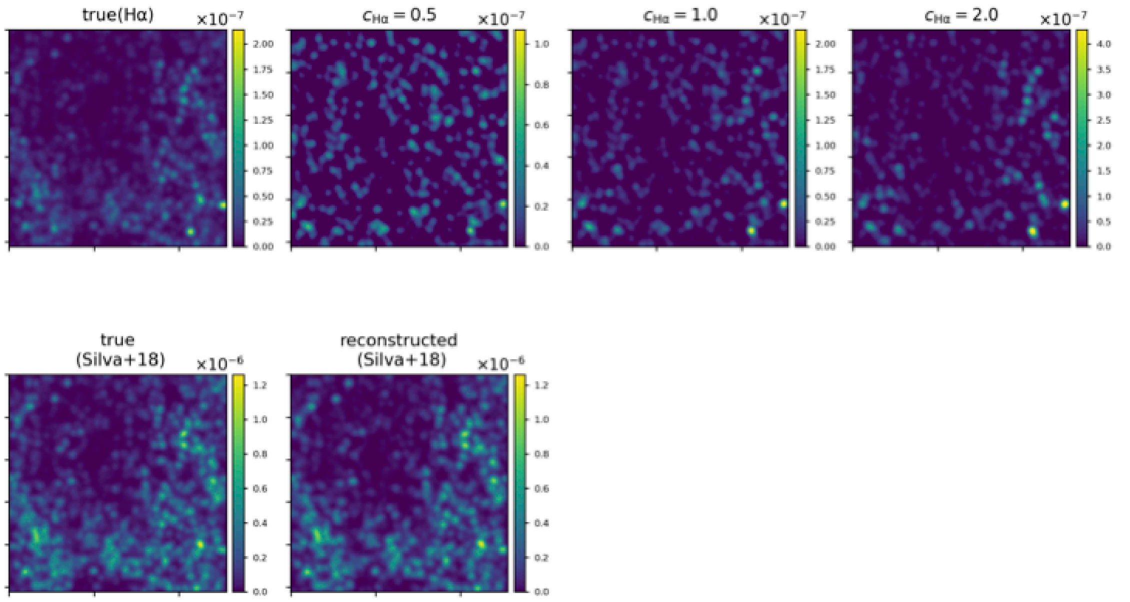


Figure 4.7: Examples of the reconstructed maps for the different line models. The top left panel is the true map of the original dataset, and then the reconstructed maps for $c_{\mathrm{H}\alpha} = 0.5$, 1.0 (original), and 2.0 are shown from left to right. The color bars are adjusted so that their true maps look the same as the true map of the original dataset. Bottom panels are the true and reconstructed maps for the line model by Silva et al. (2018). They are generated with the same cosmological realization.

where they find that a set of parameters $a = -0.85$, $b = 2.7$, $c = -3.0$, $\log M_1/\mathrm{M}_\odot = 8$, and $\log M_2/\mathrm{M}_\odot = 11.7$ reproduces the simulation results well at $z = 1$, and compute the Hα luminosity as (Kennicutt 1998)

$$L_{\mathrm{H}\alpha} = 1.26 \times 10^{41} \frac{\mathrm{SFR}}{\mathrm{M}_\odot/\mathrm{yr}} \times 10^{-A_{\mathrm{H}\alpha}/2.5} \ \mathrm{erg/s}, \tag{4.31}$$

where $A_{\mathrm{H}\alpha} = 1$ is the dust extinction magnitude. To generate realistic intensity maps, we add a 0.2 dex scatter[4] to the line luminosity. The obtained luminosity-halo mass relation is flatter than ours at large halo masses, and there are more bright Hα emitters. We refer to this model as the Silva18 model in the following.

Fig.4.7 shows examples of the "small" maps reconstructed by the network trained with the original training dataset. We use maps generated with an identical realization. For the first two datasets ($c_{\mathrm{H}\alpha} = 0.5$ and 2), the color bars are adjusted so that their true maps look the same as that of the original dataset shown in the leftmost panel on the top row. The bright signals are properly reconstructed regardless of the model details. This indicates that the network learns the property of the noise rather than the signal.

We also compare the reproducibility of the PDFs (Fig.4.8). We find that the networks reproduce the bright end of the PDFs except for the most faint model ($c_{\mathrm{H}\alpha} = 0.5$). We can see

---

[4] This value is based on the scatter in the SFR-halo mass relation derived by Guo et al. (2013).

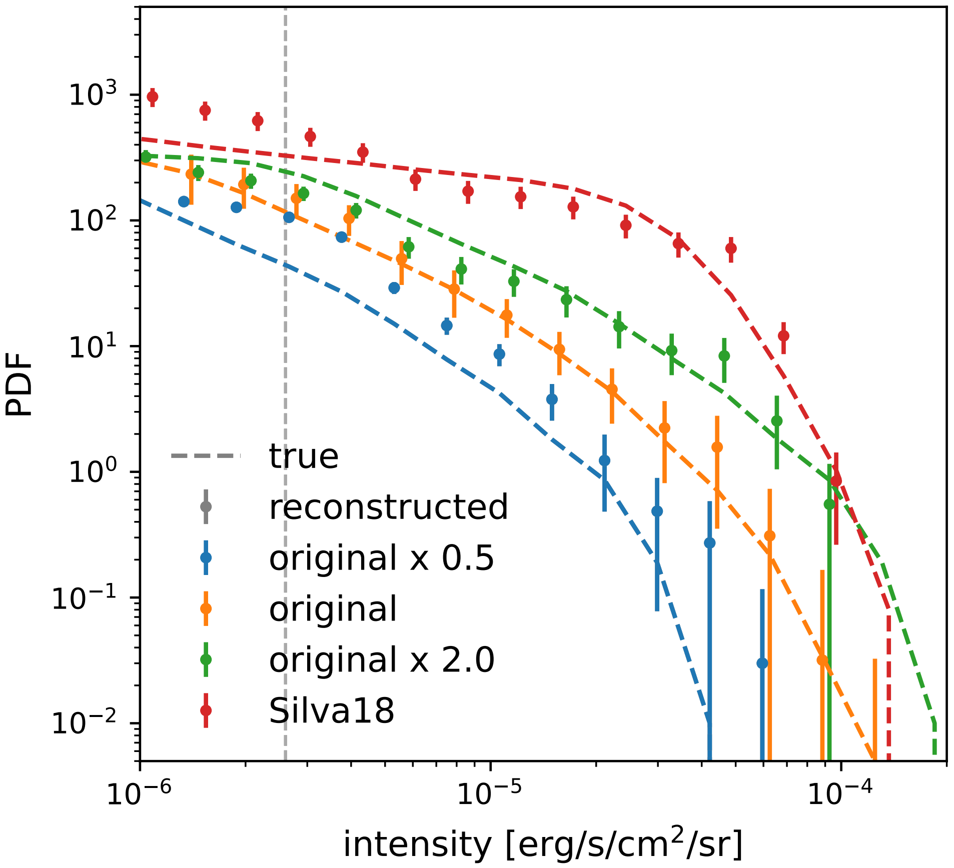


Figure 4.8: The probability distribution functions of the true (dashed line) and reconstructed (error bars) maps for the different line models. The dashed gray vertical line indicates the 1-σ noise level.

a clear difference between those of the original and Silva18 models even at the noise dominant fainter end at $\sim 3 \times 10^{-6}$ erg/s/cm^2/sr. It indicates that our network distinguishes different emission models to some extent. The reconstructed PDFs can thus be used to study the population of line emitters with a method proposed by Breysse et al. (2017). If we assume that individual pixels are dominated by only one halo (though this is not always the case), we can reproduce the luminosity function. The intensity 3×10^{-6} erg/s/cm^2/sr corresponds to $\sim 7 \times 10^{41}$ erg/s at $z = 1.3$, which is comparable or lower than the luminosity limit explored by galaxy surveys at similar redshifts (e.g., Khostovan et al. 2015). As the LIM observations cover much larger volumes, obtained information could be complementary to the results of galaxy surveys.

On the contrary to the "small" dataset, the reproducibility is considerably degraded when we perform the same test for the "large" dataset. When we input test data with emission line models significantly different from the original one, the network only detects the approximate locations of large clusters and voids but does not properly reconstruct small-scale distributions or the overall intensities of the signals. The statistics of the maps such as the power spectra and PDFs are also not reproduced. We find that the reproduced large-scale power spectra for the emission line models with $c_{\mathrm{H}\alpha} = 0.5$ and 2 have similar amplitudes as the training data with $c_{\mathrm{H}\alpha} = 1$. This can be attributed to the overfitting of the network to the training data. To apply our network to actual observational data and derive cosmological information from it, we should address such a problem.

Increasing the variety of data used for training may solve this problem. However, the training generally becomes more difficult for more diverse training data, requiring higher computational cost and detailed parameter tuning. One of the simpler ways to tackle this problem is to use multiple networks. A demonstration of such a method has been done by Acquaviva et al. (2020), with a simple regression task in which they estimate the stellar mass from the observed galaxy spectrum. They train multiple networks with multiple datasets generated with different physical models and then build another machine learning model that measures how far the observed data is from each training dataset. They find that the systematic errors of the outputs of each network can be estimated from the learned distance measure. Such techniques will allow us to select the appropriate network (training dataset) and to determine whether the observed signals are those expected or not. To apply the networks to actual observational data in the future, it is crucial to train the networks with various emission line and noise models as well as various cosmological models.

4.4.2 Choice of Training Data

We have looked at the differences between using small images with fine resolution and large images with coarse resolution. There are advantages and disadvantages of each. One might consider that the best performance is achieved when we train the network using large images with fine resolution. However, the size of the images (number of pixels) that can be dealt with is limited by the available computational power. The computational memory required for training depends on the number of pixels as well as details of the network architecture such as the number of convolutional filters and layers. After some experiments, we find that our network can be trained on our GPU with 24.5 GB memory if the image size is 2048^2 or smaller.

Also, to train a machine with large images, it would take a much longer time to complete a sufficient number of training steps. The training time is roughly proportional to the total number of pixels if the network architecture is fixed. For example, it takes several days for our network to complete 8 epochs for 2048^2 images. Such problems could partly be solved by carefully tuning the architecture of the network, but then its performance could be degraded.

To let a machine learn the long-range dependencies of the images more efficiently, tuning stride parameters of convolutional layers could be important. We may also use other machine learning models such as the state-of-the-art Transformer model (Dosovitskiy et al. 2020; Vaswani et al. 2017), which captures relationships between distant pixels with fewer layers. In any case, with the current generation computers and currently available network architecture, one should carefully choose the training data depending on one's purpose. One should use fine resolution data to obtain a robust distribution of bright emission sources, and large images to detect faint extended signals.

4.5 Conclusion

We find that cGANs can be used to separate signals from noisy LIM observed data. In particular, by using a high-resolution dataset for training, we can detect point sources with better accuracy than simply picking up pixels with more than a few-σ. The detection limit corresponds to $\sim 10^{42}$ erg/s. The reproduced PDFs can be used to study the population of emission-line galaxies. They will provide data that reinforce the star formation history traced by UV and far-infrared observations and information on the ISM of the galaxies at the intermediate redshifts.

We note that a conventional noise reduction method using a Wiener filter (Zaroubi et al. 1995) does not work well for denoising LIM data. It removes the noise of known property using the local information around each pixel of the image. The LIM signals are sparse and locally not distinguishable from the Gaussian noise. Our network uses deep layers to capture the features extended over the larger scales, which may allow proper extraction of the signals.

As well as the reconstructed maps for the "large" training dataset, the reconstructed bright sources in the "small" dataset trace the LSS of the universe. We can study cosmology by detecting the BAO and RSD on large scales and possibly the scale dependent bias on ultra-large scales owing to primordial non-Gaussianity. We note that when we reconstruct the larger-scale intensity distribution by connecting the reconstructed "small" maps, somewhat flat maps are reconstructed, and the faint filaments and voids are not as clear as those reconstructed by the "large" network. Therefore, it is better to use a larger dataset if one wants to do analyses using these faint structures.

There is a tradeoff between the resolution and field size of the training data. If we could increase the number of pixels that a network can deal with, the network may be able to robustly detect both the individual bright sources and faint extended sources at the same time. If we have ~ 1 TB GPU memory, we can deal with 8192^2 pixels (i.e., ~ 200 deg^2 with a resolution of 0.1 arcmin) with our current network architecture. It is also required to tune the hyperparameters (e.g., the stride and number of layers) and to explore a parallel GPU implementation to reduce training costs in future studies.

Our network may learn the properties of the noise as well as those of the signal. Therefore, the validity of the noise model used for the training data could be important. In this study, we only use a simple noise model since the properties of the actual observational noise of SPHEREx are not yet known. In the SPHEREx observation, the all-sky survey will be performed by shifting a 3.5 deg × 3.5 deg field of view consisting of linear variable filters gradually (Doré et al. 2014). Because of this observational strategy, there would be fluctuations of the noise level with scales corresponding to each shift (~ 10 arcmin). Other contaminations such as diffuse Galactic light and Zodiacal light should also be incorporated into the noise model. We should take into account such details of the observational noise to apply the networks to actual observation data in the future. To this end, methods of constructing training data using actual observation data (e.g., Chen et al. 2018) could be useful. To construct a proper noise model, it will also be useful to check if the training data is statistically similar to the actual observational data using methods such as those proposed by Acquaviva et al. (2020) (see Section 4.4.1). If the noise level of each pointing can be properly estimated in the observation, it would also be an interesting avenue to incorporate such a noise level map as one of the inputs to the network so that it can perform the noise reduction based on the actual survey design.

In this section, we have considered only one emission line. Even if there are multiple emission lines, we may be able to reconstruct the locations of the bright signals or the superposition of the LSSs of multiple redshifts by training machines with appropriate training datasets. After extracting emission signals from noisy maps, however, it is not straightforward to separate the individual emission line signals. Separation of them is important to avoid systematic errors in the obtained cosmological and astrophysical constraints. In the next section, we will consider the separation of multiple line signals.

Chapter 5

Signal Separation from Confused LIM Data

In the previous chapter, we have considered extracting emission line signals from the noisy LIM maps. While we have only considered one emission line, noise reduction can also be done in essentially the same manner, even in the presence of multiple emission lines. In conventional galaxy surveys, the emission lines are identified by observing the individual galaxies in detail to detect multiple emission lines. In the LIM observations, such an identification is difficult because of the lower sensitivity and coarser angular/spectral resolutions. There is a so-called line confusion problem.

Several solutions have been proposed for the line confusion problem, but they are mostly statistical methods. While they provide an average picture of the universe, it would be more informative if signal distribution itself can be extracted. To study environmental effects on galaxy formation and evolution, for example, it is crucial to directly detect large-scale structures. Even when we take a cross-correlation with the other maps such as faint 21-cm intensity maps at the EoR, it would be better to have no interlopers in the line intensity maps to avoid systematics.

In this chapter, we develop a machine learning application to solve the line confusion problem. The LSSs at different redshifts have different properties including typical distances between galaxies, typical intensity, and amplitude of the fluctuation. A machine can separate them if it learns these differences appropriately. We first describe the line confusion problem and several proposed solutions in Section 5.1. After that, we describe our methods in Section 5.2 and the obtained results in Section 5.3. We then discuss the results in Section 5.4. The contents of this chapter are based on Moriwaki et al. (2020).

5.1 Line Confusion Problem in Line Intensity Mapping Observations

Line confusion is a serious problem in LIM observations, in which several different emission lines are redshifted into the same observed wavelength. For two emission lines with rest-frame

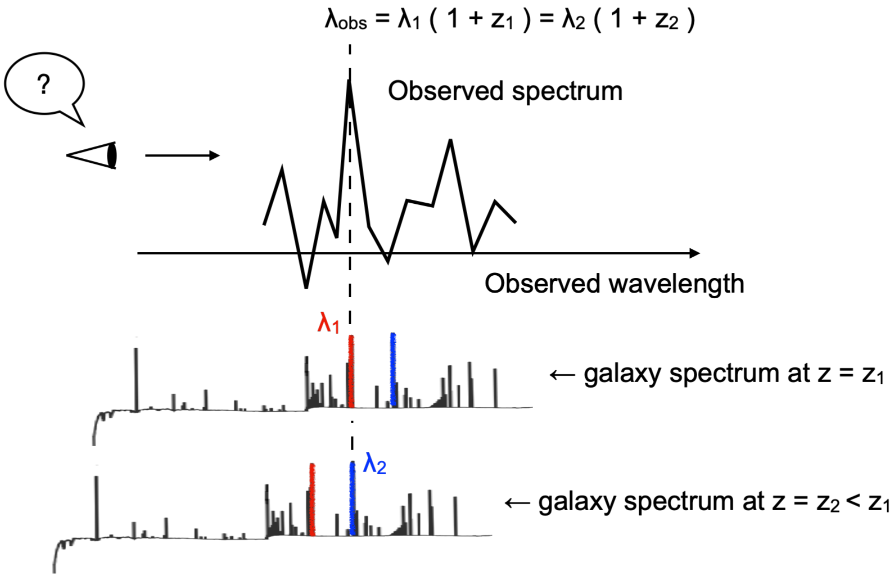


Figure 5.1: Two emission lines with rest-frame wavelengths λ_1 and λ_2 emitted from redshifts z_1 and z_2 observed at the same wavelength if they satisfy Eq.5.1. It is difficult to distinguish their contribution in the LIM observation because of the low sensitivity and low resolution.

wavelengths λ_1 and λ_2 emitted from redshifts z_1 and z_2, if they satisfy

$$\lambda_1(1+z_1) = \lambda_2(1+z_2), \tag{5.1}$$

then they are observed at the same wavelength (Fig.5.1). Several interlopers exist in most of the spectral regimes (Fonseca et al. 2017). For instance, the Hα emissions from $z = 1.3$ and the [OIII] emissions from $z = 2.0$ are both observed at $\lambda = 1.5$ μm, contaminating each other (Fig.1.2). Because of the low sensitivity and low resolution, it is generally difficult to separate them individually.

One of the most common methods to reduce the contribution of the interlopers is to take cross-correlation. The cross-correlation signal is obtained by combining two different observations covering the same region. For two intensity distributions $I_1(\boldsymbol{x})$ and $I_2(\boldsymbol{x})$, the cross-power spectrum is defined as

$$\langle \tilde{I}_1(\boldsymbol{k})\tilde{I}_2(\boldsymbol{k}')\rangle = (2\pi)^3\delta_{\rm D}^3(\boldsymbol{k}+\boldsymbol{k}')P_{1,2}(k). \tag{5.2}$$

Suppose that the signals $I_1(\boldsymbol{x})$ are contaminated by observational noises or signals from a different redshift, $I_n(\boldsymbol{x})$, and that the other map $I_2(\boldsymbol{x})$ (e.g., galaxy distribution or intensity map obtained from different observations) trace the large-scale distribution at the same redshift as $I_1(\boldsymbol{x})$. Since the contaminant $I_n(\boldsymbol{x})$ and the other signal $I_2(\boldsymbol{x})$ do not have any correlation, their cross-power spectrum is reduced to zero

$$P_{1,2}(k) = P_{1,2}(k) + P_{n,2}(k) \sim P_{1,2}(k) \tag{5.3}$$

if we take an average over many modes. Many theoretical predictions of LIM cross-correlation signals have been made (e.g., Cheng & Chang 2021; Heneka et al. 2017; Lidz et al. 2009), and several observations have already detected the cross-correlation signals between LIM data and galaxies (e.g., Chang et al. 2010; Keenan et al. 2021).

Masking bright pixels is also an effective method to detect distant (faint) signals (Gong et al. 2014; Silva et al. 2018). The auto-power spectrum of such distant signals can be recovered reasonably well when only small regions are masked compared to the entire survey area, but it is also known that the results could strongly depend on the masking threshold. A combination of masking and cross-correlation is also often used. In such analysis, null tests using completely uncorrelated sources (e.g., sources at different redshifts) are also performed to check the validity of the method (e.g., Kakuma et al. 2021). Another method to separate interlopers is to use the anisotropy of the power spectra. The wavenumbers along the angular and redshift directions ($k_\perp$, $k_{||}$) are measured using the angular diameter distance and Hubble parameter. We observe an isotropic power spectrum[1] only when appropriate distance measures are adopted. If we adopt distance measures defined at redshift z_1 to measure the signals at redshift z_1, we observe an isotropic power spectrum. If the signals are contaminated by interloper signals at redshift $z_2 \neq z_1$, we then observe an anisotropy due to the z_2 signals. This anisotropy can be used to separate signals from two redshifts that are far enough from each other (Cheng et al. 2016; Lidz & Taylor 2016).

Recently, Cheng et al. (2020) devised a pixel-by-pixel line de-confusion method that uses spectral information. They demonstrate that they can detect signals brighter than a given threshold in noisy LIM data by detecting multiple emission lines in a spectrum under the sparse approximation. Their method is only applicable when a sufficiently wide wavelength range is covered by the observation and more than two emission lines from the same galaxy are observed. We also note that they do not include the clustering information in their analysis.

5.2 Methods: One-to-Many Translation Network Architecture

We consider the LIM data at a single spectral filter $\lambda = 1.5$ μm. As we have seen in Fig.3.2, the Hα line from $z = 1.3$ and [OIII] line from $z = 2.0$ are dominant at this wavelength and the [OIII] intensity is smaller than the Hα on average. We ignore the other subdominant lines and the observational noises. We generate the training data with an area of 1.7 deg on a side and a resolution of 0.4 arcmin. By adopting such a wide area with a coarse resolution, we aim at letting the machine learn the difference between the LSSs at two different redshifts.

Our task is to separate two emission signals from their superpositions. We build a machine that equally treats each emission. We use two cGANs that reconstruct Hα and [OIII] from the observed map and train them together. The pairs of the generator and discriminator are denoted by (G_1, D_1) for Hα and (G_2, D_2) for [OIII]. For each cGAN, we adopt the same network architecture as the previous chapter (Table 4.1). The two generators are supposed to retrieve similar information at deep layers, so the encoder of the generators is shared (see Fig.5.2). Such one-to-many reconstruction networks have also been proposed in previous studies on, e.g., the

---

[1] We ignore the RSD here.

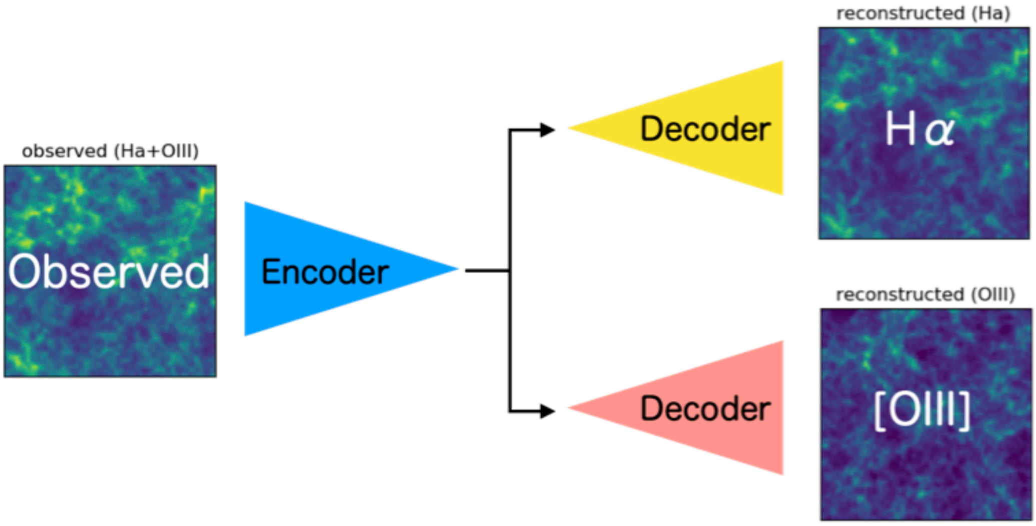


Figure 5.2: A schematic picture of the generators. They share the encoder and have independent decoders. During the training, the learnable parameters of the generators are updated simultaneously.

separation of reflected and transmitted scenes in images (Lee et al. 2018). With this network architecture, we can easily add other emission lines and other contaminant sources in future studies.

We find that the pixel-wise reproducibility is degraded when we use raw data (see appendix A.2). We thus apply Gaussian smoothing with a beam size of 1.2 arcmin as a preprocessing of the training and test data. In addition to the cGAN loss function (Eq. 4.25), we use the L1 norm between the observed data and the reconstructed Hα +[O III] data

$$\mathcal{L}_{\rm tot} = \frac{1}{n}\sum_{i=1}^{n} |x - G_1(x) - G_2(x)|, \tag{5.4}$$

as a loss function, where x is the observed map. The indexes $i = 1$ and 2 correspond to Hα and [O III], respectively. The total loss function is

$$\mathcal{L} = \sum_{i=1,2} [\mathcal{L}_{\rm cGAN}[G_i, D_i] + \lambda_i \mathcal{L}_{L1}[G_i]] + \lambda_{\rm tot}\mathcal{L}_{\rm tot}. \tag{5.5}$$

After several experiments, we choose $\lambda_1 = \lambda_2 = \lambda_{\rm tot} = 100$. The images are normalized with 2.0×10^{-7} erg/s/cm^2/sr.

5.3 Separation of Multiple Emission Line Signals

Here we describe the results of the separation of two emission line signals. Figure 5.3 shows an example of reconstructed maps. The left panel in the middle row is the sum of the two reconstructed maps, and the bottom panels show the difference between the true and reconstructed maps. The observed map is dominated by Hα signals as we have expected from the mean

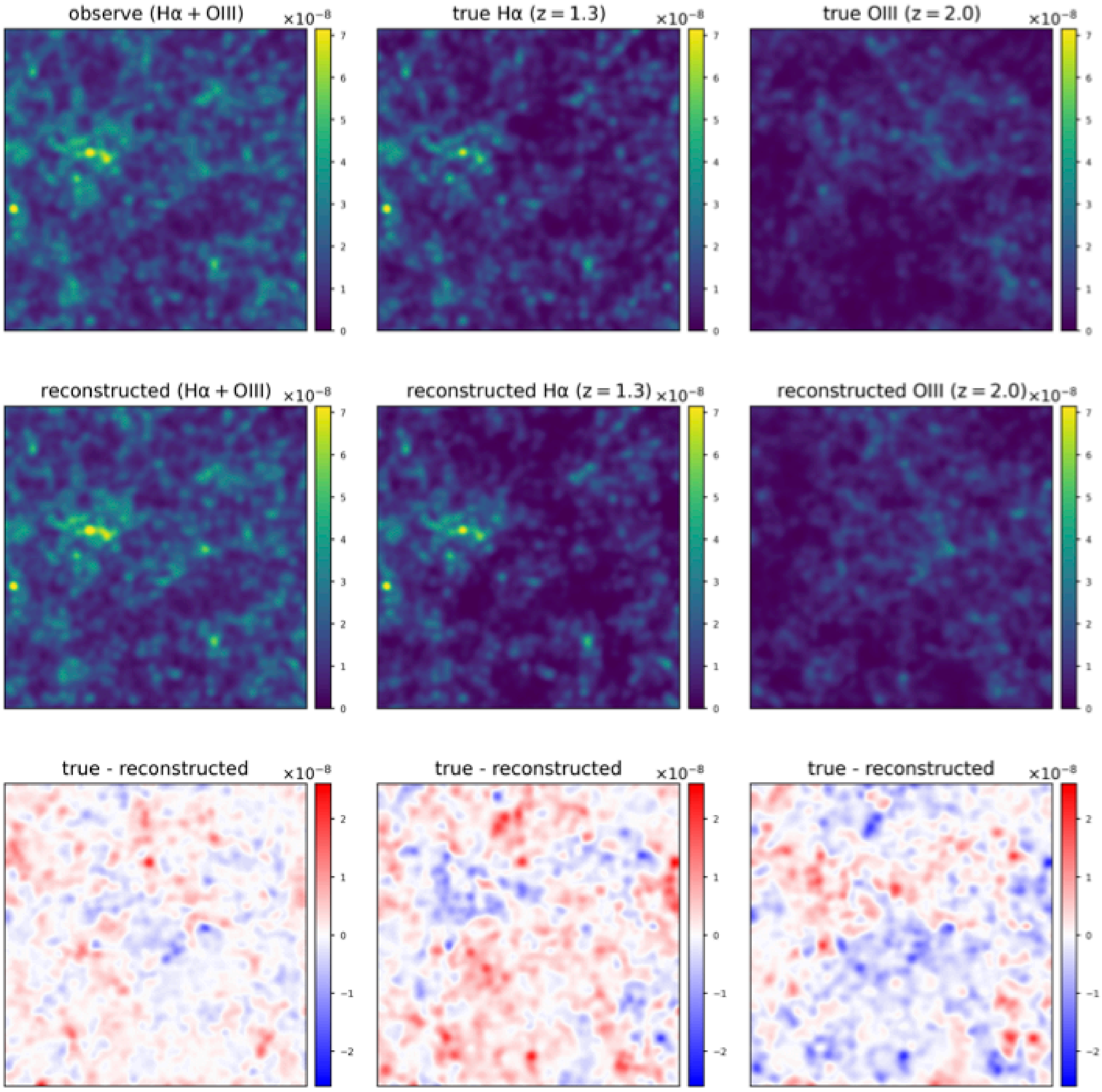


Figure 5.3: An example of signal reconstruction. 1.7 deg on a side. The intensities are in units of erg/s/cm^2/sr. *Top*: from left to right, the observed (Hα + [OIII]) intensity map and the true Hα and [OIII] intensity maps. *Middle*: the reconstructed Hα (middle) and [OIII] (right) intensity maps and the sum of them (left). *Bottom*: the difference between the true and reconstructed intensity maps.

intensity. In the reconstructed Hα map, the cluster and void regions are more clearly defined than the observed map. Even the Hα signals with comparable intensities to the bright [OIII] structures are well separated, indicating that the networks separate the signals by capturing not only the intensity of each pixel but also the extended structure. The brightest peaks of [OIII] are also well reconstructed, especially in the regions where there are only weak Hα signals. While overall separation is done very well, fainter structures including filaments traced by [OIII] emissions are not always well reproduced, so we need to be careful when studying such structures using the reconstructed maps.

Locations of the bright peaks are important probes of the LSS. We thus examine the detectability of these peaks. We define the peaks as the local maxima larger than a certain threshold: 4×10^{-8} erg/s/cm^2/sr for Hα and 2×10^{-8} erg/s/cm^2/sr for [OIII]. With these thresholds, ~ 10 peaks are detected in the true map per 1 deg^2 on average. We regard that the true and reconstructed peaks are matched when they are found within 4 arcmin of each other, which corresponds to ~ 2 Mpc at $z = 1 - 2$. The numbers of peaks found in true and reconstructed Hα ([OIII]) maps are $N_{\rm true} = 20653$ (26770) and $N_{\rm rec} = 22892$ (21480) over 1000 test data. Among them, there are $N_{\rm correct} = 17279$ (6970) peaks matched correctly. The recall ($N_{\rm correct}/N_{\rm true}$) is 0.84 (0.26) and the precision ($N_{\rm correct}/N_{\rm rec}$) is 0.76 (0.32) for H$\alpha$ ([OIII]).

We then compare the reconstructed statistics. We compute power spectra and PDFs using bagging analysis with five different networks as in the previous chapter. Fig.5.4 shows the power spectra of the true and reconstructed maps.[2] The networks properly reproduce the large-scale power at ~ 0.1 arcmin^{-1}, where the 2-halo term (the clustering of neighbor halos) is observed. The small scale powers are also well reproduced. This indicates that the network learns the point spread function (Gaussian smoothing) as well as the mean number density of bright sources.

The statistics of the test data are widely distributed because of the field variance. We thus further examine whether the network tells the difference between individual data. In the top panel of Fig.5.4, we show the difference between reconstructed ($P_{\rm rec}$) and true ($P_{\rm true}$) power spectra divided by the square root of the variance of the true power spectrum ($\sigma_{\rm true}$). The error bars indicate the variance of the difference between the reconstructed and true power spectrum, $\sigma_{\rm rec-true} = \sigma(P_{\rm rec} - P_{\rm true})$, divided by $\sigma_{\rm true}$. The variance of the difference, $\sigma_{\rm rec-true}$, is given by

$$\sigma^2_{\rm rec-true} = \sigma^2_{\rm rec} + \sigma^2_{\rm true} - 2C_{\rm rec,true}, \tag{5.6}$$

where $C_{\rm rec,true}$ is the covariance between the reconstructed and true power spectra. As seen from the bottom panel in Fig.5.4, the distribution of the power spectra are in good agreement, i.e., $\sigma_{\rm rec} \sim \sigma_{\rm true}$. If the reconstruction strategy of the networks was such that they randomly reproduce the statistics regardless of the actual properties of individual data, i.e., $C_{\rm rec,true} = 0$,

---

[2] Unlike what we have seen in the previous chapter, the shot noise is not seen because of the smoothing. We note that for a known point spread function, we can recover the unsmoothed power spectrum $P(k)$ from the smoothed power spectrum $P_{\rm sm}(k)$. For Gaussian beam with smoothing scale σ, it is given by $P(k) = \exp(k^2\sigma^2)P_{\rm sm}(k)$.

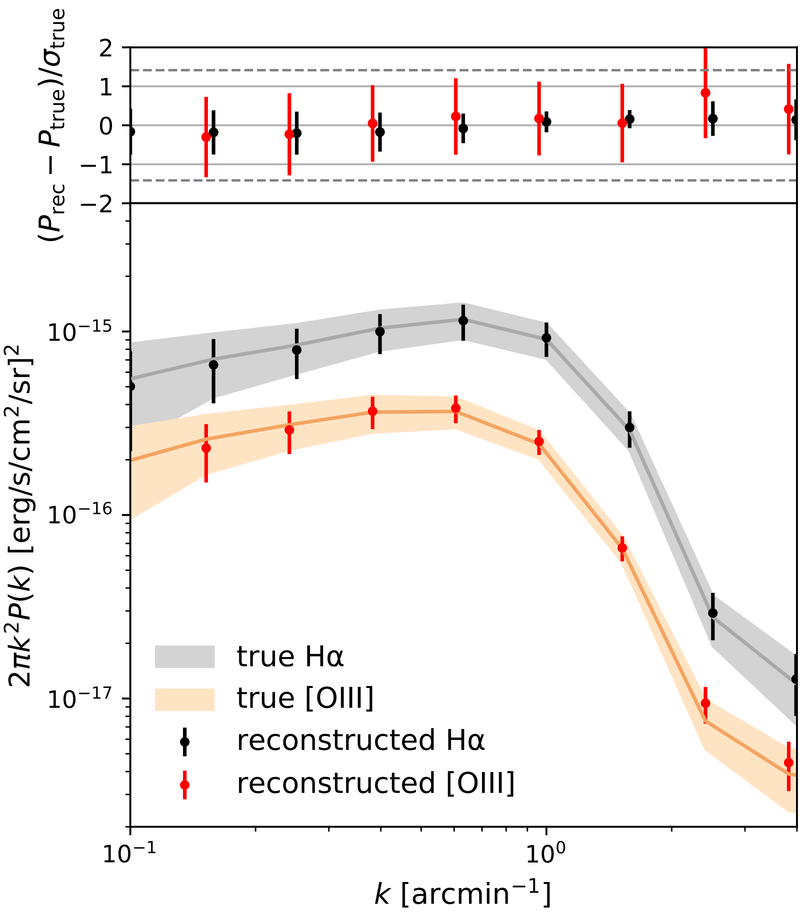


Figure 5.4: The true and reconstructed power spectra. The shaded areas indicate the distribution of the power spectra of the true Hα (gray) and [OIII] (orange) maps over the 1000 test data, and the error bars show those of the reconstructed Hα (black) and [OIII] (red) maps. The error bars on the top panel indicate the variance of the difference between the reconstructed and true power spectrum, $\sigma_{\rm rec-true} = \sigma(P_{\rm rec} - P_{\rm true})$, divided by $\sigma_{\rm true}$. The horizontal dashed lines indicate $\pm\sqrt{2}$.

then the variance of the difference would be $\sigma_{\rm rec-true} \sim \sqrt{2}\sigma_{\rm true}$. The dashed lines in the top panels of Fig.5.4 indicate $\pm\sqrt{2}$. On large scales, the error bars are smaller than this for both Hα and [OIII]. This means that the networks do capture the characteristics of individual test data to some extent.

The same is true for the reconstructed PDFs, shown in Fig.5.5. The deviation (upper panel) is smaller than $\pm\sqrt{2}$ over a wide intensity range. We also examine the reproducibility of the mean intensities. Fig.5.6 shows the comparison between the true and reconstructed mean intensities. The mean intensities of individual data are reproduced with statistical errors of a few percent for Hα and ~ 20 percent for [OIII]. Such errors are reduced by observing a larger area (e.g., 200 deg^2 for SPHEREx). The successful reproduction of the faint end of the PDFs and the mean intensities are partly owing to the absence of the noise, but we emphasize that

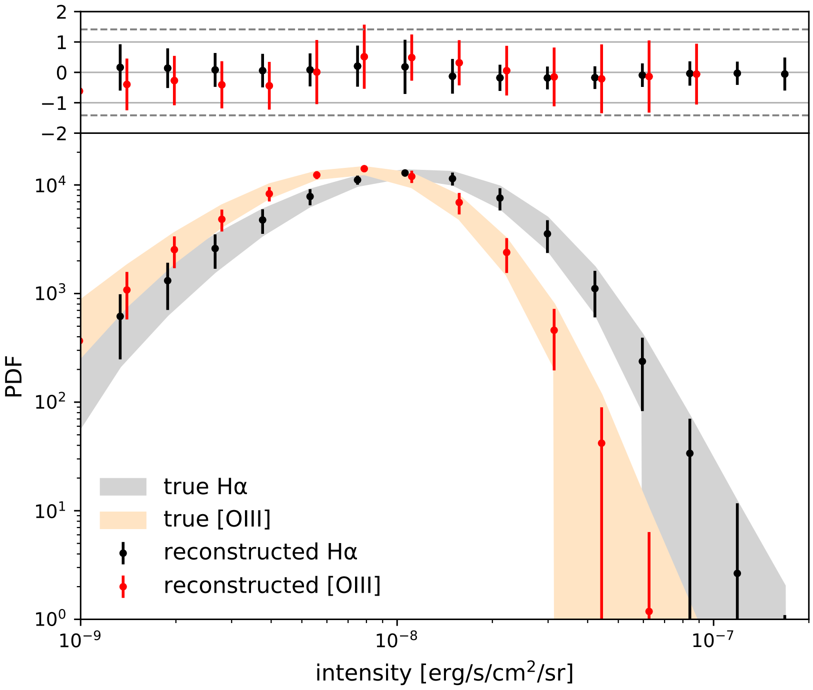


Figure 5.5: The true and reconstructed PDFs. The symbols indicate the same as Fig.5.4.

such signal separation is not possible with any method other than ours even in the absence of the noise.

5.4 Discussions

5.4.1 Different Emission Line Models

As we have seen in Chapter 4, the networks could fail to deal with a dataset that is quite different from the training dataset. We thus test the networks with different emission line models. To this end, we uniformly increase the line emissivities by 10 percent. Fig.5.7 shows the reproduced mean intensities of Hα (top) and [OIII] (bottom) for 100 test data. The intensities are successfully reproduced for both Hα and [OIII]. We also find that the detectability of the peaks does not significantly change even when the line emissivities are changed by up to 10 percent. Such robustness may be partly because of the intrinsic scatter in the training dataset, but we note that the mean intensities are appropriately reproduced even for the intensity maps that come outside the range of the training data. While the reconstructions are performed well for $\sim$ 10 percent change, we find that the reconstructed maps deviate from the true maps when we make a larger change. This is consistent with the results of the "large" dataset in Chapter 4. To properly reconstruct such outliers, we should use a dataset with a wider model range and/or multiple networks as discussed in Section 4.4.1.

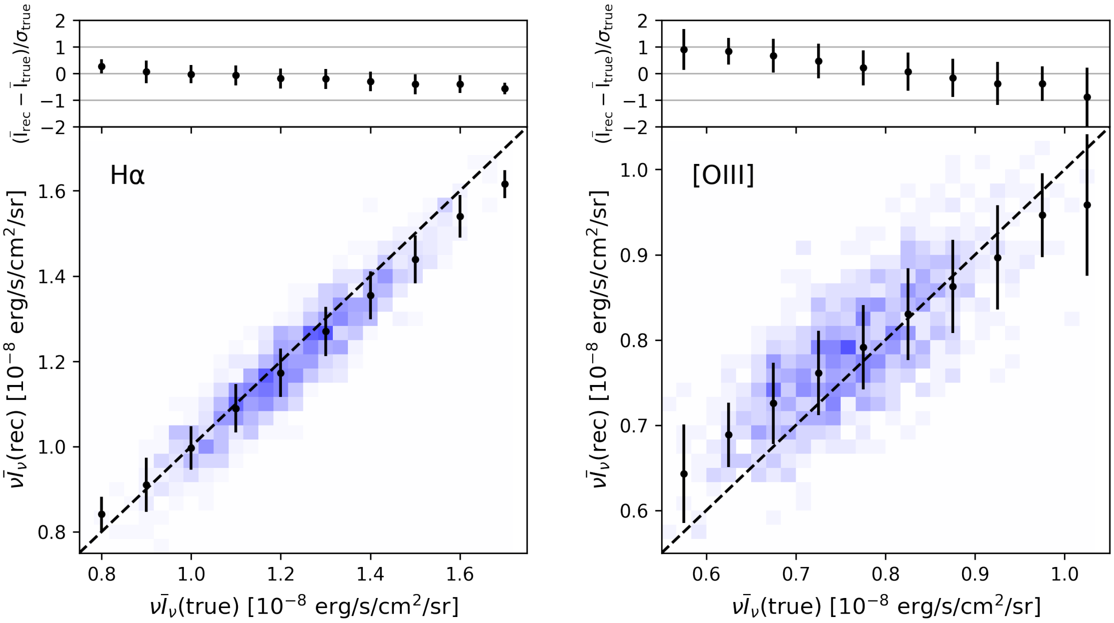


Figure 5.6: The true vs reconstructed mean intensity for Hα (left) and [OIII] (right) maps. The difference between them divided by the square root of the variance of the true intensities are shown on the top.

5.4.2 Combining Multiple Networks

Our networks detect peaks with precision as high as ~ 0.8 for the Hα maps and ~ 0.3 for the [OIII] maps. Detecting bright peaks, i.e., the locations of the clusters, is important for studying the environmental effects at the high-redshift universe. Also, reconstructed peaks will be good targets for deeper follow-up observations. For such purposes, peak catalogs with higher precision are required. In other words, we need to know the reliability of the detected signals.

Previous works develop several methods to evaluate the uncertainties of the individual reconstructed pixels. Kendall & Gal (2017), for example, devise a machine that estimates aleatoric and epistemic uncertainty on a task to estimate the depth of the objects in images. The aleatoric uncertainty is the uncertainty due to the observational noise. The epistemic uncertainty is caused by the randomness of the model and its training process. Kendall & Gal (2017) train the machine to estimate the aleatoric uncertainty (at the same time as it estimates the depth) and compute the epistemic uncertainty by manually varying the model parameters.

We can do a similar analysis to the epistemic uncertainty prediction method of Kendall & Gal (2017) by making use of the five networks used for bagging analysis. We perform signal detection with each of the five networks trained with different datasets and select the peaks that are detected by all the networks. These peaks are considered to be more reliable (more precise) than the peaks detected by a single network. With this method, 13453 (2539) peaks are selected for Hα ([OIII]) maps. Among them, 12157 (1434) peaks are matched with the

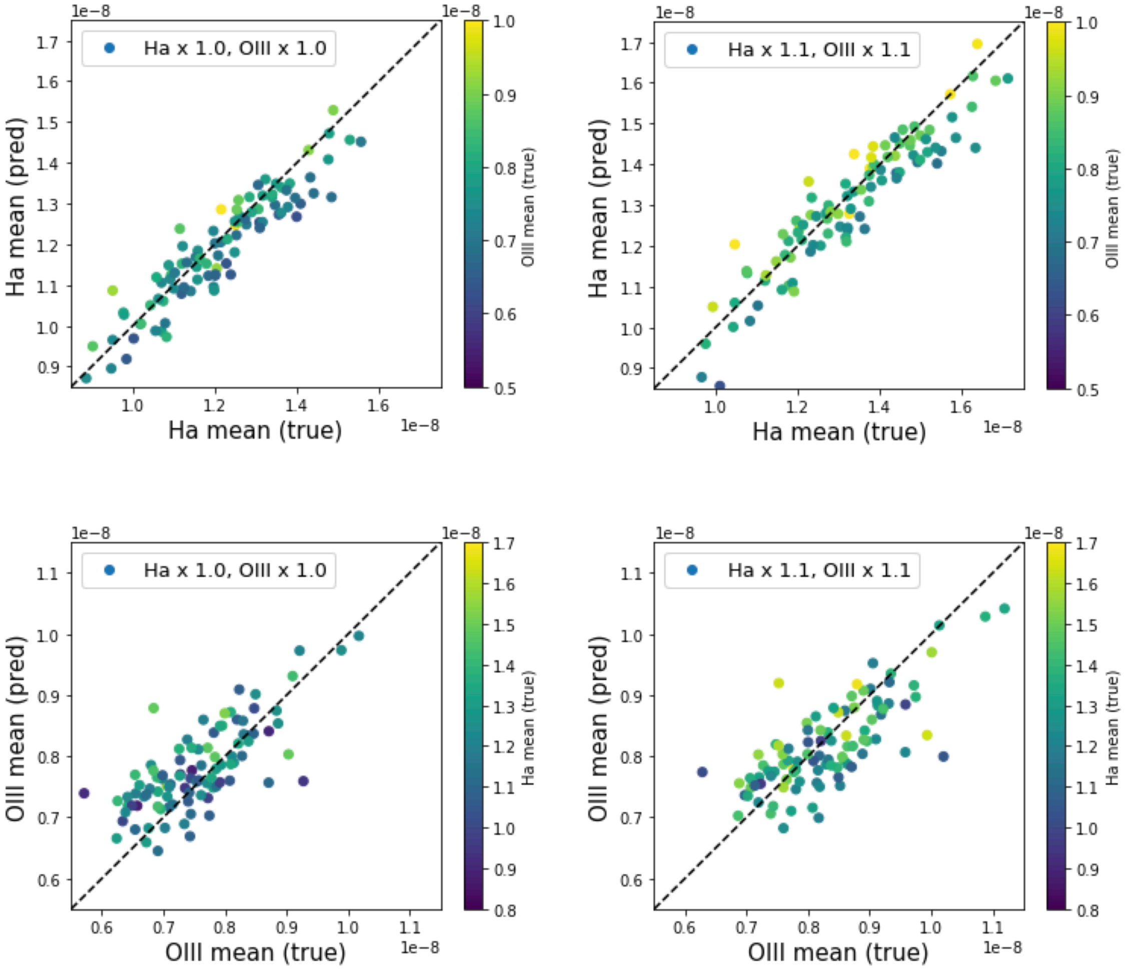


Figure 5.7: The reproduced mean intensities of Hα (top) and [Oɪɪɪ] (bottom) for 100 test data. The left panels show the results for the original dataset, and the right panels show the results when the line intensities are uniformly increased by 10 percent. The colors indicate the mean intensity of the other line emission.

true peaks, which corresponds to a precision of 0.90 (0.57). We find that the precision of the peak detection is indeed improved, and the obtained peak catalogs can be used for studying environmental effects and individual galaxies.

5.5 Convolutional Filters and Hidden Layers

So far, we have seen that the cGANs can separate two LSSs very well. Here, we investigate how a machine learns the features of the LSS. Looking at the convolutional filters and hidden layers is one of the most primitive ways of understanding CNNs. Fig.5.8 shows the convolutional filters and the outputs of the first layer. The input map is convolved with 64 convolutional filters with kernel size (5, 5). The red and blue colors of the filters correspond to the positive and negative values, respectively. The circularly symmetric filters are considered to be capturing

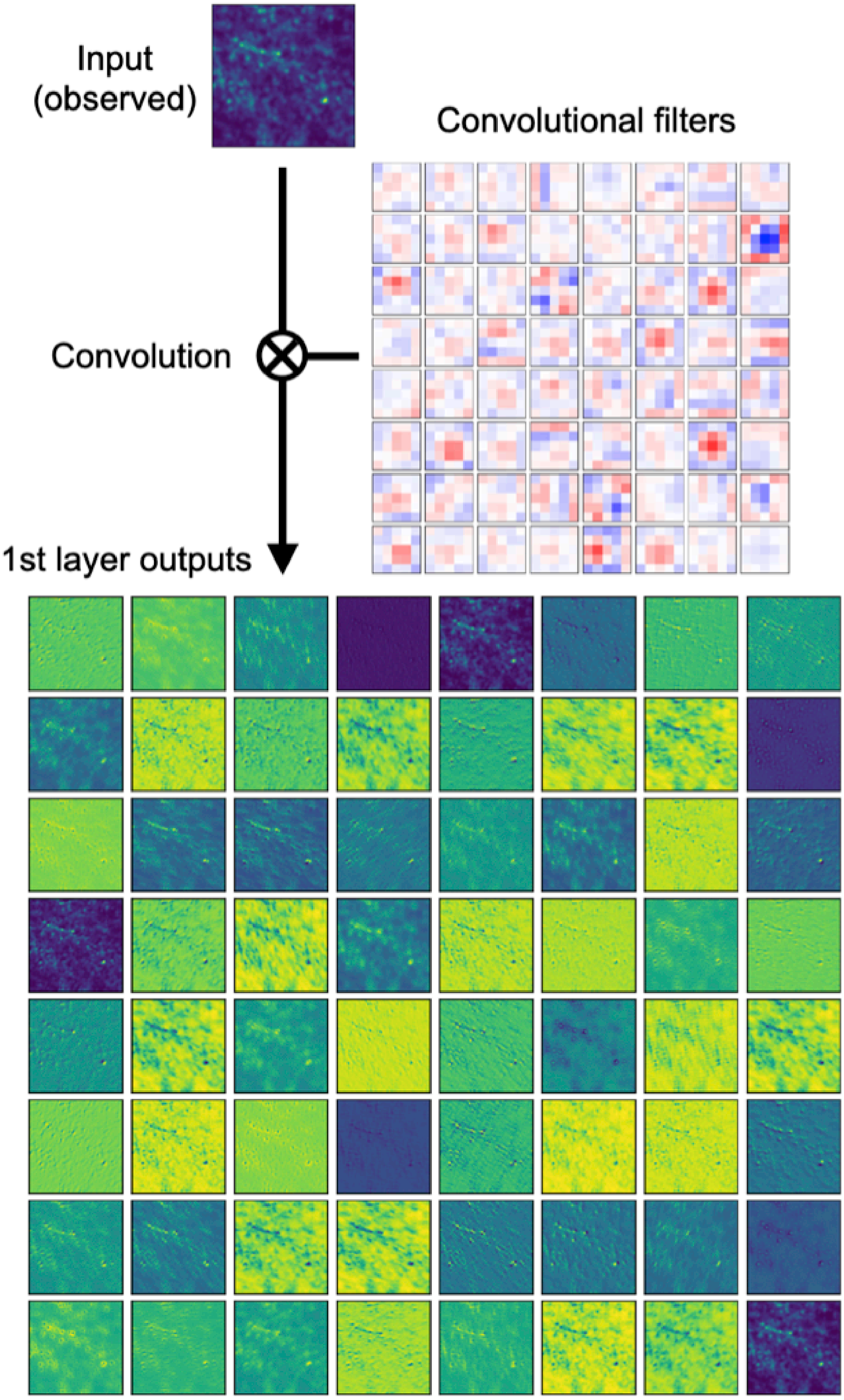


Figure 5.8: The convolutional filters of the first layer of the generator. Red and blue colors correspond to the positive and negative values, respectively. An example of the outputs of the first layer is also shown. Subtraction of biases and application of an activation function is omitted in the figure. We note that large pixel values of a given filter do not always imply the importance of the filter in the final output.

the Gaussian filter that we use in image preprocessing. In the output, we observe various feature maps: some have bright peaks highlighted, some are smeared in certain directions, some have inverted pixel values, and some are almost identical to the input. The filamentary structures are also emphasized in several feature maps. The feature maps that are almost identical to the input may be used to reproduce the small-scale structures in the outputs by being conveyed directly to the last layer through the skip connection. The feature maps in the first layer are also passed to the deeper layers and more complicated and/or larger-scale features are extracted. For instance, the fluctuations with small amplitude can be solely extracted if the network subtracts the feature maps with bright peaks highlighted from those that are almost identical to the input. Information on the distances between individual clusters (bright peaks) can also be extracted in deeper layers. The network may use the extracted features to tell differences between the LSSs at two redshifts in, e.g., the typical intensities, amplitude of the fluctuations, typical distances between galaxies, and typical size of voids.

5.6 Conclusion

In this chapter, we have proposed to solve the line confusion problem in the LIM observations using machine learning. We have devised, for the first time, a method for pixel-by-pixel separation of the LSSs at different redshifts. We have developed a one-to-many CNN that equally deals with individual emission lines. It allows us to easily extend our method to apply data with more emission lines, noise, and other contaminants. Our method does not require spectral information and thus can be applied to every observed wavelength. Such signal separation is only possible with our methods.

We demonstrate our method using mock observational data assuming SPHEREx observation. Our networks properly assign the observed peaks to each redshift with precision of 0.76 (Hα at $z = 1.3$) and 0.32 ([O III] at $z = 2.0$). By combining multiple networks, we can build peak catalogs with higher precision (0.90 and 0.57). The networks reproduce the statistics such as the power spectra, PDFs, and mean intensities as well. We find that the networks learn to distinguish individual $(1.7\ \mathrm{deg})^2$ maps rather than just learn the typical properties of the training dataset.

We look at the convolutional filters of the first layer of the generator. The networks extract various features from the observed map, which may be able to capture the differences between the LSSs at two redshifts. We also find that the reproducibility is degraded when the overall intensity is considerably changed. Conversely, the mean intensity or the amplitude of the fluctuation may play an important role in the networks' separation strategy.

Our method can be used for future observational data by SPHEREx and any other telescopes targeting different wavelength regimes after removing the observational noises using the method devised the previous chapter. It may also be directly applied to data from more future observations such as CDIM. The data extracted by our method can be used for studying galaxy formation and evolution as well as cosmology. Obtained peak catalogs of high precision will provide us good targets for deeper follow-up observations and allow us to study environmental effects on galaxy formation and evolution.

Chapter 6

Signal Extraction from 3D LIM Data

In the previous chapters, we have used two-dimensional LIM information. In actual observations, spectral data is also obtained, which can provide additional information. In this chapter, we use three-dimensional CNNs to extract signals from three-dimensional intensity data cubes containing spectral and angular distribution information. Three-dimensional CNNs are also used in recent studies for analyzing simulation data (e.g., Zhang et al. 2019) and the 21-cm line intensity maps (e.g., Prelogović et al. 2021).

In the application of machine learning to astronomical and physical data, better performance can be achieved by manually giving physical information to the machine (Breen et al. 2020; Karniadakis et al. 2021). In this study, we inform physical information. In particular, we inform the difference between rest-frame wavelengths of two emission lines to a network so that it learns the existence of two emission lines at different wavelengths. The contents of this chapter are based on Moriwaki & Yoshida (2021).

6.1 Methods

6.1.1 Data Preparation

We consider noisy line intensity maps within 1 - 2.5 μm observed by SPHEREx and consider two line-emission signals, Hα and [OIII]. We do not include the other subdominant interlopers, but if they are included, they could provide additional information. To prepare mock multi-spectral observation data cubes (light cones), we use the past-light-cone algorithm in PINOCCHIO. This algorithm replicates the simulation box and arranges the boxes by connecting the periodic boundaries to fill a light cone. We carefully choose the line of sight direction so that the same halo does not appear more than once in the light cone. The line luminosities of halos are computed as described in section 3.3. We generate the halo mass-luminosity tables at $z = 1.0$, 1.3, 2.0, 3.0, and 4.4 with IllustrisTNG, and interpolate them to compute the luminosities of the halos at redshifts between them.

We adopt the original SPHEREx angular resolution (0.1 arcmin) to put more focus on point

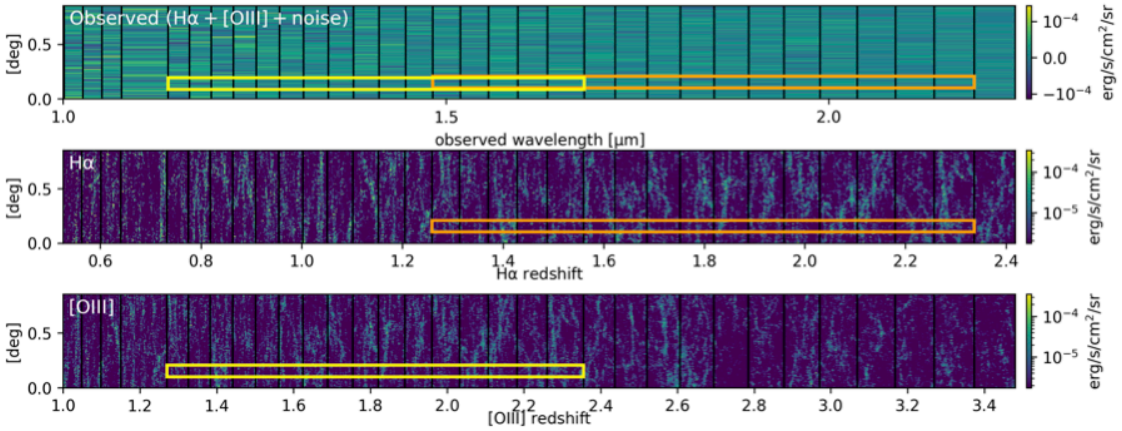


Figure 6.1: An example of the light cones generated by PINOCCHIO. We generate the Hα (middle) and [OIII] (bottom) intensity distributions with the same realization. The same LSSs are seen in each of them. By binning the sum of them with the spectral resolution of SPHEREx and adding thermal noise, we obtain the mock observational intensity maps (top). The black solid lines indicate the spectral resolution of SPHEREx. The wavelength range of the training data are indicated by yellow and orange boxes. They cover roughly the same redshift range of Hα and [OIII] lines, respectively.

source detection (see discussion in Chapter 4). We adopt 64 pixels for each angular direction. The angular size of the training data is thus 6.4 arcmin. After binning the sum of the Hα and [OIII] intensity maps, we add instrumental noises. We adopt the resolutions and noise levels of the SPHEREx observation.[1] Fig.6.1 shows an example of the obtained mock observational light cones. The black vertical lines represent the borders of the SPHEREx filter. The spectral resolution of the SPHEREx is mostly $R \sim 41$ in this wavelength regime. We avoid using the spectral bins that have different resolutions, such as those at $\lambda_{\rm obs} \sim 1.1\ \mu$m.

6.1.2 Physics-Informed Network Architecture

In general, the distribution of galaxies is correlated not only in the angular direction but also in the wavelength direction. However, the physical length corresponding to the spectral resolution is much larger than that of the angular resolution. The spectral resolution $R \sim 41$ corresponds to ~ 50 Mpc, while the angular resolution is ~ 50 kpc at the redshifts of our interest. Thus the correlation along the spectral direction is considered to provide different information from the angular correlation. We note that in the CDIM observation, the spectral resolution is ~ 10 times better, and it could be more informative.

The multi-spectral data contains another important information. Multiple emission lines from a galaxy are observed at multiple wavelengths. The observed Hα and [OIII] line signals from the same galaxy at redshift z are separated by $d\lambda = (1+z)(\lambda_{\rm H\alpha} - \lambda_{\rm [OIII]}) = 1556\text{Å}(1+z)$ in the spectral direction. For an observation that adopts a constant spectral resolution $R \equiv$

---

[1] https://github.com/SPHEREx/Public-products

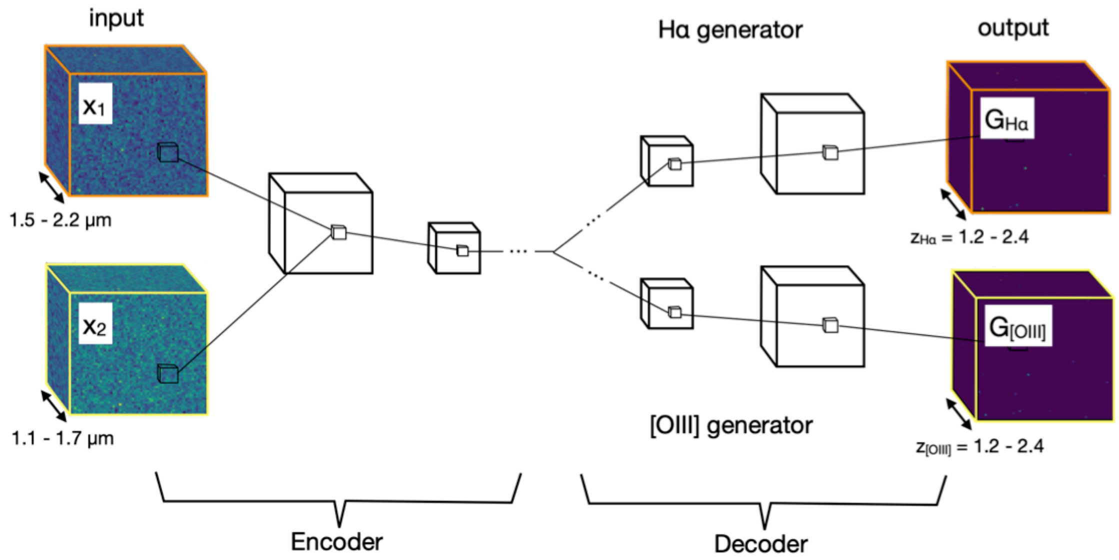


Figure 6.2: The generator architecture. They take two data cubes as input with different wavelength ranges: 1.48 - 2.19 μm and 1.14 - 1.68 μm. The Hα and [OIII] generators share the encoder consisting of four convolutional layers and have individual decoders consisting of four deconvolutional layers. We include the skip connections as well as the dropout and batch normalization.

$\lambda_{\rm obs}/d\lambda_{\rm bin}$ = const., where $d\lambda_{\rm bin}$ is the width of the spectral bin, the two emission lines from the same galaxy are always separated by the same number of filters regardless of the redshift of the galaxy. If the networks make use of this co-existence of two emissions, it may more easily separate the signals.

Such long-range dependencies can in principle be learned by sufficiently deep CNNs, but we find, after some experiments, that it is difficult to extract the signals by simply applying a three-dimensional CNN to the observed data cube. One way to make the networks more efficiently learn the long-range dependencies is to adopt convolutional kernels that are long in the spectral direction, but such a method would require a large amount of computational memory. It is also difficult to extend this method to deal with other emission lines. We thus take a different approach here.

Fig.6.2 illustrates our generator architecture. As in Chapter 5, two generators for Hα and [OIII] share the encoder part. Our generators take two portions of the observed data with different spectral ranges. The two inputs are separated by a wavelength $d\lambda = (1+z)(\lambda_{\rm H\alpha} - \lambda_{\rm [OIII]})$. In particular, we use the observed data cubes within 1.48 - 2.19 μm and 1.14 - 1.68 μm, covered by 16 spectral bins as input data. We denote them x_1 and x_2, respectively. The input data, (x_1, x_2), has $64 \times 64 \times 16 \times 2$ pixels in total. The wavelength ranges of x_1 and x_2 correspond to almost the same redshift range $z = 1.25 - 2.4$ when they are converted with the rest-frame wavelengths of the Hα and [OIII] lines, respectively. We note that the corresponding spectral bins do not always cover the same redshifts because of the finite spectral resolution. In SPHEREx observation, the redshifts of corresponding spectral bins correspond to each other with $dz/z < 0.01$ over the wavelengths of our interest. The two input data cubes x_1 and x_2 are indicated by the orange and yellow boxes in Fig.6.1. They contain the same underlying large-scale structure traced by Hα and [OIII] lines. From these inputs, the generators reconstruct the Hα and [OIII] intensities.

The network architectures are specified in Table 6.1. We first use the vanilla cGAN, but we find that learning is not stable and the reproducibility varies greatly depending on when learning is stopped and on small differences in hyperparameters. We thus develop a conditional extension of WGAN, which is known to learn more steadily. For the loss function, we use Eq.4.26 with an additional L1 term

$$\mathcal{L}_i = D_i(x_i, y_i) - D(x_i, G(x_1, x_2)) + \lambda_i |y_i - G_i(x_1, x_2)|, \tag{6.1}$$

where y_i is the true Hα ($i = 1$) or [OIII] ($i = 2$) intensity map. We adopt $\lambda_1 = \lambda_2 = 100$ after some experiments (Appendix A.2). We find that the results do not significantly change depending on the choice of these hyperparameters. The line intensities are normalized with 1.0×10^{-4} erg/s/cm^2/sr. We impose the Lipschitz condition by restricting the range of the learnable parameters of the critics to [-0.01, 0.01] as in the original proposal by Arjovsky et al. (2017). We run the training over 50 epochs with batch size 50.

Table 6.1: The architecture of the generators and discriminators: the number of filters N_{filter}, filter size, stride, presence of the batch normalization, dropout rate, and activation function. The tensor shapes are in a form of (width x, height y, depth z, features).

Generator

Layer	N_{filter}	Filter size	Stride	BN	Dropout	Activation
Conv. 1	64	(3, 3, 3, 2)	(2, 2, 2)	×	×	Leaky ReLU
Conv. 2	128	(3, 3, 3, 64)	(2, 2, 2)	○	×	Leaky ReLU
Conv. 3	256	(3, 3, 3, 128)	(2, 2, 2)	○	×	Leaky ReLU
Conv. 4	512	(3, 3, 3, 256)	(2, 2, 2)	○	×	Leaky ReLU
Deconv. 1	256	(3, 3, 3, 512)	(2, 2, 2)	○	0.5	Leaky ReLU
Deconv. 2	128	(3, 3, 3, 256)	(2, 2, 2)	○	×	Leaky ReLU
Deconv. 3	64	(3, 3, 3, 128)	(2, 2, 2)	○	×	Leaky ReLU
Deconv. 4	1	(3, 3, 3, 64)	(2, 2, 2)	×	×	tanh

Discriminator

Layer	N_{filter}	Filter size	Stride	BN	Dropout	Activation
Conv. 1	64	(3, 3, 3, 2)	(2, 2, 2)	×	×	Leaky ReLU
Conv. 2	128	(3, 3, 3, 64)	(2, 2, 2)	○	×	Leaky ReLU
Conv. 3	256	(3, 3, 3, 128)	(2, 2, 2)	○	×	Leaky ReLU
Conv. 4	512	(3, 3, 3, 256)	(2, 2, 2)	○	×	Leaky ReLU
Dense	1[a]			×	×	×

[a] Dimension of the output.

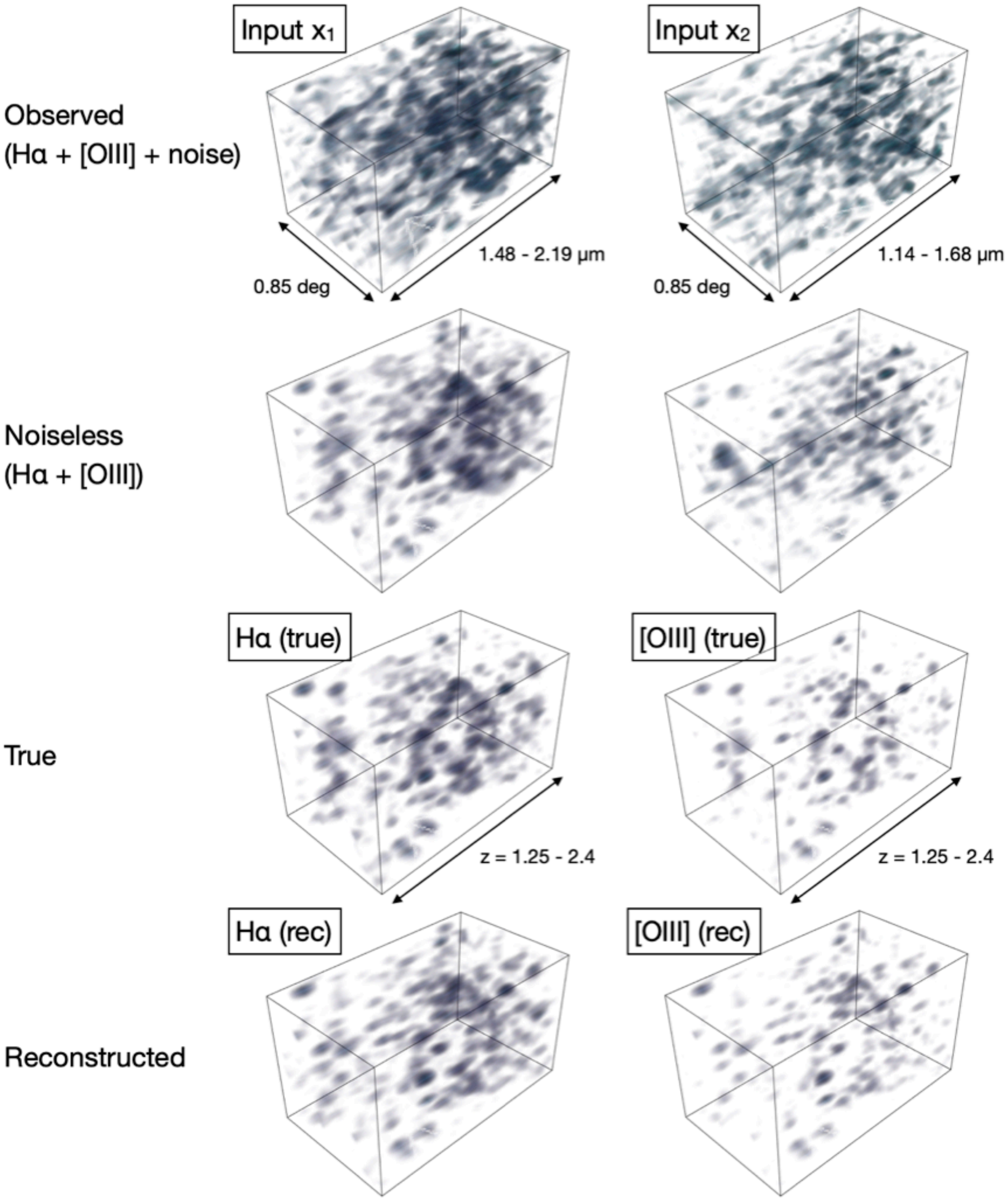


Figure 6.3: The observed intensity maps (Hα + [OIII] + noise; first row), Hα + [OIII] intensity maps (second row), true intensity maps (third row), and reconstructed intensity maps (fourth row). Two different wavelength ranges, 1.48 - 2.19 μm and 1.14 - 1.68 μm, are shown on the left and right, respectively.

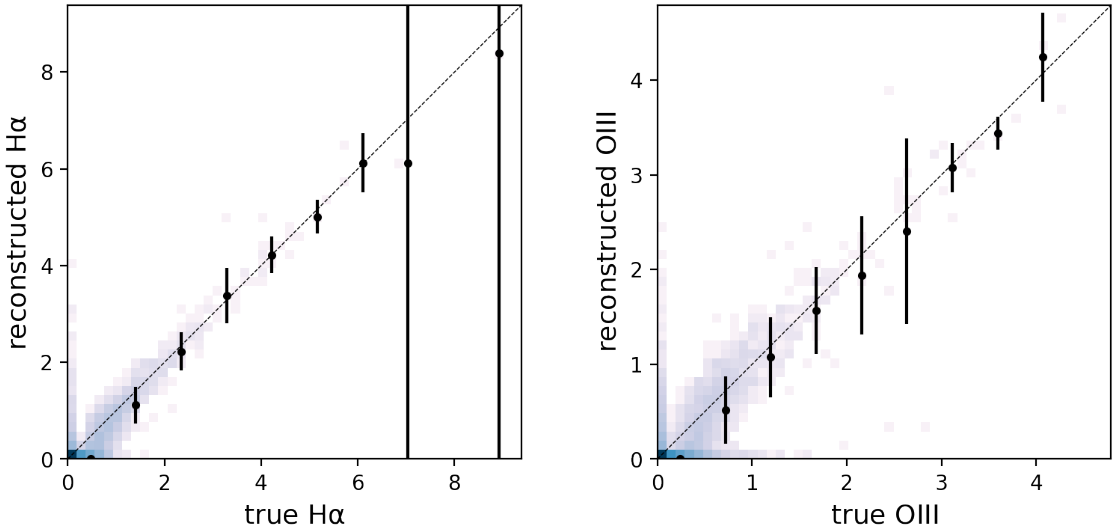


Figure 6.4: Pixel-by-pixel comparison between true and reconstructed intensities for Hα (left) and [OIII] (right) lines.

6.2 Reconstruction of Three-Dimensional Large-Scale Structures

For test data, we generate another 1000 light cones of 0.85 deg on a side and divide them into 8×8 sections in the angular direction. We input them into the generators and rearrange the outputs to obtain reconstructed (0.85 deg)^2 data.

An example of the reconstructed three-dimensional intensity maps is shown in Fig.6.3. The first, third, and fourth rows are the observed (input), true, and reconstructed maps. For reference, we also show the noiseless observed maps (only Hα + [OIII]) in the second row. The left and right columns show the data cubes at $\lambda_{\rm obs} = 1.48 - 2.19$ μm and $1.14 - 1.68$ μm, respectively, where there are underlying Hα and [OIII] signals at $z = 1.25 - 2.4$. The data is smoothed for visibility. Fig.6.4 shows the pixel-by-pixel comparison between unsmoothed maps. One can see that the bright sources are properly reconstructed, and the networks successfully remove the foreground/background line interlopers. To check their detectability, we count the numbers of bright pixels with $> 3\overline{\sigma}_n$, where $\overline{\sigma}_n$ is the average noise level over the 16 spectral bins and compute the recall ($N_{\rm correct}/N_{\rm true}$) and precision ($N_{\rm correct}/N_{\rm rec}$). We find that the networks achieve a high detectability, with a recall of 0.67 (0.78) and a precision of 0.84 (0.68) for the Hα ([OIII]) intensity. On the other hand, faint extended sources are not reproduced well because they are compromised by large observational noise. This result is similar to the result of the "small" dataset in Chapter 4.

The reproducibility of the bright pixels can also be confirmed by looking at the reconstructed PDFs. In Fig.6.5, we show the PDFs of Hα (left column) and [OIII] (right column) intensities at several spectral bins. In every spectral bin, the bright end of the PDF is properly reproduced. The bright end of the [OIII] PDFs are also reproduced well even though the contributions of the

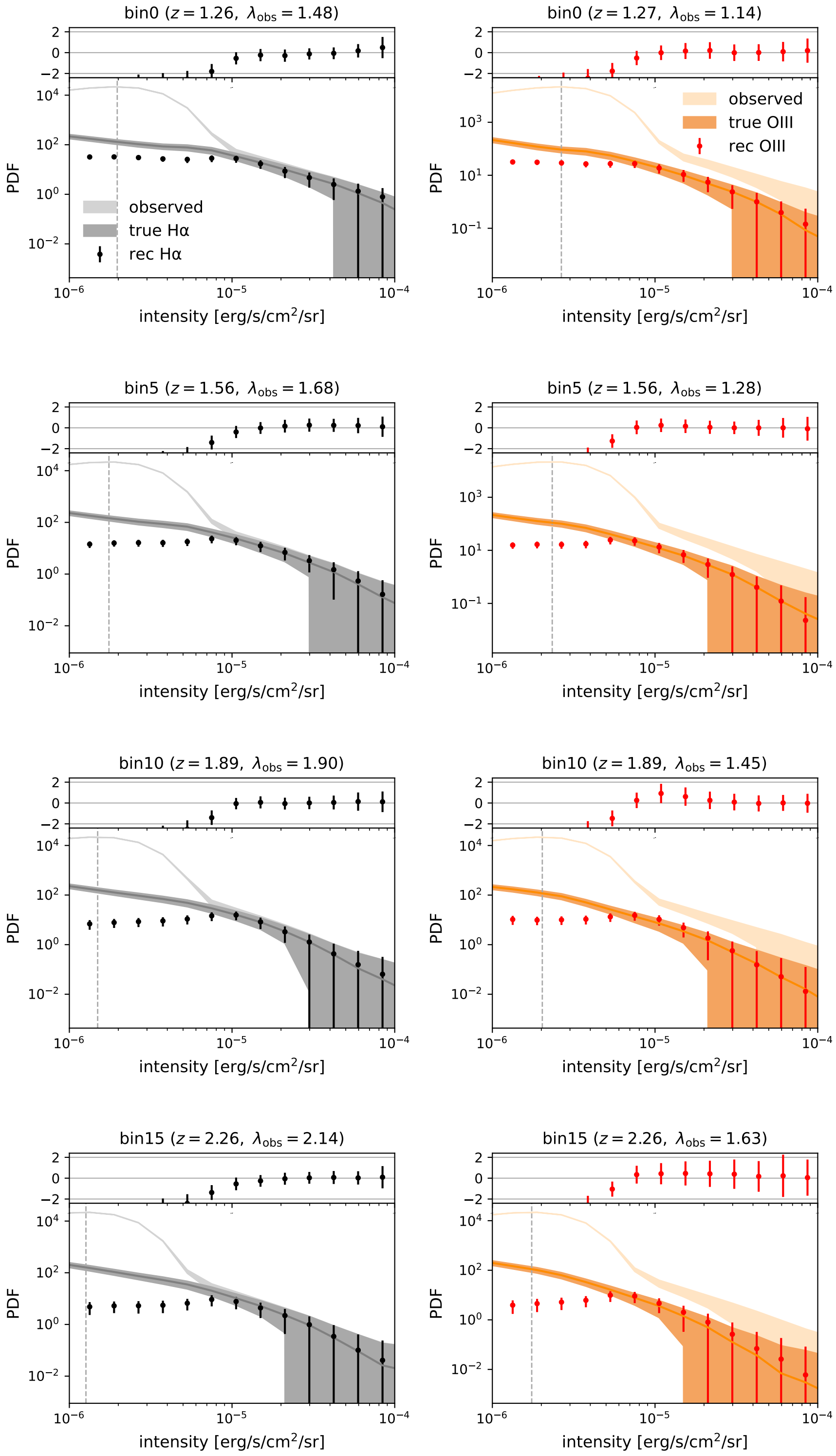


Figure 6.5: The PDFs at four spectral bins ($z \sim 1.26$, 1.56, 1.89, and 2.26) for Hα (left) and [OIII] (right) lines. The shaded regions and error bars represent the variance of the true and reconstructed PDFs over 1000 test data. The observed PDFs and the 1-σ_n noise level are also indicated by the light-colored shaded regions and the dashed vertical lines. The upper panels show the differences between true and reconstructed PDFs divided by the square root of the variance of the true PDFs.

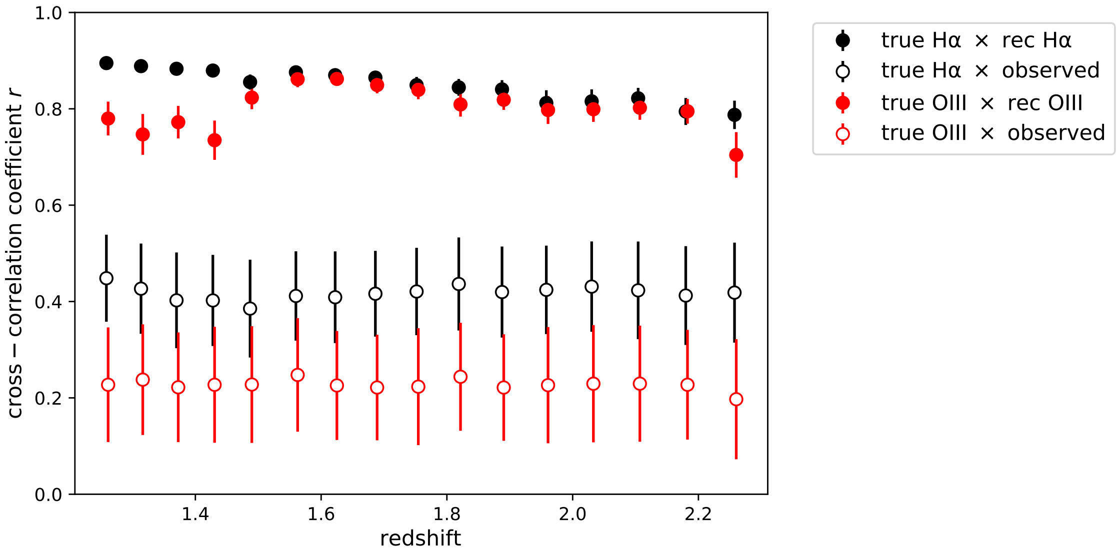


Figure 6.6: The cross-correlation coefficients between the true and reconstructed intensity maps for Hα (black) and [OIII] (red) as a function of redshift. Those between the true and observed intensity maps are also shown for reference.

foreground Hα emissions are dominant there. The small differences between individual true and reconstructed PDFs (upper panels) indicate that the networks learn to capture the properties of individual (8.5deg)^2 areas rather than just learn the typical property of the training data (see discussion after Eq.5.6).

The networks reproduce only a few faint pixels at $\lesssim 10^{-5}$ erg/s/cm^2/sr. We find that the reproducibility of the PDFs does not change if we perform bagging analysis. This behavior is different from the previous results in Chapter 4, where the network recovers the faint ends of the PDFs to some extent even in the presence of the noise. One possible reason for the difference is the usage of WGAN instead of the vanilla GAN. When the discriminator is trained optimally, the generator of the vanilla GAN sometimes updates parameters randomly (see Appendix A.1). Errors in individual outputs of the generators caused by such randomness are canceled out through the bagging procedure. Contrastingly, in WGAN, the generator is always updated based on a certain guideline given by the critic. Our generators may learn not to reconstruct the faint pixels that are almost impossible to properly reconstruct. To make the generators learn the faint end property as in Chapter 4, we may need to add some other constraints to the learning process.

We also examine the reproducibility of the large-scale clustering information. We find that the reconstructed power spectra are underestimated on large scales for the same reason as the result using the "small" dataset in Chapter 4, but the reconstructed intensity maps are still considered to have large-scale clustering information. To examine to what extent the reconstructed maps trace the large-scale distribution, we compute the cross-correlation

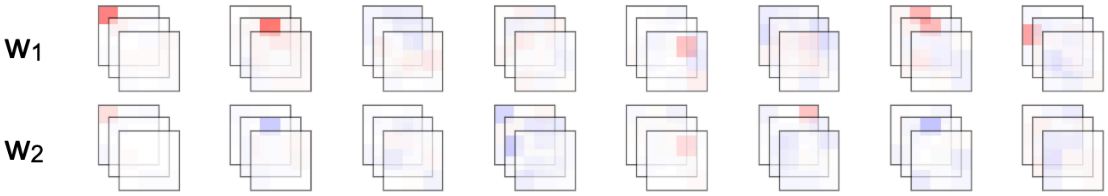


Figure 6.7: The three-dimensional convolutional filters of the first layer of the generator. Randomly chosen eight filters among $N_{\rm filter}(= 64)$ filters are shown. A set of three overlapping 3×3 maps represents a three-dimensional convolutional filter.

coefficient between the true and reconstructed maps

$$r_{\rm true,rec}(k) \equiv \frac{P_{\rm true,rec}(k)}{\sqrt{P_{\rm true}(k)P_{\rm rec}(k)}}, \tag{6.2}$$

where $P_{\rm true}$ and $P_{\rm rec}$ are the auto-power spectra of the true and reconstructed maps. It would take 1 (-1) when the two fluctuations are perfectly correlated (anti-correlated), and 0 when they are uncorrelated. Fig.6.6 shows those at $k = 0.3$ arcmin^{-1} ($\sim 0.3h$ Mpc^{-1}) at each redshift bin. The error bars show the variance of the 1000 test data. The correlation coefficients between the true and observed maps are also shown for reference. They are ~ 0.8 for all redshifts, much larger than those with observed maps. This suggests that the reconstructed maps are useful tracers of the large-scale structure at the specific redshifts. We find that the reproducibility of the above statistics and the detectability of bright sources do not largely change when data with different line intensities are input.

6.3 Understanding the Networks

In our newly developed network architecture, the generators receive two observational data cubes where the same large-scale structure is imprinted. We expect that the generators use the information of both data cubes and capture the co-existence of the two emission lines in them. In the following, we perform three tests to investigate how the networks extract signals.

Convolutional Filters

We first look at the convolutional filters to understand the function of the network. Fig.6.7 shows several convolutional filters of the first layer of the generator. The size of individual filters is (3, 3, 3, 2). In the figure, a set of three overlapping 3×3 maps represents a three-dimensional convolutional filter. We have a pair of them, (w_1, w_2), corresponding to the two feature maps of the input x_1 and x_2. The i-th feature of the next layer is computed as

$$y_{\rm i} = a(w_{1\rm i} * x_1 + w_{2\rm i} * x_2 + b_{\rm i}), \tag{6.3}$$

where a and $b_{\rm i}$ are the activation function and bias. We find that in several filter pairs, only two corresponding pixels in w_1 and w_2 have large values (e.g., leftmost one in Fig.6.7). Such

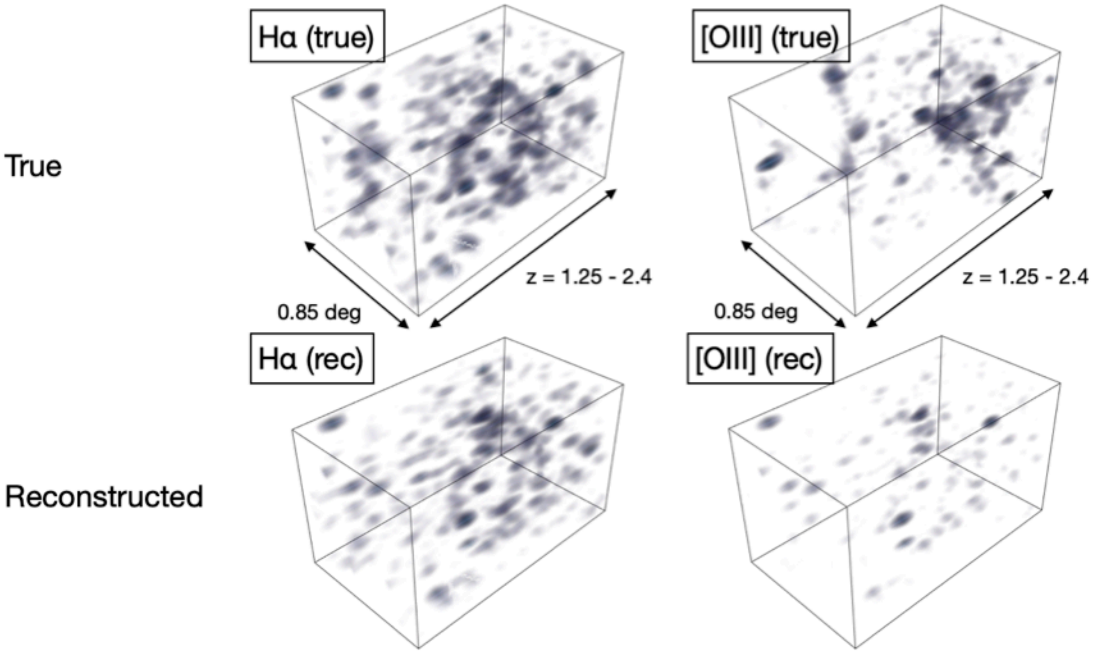


Figure 6.8: The true (top) and reconstructed (bottom) intensity maps when using different realizations for generating Hα and [OIII] maps. The same realization with Fig.6.3 is used for the Hα map.

filters are sensitive to the signals co-existing in the corresponding pixels of x_1 and x_2 and thus may allow the generators to properly extract the Hα and [OIII] signals from the same galaxies. Although it is not clear how the individual filters affect the final output of the network, this indicates that the networks make use of the multi-wavelength information.

Random Realization Test

Here, we investigate how the networks respond to uncorrelated Hα and [OIII] signals (i.e., data without any co-existing signals). We generate a mock observational map using different realizations for Hα and [OIII] and input this into the networks. The resulting reconstructed maps are shown in Fig. 6.8. We use the same realization for Hα intensity as that shown in Fig.6.3. For the [OIII] intensity map, it is clear that the network does not reproduce the true intensity distribution, indicating that the [OIII] generator relies on the Hα signals underlying in one of the input data cubes, x_1. It should be noted that the [OIII] generator is not completely dependent on the data cubes x_1. If so, the reconstructed [OIII] intensity would always look the same as the Hα intensity regardless of the input [OIII] intensity. We find that the reconstructed Hα map also differs from that in Fig.6.3. It is indicated that the Hα generator also partly relies on the data cube x_2.

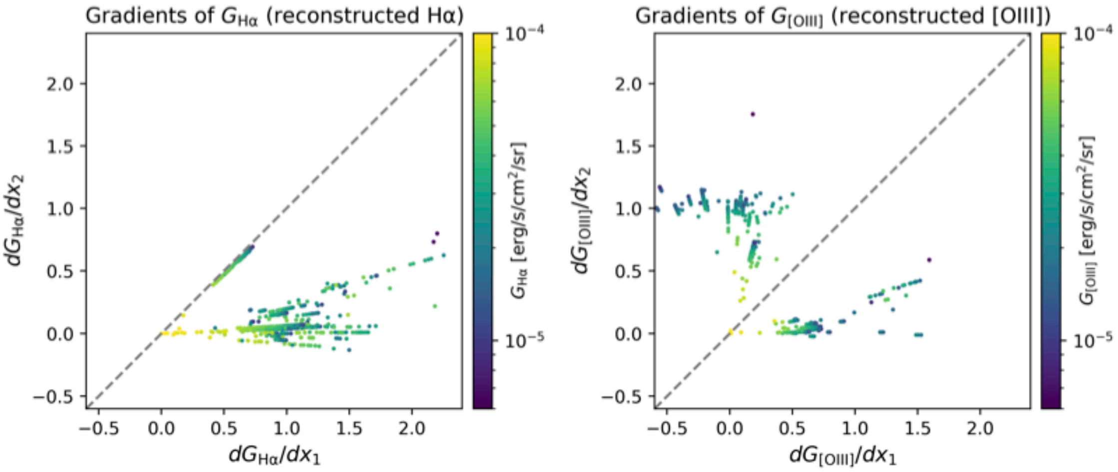


Figure 6.9: The gradients of the reconstructed Hα (left) and [OIII] (right) maps with respect to the intensities at pixels with brightest Hα intensities $(i_\mathrm{b}, j_\mathrm{b}, k_\mathrm{b})$. The horizontal and vertical axes show the gradients with respect to the first and second inputs, x_1 and x_2, respectively. The colors show the reconstructed intensities.

Saliency Analysis

To further investigate the dependence on the two inputs, x_1 and x_2, we perform a saliency analysis. The saliency analysis quantifies which pixels of the input data have large impacts on the output. One of the simplest saliency methods, the so-called vanilla gradient, computes the derivative of the output value with respect to the input pixels (Simonyan et al. 2014). Pixels with a larger gradient are considered to have a greater impact on the output. Among all the pixels in a reconstructed map, we choose a pixel with the brightest Hα intensity,[2] which is denoted by $(i_\mathrm{b}, j_\mathrm{b}, k_\mathrm{b})$. We compute the partial derivative of the intensities $G_{\mathrm{H}\alpha}^{i_\mathrm{b} j_\mathrm{b} k_\mathrm{b}}$ and $G_{\mathrm{[OIII]}}^{i_\mathrm{b} j_\mathrm{b} k_\mathrm{b}}$ with respect to the intensities of the individual pixels of the inputs x_1 and x_2

$$\frac{\partial G_a^{i_\mathrm{b} j_\mathrm{b} k_\mathrm{b}}}{\partial x_1^{ijk}}, \quad \frac{\partial G_a^{i_\mathrm{b} j_\mathrm{b} k_\mathrm{b}}}{\partial x_2^{ijk}} \quad (a = \mathrm{H}\alpha,\ \mathrm{[OIII]}) \tag{6.4}$$

for 1000 brightest pixels among 1000 test data.

As expected, we find that the most important pixels are those with $(i, j, k) = (i_\mathrm{b}, j_\mathrm{b}, k_\mathrm{b})$ in x_1 and x_2 maps. Fig.6.9 shows the gradients of the reconstructed Hα (left) and [OIII] (right) intensities with respect to the intensity at these pixels. The horizontal and vertical axes indicate the importance of the input pixels of x_1 and x_2. For reconstruction of the Hα intensity maps (left panel), we find $dG_{\mathrm{H}\alpha}/dx_1 > dG_{\mathrm{H}\alpha}/dx_2$ for most of the pixels, indicating that the network put more focus on x_0, which is consistent with what we have seen above. For reconstruction of the [OIII] intensity maps (right panel), the values are clearly divided into two groups: those with $dG_{\mathrm{[OIII]}}/dx_1 > dG_{\mathrm{[OIII]}}/dx_2$ and with $dG_{\mathrm{[OIII]}}/dx_1 < dG_{\mathrm{[OIII]}}/dx_2$. We investigate how

---

[2] We find that the results do not significantly change when we consider the brightest pixel of the reconstructed [OIII] map.

the reconstructed intensity, observed intensity, and redshifts of the pixels would affect whether the [OIII] generator relies on x_1 or x_2, but found no clear trend. In Fig.6.9, the color represents the reconstructed intensity.

The input pixels other than those at $(i_\mathrm{b}, j_\mathrm{b}, k_\mathrm{b})$ also affect the reconstruction. We find that the neighboring pixels along the spectral direction have a particular influence on the reconstruction. The networks may learn that the Hα and [OIII] emissions from the same galaxy do not always appear in the corresponding spectral bins because of the alignment of the spectral filters. In the angular direction, only the intensities at a few pixels away from $(i_\mathrm{b}, j_\mathrm{b}, k_\mathrm{b})$ have a slight effect on the reconstructed intensity, and those much farther away do not affect at all. This tells us that the networks put more focus on the information in spectral direction than the large-scale distributions extended over the angular direction. This is expected because we only adopt a dataset with a very small angular size. To let a machine learn such large-scale clustering information, we should adopt a larger boxsize, which is currently not available due to computational limitations.

6.4 Conclusion

We have developed three-dimensional CNNs that extract and separate multiple signals (Hα and [OIII]) from noisy LIM data cubes. We *inform* the physical values (the rest-frame wavelengths of emission lines) by using multiple inputs where emission lines from the same galaxy population are imprinted. We use WGAN instead of the vanilla GAN and find the training is more stable. The bright sources of both the Hα and [OIII] lines are properly reproduced. The reconstruction limit is at a few-σ_n level. A similar detection limit is found by Cheng et al. (2020).

Our method allows us to reconstruct intensity maps quite well as long as the observation covers a sufficient wavelength range. Although we have ignored the interlopers other than Hα and [OIII] lines, such as the [OII] line, our network architecture can easily be extended to include such lines. If they are added, they would provide additional information to the network, which makes it easier to reconstruct intensity maps at designated redshifts. While we only considered the intermediate redshift and the near-infrared wavelength range, our method can be applied to any other ranges in principle.

To understand how the networks reconstruct the signals, we look at the convolutional filters, conduct random realization test, and perform saliency analysis. They all suggest that the networks efficiently extract the co-exsisting signals underlying in the two input data. It is also indicated that the large-scale clustering information is not used by the networks that much. Such information can only be taught by using a dataset with a larger boxsize.

As we have discussed in the previous chapters, the obtained LSSs can be used for various studies in cosmology. They can also be used to study the astrophysical phenomena such as cosmic reionization. For example, we can compute the cross-correlation signals between the 21-cm intensity maps and the galaxy distribution to detect the EoR signals. If we use the reconstructed LIM data by our methods for such analysis, we may be able to detect the EoR signals more robustly with less systematics.

Chapter 7

Application of LIM Data for Studying Cosmic Reionization

The observed LSSs can be used for various studies in cosmology and astrophysics. As one of these applications, we investigate using them to study the EoR. The details of the cosmic reionization and IGM heating phenomena are not well known. The distant 21-cm line from the IGM is the key to elucidating them. However, the 21-cm emission is very weak and contaminated by bright foregrounds. Cross-correlation is one of the important observational techniques to investigate the EoR.

In this study, we predict the cross-power spectra between the 21-cm and [OIII] intensity maps. The [OIII] intensity maps at the EoR can be obtained in the SPHEREx-like observations including CDIM (Cooray et al. 2019). As we have discussed in the previous chapters, the bright [OIII] emitters extracted from the intensity maps by the machine learning method could also be good (or better) tracers of the LSS. We thus study the cross-correlation using the number density of the bright [OIII] emitters as well.

In Section 7.1, we describe our methods to generate the 21-cm and [OIII] maps. We then compute the cross-correlation signals between the 21-cm intensity maps and the distribution of the [OIII] emitters (number density maps or intensity maps) and show them in Section 7.2. We then discuss the results in Section 7.3. The contents of this chapter are based on Moriwaki et al. (2019).

7.1 Methods

7.1.1 Reionization Simulation

To study reionization, we use simulation outputs obtained by Eide et al. (2018). They use the output of the cosmological hydrodynamics simulation MassiveBlack-II (Khandai et al. 2015).[1] In this simulation, physical processes such as star formation (Springel & Hernquist 2003), feedback effects (Croft et al. 2009; Degraf et al. 2010; Di Matteo et al. 2008, 2012),

---

[1] The simulation adopts the ΛCDM model with $\Omega_\Lambda = 0.735$, $\Omega_m = 0.275$, $\Omega_b = 0.046$, and $h = 0.701$ (Komatsu et al. 2011).

and the formation and evolution of black holes (Di Matteo et al. 2005; Springel 2005) are included. The simulation boxsize is $100h^{-1}$ comoving Mpc on a side and the mass resolution is $m_{\rm DM} = 1.1 \times 10^7 h^{-1}$ M$_\odot$ for dark matter particles and $m_{\rm gas} = 2.2 \times 10^8$ M$_\odot$ for gas particles. The smallest halo mass is $\sim 10^8$ M$_\odot$.

Table 7.1:

	SED model
Stars	Eldridge & Stanway (2012)[a]
X-ray binary	Fragos et al. (2013a,b); Madau & Fragos (2017)[b]
ISM	Mineo et al. (2012); Pacucci et al. (2014)[c]
AGN	Krawczyk et al. (2013)[d]

[a] Include formation of binary stars (BPASS model).
[b] Include emissions from young massive stars and low mass stars.
[c] Assume that luminosity scales with SFR.
[d] Assume that luminosity scales with accretion rate. Radiation efficiency $\eta = L/\dot{M}c^2$ is set to be 0.1.

Eide et al. (2018) use the Monte Carlo radiation transfer code CRASH (Ciardi et al. 2001; Graziani et al. 2013; Maselli et al. 2009, 2003) to perform the radiation transfer calculation with a post-processing method and compute the ionization degree and gas temperature of the IGM. In the post-processing, the ionizing photons from 13.6 eV to 2 keV are divided into 82 frequency bins. The number of photons at frequency ν during a time step $\Delta t_{\rm em}$ is given by

$$N_i(\nu, z) = \hat{S}_i(\nu, z) f_{\rm esc}(\nu) \epsilon_i(z) \Delta t_{\rm em}(z), \tag{7.1}$$

where $\hat{S}_i$ is the normalized SED and ϵ_i is the total number of photons emitted from source i. The escape fraction of the ionizing photons is set to be $f_{\rm esc} = 0.15$. They consider four different ionizing/heating sources: stars, X-ray binaries, ISM, and AGNs. Table 7.1 list the models of the sources adopted in their computation. Eide et al. (2018) find that the ionization is mainly caused by stars. The X-ray binaries and hot ISM contributes to the IGM heating outside the ionized bubbles. There are few AGNs in the simulation and their contribution to the ionization and heating is small. The simulation yields Thomson scattering optical depth of the CMB, $\tau_{\rm e} = 0.055$, which is consistent with the recent CMB observation (Planck Collaboration VI 2018).

The brightness temperature of the 21-cm line relative to the background radiation field is given by (see Appendix B.1)

$$\delta T_{\rm b}^{\rm obs}(\nu) \sim 28.5 x_{\rm HI}(1+\delta_\rho)\left(\frac{\Omega_{\rm b}h^2}{0.02}\right)\left(\frac{0.15}{\Omega_{\rm m}h^2}\frac{1+z}{10}\right)^{1/2}\left(1 - \frac{T_{\rm R}}{T_{\rm s}}\right) \text{ mK}, \tag{7.2}$$

where $x_{\rm HI}$ is the neutral fraction, δ_ρ is the IGM density fluctuation, $T_R = T_{\rm CMB}$ is the background radiation temperature, and T_s is the spin temperature of the 21-cm emission. When the spin temperature is higher (lower) than the background, the 21-cm line is observed as emission (absorption). At $z < 10$, we can assume that it is the same as the kinetic temperature of the IGM, $T_s = T_{\rm kin}$.

It is crucial to predict the observed signals with various ionization and heating models. Recent 21-cm observations disfavor models that the IGM is heated only at a very late stage of the EoR but all the other models survive. Though we only have a single reionization simulation, we can study another simple model, the saturated heating model, without additional costly radiation simulation. In this model, we assume that more high-energy photons are emitted in the early universe than our fiducial model, and thus the IGM temperature is already sufficiently high, $T_s = T_{\rm kin} \gg T_{\rm R}$, at the redshift of our interest ($z \leq 10$). In this case, the 21-cm brightness temperature only depends on the fluctuations of density and the neutral fraction

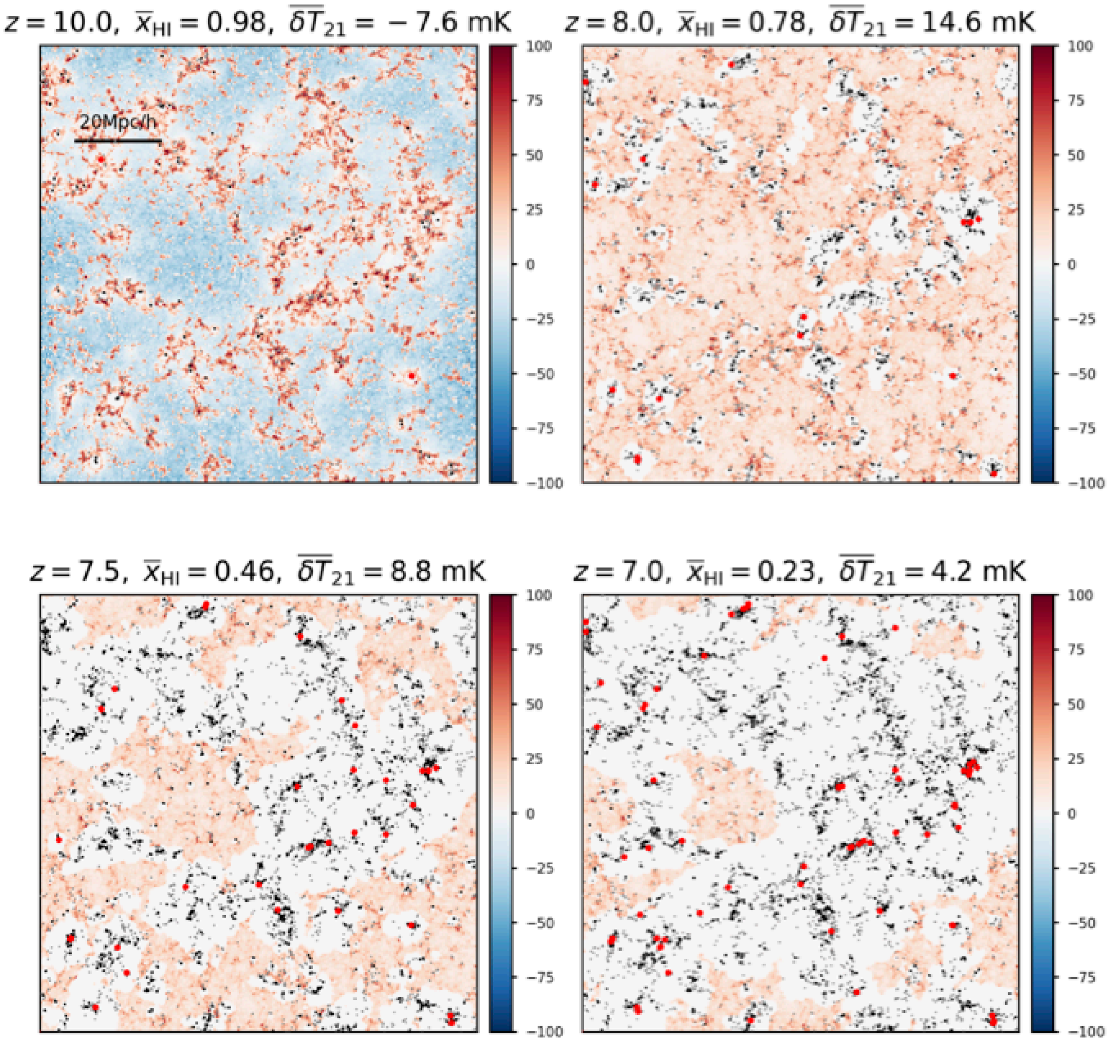
$$\delta T_{\rm b}^{\rm obs} \propto x_{\rm HI}(1+\delta_\rho), \tag{7.3}$$

The presence of such energetic sources may also affect the ionization state of the IGM, but the effect is considered to be small (Eide et al. 2018). In the following, we compare the cross-power spectra with fiducial and saturated heating models.

7.1.2 [OIII] Line Emission

We examine two kinds of cross-correlation signals: those between the 21-cm intensity maps and number density maps of the [OIII] emitters, and those between the 21-cm and [OIII] intensity maps. To compute the [OIII] luminosity, we use Eq.3.21. We ignore the dust extinction ($A_{\rm line} = 0$) and adopt the escape fraction $f_{\rm esc} = 0.15$. The [OIII] emitters trace the ionizing sources. We find that the population of galaxies with $L_{\rm [OIII]} \sim 10^{41}$ erg/s contributes the most to the total number of ionizing photons as well as the total [OIII] intensity. In our model, the heating sources reside in the star-forming and massive galaxies, which are also traced by [OIII] emitters. The [OIII] line intensity maps are generated with Eq.2.17. As for the [OIII] emitters, we consider the galaxies with $L_{\rm [OIII]} > 10^{42}$ erg/s. This threshold is comparable to the expected noise level of CDIM (Cooray et al. 2019). Thus these galaxies can be precisely detected if we apply our machine learning method to the observed data of CDIM. In the following, we ignore the contributions from interlopers for simplicity, but we note that the interloper removal methods we have devised in the previous chapters are crucial to isolate the high-redshift intensity signals at a particular redshift and to robustly detect the cross-correlation signals.

In this study, we use the snapshots of the simulation at $z = 7$, 7.5, 8, 9, and 10. Fig 7.1 and Fig 7.2 show the distributions of the 21-cm line intensity (colormaps) and [OIII] emitters (dots) for the fiducial and saturated heating model, respectively. The red and black dots are the [OIII] emitters with $L_{\rm [OIII]} > 10^{42}$ erg/s and $> 10^{41}$ erg/s, respectively. In the fiducial model, at $z = 10$, the regions closer to the [OIII] emitters are more heated and show 21-cm emission (positive signals), while the regions farther from them are still colder than the background CMB, showing absorption signals. From high redshift to low redshift, the IGM is heated up and the ionized bubbles around bright [OIII] emitters expand. In the saturated heating model, the gas temperature is sufficiently high everywhere even at $z = 10$, and the 21-cm signals



$z = 10.0,\ \overline{x}_{\rm HI} = 0.98,\ \overline{\delta T}_{21} = -7.6$ mK

$z = 8.0,\ \overline{x}_{\rm HI} = 0.78,\ \overline{\delta T}_{21} = 14.6$ mK

$z = 7.5,\ \overline{x}_{\rm HI} = 0.46,\ \overline{\delta T}_{21} = 8.8$ mK

$z = 7.0,\ \overline{x}_{\rm HI} = 0.23,\ \overline{\delta T}_{21} = 4.2$ mK

Figure 7.1: The distributions of the 21-cm line intensity and [OIII] emitters at $z = 10$, 8, 7.5, and 7. The maps are generated with slices with $1.2h^{-1}$ Mpc. The colormaps show the 21-cm line intensities and the red and black dots are the [OIII] emitters with $L_{\rm [OIII]} > 10^{42}$ erg/s and $> 10^{41}$ erg/s.

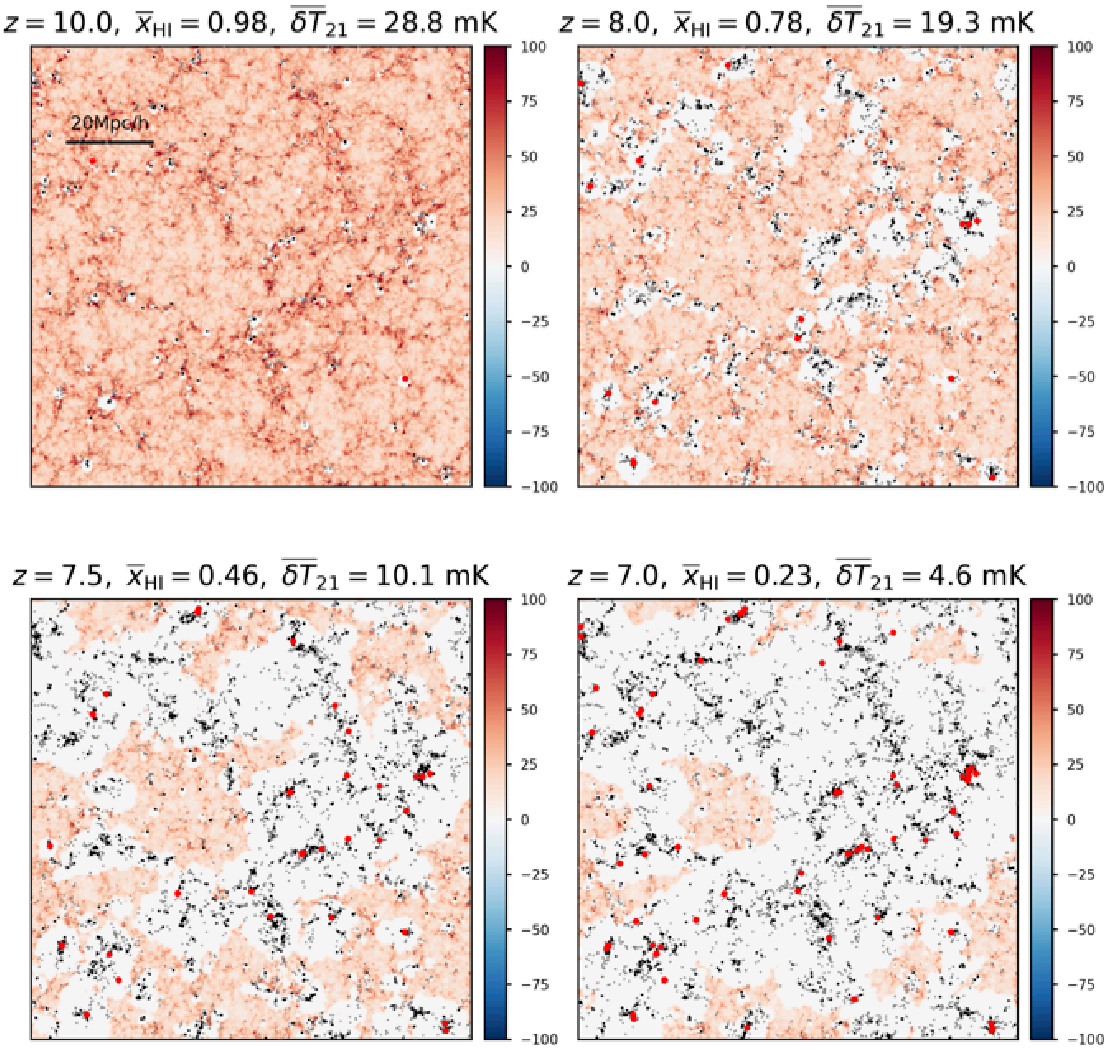


Figure 7.2: Same as Fig.7.1, but for the saturated heating model.

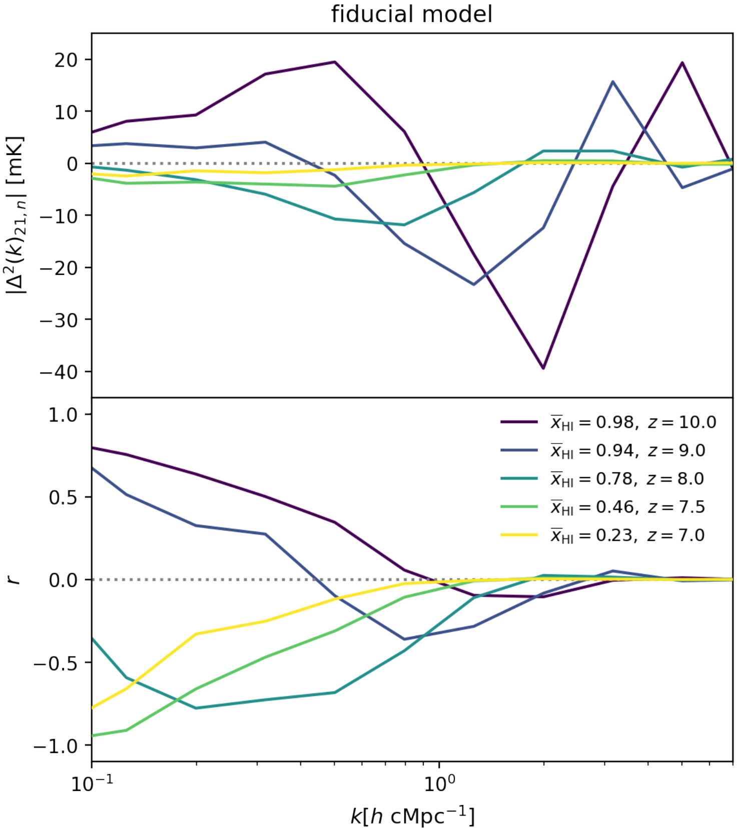


Figure 7.3: *Top:* the cross-power spectrum between the fluctuations of the 21-cm line intensities and the number densities of the [OIII] emitters with $L_{\rm [OIII]} > 10^{42}$ erg/s at $z = 10, 9, 8, 7.5, 7$. *Bottom:* the cross-correlation coefficient.

almost only trace the density fluctuations. The expansion of the ionization bubbles is the same as in the fiducial model.

7.2 Cross-Power Spectra

We first compute the cross-power spectra between the 21-cm signals and the number density of the [OIII] emitters for the fiducial model. The top panel of Fig.7.3 shows the normalized cross-power spectra

$$\Delta^2(k) = \frac{k^3}{2\pi^2} P(k). \tag{7.4}$$

The cross-correlation coefficients (Eq.6.2) are also shown in the bottom panel. At high redshifts, they are positively correlated on large scales ($k \sim 0.1h$ Mpc^{-1}) because of the positive correlation between the gas temperature and the distribution of the [OIII] emitters, and they are anti-correlated on small scales ($k \sim 2h$ Mpc^{-1}) because of the anti-correlation between the neutral fraction and the [OIII] emitters. At $z \leq 8$, the gas temperature fluctuations dissapear and the ionized bubbles expand. Thus the anti-correlations are observed on large scales

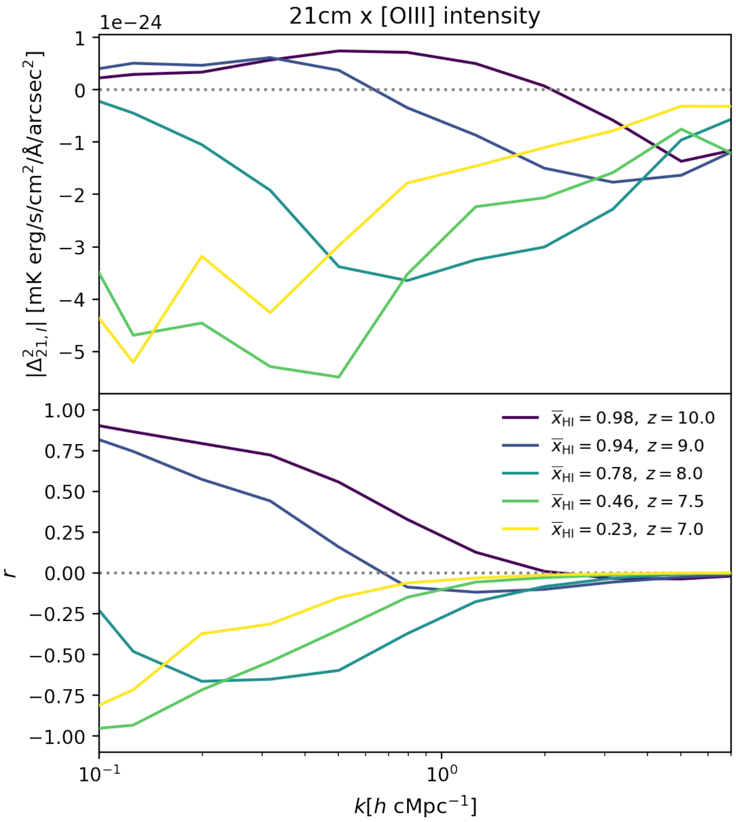


Figure 7.4: Same as the Fig.7.3, but using the [OIII] line intensity instead of the number density of [OIII] emitters to compute the cross-power spectra.

($k \lesssim 1h$ Mpc^{-1}). Such behavior of the cross-correlation signals is consistent with previous studies (e.g., Lidz et al. 2009).

We also examine the raw [OIII] LIM data. Fig.7.4 shows the cross-power signals between the 21-cm and [OIII] intensity maps. We note that the observed cross-power spectrum scales with the mean [OIII] intensity as well as that of the 21-cm intensity. Their shapes differ from those using the number density of [OIII] emitters. At $z = 9$ and 10, for example, anti-correlation is observed at much smaller scales (large k). This is because the [OIII] intensity is dominated by galaxies much fainter than 10^{42} erg/s, which are surrounded by smaller ionized bubbles. We find that when we use $L_{\rm [OIII]} > 10^{41}$ erg/s galaxy samples, the cross-power spectra have similar shapes as those using [OIII] line intensity.

We then examine the saturated model. Fig. 7.5 shows the cross-power spectrum using the number density of the [OIII] emitters for the saturated heating model. In this model, significant differences from the fiducial model are only observed on large scales and at high redshifts. We thus compare the evolutions of the cross-power spectrum amplitude at $k = 0.1h$ Mpc^{-1} in Fig. 7.6. In the saturated heating model (dashed), the positive correlation are still observed at high redshift, but their amplitudes are smaller than the fiducial model (solid) because the matter density fluctuations are much smaller than the temperature fluctuations.

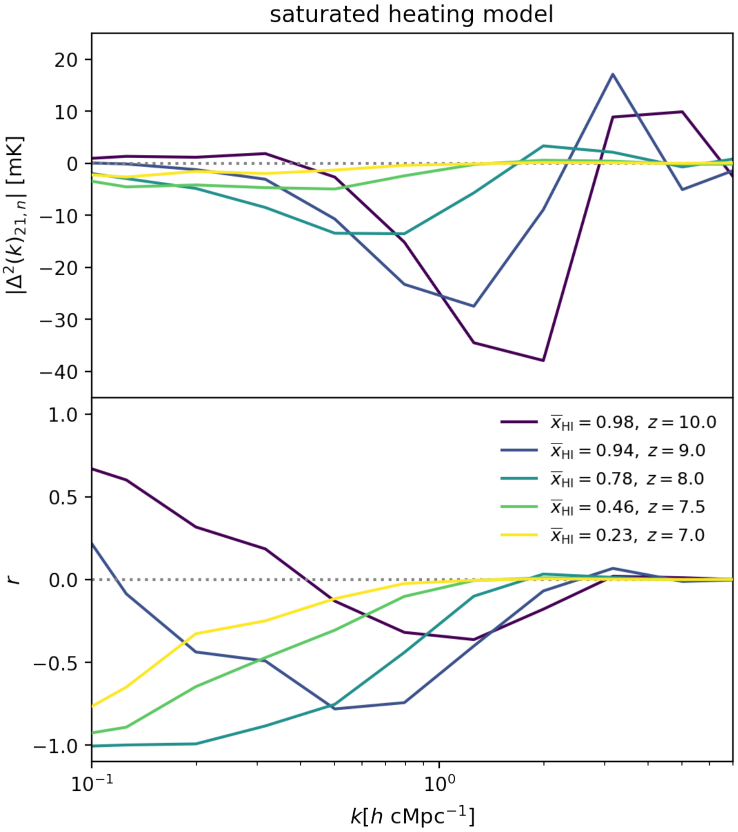


Figure 7.5: Same as the Fig.7.3, but for the saturated heating model.

7.3 Discussions

7.3.1 Small-Scale Signals

As seen in the bottom panels (correlation coefficients) of Fig.7.3 and 7.5, there are transitions from anti-correlation to no-correlation on small scales. At $z = 10$, it becomes zero at around $k = 3h$ cMpc^{-1}($\equiv$ k$_{\rm tr}$) corresponding to the physical scales $r = 2h^{-1}$ cMpc in both fiducial and saturated heating models. We find that such transiting scales $k_{\rm tr}$ correspond to the typical ionized bubble sizes. Previous studies have found that positive correlations are observed on small scales (e.g., $k > 3h$ Mpc^{-1}) (e.g., Lidz et al. 2009). Similar trends are also found by a recent study using a high-resolution cosmological simulation (Kannan et al. 2021).[2] Such small-scale positive correlations are considered to be caused by the correlation between the galaxies and self-shielded Lyman limit systems (dense gas cloud) that remain neutral. In our simulation, sub-resolution structures are not assumed while other studies incorporate the clumping factor $C = \overline{\rho^2}/\overline{\rho}^2 > 1$ for un-resolved gas clouds (e.g., Kannan et al. 2021), which would produce dense and neutral regions even within the ionized bubbles. We note that the small-scale IGM structures at high redshifts are not well known (e.g., Park et al. 2016), which can be studied

---

[2] Thus the positive correlations may not be fake signals due to the low resolutions of the previous studies.

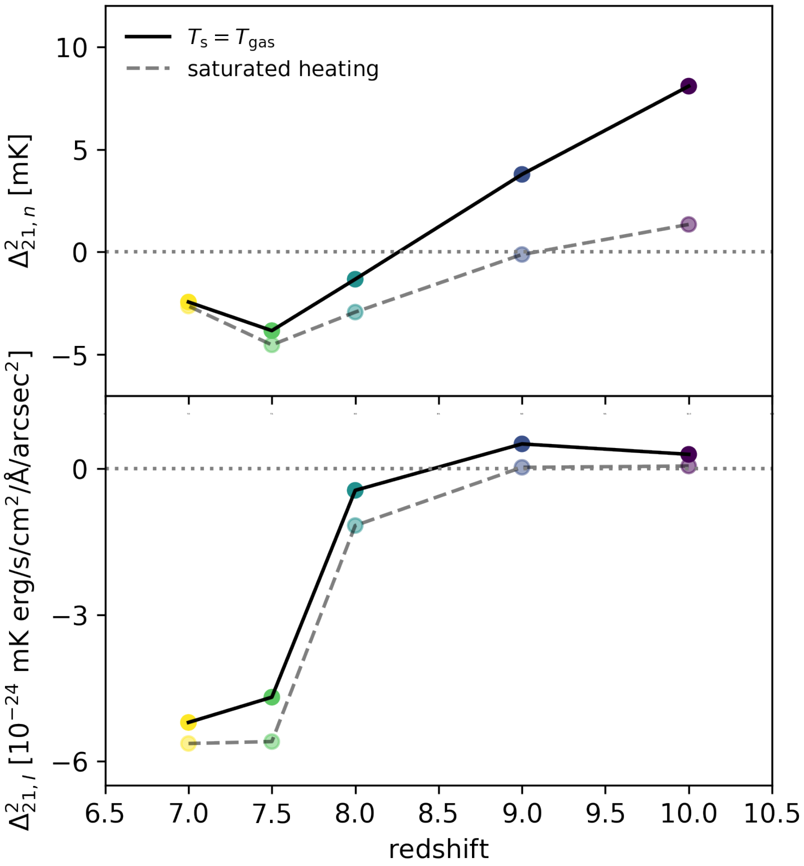


Figure 7.6: The evolution of the large-scale cross-power spectra between 21-cm and number density of the [OIII] emitters (top) or [OIII] intensity (bottom) for the fiducial (solid) and saturated heating (dashed) models.

with such small-scale cross-power signals once they are detected.

7.3.2 Large-Scale Signals

We find that the large-scale signals have different redshift evolutions depending on the heating models. From such large-scale signals obtained in future observations, we may be able to distinguish the reionization and heating models. The two phenomena give opposite contributions to the cross-correlation signals. When IGM heating is ongoing, large positive correlation signals are observed. During reionization, negative correlation signals are observed on large scales. Such different behaviors make it easier for us to determine at which period heating and reionization proceed. For this reason, the cross-power spectra could be more informative than the auto-power spectra of the 21-cm signals to distinguish different models. In particular, the period when the large-scale power turns from positive to negative (”turnover redshift” $z_{\rm to}$) can be an important probe. Fig.7.6 infers that the choice of the LSS tracer, i.e., the [OIII] intensity maps or number density maps, makes no significant change in $z_{\rm to}$. We also find that it may not significantly change when we adopt 10^{41} erg/s for the threshold value. Thus the turnover redshift will serve as a good indicator of the beginng of the reionization and the end of the heating.

7.3.3 Detectability of the Signals

We evaluate the detectability of the cross-power spectra with future observations. Under the Gaussian approximation, the variance of the cross-power spectrum between three-dimensional maps A and B is given by

$$\sigma^2_{\mathrm{A,B}}(\boldsymbol{k}) = \frac{1}{2}(P^2_{\mathrm{A,B}}(\boldsymbol{k}) + \sigma_{\mathrm{A}}(\boldsymbol{k})\sigma_{\mathrm{B}}(\boldsymbol{k})), \tag{7.5}$$

where

$$\sigma^2_{\mathrm{A}}(\boldsymbol{k}) = (P_{\mathrm{A}}(\boldsymbol{k}) + P_{\mathrm{noise,A}}(\boldsymbol{k}))^2, \tag{7.6}$$

is the variance of the auto-power spectrum of A. P_{A} is the auto-power spectrum of A, and $P_{\mathrm{noise,A}}$ is the noise power spectrum. The variance of the spherically averaged power spectrum is computed as

$$\frac{1}{\sigma^2_{\mathrm{A,B}}(k)} = N_k \int \frac{\mathrm{d}\mu}{\sigma^2_{\mathrm{A,B}}(\boldsymbol{k})}, \tag{7.7}$$

where μ is the cosine of the angle between the line-of-sight and $\boldsymbol{k}$, and N_k is the number of modes within a hemispherical shell of width δk. The number of modes is given by

$$N_k = \frac{1}{2}\frac{4\pi k^2 \delta k}{(2\pi)^3/V_{\mathrm{survey}}} = \frac{k^2 \delta k V_{\mathrm{survey}}}{4\pi^2}, \tag{7.8}$$

where V_{survey} is the survey volume.

The expected noise power spectra of the 21-cm maps obtained from SKA observations are given by Eq.B.15. When we use the galaxy number density, the noise term is the combination of the galaxy shot noise and the uncertainty of the redshift (Tegmark 1997)

$$P_{\mathrm{noise,gal}} = \frac{1}{\overline{n}_{\mathrm{gal}}} \exp(k^2_{||}\sigma^2_r), \tag{7.9}$$

where $\overline{n}_{\mathrm{gal}}$ is the mean number density of galaxies, $k_{||} = \mu k$ is the parallel component of the wavenumber vector, $\sigma_r = c\sigma_z/H(z)$, and σ_z is the uncertainty of the redshift. We adopt $\sigma_z = 0.01$.[3]

We consider that SKA and the [OIII] surveys (e.g., CDIM) have an overlapping survey region with $V_{\mathrm{surv}} = 10^6 h^{-3}$ Mpc^3. This corresponds to 1 deg^2 with $dz = 0.4$ at $z = 10$. Fig.7.7 shows the absolute amplitude of the cross-power spectrum (solid) and the errors obtained in Eq.7.5 (single-dotted lines). One can see that the large-scale power spectra can be detected with S/N = 5-20 even at $z = 10$. We find that the sample variance (P_{21}, P_{gal}) dominates at

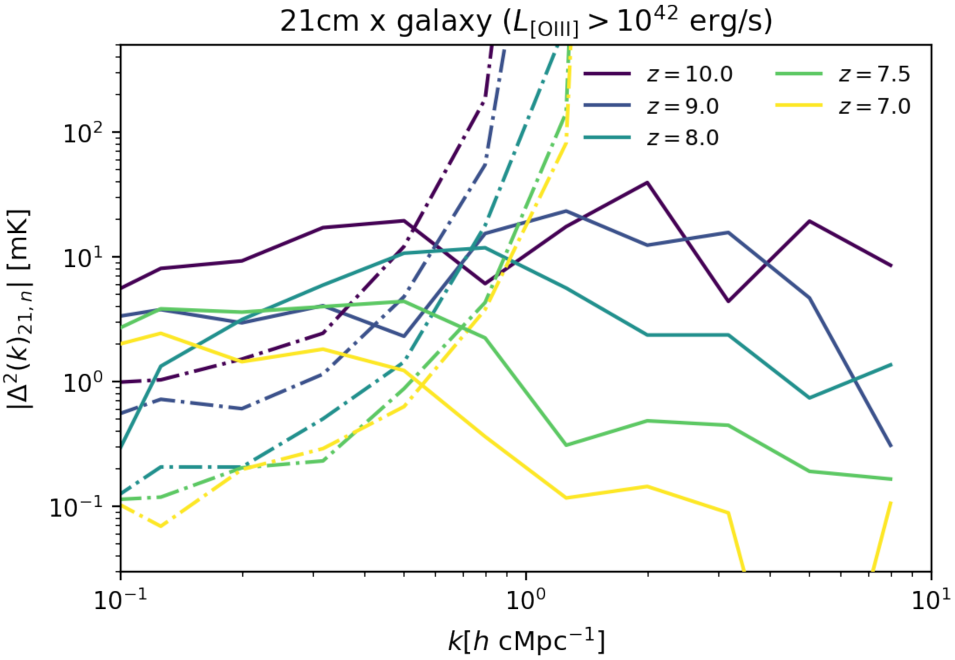


Figure 7.7: The amplitude of the cross-power spectra and their variance (Eq.7.7) are indicated by the solid and dashed lines.

large scales and the noise terms ($P_{\mathrm{noise},21}$, $P_{\mathrm{noise,gal}}$) dominate at small scales. We have almost the same S/N of the large-scale power in the saturated heating model.

Fig.7.7 shows a rapid increase of the noise terms at $k \gtrsim 0.5h$ Mpc^{-1}, indicating the difficulty of detecting small-scale signals. This increase is mainly due to the finite angular resolution of the 21-cm observations and the spectral resolution of the [OIII] observation. Fig7.8 shows how the errors change for different values of redshift uncertainty σ_z at $z = 10$. While the steep increase of the error is mitigated, still the signals on the smallest scales are not detectable. Detecting such signals is difficult even with the [OIII] surveys with a higher spectral resolution, and for that, we need higher resolutions for the 21-cm observations.

7.4 Conclusion

In this chapter, we have investigated how the high-redshift LSSs traced by the emission line galaxies can be used for studying the cosmic reionization and heating. While there have been several studies of the cosmic reionization using LIM data of [CII] (e.g., Dumitru et al. 2019), we propose to use [OIII] emissions for the first time. We use the post-processing simulation data to compute the 21-cm signals and compute the cross-power spectra using the distributions of the [OIII] emitters.

We find that the future observations of the 21-cm LIM by SKA and [OIII] LIM by CDIM-like

---

[3] This is comparable to the expected spectral resolution of the CDIM.

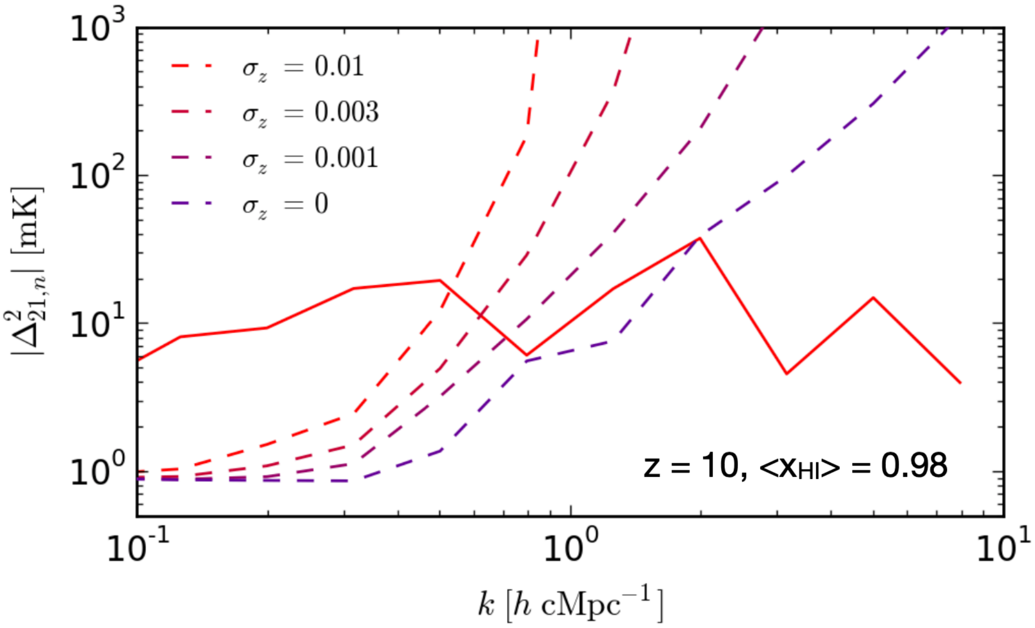


Figure 7.8: The variance of the cross-power spectrum at $z = 10$ for different redshift uncertainties.

telescope can detect the large-scale cross-power spectra. We note that we do not include the foreground and the interlopers, and de-confusion methods like what we devised in the previous chapters could be very important for robustly detecting such signals. The large-scale power spectra depend on the progress of the IGM heating as well as the reionization. The cross-power signals may provide more information than the auto-power spectra. In particular, we can use the positive-to-negative transition redshifts z_{to} as an indicator of the end of the heating and the beginning of the reionization. Once we detect the signals, we can also study the ionizing and heating sources in the early universe, including the first galaxies. To properly interpret the signals obtained in the future, it is crucial to examine various physical models as has been done in the case of the auto-power spectrum (e.g., Cohen et al. 2017).

Chapter 8

Summary and Outlook

Line intensity mapping (LIM) is an emerging observational technique to probe the three-dimensional large-scale structure (LSS) of the universe over a large volume efficiently. While it is a promising method, it suffers from large observational noise and line confusion. Several methods have been proposed to solve these problems in previous studies, but they mostly aim at detecting statistical signals.

In this thesis, we have proposed, for the first time, to use machine learning to extract line intensity maps from contaminated observational data. We have generated mock observational data assuming future near-infrared LIM observations (e.g., SPHEREx and CDIM) and demonstrated the performance of our methods.

In Chapter 4, we have trained conditional generative adversarial networks (cGANs) to extract the emission line signals from the noisy observational data. Our network is able to extract signals with higher precision than simply extracting the bright pixels from the observed maps. We find that the reproducibility depends on the choice of the resolution and size of the training data. When we use a training dataset with a fine resolution, bright point sources are extracted well. When a dataset with a larger area is used, the distributions of faint extended signals are properly extracted. Due to the limitation of computational cost, it is necessary to carefully choose the training data according to one's purpose.

In Chapter 5, we have constructed cGANs to separate the emission line interlopers from LIM data at a specific observational wavelength. The networks detect bright peaks of two different emission signals, Hα from $z = 1.3$ and [OIII] from $z = 2.0$, with 76 and 32 percent precision. The precision of the peak detection is improved to 0.90 and 0.57 when using multiple networks. We have investigated how the networks separate different emission signals by visualizing the convolutional filters. We have found that the networks extract characteristic features of the LSSs such as bright peaks and filaments. The extracted features may be used by the networks to distinguish the signals from different redshifts.

In Chapter 6, we have developed cGANs that can be applied to the three-dimensional LIM data. We have configured a network architecture that provides the machine with physical information. In particular, we input two observational data cubes in which there are the same underlying LSSs traced by two different emission lines, Hα and [OIII]. We have found that the networks can extract the bright peaks of individual emission lines with 84 and 68 percent precision. We have also investigated how the networks extract the signals by performing various analyses and it was suggested that the networks focus on the co-existence of multiple emission line signals in the spectral direction.

The extracted intensity maps tracing a specific redshift can be used for various studies in cosmology and astrophysics. As one of such applications of the LIM data, in Chapter 7, we have examined using the [OIII] intensity maps to study the cosmic reionization. We have generated line intensity maps using outputs of cosmological simulations and radiation transfer conducted by Eide et al. (2018), and computed the cross-correlation signals between the 21-cm line from the intergalactic medium (IGM) and the [OIII] emission lines from galaxies at the epoch of reionization. We have predicted the detectability of the signals and found that the future observations of the 21-cm intensity maps such as SKA and near-infrared intensity maps such as CDIM can detect the large-scale signals. Detected large-scale signals can be used to identify the epoch of reionization and IGM heating. We have found that the redshift where the large-scale power turns from positive to negative serves as a good indicator of the end of the heating and the beginning of the reionization. Our signal extraction methods are crucial for detecting such distant signals.

Future Prospects

One of the directions of future development of our methods is to include more than two emission lines. Inclusion of subdominant lines is important to extract individual signals more accurately. The networks that deal with every possible emission line would also allow us to extract faint but important signals such as the Lyα emissions from the epoch of reionization. In Chapter 6, we have found that the coexistence of multiple emission lines contributes a good reproducibility of the signals. Thus including additional emission lines could make the separation easier. If trained properly, the networks would also be able to remove spatially and spectrally smooth noise components such as the diffuse Galactic light and Zodiacal light.

Ultimately, we will require machines that can properly distinguish between various cosmological and astrophysical models so that the extracted data can be used to precisely constrain the physical models. For this purpose, it is crucial to train the networks with data of a large volume. This could be achieved by using a number of GPUs with larger memory or by tuning the architecture of the networks more carefully in the future.

There are several challenges in applying machine learning methods to actual observational data. One of the most serious problem is that it is often uncertain whether the training data really represents the observed data. In addition to the line emission model, the validity of the noise model should also be carefully examined as we have discussed in Section 4.5. We can tackle these problems with various approaches proposed in recent years. We may be able to distinguish differences between observational data with different physical models by using multiple machines trained with different models (e.g., Acquaviva et al. 2020). Methods for estimating uncertainties of outputs of the machine (e.g., Kendall & Gal 2017) could also be important to deal with the actual observational data. Another robust method is to generate the training data using a part of the actual observational data (e.g., Chen et al. 2018).

Understanding the machine's strategy is also important. Looking at the convolutional filters is informative, but it is difficult to understand the functions of all the filters in deep layers. Symbolic representation approaches (e.g., Cranmer et al. 2020), which reduce the effective number of features in hidden layers, would help us to perceive how a network separates the signals. Activation maximization methods (e.g., Mordvintsev et al. 2015), which visualize the features that maximize the activation in the hidden layers, may also allow us to capture the important large-scale features of the intensity maps. Such understanding methods would not only validate the strategy of the machine but also provide some hints about new ways of analyzing the LSS data.

Our machine learning methods are crucial to fully utilize future LIM observational data. In the observations of the LSSs, the size of the observable regions will determine the uncertainties of the resulting constraints once sufficient sensitivities are achieved. In this sense, LIM observations will have an important impact on future precision cosmology. The research in this thesis will contribute to solving various issues in cosmology and astrophysics, including the properties of dark energy and dark matter, physics in the early universe, the evolution of the universe, the formation and evolution of galaxies, and other astronomical phenomena.

Appendix A

Training of the Generative Models

A.1 Loss Functions of GANs

A generative model learns a probability distribution P_r that the training data follows. Some models directly estimate P_r, while others including VAE and GAN use a parametric function g_θ. Here, we consider the latter model. The parametric function such as a neural network generates data x from a random input z

$$G_\theta : z \in \mathcal{Z} \rightarrow x \in P_\theta, \tag{A.1}$$

where $\mathcal{Z}$ is the distribution that z follows. The model parameters θ are updated in training so that the probability distribution function of the output data $P_\theta(x)$ gets close to the true distribution function $P_r(x)$.

There are various ways to measure how close the two probability distributions are. We consider two of them, the Jensen-Shannon divergence and the Earth-Mover distance, which vanilla GAN (Goodfellow et al. 2014) and WGAN (Arjovsky et al. 2017) use for optimization, and discuss the difference between them. Below, $\mathcal{X}$ denotes the phase space of the data of our interest. For two-dimensional images with the number of pixels $N \times N$, for instance, it is $\mathcal{X} = \boldsymbol{R}^{N\times N}$.

- The Jensen-Shannon (JS) divergence is defined by

  $$JS(P_r, P_g) = KL(P_r||P_m) + KL(P_g||P_m), \tag{A.2}$$

  where P_m is the mixed distribution of the two, $P_m = (P_r + P_m)/2$, and

  $$KL(P_1||P_2) = \int P_1(x) \log\left(\frac{P_1(x)}{P_2(x)}\right) dx \tag{A.3}$$

  is the Kullback-Leibler (KL) divergence

- The Earth-Mover (EM) distance is defined by

  $$W(P_r, P_g) = \inf_{\gamma\in\Pi(P_r,P_g)} E_{(x,y)\sim\gamma}[||x - y||], \tag{A.4}$$

where $\Pi(P_r, P_g)$ is the set of all possible joint probability distributions $\gamma(x, y)$ that are reduced to $P_r(x)$ and $P_g(y)$ when integrated along the x-/y- axes

$$\int \gamma(x,y)dx = P_r(x), \quad \int \gamma(x,y)dy = P_g(y). \tag{A.5}$$

$E_{x\sim P}[*]$ represents an operation to compute the average of $[*]$ over data x generated from a probability distribution P.

If one wants to optimize a generator based on the JS divergence, one brute-force method is to generate a sufficiently large amount of data to compute $P_g(x)$ and take the gradient of Eq.A.2 using it. However, it is not practical to repeat such an enormous operation for each training step. Another possibility is to approximate Eq.A.2 using a discriminator. The true data generated from P_r and data generated from P_g are labeled as "T" and "F", respectively. From Bayes' theorem, the probability distribution functions are rewritten as

$$P_r(x) = P(\ x \mid \text{label} = \text{T}\ ) = \frac{P(x)P(\ \text{label} = \text{T} \mid x\ )}{P(\ \text{label} = \text{T}\ )}, \tag{A.6}$$

$$P_g(x) = P(\ x \mid \text{label} = \text{F}\ ) = \frac{P(x)P(\ \text{label} = \text{F} \mid x\ )}{P(\ \text{label} = \text{F}\ )}, \tag{A.7}$$

where $P(x) = P_r(x)+P_g(x)$. Here, $P(\ \text{label} = \text{T} \mid x\ )$ and $P(\ \text{label} = \text{F} \mid y\ )$ are the probabilities that the given data x is true and generated data, respectively. That is, they can be replaced with a discriminator D

$$P(\ \text{label} = \text{T} \mid x\ ) \sim D(x), \quad P(\ \text{label} = \text{F} \mid x\ ) \sim 1 - D(x) \tag{A.8}$$

as long as the discriminator is functioning appropriately. When we have the same amount of true and generated data, i.e., $P(\ \text{label} = \text{T}\ ) = P(\ \text{label} = \text{F}\ ) = 1/2$, then we get

$$\frac{P_r(x)}{P(x)} \sim 2D(x), \quad \frac{P_g(x)}{P(x)} \sim 2(1 - D(x)). \tag{A.9}$$

Therefore, we can rewrite the JS divergence as

$$JS(P_r, P_g) = \int \Big[P_r(x) \log\Big(\frac{2P_r(x)}{P_r(x) + P_g(x)}\Big) + P_g(x) \log\Big(\frac{2P_g(x)}{P_r(x) + P_g(x)}\Big)\Big]dx \tag{A.10}$$

$$\sim 2\log 4 + \int \Big[P_r(x) \log D(x) - P_g(x) \log(1 - D(x))\Big]dx \tag{A.11}$$

$$= 2\log 4 + E_{x\sim P_r} \log D(x) - E_{z\sim \mathcal{Z}} \log(1 - D(G(z))). \tag{A.12}$$

A practical version of this is the loss function defined in Eq.4.22. We now see that the generator of the vanilla GAN is trained to reduce the JS divergence approximated with the discriminator.

Next, let us think about optimizing the generator using the EM distance. For this, we make use of the fact that it can be rewritten as (Villani 2008)

$$W(P_r, P_g) = \sup_{||f||_L \leq 1} \Big[E_{x\sim P_r} f(x) - E_{x\sim P_g} f(x)\Big], \tag{A.13}$$

where $||f||_L \leq 1$ the 1-Lipschitz condition on a function $f : \mathcal{X} \rightarrow \boldsymbol{R}$

$$\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \leq 1. \tag{A.14}$$

The question is for what function f is the content of the brackets in Eq.A.13 (nearly) equal to the supremum? It might not be so easy to find such a function, but we can train a function $f = D$ to work appropriately, just as we train a discriminator D in the case of the vanilla GAN. When we use a function D that works appropriately, the EM distance is given by

$$W(P_r, P_g) \sim \Big[E_{x \sim P_r} D(x) - E_{z \sim \mathcal{Z}} D(G_\theta(z))\Big], \tag{A.15}$$

which is equivalent to the loss function of WGAN in Eq.4.26. The function D is called a critic.

So far, we have seen that the generators of the GAN and WGAN are optimized based on the JS divergence and the EM distance, respectively. We then compare these distance measures based on Arjovsky et al. (2017). Let us consider a one-dimensional space $\mathcal{X} = \boldsymbol{R}$, and probability distributions $P_r(x) = \delta(x)$ and $P_\theta = \delta(x - \theta)$, where δ is the one-dimensional Dirac delta function. In this case,

$$JS(P_r, P_\theta) = \begin{cases} \log 2 & (\theta \neq 0), \\ 0 & (\theta = 0), \end{cases} \tag{A.16}$$

$$W(P_r, P_\theta) = |\theta|. \tag{A.17}$$

The JS divergence is not continuous at $\theta = 0$, i.e., when $\theta \rightarrow 0$, $P_\theta \rightarrow \log 2 \neq P_{\theta=0}$, and thus it is not differentiable at $\theta = 0$. This discontinuity makes it impossible for the generator to reach the optimal parameter $\theta = 0$. Even in more mild situations, the gradient of the JS divergence with respect to the parameter θ tends to vanish $\nabla_\theta JS \sim 0$ when the discriminator learns very quickly. In such a case, the generator gets less information on how to optimize the parameters. On the other hand, it can be proven that the EM distance is always continuous everywhere in θ as long as the generator G_θ is continuous in θ (Arjovsky et al. 2017). Therefore, for WGAN, a clear gradient exists everywhere even after the critic is trained till optimality. This is thought to contribute to the more stable behavior of the WGAN.

A.2 Choice of Training Models and Datasets

Comparison with the Conventional Loss Function

In our study, we use GANs, in which discriminators are trained so that they distinguish the true and generated images and the generators are trained through adversarial loss functions. We can also use a more conventional loss function such as the L1 norm. In the signal extraction task from noisy LIM maps (Chapter 4), we find that the generator reconstructs maps with almost no signal when it is optimized only with the L1 loss function between the true and reconstructed maps (Eq.4.24). We also examine it in the signal separation task (Chapter 5). Fig.A.1 shows the result. The result is better than what we observed in the first test because the images are

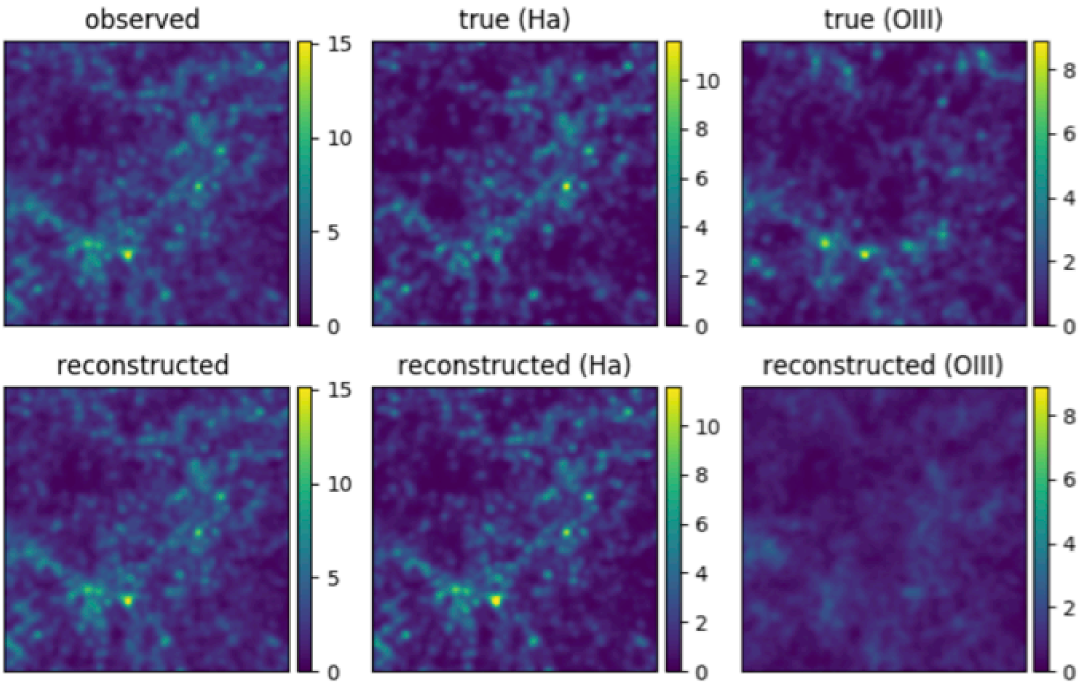


Figure A.1: The result of the signal separation (Chapter 5) trained with L1 norm. *Top*: the observed (Hα + [O III]), true Hα, and true [O III] line intensity maps. *Bottom*: the reconstructed Hα (middle) and [O III] (right) maps and the sum of them (left).

smoothed beforehand, the reconstructed [O III] map is blurry and the reproducibility of the Hα map is also degraded.

Such a blurring effect is known to be observed when using the VAEs as well and can be explained as follows.[1] Let us consider a simple example, where the generator is trained to reconstruct one-dimensional data with five pixels (Fig.A.2). We assume that the true LIM signals are sparse, i.e., most of the pixels have very small values and only a few pixels have bright signals (red lines). In the left panel, the generator learns to have a single bright pixel but mispredicts its location by only one pixel. while in the right panel, it distributes the signals uniformly across all pixels. The L1 norm is smaller for the right panel than the left. Therefore the generator would be optimized to take the strategy on the right panel unless it finds a way to perfectly predict the location of the bright pixel. In our signals extraction task, it is more important to detect individual peaks and to reproduce the overall statistics such as power spectra and PDFs. We thus use GANs with the L1 norm in our study.

---

[1] The effect is greater when the mean squared error (L2 norm, Eq.4.7) is adopted as a loss function (Isola et al. 2016).

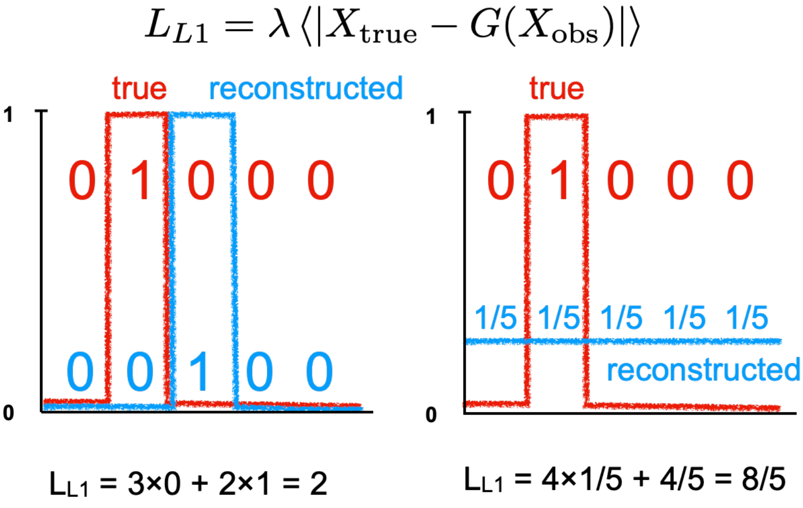


Figure A.2: Two different strategies of the generator for the reconstruction task of the one-dimensional data with five pixels. In the left, it predicts one bright signal but fails to reconstruct its location, while in the right, it predicts uniform density distribution with an expected mean value (1/5). They achieve $\mathcal{L}_{\rm L1} = 2$ (left) and 8/5 (right), respectively.

Hyperparameters

The coefficient of the L1 norm λ should be determined together with the normalization factor of the images $I_{\rm norm}$. We adopt the tanh activation function at the output layer of the generators, so pixel values of the output images are restricted to [-1, 1]. We adopt a value for $I_{\rm norm}$ that is barely larger than the largest pixel value in the training data. The L1 norm is considered to speed up learning in the early stages of training. We test several values of λ and find that too large λ results in blurry images. In the tasks of Chapters 5 and 6, the choice of λ does not change the final results a lot. In the tasks of Chapter 4, the results vary depending on the λ, but the variance is about the same as the variance obtained by bagging. Thus the variation is considered to be due to the randomness of the training process rather than the choice of λ.

We then examine how the result changes depending on the number of training data, batch size, and epoch. Fig.A.3 shows the evolution of the mean squared errors of the validation data in the task of Chapter 5 for several sets of hyperparameters. We stop some of the training (adopt small epochs) when no further change is expected. The number of training data is changed by the number of realizations $n_{\rm rea}$ as $n = 100 n_{\rm rea}$. We find that the final error does not depend on the choice of this number (see the green and orange lines, where we adopt the same values for other parameters). When the batch size is 1 (purple and red lines), the final error is worse than the other cases. The batch normalization with batch size 1 is known as instance normalization. It is known that the outputs tend to be independent of the contrast of the inputs when we adopt instance normalization (Ulyanov et al. 2016). The worse results for batch size 1 could be due to such a nature.[2] For the batch sizes > 1, the initial learning speed

---

[2] We perform the batch normalization for the validation using the moving mean and variance calculated

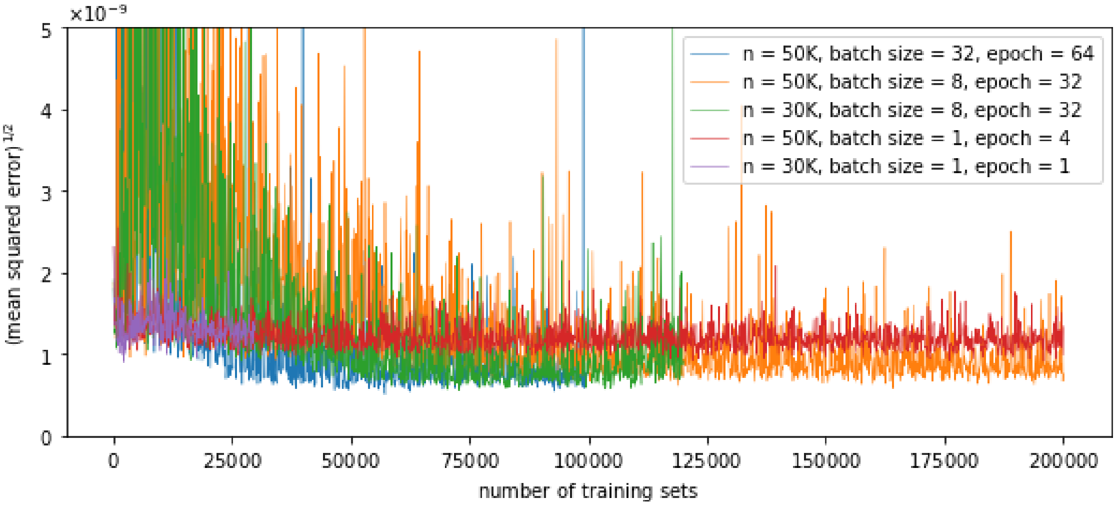


Figure A.3: The evolution of the mean squared error of the validation data for several sets of hyperparameters: the total number of training data n, batch size, and epoch.

becomes faster for larger sizes, but we get similar final errors. Such a result for a fixed learning rate is consistent with the previous more detailed study of the hyperparameters (Smith 2018). We also find that the error converges well when the epoch is about twice the batch size.

Pre-processing

In Chapter 5, we preprocess the images with Gaussian smoothing. Fig.A.4 shows the result without training. We find that the positions of clusters and filaments are reconstructed to the same extent as with smoothing, but the pixel-by-pixel reproducibility is reduced. Positions of the bright pixels in the [OIII] maps are reconstructed almost randomly within each cluster region. This may be because the generator mainly focuses on the LSS to separate the individual signals.

---

during the training, but the effect of the instance normalization could still affect the training process.

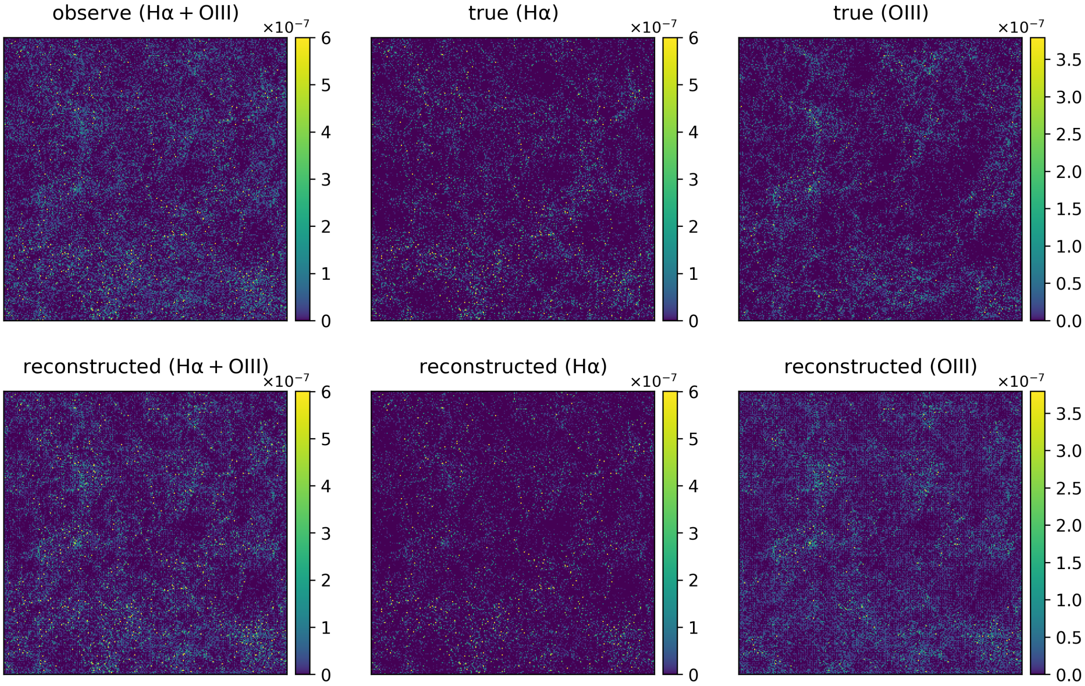


Figure A.4: The result of signal separation (Chapter 5) without smoothing.

Appendix B

21-cm Line from Intergalactic Medium

B.1 Brightness Temperature

The 21-cm line is radiated by the transition between the parallel (state 1) and antiparallel (state 0) states of the spin of the neutral hydrogen nucleus and electrons. It has an energy $E_{10} = 5.9 \times 10^{-6}$ eV corresponding to the rest-frame frequency $\nu_0 = 1.42$ GHz and the rest-frame wavelength $\lambda_0 = 21.1$ cm. The 21-cm line is observed as emission or absorption against the background radiation field. We consider the CMB as the background in general. At the observed frequency of the 21-cm line, the CMB obeys the Rayleigh-Jeans law, and the intensity I_ν is conventionally represented by the brightness temperature

$$T = \frac{c^2 I_\nu}{2 k_{\rm B} \nu^2}. \tag{B.1}$$

The spontaneous transition probability from state 1 to 0 (Einstein A coefficient) is as small as (Wild 1952)

$$A_{10} = 2.85 \times 10^{-15}\ {\rm s}^{-1}. \tag{B.2}$$

The emissivity of the 21-cm line is determined by the ratio between the number densities of atoms in the two hyperfine states. The spin temperature T_s is defined by

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} \exp\left(-\frac{h\nu_0}{k_{\rm B} T_s}\right) = 3 \exp\left(-\frac{T_*}{T_s}\right), \tag{B.3}$$

where $T_* = h\nu_0 / k_{\rm B} = 0068$ K is the excitation temperature.

The spin temperature is determined by three processes: the absorption of CMB photons, collisions with hydrogen atoms and electrons, and Lyα coupling (Wouthuysen-Field process; Field 1958; Wouthuysen 1952). In the equilibrium state,

$$n_1 (C_{10} + P_{10} + A_{10} + B_{10} I_{\rm CMB}) = n_0 (C_{01} + P_{01} + B_{01} I_{\rm CMB}), \tag{B.4}$$

where $C_{\rm ij}$ is the (de-)excitation rate via collisions, $P_{\rm ij}$ is the (de-)excitation rate via the Wouthuysen-Field process, $B_{\rm ij}$ is the Einstein B coefficient, and $I_{\rm CMB}$ is the CMB intensity. In the Rayleigh-Jeans limit, this equation is reduced to (Wild 1952)

$$T_s^{-1} = \frac{T_{\rm CMB}^{-1} + x_c T_{\rm kin}^{-1} + x_\alpha T_c^{-1}}{1 + x_c + x_\alpha}, \tag{B.5}$$

where $T_{\rm kin}$ is the kinetic temperature of the gas, T_c is the color temperature of UV light defined by $P_{01}/P_{10} = g_1/g_0 \exp(-T_*/T_c)$, and x_c and x_α are the coupling constants. After decoupling, the photon cools down as $T_\gamma \propto (1+z)$ by adiabatic expansion. The gas temperature also remains at the same temperature as the photon until $z \sim 100$ because the small amount of remaining ionized component exchanges energy with the photon through Compton scattering. As the universe expands, photons and baryons eventually decouple and the gas temperature drops as $T_{\rm gas} \propto (1+z)^2$. At this moment, the collision term dominates, resulting in $T_s \sim T_{\rm kin} < T_{\rm CMB}$, but it gets inefficient as the expansion proceeds further. After the formation of the first galaxies, Lyα coupling becomes dominant. As the Lyα photons exchange energy with the gas via the recoil of the hydrogen atoms, the Lyα spectrum is thermalized, and we have $T_s \sim T_c \sim T_{\rm kin}$. The IGM is then warmed by high-energy photons from the galaxy, and the spin temperature increases as well. At $z < 10$, we can assume that the Lyα coupling is sufficiently effective, so we can assume $T_s = T_{\rm kin}$.

Now, consider the case where the background radiation with brightness T_R passes through a uniform cloud with a number density of neutral hydrogen $n_{\rm HI}$ and spin temperature T_s at redshift z. The cloud absorbs some of the photons with $\nu = \nu_0$ and re-radiates them. The radiation transfer equation is

$$T_b(\nu) = T_s(1 - {\rm e}^{-\tau_\nu}) + T_R(\nu){\rm e}^{-\tau_\nu}, \tag{B.6}$$

where τ_ν is the optical depth. It is given by the integral of the absorption coefficient (e.g. Field 1958; Furlanetto et al. 2006)

$$\tau_\nu = \int (1 - {\rm e}^{-T_*/T_s})\sigma(\nu) n_0 \, ds, \tag{B.7}$$

where

$$\sigma(\nu) = \frac{3c^2 A_{10}}{8\pi\nu^2}\phi(\nu) \tag{B.8}$$

is the cross-section, and $\phi(\nu)$ is the line shape. If we only consider the line broadening due to the bulk motion, we get

$$\phi(\nu) ds \sim \frac{c/\nu}{(1+z)(dv_{||}/dr_{||})}\delta(\nu - \nu_0) d\nu, \tag{B.9}$$

where $dv_{||}/dr_{||}$ is the velocity gradient of the cloud along the travel direction of the photon, and δ is the one-dimensional Dirac delta function. In general, the spin temperature is much larger than T_*, and we have $n_0 \sim n_{\rm HI}/4$ from Eq. B.3. The optical depth is then written as

$$\tau_\nu = \frac{3c^2 h A_{10} n_{\rm HI}}{32\pi\nu_0^2 k_{\rm B} T_s (1+z)(dv_{||}/dr_{||})}. \tag{B.10}$$

In the following, we consider that only the Hubble expansion contributes to the line-of-sight velocity $v_{||}$ and replace the gradient term $(1+z)(dv_{||}/dr_{||})$ with the Hubble constant $H(z)$, which is given by

$$H(z) = H_0\sqrt{(1+z)^3\Omega_{\mathrm{m}}} \tag{B.11}$$

in the matter-dominated era.

When $\tau_\nu \ll 1$, the brightness temperature measured with respect to the background is given by

$$\begin{aligned}\delta T_{\mathrm{b}}^{\mathrm{obs}}(\nu) &\equiv \frac{T_{\mathrm{b}}^{\mathrm{obs}}(\nu) - T_{\mathrm{R}}}{1+z} = \frac{T_{\mathrm{s}} - T_{\mathrm{R}}}{1+z}(1-\mathrm{e}^{-\tau_\nu}) \\ &\sim \frac{T_{\mathrm{s}} - T_{\mathrm{R}}}{1+z}\tau_\nu. \end{aligned} \tag{B.12}$$

Here, we define the hydrogen neutral fraction x_{HI} as

$$n_{\mathrm{HI}} = x_{\mathrm{HI}} n_{\mathrm{H}}, \tag{B.13}$$

where n_{H} is the number density of the hydrogen atoms. The mean number density of the hydrogen atoms in the primordial IGM[1] at redshift z is given by

$$\overline{n}_{\mathrm{H}} \sim \frac{\rho_{\mathrm{b}}}{1.3m_{\mathrm{H}}} \sim 8.6\times10^{-6}\Omega_{\mathrm{b}}h^2(1+z)^3 \ \mathrm{cm}^{-3}. \tag{B.14}$$

Therefore the brightness temperature is given by Eq.7.2.

B.2 Noise Power Spectrum

In 21-cm observations using the interferometer, the noise in the power spectrum is given by (McQuinn et al. 2006)

$$P_{\mathrm{noise,21}} = \frac{T_{\mathrm{sys}}^2}{Bt_{\mathrm{int}}}\frac{D^2\Delta D}{n(k_\perp)}\left(\frac{\lambda^2}{A_e}\right)^2, \tag{B.15}$$

where $k_\perp = (1-\mu^2)^{1/2}k$ is the component of the wavenumber vector perpendicular to the line-of-sight direction, t_{int} is the observational time, T_{sys} is the system temperature, D is the comoving distance to observation target, ΔD is the comoving depth of the survey volume, B is the corresponding survey bandpass, $n(k_\perp)$ is the number density of baselines that observes $k_\perp$ mode, $\lambda = 21(1+z)$ cm is the observed wavelength, and A_e is the effective area of the telescope. ΔD and B are related as

$$\Delta D = 1.7\left(\frac{B}{0.1\mathrm{MHz}}\right)\left(\frac{1+z}{10}\right)^{1/2}\left(\frac{\Omega_{\mathrm{m}}h^2}{0.15}\right)^{-1/2}. \tag{B.16}$$

---

[1] We ignore heavy elements.

We consider SKA observation with $t_{\rm int} = 1000$ hours. We adopt $T_{\rm sys} = 280[(1+z)/7.5]^{2.3}$ K and $a_e = 462(1+z)^2/(1+8)^2$ following Kubota et al. (2018). For $n(k_\perp)$, we adopt a simple decreasing function

$$n(k_\perp) \propto \begin{cases} r^{-2} & (20 \text{ m} < r < 1000 \text{ m}) \\ 0 & (\text{otherwise}) \end{cases} \tag{B.17}$$

with 670 antenna tiles in total.

Acknowledgement

First of all, I would like to express my sincere gratitude to my supervisor, Professor Naoki Yoshida, for his continuous support and encouragement. His insightful feedback and ideas brought my research to a higher level. I learned how to conduct research, find a new topic, write a paper, and give a presentation from him. Even when I was depressed about my research not going well, discussions with him helped me regain my motivation. He also gave me opportunities to meet many researchers, which not only made a significant contribution to this thesis but will be invaluable in my future research life.

I thank Marius Berge Eide and Benedetta Ciardi for providing simulation data and fruitful discussion on cosmic reionization. I am grateful to Masato Shirasaki for giving many helpful comments on my research. He also taught me how to train a machine learning model. I would also like to thank Yasuhiro Imoto for his comments on machine learning. I would like to express my appreciation to Anastasia Fialkov for accepting me as a guest remote member of her group at the Institute of Astronomy, Cambridge, and giving many insightful comments.

I thank the committee members of this dissertation, Masahiro Takada, Hideyuki Tagoshi, Satoshi Yamamoto, Kenta Hotokezaka, and Yoshiyuki Kabashima, for their insightful comments. I also thank the following people: Ikkoh Shimizu and Shigeki Inoue, who taught me how to analyze the simulation data; Tohru Nagao, Akio Inoue, Takuya Hashimoto, Yoichi Tamura, Hiroshi Matsuo, Yuichi Matsuda, and Yuichi Harikane, whose insight and knowledge helped me to build the line emission model; Members of the EoR working group of the Japan SKA Consortium, whose comments deepened my research.

I would like to express my gratitude to the following people: Kazumi Kashiyama, for his many comments on my research and casual but actually kind encouragement; Tilman Hartwig, for the fruitful discussion on machine learning and helping me to use the GPU; Yuta Tarumi, for lots of fun discussions about research and other things. I also thank all the other members of the University of Tokyo Theoretical Astrophysical Group and Research Center for the Early Universe for many fruitful discussions and good times we shared. I really enjoyed my five years as a graduate student very much with you. I would like to give a special thanks to Rui Lan Zhang for her help in proofreading this thesis.

Finally, I deeply thank my parents, Katsumi and Kazunori, and husband, Ryo, for their patience and encouragement. I would not have been able to continue my research without them.

I acknowledge the support from JSPS KAKENHI Grant Number JP19J21379, the Advanced

Leading Graduate Course for Photon Science (ALPS) of the University of Tokyo, and the JSR Fellowship, the University of Tokyo. I appreciate the support by First Light School in Sao Paulo. Part of the calculations in this work were carried out on Cray XC50 at Center for Computational Astrophysics, National Astronomical Observatory of Japan.

Bibliography

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. 2016, in 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) (Savannah, GA: USENIX Association), 265–283

Ackermann, S., Schawinski, K., Zhang, C., Weigel, A. K., and Turp, M. D., Using transfer learning to detect galaxy mergers, Sep. 2018, MNRAS, 479, 415

Acquaviva, V., Lovell, C., and Ishida, E., Debunking Generalization Error or: How I Learned to Stop Worrying and Love My Training Set, Nov. 2020, arXiv e-prints, arXiv:2012.00066

Ade, P. A. R., Anderson, C. J., Barrentine, E. M., Bellis, N. G., Bolatto, A. D., Breysse, P. C., Bulcha, B. T., Cataldo, G., Connors, J. A., Cursey, P. W., Ehsan, N., Grant, H. C., Essinger-Hileman, T. M., Hess, L. A., Kimball, M. O., Kogut, A. J., Lamb, A. D., Lowe, L. N., Mauskopf, P. D., McMahon, J., Mirzaei, M., Moseley, S. H., Mugge-Durum, J. W., Noroozian, O., Pen, U., Pullen, A. R., Rodriguez, S., Shirron, P. J., Somerville, R. S., Stevenson, T. R., Switzer, E. R., Tucker, C., Visbal, E., Volpert, C. G., Wollack, E. J., and Yang, S., The Experiment for Cryogenic Large-Aperture Intensity Mapping (EXCLAIM), Jan. 2020, Journal of Low Temperature Physics, 199, 1027

Ali, Z. S., Parsons, A. R., Zheng, H., Pober, J. C., Liu, A., Aguirre, J. E., Bradley, R. F., Bernardi, G., Carilli, C. L., Cheng, C., DeBoer, D. R., Dexter, M. R., Grobbelaar, J., Horrell, J., Jacobs, D. C., Klima, P., MacMahon, D. H. E., Maree, M., Moore, D. F., Razavi, N., Stefan, I. I., Walbrugh, W. P., and Walker, A., PAPER-64 Constraints on Reionization: The 21 cm Power Spectrum at z = 8.4, Aug. 2015, ApJ, 809, 61

Arjovsky, M., Chintala, S., and Bottou, L., Wasserstein GAN, Jan. 2017, arXiv e-prints, arXiv:1701.07875

Ball, N. M. and Brunner, R. J., Data Mining and Machine Learning in Astronomy, Jan. 2010, International Journal of Modern Physics D, 19, 1049

Baron, D. and Poznanski, D., The weirdest SDSS galaxies: results from an outlier detection algorithm, Mar. 2017, MNRAS, 465, 4530

Battye, R., Browne, I., Chen, T., Dickinson, C., Harper, S., Olivari, L., Peel, M., Remazeilles, M., Roychowdhury, S., Wilkinson, P., Abdalla, E., Abramo, R., Ferreira, E., Wuensche, A., Vilella, T., Caldas, M., Tancredi, G., Refregier, A., Monstein, C., Abdalla, F., Pourtsidou, A., Maffei, B., Pisano, G., and Ma, Y.-Z., Update on the BINGO 21cm intensity mapping experiment, Oct. 2016, arXiv e-prints, arXiv:1610.06826

Bigot-Sazy, M. A., Ma, Y. Z., Battye, R. A., Browne, I. W. A., Chen, T., Dickinson, C., Harper, S., Maffei, B., Olivari, L. C., and Wilkinsondagger, P. N. 2016, in Astronomical Society of the Pacific Conference Series, Vol. 502, Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015, ed. L. Qain & D. Li, 41

Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., and Mahesh, N., An absorption profile centred at 78 megahertz in the sky-averaged spectrum, Mar. 2018, Nature, 555, 67

Bowman, J. D., Cairns, I., Kaplan, D. L., Murphy, T., Oberoi, D., Staveley-Smith, L., Arcus, W., Barnes, D. G., Bernardi, G., Briggs, F. H., Brown, S., Bunton, J. D., Burgasser, A. J., Cappallo, R. J., Chatterjee, S., Corey, B. E., Coster, A., Deshpande, A., deSouza, L., Emrich, D., Erickson, P., Goeke, R. F., Gaensler, B. M., Greenhill, L. J., Harvey-Smith, L., Hazelton, B. J., Herne, D., Hewitt, J. N., Johnston-Hollitt, M., Kasper, J. C., Kincaid, B. B., Koenig, R., Kratzenberg, E., Lonsdale, C. J., Lynch, M. J., Matthews, L. D., McWhirter, S. R., Mitchell, D. A., Morales, M. F., Morgan, E. H., Ord, S. M., Pathikulangara, J., Prabu, T., Remillard, R. A., Robishaw, T., Rogers, A. E. E., Roshi, A. A., Salah, J. E., Sault, R. J., Shankar, N. U., Srivani, K. S., Stevens, J. B., Subrahmanyan, R., Tingay, S. J., Wayth, R. B., Waterson, M., Webster, R. L., Whitney, A. R., Williams, A. J., Williams, C. L., and Wyithe, J. S. B., Science with the Murchison Widefield Array, Apr. 2013, PASA, 30, e031

Breen, P. G., Foley, C. N., Boekholt, T., and Portegies Zwart, S., Newton versus the machine: solving the chaotic three-body problem using deep neural networks, May. 2020, MNRAS, 494, 2465

Breiman, L., Bagging Predictors. 1996, Machine Learning, 24, 123

Breysse, P. C., Kovetz, E. D., Behroozi, P. S., Dai, L., and Kamionkowski, M., Insights from probability distribution functions of intensity maps, May. 2017, MNRAS, 467, 2996

Bruzual, G. and Charlot, S., Stellar population synthesis at the resolution of 2003, Oct. 2003, MNRAS, 344, 1000

Bull, P., Extending Cosmological Tests of General Relativity with the Square Kilometre Array, Jan. 2016, ApJ, 817, 26

Caldeira, J., Wu, W. L. K., Nord, B., Avestruz, C., Trivedi, S., and Story, K. T., DeepCMB: Lensing reconstruction of the cosmic microwave background with deep neural networks, Jul. 2019, Astronomy and Computing, 28, 100307

Carilli, C. 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14), 171

Chabrier, G., Galactic Stellar and Substellar Initial Mass Function, Jul. 2003, PASP, 115, 763

Chang, T.-C., Pen, U.-L., Bandura, K., and Peterson, J. B., An intensity map of hydrogen 21-cm emission at redshift z~0.8, Jul. 2010, Nature, 466, 463

Chen, J., Chen, J., Chao, H., and Yang, M. 2018, in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3155–3164

Cheng, Y.-T. and Chang, T.-C., Cosmic Near-Infrared Background Tomography with SPHEREx Using Galaxy Cross-Correlations, Sep. 2021, arXiv e-prints, arXiv:2109.10914

Cheng, Y.-T., Chang, T.-C., Bock, J., Bradford, C. M., and Cooray, A., Spectral Line Deconfusion in an Intensity Mapping Survey, Dec. 2016, ApJ, 832, 165

Cheng, Y.-T., Chang, T.-C., and Bock, J. J., Phase-space Spectral Line Deconfusion in Intensity Mapping, Oct. 2020, ApJ, 901, 142

Ciardi, B., Ferrara, A., Marri, S., and Raimondo, G., Cosmological reionization around the first stars: Monte Carlo radiative transfer, Jun. 2001, MNRAS, 324, 381

Cohen, A., Fialkov, A., Barkana, R., and Lotem, M., Charting the parameter space of the global 21-cm signal, Dec. 2017, MNRAS, 472, 1915

Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N. J. G., Dalton, G., Eke, V. R., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Jenkins, A., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Peterson, B. A., Sutherland, W., and Taylor, K., The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications, Sep. 2005, MNRAS, 362, 505

Colless, M., Dalton, G., Maddox, S., Sutherland, W., Norberg, P., Cole, S., Bland-Hawthorn, J., Bridges, T., Cannon, R., Collins, C., Couch, W., Cross, N., Deeley, K., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Madgwick, D., Peacock, J. A., Peterson, B. A., Price, I., Seaborne, M., and Taylor, K., The 2dF Galaxy Redshift Survey: spectra and redshifts, Dec. 2001, MNRAS, 328, 1039

Concerto Collaboration, Ade, P., Aravena, M., Barria, E., Beelen, A., Benoit, A., Béthermin, M., Bounmy, J., Bourrion, O., Bres, G., De Breuck, C., Calvo, M., Cao, Y., Catalano, A., Désert, F. X., Durán, C. A., Fasano, A., Fenouillet, T., Garcia, J., Garde, G., Goupy, J., Groppi, C., Hoarau, C., Lagache, G., Lambert, J. C., Leggeri, J. P., Levy-Bertrand, F., Macías-Pérez, J., Mani, H., Marpaud, J., Mauskopf, P., Monfardini, A., Pisano, G., Ponthieu, N., Prieur, L., Roni, S., Roudier, S., Tourres, D., and Tucker, C., A wide field-of-view low-resolution spectrometer at APEX: Instrument design and scientific forecast, Oct. 2020, A&A, 642, A60

Cooray, A., Chang, T.-C., Unwin, S., Zemcov, M., Coffey, A., Morrissey, P., Raouf, N., Lipscy, S., Shannon, M., Wu, G., Cen, R., Ram Chary, R., Doré, O., Fan, X., Fazio, G. G., Finkelstein, S. L., Heneka, C., Lee, B., Linden, P., Nayyeri, H., Rhodes, J., Sadoun, R., Silva, M. B., Trac, H., Wu, H.-Y., and Zheng, Z., CDIM: Cosmic Dawn Intensity Mapper Final Report, Mar. 2019, arXiv e-prints, arXiv:1903.03144

Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., and Ho, S., Discovering Symbolic Models from Deep Learning with Inductive Biases, Jun. 2020, arXiv e-prints, arXiv:2006.11287

Creque-Sarbinowski, C. and Kamionkowski, M., Searching for decaying and annihilating dark matter with line intensity mapping, Sep. 2018, Phys. Rev. D, 98, 063524

Crites, A. T., Bock, J. J., Bradford, C. M., Chang, T. C., Cooray, A. R., Duband, L., Gong, Y., Hailey-Dunsheath, S., Hunacek, J., Koch, P. M., Li, C. T., O'Brient, R. C., Prouve, T., Shirokoff, E., Silva, M. B., Staniszewski, Z., Uzgil, B., and Zemcov, M. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9153, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII, ed. W. S. Holland & J. Zmuidzinas, 91531W

Croft, R. A. C., Di Matteo, T., Springel, V., and Hernquist, L., Galaxy morphology, kinematics and clustering in a hydrodynamic simulation of a Λ cold dark matter universe, Nov. 2009, MNRAS, 400, 43

Cui, K., Liu, J., Feng, F., and Liu, J., Identify Light Curve Signals with Deep Learning Based Object Detection Algorithm. I. Transit Detection, Aug. 2021, arXiv e-prints, arXiv:2108.00670

Dalal, N., Doré, O., Huterer, D., and Shirokov, A., Imprints of primordial non-Gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects, Jun. 2008, Phys. Rev. D, 77, 123514

Davis, M., Faber, S. M., Newman, J., Phillips, A. C., Ellis, R. S., Steidel, C. C., Conselice, C., Coil, A. L., Finkbeiner, D. P., Koo, D. C., Guhathakurta, P., Weiner, B., Schiavon, R., Willmer, C., Kaiser, N., Luppino, G. A., Wirth, G., Connolly, A., Eisenhardt, P., Cooper, M., and Gerke, B. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4834, Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II, ed. P. Guhathakurta, 161–172

Dawson, K. S., Schlegel, D. J., Ahn, C. P., Anderson, S. F., Aubourg, É., Bailey, S., Barkhouser, R. H., Bautista, J. E., Beifiori, A., Berlind, A. A., Bhardwaj, V., Bizyaev, D., Blake, C. H., Blanton, M. R., Blomqvist, M., Bolton, A. S., Borde, A., Bovy, J., Brandt, W. N., Brewington, H., Brinkmann, J., Brown, P. J., Brownstein, J. R., Bundy, K., Busca, N. G., Carithers, W., Carnero, A. R., Carr, M. A., Chen, Y., Comparat, J., Connolly, N., Cope, F., Croft, R. A. C., Cuesta, A. J., da Costa, L. N., Davenport, J. R. A., Delubac, T., de Putter, R., Dhital, S., Ealet, A., Ebelke, G. L., Eisenstein, D. J., Escoffier, S., Fan, X., Filiz Ak, N., Finley, H., Font-Ribera, A., Génova-Santos, R., Gunn, J. E., Guo, H., Haggard, D., Hall, P. B., Hamilton, J.-C., Harris, B., Harris, D. W., Ho, S., Hogg, D. W., Holder, D., Honscheid, K., Huehnerhoff, J., Jordan, B., Jordan, W. P., Kauffmann, G., Kazin, E. A., Kirkby, D., Klaene, M. A., Kneib, J.-P., Le Goff, J.-M., Lee, K.-G., Long, D. C., Loomis, C. P., Lundgren, B., Lupton, R. H., Maia, M. A. G., Makler, M., Malanushenko, E., Malanushenko, V., Mandelbaum, R., Manera, M., Maraston, C., Margala, D., Masters, K. L., McBride, C. K., McDonald, P., McGreer, I. D., McMahon, R. G., Mena, O., Miralda-Escudé, J., Montero-Dorta, A. D., Montesano, F., Muna, D., Myers, A. D., Naugle, T., Nichol, R. C., Noterdaeme, P., Nuza, S. E., Olmstead, M. D., Oravetz, A., Oravetz, D. J., Owen, R., Padmanabhan, N., Palanque-Delabrouille, N., Pan, K., Parejko, J. K., Pâris, I., Percival, W. J., Pérez-Fournon, I., Pérez-Ràfols, I., Petitjean, P., Pfaffenberger, R., Pforr, J., Pieri, M. M., Prada, F., Price-Whelan, A. M., Raddick, M. J., Rebolo, R., Rich, J., Richards, G. T., Rockosi, C. M., Roe, N. A., Ross, A. J., Ross, N. P., Rossi, G., Rubiño-Martin, J. A., Samushia, L., Sánchez, A. G., Sayres, C., Schmidt, S. J., Schneider, D. P., Scóccola, C. G., Seo, H.-J., Shelden, A., Sheldon, E., Shen, Y., Shu, Y., Slosar, A., Smee, S. A., Snedden, S. A., Stauffer, F., Steele, O., Strauss, M. A., Streblyanska, A., Suzuki, N., Swanson, M. E. C., Tal, T., Tanaka, M., Thomas, D., Tinker, J. L., Tojeiro, R., Tremonti, C. A., Vargas Magaña, M., Verde, L., Viel, M., Wake, D. A., Watson, M., Weaver, B. A., Weinberg, D. H., Weiner, B. J., West,

A. A., White, M., Wood-Vasey, W. M., Yeche, C., Zehavi, I., Zhao, G.-B., and Zheng, Z., The Baryon Oscillation Spectroscopic Survey of SDSS-III, Jan. 2013, AJ, 145, 10

de Lapparent, V., Geller, M. J., and Huchra, J. P., A Slice of the Universe, Mar. 1986, ApJ, 302, L1

DeBoer, D. R., Parsons, A. R., Aguirre, J. E., Alexander, P., Ali, Z. S., Beardsley, A. P., Bernardi, G., Bowman, J. D., Bradley, R. F., Carilli, C. L., Cheng, C., de Lera Acedo, E., Dillon, J. S., Ewall-Wice, A., Fadana, G., Fagnoni, N., Fritz, R., Furlanetto, S. R., Glendenning, B., Greig, B., Grobbelaar, J., Hazelton, B. J., Hewitt, J. N., Hickish, J., Jacobs, D. C., Julius, A., Kariseb, M., Kohn, S. A., Lekalake, T., Liu, A., Loots, A., MacMahon, D., Malan, L., Malgas, C., Maree, M., Martinot, Z., Mathison, N., Matsetela, E., Mesinger, A., Morales, M. F., Neben, A. R., Patra, N., Pieterse, S., Pober, J. C., Razavi-Ghods, N., Ringuette, J., Robnett, J., Rosie, K., Sell, R., Smith, C., Syce, A., Tegmark, M., Thyagarajan, N., Williams, P. K. G., and Zheng, H., Hydrogen Epoch of Reionization Array (HERA), mar. 2017, Publications of the Astronomical Society of the Pacific, 129, 045001

Degraf, C., Di Matteo, T., and Springel, V., Faint-end quasar luminosity functions from cosmological hydrodynamic simulations, Mar. 2010, MNRAS, 402, 1927

DESI Collaboration, Aghamousa, A., Aguilar, J., Ahlen, S., Alam, S., Allen, L. E., Allende Prieto, C., Annis, J., Bailey, S., Balland, C., Ballester, O., Baltay, C., Beaufore, L., Bebek, C., Beers, T. C., Bell, E. F., Bernal, J. L., Besuner, R., Beutler, F., Blake, C., Bleuler, H., Blomqvist, M., Blum, R., Bolton, A. S., Briceno, C., Brooks, D., Brownstein, J. R., Buckley-Geer, E., Burden, A., Burtin, E., Busca, N. G., Cahn, R. N., Cai, Y.-C., Cardiel-Sas, L., Carlberg, R. G., Carton, P.-H., Casas, R., Castander, F. J., Cervantes-Cota, J. L., Claybaugh, T. M., Close, M., Coker, C. T., Cole, S., Comparat, J., Cooper, A. P., Cousinou, M. C., Crocce, M., Cuby, J.-G., Cunningham, D. P., Davis, T. M., Dawson, K. S., de la Macorra, A., De Vicente, J., Delubac, T., Derwent, M., Dey, A., Dhungana, G., Ding, Z., Doel, P., Duan, Y. T., Ealet, A., Edelstein, J., Eftekharzadeh, S., Eisenstein, D. J., Elliott, A., Escoffier, S., Evatt, M., Fagrelius, P., Fan, X., Fanning, K., Farahi, A., Farihi, J., Favole, G., Feng, Y., Fernandez, E., Findlay, J. R., Finkbeiner, D. P., Fitzpatrick, M. J., Flaugher, B., Flender, S., Font-Ribera, A., Forero-Romero, J. E., Fosalba, P., Frenk, C. S., Fumagalli, M., Gaensicke, B. T., Gallo, G., Garcia-Bellido, J., Gaztanaga, E., Pietro Gentile Fusillo, N., Gerard, T., Gershkovich, I., Giannantonio, T., Gillet, D., Gonzalez-de-Rivera, G., Gonzalez-Perez, V., Gott, S., Graur, O., Gutierrez, G., Guy, J., Habib, S., Heetderks, H., Heetderks, I., Heitmann, K., Hellwing, W. A., Herrera, D. A., Ho, S., Holland, S., Honscheid, K., Huff, E., Hutchinson, T. A., Huterer, D., Hwang, H. S., Illa Laguna, J. M., Ishikawa, Y., Jacobs, D., Jeffrey, N., Jelinsky, P., Jennings, E., Jiang, L., Jimenez, J., Johnson, J., Joyce, R., Jullo, E., Juneau, S., Kama, S., Karcher, A., Karkar, S., Kehoe, R., Kennamer, N., Kent, S., Kilbinger, M., Kim, A. G., Kirkby, D., Kisner, T., Kitanidis, E., Kneib, J.-P., Koposov, S., Kovacs, E., Koyama, K., Kremin, A., Kron, R., Kronig, L., Kueter-Young, A., Lacey, C. G., Lafever, R., Lahav, O., Lambert, A., Lampton, M., Landriau, M., Lang, D., Lauer, T. R., Le Goff, J.-M., Le Guillou, L., Le Van Suu, A., Lee, J. H., Lee, S.-J., Leitner, D., Lesser, M., Levi, M. E., L'Huillier, B., Li, B., Liang, M., Lin, H., Linder, E., Loebman, S. R., Lukić, Z., Ma, J., MacCrann, N., Magneville, C., Makarem, L., Manera, M., Manser, C. J., Marshall, R., Martini, P., Massey, R., Matheson, T., McCauley, J., McDonald, P., McGreer, I. D., Meisner, A., Metcalfe, N., Miller, T. N., Miquel, R., Moustakas, J., Myers, A., Naik, M., Newman, J. A., Nichol, R. C., Nicola, A., Nicolati da Costa, L., Nie, J., Niz, G., Norberg, P., Nord, B., Norman, D., Nugent, P., O'Brien, T., Oh, M., Olsen, K. A. G., Padilla, C., Padmanabhan, H., Padmanabhan, N., Palanque-Delabrouille, N., Palmese, A., Pappalardo, D., Pâris, I., Park, C., Patej, A., Peacock, J. A., Peiris, H. V., Peng, X., Percival, W. J., Perruchot, S., Pieri, M. M., Pogge, R., Pollack, J. E., Poppett, C., Prada, F., Prakash, A., Probst, R. G., Rabinowitz, D., Raichoor, A., Ree, C. H., Refregier, A., Regal, X., Reid, B., Reil, K., Rezaie, M., Rockosi, C. M., Roe, N., Ronayette, S., Roodman, A., Ross, A. J., Ross, N. P., Rossi, G., Rozo, E., Ruhlmann-Kleider, V., Rykoff, E. S., Sabiu, C., Samushia, L., Sanchez, E., Sanchez, J., Schlegel, D. J., Schneider, M., Schubnell, M., Secroun, A., Seljak, U., Seo, H.-J., Serrano, S., Shafieloo, A., Shan, H., Sharples, R., Sholl, M. J., Shourt, W. V., Silber, J. H., Silva, D. R., Sirk, M. M., Slosar, A., Smith, A., Smoot, G. F., Som, D., Song, Y.-S., Sprayberry, D., Staten, R., Stefanik, A., Tarle, G., Sien Tie, S., Tinker, J. L., Tojeiro, R., Valdes, F., Valenzuela, O., Valluri, M., Vargas-Magana, M., Verde, L., Walker, A. R., Wang, J., Wang, Y., Weaver, B. A., Weaverdyck, C., Wechsler, R. H., Weinberg, D. H., White, M., Yang, Q., Yeche, C., Zhang, T., Zhao, G.-B., Zheng, Y., Zhou, X., Zhou, Z., Zhu, Y., Zou, H., and Zu, Y., The DESI Experiment Part I: Science,Targeting, and Survey Design, Oct. 2016, arXiv e-prints, arXiv:1611.00036

Di Matteo, T., Colberg, J., Springel, V., Hernquist, L., and Sijacki, D., Direct Cosmological Simulations of the Growth of Black Holes and Galaxies, Mar. 2008, ApJ, 676, 33

Di Matteo, T., Khandai, N., DeGraf, C., Feng, Y., Croft, R. A. C., Lopez, J., and Springel, V., Cold Flows and the First Quasars, Feb. 2012, ApJ, 745, L29

Di Matteo, T., Springel, V., and Hernquist, L., Energy input from quasars regulates the growth and activity of black holes and their host galaxies, Feb. 2005, Nature, 433, 604

Doré, O., Bock, J., Ashby, M., Capak, P., Cooray, A., de Putter, R., Eifler, T., Flagey, N., Gong, Y., Habib, S., Heitmann, K., Hirata, C., Jeong, W.-S., Katti, R., Korngut, P., Krause, E., Lee, D.-H., Masters, D., Mauskopf, P., Melnick, G., Mennesson, B., Nguyen, H., Öberg, K., Pullen, A., Raccanelli, A., Smith, R., Song, Y.-S., Tolls, V., Unwin, S., Venumadhav, T., Viero, M., Werner, M., and Zemcov, M., Cosmology with the SPHEREX All-Sky Spectral Survey, Dec. 2014, arXiv e-prints, arXiv:1412.4872

Doré, O., Werner, M. W., Ashby, M. L. N., Bleem, L. E., Bock, J., Burt, J., Capak, P., Chang, T.-C., Chaves-Montero, J., Chen, C. H., Civano, F., Cleeves, I. I., Cooray, A., Crill, B., Crossfield, I. J. M., Cushing, M., de la Torre, S., DiMatteo, T., Dvory, N., Dvorkin, C., Espaillat, C., Ferraro, S., Finkbeiner, D., Greene, J., Hewitt, J., Hogg, D. W., Huffenberger, K., Jun, H.-S., Ilbert, O., Jeong, W.-S., Johnson, J., Kim, M., Kirkpatrick, J. D., Kowalski, T., Korngut, P., Li, J., Lisse, C. M., MacGregor, M., Mamajek, E. E., Mauskopf, P., Melnick, G., Ménard, B., Neyrinck, M., Öberg, K., Pisani, A., Rocca, J., Salvato, M., Schaan, E., Scoville, N. Z., Song, Y.-S., Stevens, D. J., Tenneti, A., Teplitz, H., Tolls, V., Unwin, S., Urry, M., Wandelt, B., Williams, B. F., Wilner, D., Windhorst, R. A., Wolk, S., Yorke, H. W., and Zemcov, M., Science Impacts of the SPHEREx All-Sky Optical to Near-Infrared Spectral Survey II: Report of a Community Workshop on the Scientific Synergies Between the SPHEREx Survey and Other Astronomy Observatories, May. 2018, arXiv e-prints, arXiv:1805.05489

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N., An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, Oct. 2020, arXiv e-prints, arXiv:2010.11929

du Buisson, L., Sivanandam, N., Bassett, B. A., and Smith, M., Machine learning classification of SDSS transient survey images, Dec. 2015, MNRAS, 454, 2026

Dumitru, S., Kulkarni, G., Lagache, G., and Haehnelt, M. G., Predictions and sensitivity forecasts for reionization-era [C II] line intensity mapping, May. 2019, MNRAS, 485, 3486

Eide, M. B., Graziani, L., Ciardi, B., Feng, Y., Kakiichi, K., and Di Matteo, T., The epoch of cosmic heating by early sources of X-rays, May. 2018, MNRAS, 476, 1174

Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., Scranton, R., Seo, H.-J., Tegmark, M., Zheng, Z., Anderson, S. F., Annis, J., Bahcall, N., Brinkmann, J., Burles, S., Castander, F. J., Connolly, A., Csabai, I., Doi, M., Fukugita, M., Frieman, J. A., Glazebrook, K., Gunn, J. E., Hendry, J. S., Hennessy, G., Ivezić, Z., Kent, S., Knapp, G. R., Lin, H., Loh, Y.-S., Lupton, R. H., Margon, B., McKay, T. A., Meiksin, A., Munn, J. A., Pope, A., Richmond, M. W., Schlegel, D., Schneider, D. P., Shimasaku, K., Stoughton, C., Strauss, M. A., SubbaRao, M., Szalay, A. S., Szapudi, I., Tucker, D. L., Yanny, B., and York, D. G., Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies, Nov. 2005, ApJ, 633, 560

Eldridge, J. J. and Stanway, E. R., The effect of stellar evolution uncertainties on the rest-frame ultraviolet stellar lines of C IV and He II in high-redshift Lyman-break galaxies, Jan. 2012, MNRAS, 419, 479

Ellis, R. S., Colless, M., Broadhurst, T., Heyl, J., and Glazebrook, K., Autofib Redshift Survey - I. Evolution of the galaxy luminosity function, May. 1996, MNRAS, 280, 235

Erhan, D., Bengio, Y., Courville, A., and Vincent, P., Visualizing Higher-Layer Features of a Deep Network, 01. 2009, Technical Report, Univeristé de Montréal

Fan, X., Carilli, C. L., and Keating, B., Observational Constraints on Cosmic Reionization, Sep. 2006, ARA&A, 44, 415

Ferland, G. J., Chatzikos, M., Guzmán, F., Lykins, M. L., van Hoof, P. A. M., Williams, R. J. R., Abel, N. P., Badnell, N. R., Keenan, F. P., Porter, R. L., and Stancil, P. C., The 2017 Release Cloudy, Oct. 2017, Rev. Mexicana Astron. Astrofis., 53, 385

Field, G. B., Excitation of the Hydrogen 21-CM Line, Jan. 1958, Proceedings of the IRE, 46, 240

Field, G. B., The Spin Temperature of Intergalactic Neutral Hydrogen., May. 1959, ApJ, 129, 536

Fluke, C. J. and Jacobs, C., Surveying the reach and maturity of machine learning and artificial intelligence in astronomy, Jan. 2020, WIREs Data Mining and Knowledge Discovery, 10, e1349

Fonseca, J., Camera, S., Santos, M. G., and Maartens, R., Hunting Down Horizon-scale Effects with Multi-wavelength Surveys, Oct. 2015, ApJ, 812, L22

Fonseca, J., Silva, M. B., Santos, M. G., and Cooray, A., Cosmology with intensity mapping techniques using atomic and molecular lines, Jan. 2017, MNRAS, 464, 1948

Fragos, T., Lehmer, B. D., Naoz, S., Zezas, A., and Basu-Zych, A., Energy Feedback from X-Ray Binaries in the Early Universe, Oct. 2013a, ApJ, 776, L31

Fragos, T., Lehmer, B., Tremmel, M., Tzanavaris, P., Basu-Zych, A., Belczynski, K., Hornschemeier, A., Jenkins, L., Kalogera, V., Ptak, A., and Zezas, A., X-Ray Binary Evolution Across Cosmic Time, Feb. 2013b, ApJ, 764, 41

Furlanetto, S. R., Oh, S. P., and Briggs, F. H., Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe, Oct. 2006, Phys. Rep., 433, 181

Gong, Y., Silva, M., Cooray, A., and Santos, M. G., Foreground Contamination in Lyα Intensity Mapping during the Epoch of Reionization, Apr. 2014, ApJ, 785, 72

Goodfellow, I., Bengio, Y., and Courville, A. 2016, Deep Learning (MIT Press), http://www.deeplearningbook.org

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., Generative Adversarial Networks, Jun. 2014, arXiv e-prints, arXiv:1406.2661

Graziani, L., Maselli, A., and Ciardi, B., CRASH3: cosmological radiative transfer through metals, May. 2013, MNRAS, 431, 722

Gregory, S. A. and Thompson, L. A., The Coma/A1367 supercluster and its environs., Jun. 1978, ApJ, 222, 784

Guo, Q., White, S., Angulo, R. E., Henriques, B., Lemson, G., Boylan-Kolchin, M., Thomas, P., and Short, C., Galaxy formation in WMAP1 and WMAP7 cosmologies, Jan. 2013, MNRAS, 428, 1351

Gupta, A., Zorrilla Matilla, J. M., Hsu, D., and Haiman, Z., Non-Gaussian information from weak lensing data via deep learning, May. 2018, Phys. Rev. D, 97, 103515

Hassan, S., Liu, A., Kohn, S., and La Plante, P., Identifying reionization sources from 21 cm maps using Convolutional Neural Networks, Feb. 2019, MNRAS, 483, 2524

Heneka, C., Cooray, A., and Feng, C., Probing the Intergalactic Medium with Lyα and 21 cm Fluctuations, Oct. 2017, ApJ, 848, 52

Heneka, C. and Mesinger, A., The spin-temperature dependence of the 21-cm-LAE cross-correlation, Jul. 2020, MNRAS, 496, 581

Hill, G. J., Gebhardt, K., Komatsu, E., Drory, N., MacQueen, P. J., Adams, J., Blanc, G. A., Koehler, R., Rafal, M., Roth, M. M., Kelz, A., Gronwall, C., Ciardullo, R., and Schneider, D. P. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 399, Panoramic Views of Galaxy Formation and Evolution, ed. T. Kodama, T. Yamada, & K. Aoki, 115

Hong, S. E., Jeong, D., Hwang, H. S., and Kim, J., Revealing the Local Cosmic Web from Galaxies by Deep Learning, May. 2021, ApJ, 913, 76

Inoue, A. K., Rest-frame ultraviolet-to-optical spectral characteristics of extremely metal-poor and metal-free galaxies, Aug. 2011, MNRAS, 415, 2920

Ioffe, S. and Szegedy, C., Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, Feb. 2015, arXiv e-prints, arXiv:1502.03167

Isola, P., Zhu, J., Zhou, T., and Efros, A. A., Image-to-Image Translation with Conditional Adversarial Networks. 2016, CoRR, abs/1611.07004, arXiv:1611.07004

Kaiser, N., Clustering in real space and in redshift space, Jul. 1987, MNRAS, 227, 1

Kakuma, R., Ouchi, M., Harikane, Y., Ono, Y., Inoue, A. K., Komiyama, Y., Kusakabe, H., Lee, C.-H., Matsuda, Y., Matsuoka, Y., Mawatari, K., Momose, R., Shibuya, T., and Taniguchi, Y., SILVERRUSH. IX. Lyα Intensity Mapping with Star-forming Galaxies at z = 5.7 and 6.6: A Possible Detection of Extended Lyα Emission at $\gtrsim$100 Comoving Kiloparsecs around and beyond the Virial-radius Scale of Galaxy Dark Matter Halos, Jul. 2021, ApJ, 916, 22

Kannan, R., Smith, A., Garaldi, E., Shen, X., Vogelsberger, M., Pakmor, R., Springel, V., and Hernquist, L., The THESAN project: predictions for multi-tracer line intensity mapping in the Epoch of Reionization, Nov. 2021, arXiv e-prints, arXiv:2111.02411

Karkare, K. S., Anderson, A. J., Barry, P. S., Benson, B. A., Carlstrom, J. E., Cecil, T., Chang, C. L., Dobbs, M. A., Hollister, M., Keating, G. K., Marrone, D. P., McMahon, J., Montgomery, J., Pan, Z., Robson, G., Rouble, M., Shirokoff, E., and Smecher, G., SPT-SLIM: A Line Intensity Mapping Pathfinder for the South Pole Telescope, Nov. 2021, arXiv e-prints, arXiv:2111.04631

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L., Physics-informed machine learning. 2021, Nature Reviews Physics, 3, 422

Keating, G. K., Marrone, D. P., Bower, G. C., and Keenan, R. P., An Intensity Mapping Detection of Aggregate CO Line Emission at 3 mm, Oct. 2020, ApJ, 901, 141

Keating, G. K., Marrone, D. P., Bower, G. C., Leitch, E., Carlstrom, J. E., and DeBoer, D. R., COPSS II: The Molecular Gas Content of Ten Million Cubic Megaparsecs at Redshift z $\sim$ 3, Oct. 2016, ApJ, 830, 34

Keenan, R. P., Keating, G. K., and Marrone, D. P., An Intensity Mapping Constraint on the CO-Galaxy Cross Power Spectrum at Redshift ˜3, Oct. 2021, arXiv e-prints, arXiv:2110.02239

Kendall, A. and Gal, Y., What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? 2017, arXiv e-prints, arXiv:1703.04977

Kennicutt, Jr., R. C., The Global Schmidt Law in Star-forming Galaxies, May. 1998, ApJ, 498, 541

Khandai, N., Di Matteo, T., Croft, R., Wilkins, S., Feng, Y., Tucker, E., DeGraf, C., and Liu, M.-S., The MassiveBlack-II simulation: the evolution of haloes and galaxies to z $\sim$ 0, Jun. 2015, MNRAS, 450, 1349

Khostovan, A. A., Sobral, D., Mobasher, B., Best, P. N., Smail, I., Stott, J. P., Hemmati, S., and Nayyeri, H., Evolution of the H β + [O III] and [O II] luminosity functions and the [O II] star formation history of the Universe up to z$\sim$ 5 from HiZELS, Oct. 2015, MNRAS, 452, 3948

Kingma, D. P. and Ba, J., Adam: A Method for Stochastic Optimization. 2014, arXiv e-prints, arXiv:1412.6980

Kingma, D. P. and Welling, M., Auto-Encoding Variational Bayes, Dec. 2013, arXiv e-prints, arXiv:1312.6114

Kohno, K., Kawabe, R., Tamura, Y., Endo, A., Baselmans, J. J. A., Karatsu, K., Inoue, A. K., Moriwaki, K., Hayatsu, N. H., Yoshida, N., Yoshimura, Y., Hatsukade, B., Umehata, H., Oshima, T., Takekoshi, T., Taniguchi, A., Klaassen, P. D., Mroczkowski, T., Cicone, C., Bertoldi, F., Dannerbauer, H., and Tosaki, T. 2020, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 11453, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 114530N

Kojima, T., Ouchi, M., Rauch, M., Ono, Y., Nakajima, K., Isobe, Y., Fujimoto, S., Harikane, Y., Hashimoto, T., Hayashi, M., Komiyama, Y., Kusakabe, H., Kim, J. H., Lee, C.-H., Mukae, S., Nagao, T., Onodera, M., Shibuya, T., Sugahara, Y., Umemura, M., and Yabe, K., Extremely Metal-poor Representatives Explored by the Subaru Survey (EMPRESS). I. A Successful Machine-learning Selection of Metal-poor Galaxies and the Discovery of a Galaxy with M* ¡ 10^6 $M_\odot$ and 0.016 $Z_\odot$, Aug. 2020, ApJ, 898, 142

Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolta, M. R., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L., Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation, Feb. 2011, ApJS, 192, 18

Konno, A., Ouchi, M., Ono, Y., Shimasaku, K., Shibuya, T., Furusawa, H., Nakajima, K., Naito, Y., Momose, R., Yuma, S., and Iye, M., Accelerated Evolution of the Lyα Luminosity Function at z ¿˜7 Revealed by the Subaru Ultra-deep Survey for Lyα Emitters at z = 7.3, Dec. 2014, ApJ, 797, 16

Konno, A., Ouchi, M., Shibuya, T., Ono, Y., Shimasaku, K., Taniguchi, Y., Nagao, T., Kobayashi, M. A. R., Kajisawa, M., Kashikawa, N., Inoue, A. K., Oguri, M., Furusawa, H., Goto, T., Harikane, Y., Higuchi, R., Komiyama, Y., Kusakabe, H., Miyazaki, S., Nakajima, K., and Wang, S.-Y., SILVERRUSH. IV. Lyα luminosity functions at z = 5.7 and 6.6 studied with $\sim$1300 Lyα emitters on the 14-21 deg^2 sky, Jan. 2018, PASJ, 70, S16

Krawczyk, C. M., Richards, G. T., Mehta, S. S., Vogeley, M. S., Gallagher, S. C., Leighly, K. M., Ross, N. P., and Schneider, D. P., Mean Spectral Energy Distributions and Bolometric Corrections for Luminous Quasars, May. 2013, ApJS, 206, 4

Kriek, M., Shapley, A. E., Reddy, N. A., Siana, B., Coil, A. L., Mobasher, B., Freeman, W. R., de Groot, L., Price, S. H., Sanders, R., Shivaei, I., Brammer, G. B., Momcheva, I. G., Skelton, R. E., van Dokkum, P. G., Whitaker, K. E., Aird, J., Azadi, M., Kassis, M., Bullock, J. S., Conroy, C., Davé, R., Kereš, D., and Krumholz, M., The MOSFIRE Deep

Evolution Field (MOSDEF) Survey: Rest-frame Optical Spectroscopy for ~1500 H-selected Galaxies at 1.37 ¡ z ¡ 3.8, Jun. 2015, ApJS, 218, 15

Kubota, K., Yoshiura, S., Takahashi, K., Hasegawa, K., Yajima, H., Ouchi, M., Pindor, B., and Webster, R. L., Detectability of the 21-cm signal during the epoch of reionization with 21-cm Lyman α emitter cross-correlation - I, Sep. 2018, MNRAS, 479, 2754

Laureijs, R., Amiaux, J., Arduini, S., Augueres, J. L., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., Garilli, B., Gondoin, P., Guzzo, L., Hoar, J., Hoekstra, H., Holmes, R., Kitching, T., Maciaszek, T., Mellier, Y., Pasian, F., Percival, W., Rhodes, J., Saavedra Criado, G., Sauvage, M., Scaramella, R., Valenziano, L., Warren, S., Bender, R., Castander, F., Cimatti, A., Le Fèvre, O., Kurki-Suonio, H., Levi, M., Lilje, P., Meylan, G., Nichol, R., Pedersen, K., Popa, V., Rebolo Lopez, R., Rix, H. W., Rottgering, H., Zeilinger, W., Grupp, F., Hudelot, P., Massey, R., Meneghetti, M., Miller, L., Paltani, S., Paulin-Henriksson, S., Pires, S., Saxton, C., Schrabback, T., Seidel, G., Walsh, J., Aghanim, N., Amendola, L., Bartlett, J., Baccigalupi, C., Beaulieu, J. P., Benabed, K., Cuby, J. G., Elbaz, D., Fosalba, P., Gavazzi, G., Helmi, A., Hook, I., Irwin, M., Kneib, J. P., Kunz, M., Mannucci, F., Moscardini, L., Tao, C., Teyssier, R., Weller, J., Zamorani, G., Zapatero Osorio, M. R., Boulade, O., Foumond, J. J., Di Giorgio, A., Guttridge, P., James, A., Kemp, M., Martignac, J., Spencer, A., Walton, D., Blümchen, T., Bonoli, C., Bortoletto, F., Cerna, C., Corcione, L., Fabron, C., Jahnke, K., Ligori, S., Madrid, F., Martin, L., Morgante, G., Pamplona, T., Prieto, E., Riva, M., Toledo, R., Trifoglio, M., Zerbi, F., Abdalla, F., Douspis, M., Grenet, C., Borgani, S., Bouwens, R., Courbin, F., Delouis, J. M., Dubath, P., Fontana, A., Frailis, M., Grazian, A., Koppenhöfer, J., Mansutti, O., Melchior, M., Mignoli, M., Mohr, J., Neissner, C., Noddle, K., Poncet, M., Scodeggio, M., Serrano, S., Shane, N., Starck, J. L., Surace, C., Taylor, A., Verdoes-Kleijn, G., Vuerli, C., Williams, O. R., Zacchei, A., Altieri, B., Escudero Sanz, I., Kohley, R., Oosterbroek, T., Astier, P., Bacon, D., Bardelli, S., Baugh, C., Bellagamba, F., Benoist, C., Bianchi, D., Biviano, A., Branchini, E., Carbone, C., Cardone, V., Clements, D., Colombi, S., Conselice, C., Cresci, G., Deacon, N., Dunlop, J., Fedeli, C., Fontanot, F., Franzetti, P., Giocoli, C., Garcia-Bellido, J., Gow, J., Heavens, A., Hewett, P., Heymans, C., Holland, A., Huang, Z., Ilbert, O., Joachimi, B., Jennins, E., Kerins, E., Kiessling, A., Kirk, D., Kotak, R., Krause, O., Lahav, O., van Leeuwen, F., Lesgourgues, J., Lombardi, M., Magliocchetti, M., Maguire, K., Majerotto, E., Maoli, R., Marulli, F., Maurogordato, S., McCracken, H., McLure, R., Melchiorri, A., Merson, A., Moresco, M., Nonino, M., Norberg, P., Peacock, J., Pello, R., Penny, M., Pettorino, V., Di Porto, C., Pozzetti, L., Quercellini, C., Radovich, M., Rassat, A., Roche, N., Ronayette, S., Rossetti, E., Sartoris, B., Schneider, P., Semboloni, E., Serjeant, S., Simpson, F., Skordis, C., Smadja, G., Smartt, S., Spano, P., Spiro, S., Sullivan, M., Tilquin, A., Trotta, R., Verde, L., Wang, Y., Williger, G., Zhao, G., Zoubian, J., and Zucca, E., Euclid Definition Study Report, Oct. 2011, arXiv e-prints, arXiv:1110.3193

Lee, D., Yang, M.-H., and Oh, S., Generative Single Image Reflection Separation. 2018, arXiv e-prints, arXiv:1801.04102

Li, T. Y., Wechsler, R. H., Devaraj, K., and Church, S. E., Connecting CO Intensity Mapping to Molecular Gas and Star Formation in the Epoch of Galaxy Assembly, Feb. 2016, ApJ, 817, 169

Li, W., Xu, H., Ma, Z., Zhu, R., Hu, D., Zhu, Z., Gu, J., Shan, C., Zhu, J., and Wu, X.-P., Separating the EoR signal with a convolutional denoising autoencoder: a deep-learning-based method, 02. 2019, Monthly Notices of the Royal Astronomical Society, 485, 2628

Lidz, A. and Taylor, J., On Removing Interloper Contamination from Intensity Mapping Power Spectrum Measurements, Jul. 2016, ApJ, 825, 143

Lidz, A., Zahn, O., Furlanetto, S. R., McQuinn, M., Hernquist, L., and Zaldarriaga, M., Probing Reionization with the 21 cm Galaxy Cross-Power Spectrum, Jan. 2009, ApJ, 690, 252

Lilly, S. J., Hammer, F., Le Fevre, O., and Crampton, D., The Canada-France Redshift Survey. III. "Single Emission-Line" Objects, Analysis of Repeat Observations, and Spectroscopic Identifications in the 1415+52 and 2215+00 Fields, Dec. 1995, ApJ, 455, 75

Liu, A., Parsons, A. R., and Trott, C. M., Epoch of reionization window. II. Statistical methods for foreground wedge reduction, Jul. 2014, Phys. Rev. D, 90, 023019

Madau, P. and Fragos, T., Radiation Backgrounds at Cosmic Dawn: X-Rays from Compact Binaries, May. 2017, ApJ, 840, 39

Marinacci, F., Vogelsberger, M., Pakmor, R., Torrey, P., Springel, V., Hernquist, L., Nelson, D., Weinberger, R., Pillepich, A., Naiman, J., and Genel, S., First results from the IllustrisTNG simulations: radio haloes and magnetic fields, Nov. 2018, MNRAS, 480, 5113

Maselli, A., Ciardi, B., and Kanekar, A., CRASH2: coloured packets and other updates, Feb. 2009, MNRAS, 393, 171

Maselli, A., Ferrara, A., and Ciardi, B., CRASH: a radiative transfer scheme, Oct. 2003, MNRAS, 345, 379

Matilla, J. M. Z., Sharma, M., Hsu, D., and Haiman, Z., Interpreting deep learning models for weak lensing, Dec. 2020, Phys. Rev. D, 102, 123506

McGreer, I. D., Mesinger, A., and D'Odorico, V., Model-independent evidence in favour of an end to reionization by z $\approx$ 6, Feb. 2015, MNRAS, 447, 499

McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., and Furlanetto, S. R., Cosmological Parameter Estimation Using 21 cm Radiation from the Epoch of Reionization, Dec. 2006, ApJ, 653, 815

Mertens, F. G., Ghosh, A., and Koopmans, L. V. E., Statistical 21-cm signal separation via Gaussian Process Regression analysis, Aug. 2018, MNRAS, 478, 3640

Mineo, S., Gilfanov, M., and Sunyaev, R., X-ray emission from star-forming galaxies - II. Hot interstellarmedium, Nov. 2012, MNRAS, 426, 1870

Monaco, P., Sefusatti, E., Borgani, S., Crocce, M., Fosalba, P., Sheth, R. K., and Theuns, T., An accurate tool for the fast generation of dark matter halo catalogues, Aug. 2013, MNRAS, 433, 2389

Mordvintsev, A., Olah, C., and Tyka, M. 2015, Inceptionism: going deeper into neural networks., Google AI Blog

Moriwaki, K., Filippova, N., Shirasaki, M., and Yoshida, N., Deep learning for intensity mapping observations: Component extraction, May. 2020, MNRAS, arXiv:2002.07991

Moriwaki, K., Shirasaki, M., and Yoshida, N., Deep Learning for Line Intensity Mapping Observations: Information Extraction from Noisy Maps, Jan. 2021, ApJ, 906, L1

Moriwaki, K. and Yoshida, N., Deep-learning Reconstruction of Three-dimensional Galaxy Distributions with Intensity Mapping Observations, Dec. 2021, ApJ, 923, L7

Moriwaki, K., Yoshida, N., Eide, M. B., and Ciardi, B., Cross-correlation between the 21-cm signal and [O III] emitters during early cosmic reionization, Oct. 2019, MNRAS, 489, 2471

Moriwaki, K., Yoshida, N., Shimizu, I., Harikane, Y., Matsuda, Y., Matsuo, H., Hashimoto, T., Inoue, A. K., Tamura, Y., and Nagao, T., The distribution and physical properties of high-redshift [O III] emitters in a cosmological hydrodynamics simulation, Nov. 2018, MNRAS, 481, L84

Moutarde, F., Alimi, J. M., Bouchet, F. R., Pellat, R., and Ramani, A., Precollapse Scale Invariance in Gravitational Instability, Dec. 1991, ApJ, 382, 377

Naiman, J. P., Pillepich, A., Springel, V., Ramirez-Ruiz, E., Torrey, P., Vogelsberger, M., Pakmor, R., Nelson, D., Marinacci, F., Hernquist, L., Weinberger, R., and Genel, S., First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium, Jun. 2018, MNRAS, 477, 1206

Nakajima, K., Ouchi, M., Shimasaku, K., Hashimoto, T., Ono, Y., and Lee, J. C., First Spectroscopic Evidence for High Ionization State and Low Oxygen Abundance in Lyα Emitters, May. 2013, ApJ, 769, 3

Nelson, D., Pillepich, A., Springel, V., Weinberger, R., Hernquist, L., Pakmor, R., Genel, S., Torrey, P., Vogelsberger, M., Kauffmann, G., Marinacci, F., and Naiman, J., First results from the IllustrisTNG simulations: the galaxy colour bimodality, Mar. 2018, MNRAS, 475, 624

Nelson, D., Springel, V., Pillepich, A., Rodriguez-Gomez, V., Torrey, P., Genel, S., Vogelsberger, M., Pakmor, R., Marinacci, F., Weinberger, R., Kelley, L., Lovell, M., Diemer, B., and Hernquist, L., The IllustrisTNG simulations: public data release, May. 2019, Computational Astrophysics and Cosmology, 6, 2

Newburgh, L. B., Bandura, K., Bucher, M. A., Chang, T. C., Chiang, H. C., Cliche, J. F., Davé, R., Dobbs, M., Clarkson, C., Ganga, K. M., Gogo, T., Gumba, A., Gupta, N., Hilton, M., Johnstone, B., Karastergiou, A., Kunz, M., Lokhorst, D., Maartens, R., Macpherson, S., Mdlalose, M., Moodley, K., Ngwenya, L., Parra, J. M., Peterson, J., Recnik, O., Saliwanchik, B., Santos, M. G., Sievers, J. L., Smirnov, O., Stronkhorst, P., Taylor, R., Vanderlinde, K., Van Vuuren, G., Weltman, A., and Witzemann, A. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9906, Ground-based and Airborne Telescopes VI, ed. H. J. Hall, R. Gilmozzi, & H. K. Marshall, 99065X

Ntampaka, M., Trac, H., Sutherland, D. J., Battaglia, N., Póczos, B., and Schneider, J., A Machine Learning Approach for Dynamical Mass Measurements of Galaxy Clusters, Apr. 2015, ApJ, 803, 50

Ntampaka, M., Avestruz, C., Boada, S., Caldeira, J., Cisewski-Kehe, J., Di Stefano, R., Dvorkin, C., Evrard, A. E., Farahi, A., Finkbeiner, D., Genel, S., Goodman, A., Goulding, A., Ho, S., Kosowsky, A., La Plante, P., Lanusse, F., Lochner, M., Mandelbaum, R., Nagai, D., Newman, J. A., Nord, B., Peek, J. E. G., Peel, A., Poczos, B., Rau, M. M., Siemiginowska, A., Sutherland, D. J., Trac, H., and Wandelt, B., The Role of Machine Learning in the Next Decade of Cosmology, May. 2019, BAAS, 51, 14

Osterbrock, D. E. and Ferland, G. J. 2006, Astrophysics of gaseous nebulae and active galactic nuclei ()

Ouchi, M., Hamana, T., Shimasaku, K., Yamada, T., Akiyama, M., Kashikawa, N., Yoshida, M., Aoki, K., Iye, M., Saito, T., Sasaki, T., Simpson, C., and Yoshida, M., Definitive Identification of the Transition between Small- and Large-Scale Clustering for Lyman Break Galaxies, Dec. 2005, ApJ, 635, L117

Ouchi, M., Mobasher, B., Shimasaku, K., Ferguson, H. C., Fall, S. M., Ono, Y., Kashikawa, N., Morokuma, T., Nakajima, K., Okamura, S., Dickinson, M., Giavalisco, M., and Ohta, K., Large Area Survey for z = 7 Galaxies in SDF and GOODS-N: Implications for Galaxy Formation and Cosmic Reionization, Dec. 2009, ApJ, 706, 1136

Ouchi, M., Harikane, Y., Shibuya, T., Shimasaku, K., Taniguchi, Y., Konno, A., Kobayashi, M., Kajisawa, M., Nagao, T., Ono, Y., Inoue, A. K., Umemura, M., Mori, M., Hasegawa, K., Higuchi, R., Komiyama, Y., Matsuda, Y., Nakajima, K., Saito, T., and Wang, S.-Y., Systematic Identification of LAEs for Visible Exploration and Reionization Research Using Subaru HSC (SILVERRUSH). I. Program strategy and clustering properties of $\sim$2000 Lyα emitters at z = 6-7 over the 0.3-0.5 Gpc^2 survey area, Jan. 2018, PASJ, 70, S13

Paciga, G., Chang, T.-C., Gupta, Y., Nityanada, R., Odegova, J., Pen, U.-L., Peterson, J. B., Roy, J., and Sigurdson, K., The GMRT Epoch of Reionization experiment: a new upper limit on the neutral hydrogen power spectrum at z$\approx$ 8.6, May. 2011, MNRAS, 413, 1174

Pacucci, F., Mesinger, A., Mineo, S., and Ferrara, A., The X-ray spectra of the first galaxies: 21 cm signatures, Sep. 2014, MNRAS, 443, 678

Papernot, N. and McDaniel, P., Deep k-Nearest Neighbors: Towards Confident, Interpretable and Robust Deep Learning, Mar. 2018, arXiv e-prints, arXiv:1803.04765

Park, C., Choi, Y.-Y., Vogeley, M. S., Gott, J. Richard, I., Kim, J., Hikage, C., Matsubara, T., Park, M.-G., Suto, Y., Weinberg, D. H., and SDSS Collaboration, Topology Analysis of the Sloan Digital Sky Survey. I. Scale and Luminosity Dependence, Nov. 2005, ApJ, 633, 11

Park, H., Shapiro, P. R., Choi, J.-h., Yoshida, N., Hirano, S., and Ahn, K., The Hydrodynamic Feedback of Cosmic Reionization on Small-scale Structures and Its Impact on Photon Consumption During the Epoch of Reionization, Nov. 2016, ApJ, 831, 86

Parsons, A. R., Backer, D. C., Foster, G. S., Wright, M. C. H., Bradley, R. F., Gugliucci, N. E., Parashare, C. R., Benoit, E. E., Aguirre, J. E., Jacobs, D. C., Carilli, C. L., Herne, D., Lynch, M. J., Manley, J. R., and Werthimer, D. J., THE PRECISION ARRAY FOR PROBING THE EPOCH OF RE-IONIZATION: EIGHT STATION RESULTS, mar. 2010, The Astronomical Journal, 139, 1468

Pearson, W. J., Wang, L., Trayford, J. W., Petrillo, C. E., and van der Tak, F. F. S., Identifying galaxy mergers in observations and simulations with deep learning, Jun. 2019, A&A, 626, A49

Petroff, M. A., Addison, G. E., Bennett, C. L., and Weiland, J. L., Full-sky Cosmic Microwave Background Foreground Cleaning Using Machine Learning, Nov. 2020, ApJ, 903, 104

Pillepich, A., Springel, V., Nelson, D., Genel, S., Naiman, J., Pakmor, R., Hernquist, L., Torrey, P., Vogelsberger, M., Weinberger, R., and Marinacci, F., Simulating galaxy formation with the IllustrisTNG model, Jan. 2018, MNRAS, 473, 4077

Planck Collaboration IX, Planck 2018 results. IX. Constraints on primordial non-Gaussianity, Sep. 2020, A&A, 641, A9

Planck Collaboration VI, Planck 2018 results. VI. Cosmological parameters, Jul. 2018, arXiv e-prints, arXiv:1807.06209

Planck Collaboration XIII, Planck 2015 results. XIII. Cosmological parameters, Sep. 2016, A&A, 594, A13

Planck Collaboration XIV, Planck 2015 results. XIV. Dark energy and modified gravity, Sep. 2016, A&A, 594, A14

Prakash, A., Licquia, T. C., Newman, J. A., Ross, A. J., Myers, A. D., Dawson, K. S., Kneib, J.-P., Percival, W. J., Bautista, J. E., Comparat, J., Tinker, J. L., Schlegel, D. J., Tojeiro, R., Ho, S., Lang, D., Rao, S. M., McBride, C. K., Ben Zhu, G., Brownstein, J. R., Bailey, S., Bolton, A. S., Delubac, T., Mariappan, V., Blanton, M. R., Reid, B., Schneider, D. P., Seo, H.-J., Carnero Rosell, A., and Prada, F., The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Luminous Red Galaxy Target Selection, Jun. 2016, ApJS, 224, 34

Prelogović, D., Mesinger, A., Murray, S., Fiameni, G., and Gillet, N., Machine learning galaxy properties from 21 cm lightcones: impact of network architectures and signal contamination, Jun. 2021, arXiv e-prints, arXiv:2107.00018

Price, D. C., Greenhill, L. J., Fialkov, A., Bernardi, G., Garsden, H., Barsdell, B. R., Kocz, J., Anderson, M. M., Bourke, S. A., Craig, J., Dexter, M. R., Dowell, J., Eastwood, M. W., Eftekhari, T., Ellingson, S. W., Hallinan, G., Hartman, J. M., Kimberk, R., Lazio, T. J. W., Leiker, S., MacMahon, D., Monroe, R., Schinzel, F., Taylor, G. B., Tong, E., Werthimer, D., and Woody, D. P., Design and characterization of the Large-aperture Experiment to Detect the Dark Age (LEDA) radiometer systems, Aug. 2018, MNRAS, 478, 4193

Pullen, A. R., Serra, P., Chang, T.-C., Doré, O., and Ho, S., Search for C II emission on cosmological scales at redshift Z $\sim$ 2.6, Aug. 2018, MNRAS, 478, 1911

Raichoor, A., Comparat, J., Delubac, T., Kneib, J. P., Yèche, C., Dawson, K. S., Percival, W. J., Dey, A., Lang, D., Schlegel, D. J., Gorgoni, C., Bautista, J., Brownstein, J. R., Mariappan, V., Seo, H. J., Tinker, J. L., Ross, A. J., Wang, Y., Zhao, G. B., Moustakas, J., Palanque-Delabrouille, N., Jullo, E., Newmann, J. A., Prada, F., and Zhu, G. B., The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: final emission line galaxy target selection, Nov. 2017, MNRAS, 471, 3955

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

Raveri, M., Bull, P., Silvestri, A., and Pogosian, L., Priors on the effective dark energy equation of state in scalar-tensor theories, Oct. 2017, Phys. Rev. D, 96, 083509

Rutkowski, M. J., Scarlata, C., Haardt, F., Siana, B., Henry, A., Rafelski, M., Hayes, M., Salvato, M., Pahl, A. J., Mehta, V., Beck, M., Malkan, M., and Teplitz, H. I., Lyman Continuum Escape Fraction of Star-forming Dwarf Galaxies at z $\sim$ 1, Mar. 2016, ApJ, 819, 81

Salpeter, E. E., The Luminosity Function and Stellar Evolution., Jan. 1955, ApJ, 121, 161

Schaye, J., Crain, R. A., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., Jenkins, A., Rosas-Guevara, Y. M., White, S. D. M., Baes, M., Booth, C. M., Camps, P., Navarro, J. F., Qu, Y., Rahmati, A., Sawala, T., Thomas, P. A., and Trayford, J., The EAGLE project: simulating the evolution and assembly of galaxies and their environments, Jan. 2015, MNRAS, 446, 521

Seljak, U., Extracting Primordial Non-Gaussianity without Cosmic Variance, Jan. 2009, Phys. Rev. Lett., 102, 021302

Shimabukuro, H. and Semelin, B., Analysing the 21 cm signal from the epoch of reionization with artificial neural networks, Jul. 2017, MNRAS, 468, 3869

Shirasaki, M., Moriwaki, K., Oogi, T., Yoshida, N., Ikeda, S., and Nishimichi, T., Noise reduction for weak lensing mass mapping: an application of generative adversarial networks to Subaru Hyper Suprime-Cam first-year data, Jun. 2021, MNRAS, 504, 1825

Shirasaki, M., Yoshida, N., and Ikeda, S., Denoising weak lensing mass maps with deep learning, Aug. 2019, Phys. Rev. D, 100, 043527

Silva, B. M., Zaroubi, S., Kooistra, R., and Cooray, A., Tomographic intensity mapping versus galaxy surveys: observing the Universe in H α emission with new generation instruments, Apr. 2018, MNRAS, 475, 1587

Simonyan, K., Vedaldi, A., and Zisserman, A., Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, Dec. 2013, arXiv e-prints, arXiv:1312.6034

Simonyan, K., Vedaldi, A., and Zisserman, A. 2014, in In Workshop at International Conference on Learning Representations

Smith, L. N., A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay, Mar. 2018, arXiv e-prints, arXiv:1803.09820

Sobral, D., Smail, I., Best, P. N., Geach, J. E., Matsuda, Y., Stott, J. P., Cirasuolo, M., and Kurk, J., A large Hα survey at z = 2.23, 1.47, 0.84 and 0.40: the 11 Gyr evolution of star-forming galaxies from HiZELS ★, Jan. 2013, MNRAS, 428, 1128

Sobral, D., Stroe, A., Koyama, Y., Darvish, B., Calhau, J., Afonso, A., Kodama, T., and Nakata, F., The nature of Hα star-forming galaxies at z $\sim$ 0.4 in and around Cl 0939+4713: the environment matters, Jun. 2016, MNRAS, 458, 3443

Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B. S., Greene, T., Guyon, O., Hirata, C., Kalirai, J., Kasdin, N. J., Macintosh, B., Moos, W., Perlmutter, S., Postman, M., Rauscher, B., Rhodes, J., Wang, Y., Weinberg, D., Benford, D., Hudson, M., Jeong, W. S., Mellier, Y., Traub, W., Yamada, T., Capak, P., Colbert, J., Masters, D., Penny, M., Savransky, D., Stern, D., Zimmerman, N., Barry, R., Bartusek, L., Carpenter, K., Cheng, E., Content, D., Dekens, F., Demers, R., Grady, K., Jackson, C., Kuan, G., Kruk, J., Melton, M., Nemati, B., Parvin, B., Poberezhskiy, I., Peddie, C., Ruffa, J., Wallace, J. K., Whipple, A., Wollack, E., and Zhao, F., Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report, Mar. 2015, arXiv e-prints, arXiv:1503.03757

Springel, V., The cosmological simulation code GADGET-2, Dec. 2005, MNRAS, 364, 1105

Springel, V., E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh, 01. 2010, Monthly Notices of the Royal Astronomical Society, 401, 791

Springel, V. and Hernquist, L., Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation, Feb. 2003, MNRAS, 339, 289

Springel, V., Pakmor, R., Pillepich, A., Weinberger, R., Nelson, D., Hernquist, L., Vogelsberger, M., Genel, S., Torrey, P., Marinacci, F., and Naiman, J., First results from the IllustrisTNG simulations: matter and galaxy clustering, Mar. 2018, MNRAS, 475, 676

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., Dropout: A Simple Way to Prevent Neural Networks from Overfitting. 2014, Journal of Machine Learning Research, 15, 1929

Stacey, G. J., Aravena, M., Basu, K., Battaglia, N., Beringue, B., Bertoldi, F., Bond, J. R., Breysse, P., Bustos, R., Chapman, S., Chung, D. T., Cothard, N., Erler, J., Fich, M., Foreman, S., Gallardo, P., Giovanelli, R., Graf, U. U., Haynes, M. P., Herrera-Camus, R., Herter, T. L., Hložek, R., Johnstone, D., Keating, L., Magnelli, B., Meerburg, D., Meyers, J., Murray, N., Niemack, M., Nikola, T., Nolta, M., Parshley, S. C., Riechers, D. A., Schilke, P., Scott, D., Stein, G., Stevens, J., Stutzki, J., Vavagiakis, E. M., and Viero, M. P. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10700, Ground-based and Airborne Telescopes VII, ed. H. K. Marshall & J. Spyromilio, 107001M

Sun, G., Chang, T. C., Uzgil, B. D., Bock, J. J., Bradford, C. M., Butler, V., Caze-Cortes, T., Cheng, Y. T., Cooray, A., Crites, A. T., Hailey-Dunsheath, S., Emerson, N., Frez, C., Hoscheit, B. L., Hunacek, J., Keenan, R. P., Li, C. T., Madonia, P., Marrone, D. P., Moncelsi, L., Shiu, C., Trumper, I., Turner, A., Weber, A., Wei, T. S., and Zemcov, M., Probing Cosmic Reionization and Molecular Gas Growth with TIME, Jul. 2021, ApJ, 915, 33

Switzer, E. R., Masui, K. W., Bandura, K., Calin, L. M., Chang, T. C., Chen, X. L., Li, Y. C., Liao, Y. W., Natarajan, A., Pen, U. L., Peterson, J. B., Shaw, J. R., and Voytek, T. C., Determination of z ~0.8 neutral hydrogen fluctuations using the 21cm intensity mapping autocorrelation., Jul. 2013, MNRAS, 434, L46

Takada, M., Ellis, R. S., Chiba, M., Greene, J. E., Aihara, H., Arimoto, N., Bundy, K., Cohen, J., Doré, O., Graves, G., Gunn, J. E., Heckman, T., Hirata, C. M., Ho, P., Kneib, J.-P., Le Fèvre, O., Lin, L., More, S., Murayama, H., Nagao, T., Ouchi, M., Seiffert, M., Silverman,

J. D., Sodré, L., Spergel, D. N., Strauss, M. A., Sugai, H., Suto, Y., Takami, H., and Wyse, R., Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph, Feb. 2014, PASJ, 66, R1

Tegmark, M., Measuring Cosmological Parameters with Galaxy Surveys, Nov. 1997, Phys. Rev. Lett., 79, 3806

The HERA Collaboration, Abdurashidova, Z., Aguirre, J. E., Alexander, P., Ali, Z., Balfour, Y., Barkana, R., Beardsley, A., Bernardi, G., Billings, T., Bowman, J., Bradley, R., Bull, P., Burba, J., Carey, S., Carilli, C., Cheng, C., DeBoer, D., Dexter, M., de Lera Acedo, E., Dillon, J., Ely, J., Ewall-Wice, A., Fagnoni, N., Fialkov, A., Fritz, R., Furlanetto, S., Gale-Sides, K., Glendenning, B., Gorthi, D., Greig, B., Grobbelaar, J., Halday, Z., Hazelton, B., Heimersheim, S., Hewitt, J., Hickish, J., Jacobs, D., Julius, A., Kern, N., Kerrigan, J., Kittiwisit, P., Kohn, S., Kolopanis, M., Lanman, A., La Plante, P., Lekalake, T., Lewis, D., Liu, A., Ma, Y.-Z., MacMahon, D., Malan, L., Malgas, C., Maree, M., Martinot, Z., Matsetela, E., Mesinger, A., Mirocha, J., Molewa, M., Morales, M., Mosiane, T., Munoz, J., Murray, S., Neben, A., Nikolic, B., Devi Nunhokee, C., Parsons, A., Patra, N., Pieterse, S., Pober, J., Qin, Y., Razavi-Ghods, N., Reis, I., Ringuette, J., Robnett, J., Rosie, K., Santos, M., Sikder, S., Sims, P., Smith, C., Syce, A., Thyagarajan, N., Williams, P., and Zheng, H., HERA Phase I Limits on the Cosmic 21-cm Signal: Constraints on Astrophysics and Cosmology During the Epoch of Reionization, Aug. 2021, arXiv e-prints, arXiv:2108.07282

Ulyanov, D., Vedaldi, A., and Lempitsky, V. 2016, Instance Normalization: The Missing Ingredient for Fast Stylization, arXiv:1607.08022

van Haarlem, M. P., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W. T., de Bruyn, A. G., Nijboer, R., Swinbank, J., Fallows, R., Brentjens, M., Nelles, A., Beck, R., Falcke, H., Fender, R., Hörandel, J., Koopmans, L. V. E., Mann, G., Miley, G., Röttgering, H., Stappers, B. W., Wijers, R. A. M. J., Zaroubi, S., van den Akker, M., Alexov, A., Anderson, J., Anderson, K., van Ardenne, A., Arts, M., Asgekar, A., Avruch, I. M., Batejat, F., Bähren, L., Bell, M. E., Bell, M. R., van Bemmel, I., Bennema, P., Bentum, M. J., Bernardi, G., Best, P., Bîrzan, L., Bonafede, A., Boonstra, A. J., Braun, R., Bregman, J., Breitling, F., van de Brink, R. H., Broderick, J., Broekema, P. C., Brouw, W. N., Brüggen, M., Butcher, H. R., van Cappellen, W., Ciardi, B., Coenen, T., Conway, J., Coolen, A., Corstanje, A., Damstra, S., Davies, O., Deller, A. T., Dettmar, R. J., van Diepen, G., Dijkstra, K., Donker, P., Doorduin, A., Dromer, J., Drost, M., van Duin, A., Eislöffel, J., van Enst, J., Ferrari, C., Frieswijk, W., Gankema, H., Garrett, M. A., de Gasperin, F., Gerbers, M., de Geus, E., Grießmeier, J. M., Grit, T., Gruppen, P., Hamaker, J. P., Hassall, T., Hoeft, M., Holties, H. A., Horneffer, A., van der Horst, A., van Houwelingen, A., Huijgen, A., Iacobelli, M., Intema, H., Jackson, N., Jelic, V., de Jong, A., Juette, E., Kant, D., Karastergiou, A., Koers, A., Kollen, H., Kondratiev, V. I., Kooistra, E., Koopman, Y., Koster, A., Kuniyoshi, M., Kramer, M., Kuper, G., Lambropoulos, P., Law, C., van Leeuwen, J., Lemaitre, J., Loose, M., Maat, P., Macario, G., Markoff, S., Masters, J., McFadden, R. A., McKay-Bukowski, D., Meijering, H., Meulman, H., Mevius, M., Middelberg, E., Millenaar, R., Miller-Jones, J. C. A., Mohan, R. N., Mol, J. D., Morawietz, J., Morganti, R., Mulcahy, D. D., Mulder, E., Munk, H., Nieuwenhuis, L., van Nieuwpoort, R., Noordam, J. E., Norden, M., Noutsos, A., Offringa, A. R., Olofsson, H., Omar, A., Orrú, E., Overeem, R., Paas, H., Pandey-Pommier, M., Pandey, V. N., Pizzo, R., Polatidis, A., Rafferty, D., Rawlings, S., Reich, W., de Reijer, J. P., Reitsma, J., Renting, G. A., Riemers, P., Rol, E., Romein,

J. W., Roosjen, J., Ruiter, M., Scaife, A., van der Schaaf, K., Scheers, B., Schellart, P., Schoenmakers, A., Schoonderbeek, G., Serylak, M., Shulevski, A., Sluman, J., Smirnov, O., Sobey, C., Spreeuw, H., Steinmetz, M., Sterks, C. G. M., Stiepel, H. J., Stuurwold, K., Tagger, M., Tang, Y., Tasse, C., Thomas, I., Thoudam, S., Toribio, M. C., van der Tol, B., Usov, O., van Veelen, M., van der Veen, A. J., ter Veen, S., Verbiest, J. P. W., Vermeulen, R., Vermaas, N., Vocks, C., Vogt, C., de Vos, M., van der Wal, E., van Weeren, R., Weggemans, H., Weltevrede, P., White, S., Wijnholds, S. J., Wilhelmsson, T., Wucknitz, O., Yatawatta, S., Zarka, P., Zensus, A., and van Zwieten, J., LOFAR: The LOw-Frequency ARray, Aug. 2013, A&A, 556, A2

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I., Attention Is All You Need, Jun. 2017, arXiv e-prints, arXiv:1706.03762

Vieira, J., Aguirre, J., Bradford, C. M., Filippini, J., Groppi, C., Marrone, D., Bethermin, M., Chang, T.-C., Devlin, M., Dore, O., Fu, J. F., Hailey Dunsheath, S., Holder, G., Keating, G., Keenan, R., Kovetz, E., Lagache, G., Mauskopf, P., Narayanan, D., Popping, G., Shirokoff, E., Somerville, R., Trumper, I., Uzgil, B., and Zmuidzinas, J., The Terahertz Intensity Mapper (TIM): a Next-Generation Experiment for Galaxy Evolution Studies, Sep. 2020, arXiv e-prints, arXiv:2009.14340

Villaescusa-Navarro, F., Bull, P., and Viel, M., Weighing Neutrinos with Cosmic Neutral Hydrogen, Dec. 2015, ApJ, 814, 146

Villani, C. 2008, Optimal transport – Old and new, Vol. 338, xxii+973

Villanueva-Domingo, P. and Villaescusa-Navarro, F., Removing Astrophysics in 21 cm Maps with Neural Networks, Jan. 2021, ApJ, 907, 44

Visbal, E., Haiman, Z., and Bryan, G. L., Looking for Population III stars with He II line intensity mapping, Jul. 2015, MNRAS, 450, 2506

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Nelson, D., and Hernquist, L., Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe, Oct. 2014, MNRAS, 444, 1518

Voytek, T. C., Natarajan, A., Jáuregui García, J. M., Peterson, J. B., and López-Cruz, O., Probing the Dark Ages at z ~20: The SCI-HI 21 cm All-sky Spectrum Experiment, Feb. 2014, ApJ, 782, L9

Vrbanec, D., Ciardi, B., Jelić, V., Jensen, H., Zaroubi, S., Fernandez, E. R., Ghosh, A., Iliev, I. T., Kakiichi, K., Koopmans, L. V. E., and Mellema, G., Predictions for the 21 cm-galaxy cross-power spectrum observable with LOFAR and Subaru, Mar. 2016, MNRAS, 457, 666

Weinberger, R., Springel, V., Hernquist, L., Pillepich, A., Marinacci, F., Pakmor, R., Nelson, D., Genel, S., Vogelsberger, M., Naiman, J., and Torrey, P., Simulating galaxy formation with black hole driven thermal and kinetic feedback, Mar. 2017, MNRAS, 465, 3291

Wiersma, R. P. C., Ciardi, B., Thomas, R. M., Harker, G. J. A., Zaroubi, S., Bernardi, G., Brentjens, M., de Bruyn, A. G., Daiboo, S., Jelic, V., Kazemi, S., Koopmans, L. V. E., Labropoulos, P., Martinez, O., Mellema, G., Offringa, A., Pandey, V. N., Schaye, J., Veligatla, V., Vedantham, H., and Yatawatta, S., LOFAR insights into the epoch of reionization from the cross-power spectrum of 21 cm emission and galaxies, Jul. 2013, MNRAS, 432, 2615

Wild, J. P., The Radio-Frequency Line Spectrum of Atomic Hydrogen and its Applications in Astronomy., Mar. 1952, ApJ, 115, 206

Wouthuysen, S. A., On the excitation mechanism of the 21-cm (radio-frequency) interstellar hydrogen emission line., Jan. 1952, AJ, 57, 31

Yue, B. and Ferrara, A., Studying high-z galaxies with [C II] intensity mapping, Dec. 2019, MNRAS, 490, 1928

Zaroubi, S., Hoffman, Y., Fisher, K. B., and Lahav, O., Wiener Reconstruction of the Large-Scale Structure, Aug. 1995, ApJ, 449, 446

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