

# Foreground removal of CO intensity mapping using deep learning

Xingchen Zhou,<sup>1,2</sup> Yan Gong<sup>1,3</sup>★, Furen Deng,<sup>1,2</sup> Meng Zhang,<sup>1,2</sup> Bin Yue<sup>1</sup> and Xuelei Chen<sup>1,2,4,5</sup>

<sup>1</sup>National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China

<sup>2</sup>School of Astronomy and Space Science, University of Chinese Academy of Science, Beijing 100049, China

<sup>3</sup>Science Center for China Space Station Telescope, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100101, China

<sup>4</sup>Department of Physics, College of Sciences, Northeastern University, Shenyang 110819, China

<sup>5</sup>Center for High Energy Physics, Peking University, Beijing 100871, China

Accepted 2023 February 17. Received 2023 February 14; in original form 2022 December 30

## ABSTRACT

Line intensity mapping (LIM) is a promising probe to study star formation, the large-scale structure of the Universe, and the epoch of reionization (EoR). Since carbon monoxide (CO) is the second most abundant molecule in the Universe except for molecular hydrogen H<sub>2</sub>, it is suitable as a tracer for LIM surveys. However, just like other LIM surveys, CO intensity mapping also suffers strong foreground contamination that needs to be eliminated for extracting valuable astrophysical and cosmological information. In this work, we take <sup>12</sup>CO(*J* = 1–0) emission line as an example to investigate whether deep learning method can effectively recover the signal by removing the foregrounds. The CO(1–0) intensity maps are generated by N-body simulations considering CO luminosity and halo mass relation, and we discuss two cases with median and low CO signals by comparing different relations. We add foregrounds generated from real observations, including thermal dust, spinning dust, free-free, synchrotron emission, and cosmic microwave background anisotropy. The beam with sidelobe effect is also considered. Our deep learning model is built upon ResUNet, which combines image generation algorithm UNet with the state-of-the-art architecture of deep learning, ResNet. The principal component analysis (PCA) method is employed to preprocess data before feeding it to the ResUNet. We find that, in the case of low instrumental noise, our UNet can efficiently reconstruct the CO signal map with correct line power spectrum by removing the foregrounds and recovering PCA signal loss and beam effects. Our method also can be applied to other intensity mappings like neutral hydrogen 21-cm surveys.

**Key words:** methods: data analysis – diffuse radiation – large-scale structure of Universe.

## 1 INTRODUCTION

Line intensity mapping (LIM) is a promising technique that can efficiently explore huge survey volume in an affordable observational time. It integrates the emission line intensity from galaxies and intergalactic medium (IGM), and does not require resolving individual sources (Gong et al. 2011; Visbal, Trac & Loeb 2011; Fonseca et al. 2017; Kovetz et al. 2017; Bernal & Kovetz 2022). LIM has been widely proposed in studying the epoch of reionization (EoR), galaxy star formation, cosmic large-scale structure, etc. It offers an opportunity to comprehensively explore our Universe in redshifts beyond the reach of traditional galaxy surveys.

Several emission lines are commonly proposed and used as tracers in LIM, such as carbon monoxide (CO), singly ionized carbon [C II], H $\alpha$ , and neutral hydrogen 21-cm emission lines (Visbal & Loeb 2010; Carilli 2011; Gong et al. 2011, 2012, 2014, 2017; Lidz et al. 2011; Gong, Cooray & Santos 2013; Silva et al. 2013, 2015; Pullen, Doré & Bock 2014; Uzgil et al. 2014; Fonseca et al. 2017; Gong, Chen & Cooray 2020). Among these lines, CO is the second most abundant molecule in the Universe, except for molecular hydrogen

(H<sub>2</sub>). At high redshift ( $z \gtrsim 6$ ) around EoR, CO can trace the star-forming regions within galaxies to probe cosmic reionization history (Carilli 2011; Gong et al. 2011; Lidz et al. 2011). At low redshift with  $z \lesssim 3$ , CO LIM can be used to investigate star formation rate, and map the large-scale matter field to study the nature of dark energy and dark matter (Kovetz et al. 2017). The CO rotational line emission is at ladder of frequency  $\nu_{J \rightarrow J-1} = J \times 115.27$  GHz for  $J \rightarrow J-1$  transition. Since CO lines are usually bright in galaxies, ground-based telescopes targeting sub-mm wavelength can efficiently observe them (Bernal & Kovetz 2022). Currently, several ongoing and planning probes are targeting at CO lines, such as COPSS (Keating et al. 2015, 2016), mmiME (Keating et al. 2020), COMAP (Breyse et al. 2022; Cleary et al. 2022), FYST (CCAT-Prime Collaboration 2021), and SPT-SLIM (Karkare et al. 2022). COMAP recently has published their first-stage results and shown promising prospects (Chung et al. 2022; Ihle et al. 2022).

On the other hand, foreground removal is a big problem in CO and other LIM surveys, that can greatly affect the signal extraction. The LIM foregrounds include the continuum components, such as thermal dust, free-free and synchrotron emissions, and non-continuum ones, e.g. cosmic microwave background (CMB) anisotropy, spinning dust emission, and contaminations of other emission lines from different redshifts at the same observed frequency. The continuum foreground

★ E-mail: [gongyan@bao.ac.cn](mailto:gongyan@bao.ac.cn)

usually is attempted to be removed by blind subtraction methods, such as principal component analysis (PCA; Wold, Esbensen & Geladi 1987; Abdi & Williams 2010), fast independent component analysis (fastICA; Hyvärinen & Oja 2000), etc. PCA is one of the mostly used methods. It can extract the principal components (PCs) with large eigenvalues of covariance matrix, and reduce the contamination by removing the contribution from these modes. Although PCA is effective in removing continuum foregrounds, it may corrupt the information and cause signal loss at large scales (Alonso et al. 2014; Deng et al. 2022). So algorithms that can recover the power spectrum usually need to be applied after PCA subtraction.

The instrumental effects will further complicate the foreground removal of LIM. Gaussian white noise induced by system temperature of telescope usually can affect the signal extraction from observational maps. Although it is relatively easy to be evaluated and removed in power spectra since it should be a constant at different scales, this noise should be suppressed as much as possible to avoid overwhelming the signal. Another instrumental effect called beam effect should also be carefully considered. The beam effect is caused by main and side lobes of radio antenna (Matshawule et al. 2021), and has frequency-dependent feature. On one hand, it can affect the spatial resolution and result in a drop in the power spectrum at small scales. On the other hand, beam effect can change the shape of foreground spectrum in frequency space. Hence, the foreground removal by blind subtraction methods, e.g. PCA, usually cannot obtain satisfactory performance when beam effect is considered (Ni et al. 2022).

Deep learning is widely used in information extraction of images, that probably can be applied in foreground removal and instrumental effect correction in LIM. Deep learning is one of the most prosperous branches of machine learning. The development and optimization of deep learning algorithm are in active research. Convolutional neural networks (CNN) introduced by Fukushima & Miyake (1982) are the most widely used architecture. The performance of CNN improves with increase of convolutional layers, however, deep layers will suffer gradient vanishing problem and cannot learn effectively. He et al. (2015) proposes ResNet that can address this problem using identity mapping. Applying ResNet block, CNN can be built very deep, and even hundreds of convolutional layers will not cause gradient vanishing problem, greatly improving the learning ability of CNN.

For image generation tasks like foreground removal, UNet and generative adversarial network (GAN) are the suitable networks. UNet is originated from autoencoder (Rumelhart, Hinton & Williams 1985) and improve its generative ability through concatenation with features obtained in encoding path (Ronneberger, Fischer & Brox 2015). Autoencoders aim to transform inputs to outputs with small distortions. However, with skip connection introduced in UNet and deep convolutional layers, the transformation ability of autoencoder significantly increases. This structure can be used for segmentation of filaments, walls and voids in large scale structure (Aragon-Calvo 2019), mapping galaxies from dark matter halos (Zhang et al. 2019; Kasmanoff et al. 2020), removal of foreground (Makinen et al. 2021; Ni et al. 2022), and many other interesting applications in astronomical and cosmological field.

GAN is a generative model proposed in recent years (Goodfellow et al. 2014), and it is more complicated than common networks. It has a critic or discriminator that tells the realness of generated images, and guides the generator to output more realistically. One structure of GAN family that can perform image-to-image generation is called Pix2PixGAN (Isola et al. 2016). This GAN can be regarded as conditional GAN (cGAN; Mirza & Osindero 2014) with input images as conditions. GAN structures can simulate realistic cosmic

web in extremely short time (Rodríguez et al. 2018), generate 3D N-body simulations (Perraudin et al. 2019; Feder, Berger & Stein 2020), and polish simulations from existing low resolution ones (Kodi Ramanah et al. 2020; Li et al. 2021). Therefore it can be built into emulators, and hopefully substitute the expensive N-body or hydrodynamical simulations. Denoising LIM is also a promising application (Moriwaki, Shirasaki & Yoshida 2021; Moriwaki & Yoshida 2021).

In this work, we take  $^{12}\text{CO}(J=1-0)$  line as an example, which has a rest-frame wavelength of 2.6 mm and frequency of 115.27 GHz, to study foreground removal using deep learning. We consider low-redshift CO(1-0) LIM at  $z \sim 1$  contaminated by foregrounds and instrumental noise, including thermal dust, spinning dust, free-free emission, synchrotron emission, CMB anisotropy, beam effect, thermal noise, etc. CO line intensity maps are mocked by N-body simulation and the relation of CO luminosity and halo mass. Foreground components at frequencies of CO signals are obtained from PYSM3 PYTHON library, which are fitted by publicly available data from WMAP and Planck surveys and extended to small scale (Thorne et al. 2017). The instrumental effects we consider are based on COMAP, and the Bessel beam model is assumed in the analysis. PCA method is tested and used as preprocessing for input of our UNet, and finally we evaluate the foreground removal performance by UNet.

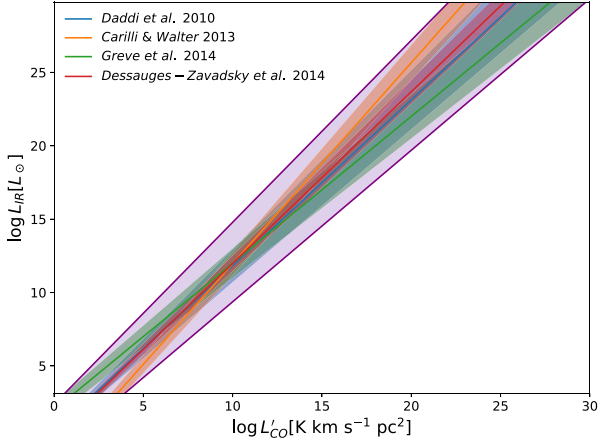
The paper is organized as follows: we discuss our mock data including CO intensity maps, foregrounds, instrumental noise, and beam effect in Section 2. In Section 3, we introduce our deep learning model and training process. The results are shown in Section 4 and we also have some discussions on instrumental and PCA effects. We conclude this work in Section 5.

## 2 MOCK DATA

We consider CO(1-0) intensity mapping at  $z \sim 1$  contaminated by several foreground components, and instrumental effects including white noise and beam effects induced by radio telescope are also simulated. We use a N-body simulation at redshift around  $z = 1$  to generate our CO mock maps. The foreground components considered are thermal dust, spinning dust, synchrotron, free-free emissions, and CMB anisotropy. The white noise is simulated based on and modified from instrumental parameters of COMAP, and the Bessel beam model is employed to mock the beam effects of radio antenna.

### 2.1 CO intensity maps

We generate CO mock maps based on MDPL2 from CosmoSim simulation sets (Klypin et al. 2016). This N-body simulation uses cosmological parameters from Planck2013 (Planck Collaboration 2014) and is performed by L-GADGET2 code (Springel et al. 2005). The box size is  $1 \text{ Gpc h}^{-1}$ , number of particles is  $3840^3$ , with a mass resolution of  $1.51 \times 10^9 h^{-1} M_{\odot}$  of single dark matter particle. This resolution is large enough for our LIM study. We make use of the halo catalogue obtained by friend-of-friend method from snapshot of  $z = 1.03$  to be our basis for CO intensity maps. We split the whole box to subboxes of  $300 \text{ Mpc h}^{-1}$  to generate lightcones with angles of  $7.45 \times 7.45 \text{ deg}^2$ . After generating the lightcones, we slice the planes according to the resolution of frequency, taken to be 0.1 GHz in this work. Since we set  $z = 1$  at the middle of subbox, the frequency range is from 55.2 to 60.2 GHz with 0.1 GHz spacing. We find N halos inside each spatial pixel in each frequency plane, and calculate



**Figure 1.**  $L_{\text{IR}} - L'_{\text{CO}}$  relations used in this work. We consider the red line with  $(\alpha, \beta) = (1.17, 0.28)$  as our fiducial model (Dessauges-Zavadsky et al. 2015). The lower and upper bounds of these luminosity relations (purple lines) are also derived with  $(\alpha, \beta) = (1.24, 2.39)$  and  $(1.03, -0.96)$ , respectively.

the observed CO intensity by

$$I_{\text{CO}}(\nu) = \frac{1}{(\Delta\theta)^2} \sum_i^N \frac{1}{\Delta\nu} \frac{L_{\text{CO}}^i}{4\pi r_i^2 (1+z_i)^2}, \quad (1)$$

where  $\Delta\theta$  is the angular size of spatial pixel,  $\Delta\nu$  is the resolution of frequency,  $L_{\text{CO}}^i$ ,  $r_i$ , and  $z_i$  are CO luminosity, comoving distance, and redshift of the  $i$ th halo, respectively. The CO luminosity can be calculated from  $L_{\text{CO}} - M_{\text{h}}$  relation, and we obtain this relation with the help of infrared (IR) luminosity.

Since CO and IR luminosity are both star formation tracers, they can be closely related by empirical relation,

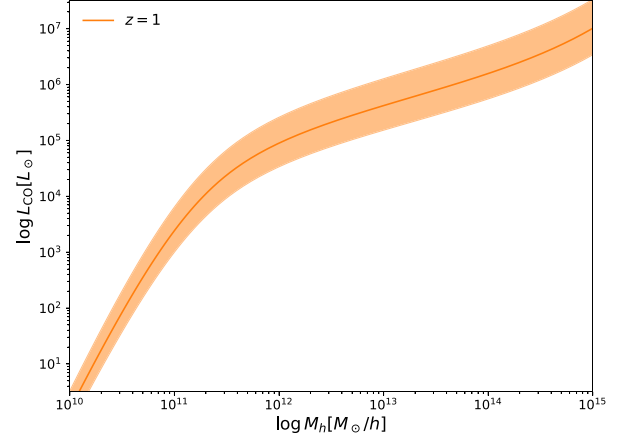
$$\log(L_{\text{IR}}) = \alpha \log(L'_{\text{CO}}) + \beta, \quad (2)$$

where  $L_{\text{IR}}$  is in units of  $L_{\odot}$  and  $L'_{\text{CO}}$  is in units of  $\text{K km s}^{-1} \text{ pc}^2$ .  $\alpha$  and  $\beta$  are fitted from galaxy observations, and we consider  $(\alpha, \beta) = (1.13 \pm 0.09, 0.53 \pm 0.22)$  (Daddi et al. 2010),  $(1.37 \pm 0.04, -1.74 \pm 0.40)$  (Carilli & Walter 2013),  $(1.00 \pm 0.05, 2.00 \pm 0.50)$  (Greve et al. 2014), and  $(1.17 \pm 0.03, 0.28 \pm 0.23)$  (Dessauges-Zavadsky et al. 2015) in this work. These relations are measured by observations, and are available at  $L_{\text{IR}} \lesssim 10^{14} L_{\odot}$  (see e.g. Aravena et al. 2012), which should cover the halo mass range with  $M \lesssim 10^{15} M_{\odot}$  in our simulation. The relations are shown in Fig. 1, and the lower and upper bounds of the luminosity relations are also estimated as  $(\alpha, \beta) = (1.24, 2.39)$  and  $(1.03, -0.96)$ , respectively, shown in purple lines. We take the median values  $(\alpha, \beta) = (1.17, 0.28)$  as our fiducial model and the lower bound with  $(1.24, 2.39)$  as the low CO signal model to check the performance of our foreground removal method. The unit conversion from  $L'_{\text{CO}}$  to  $L_{\text{CO}}$  in units of  $L_{\odot}$  is

$$L_{\text{CO}} = 4.9 \times 10^{-5} L_{\odot} \left( \frac{\nu'_{\text{CO}}}{115.27 \text{ GHz}} \right)^3 \left( \frac{L'_{\text{CO}}}{\text{K km s}^{-1} \text{ pc}^2} \right), \quad (3)$$

where  $\nu'_{\text{CO}} = 115.27 \text{ GHz}$  is the rest-frame frequencies of CO(1-0) transition.

Then we need to derive the relation between IR luminosity  $L_{\text{IR}}$  and halo mass  $M_{\text{h}}$ . By assuming the ‘abundance matching’ technique can be applied, this relation could be obtained. Here we also assume that the subhalos contribute little to the IR luminosity and can be neglected (Yue et al. 2015). The abundance matching equates the number density of galaxies with IR luminosity above a luminosity



**Figure 2.**  $L_{\text{CO}} - M_{\text{h}}$  relation at  $z = 1$  derived from our fiducial  $L_{\text{IR}} - L'_{\text{CO}}$  model. The scatter of the relation is also shown in orange bands, which is estimated by the uncertainties of  $\alpha$  and  $\beta$  in the CO and IR luminosity relation.

$L_{\text{IR}}$  and the number density of halos above a halo mass  $M_{\text{h}}$  by

$$\int_{>L_{\text{IR}}} \Phi(L_{\text{IR}}, z) dL_{\text{IR}} = \int_{>M_{\text{h}}} \frac{dn}{dM_{\text{h}}} dM_{\text{h}}, \quad (4)$$

where  $\Phi(L_{\text{IR}}, z)$  is the IR luminosity function, and  $dn/dM_{\text{h}}$  is the halo mass function given in Sheth & Tormen (1999). Based on the simulation, we set the ranges of the halo mass and IR luminosity as  $M \in (10^{10}, 10^{15} M_{\odot}/h)$  and  $L_{\text{IR}} \in (10^5, 10^{14} L_{\odot})$ , respectively, to perform the abundance matching.

The IR luminosity functions of spiral galaxies, starburst galaxies, and star-forming galaxies including active galactic nuclei are comprehensively studied in literatures (e.g. Magnelli et al. 2009; Gruppioni et al. 2013; Magnelli et al. 2014). Here we choose to use an IR luminosity function given in Gruppioni et al. (2013), which can be written as

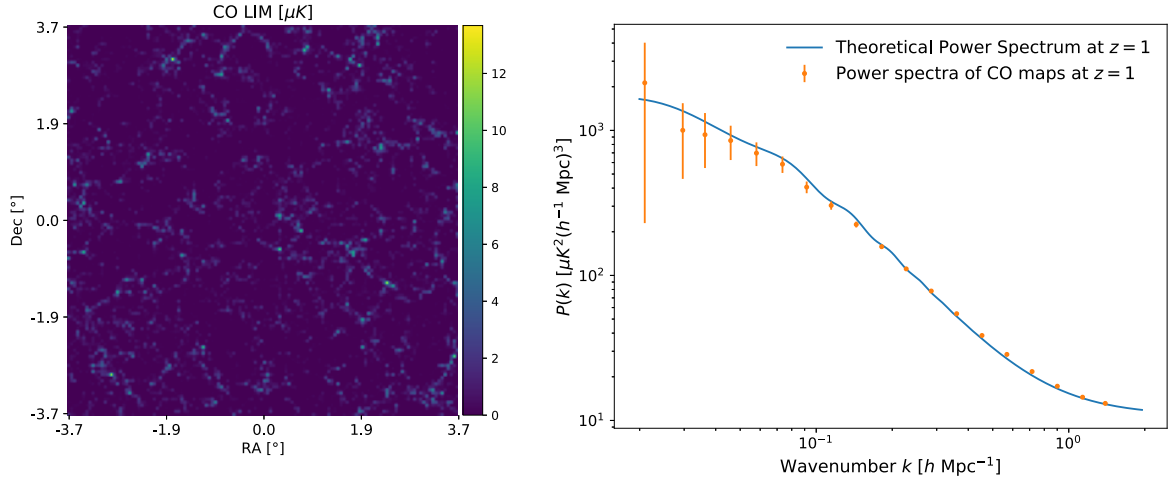
$$\Phi = \Phi_* \left( \frac{L_{\text{IR}}}{L_{\text{IR}}^*} \right)^{1-\alpha} \exp \left[ -\frac{1}{2\sigma^2} \log^2 \left( 1 + \frac{L_{\text{IR}}}{L_{\text{IR}}^*} \right) \right]. \quad (5)$$

Here the redshift evolution of parameters  $\Phi_*$ ,  $L_{\text{IR}}^*$ ,  $\alpha$ , and  $\sigma$  are found to be

$$\begin{aligned} \Phi_* &= \begin{cases} 5.7 \times 10^{-3} (1+z)^{-0.57} & z \leq 1.1 \\ 6.81 \times 10^{-2} (1+z)^{-3.92} & z > 1.1 \end{cases} \\ L_{\text{IR}}^* &= \begin{cases} 7.68 \times 10^9 (1+z)^{3.55} & z \leq 1.85 \\ 5.80 \times 10^{10} (1+z)^{1.62} & z > 1.85 \end{cases} \\ \alpha &= \begin{cases} 1.15 & z \leq 0.3 \\ 1.2 & z > 0.3 \end{cases} \\ \sigma &= \begin{cases} 0.52 & z \leq 0.3 \\ 0.5 & z > 0.3 \end{cases}. \end{aligned}$$

Considering abundance matching (equation 4), relation between CO and IR luminosity (equation 2) and unit conversion (equation 3), the relation between CO luminosity  $L_{\text{CO}}$  and halo mass  $M_{\text{h}}$  can be derived. We show this relation for our fiducial model at  $z = 1$  in Fig. 2. The uncertainty of this relation is also shown in orange band, which is derived from the errors of  $\alpha$  and  $\beta$ . This result is also consistent with other studies (e.g. Padmanabhan 2018). After obtaining  $L_{\text{CO}} - M_{\text{h}}$  relation, we can derive the CO line intensity given by equation (1), and generate the CO signal intensity maps.

Here we take spatial resolution as  $1024 \times 1024$  with angular size  $\Delta\theta$  of each pixel as 0.44 arcmin. Note that this large resolution is prepared for adding beam effect that will be discussed later.



**Figure 3.** Left-hand panel: An example of CO(1-0) signal mock maps at 57.6 GHz (slice of  $z \sim 1$ ). Right-hand panel: The corresponding CO(1-0) signal power spectrum derived from the mock map, and the theoretical power spectrum assuming the same parameter values is also shown in blue curve for comparison.

Then each CO intensity map is a three-dimensional matrix with size  $1024 \times 1024 \times 51$ , and the third dimension indicates the number of frequency slices. Considering the six different directions of line-of-sight and 27 subboxes cut from the whole simulation box, we obtain 162 CO signal intensity maps in total. We show an example of CO intensity maps downsampled to  $128 \times 128$  at 57.6 GHz (slice at  $z \sim 1$ ) in the left-hand panel of Fig. 3. We also compare the derived line power spectrum from the mock map with the theoretical one using the same parameter values in the right-hand panel (Gong et al. 2011). The errorbars of the line power spectrum are calculated by the root-mean-square of errors from power spectrum of each map. We can see that the cosmic large-scale structure can be clearly recognized in the CO map, and the power spectrum of the simulated CO map is in good agreement with the theoretical one.

## 2.2 Foregrounds

From the brightness temperature rms of astrophysical components with respect to frequency in Planck 2015 results (Planck Collaboration 2016), we find that our CO maps within about 55 ~ 60 GHz can be contaminated by several foreground components, including thermal dust, spinning dust, synchrotron, free-free emissions, and CMB anisotropy. We use PYSM3 PYTHON package (Thorne et al. 2017) to generate all these foreground components. PYSM3 provides an easy-to-use interface to simulate the sky based on publicly available data from WMAP and Planck surveys. The sky components are extrapolated to user-specified frequencies using scaling law and maps of spectral parameters. And the high resolution sky can be obtained through extrapolating the angular power spectrum of available data to smaller scales. We obtain the sky maps of foreground components using this package with  $n_{\text{side}}$  as 8192. This resolution corresponds to an angle of 0.43 arcmin, which is approximately equivalent to the pixel resolution 0.44 arcmin of the CO signal maps. Therefore, the slicing of foreground can be performed without smoothing the information at small scale, and this resolution is high enough for simulating the beam effect in the survey.

After obtaining the sky maps of the foreground components, we generate square maps using HEALPIX with the same spatial pixels as the CO intensity maps in declinations higher than 50 degree and lower than -50 degree. This criteria is to ensure that the foreground emissions are not too large compared to that around galactic disc.

For sky area of  $7.45 \times 7.45 \text{ deg}^2$  in the simulation, we can obtain 48 foreground maps surrounding one circle of the whole sky and in total 96 maps for northern and southern hemispheres at frequencies the same as the CO maps. Each foreground component and the sum of them at 57.6 GHz are shown in Fig. 4. We notice that the spinning dust and free-free emissions have some grid effects, and they are caused by extrapolation of angular power spectra to small scales.

Note that we do not consider the contamination from other emission lines, such as CO(2-1) and CO(3-2) that can contaminate the CO(1-0) signal from higher redshifts. Since CO(1-0) line is luminous and has a long wavelength, the main contamination from other lines usually comes from higher redshifts, and these lines are then redshifted and become relatively not very strong. On the other hand, the contamination from interloper lines also can be effectively distinguished by the shape difference from the signal line in a redshift-space line intensity power spectrum (Visbal & Loeb 2010; Gong et al. 2014; Lidz & Taylor 2016), or can be modeled as cosmological signals from other redshifts (Gong et al. 2020). For simplicity, we ignore the contamination from these interloper lines, and only focus on the foregrounds mentioned above.

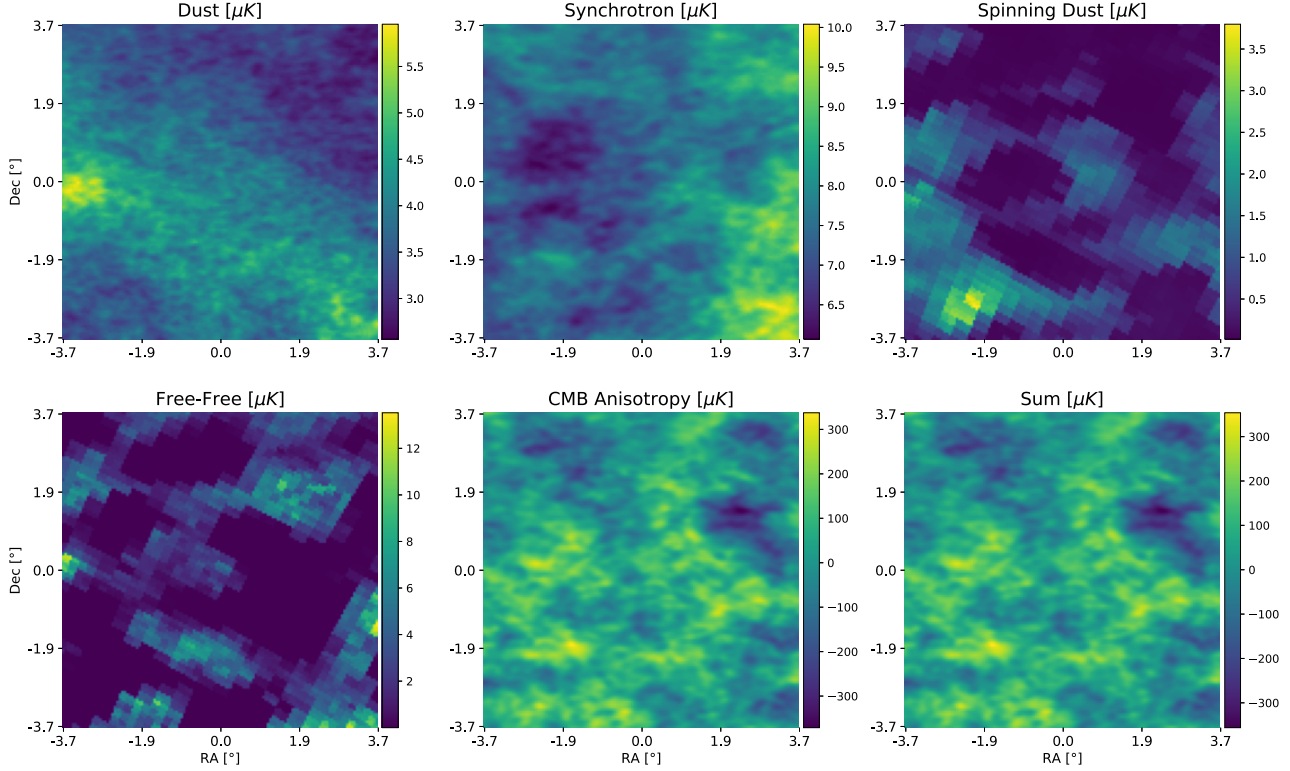
## 2.3 Instrumental effects

The instrumental effects we consider are white noise and beam effects of radio telescope. Here we derive the white noise from an assumed instrument with some of parameters similar to the CO Mapping Array Project (COMAP; Cleary et al. 2022). The instrumental noise per voxel can be estimated by

$$\sigma_T = \frac{T_{\text{sys}}}{\sqrt{\tau \Delta \nu}} = \frac{T_{\text{sys}} \sqrt{N_{\text{pixel}}}}{\sqrt{\tau_{\text{tot}} e_{\text{obs}} N_{\text{feed}} \Delta \nu}}. \quad (6)$$

We assume the total observing time  $\tau_{\text{tot}} = 9000$  hrs, efficiency of observation  $e_{\text{obs}} = 1$ ,  $N_{\text{feeds}} = 100$ , the frequency resolution  $\Delta \nu = 0.1$  GHz, and the system temperature  $T_{\text{sys}} = 40$  K. The number of pixels  $N_{\text{pixel}}$  can be estimated by the survey area and beam size, and we find  $N_{\text{pixel}} = 260 \times 260$  at  $z = 1$ . The instrumental parameters are summarized in Table 1. Then we calculate the instrumental noise per voxel and find that  $\sigma_T = 18.3 \mu\text{K}$  at 57.6 GHz. The noise is evenly added over voxels of the CO maps by Gaussian distribution. In the following discussion, we will firstly ignore the instrumental noise, and then discuss its effect on foreground removal.





**Figure 4.** The foreground components at 57.6 GHz, including thermal dust, synchrotron, spinning dust, free-free emissions, CMB anisotropy, and the sum of them. They are sliced by HEALPIX with similar angular resolution of the CO signal maps. We notice that the spinning dust and free-free emissions have some grid patterns, which are caused by the extrapolation of angular power spectra to small scales.

**Table 1.** Instrumental parameters considered in this work.

Parameter	Value
Telescope aperture, $D$ [m]	10.4
Survey area, $A_S$ [deg <sup>2</sup> ]	$7.45 \times 7.45$
Total observing time, $\tau_{\text{tot}}$ [hrs]	9000
Beam size at 57.6 GHz, $\theta_d$ [arcmin]	1.72
Number of feeds, $N_{\text{feed}}$	100
Frequency range, $B$ [GHz]	55–60
Frequency resolution, $\Delta\nu$ [GHz]	0.1
System temperature, $T_{\text{sys}}$ [K]	40

The beam characterizes the power response of a radio telescope to the sky signal, and the observed sky map is the convolution of the real sky and the beam of telescope. The beam of a telescope is generally chromatic, i.e. the shape of beam will change with respect to frequency. The chromaticity will cause fluctuation in the spectrum of foreground and increase the degeneracy between signal and foreground, and therefore potentially cause more signal loss or foreground leakage in foreground removal.

Gaussian function is widely used to model the beam. However, this simplified model lacks sidelobe and therefore underestimate the chromaticity. Other beam model with sidelobe of different level is proposed, e.g. Cosine model (Condon & Ransom 2016), gaussian tapered Airy dish (Harper et al. 2018), and Bessel or Jinc beam (Wilson, Rohlf & Hüttemeister 2013; Matshawule et al. 2021), based on different aperture tapering. Beam models based on electromagnetic simulation and astro-holographic observation are also proposed (Asad et al. 2021; Matshawule et al. 2021; Sun et al. 2022). Here we use the Bessel model corresponding to uniformly

illuminated circular aperture. We notify that the Bessel model is the worst case of sidelobe effect, because any tapering toward the edge of aperture will generally suppress the level of sidelobe.

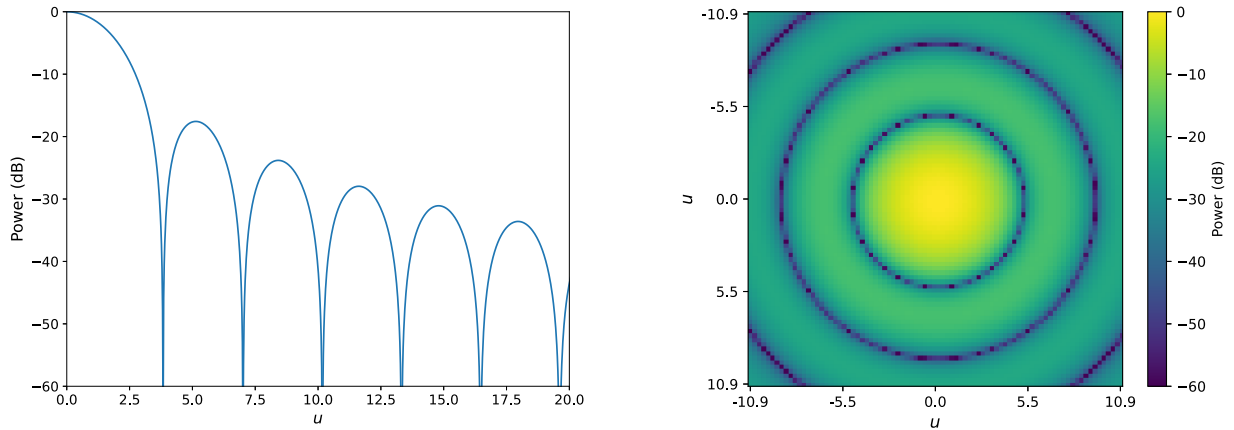
The power response of the Bessel model can be written as

$$F(\sin \theta) = \left( 2 \frac{J_1(u)}{u} \right)^2, \quad (7)$$

where

$$u = \frac{\sin \theta}{(\theta_d/\pi)} \text{ and } \theta_d = \frac{\lambda}{D}.$$

Here  $\lambda$  is the observed wavelength,  $D$  is the aperture of the antenna assumed to be 10.4 m the same as COMAP, and  $J_1$  is the Bessel function of the first kind. We plot the Bessel model and two-dimensional Bessel kernel at 57.6 GHz in Fig. 5. We notice that the first sidelobe is at  $u \sim 5$  or 2.86 arcmin, which is larger than the pixel resolutions of CO maps and foregrounds, and thus the sidelobe effects can be well simulated. We construct a beam kernel with the same resolution as the CO and foreground maps by including the first sidelobe, and convolve with the CO intensity maps and foregrounds to simulate the beam effect. Then we downsample these maps to spatial resolution of  $128 \times 128$  by averaging adjacent 8 pixels, and the instrumental noise level becomes  $\sigma = 9.0 \mu\text{K}$  per voxel. The spatial pixels as powers of 2 are easy for building and training of neural networks. We also notice that the peculiar velocity can degenerate with the beam effect, that will suppress the line power spectrum at small scales. Since the intensity mapping mainly explores the large scales and this work is focusing on the foreground removal, for simplicity, we ignore the peculiar velocity and other redshift-space distortion effect when generating the mock maps.



**Figure 5.** The Bessel model (left-hand panel) and the two-dimensional Bessel kernel (right-hand panel) at 57.6 GHz we assume in this work.

### 3 DEEP LEARNING METHOD

We use PCA as benchmark model to remove foregrounds and compare it with our deep learning method. PCA decomposes multivariate data set to the components which have largest variances (Wold et al. 1987; Abdi & Williams 2010). These components are called PCs or PCA modes, corresponding to eigenvectors of covariance matrix of the data. The two-dimensional Gaussian distributions can be decomposed to two PCs forming an ellipsoid for example. PCA subtraction finds the largest eigenvalues of covariance of data with largest variances and subtracts these components. This method tends to subtract foreground along with some signal information. And since it depends on the assumption that the foregrounds are smooth across frequencies, the foreground removal is further complicated considering the beam effect that is also correlated with frequencies. These two effects will lead to deviations of power spectra with true powers. Deep learning algorithm concerning image-to-image generation may recover these deviations by learning features from images. Our networks are implemented by Keras<sup>1</sup> with TensorFlow as backend.<sup>2</sup>

#### 3.1 ResUNet

Our deep learning model is built upon ResUNet (Diakogiannis et al. 2020), which effectively combines UNet with ResNet. UNet is introduced in biomedical image segmentation (Ronneberger et al. 2015), and is based on CNN. To perform image-to-image generation task, UNet uses a bottleneck structure. It convolves and downsamples the maps to learn the relevant features in encoding path. In decoding path where it generates images, UNet uses transposed convolution or upsampling-convolution to increase the spatial dimensions of feature maps to output size. And it uses a bridge to connect the features learned from encoding path with decoding path to increase its generative ability.

Our UNet uses modified structure called ResUNet (Diakogiannis et al. 2020), which is constructed based on ResNet (He et al. 2015). ResNet is a milestone in development of deep learning field, and it helps tackling the vanishing gradient problem using identity mapping method. Since CNN increases its ability by stacking more convolutional layers, we can build very deep networks using ResNet block without worrying the training problem induced by vanishing

gradient problem. Two types of ResNet blocks are employed in our work, called ResNet1 and ResNet2. Using a scaling layer, ResNet1 can downsample the spatial dimension of maps by a factor of 2, and the inputs are also downsampled and added to the outputs from the main convolution path. On the other hand, ResNet2 keeps the spatial dimension using the same-padding convolutional layers. In decoding path, ResNet1 is altered to upsample the spatial dimension. He et al. (2015) employs two convolutional layers using same kernel sizes in his originally implemented ResNet block, however, here we alter the kernel size of first layer to be  $1 \times 1$  for decreasing the channels and reduce the computing expense. We have proven that this modified structure has no influence on performance. The upsampling can be performed by UpSampling or transposed convolutional layers. Here we choose UpSampling to avoid the potential checkerboard effect (Odena, Dumoulin & Olah 2016). BatchNormalization (Ioffe & Szegedy 2015) and LeakyReLU (Maas et al. 2013) are employed to reduce overfitting and apply non-linearity after convolution, respectively. LeakyReLU is improved from ReLU and allows a small gradient when some units are below 0, and the negative slope parameter of LeakyReLU is taken to be 0.3.

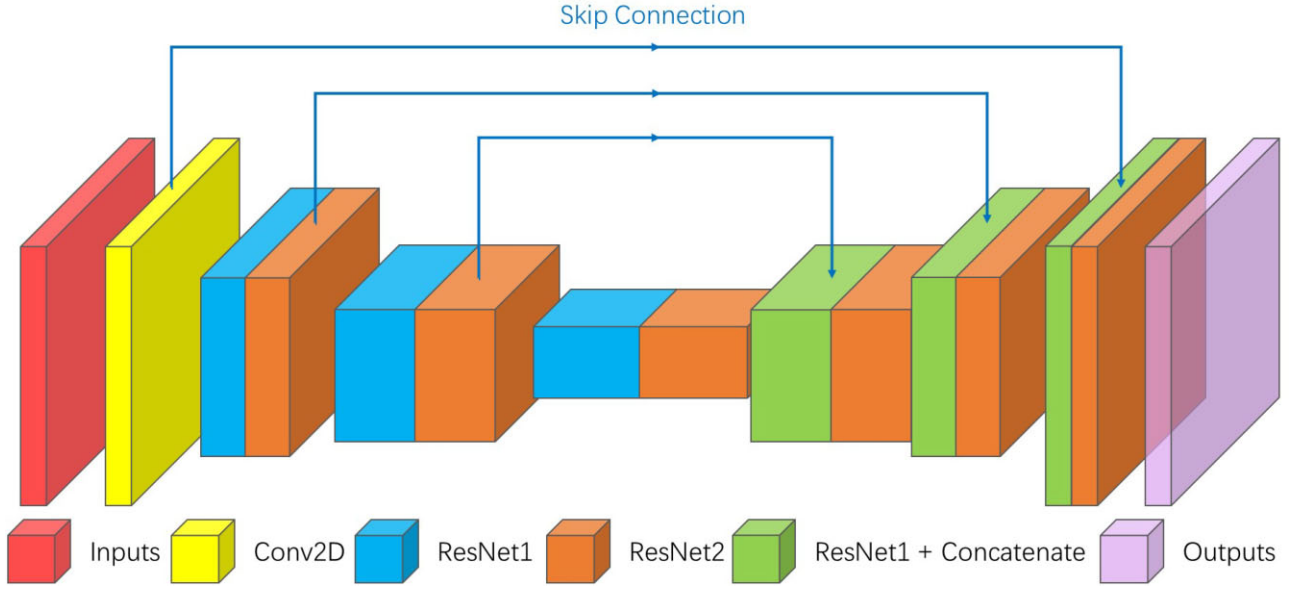
The inputs are firstly convolved by a same-padding convolutional layer with 64 kernels of size 3, obtaining a feature map with the same spatial size as inputs. After convolution, we use LeakyReLU as our activation function. Next we stack five ResNet1 and ResNet2 blocks crossly. The feature maps are downsampled by 2 every time they pass ResNet1, and ResNet2 is employed to increase the learning ability by stacking more convolutional layers. Finally, the inputs are encoded in 3d feature matrix of  $4 \times 4 \times 1024$ . The decoding path is almost mirror of encoding path. We upsample the feature maps by ResNet1 with UpSampling, concatenate with maps obtained in encoding path, and further process them with ResNet2. After upsampling five times, maps reach the same spatial dimension as inputs. We use a convolutional layer to scale the channels to 51 to obtain the outputs with the same size of inputs. Finally we apply square function to the outputs to restrict positive maps since the CO signals are all positive. The schematic diagrams of ResUNet and ResNet blocks are displayed in Figs 6 and 7, respectively. The details of our ResUNet architecture can be found in Table 2.

#### 3.2 Training

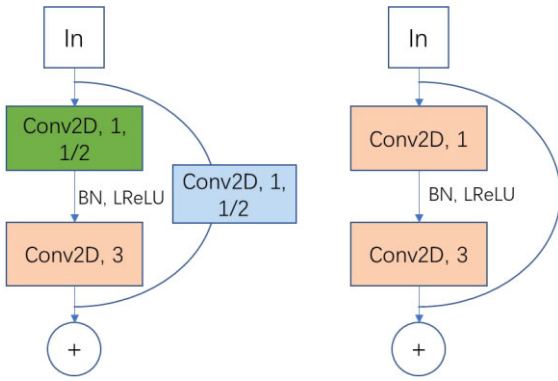
Before training, we need to split the mock data and generate the training and testing sets. We randomly select 130 of the 162 mock CO signal intensity maps and 70 of the 96 total foreground maps,

<sup>1</sup><https://keras.io>

<sup>2</sup><https://www.tensorflow.org/>



**Figure 6.** Schematic diagram of ResUNet. The ResNet1 and ResNet2 block are built crossly. The skip connection concatenates features learned in encoding path with the ones in decoding path.



**Figure 7.** ResNet1 (left) and ResNet2 (right) used in our work. ResNet1 downsamples the spatial dimension of feature maps by 2 using strided convolution, and the inputs are also downsampled and added to the output of the main convolutional layer. ResNet2 will keep the spatial dimension. BatchNormalization and LeakyReLU are employed to reduce overfitting and apply non-linearity after convolution respectively.

with beam effect included, and combine them together to obtain 9100 maps as the training set. The rest of them are used to create 832 mock observational maps as the testing set. Note that the number of the training and testing data are selected randomly here. As a test, we also try to reduce the number of the training data, and find that at least 2000 training data are needed to obtain a similar result.

Our inputs and targets for UNet are beam-convolved CO intensity maps contaminated by foregrounds and CO signal intensity maps without beam effect, respectively, and they are all data cubes with size  $128 \times 128 \times 51$ . Before feeding the UNet, these inputs are preprocessed by PCA for preliminary foreground removal and then our UNet tries to accurately recover the outputs (i.e. CO signal maps without beam effect) by removing residual foreground contamination and eliminating certain effects after PCA procedure like signal loss. We also discuss the result when inputs are not preprocessed by PCA in Section 4. We employ LogCosh as our loss function as recommended in previous studies (Makinen et al. 2021; Ni et al.

**Table 2.** Details of ResUNet architecture.

Layers	Output status <sup>a</sup>	Number of params. <sup>b</sup>
Input	(128, 128, 51)	0
Conv2d	(128, 128, 64)	29,440
LeakyReLU ①	(128, 128, 64)	0
ResBlock ② <sup>c</sup>	(64, 64, 128)	130,656
ResBlock ③	(32, 32, 256)	519,360
ResBlock ④	(16, 16, 512)	2,070,912
ResBlock ⑤	(8, 8, 512)	8,270,592
ResBlock ⑥	(4, 4, 1024)	8,928,000
ResBlock + ⑤ <sup>d</sup>	(8, 8, 1536)	11,290,112
ResBlock + ④	(16, 16, 1024)	9,192,448
ResBlock + ③	(32, 32, 512)	2,434,560
ResBlock + ②	(64, 64, 256)	611,072
ResBlock + ①	(128, 128, 128)	153,984
Conv2d	(128, 128, 51)	58,803
Square	(128, 128, 51)	0

Notes. <sup>a</sup>Format: (dimension, dimension, channel).

<sup>b</sup>Total number of parameters: 43 689 939.

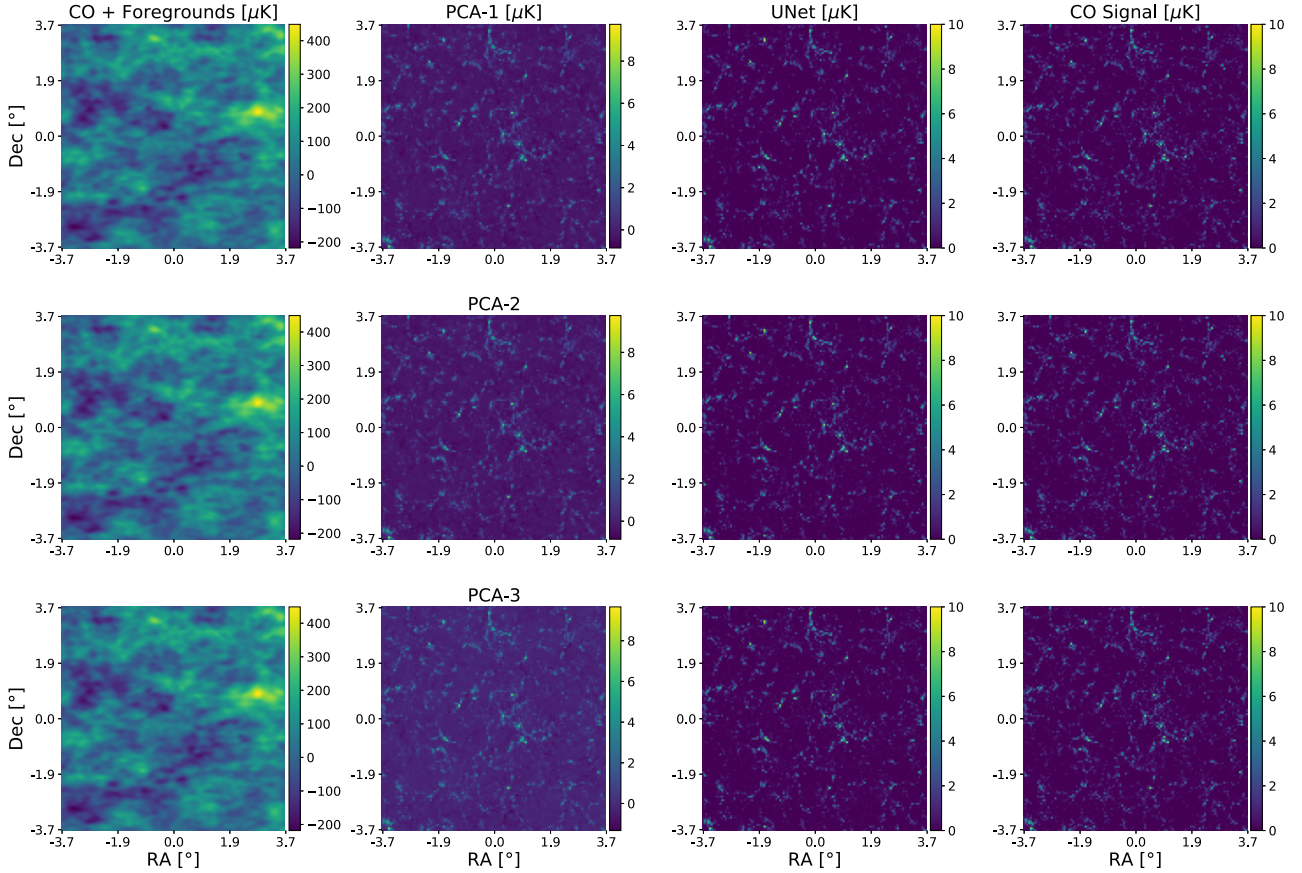
<sup>c</sup>ResNet1 + ResNet2.

<sup>d</sup>ResNet1 + concatenation + ResNet2.

2022). LogCosh calculates log hyperbolic cosine of deviations and is similar to the absolute function but with a smooth gradient around 0. Adam optimizer (Kingma & Ba 2014) with initial learning rate  $1 \times 10^{-4}$  is employed to optimize our network, and this learning rate can be adaptively adjusted during the training process. The batch size is 8, considering the memory limit of Graphical Processing Unit.

For low CO signal case, the amplitude of our CO signal maps are approximately 100 times lower than that in the fiducial model. To avoid extremely small gradients which can slow down the training, we scale both the PCA preprocessed inputs and targets by 100 times larger and then scale them back after training. Note that this scaling does not affect the results but only the training speed. Then the





**Figure 8.** The CO foreground contaminated, PCA preprocessed, UNet generated, and signal (or target) intensity maps for our fiducial CO model at 57.6 GHz (slice at  $z \sim 1$ ). The first, second, and third rows denote the cases of removing the first one (PCA-1), two (PCA-2), and three (PCA-3) largest modes in the PCA procedure, respectively. The maps are truncated at 10  $\mu\text{K}$  for better comparison. We can see that our UNet can excellently recover the web structure of CO signal map, no matter how many modes are subtracted in PCA.

training process is fast and can reach steady losses after about 15 epochs, and we save our UNet model with the lowest loss value.

#### 4 RESULT AND DISCUSSION

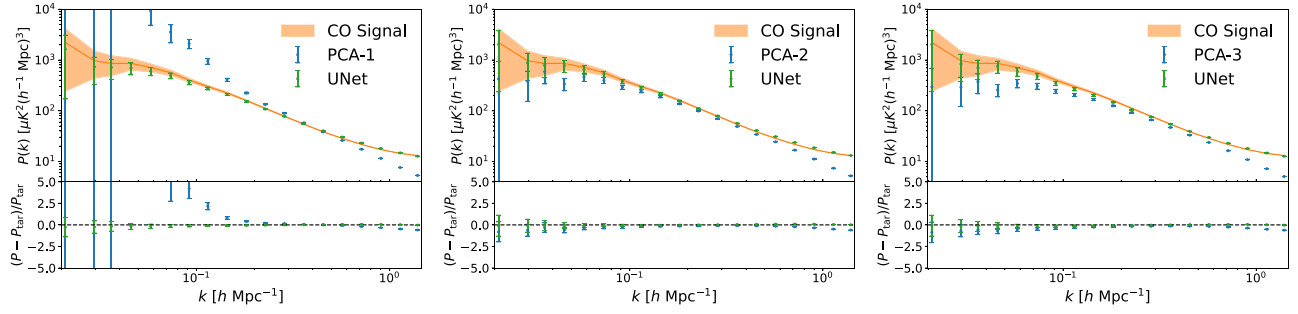
After the training process, we feed the testing data to our UNet model for testing the network. Similar to the training process, we first preprocess the testing data with PCA for subtracting the foregrounds. Fig. 8 shows the relevant maps at 57.6 GHz (slice of  $z \simeq 1$ ) for our CO fiducial model mentioned in Section 2.1. The maps are truncated at 10  $\mu\text{K}$  for better comparison. We can see that our UNet can effectively recover the CO signal map no matter how many modes are subtracted in the PCA procedure.

For further comparison and assessment, we calculate and show the CO power spectra of PCA preprocessed, UNet generated and signal or target maps in Fig. 9. The relative deviation  $(P - P_{\text{tar}})/P_{\text{tar}}$  are also shown in the lower panels. The errors of the power spectra are calculated by root-mean-square (rms) of all the power spectra. We notice that the PCA-1, i.e. the case of removing the first largest mode, cannot effectively remove the foregrounds, resulting in large deviations compared to CO signal power spectrum. Then the first two or three largest modes need to be removed (i.e. PCA-2 and PCA-3) for correctly removing the foregrounds. However, we find that the signal loss effect still can be clearly recognized at large and small scales induced by PCA and beam effect as expected. As we mentioned, although the PCA blind subtraction can remove most

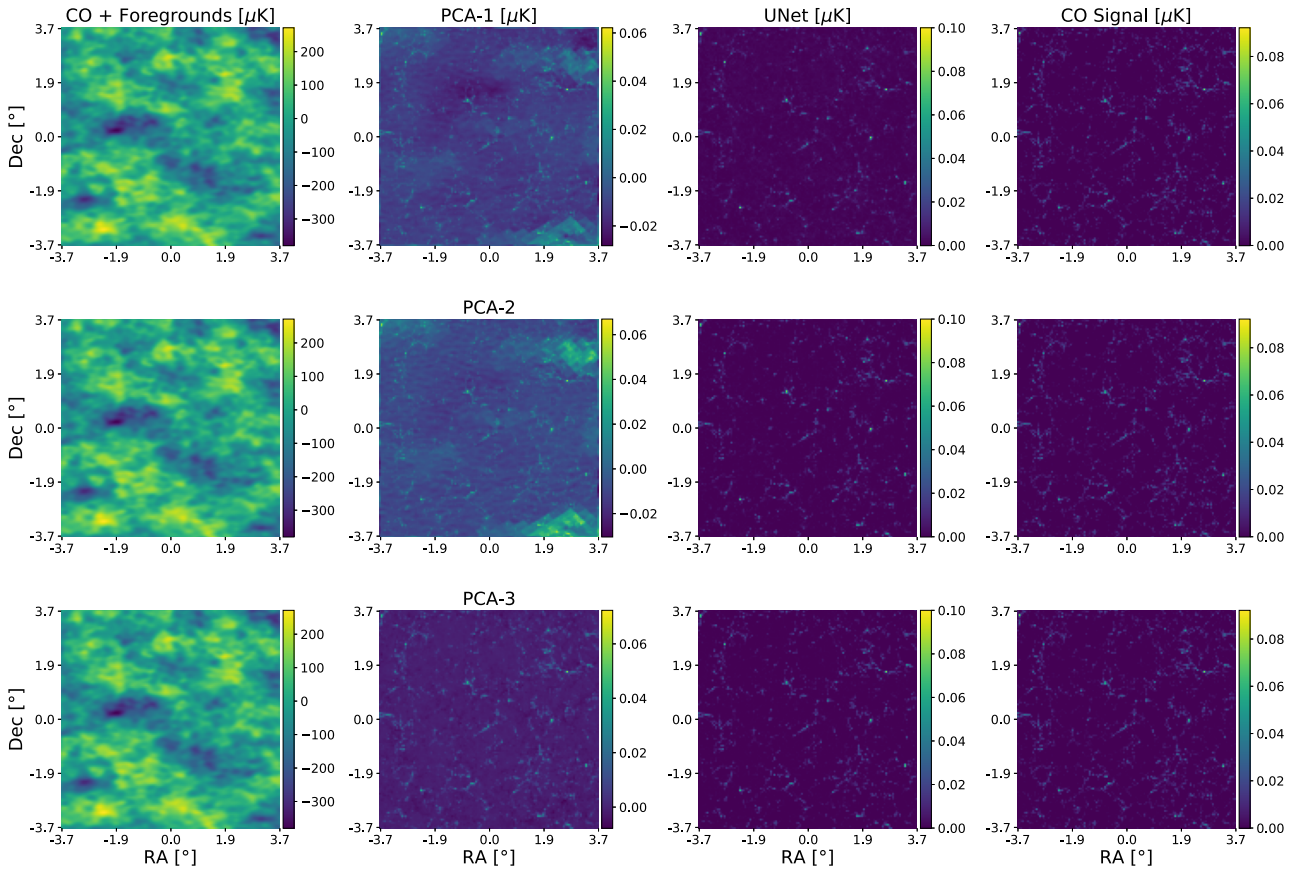
continuum foregrounds from CO intensity maps, it also can corrupt the information at large scale, making deviations in power spectra compared to the signal. Usually, the more modes PCA subtracts, the more signal loses at large scales, and the beam effect further complicates this situation that information at both large and small scales will be corrupted because of the correlation of beam and frequencies. On the other hand, in Fig. 9, we can find that our UNet can excellently recover the power spectra regardless the number of modes subtracted by PCA, even in the non-linear regime at  $k \gtrsim 1 \text{ Mpc}^{-1}h$ . The deviations at large and small scale induced by PCA and beam effect can be corrected, and all derived data points are consistent with the CO signal power spectrum within  $1\sigma$  confidence level.

For the low CO signal model with CO signal approximately 100 times lower (see Fig. 1), we find that more modes of PCA are required to clean the foregrounds. Similar to Figs 8 and 9, we show the relevant maps and power spectra of the low CO signal model at 57.6 GHz in Figs 10 and Fig. 11, respectively. The maps are truncated at 0.1  $\mu\text{K}$  for better comparison. As can be seen, the PCA-1 and PCA-2 cannot effectively remove the foregrounds in this case, and the web structures from the CO signal are still deeply buried in foreground residuals. This is also indicated in Fig. 11, that there is large discrepancy between the power spectra of PCA-1 or PCA-2 (blue data points) and the CO signal (orange curves). Our UNet manages to extract the signal from the PCA-1 map, and some deviation still exists in the derived





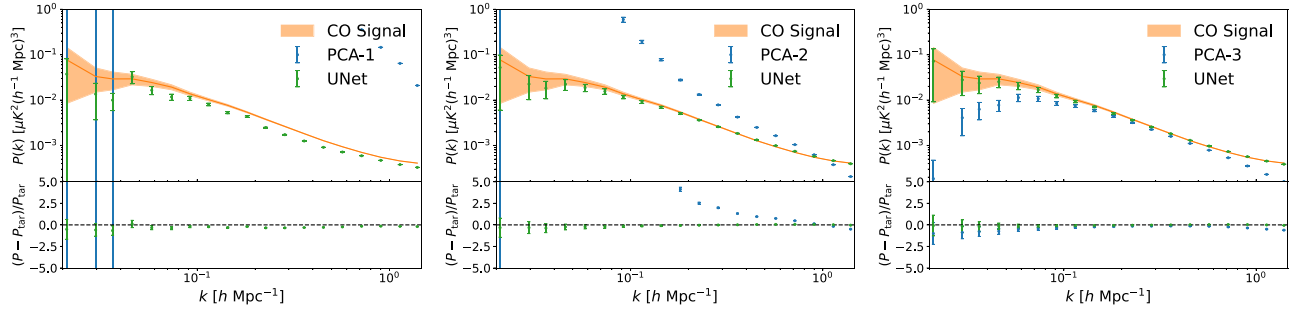
**Figure 9.** The CO power spectra of the PCA preprocessed, UNet generated and signal or target maps for our fiducial CO model. The errorbars are calculated by root-mean-square of all power spectra of the testing maps. The relative difference of the power spectra  $(P - P_{\text{tar}})/P_{\text{tar}}$  is also shown in the lower panels. We notice that PCA-1 cannot completely remove the foregrounds, resulting in large deviations compared to the signal power spectrum. PCA-2 and PCA-3 can effectively clean the foregrounds but with a loss of information at both large and small (due to beam effect) scales. Our UNet can recover the power spectra excellently by correcting the residual foregrounds and beam effect, regardless of the number of modes PCA subtracts, and the derived power spectra are consistent with the CO signal power spectrum within  $1\sigma$  confidence level.



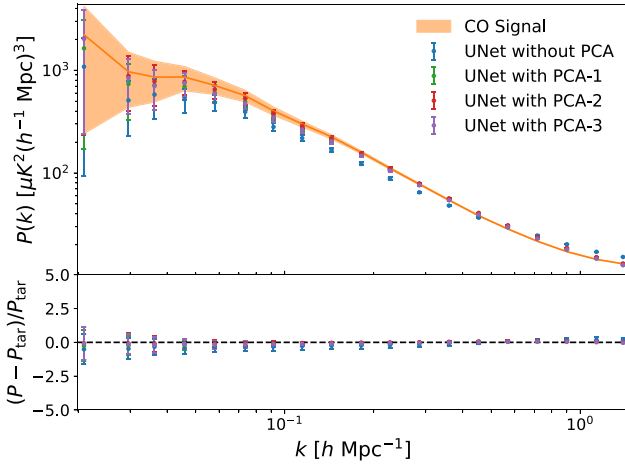
**Figure 10.** Same as Fig. 8 but for our low CO signal model with the signal approximately 100 times lower compared to the fiducial one. We can see that more modes need to be removed in the PCA foreground removal, and our UNet can still correctly recover the CO signal map in different PCA mode-removing cases.

power spectrum (green data points) compared to the CO signal power spectrum. For the PCA-2 maps, the UNet can excellently recover the signals both in the derived maps and power spectrum, which is consistent with the CO signal power spectrum in  $1\sigma$ . In the PCA-3 case, although the result after the mode subtraction procedure is much better than the PCA-1 and PCA-2 cases, the signal loss effect is still strong at large scales. Here our UNet can perfectly reconstruct the CO signal in the derived map, and recover the signal power spectra even better than the result of the PCA-2 case.

We also investigate the cases without PCA preprocessing. For our fiducial CO model, we find that the UNet can generate quite similar CO maps with more training epochs compared to the result with PCA preprocessing. The derived CO power spectra with and without PCA preprocessing are shown in Fig. 12. We notice that the power spectrum derived from the network without PCA preprocessing (blue data points) has more than  $1\sigma$ – $2\sigma$  deviations compared to the CO signal power spectrum (orange curves) at all scales. On the other hand, the results with PCA preprocessing can match the CO signal power spectrum in  $1\sigma$ . For the case of our low CO signal model,



**Figure 11.** Same as Fig. 9 but for our low CO signal model. There are large discrepancies between the PCA-1 or PCA-2 (blue data points) and the CO signal power spectra (orange curve) in this case. Our UNet cannot accurately obtain the CO power spectrum in the PCA-1 case, and some deviations still exist. However, the CO signal power spectra can be reconstructed within  $1\sigma$  error in the PCA-2 and PCA-3 cases, and the signal loss effect also can be effectively corrected.



**Figure 12.** Comparison between the power spectra of the maps generated by UNet with (green, red, and purple data points) and without (blue data points) PCA preprocessing for the case of our fiducial CO model. We notice that the power spectrum reconstructed by the network suffers deviations at all scales without PCA preprocessing, and it is even worse in the low CO signal model. Hence, PCA is a necessary procedure in our neural network analysis.

the CO signals is completely dominated by foregrounds, and our UNet cannot derive correct CO map and the power spectrum without PCA preprocessing. Therefore, PCA preprocessing is a necessary procedure in our analysis using the neural network.

In addition, we try to use the neural network to extract the CO signal considering significant instrumental noise. Our assumed instrumental parameters are described in Section 2.3, and we generate the Gaussian noise maps and add them into the total CO maps. We find that the amplitude of this noise power spectrum is about  $2600 \mu\text{K}^2(h^{-1}\text{Mpc})^3$ , which is higher than the signal power spectrum at all scales for our fiducial CO model (see the power spectra of Fig. 3). In this case, we cannot obtain correct CO signal maps and power spectra using the neural network, since the CO signal is still totally dominated by the Gaussian noise after the PCA preprocessing. We find that only if the Gaussian noise can be  $1/10$  or lower compared to the current level, our network can derive the correct results. This indicates that low instrumental noise is required in deep learning analysis for obtaining correct line signal maps and power spectra, which is also the requirement in traditional signal extraction methods. Since the instrumental noise can be reduced by increasing total observational time, adding more instrumental feeds, or decreasing the system temperature, this problem is fortunately solvable in the intensity mapping surveys. Therefore, deep learning could be an

effective method in map foreground removal of CO and other LIM with low instrumental noise, that also has been proven in previous studies (e.g. Moriwaki et al. 2021; Ni et al. 2022).

## 5 SUMMARY AND CONCLUSION

In this work, we investigate the foreground removal of CO intensity mapping at  $z \simeq 1$  by deep learning methods. The CO intensity maps are simulated based on N-body simulations and the relation of CO luminosity and halo mass  $L_{\text{CO}}-M_{\text{h}}$  from CO and IR observations. We derive the upper, median, and lower relations of the current  $L_{\text{CO}}-L_{\text{IR}}$  measurements, and set the median and lower cases as our fiducial and low CO models, respectively. The CO intensity is approximately 100 times smaller for the low CO model than the fiducial one.

We also include foreground contamination, instrumental noise, and beam effect in the mock data. The foreground components we consider are thermal dust, spinning dust, synchrotron, free-free emissions, and CMB anisotropy at frequencies of CO signals generated by PYSM3 PYTHON package. The instrumental effects, such as instrumental noise and beam effects of radio telescope are also considered. The instrumental noise is induced by system temperature and it is assumed to be Gaussian-like. The beam model we investigate is Bessel or Jinc beam, which has the largest sidelobes among models studied in literature. CO maps and foregrounds are all convolved with this beam model to simulate the beam effects. We find the traditional method such as PCA will lead to deviations of power spectra by signal loss effect, and considering the beam effect, it becomes even worse. Then we try to use deep learning for accurately recovering the CO signal map with correct line power spectrum.

Our deep learning model is based on ResUNet, which combines the image-to-image generation UNet with ResNet architecture. Since PCA can effectively remove most foregrounds, we preprocess the contaminated CO maps with PCA before feeding our UNet. We find that for our fiducial CO model, removing the first two or more largest modes (i.e. PCA-2 and PCA-3 in our analysis) in PCA can effectively subtract the foregrounds, but resulting in some deviations in power spectra due to signal loss effect. Our UNet can accurately recover the CO signal maps with correct line power spectra, no matter how many modes are subtracted. For the low CO signal model, at least the first three largest modes (i.e. PCA-3) need to be removed in the PCA procedure for effectively subtracting the foregrounds but with large signal loss. However, we find that our UNet can obtain the correct CO signal maps and recover the line power spectra in both PCA-2 and PCA-3 cases. We also investigate the effects of PCA preprocessing and instrumental noise on the results from deep learning. We find that PCA preprocessing for foreground removal is necessary, especially

for low signal case, and the instrumental noise also need to be limited in relatively low levels for accurate signal extraction. Besides, our method also can be easily applied in other LIM like neutral hydrogen 21-cm emission. Therefore, the deep learning method could be an effective way for foreground removal in CO and other intensity mapping, and can obtain accurate line power spectrum for exploring the cosmic large-scale structure, star formation, and galaxy evolution.

## ACKNOWLEDGEMENTS

XCZ and YG acknowledge the support of 2020SKA0110402, MOST-2018YFE0120800, National Key R&D Program of China No. 2022YFF0503404, NSFC-11822305, NSFC-11773031, and NSFC-11633004. XLC acknowledges the support of the National Natural Science Foundation of China through grant No. 11473044, 11973047, and the Chinese Academy of Science grants QYZDJ-SSW-SLH017, XDB 23040100, XDA15020200. This work is also supported by the science research grants from the China Manned Space Project with No. CMS-CSST-2021-B01 and CMS-CSST-2021-A01. This work was funded by the National Natural Science Foundation of China (NSFC) under No. 11080922.

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

## REFERENCES

- Abdi H., Williams L. J., 2010, *Wiley Interdiscip. Rev.: Comput. Stat.*, 2, 433
- Alonso D., Bull P., Ferreira P. G., Santos M. G., 2014, *MNRAS*, 447, 400
- Aragon-Calvo M. A., 2019, *MNRAS*, 484, 5771
- Aravena M. et al., 2012, *MNRAS*, 426, 258
- Asad K. et al., 2021, *MNRAS*, 502, 2970
- Bernal J. L., Kovetz E. D., 2022, *AAPR*, 30, 5
- Breyse P. C. et al., 2022, *ApJ*, 933, 188
- CCAT-Prime Collaboration, 2023, *ApJS*, 264, 7
- Carilli C. L., 2011, *ApJ*, 730, L30
- Carilli C. L., Walter F., 2013, *ARA&A*, 51, 105
- Chung D. T. et al., 2022, *ApJ*, 933, 186
- Cleary K. A. et al., 2022, *ApJ*, 933, 182
- Condon J. J., Ransom S. M., 2016, *Essential Radio Astronomy*, Princeton University Press, Princeton, NJ
- Daddi E. et al., 2010, *ApJ*, 714, L118
- Deng F., Gong Y., Wang Y., Dong S., Cao Y., Chen X., 2022, *MNRAS*, 515, 5894
- Dessauges-Zavadsky M. et al., 2015, *A&A*, 577, A50
- Diakogiannis F. I., Waldner F., Caccetta P., Wu C., 2020, *ISPRS J. Photogramm. Remote Sens.*, 162, 94
- Feder R. M., Berger P., Stein G., 2020, *Phys. Rev. D*, 102, 103504
- Fonseca J., Silva M. B., Santos M. G., Cooray A., 2017, *MNRAS*, 464, 1948
- Fukushima K., Miyake S., 1982, *Pattern Recognit.*, 15, 455
- Gong Y., Cooray A., Silva M. B., Santos M. G., Lubin P., 2011, *ApJ*, 728, L46
- Gong Y., Cooray A., Silva M., Santos M. G., Bock J., Bradford C. M., Zemcov M., 2012, *ApJ*, 745, 49
- Gong Y., Cooray A., Santos M. G., 2013, *ApJ*, 768, 130
- Gong Y., Silva M., Cooray A., Santos M. G., 2014, *ApJ*, 785, 72
- Gong Y., Cooray A., Silva M. B., Zemcov M., Feng C., Santos M. G., Dore O., Chen X., 2017, *ApJ*, 835, 273
- Gong Y., Chen X., Cooray A., 2020, *ApJ*, 894, 152
- Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014, preprint ([arXiv:1406.2661](https://arxiv.org/abs/1406.2661))
- Greve T. R. et al., 2014, *ApJ*, 794, 142
- Grupponi C. et al., 2013, *MNRAS*, 432, 23
- Harper S. E., Dickinson C., Battye R. A., Roychowdhury S., Browne I. W. A., Ma Y.-Z., Olivari L. C., Chen T., 2018, *MNRAS*, 478, 2416
- He K., Zhang X., Ren S., Sun J., 2015, preprint ([arXiv:1512.03385](https://arxiv.org/abs/1512.03385))
- Hyvärinen A., Oja E., 2000, *Neural Netw.*, 13, 411
- Ihle H. T. et al., 2022, *ApJ*, 933, 185
- Ioffe S., Szegedy C., 2015, preprint ([arXiv:1502.03167](https://arxiv.org/abs/1502.03167))
- Isola P., Zhu J.-Y., Zhou T., Efros A. A., 2016, preprint ([arXiv:1611.07004](https://arxiv.org/abs/1611.07004))
- Karkare K. S. et al., 2022, *J. Low Temp. Phys.*
- Kasmanoff N., Villaescusa-Navarro F., Tinker J., Ho S., 2020, preprint ([arXiv:2012.00186](https://arxiv.org/abs/2012.00186))
- Keating G. K. et al., 2015, *ApJ*, 814, 140
- Keating G. K., Marrone D. P., Bower G. C., Leitch E., Carlstrom J. E., DeBoer D. R., 2016, *ApJ*, 830, 34
- Keating G. K., Marrone D. P., Bower G. C., Keenan R. P., 2020, *ApJ*, 901, 141
- Kingma D. P., Ba J., 2014, preprint ([arXiv:1412.6980](https://arxiv.org/abs/1412.6980))
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340
- Kodi Ramanah D., Charnock T., Villaescusa-Navarro F., Wandelt B. D., 2020, *MNRAS*, 495, 4227
- Kovetz E. D. et al., 2017, preprint ([arXiv:1709.09066](https://arxiv.org/abs/1709.09066))
- Li Y., Ni Y., Croft R. A. C., Di Matteo T., Bird S., Feng Y., 2021, *Proc. Natl. Acad. Sci.*, 118, e2022038118
- Lidz A., Taylor J., 2016, *ApJ*, 825, 143
- Lidz A., Furlanetto S. R., Oh S. P., Aguirre J., Chang T.-C., Doré O., Pritchard J. R., 2011, *ApJ*, 741, 70
- Maas A. L. et al., 2013, *Proc. ICML*, 30, 3
- Magnelli B., Elbaz D., Chary R., Dickinson M., Borgne D., Frayer D., Willmer C., 2009, *A&A*, 496
- Magnelli B. et al., 2014, *A&A*, 561, A86
- Makinen T. L., Lancaster L., Villaescusa-Navarro F., Melchior P., Ho S., Perreault-Levasseur L., Spergel D. N., 2021, *J. Cosmology Astropart. Phys.*, 2021, 081
- Matshawule S. D., Spinelli M., Santos M. G., Ngobese S., 2021, *MNRAS*, 506, 5075
- Mirza M., Osindero S., 2014, preprint ([arXiv:1411.1784](https://arxiv.org/abs/1411.1784))
- Moriwaki K., Yoshida N., 2021, *ApJ*, 923, L7
- Moriwaki K., Shirasaki M., Yoshida N., 2021, *ApJ*, 906, L1
- Ni S., Li Y., Gao L.-Y., Zhang X., 2022, *ApJ*, 934, 83
- Odena A., Dumoulin V., Olah C., 2016, *Distill*, 1, e3
- Padmanabhan H., 2018, *MNRAS*, 475, 1477
- Perraudin N., Srivastava A., Lucchi A., Kacprzak T., Hofmann T., Réfrégier A., 2019, *Comput. Astrophys. Cosmol.*, 6, 5
- Planck Collaboration, 2014, *A&A*, 571, A16
- Planck Collaboration, 2016, *A&A*, 594, A10
- Pullen A. R., Doré O., Bock J., 2014, *ApJ*, 786, 111
- Rodríguez A. C., Kacprzak T., Lucchi A., Amara A., Sgier R., Fluri J., Hofmann T., Réfrégier A., 2018, *Comput. Astrophys. Cosmol.*, 5, 4
- Ronneberger O., Fischer P., Brox T., 2015, preprint ([arXiv:1505.04597](https://arxiv.org/abs/1505.04597))
- Rumelhart D. E., Hinton G. E., Williams R. J., 1986, *Nature*, 323, 533
- Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
- Silva M. B., Santos M. G., Gong Y., Cooray A., Bock J., 2013, *ApJ*, 763, 132
- Silva M., Santos M. G., Cooray A., Gong Y., 2015, *ApJ*, 806, 209
- Springel V. et al., 2005, *Nature*, 435, 629
- Sun S. et al., 2022, *Res. Astron. Astrophys.*, 22, 065020
- Thorne B., Dunkley J., Alonso D., Naess S., 2017, *MNRAS*, 469, 2821
- Uzgil B. D., Aguirre J. E., Bradford C. M., Lidz A., 2014, *ApJ*, 793, 116
- Visbal E., Loeb A., 2010, *J. Cosmol. Astropart. Phys.*, 2010, 016
- Visbal E., Trac H., Loeb A., 2011, *J. Cosmology Astropart. Phys.*, 2011, 010
- Wilson T. L., Rohlfs K., Hüttemeister S., 2013, *Tools Radio Astron.*
- Wold S., Esbensen K., Geladi P., 1987, *Chemometr. Intell. Lab. Syst.*, 2, 37
- Yue B., Ferrara A., Pallottini A., Gallerani S., Vallini L., 2015, *MNRAS*, 450, 3829
- Zhang X., Wang Y., Zhang W., Sun Y., He S., Contardo G., Villaescusa-Navarro F., Ho S., 2019, preprint ([arXiv:1902.05965](https://arxiv.org/abs/1902.05965))

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