



Eliminating Primary Beam Effect in Foreground Subtraction of Neutral Hydrogen Intensity Mapping Survey with Deep Learning

Shulei Ni (倪书磊)¹ , Yichao Li (李毅超)¹ , Li-Yang Gao (高立扬)¹ , and Xin Zhang (张鑫)^{1,2,3}

¹ Department of Physics, College of Sciences, Northeastern University, Shenyang 110819, People's Republic of China; zhangxin@mail.neu.edu.cn

² Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang 110819, People's Republic of China

³ Key Laboratory of Data Analytics and Optimization for Smart Industry (Ministry of Education), Northeastern University, People's Republic of China

Received 2022 April 12; revised 2022 June 15; accepted 2022 June 16; published 2022 July 27

Abstract

In neutral hydrogen (HI) intensity mapping (IM) survey, foreground contamination on cosmological signal is extremely severe, and systematic effects caused by radio telescopes further aggravate the difficulties in subtracting foreground. We investigate whether the deep-learning method, the 3D U-Net algorithm, can play a crucial role in foreground subtraction when considering the systematic effect caused by the telescope's primary beam. We consider two beam models, i.e., the Gaussian beam and Cosine beam models. The traditional principal component analysis (PCA) method is employed as a preprocessing step for the U-Net method to reduce the map dynamic range. We find that in the case of the Gaussian beam, the PCA method can effectively clean the foreground. However, the PCA method cannot handle the systematic effect induced by the Cosine beam, and the additional U-Net method can improve the result significantly. To show how well the PCA and U-Net methods can recover the HI signal, we also derive the HI angular power spectrum and HI 2D power spectrum after performing foreground subtraction. It is found that in the case of Gaussian beam, the concordance with the original HI map using U-Net is better than that using PCA by 27.4%, and in the case of Cosine beam, the concordance using U-Net is better than that using PCA by 144.8%. Therefore, the U-Net-based foreground subtraction can efficiently eliminate the telescope primary beam effect and shed new light on recovering HI power spectrum for future HI IM experiments.

Unified Astronomy Thesaurus concepts: H I line emission (690); Sky surveys (1464); Radio telescopes (1360); Neural networks (1933); Large-scale structure of the universe (902); Observational cosmology (1146)

1. Introduction

The cosmic large-scale structure (LSS) carries crucial information about the evolutionary history of the universe. The neutral hydrogen (HI) intensity mapping (IM) has been proposed as a promising technique for the LSS survey (e.g., Battye et al. 2004; McQuinn et al. 2006; Pritchard & Loeb 2008). HI IM relies on observing integrated HI 21 cm emission of many galaxies in large voxels and can be quickly carried out to cover a very large survey volume, which is ideal for cosmological studies (Chang et al. 2008; Loeb & Wyithe 2008; Mao et al. 2008; Wyithe & Loeb 2008; Wyithe et al. 2008; Bagla et al. 2010; Seo et al. 2010; Lidz et al. 2011; Ansari et al. 2012; Battye et al. 2013).

Chang et al. (2010) for the first time explored the HI IM technique with Green Bank Telescope (GBT) by measuring the cross-correlation function between an HI IM survey and an optical galaxy survey. So far, several other explorations of HI IM have been done by measuring the cross-correlation power spectrum between the HI IM survey and the optical galaxy survey, e.g., using the GBT or Parkes telescope (Masui et al. 2013; Wolz et al. 2017, 2022; Anderson et al. 2018), but the HI IM autocorrelation power spectrum remains undetected (Switzer et al. 2013). Besides, a number of specially designed radio telescopes or telescope arrays targeting at HI IM LSS survey, such as Tianlai (Chen 2012), CHIME (Bandura et al. 2014), BINGO (Peel et al. 2019), FAST (Nan et al. 2011), HIRAX (Newburgh et al. 2016), are either under construction

or collecting data (Li et al. 2020; Abdalla et al. 2021; Wuensche et al. 2021; Wu et al. 2021; Amiri et al. 2022). The HI IM is proposed as the major technique for cosmological exploration with the Square Kilometre Array (SKA; Bull et al. 2015; Santos et al. 2015; Bacon et al. 2020). Moreover, with the newly built MeerKAT array, it is also proposed to have an HI IM survey in single-dish mode (Santos et al. 2017; Li et al. 2021; Wang et al. 2021).

However, there are a couple of technical challenges for HI IM data analysis. The major challenge comes from extracting the weak HI brightness fluctuation from bright foreground radio emission in the same frequency range, i.e., the synchrotron or free-free emission of the Galaxy and nearby radio galaxy point sources. It is shown that the intensity of foreground emission, especially synchrotron emission from the Galaxy, is about 5 orders of magnitude stronger than the HI signal. Additionally, the free-free emission of the Galaxy, as well as the extragalactic point sources, can be about 3 orders of magnitude stronger than the HI signal.

Considerable work has been done to subtract foreground, proposing some methods for deducting HI foreground. There are advantages and disadvantages to each of these methods, which are mainly divided into two categories: nonblind and blind algorithms. Nonblind algorithms mainly rely on the assumption of a smooth foreground spectrum model by fitting a few low-order polynomials to extract the smooth foreground components. However, the complex systematic effects of the actual data can seriously undermine the assumption of the smooth foreground. The nonblind algorithm performs poorly in the actual data analysis, and it is difficult to eliminate the foreground interference. Blind algorithms are commonly used foreground subtraction methods in the analysis of HI IM real



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

data, mainly including principal component analysis (PCA; de Oliveira-Costa et al. 2008), fast independent component analysis (fastICA; Hothi et al. 2020), correlated component analysis (CCA; Bonaldi & Brown 2015), Gaussian process (Mertens et al. 2018; Soares et al. 2022), etc. The blind algorithms can cope with the simple systematic effects and improve the accuracy of foreground subtraction to a certain extent. But they usually cause serious signal loss and have some randomness and instability when dealing with complex beams.

These foreground subtraction methods all rely on the assumption of smoothness of the foreground emission spectrum. The instrument's systematic effects are entangled with the foreground, leading to unsatisfactory foreground subtraction by these algorithms. Therefore, the development of new algorithms for HI foreground subtraction is imminent. Deep learning forms more abstract high-level features through the underlying features to discover how the instrumental effects impact the signal. Therefore, we wish to use the deep-learning approach to deal with this challenge.

Deep learning has unique advantages in feature recognition of complex systems (Makinen et al. 2021; Villanueva-Domingo & Villaescusa-Navarro 2021). In the work of Makinen et al. (2021), the foreground subtraction with U-Net, one of the convolution network deep-learning methods, has been generally discussed. The main objective of this study is to apply deep-learning algorithms in foreground subtraction to eliminate instrumental effects and improve foreground subtraction. We plan to investigate the application of deep learning in eliminating instrumental effects in foreground subtraction through a series of simulations. The instrumental effects considered in this work mainly include the frequency dependence of the antenna beam and the sidelobe leakage. At present, the analysis is based primarily on the antenna beam model of MeerKAT (Matshawule et al. 2021). We use simulation software to simulate the instrumental effects from the HI observed signal, foreground, and the introduction of beam models of different complexities. These simulated data are used as a training set, test set, and validation set, respectively, to study applying the U-Net deep-learning method to eliminate the instrumental effects and optimize the foreground reduction.

The rest of the paper is organized as follows: In Section 2 we introduce our simulation. In Section 3 we consider the frequency dependence of the antenna beam and the sidelobe leakage. In Section 4 we explain how PCA and deep learning work to remove the foreground. In Section 5 we show the results in our analysis and make some discussions. We conclude in Section 6.

2. Sky Simulation

In this section, we describe the simulated sky components used in our analysis, including the cosmological HI signal (Section 2.1), foreground emission (Section 2.2), and white Gaussian noise (Section 2.3). The simulated sky components are generated using the publicly available package, i.e., Cosmological Realizations for Intensity Mapping Experiments (CRIME⁴; Alonso et al. 2014). We briefly introduce the simulation method. The simulated sky components are all given in full sky using the HEALPix pixelization strategy (Górski et al. 2005). Given the low angular resolution requirement, we use the map with $N_{\text{side}} = 256$ corresponding to an angular resolution of 13.73 arcmin, which is good enough for the HI IM experiments in the near future. Meanwhile, we

consider the frequency range spanning between 900 and 1050 MHz.

2.1. Cosmological Signal

The mean brightness temperature of HI 21 cm emission from redshift z is

$$\bar{T}_{\text{HI}}(z) = 190.55 \text{ mK} \frac{\Omega_b h(1+z)^2 x_{\text{HI}}(z)}{\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}}, \quad (1)$$

where $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$ is the dimensionless Hubble constant, x_{HI} is the fraction of the HI mass in the total baryons, and Ω_b , Ω_m , and Ω_Λ are the baryon, total matter, and dark energy density fractions, respectively. The HI brightness temperature fluctuation is expressed as

$$T_{\text{HI}}(z, \hat{n}) = \bar{T}_{\text{HI}}(z)(1 + \delta_{\text{HI}}(z, \hat{n})), \quad (2)$$

where $\delta_{\text{HI}}(z, \hat{n})$ is the HI overdensity field in redshift space. The final simulated maps are generated using uniform frequency slices according to $z+1 = \nu_0/\nu$, where $\nu_0 = 1420.406 \text{ MHz}$ is the HI 21 cm emission rest-frame frequency. The HI fluctuation signal is simulated using the lognormal realization with the power spectrum of (Alonso et al. 2014; Olivari et al. 2018)

$$C_\ell^{\text{HI}}(\nu_1, \nu_2) = \frac{2}{\pi} \int dk k^2 P_{\text{DM}}(k) W_{\ell, \nu_1}(k) W_{\ell, \nu_2}(k), \quad (3)$$

where $P_{\text{DM}}(k)$ is the underlying dark matter power spectrum at redshift $z=0$, and $W_{\ell, \nu}$ is the window function defined as

$$W_{\ell, \nu_i}(k) = \int dz \bar{T}_{\text{HI}}(z) \phi(\nu_i, z) D(z) \times (b(z)j_\ell(k\chi) - f(z)j''_\ell(k\chi)), \quad (4)$$

where j_ℓ is the ℓ -th spherical Bessel function, $\phi(\nu_i, z)$ is the frequency selection function centering at ν_i , $D(z)$ is the growth factor, $f(z)$ is the growth rate and $b(z)$ is the linear bias of the HI density with respect to the underlying dark matter density, i.e., $\delta_{\text{HI}} = b(z)\delta_{\text{DM}}$.

In our simulation, we use the best-fit cosmological parameters from the Planck 2018 Λ CDM result (Aghanim et al. 2020), i.e., $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_b = 0.049$, $\Omega_m = 0.311$, $\Omega_\Lambda = 0.689$, and $\sigma_8 = 0.81$. As an example, one simulated HI map of frequency slice at $\nu = 1000 \text{ MHz}$ is shown in Figure 1(a).

2.2. Foreground Components

The foreground consists of many distinct components that emit via different emission mechanisms. Following the analysis in the literature, we consider four different foreground components in our simulation, i.e., the Galactic synchrotron emission, the Galactic free-free emission, the extragalactic free-free emission, and the extragalactic point sources, which are known as the major contamination of the HI signal.

The foreground is simulated using two different methods according to two different angular distributions, i.e., the anisotropic and isotropic ones. The Galactic synchrotron is produced by the accelerated motion of energetic charged particles dispersed in the Galactic magnetic field. Its angular structure is highly anisotropic, where the brightness temperature grows steeply toward the Galactic plane.

⁴ <http://intensitymapping.physics.ox.ac.uk/CRIME.html>

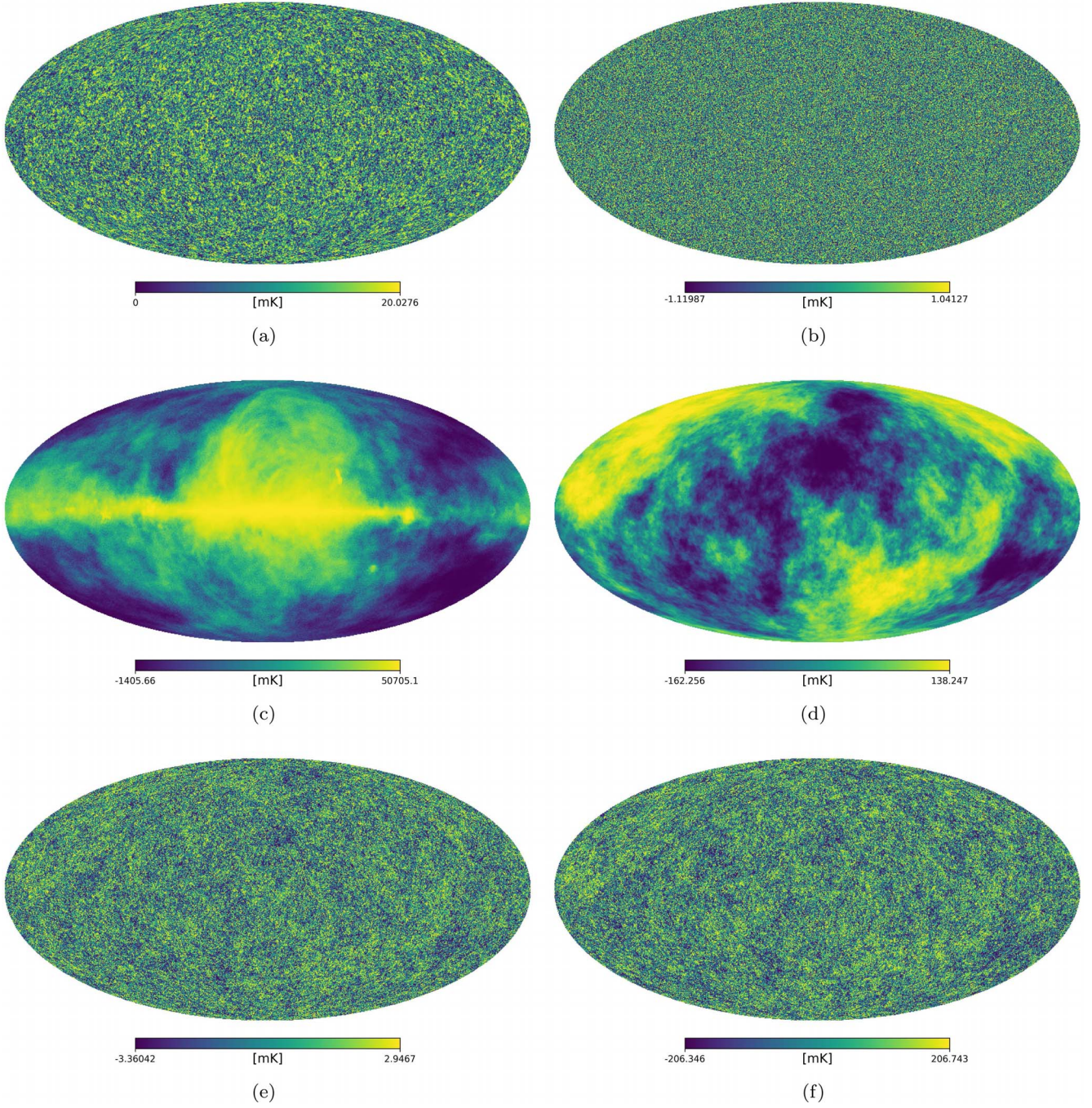


Figure 1. Simulated sky maps from different components. Panels (a)–(f) are the maps of H I, white noise, Galactic synchrotron, Galactic free–free, extragalactic free–free, and extragalactic point source, respectively. The maps are shown using the histogram equalized color mapping to amplify the contrast of the image.

In the simulations, the Galactic synchrotron amplitude template is based on the Haslam map at 408 MHz (Haslam et al. 1982). The Galactic synchrotron emission is simulated by extrapolating the Haslam map’s brightness temperature $T_{\text{Haslam}}(\hat{n})$ to other frequencies with a direction-dependent spectral index. The spectral index, $\gamma(\hat{n})$, is generated with the Planck Sky Model (PSM; Delabrouille et al. 2013). The anisotropic Galactic synchrotron emission is simulated via

$$T_{\text{syn},0}(\nu, \hat{n}) = T_{\text{Haslam}}(\hat{n}) \left(\frac{408 \text{ MHz}}{\nu} \right)^{\gamma(\hat{n})}. \quad (5)$$

Due to the low resolution of the Haslam sky map, the angular structure information is lost at small scales. We assume

that the anisotropy of Galactic synchrotron emission can be neglected at small scales. Therefore, based on the following form of spectrum,

$$C_{\ell}^{\text{fg}}(\nu_1, \nu_2) = A \left(\frac{\ell_{\text{ref}}}{\ell} \right)^{\beta} \left(\frac{\nu_{\text{ref}}^2}{\nu_1 \nu_2} \right)^{\alpha} \exp \left(-\frac{\log^2(\nu_1/\nu_2)}{2\xi^2} \right), \quad (6)$$

we generate the isotropic structure of synchrotron emission using the method implemented by Santos, Cooray, and Knox (SCK; Santos et al. 2005) with random realizations, $\delta T_{\text{syn}}(\nu, \hat{n})$. In Equation (6), A is the amplitude of the power spectrum, and ξ is the frequency-space correlation length of the emission. We use the reference scale $\ell_{\text{ref}} = 1000$ and the reference frequency

Table 1
Parameters of Isotropic Foreground Power Spectrum Model

Foreground Component	A (mK ²)	β	α	ξ
Iso. Gal. Syn.	700	2.4	2.80	4.0
Point sources	57.0	1.1	2.07	1.0
Galactic free-free	0.088	3.0	2.15	35
Extragalactic free-free	0.014	1.0	2.10	35

Note. The pivot values are $\ell_{\text{ref}} = 1000$ and $\nu_{\text{ref}} = 130$ MHz. See Equation (6) for different components.

$\nu_{\text{ref}} = 130$ MHz. The rest parameters are listed in Table 1. In order to restrict the synchrotron simulation to only the small scales, the random realizations are rescaled to 0 on the scales overlapping with the Haslam map’s resolution scale before adding to the final Galactic synchrotron map,

$$T_{\text{syn}}(\nu, \hat{n}) = T_{\text{syn},0}(\nu, \hat{n}) + \delta T_{\text{syn}}(\nu, \hat{n}). \quad (7)$$

Figure 1(c) shows an example of the simulated Galactic synchrotron map at the frequency of 1000 MHz.

Additionally, the SCK model is also used to simulate free-free emission and extragalactic point sources.

Extragalactic point sources are radio galaxies radiating from outside the Milky Way. Both neighboring radio galaxies and high-redshift radio galaxies are included. For high-redshift galaxies, their association with cosmological signals needs to be considered. We use the SCK model to simulate radio galaxies with certain clustering properties by introducing an LSS power spectrum model of radio galaxies through Equation (6), and the specific simulation parameters are shown in Table 1. Strictly speaking, high-redshift radio galaxies are the background components of the HI signal. However, their interference effect on the HI is the same as the foreground, and they are also used as foreground components in the simulation. Figure 1(c) shows an example of the simulated extragalactic point-source map at the frequency of 1000 MHz.

Free-free emission is the electromagnetic radiation produced by the deceleration of high-speed electrons when deflected by an interstellar ion, i.e., bremsstrahlung. Free-free emission is not homogeneous, and considering its exceptionally smooth frequency dependence and its subdominant amplitude, to simplify the simulation we approximate the free-free emission as isotropic and introduce its distribution power spectrum model through Equation (6) to simulate the free-free emission sky map. An example of Galactic and extragalactic free-free emission maps at 1000 MHz are shown in Figures 1(d) and (e).

The total foreground map is generated by combining all different components,

$$T_{\text{fg}}(\nu, \hat{n}) = T_{\text{syn}}(\nu, \hat{n}) + T_{\text{PS}}(\nu, \hat{n}) + T_{\text{GFF}}(\nu, \hat{n}) + T_{\text{EGFF}}(\nu, \hat{n}), \quad (8)$$

in terms of Galactic synchrotron map $T_{\text{syn}}(\nu, \hat{n})$, extragalactic point-source map $T_{\text{PS}}(\nu, \hat{n})$, Galactic free-free emission map $T_{\text{GFF}}(\nu, \hat{n})$, and extragalactic free-free emission map $T_{\text{EGFF}}(\nu, \hat{n})$.

2.3. White Noise

We also consider the instrumental noise and its impact on the observed HI signal. The instrument generates both the uncorrelated white noise and the correlated $1/f$ noise (Li et al. 2021). The effect of $1/f$ noise is related to the detailed scanning strategy. To simplify the simulation, we only consider the white

noise and add onto our simulated maps a Gaussian random realization, $T_{\text{noise}}(\nu, \hat{n})$, with rms

$$\sigma_{\text{noise}} = T_{\text{sys}} \sqrt{\frac{4\pi f_{\text{sky}}}{\Omega_{\text{beam}} N_{\text{dish}} t_{\text{obs}} \Delta\nu}}, \quad (9)$$

where N_{dish} is the number of telescope dishes or feeds, $\Omega_{\text{beam}} \approx (\lambda/D)^2$ is the solid angle of the telescope main beam, t_{obs} is the total observation time, and $\Delta\nu$ is the frequency resolution.

Since there is no full-sky experiment for HI, we refer to the observational parameters of a currently available experiment, the MeerKAT telescope, for a simple simulation of the noise. The analysis reveals that the instrumental noise is much smaller relative to the foreground, so we only briefly consider the instrumental noise. We fix the sensitivity of telescope $T_{\text{sys}} = 16$ K, the number of dishes $N_{\text{dish}} = 64$, the observation time $t_{\text{obs}} = 8000$ h, the sky coverage $f_{\text{sky}} = 1$, the frequency resolution $\Delta\nu = 2.34375$ MHz, and the dish diameter $D = 13.5$ m. So, the uncorrelated Gaussian white noise with an rms $\sigma_{\text{noise}} = 0.25$ mK. As an example, one simulated noise map of frequency slice at $\nu = 1000$ MHz is shown in Figure 1(b).

The foreground parameters listed in Table 1 are fixed in our simulations. In practice, these parameters may differ, resulting in different foreground features. However, such variation of parameters does not increase the degrees of freedom of the foreground components. A preprocessing with blind foreground subtraction, such as PCA, can remove the major foreground components without accurate measurements of the foreground parameters. The residuals have subdominant effects when the foreground parameters are changed. To make the analysis easier, we skip generating the robust training set simulations by varying the foreground parameters.

3. Primary Beam Model

The final simulated sky map is the linear combination of the different components,

$$T_{\text{sky}}(\nu, \hat{n}) = T_{\text{HI}}(\nu, \hat{n}) + T_{\text{fg}}(\nu, \hat{n}) + T_{\text{noise}}(\nu, \hat{n}). \quad (10)$$

The observed sky map is the observational sky map convolved with the telescope’s primary beam.

The frequency dependence of the primary beam, such as the frequency evolution of the beam size, can complicate the foreground emission. Such a feature has been noticed in most of the HI IM experiments and can be overcome by degrading the beam size to the largest one (Switzer et al. 2013). To simulate such an effect, we use a simple Gaussian beam model with frequency-dependent beam size,

$$B_G(\nu, \theta) = \exp \left[-4 \ln 2 \left(\frac{\theta}{\Delta\theta_G(\nu)} \right)^2 \right], \quad (11)$$

where $\Delta\theta_G(\nu)$ is the full width at half maximum (FWHM) of the primary beam, i.e., frequency-dependent beam size. Such a frequency dependence is modeled as (Matshawule et al. 2021)

$$\Delta\theta_G(\nu) = 1.16 \frac{\lambda(\nu)}{D}, \quad (12)$$

where $\lambda(\nu) = c/\nu$ is the wavelength at frequency ν and c is the speed of light. The factor of 1.16 is a good approximation to the MeerKAT primary beam. We take the MeerKAT beam as the example throughout this work and use $D = 13.5$ m.

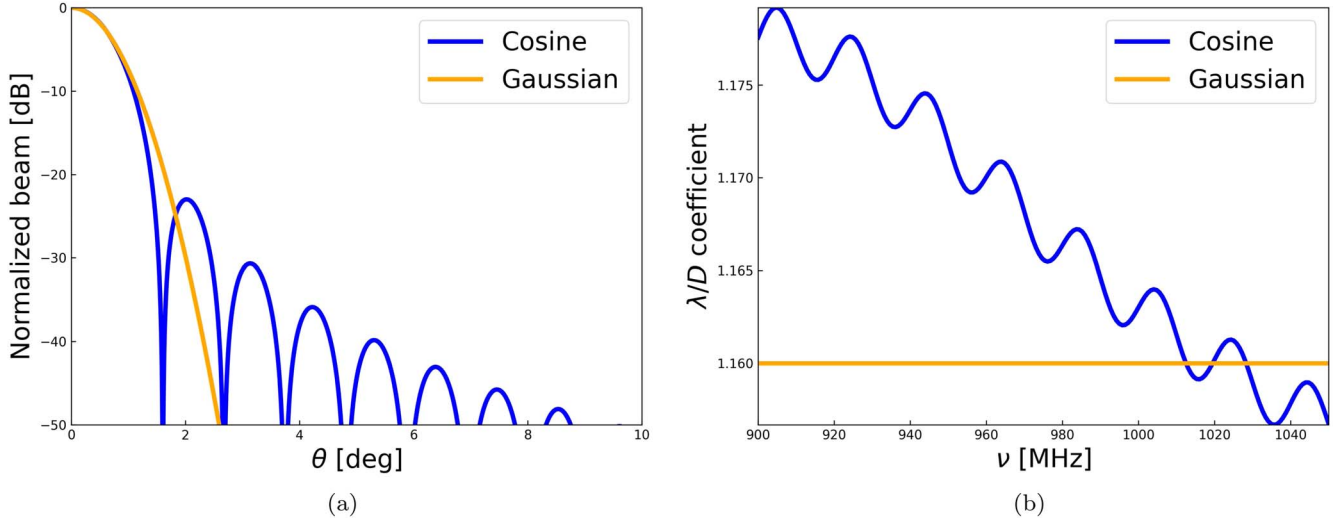


Figure 2. MeerkAT beam model (Matshawule et al. 2021). Panel (a): Comparison of primary beam models, the standard Gaussian model (orange) and the Cosine model (blue), at 1000 MHz. Panel (b): The λ/D normalized beam size as a function of frequency for the Gaussian model [orange, see Equation (12)] and the Cosine model [blue, see Equation (14)].

The Gaussian beam model only includes the frequency variation of the primary beam size and is not accurate enough to recover the actual primary beam feature. Using the MeerkAT *L*-band full-polarization “astrophotographic” observation, Asad et al. (2021) reconstructed a Zernike-based beam model and provided a software tool named EIDOS⁵, which is the most accurate beam model for MeerkAT dishes. With such a model, two important features can be seen clearly. One is the primary beam sidelobes. The EIDOS software provides the accurate primary beam sidelobe feature within a 5° radius. To extend the sidelobes to a large separation angle, following Matshawule et al. (2021), we use a cosine-tapered field illumination function to produce the sidelobe pattern,

$$B_C(\nu, \theta) = \left[\frac{\cos(1.189\pi/\Delta\theta_C(\nu))}{1 - 4(1.189\theta/\Delta\theta_C^2(\nu))} \right]^2, \quad (13)$$

which is called the Cosine beam model in this work. The primary beam patterns of the Gaussian and Cosine beam models at the frequency of 1000 MHz are shown in Figure 2(a).

The Cosine beam model also has the frequency-dependent beam size, $\Delta\theta_C(\nu)$. Another important primary beam feature produced with the EIDOS software is that the beam size follows the frequency evolution as expressed in Equation (12) but also exhibits a low-level frequency-dependent ripple. Such a ripple can be well modeled with a sinusoidal oscillation on top of a high-order polynomial function (Matshawule et al. 2021),

$$\Delta\theta_C(\nu) = \frac{\lambda(\nu)}{D} \left[\sum_{d=0}^8 a_d \hat{\nu}^d + A \sin\left(\frac{2\pi\hat{\nu}}{T}\right) \right], \quad (14)$$

where $\hat{\nu} = \nu/\text{MHz}$, $A=0.1$, $T=20$, and $a_d = \{3.4 \times 10^{-21}, -3.0 \times 10^{-17}, -1.2 \times 10^{-11}, -2.6 \times 10^{-10}, 3.5 \times 10^{-7}, -3.0 \times 10^{-4}, 0.16449, -50.37020, 6704.28133\}$. The λ/D normalized beam sizes for both Gaussian and Cosine beam models are shown in Figure 1(b).

In this work, we use the Gaussian beam model as a simple case and consider the Cosine beam model as a realistic case. Although the model is based on the MeerkAT primary beam measurements,

the frequency-dependent effects addressed here exist in most H I IM experiments, as most of the radio telescope dishes have similar sidelobe and frequency-dependent ripple features.

The measured sky brightness temperature map is the convolution of the original sky map with one of the beam models (see Figure 3).

4. U-Net Network

This section introduces the deep neural network architecture (3D U-Net) used in our foreground subtraction analysis (Cićek et al. 2016; Fabian & Klaus 2019; Makinen et al. 2021). The U-Net network is a convolutional neural network (CNN) originally developed for biomedical image segmentation (Ronneberger et al. 2015). It is based on the CNN, but with significant structural modifications. The basic U-Net architecture is shown in Figure 4. In addition to the standard CNN architecture, U-Net has many feature channels in the upsampling part, allowing the network to propagate contextual information to higher resolution layers through a series of transpose convolutions. The main idea is to add successive layers to the traditional contractual network, with an upsampling operation replacing the convergence operation. As a result, these layers increase the output resolution. The extended path is more or less symmetric with the contracted part and produces a U-shaped structure (Ronneberger et al. 2015).

The U-Net network is mainly used to process 2D images. In contrast, H I observations can be obtained for multiple frequency channels in a specific frequency range. The correlation between frequencies is the main feature distinguishing the H I signal from the foreground. It is shown that the foreground signal has a stronger frequency correlation than the H I signal. In addition, the complex variation of the primary beam shape of the telescope with frequency is the main difficulty of foreground subtraction. Therefore, introducing frequency information in deep learning is necessary to eliminate systematic effects and extract foreground information accurately. For the H I foreground subtraction problem, we adopt the generalized model of the U-Net network, i.e., 3D U-Net, which can include both angular and frequency information (Ronneberger et al. 2015; Cićek et al. 2016; Villanueva-Domingo & Villaescusa-Navarro 2021; Fabian & Klaus 2019; Makinen et al. 2021).

⁵ <https://github.com/ratt-ru/eidos>

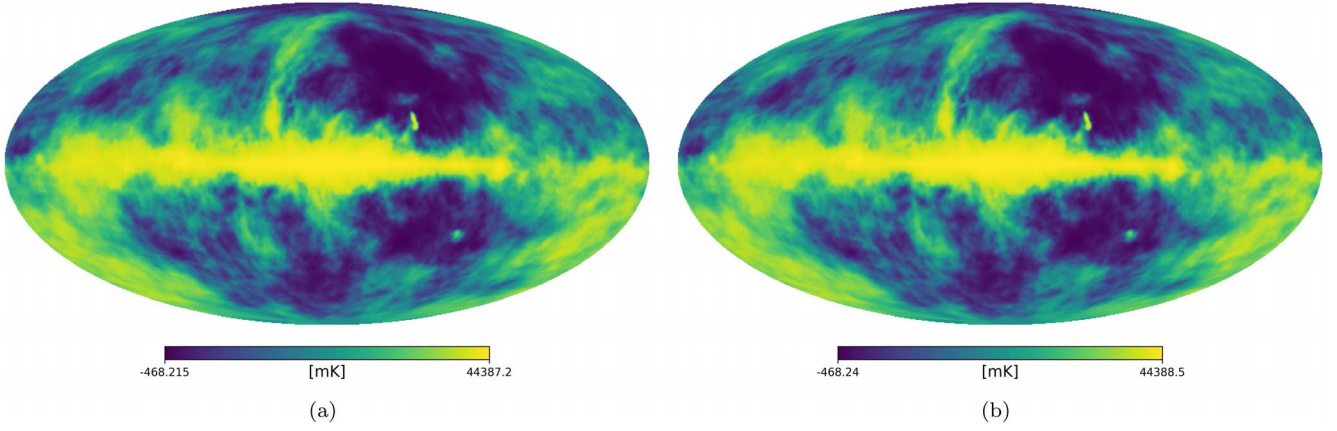


Figure 3. Different beam convolved full-sky maps at 1000 MHz: Gaussian beam convolved (a) and Cosine beam convolved (b) maps. The maps are shown using the histogram equalized color mapping to amplify the image's contrast.

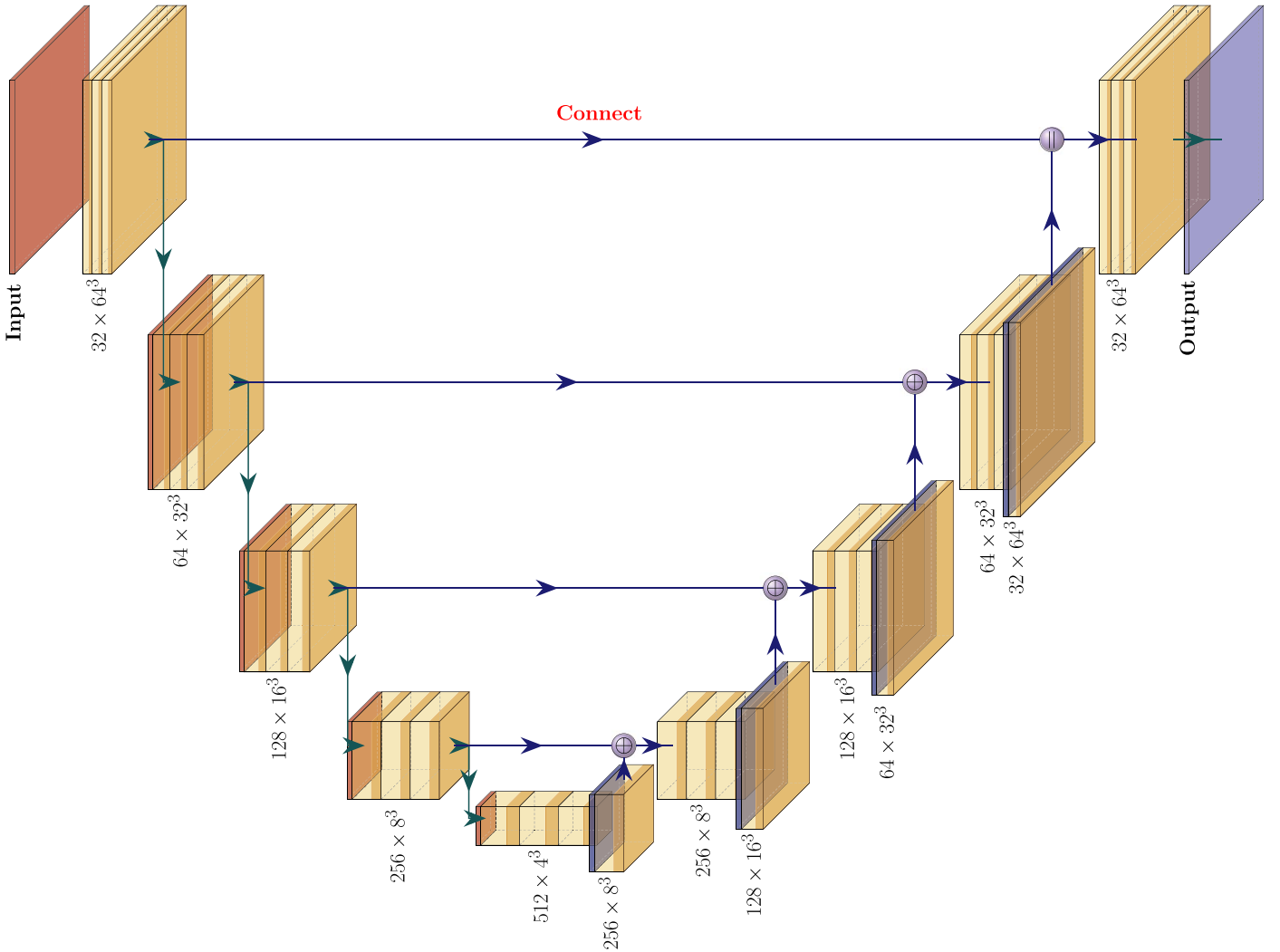


Figure 4. The U-Net architecture. Each orange box corresponds to a multichannel feature map and a series of convolutions. The bottom of each box illustrates the number of channels and the size of the output. The dark red box and the dark blue box indicate the input and output. The light blue box shows the connected portion of the box during transpose convolutions. The down, right, and up arrows indicate maximum pooling, skip connection, and transpose convolutions, respectively. The last arrow operation denotes a $1 \times 1 \times 1$ convolution for mapping 64 features to the final 3D data. This visualization is made with the PlotNeuralNet library⁶.

4.1. Data Preprocessing

The core algorithm of U-Net is semantic segmentation, which is an approach to detecting the belonging class of the object for every pixel (Ronneberger et al. 2015; Makinen et al. 2021;

Villanueva-Domingo & Villaescusa-Navarro 2021). It can only handle 2D flat images or 3D sky patches, but not curved images

⁶ <https://github.com/HarisIqbal88/PlotNeuralNet>

or sky patches directly. As shown in the literature (Makinen et al. 2021), the smooth components of the foreground contamination can be removed efficiently. However, the non-smooth foreground residual, which is mainly caused due to the systematic effect, challenges the detection of the H I LSS signal. The foreground residual can be further reduced with the model-independent foreground subtraction method, such as PCA. Since the smooth components of the foreground are not the major problem and can be efficiently removed with the current foreground cleaning method, we first perform PCA on the simulated data before feeding it to the U-Net architecture.

On the other hand, preprocessing with PCA is necessary. Without the preprocessing step, the U-Net performs poorly in recovering the H I fluctuation signal. It is primarily due to the large dynamic range of the map amplitude (Di Matteo et al. 2002; de Oliveira-Costa et al. 2008; Liu et al. 2014), which makes extracting correct information with the U-Net network difficult. However, an aggressive PCA mode subtraction can result in H I signal loss (Switzer et al. 2015). According to the simulation in the literature, the smooth foreground contamination can be significantly removed with a few PCA modes (Spinelli et al. 2022). To avoid any H I signal loss, we subtract only the first three PCA modes. Subtracting a few more PCA modes, such as the first six PCA modes in Makinen et al. (2021), can potentially improve the benchmark. As long as the U-Net can handle the residuals, we wish to keep the preprocessing step to a minimum of mode subtraction. However, in practice, the preprocessing needs to be further improved because the actual foreground contamination is more complex than the simulated one.

The 3D U-Net architecture requires equal-area square image data with HEALPix pixelization. The full-sky map needs to be divided into several small sky patches. We split the full-sky map into 192 equal-area patches corresponding to the pixel area of HEALPix Nside of 4. Each patch contains 64×64 pixels. The full frequency range, i.e., $900 \sim 1050$ MHz, is binned into 64 frequency bands. Finally, the simulated full sky is split into 192 sky patches with a shape of $64 \times 64 \times 64$. Three sky patches, as an example, are illustrated in Figure 5. The sky patches from the raw simulated sky map, Gaussian beam convolved sky map, and Cosine beam convolved sky map are shown in the top, middle and bottom panels, respectively. We use the same sky segmentation method for the training, test, and validation data sets.

The direction-dependent spectral index is used in our simulation. The PCA is performed by taking the Singular Value Decomposition (SVD) of the input data cube. The data cube has already been reshaped into $N_\nu \times N_\theta$, where N_ν is the number of frequency channels and N_θ is the number of pointing directions. The singular modes are used as the basis functions as expressed in Equation (3.2) of Makinen et al. (2021) and the first few modes represent the foreground contamination. It is not necessary to assume each of the foreground components is separable in frequency and angular direction. The foreground modes estimated with PCA are not exactly the same sets of foreground components as the simulation input. Instead, it is the linear combination of the input components. As long as the foreground components can be expressed in the form of power-law spectrum, they can be significantly subtracted.

4.2. Loss Function

Although the dynamic amplitude range is significantly reduced with the PCA mode subtraction, there is still large variability between the residual foreground and H I signal. Following Makinen et al. (2021), the Logcosh loss function is used in our analysis,

$$L(p, t) = \sum_i \log \cosh(p_i - t_i), \quad (15)$$

where $p_i \in p$ denotes prediction and $t_i \in t$ denotes simulation target. The Logcosh function behaves much like the L1 norm in that it is more robust and less sensitive to outliers, which is the main reason we chose it.

4.3. Training and Testing

The core of any CNN is the convolutional layers, which are defined by the convolution of a number of filters on the input. The network is basically a stack of layers. Following Makinen et al. (2021), we fixed the number of convolution kernels to be 32 in the first convolution layer. The kernel size defines the field of view of the convolution, which is fixed to $3 \times 3 \times 3$. To obtain the required output dimensionality, we used “same” padding to handle the boundaries of the samples in the convolutions and transpose convolutions. The stride defines the step size of the kernel traversing the images. We use the default settings of stride equal to 1 in the convolutions and stride equal to 2 in the transpose convolutions.

Our selected U-Net architecture is trained end-to-end to the foreground removal using a set of H I foreground simulations. Table 2 presents the details of the hyperparameters used in this network. The NAdam optimizer is used in this analysis with the default TensorFlow parameters (Reddi et al. 2019). The hyperparameters are carefully fine-tuned to optimize the network. The batch size is optimized to 16 and the number of initial convolution filters is optimized to 32, both of which are limited by the GPU memory. The initial learning rate is set to 10^{-2} and decayed by a factor of 10 if the validation loss function is not improved after 20 epochs. Meanwhile, the weight decay is also tested with a set of prior values as listed in Table 2. Finally, we find that optimized values for learning rate and weight decay are 10^{-4} and 10^{-5} , respectively. The total number of trainable parameters is 7.4×10^7 . Following Makinen et al. (2021), we apply a rectified linear unit (ReLU) activation in every convolution. In Figure 6, we show the evolution of the loss function. The dark blue solid line indicates the training set loss function evolution and the light blue dashed line indicates the validation set loss function evolution. The results for Gaussian beam convolved maps and Cosine beam convolved maps are shown in the left and right panels, respectively.

The loss function of the validation set is converged after about 20 epochs and 50 epochs for Gaussian and Cosine beam convolved maps, respectively. The recovered power spectra in the cases with 20 and 100 epochs are shown in Figure 7, where the results for Gaussian maps are shown in the left panel and the results for Cosine maps are shown in the right panel. Significantly, the recovered angular power spectra from Gaussian beam convolved maps are consistent between the cases with 20 and 100 epochs, which is because

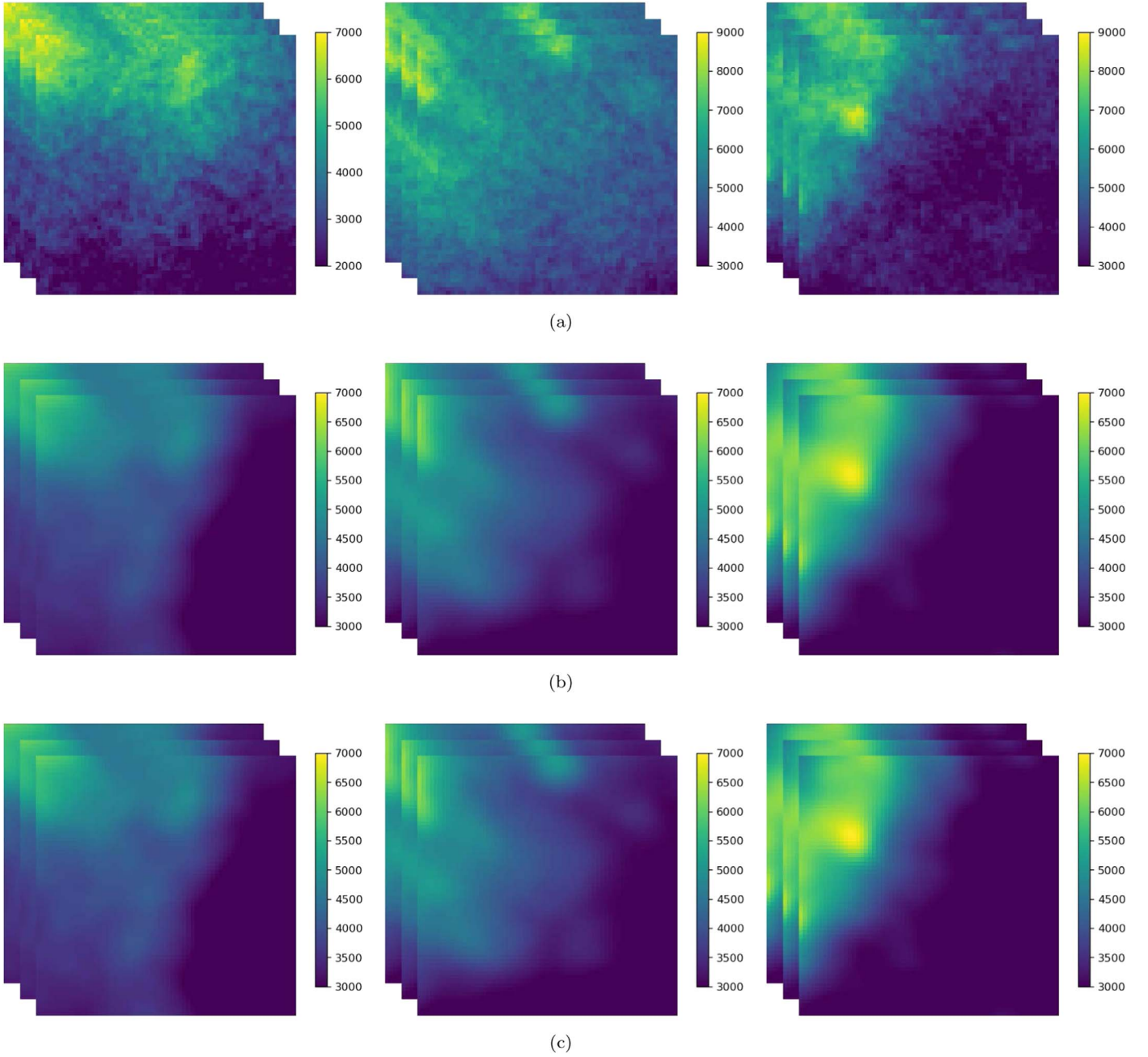


Figure 5. Data cubes of 214.86 deg^2 sky patches. From left to right, the center of the sky patches are $(l, b) = (179^\circ 648, 33^\circ 331)$, $(l, b) = (224^\circ 824, 9^\circ 594)$, and $(l, b) = (303^\circ 398, 2^\circ 836)$, where l and b are the Galactic longitude and Galactic latitude, respectively. The redshift range of the sky patches is $z \in [0.353, 0.578]$. The unit of the sky patches is mK. Panel (a) shows the raw simulated sky patch without beam convolved, Panel (b) shows the Gaussian beam convolved sky patch, and Panel (c) shows the Cosine beam convolved sky patch.

Table 2
Description of the Hyperparameters in the U-Net Architecture Design

Hyperparameter	Description	Prior Value	Optimum Value
η	learning rate for optimizer	$10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$	10^{-4}
ω	weight decay for optimizer	$10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}$	10^{-5}
n_{filter}	initial number of convolution filters	8, 16, 32	32
b	batch size, i.e., number of samples per gradient descent step	8, 16	16
Ω	optimizer for training	Adam, NAdam	NAdam

the loss function is converged after 20 epochs. Although the loss function in the Cosine case converges after about 50 epochs, the improvement in recovering the power spectrum is

not obvious after 20 epochs. Therefore, in our analysis, we adopt a 20 epoch calculation scheme to make our computations faster.

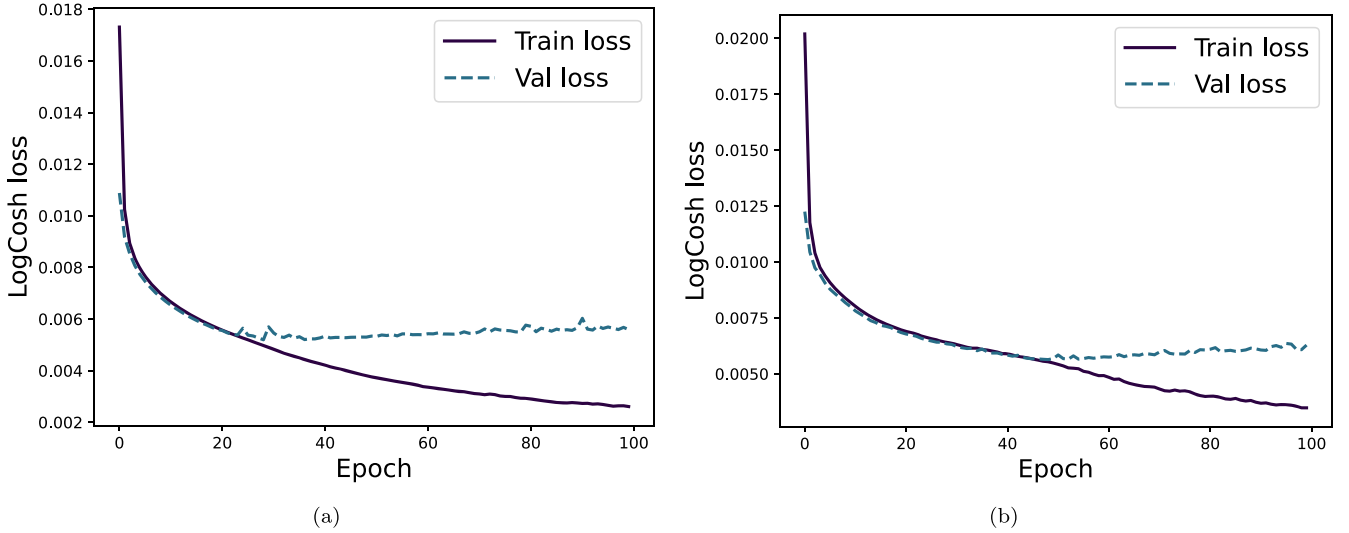


Figure 6. Loss function evolution for each network as a function of the number of epochs. The dark blue solid line indicates the training set loss function evolution and the light blue dashed line indicates the validation set loss function evolution. Panel (a) is U-Net foreground deduction for Gaussian beam convolved maps, and Panel (b) is U-Net foreground deduction for Cosine beam convolved maps.

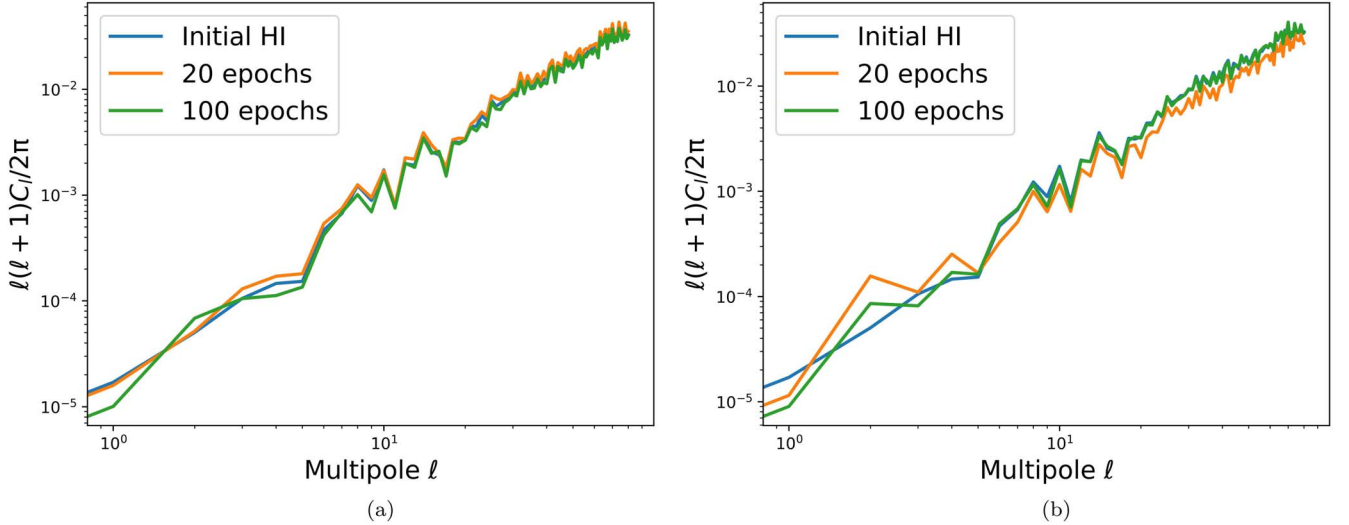


Figure 7. The angular power spectra of the simulated sky maps after PCA+U-Net foreground subtraction. Panel (a) shows the angular power spectra from the Gaussian beam convolved sky map, and Panel (b) shows the angular power spectra from the Cosine beam convolved sky map.

After completing the training, we investigated the accuracy of the network’s prediction. In Figure 8, we show the foreground subtraction results for PCA and PCA+U-Net in the presence of Gaussian and Cosine instrumental effects.

5. Results and Discussion

Three different small patches of the foreground-cleaned maps are shown in different columns of Figure 9. The different sky patches of the initial HI signal map are shown in Figure 9(a). Figures 9(b) and (c) show the sky patches of the Gaussian beam convolved maps cleaned with PCA and PCA+U-Net, respectively. Both the PCA and PCA+U-Net cleaned maps have good consistency with the initial HI map. It shows that both PCA and PCA+U-Net foreground subtraction approaches perform well and that the systematic effect caused by the simple Gaussian beam model can be effectively eliminated with both methods.

Figures 9(d) and (e) show patches of Cosine beam convolved maps cleaned with PCA and PCA+U-Net, respectively. As we can see in Figure 9(d), the PCA cleaned maps have extra structures. Such structures can be eliminated with the additional U-Net process, as shown in Figure 9(e).

We can verify the foreground subtraction efficiency by comparing the angular power spectrum to the initial HI model. The sky patches are refilled to their original positions after foreground subtraction, forming the cleaned whole sky map. The angular power spectrum is shown in Figure 8. The initial HI power spectrum model is shown with the blue line. The orange line shows the result of the simulation map cleaned with the first three PCA modes. The green line shows the result with the combination of PCA and U-Net foreground subtractions.

5.1. PCA Foreground Subtraction

We compare the results for the Gaussian beam convolved map and the Cosine beam convolved map. With the Gaussian

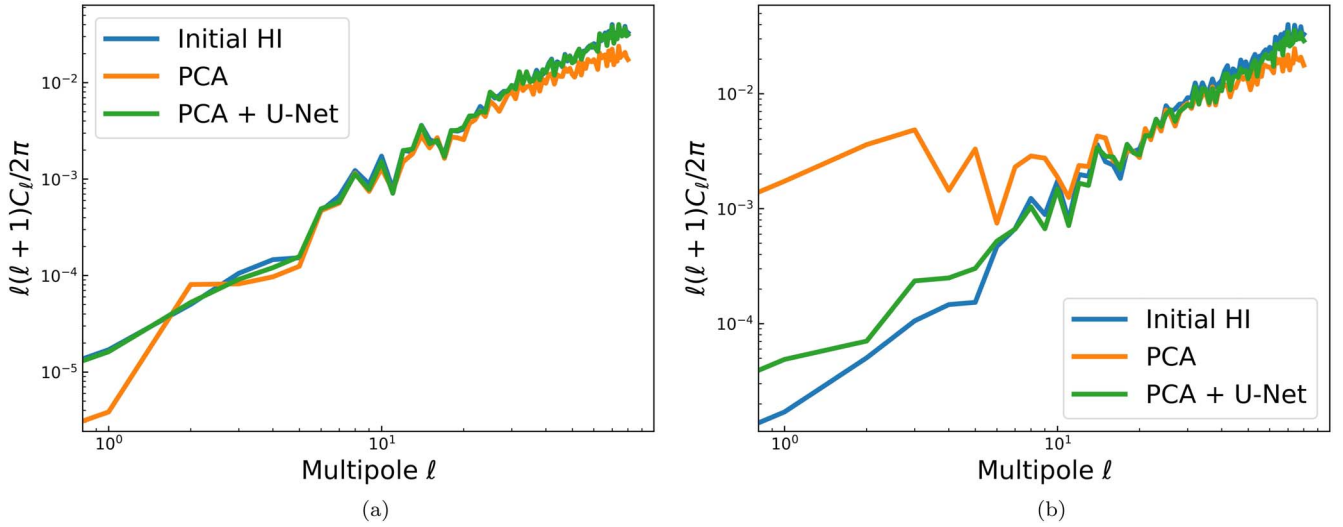


Figure 8. The angular power spectra of the simulated sky maps after foreground subtraction. The blue, orange, and green lines show the results of the initial H I model, the foreground subtraction with the first three PCA modes, and the foreground subtraction with the combination of PCA and U-Net, respectively. Panel (a) shows the angular power spectra from the Gaussian beam convolved sky map, and Panel (b) shows the angular power spectra from the Cosine beam convolved sky map.

beam convolved map, the H I signal can be recovered by simply subtracting a few of the PCA modes, e.g., three modes subtraction in this analysis. The variation on large scales is due to the effect of the signal loss during blind foreground subtraction, and the reduction on small scales is due to beam smoothing. However, if the map is convolved with the Cosine beam, the H I signal cannot be recovered through PCA foreground subtraction. The reason is that the Cosine beam model induces much more systematic effects than the Gaussian beam. As discussed in Section 3, with the Gaussian beam model, only the effect of beam size variation with frequency is induced. The Cosine beam model includes the beam sidelobes and the frequency-dependent beam size ripple. The intensity of the bright sources outside of the field of view can be observed through the beam sidelobe, which is a non-smooth component. Meanwhile, the beam size ripple directly complicates the foreground spectrum shape. Such effects can be further investigated through the 2D power spectrum.

With the data cubes of 192 sky patches, we estimate the 2D autocorrelation power spectrum and plot the ratio between the foreground-cleaned map and the initial H I model,

$$R(k_{\parallel}, k_{\perp})_{\text{auto}} = \frac{\bar{P}_{\text{cln}}(k_{\parallel}, k_{\perp})}{P_{\text{HI}}(k_{\parallel}, k_{\perp})} - 1, \quad (16)$$

where $\bar{P}_{\text{cln}}(k_{\parallel}, k_{\perp}) = \frac{1}{192} \sum_{i=1}^{192} P_{\text{cln}}^i(k_{\parallel}, k_{\perp})$ and $P_{\text{cln}}^i(k_{\parallel}, k_{\perp})$ is the 2D power spectrum of one foreground-cleaned sky patch. The results for PCA cleaned map are shown in Figure 10. The left panel shows the result of the Gaussian beam convolved map. The values of $R(k_{\parallel}, k_{\perp})_{\text{auto}}$ are generally consistent with 0 at large scales (i.e., $k_{\perp} \lesssim 0.06 \text{ Mpc}^{-1}$), which indicates that the H I signal can be recovered via PCA at large scales. The negative $R(k_{\parallel}, k_{\perp})_{\text{auto}}$ values indicate the signal reduction, especially on large k_{\perp} . Such signal reduction is mainly due to the beam smoothing effect. We cannot observe significant signal reduction at small k_{\parallel} . As shown in the literature (Switzer et al. 2015), the H I signal at such scales is always removed during the PCA foreground subtraction. However, because we

only remove the first three PCA modes, such signal loss effect is still negligible.

The result of the Cosine beam convolved map is shown in the right panel of Figure 10. There is significant extra power at $k_{\parallel} \sim 0.7 \text{ Mpc}^{-1}$. Such extra power is mainly due to the beam size ripple induced by the Cosine beam model. Our results are consistent with the simulation analysis in the literature (Matshawule et al. 2021). In order to remove such a systematic effect, it needs aggressive mode subtraction, which can cause non-negligible H I signal loss (Switzer et al. 2015).

Note that the white strips in Figure 10 do not have any physical meaning, which is due to the logarithmic binning being not uniform in the linear k -space. The same is also in Figures 11–13.

In order to check the concordance between the foreground removed H I map and the initial H I map, we estimate the cross-correlation power spectrum, $P_{\text{cln, HI}}(k_{\parallel}, k_{\perp})$. The concordance can be evaluated with the power spectrum ratio between the foreground-cleaned map and the initial H I map,

$$R(k_{\parallel}, k_{\perp})_{\text{cross}} = \frac{\bar{P}_{\text{cln, HI}}(k_{\parallel}, k_{\perp})}{P_{\text{HI}}(k_{\parallel}, k_{\perp})} - 1, \quad (17)$$

where $\bar{P}_{\text{cln, HI}}(k_{\parallel}, k_{\perp})$ represents the averaged cross-correlation power spectrum across different sky patches. The results for PCA cleaned map are shown in Figure 11. The results of the Gaussian and Cosine beam convolved maps are shown in the left and right panels, respectively. Obviously, the extra power shown in $R(k_{\parallel}, k_{\perp})_{\text{auto}}$ of the Cosine beam convolved case is eliminated through the cross-correlation power spectrum estimator. This is because the power spectrum error due to the systematic effect can be eliminated through cross-correlation. The values of $R(k_{\parallel}, k_{\perp})_{\text{cross}}$ in both Gaussian and Cosine beam convolved cases show a similar trend. The negative $R(k_{\parallel}, k_{\perp})_{\text{cross}}$ values at large k_{\perp} indicate the reduction of the correlation efficiency, which is mainly due to the beam smoothing effect. However, although the systematic effect can be removed via the cross-correlation power spectrum, it still enlarges the power spectrum estimation error and biases

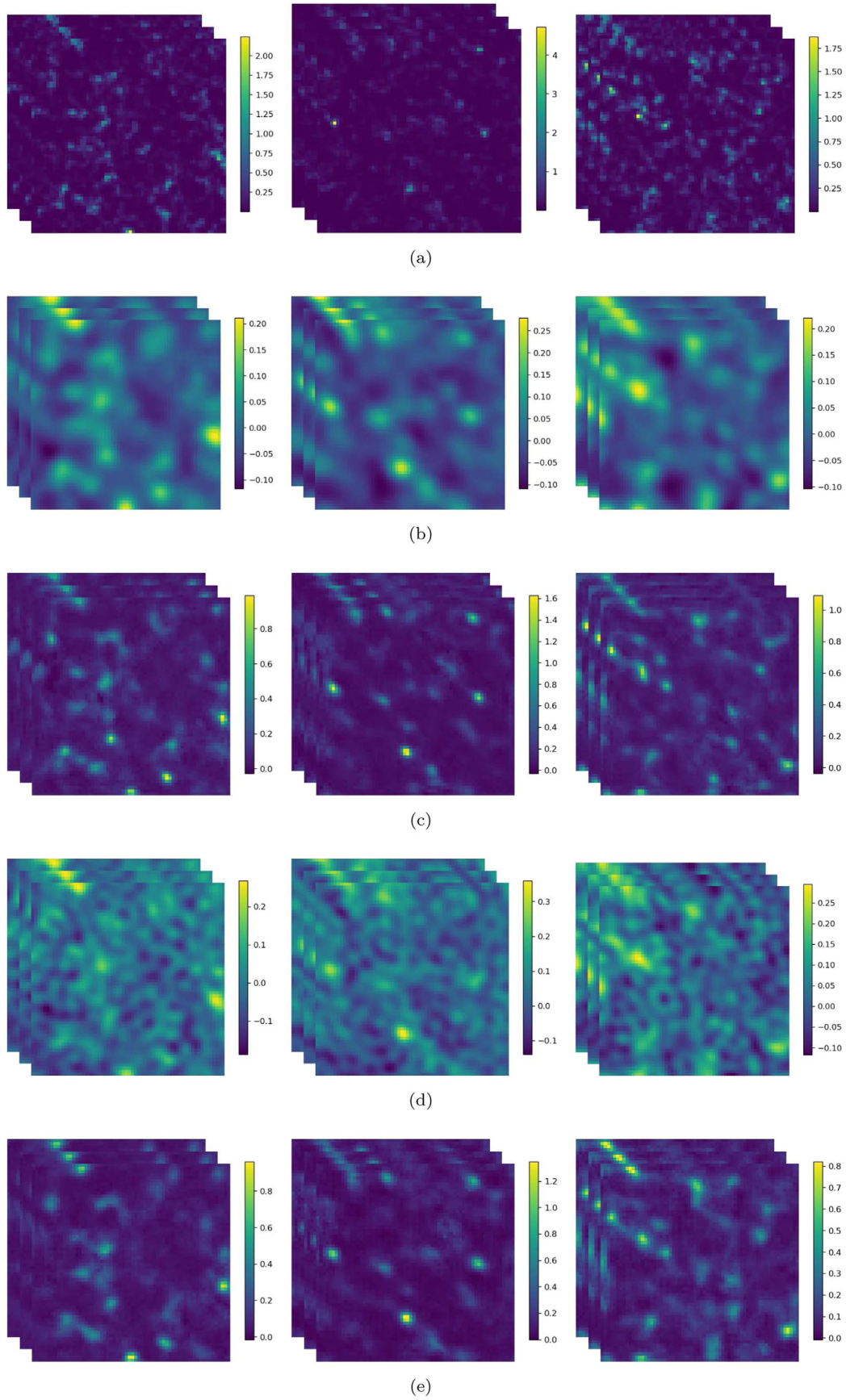


Figure 9. Data cubes of the same sky patches as shown in Figure 5, but with the foreground subtracted. From top to bottom are initial H I map (a), Gaussian beam convolved map cleaned with PCA (b), Gaussian beam convolved map cleaned with PCA+U-Net (c), Cosine beam convolved map cleaned with PCA (d), and Cosine beam convolved map cleaned with PCA+U-Net (e). The unit of the cube is mK.

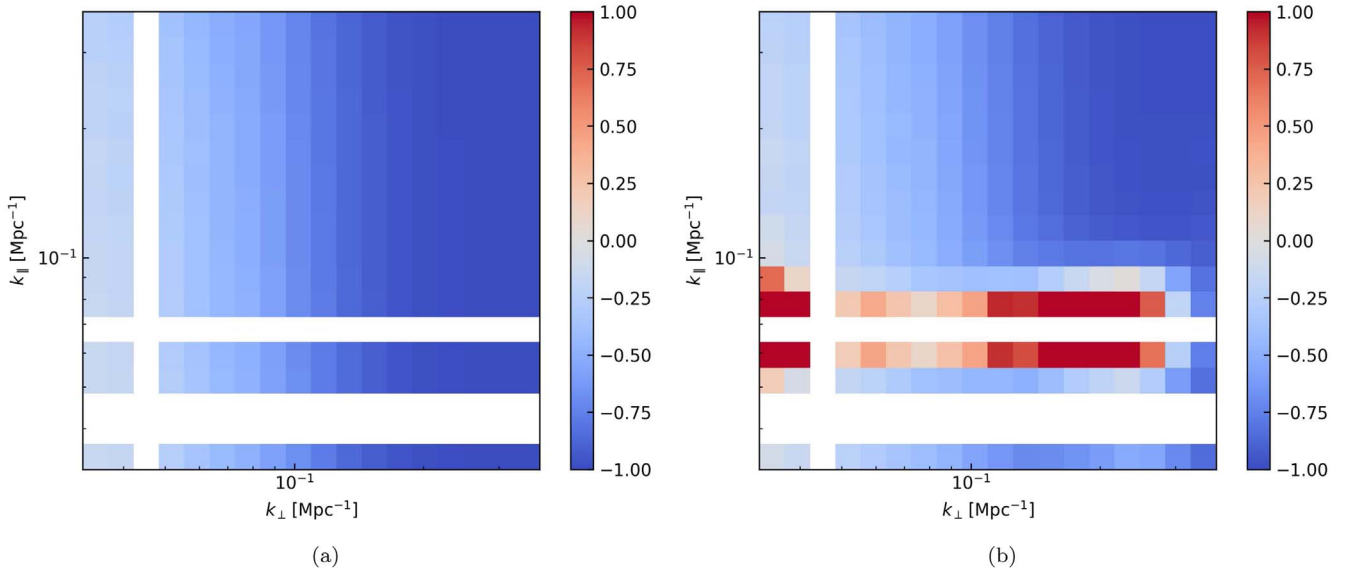


Figure 10. The 2D autocorrelation power spectrum ratio of the PCA cleaned map and the initial H I model. Panel (a) shows the result from the Gaussian convolved map and Panel (b) shows the result from the Cosine beam convolved map.

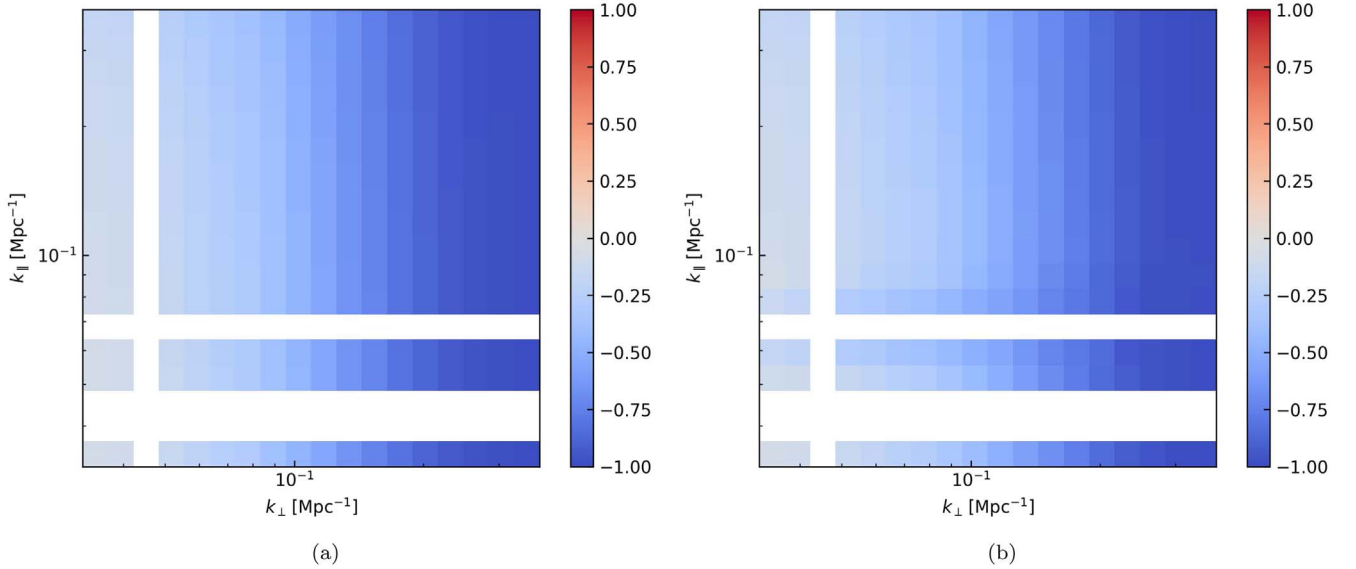


Figure 11. The 2D cross-correlation power spectrum ratio of the PCA cleaned map and the initial H I model. Panel (a) shows the result from the Gaussian convolved map and Panel (b) shows the result from the Cosine beam convolved map.

the autocorrelation power spectrum. The elimination of the systematic effect is crucial to the autocorrelation of power spectrum detection.

5.2. Additional U-Net Foreground Subtraction

Instead of applying aggressive additional mode subtraction, we feed the residual maps to the U-Net architecture. The angular power spectra of the U-Net cleaned maps are shown with green lines in Figure 8. The U-Net foreground subtraction corrects the angular power spectrum in both the Gaussian and Cosine beam convolved cases. But in the Gaussian beam convolved case, the correction from U-Net is not evident. A significant improvement can be seen from the results in the Cosine beam convolved map, as shown in Figure 8(b). With the additional U-Net foreground subtraction, the angular power spectrum deviation at large scales is significantly eliminated.

The recovered angular power spectrum is consistent with the initial H I power spectrum. Next, we can further investigate it with the 2D power spectrum.

The $R(k_{\parallel}, k_{\perp})_{\text{auto}}$ values, as specified by Equation (16), in the U-Net foreground subtracted sky patches are shown in Figure 12. The results for the Gaussian and Cosine beam convolved cases are shown in the left and right panels, respectively. These two results are generally consistent with each other, which indicates that the U-Net foreground subtraction can remove the systematic effect induced by either the simple Gaussian beam model or the complicated Cosine beam model.

The $R(k_{\parallel}, k_{\perp})_{\text{cross}}$ values, as specified by Equation (17), in the U-Net foreground subtracted sky patches are shown in Figure 13, with the Gaussian beam convolved case in the left panel and the Cosine beam convolved case in the right panel, respectively. The $R(k_{\parallel}, k_{\perp})_{\text{cross}}$ values are generally ~ 0 ,

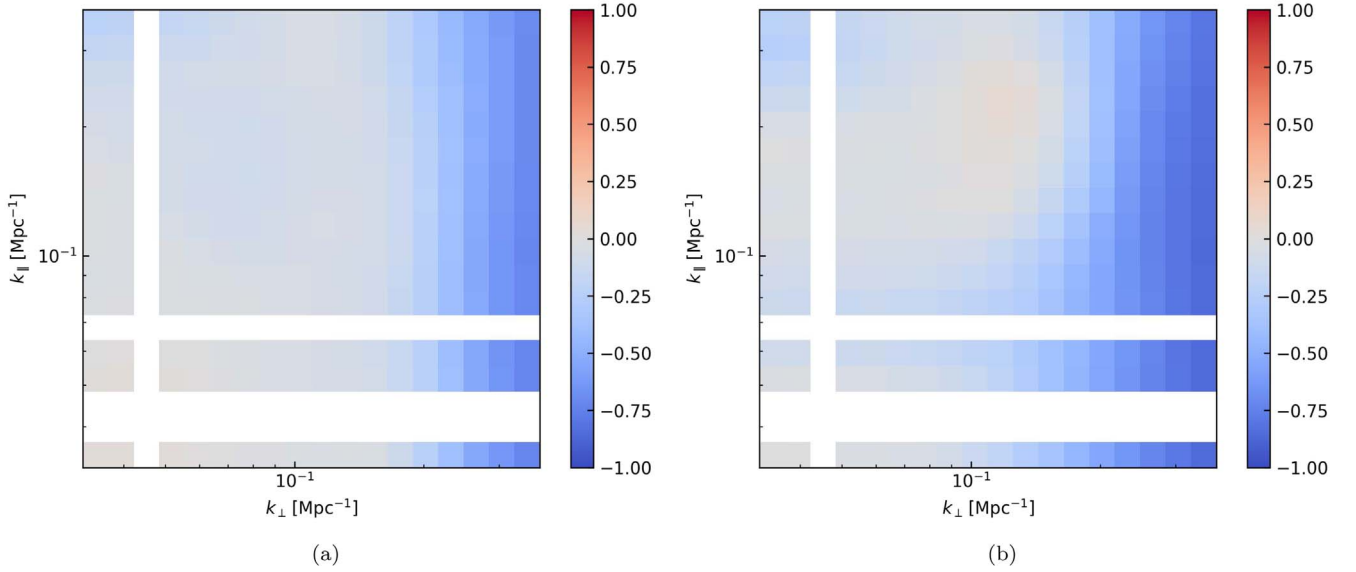


Figure 12. The 2D autocorrelation power spectrum ratio of the U-Net cleaned map and the initial H I model. Panel (a) shows the result from the Gaussian convolved map cross-correlated with the initial H I map and Panel (b) shows the result from the Cosine beam convolved map cross-correlated with the initial H I map.

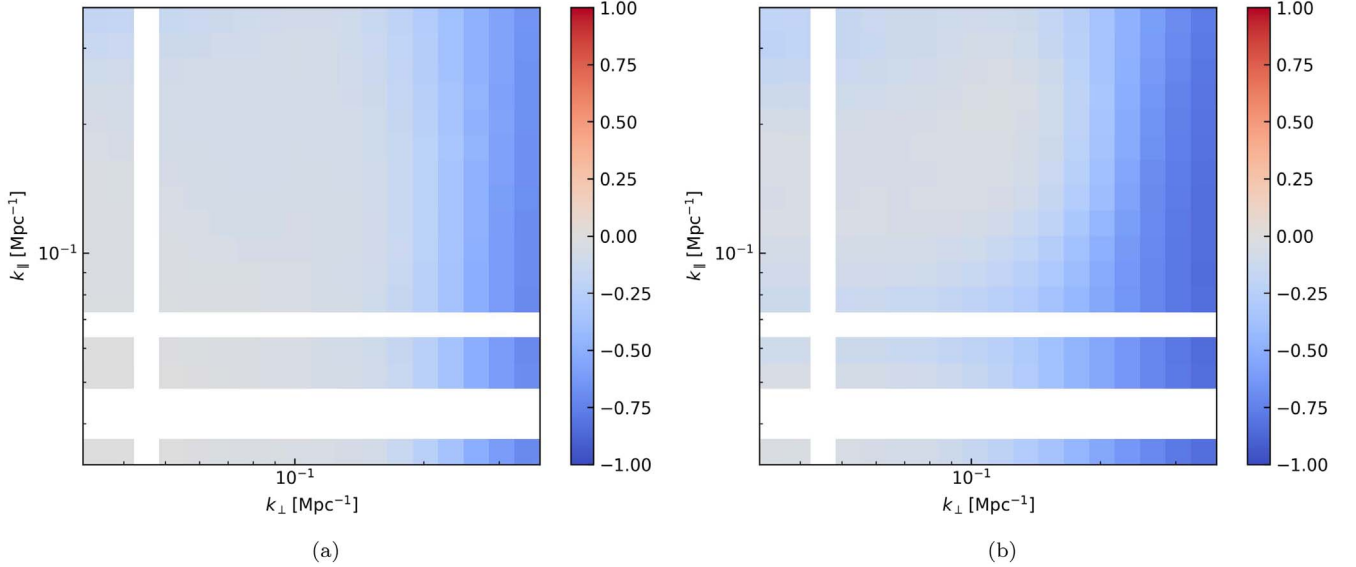


Figure 13. The 2D cross-correlation power spectrum ratio of the U-Net cleaned map and the initial H I model. Panel (a) shows the result from the Gaussian convolved map cross-correlated with the initial H I map and Panel (b) shows the result from the Cosine beam convolved map cross-correlated with the initial H I map.

indicating that the recovered H I fluctuation signal is consistent with the initial H I signal.

5.3. PCA versus U-Net

In Figure 8(a), PCA is effective in reducing the foreground with Gaussian beam, but slightly worse than the U-Net network. As can be seen in Figure 8(b), the results of PCA are much worse than those of the U-Net when dealing with the Cosine beam convolutional sky map. This indicates that PCA fails in processing the sky map with convolved complex bundles, while the U-Net network can learn large-scale information after the convolved bundles of the sky map. In contrast, the U-Net network can still recover the angular power spectrum of the original H I signal on large scales.

In Figures 10–13, we plotted average autocorrelation and cross-correlation power spectra in the lateral (k_{\perp}) and radial

(k_{\parallel}) directions. We define a measure of concordance:

$$\Gamma = \text{mean}\{R(k_{\parallel}, k_{\perp})_{\text{X}}\} + 1, \quad (18)$$

where $\text{mean}\{\dots\}$ represents the mean value across the k_{\parallel} – k_{\perp} bins and $R(k_{\parallel}, k_{\perp})_{\text{X}}$ is the spectrum ratio defined in Equations (16) and (17) with the subscript X denoting the auto or cross power spectrum.

We first discuss the case of the autocorrelation power spectrum. We obtain the concordances with the original H I map using both the PCA and U-Net methods, i.e., $\Gamma_{\text{pca}}^{\text{gau}} = 60.73\%$, $\Gamma_{\text{pca}}^{\text{cos}} = 28.46\%$, $\Gamma_{\text{unet}}^{\text{gau}} = 77.38\%$, and $\Gamma_{\text{unet}}^{\text{cos}} = 69.66\%$, respectively. So, it is clearly seen that the 3D U-Net method is much better than the PCA method in recovering the H I signal. We find that, in the case of the Gaussian beam, the concordance of using the U-Net method is better than that of PCA by 27.4%, and in the case of the Cosine beam, the concordance of using the U-Net

method is better than that of PCA by 144.8%. The results of $R(k_{\parallel}, k_{\perp})_{\text{cross}} \sim 0$ at large scales for both the PCA and U-Net foreground subtractions indicate that the recovered H I fluctuation is consistent with the initial H I signal.

It should be noticed that, although the sky map is partitioned into 192 small sky patches, the H I fluctuation on large scales can still be properly recovered. This is due to the fact that either the foreground contamination or the systematic effect considered here are highly correlated across frequencies. The results indicate that, even with small sky patch observation, the systematic effect induced by the telescope primary beam can be well eliminated using the U-Net architecture. It is ideal for most of the current stage H I IM experiments.

The U-Net architecture is one of the supervised deep-learning methods. It means that the foreground subtraction with the U-Net architecture is model dependent. The accuracy of the primary beam model affects the foreground subtraction efficiency. However, the telescope's primary beam, especially for telescope arrays, can be measured via "astroholographic" observation (Asad et al. 2021). Therefore, the U-Net-based foreground subtraction sheds new light on eliminating the primary beam effect for future H I IM experiments.

PCA is currently one of the most effective blind foreground subtraction methods that have been widely discussed in different analyses with both simulated and actual observational data. Due to the systematic effect, most of the blind foreground subtraction methods, including the PCA method, can result in either significant foreground residual or substantial signal loss. In this work, we show that the additional foreground subtraction with the U-Net network can eliminate the primary beam effect and improve the efficiency of PCA foreground subtraction. It can potentially improve other blind foreground subtraction methods, such as independent component analysis (ICA; Wolz et al. 2017), as well as nonblind foreground subtraction methods.

6. Conclusion

Deep learning, derived from artificial neural networks, can overcome the shortcomings of traditional feature extraction and matching, and has made significant breakthroughs in the field of image recognition. Deep learning has been very successful in astronomy by combining underlying features to form more abstract higher-level representations of attribute classes or features, trained with large amounts of specific data.

In the H I 21 cm IM survey, the foreground contamination on the cosmological signals is extremely severe, and the systematic effects caused by radio telescopes themselves further aggravate the difficulties in subtracting the foregrounds. These foregrounds mainly originate from Galaxy synchrotron, Galaxy free-free, extragalactic free-free, and extragalactic point sources emissions, which are in total five orders of magnitude higher than the cosmological signals in the H I IM survey. Therefore, finding an effective way to subtract foregrounds is one of the crucial challenges for the H I IM observations.

In this work, we used a similar network structure to Makinen et al. (2021). Compared with Makinen et al. (2021), we consider the frequency band of the MeerKAT telescope and, more importantly, we focus on the systematic effect induced by the radio telescope's primary beam. For the radio telescope primary beam effect, we consider the Gaussian beam model as a simple case and the Cosine beam model as a sophisticated case. We investigate whether the deep-learning method, concretely the U-Net algorithm in this work, can play a crucial

role in eliminating such primary beam-induced systematic effect and improving the foreground subtraction efficiency.

The results of PCA foreground subtraction are different between using the Gaussian beam convolved map and the Cosine beam convolved one. When processing the Gaussian beam convolved map, the recovered angular power spectrum is essentially consistent with the initial H I angular power spectrum. Using the 2D power spectrum, we find that the values of $R(k_{\parallel}, k_{\perp})_{\text{auto}}$ [see Equation (16)] are close to 0 at large scales. It indicates that PCA foreground subtraction can recover part of the H I signal. While at small scales, the H I signal reduction is found, which is due to the beam smoothing effect.

However, when dealing with the Cosine beam convolved map, the angular power spectrum with three PCA modes subtraction deviates significantly from the initial H I angular power spectrum. With the 2D power spectrum, we find that there is a significant extra power at $k_{\parallel} \sim 0.7 \text{ Mpc}^{-1}$, which is relevant to the beam size ripple induced by the Cosine beam model. Therefore, PCA foreground subtraction fails in subtracting foregrounds in the Cosine beam convolved map. It indicates that the PCA foreground subtraction cannot remove the complex systematic effect induced by a sophisticated primary beam model, such as the Cosine beam model.

Additionally, we evaluate the concordance of the results with the cross-correlation power spectrum between the foreground removed H I map and the initial H I map. The values of $R(k_{\parallel}, k_{\perp})_{\text{cross}}$ [see Equation (17)] in both the Gaussian and Cosine beam convolved cases show a similar trend. The negative $R(k_{\parallel}, k_{\perp})_{\text{cross}}$ values at large k_{\perp} indicate the reduction of the correlation efficiency, which is mainly due to the beam smoothing effect.


Instead of directly using U-Net for foreground subtraction, we conservatively feed the PCA-subtracted residuals into the U-Net network. When dealing with the map convolved with the Cosine beam, the results are significantly improved. In particular, the deviation of the large-scale angular power spectrum is significantly removed, and the recovered angular power spectrum is consistent with the initial H I angular power spectrum. By studying the 2D power spectrum, we find that the values of $R(k_{\parallel}, k_{\perp})_{\text{auto}}$ are consistent in general between using the Gaussian and Cosine beam convolved maps. We quantify the results by defining a measure of concordance, as indicated by Equation (18). We find that, in the case of Gaussian beam, the concordance with the original H I map using U-Net is better than that using PCA by 27.4%, and in the case of Cosine beam, the concordance using U-Net is better than that using PCA by 144.8%. With the cross-correlation power spectrum, we find that $R(k_{\parallel}, k_{\perp})_{\text{cross}} \sim 0$, which indicates that the recovered H I signal is consistent with the initial H I signal. Therefore, the results show that the U-Net foreground subtraction method can eliminate the systematic effects caused by the telescope's primary beam.

In this work, we find that the U-Net-based foreground subtraction method can significantly improve the foreground subtraction efficiency, especially in the case of serious primary beam-induced systematic effect. U-Net is based on a fully convolutional network and its architecture was modified and extended to work with fewer training images and to yield more precise segmentation. Such a supervised deep-learning method relies on the completeness of the simulated training set. Because the detailed primary beam feature, as well as the systematic effect, can be different between different

experiments, an accurate beam model is required if a different experiment is considered. Such beam model can be provided using either “astroholographic” observation or EM simulation (Asad et al. 2021). Therefore, the U-Net-based foreground subtraction method sheds new light on recovering the H I power spectrum for future H I IM experiments, e.g., SKA and HERA (DeBoer et al. 2017).

This work was supported by the National Natural Science Foundation of China (grants Nos. 11975072, 11835009, 11875102, and 11690021), the Liaoning Revitalization Talents Program (grant No. XLYC1905011), the Fundamental Research Funds for the Central Universities (grant No. N2005030), the National 111 Project of China (grant No. B16009), and the Science Research Grants from the China Manned Space Project (grant No. CMS-CSST-2021-B01).

ORCID iDs

Shulei Ni (倪书磊)  <https://orcid.org/0000-0002-5386-1627>
 Yichao Li (李毅超)  <https://orcid.org/0000-0003-1962-2013>
 Li-Yang Gao (高立扬)  <https://orcid.org/0000-0001-5469-5408>
 Xin Zhang (张鑫)  <https://orcid.org/0000-0002-6029-1933>

References

- Abdalla, E., Ferreira, E. G. M., Landim, R. G., et al. 2021, arXiv:2107.01633
- Aghanim, N., Akrami, Y., Ashdown, M., et al. 2020, *A&A*, **641**, A6
- Alonso, D., Ferreira, P. G., & Santos, M. G. 2014, *MNRAS*, **444**, 3183
- Amiri, M., Bandura, K., Chen, T., et al. 2022, arXiv:2202.01242
- Anderson, C. J., Luci, N. J., Li, Y., et al. 2018, *MNRAS*, **476**, 3382
- Ansari, R., Campagne, J. E., Colom, P., et al. 2012, *A&A*, **540**, A129
- Asad, K. M. B., Girard, J. N., deVilliers, M., et al. 2021, *MNRAS*, **502**, 2970
- Bacon, D. J., Battye, R. A., Bull, P., et al. 2020, *PASA*, **37**, e007
- Bagla, J. S., Khandai, N., & Datta, K. K. 2010, *MNRAS*, **407**, 567
- Bandura, K., Addison, G. E., Amiri, M., et al. 2014, *Proc. SPIE*, **9145**, 22
- Battye, R. A., Browne, I. W. A., Dickinson, C., et al. 2013, *MNRAS*, **434**, 1239
- Battye, R. A., Davies, R. D., & Weller, J. 2004, *MNRAS*, **355**, 1339
- Bonaldi, A., & Brown, M. L. 2015, *MNRAS*, **447**, 1973
- Bull, P., Ferreira, P. G., Patel, P., & Santos, M. G. 2015, *ApJ*, **803**, 21
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. 2016, arXiv:1606.06650
- Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, *Natur*, **466**, 463
- Chang, T.-C., Pen, U.-L., Peterson, J. B., & McDonald, P. 2008, *PhRvL*, **100**, 091303
- Chen, X. 2012, *Int.J. Mod. Phys. Conf. Ser.*, **12**, 256
- DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, *PASP*, **129**, 045001
- Delabrouille, J., Betoule, M., Melin, J. B., et al. 2013, *A&A*, **553**, A96
- de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. 2008, *MNRAS*, **388**, 247
- Di Matteo, T., Perna, R., Abel, T., & Rees, M. J. 2002, *ApJ*, **564**, 576
- Fabian, I., & Klaus, M.-H. 2019, arXiv:1908.02182
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, **622**, 759
- Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, *A&AS*, **47**, 1
- Hothi, I., Chapman, E., Pritchard, J. R., et al. 2020, *MNRAS*, **500**, 2264
- Li, J., Zuo, S., Wu, F., et al. 2020, *SCPMA*, **63**, 129862
- Li, Y., Santos, M. G., Grainge, K., Harper, S., & Wang, J. 2021, *MNRAS*, **501**, 4344
- Lidz, A., Furlanetto, S. R., Oh, S. P., et al. 2011, *ApJ*, **741**, 70
- Liu, A., Parsons, A. R., & Trott, C. M. 2014, *PhRvD*, **90**, 023018
- Loeb, A., & Wyithe, S. 2008, *PhRvL*, **100**, 161301
- Makinen, T. L., Lancaster, L., Villaescusa-Navarro, F., et al. 2021, *JCAP*, **04**, 081
- Mao, Y., Tegmark, M., McQuinn, M., Zaldarriaga, M., & Zahn, O. 2008, *PhRvD*, **78**, 023529
- Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, *ApJL*, **763**, L20
- Matshawule, S. D., Spinelli, M., Santos, M. G., & Ngobese, S. 2021, *MNRAS*, **506**, 5075
- McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, *ApJ*, **653**, 815
- Mertens, F. G., Ghosh, A., & Koopmans, L. V. E. 2018, *MNRAS*, **478**, 3640
- Nan, R., Li, D., Jin, C., et al. 2011, *IJMPD*, **20**, 989
- Newburgh, L. B., Bandura, K., Bucher, M. A., et al. 2016, *Proc. SPIE*, **9906**, 99065X
- Olivari, L. C., Dickinson, C., Battye, R. A., et al. 2018, *MNRAS*, **473**, 4242
- Peel, M., Wuensche, C. A., Abdalla, E., et al. 2019, *JAI*, **8**, 1940005
- Pritchard, J. R., & Loeb, A. 2008, *PhRvD*, **78**, 103511
- Reddi, S. J., Kale, S., & Kumar, S. 2019, arXiv:1904.09237
- Ronneberger, O., Fischer, P., & Brox, T. 2015, arXiv:1505.04597
- Santos, M. G., Bull, P., Alonso, D., et al. 2015, PoS, AASKA14, 019
- Santos, M. G., Cluver, M., Hilton, M., et al. 2017, arXiv:1709.06099
- Santos, M. G., Cooray, A., & Knox, L. 2005, *ApJ*, **625**, 575
- Seo, H.-J., Dodelson, S., Marriner, J., et al. 2010, *ApJ*, **721**, 164
- Soares, P. S., Watkinson, C. A., Cunningham, S., & Pourtsidou, A. 2022, *MNRAS*, **510**, 5872
- Spinelli, M., Carucci, I. P., Cunningham, S., et al. 2022, *MNRAS*, **509**, 2048
- Switzer, E. R., Chang, T.-C., Masui, K. W., Pen, U.-L., & Voytek, T. C. 2015, *ApJ*, **815**, 51
- Switzer, E. R., Masui, K. W., Bandura, K., et al. 2013, *MNRAS*, **434**, L46
- Villanueva-Domingo, P., & Villaescusa-Navarro, F. 2021, *ApJ*, **907**, 44
- Wang, J., Santos, M. G., Bull, P., et al. 2021, *MNRAS*, **505**, 3698
- Wolz, L., Blake, C., Abdalla, F. B., et al. 2017, *MNRAS*, **464**, 4938
- Wolz, L., Pourtsidou, A., Masui, K. W., et al. 2022, *MNRAS*, **510**, 3495
- Wu, F., Li, J., Zuo, S., et al. 2021, *MNRAS*, **506**, 3455
- Wuensche, C. A., Villela, T., Abdalla, E., et al. 2021, arXiv:2107.01634
- Wyithe, S., & Loeb, A. 2008, *MNRAS*, **383**, 606
- Wyithe, S., Loeb, A., & Geil, P. 2008, *MNRAS*, **383**, 1195