Deep-Learning Study of the 21-cm Differential Brightness Temperature During the Epoch of Reionization

Yungi Kwon
Department of Physics, University of Seoul, Seoul 02504, Korea

Sungwook E. Hong∗
Natural Science Research Institute, University of Seoul, Seoul 02504, Korea

Inkyu Park
Department of Physics, University of Seoul, Seoul 02504, Korea and
Natural Science Research Institute, University of Seoul, Seoul 02504, Korea

(Received 28 April 2020; revised 26 May 2020; accepted 10 June 2020)

We propose a deep learning analysis technique with a convolutional neural network (CNN) to predict the evolutionary track of the Epoch of Reionization (EoR) from the 21-cm differential brightness temperature tomography images. We use 21cmFAST, a fast semi-numerical cosmological 21-cm signal simulator, to produce mock 21-cm maps between \( z = 6 \sim 13 \). We then apply two observational effects, such as instrumental noise and limit of (spatial and depth) resolution somewhat suitable for realistic choices of the Square Kilometre Array (SKA), into the 21-cm maps. We design our deep learning model with CNN to predict the sliced-averaged neutral hydrogen fraction from the given 21-cm map. The estimated neutral fraction from our CNN model has great agreement with the true value even after coarsely smoothing with broad beam size and frequency bandwidth and heavily covered by noise with narrow beam size and frequency bandwidth. Our results show that the deep learning analyzing method has the potential to reconstruct the EoR history efficiently from the 21-cm tomography surveys in future.

PACS numbers: 32.80.Fb, 95.85.Bh, 98.58.Ge, 98.70.Vc, 98.80.Es
Keywords: Epoch of reionization, Deep learning
DOI: 10.3938/jkps.77.49

I. INTRODUCTION

About 380,000 years after the Big Bang, the Universe had cooled down enough that the floating free protons and electrons could combine to form neutral hydrogen atoms (HI). As no luminous objects such as stars and galaxies existed, this period is referred to as the cosmic dark ages. During the cosmic dark ages, most of the hydrogen atoms were neutral. Then, by the gravitational collapse of overdense regions in the intergalactic medium (IGM), more and more pronounced structures of the Universe started to form. Eventually, the first luminous objects started to form, which led to the epoch of reionization (EoR). During this epoch, the first radiating objects in the Universe heated and re-ionized the surrounding local neutral IGM by energetic radiation.

As the EoR involves the formation of the first objects and large-scale structures, numerous studies have performed to understand the epoch. For example, several observations have constrained the duration of the EoR from redshifts \( z \sim 12 \sim 6 \) [1–5]. However, due to the lack of bright objects, the current observational evidence of the EoR is rather limited. For instance, we poorly understand the main source governing the EoR, the properties leading to the ionizing process, and how all of this intimately affected the formation of subsequent structures.

Several observational methods have been suggested to understand the EoR. For example, how the cosmic microwave background (CMB) anisotropy observations could constrain the evolutionary history of the mean neutral hydrogen fraction \((\bar{x}_{\text{HI}}(z))\) was studied in Refs. 6–8. Among various observational methods for studying the EoR, one of the most promising ways is to use the 21-cm wavelength radiation emitted by the hyperfine transition of neutral hydrogen atoms, as hydrogen gas was the predominant component of the Universe during that era. Several radio-telescope experiments are being planned for observing the redshifted 21-cm signal from the EoR: the Murchison Widefield Array (MWA) [9], the Giant
Mettrewave Radio Telescope (GMRT) [10], the Low Frequency Array (LOFAR) [11], the Precision Array Probing the Epoch of Reionization (PAPER) [12], the Hydrogen Epoch of Reionization Array (HERA) [13], and the Square Kilometre Array (SKA) [14,15]. Nevertheless, due to the foregrounds from the extra-galactic radio sources and diffuse galactic foreground, ionospheric distortions, instrumental noise, etc., a challenge remains to extract the proper redshifted 21-cm signal during the EoR from radio observations. Specifically, Galactic synchrotron radiation, which has a wavelength similar to that of the redshifted 21-cm signal during the EoR, is expected to be much brighter than the cosmological 21-cm signal [16–18].

While numerous methods for subtracting foregrounds and analyzing the 21-cm signals have been suggested [19–22], brand-new analysis techniques have increasingly approved recently. One of them is deep learning, a subset of machine learning that works with artificial neural networks (ANN) and designed to imitate how humans think and learn by using so-called neurons. The neural network accepts a series of data as an input and looks for a certain patterns within the data by fitting the weights on the connections between the neurons of the network. The deep learning technique has been shown to have enormous potential in astrophysics [23–29]. For example, in Ref. 30, the deep learning technique was used to estimate the ionizing efficiency, the minimum virial temperature of halos, and the mean free path of ionizing photons by using the 21-cm power spectra at different redshifts as inputs. Also, in Ref. 31, the entire 21-cm lightcone map was used as the input of the convolutional neural network (CNN) to discriminate the re-ionizing process from the active galactic nuclei (AGNs) and star-forming galaxies. To date, however, a series of studies has been done to deal with somewhat ideal simulated data that poorly considered actual observational effects.

In this paper, we introduce a novel deep learning-based method by re-constructing the mean neutral hydrogen fraction, \( \delta \), during the EoR from mock 21-cm differential brightness temperature maps by considering various observational effects. We simulate 21-cm maps by using a semi-numerical simulation code 21cmFAST [32] at different redshifts. We convolve them with a certain beam size and frequency bandwidth under somewhat realistic observational conditions, and add the corresponding white Gaussian noise to generate noisy maps. We use those noisy 21-cm maps as inputs of our CNN model and train the model to predict the corresponding value of \( \delta \).

This paper is organized as follows: In Sec. II, we describe how to simulate the mock 21-cm differential brightness temperature maps. In Sec. III, we introduce our CNN architecture and training method. In Sec. IV, we show the performance of our method under various observational conditions. Finally, in Sec. V, we summarize our results. Throughout this paper, we adopt the background cosmological parameters best fit to the standard Lambda cold dark matter (LCDM) cosmology from Planck 2018 [33]: matter density parameter \( \Omega_m = 0.31 \), baryon density parameter \( \Omega_b = 0.048 \), cosmological constant density parameter \( \Omega_{\Lambda} = 0.69 \), spectral index \( n_s = 0.97 \), rms density fluctuation \( \sigma_8 = 0.81 \), Hubble parameter \( h = 0.68 \), and the primordial helium abundance \( Y_p = 0.245 \).

II. SIMULATION OF NOISY MOCK 21-CM MAPS

1. Simulating Undistorted Redshifted 21-cm Signals

We use the 21cmFAST [32] to generate the 21-cm signal and corresponding neutral fraction with the redshifts \( z = 6 \sim 13 \), where most of the current studies strongly have been constrained. The 21cmFAST, which is self-consistent and semi-numerical, is a publicly available simulation tool for the cosmological reionization. Specifically, this is optimized to generate the 21-cm signal during the epoch. By combinations of the excursion-set formalism [34,35], which was mentioned earlier for identifying ionized hydrogen regions and the first-order perturbation theory [36], full three-dimensional realizations of the density, ionization field, velocity, spin temperature, and finally the 21-cm differential brightness temperature with a given redshift are generated.

We first generate the density and the velocity fields with 50, 100, 200, and 300 Mpc comoving box-sizes and 800\(^3\), 1000\(^3\), and 1250\(^3\) grids between \( z = 6 \sim 13 \). We use different random seeds to generate the initial conditions for each combination of comoving box-size and grid number, and, as a result, we get 12 independent simulation sets. For each simulation set, we obtain \( 18 \sim 35 \) snapshots between \( z = 6 \sim 13 \), where the number of snapshots depends on the number of grids. Note that the spatial resolutions used in this paper (\( \Delta x = 0.04 \sim 0.375 \) Mpc) are smaller than the length scale on which the density perturbation from the 21cmFAST agrees well with the N-body simulations [32]. However, since we do not focus on small-scale details and the output 21-cm maps will be smoothed at larger scales, our choice of spatial resolution does not severely affect our result.

Then, we calculate the “undistorted” 21-cm differential brightness temperature, \( \delta T_b^0(\mathbf{x}, z) \), by downgrading the resolution to 200\(^3\) grids:

\[
\delta T_b^0 \approx (27 \text{ mK}) \delta \nu_R \left(1 + \delta \right) \left( \frac{H}{\text{d}v_R/\text{d}r + H} \right) \times \left(1 - \frac{T_{\text{CMB}}}{T_S} \right) \left(1 + z \frac{0.15}{10 \Omega_m h^2} \right)^{1/2} \left( \frac{\Omega_b h^2}{0.023} \right), \tag{1}
\]

where \( \delta \), \( T_{\text{CMB}} \), and \( T_S \) are the gas overdensity, the CMB temperature, and the gas spin temperature, respectively.
[17]. In 21cmFAST, \( T_S(z) \) is calculated from the evolution of the kinetic gas temperature and the Ly-\( \alpha \) background. Also, \( x_{\text{HI}}(\mathbf{x}, z) \) is calculated from the number of IGM ionizing photons per baryon [37]:

\[
n_{\text{ion}} = \bar{\rho}_b^{-1} \int_0^\infty \frac{dM_h}{dM_h} \frac{dn(M_h, z)}{dM_h} f_{\text{duty}} M_\star f_{\text{esc}} N_{\gamma/b},
\]

(2)

where \( \bar{\rho}_h \), \( M_h \), \( f_{\text{duty}} \), \( M_\star \), \( f_{\text{esc}} \), \( N_{\gamma/b} \) are the mean baryon density, halo mass, duty cycle related to the suppression of star formation at massive halos, stellar mass, escape fraction, and the number of ionizing photons per stellar baryon, respectively. \( f_{\text{duty}} \), \( M_\star \), and \( f_{\text{esc}} \) can be expressed, respectively, as functions of \( M_h \) as follows:

\[
f_{\text{duty}}(M_h) = \exp\left(-\frac{M_h}{M_{\text{turn}}}\right),
\]

(3)

\[
M_\star(M_h) = f_* \frac{\Omega_b}{\Omega_m} M_h = f_{*,10} \left(\frac{M_h}{10^{10} M_\odot}\right) \Omega_b \Omega_m M_h,
\]

(4)

\[
f_{\text{esc}}(M_h) = f_{\text{esc},10} \left(\frac{M_h}{10^{10} M_\odot}\right)^{\alpha_{\text{esc}}},
\]

(5)

where \( M_{\text{turn}} \) is the halo mass threshold for efficient star formation, and \( f_* \) is the fraction of galactic gas in stars. In this paper, we adopt the fiducial setup from Ref. 37 for reionization parameters: \((N_{\gamma/b}, f_{*,10}, \alpha_*, f_{\text{esc},10}, \alpha_{\text{esc}}, M_{\text{turn}}) = (5000, 0.05, 0.5, 0.1, -0.5, 5 \times 10^9 M_\odot)\). Note that we do not test different values of the reionization parameters in this paper mainly because we focus on the evolution of the mean neutral hydrogen fraction, \( \bar{x}_{\text{HI}}(z) \), rather than constraining the reionization model.

In 21cmFAST, \( T_S(z) \) is calculated from the evolution of the kinetic gas temperature and the Ly-\( \alpha \) background. Also, \( x_{\text{HI}}(\mathbf{x}, z) \) is calculated from the number of IGM ionizing photons per baryon [37]:

\[
n_{\text{ion}} = \bar{\rho}_b^{-1} \int_0^\infty \frac{dM_h}{dM_h} \frac{dn(M_h, z)}{dM_h} f_{\text{duty}} M_\star f_{\text{esc}} N_{\gamma/b},
\]

(2)

where \( \bar{\rho}_h \), \( M_h \), \( f_{\text{duty}} \), \( M_\star \), \( f_{\text{esc}} \), \( N_{\gamma/b} \) are the mean baryon density, halo mass, duty cycle related to the suppression of star formation at massive halos, stellar mass, escape fraction, and the number of ionizing photons per stellar baryon, respectively. \( f_{\text{duty}} \), \( M_\star \), and \( f_{\text{esc}} \) can be expressed, respectively, as functions of \( M_h \) as follows:

\[
f_{\text{duty}}(M_h) = \exp\left(-\frac{M_h}{M_{\text{turn}}}\right),
\]

(3)

\[
M_\star(M_h) = f_* \frac{\Omega_b}{\Omega_m} M_h = f_{*,10} \left(\frac{M_h}{10^{10} M_\odot}\right) \Omega_b \Omega_m M_h,
\]

(4)

\[
f_{\text{esc}}(M_h) = f_{\text{esc},10} \left(\frac{M_h}{10^{10} M_\odot}\right)^{\alpha_{\text{esc}}},
\]

(5)

where \( M_{\text{turn}} \) is the halo mass threshold for efficient star formation, and \( f_* \) is the fraction of galactic gas in stars. In this paper, we adopt the fiducial setup from Ref. 37 for reionization parameters: \((N_{\gamma/b}, f_{*,10}, \alpha_*, f_{\text{esc},10}, \alpha_{\text{esc}}, M_{\text{turn}}) = (5000, 0.05, 0.5, 0.1, -0.5, 5 \times 10^9 M_\odot)\). Note that we do not test different values of the reionization parameters in this paper mainly because we focus on the evolution of the mean neutral hydrogen fraction, \( \bar{x}_{\text{HI}}(z) \), rather than constraining the reionization model.

Figure 1 shows an example of undistorted 21-cm maps in a 50-cMpc-box. As expected in Eq. (1) (if \( T_S \gg T_{\text{CMB}} \), which is believed to be satisfied during the main reionization epoch), the undistorted 21-cm maps are proportional to the matter density and follow a Gaussian distribution at high \((z, \bar{x}_{\text{HI}})\). On the other hand, at low \((z, \bar{x}_{\text{HI}})\), the overdense region becomes fully ionized and shows no 21-cm signal; therefore, the distribution is highly non-Gaussian. Note that the estimate of \( \bar{x}_{\text{HI}} \) from the undistorted 21-cm maps should be rather straightforward as it is simply the fraction of area where \( x_{\text{HI}}(\mathbf{x}) \) is close to zero. However, the estimate of \( \bar{x}_{\text{HI}} \) may not be that straightforward if one adds observational effects to the 21-cm signal, as we will see in the next subsection.

2. Processing with the Observational Effects

In practice, we expect the observed signal to be very different from those in Fig. 1, because the actual observed signal will be obtained by integrating the angle, frequency, and observation time. At this point, the sizes of the angle and the frequency are the limits of the spatial and depth resolutions of the telescope. As the size of the angle or frequency of the telescope becomes larger, the observed signal becomes more smoothed. Then, small-scale information is lost and the estimate of \( \bar{x}_{\text{HI}} \), as we mentioned in the previous subsection, may fail. On the other hand, while the signal can contain small-scale information at a smaller angle or frequency, if the integration...
Fig. 3. Top panel: RMS smoothed mock 21-cm maps as a function of mean neutral hydrogen fraction. Bottom panel: sensitivity limits with 10-, 100-, 1000-, and 10000-hour integration times of SKA (solid lines, from top to bottom: [39]) compared to the maximum RMS smoothed mock 21-cm maps over time (peak value at the top panel; crosses). From left to right: $\Delta \theta = 1'$, $2'$, and $3'$. Colors: $\Delta \nu = 0.2$ MHz(red), 1 MHz(green), and 2 MHz(blue).

time is fixed, the level of noise becomes higher, and estimating $x_{\text{HI}}$ may be difficult. Therefore, for performance tests of analysis methods such as our CNN model, observational effects, i.e., smoothing and noise, must be included.

We start by smoothing the undistorted 21-cm maps for a given angle and frequency setup of the radio telescope. Here, we use three choices of the beam sizes and frequency bandwidths—$\Delta \theta = 1'$, $2'$, and $3'$, and $\Delta \nu = 0.2$ MHz, 1 MHz, and 2 MHz. The corresponding comoving length scales for given $\Delta \theta$ and $\Delta \nu$ are

$$\Delta L_\perp (\Delta \theta, z) = D_c(z) \times \Delta \theta,$$

$$\Delta L_\parallel (\Delta \nu, z) = \frac{c(1+z)^2}{H_0 \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda} \nu_0} \Delta \nu,$$

where $D_c(z)$ is the comoving distance and $\nu_0 = 1420$ MHz is the rest-frame frequency of the 21-cm line. Note that the subscripts ‘$\perp$’ and ‘$\parallel$’ emphasize that the beam size and the frequency bandwidth affect the smoothing of $\delta T_b$ in directions perpendicular and parallel to the line of sight (LOS), respectively.

The actual shape of the smoothing kernels in directions parallel and perpendicular to the LOS ($W(\nu; \Delta \nu)$ and $W(\theta; \Delta \theta)$, respectively) strongly depends on the configuration of the radio telescope, even for identical choices of $(\Delta \theta, \Delta \nu)$. Especially, the shape of $W(\theta; \Delta \theta)$ (beam shape) can vary greatly, depending on the antennas design. Some “dirty” beam shape, such as the compensated Gaussian kernel, is known to be possible [38], and in that case, the understanding of $x_{\text{HI}}(z)$ could be extremely difficult [39]. Nevertheless, for simplicity and because the configuration of future surveys is not fixed, we assume a Gaussian kernel for both the beam and the frequency smoothing kernels in this paper. We use $\Delta \theta$ and $\Delta \nu$, or identically, $\Delta L_\perp$ and $\Delta L_\parallel$, as the full width at half maximum of the beam and the frequency smoothing kernel.

Also, because interferometric radio observations might be insensitive to the large-scale fluctuation of $\delta T_b$, we subtract the mean of the smoothed 21-cm signal from
the map so that the average of the subtracted 21-cm map becomes zero. In summary, we obtain the “smoothed” 21-cm map (or, $\delta T_b^z(\theta, \nu)$, where $\nu = \nu_0/(1 + z)$), as follows:

$$
\delta T_b^z(\theta, \nu) = \delta T_b^z(\theta, \nu) - \langle \delta T_b^z(\theta', \nu) \rangle_{\theta'},
$$

(8)

$$
\delta T_b^z(\theta, \nu) = \int d\nu' \int d\theta' \delta T_b^z(\theta', \nu') \times W_G(|\theta - \theta'|; \Delta \theta) W_G(\nu - \nu'; \Delta \nu).
$$

(9)

Here, the subscript ‘G’ emphasizes the Gaussian kernel.

Figure 2 shows an example of the smoothed mock 21-cm map in a 300-cMpc box at $(z, \bar{x}_{\text{HI}}) = (12.73, 0.97)$. Compared to the undistorted 21-cm maps in Fig. 1, the smoothed maps show two key features that make the estimate of $\bar{x}_{\text{HI}}(z)$ more difficult. First of all, due to the mean subtraction after smoothing, we cannot use the mean value $\langle \delta T_b^z(\theta, \nu) \rangle_{\theta}$ as an indicator of $\bar{x}_{\text{HI}}(z)$, especially for high-$z$ cases (see the bottom-left panel of Fig. 1 as a comparison). Additionally, smoothing in both the parallel and the perpendicular directions to the LOS wipes out the small-scale features. As a result, without prior knowledge, the heavily smoothed 21-cm map at the lower-right panel of Fig. 2 might be confused to be at the middle of EoR (e.g., $\bar{x}_{\text{HI}}(z) \sim 0.5$). Note that, however, the spatial distribution of the smoothed 21-cm maps is Gaussian and highly non-Gaussian at $\bar{x}_{\text{HI}} \simeq 1$ and $1 - \bar{x}_{\text{HI}} \gg 0$, respectively. Therefore, distinguishing between the smoothed 21-cm maps before and the middle of EoR may be possible by using n-point correlation functions or Minkowski functionals [39,40], even though such methods usually require precise measurements of $\delta T_b^z$.

After smoothing, we add noise whose level depends on the integration time and the configuration of radio telescope to the mock 21-cm maps. Here, we adopt the sensitivity limit used in Ref. 39, which uses the sensitivity calculation from Ref. 41 by assuming the SKA with a core size of $\sim 1$ km. While the spatial and the intensity distributions of noise and how it convolves with the signal strongly depend on the configuration of the radio telescope [42], here we adopt a somewhat simple model as follows: We calculate the root-mean-square (RMS) noise level $\langle \sigma(DT; \Delta \theta, \Delta \nu) \rangle$ by using 10-, 100-, 1000-, and 10000-hour integration times and produce Gaussian white noise with a given RMS level and random seeds ($\epsilon$). Then, we obtain the “noisy” 21-cm map (or, $\delta T_b^z(\theta, \nu)$) by adding the noise to the smoothed 21-cm map:

$$
\delta T_b^z(\theta, \nu) = \delta T_b^z(\theta, \nu) + \sigma(\Delta T; \Delta \theta, \Delta \nu, \epsilon).
$$

(10)

Note that the mean value of the noisy 21-cm map remains zero.

Figure 3 shows a comparison between our sensitivity limit estimate and the RMS value of the smoothed mock 21-cm maps ($\delta T_{b,\text{rms}}$) for various beam sizes, frequency bandwidths, and integration times. Here, $\delta T_{b,\text{rms}}/\sigma$ can be regarded as a proxy of the signal-to-noise ratio (SNR). As commented earlier, to fully utilize the advantages of some n-points or topological analysis methods, one needs to achieve $\text{SNR} \gg 1$. For marginal choices of the beam size and the frequency bandwidth, e.g., $\Delta \theta = 1' \sim 2'$ and $\Delta \nu = 1 \sim 2$ MHz, such a SNR is possible only with a $\gtrsim$10000-hour SKA integration time [39].

Figure 4 further shows how adding noise affects the same mock 21-cm maps in Fig. 2, by assuming a 1000-hour SKA integration time. At large beam size and frequency bandwidth (e.g., bottom-right panel of Fig. 4), the RMS values of the smoothed mock 21-cm maps are higher than the sensitivity limit. Also, the overall distribution of the noisy 21-cm map is similar to that of the smoothed 21-cm map. On the other hand, at small beam size and frequency bandwidth (e.g., top-left panel of Fig. 4), the RMS value is about an order of magnitude smaller than the sensitivity limit. As a result, without an additional denoising process, sophisticated analysis methods that rely on a detailed spatial distribution of $\delta T_b^z$ might have difficulties for such maps.

Note that, however, one can see a visual pattern of patches with generally positive and negative values of $\delta T_b^z$ at the top-left panel of Fig. 4, even before any denoising process. Also, one can see that such a visual pattern is similar to the clearer patterns shown at other panels (i.e., noisy 21-cm maps with larger beam size and frequency bandwidth). However, the mathematical definition of such a “visual pattern” may not be straightforward, while binning and Gaussian smoothing can be candidates of such mathematical definition. Note that such recognition of mathematically unclear visual patterns, which we will see in the next section, is one of the strongest specialties of the deep learning technique.
III. DEEP LEARNING PREDICTION OF THE MEAN NEUTRAL HYDROGEN FRACTION

1. Deep Learning Architecture

We use the convolutional neural network (CNN) optimized to recognize visual patterns within the image dataset while efficiently lessening the computational cost for expressing complex non-linear function for multi-dimensional input data. Figure 5 and Table 1 show our CNN architecture design. The entire network is divided into two parts, feature extraction and (fully-connected; FC) linear regression. All the filters in the first part of the architecture have the same size (3,3) and are employed to extract the feature of the processed 21-cm map through a series of convolutions. Before each convolution layer, we use the max pooling (MaxPool) layer. The MaxPool layer helps to ease operations needed over the training by reducing the features resulting from convolution one and can improve the result by keeping the network from being over-fitted with a represented value (c.f. maximum or mean) of a particular feature over from the local regions, without more feature maps than necessary. Finally, the convolution and the pooling layers are followed by a linear regression part, which consists of three FC layers. The FC layers employ the high-level features from the first part of the CNN and help to teach the network produce a value close to the desired output value through the last fully-connected layer activated by using a linear function.

We apply the rectified linear unit (ReLU), which is defined as \( \text{ReLU}(s) = \max(0, s) \), as the activation functions of all convolutional and FC layers [43,44]. The weights on the fully-connected layers are initialized by applying a He uniform initialization [45], which extracts \( n_{in} \) samples from the weights initialized from a uniform distribution between \(-\sqrt{6}/n_{in}\) and \(+\sqrt{6}/n_{in}\). The He uniform initialization for the weights technique helps in reaching a global minimum of the cost function faster and more effectively.

Before each ReLU activation, we also add batch normalization (BatchNorm) layers [46]. For the situation where the inputs are fed with forwarding to the output layer, the distributions of the inputs keep changing with each iterations because the weights of the layers are changed over the training of the network. Thus, the intermediate layers have difficulty to adopting the changing distributions of inputs. To prevent such problems, BatchNorm renormalizes a mini-batch of inputs \( x_i \) into \( y_i \) as

\[
y_i = \gamma \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta,
\]

where \( \mu \) and \( \sigma \) are the mean and the standard devi-
Table 1. Outline of the convolutional neural network used for training.

| Layer | # Filters | Filter Size, Stride | Padding | Output Dimension |
|-------|-----------|---------------------|----------|------------------|
| Input | -         | -                   | -        | (3, 200, 200)    |
| Conv2D-1 | 32 | (3,3), (1,1) | valid | (32, 198, 198) |
| BatchNorm-1 (ReLU) | - | - | - | (32, 198, 198) |
| MaxPool2D-1 | 1 | (2,2), (2,2) | valid | (32, 99, 99) |
| Conv2D-2 | 32 | (3,3), (1,1) | valid | (32, 97, 97) |
| BatchNorm-2 (ReLU) | - | - | - | (32, 97, 97) |
| MaxPool2D-2 | 1 | (2,2), (2,2) | valid | (32, 48, 48) |
| Conv2D-3 | 64 | (3,3), (1,1) | valid | (64, 46, 46) |
| BatchNorm-3 (ReLU) | - | - | - | (64, 46, 46) |
| MaxPool2D-3 | 1 | (2,2), (2,2) | valid | (64, 23, 23) |
| Flatten | - | - | - | 33856 |
| FC-1 (He-uniform) | - | - | - | 64 |
| BatchNorm-4 (ReLU) | - | - | - | 64 |
| FC-2 (He-uniform) | - | - | - | 32 |
| BatchNorm-5 (ReLU) | - | - | - | 32 |
| FC-3 (linear) | - | - | - | 1 |

Table 2. Color-map scheme used to generate input images from the noisy mock 21-cm maps. Colors for \( \delta T^b \) values in between are determined using linear interpolation.

| \( \delta T^b \) [mK] | Color Name | Red \([0 \sim 1]\) | Green \([0 \sim 1]\) | Blue \([0 \sim 1]\) |
|----------------------|------------|----------------|----------------|----------------|
| −210 | Yellow | 1 | 1 | 0 |
| −105 | Red | 1 | 0 | 0 |
| 0 | Black | 0 | 0 | 0 |
| 15 | Green | 0 | 0.5 | 0 |
| 30 | Blue | 0 | 0 | 1 |

Not only can BatchNorm effectively complement some chronic problems, such as gradient vanishing problem, it allows a much higher learning rate to be used, and it can be used as a regularizer for a part.

2. Training

For the CNN training, first of all, we produce samples by splitting the noisy 21-cm maps with 200\(^3\) grids into 67 slices of 200 × 200 pixels along the LOS direction. We calculate the mean neutral hydrogen fraction for each slice. Note that, while we only use 18 ~ 35 snapshots between \( z = 6 \sim 13 \), \( \bar{x}_{HI} \) estimated from different slices at the same redshift may vary. As a result, for a given \((\Delta \theta, \Delta \nu, \Delta T)\), we have 1206 ~ 2345 samples for each box size and high-resolution grid setup. Note that the difference in the angular field of view from a fixed box size between \( z = 6 \sim 13 \) is less than 13%, which we will consider as negligible. We have also found that the performance is similar even if one uses 10318 samples from all 9 box-size- and high-resolution-grid setups; therefore, we use all 10318 samples hereafter.

We produce the inputs of our CNN architecture by converting each 200 × 200 pixel of a \( \delta T^b \) slice into a 3-channel RGB image \((i.e.,\ an\ array\ with\ (3, 200, 200)\ size)\) by applying a custom colormap scheme (see Table 2 for the definition and Fig. 6 for examples). We have also tested using the 200 × 200 slice directly as an input and found no notable difference in performance.

We then split our samples into three sets: training set used for training the model parameters, validation set used for checking the training process, and test set for evaluating the performance of the trained model. For a given \((\Delta \theta, \Delta \nu, \Delta T)\), we split 10318 samples into 7428 training samples, 826 validation samples, and 2064 test samples. We randomly split the samples into three sets while keeping the distributions of \( \bar{x}_{HI} \) similar.

We compile our CNN model with an Adam optimizer [47] and a learning rate \(10^{-3}\). Also, we adopt the mean squared error (MSE) as the loss function:

\[
MSE = \frac{1}{m} \sum_{i=1}^{m} (\bar{x}_{HI,\text{pred}} - \bar{x}_{HI,\text{true}})^2 .
\]
Here, $m$ is the minibatch size for training, which is set to 8. We have also tested several minibatch sizes between 4 and 64 and found no notable difference in performance for $m \geq 8$.

We define an epoch as an iteration of fitting the model parameters from 7428 training samples and evaluating the training loss, as well as validation loss, from 826 validation samples. Because we set the minibatch size to 8, the number of minibatches per epoch is 929. For a single run of CNN training, we set the maximum number of epochs as 500. However, for saving time and preventing overfitting, we stop the training when the validation loss is not improved for 30 epochs, and the typical number of epochs becomes $\sim 200$. We then use the model at the epoch where the validation loss is minimized. We perform our training with Keras [48] with Tensorflow GPU version a back end [49], and each run takes $\sim 2$ hours with a single NVIDIA V100 GPU card in a NVIDIA DGX-1 GPU platform.

Furthermore, we make use of the cross-validation (CV) technique to test the stability of a certain deep learning model. In the CV technique, the entire dataset is split into $K$-folds, and multiple CNN models are trained by using a certain fold as a validation set and the remaining ($K - 1$) folds as a training set. We use the shuffled 10-folded CV so that we train our network 5 times with the number of epochs and data having a different distribution for each time. We have found that the validation loss from different models agrees within less than a 75% deviation from their average.

IV. RESULTS

The bottom panel of Fig. 5 shows how our CNN model processes the input noisy 21-cm map during feature extraction. Numerous feature maps are similar to the black-and-white segmented images of the input maps with different intensity thresholds and small-scale noises. This means that our CNN model allows a strategy similar to the most straightforward estimate method of $\langle x_{\text{HI}} \rangle$, i.e., measuring the area fraction of zero $\delta T_b^0(\theta)$. Allowing different intensity thresholds is mainly because the value of $\delta T_b^0(\theta)$ at $\delta T_b^0(\theta) = 0$ is uncertain due to the renormalization of the mean value. Also, allowing different small-scale noises may be related to denoising the noisy 21-cm map in several ways. Note that the variations in the intensity thresholds and the small-scale noises depend on $(\Delta \theta, \Delta \nu, \Delta T)$.

Figure 6 shows the noisy mock 21-cm maps for the two extreme choices of beam size and frequency bandwidth. The top panel of Fig. 6 corresponds to large values of the beam size and frequency bandwidth. The observational noise is relatively small, the entire 21-cm map is too smoothed so that the 21-cm maps from three characteristic epochs ($x_{\text{HI}} \sim 1, 0.5, \text{and } 0$) look similar. On the other hand, the bottom panel, which comes from small $(\Delta \theta, \Delta \nu)$, has a more observational noise than the signal itself, which also makes the 21-cm maps from the three epochs hard to distinguish visually. Interestingly, our CNN method can clearly distinguish such noisy 21-cm maps at different epochs with a prediction error of $x_{\text{HI}}$ of less than $\sim 0.03$.

Figure 7 and Table 3 show the prediction power of our CNN method for different choices of the beam size and the frequency bandwidth by assuming a 1000-hour integration time of SKA. Overall, the predicted mean neu-
Table 3. Summary of the absolute prediction error $|\bar{x}_{\text{HI,pred}} - \bar{x}_{\text{HI,true}}|$ from our CNN method. We assume 1000-hour integration time of SKA. The best and the worst cases are marked in bold-faced type and are underlined, respectively.

| Beam size | Bandwidth | Median $1\sigma$-high | 2$\sigma$-high | Maximum |
|-----------|-----------|------------------------|----------------|---------|
|            | 0.2 MHz   | 0.0090                 | 0.0198         | 0.0409  |
| 1'         | 1 MHz     | 0.0087                 | 0.0197         | 0.0474  |
|            | 2 MHz     | 0.0100                 | 0.0227         | 0.0687  |
| 2'         | 0.2 MHz   | 0.0092                 | 0.0202         | 0.0490  |
|            | 1 MHz     | 0.0082                 | 0.0169         | 0.0349  |
|            | 2 MHz     | 0.0082                 | 0.0184         | 0.0544  |
| 3'         | 0.2 MHz   | 0.0091                 | 0.0212         | 0.0504  |
|            | 1 MHz     | 0.0075                 | 0.0162         | 0.0429  |
|            | 2 MHz     | 0.0075                 | 0.0178         | 0.0616  |

Table 4. Same as Table 3, but by varying $(\Delta \theta, \Delta T)$ while fixing $\Delta \nu = 1$ MHz.

| Beam size | Integration Time | Median $1\sigma$-high | 2$\sigma$-high | Maximum |
|-----------|------------------|------------------------|----------------|---------|
| 2'        | 1000 hours       | 0.0082                 | 0.0169         | 0.0349  |
|           | 100 hours        | 0.0416                 | 0.118          | 0.746   |
| 3'        | 1000 hours       | 0.0075                 | 0.0162         | 0.0429  |
|           | 100 hours        | 0.0227                 | 0.0530         | 0.163   |

Table 4. Same as Table 3, but by varying $(\Delta \theta, \Delta T)$ while fixing $\Delta \nu = 1$ MHz.

Fig. 8. Same as Fig. 7, but by varying $(\Delta \theta, \Delta T)$ while fixing $\Delta \nu = 1$ MHz. From left to right: $\Delta \theta = 2'$ and $3'$. From top to bottom: $\Delta T = 1000$ hours and 100 hours.

Deep-Learning Study of the 21-cm Differential Brightness Temperature · · · – Yungi Kwon et al. -57-

gration time of SKA. As the noise level is about an order of magnitude higher than the actual signal, a significant amount of deviation exists even for a large beam size of $\Delta \theta \gtrsim 3'$. Furthermore, for $\Delta \theta \lesssim 2'$ two populations with $(\bar{x}_{\text{HI,true}}, \bar{x}_{\text{HI,pred}}) \approx (0, 1)$ and vice versa are clearly observed. This means that the CNN model fails to distinguish between the 21-cm maps before EoR and those at the late stage of EoR — in other words, the number of “classification” categories that our CNN model with a large noise level tends to find becomes extremely low. Note that the SNR at $(\Delta \theta, \Delta \nu, \Delta T) = (2', 1$ MHz, 100 hours) is similar to that at $(1', 1$ MHz, 100 hours), where the prediction accuracy of $\bar{x}_{\text{HI}}$ is reasonably good. This emphasizes that, because the distribution of the 21-cm map is highly non-Gaussian in general, the SNR (or, $\delta T_{\text{rms}}/\sigma$) alone cannot fully determine the prediction power of our CNN method. Further study on the deep learning architecture or the combination with other analysis methods might be helpful to enhance the prediction power for $\bar{x}_{\text{HI}}$, which is beyond our scope.

V. CONCLUSIONS

In this paper, we introduced a novel convolutional neural network (CNN)-based deep learning technique for the prediction of the mean neutral hydrogen fraction ($\bar{x}_{\text{HI}}$) during the Epoch of Reionization (EoR) from the two-dimensional tomography of the redshifted 21-cm maps of the differential brightness temperature. We first simulated the undistorted 21-cm maps of 200$^3$ uniform grids with 50$^3$ ~ 300-cMpc box sizes between $z = 6 \sim 13$ by using the semi-numerical simulation code 21cmFAST. We then applied various instrumental conditions to the dataset by controlling the beam size and the frequency bandwidth suitable for the upcoming Square Kilometre
Array (SKA) to produce noisy mock 21-cm maps. After having converted the noisy mock 21-cm maps into RGB images, we applied them as inputs of our CNN architecture to predict the corresponding values of $\bar{x}_{\text{HI}}$.

The main results of this paper can be summarized as follows.

1. Our CNN method has a capability to predict $\bar{x}_{\text{HI}}$ from the raw noisy 21-cm maps even when the overall signal-to-noise ratio (SNR) is less than unity, depending on the radio survey configuration. This is because the 21-cm maps during the EoR are highly non-Gaussian in general, so that the CNN method can utilize many additional features other than just SNR.

For a similar reason, the overall performance of our CNN method depends on a combination of the radio-survey configuration, such as beam size, frequency bandwidth, and telescope integration time, other than just the SNR.

2. For a survey configuration with a low SNR (e.g., SNR $\lesssim 1$), a systematic bias with a wavy shape exists between the CNN prediction of $\bar{x}_{\text{HI}}$ and its true value. This might mean that the CNN method becomes closer to the classifier of discrete categories rather than the regressor of a continuous value, mainly due to the high noise level.

On the other hand, a survey configuration with large smoothing (e.g., large beam size and frequency bandwidth) may allow a large scatter of the prediction error, partly due to the slices containing large numbers of small ionized bubbles.

3. If one uses a 1000-hour integration of SKA, a beam size and frequency bandwidth of $\Delta \theta \simeq 2^\prime \sim 3^\prime$ and $\Delta \nu \simeq 1$ MHz, respectively, can be regarded as an optimal configuration for the prediction of $\bar{x}_{\text{HI}}$ with the CNN method. With such a configuration, the median and the $2\sigma$-upper bound of the absolute prediction error $|\bar{x}_{\text{HI, pred}} - \bar{x}_{\text{HI, true}}|$ are $\sim 0.008$ and $0.04$, respectively.

Although our work can successfully reconstruct the reionization history during the EoR by incorporating some of the difficulties from actual radio observations, we have found that plenty of rooms exists for improvement. For example, for a more realistic performance test, adopting a detailed information on the beam shape and the noise addition from future radio surveys, as well as the Galactic and the extragalactic HI foreground [42], might be useful. Also, including 21-cm maps with varying cosmological and reionization parameters would be helpful for understanding how our reconstruction of $\bar{x}_{\text{HI}}(z)$ depends on such parameters.

ACKNOWLEDGMENTS

The authors thank Kyungjin Ahn, Hyunbae Park, Dongsu Bak, Sangnam Park, David Parkinson, Jacobo Asorey, and an anonymous reviewer for helpful discussion and comments. The authors were supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (2018R1A6A1A06024977). Computational data were transferred through a high-speed network provided by the Korea Research Environment Open Network (KREONET).

REFERENCES

[1] J. Miralda-Escudé, Science 300, 1904 (2003); astro-ph/0307396.
[2] S. R. Furlanetto and S. P. Oh, Mon. Not. R. Astron. Soc. 363, 1031 (2005); astro-ph/0505065.
[3] X. Fan et al., Astron. J. 132, 117 (2006), astro-ph/0512082.
[4] O. Zahn et al., Astrophys. J. 756, 65 (2012); 1111.6386.
[5] Planck Collaboration et al., Astron. Astrophys. 596, A108 (2016); 1605.03507.
[6] M. J. Mortonson and W. Hu, Astrophys. J. 672, 737 (2008); 0705.1132.
[7] K. Ahn et al., Astrophys. J. Lett. 756, L16 (2012); 1206.5007.
[8] C. H. Heinrich, V. Miranda and W. Hu, Phys. Rev. D 95, 023513 (2017); 1609.04788.
[9] S. J. Tingay et al., Publ. Astron. Soc. Aust. 30, e007 (2013); 1206.6945.
[10] G. Paciga et al., Mon. Not. R. Astron. Soc. 413, 1174 (2011); 1006.1351.
[11] M. P. van Haarlem et al., Astron. Astrophys. 556, A2 (2013); 1305.3550.
[12] A. R. Parsons et al., Astron. J. 139, 1468 (2010); 0904.2334.
[13] D. R. DeBoer et al., Publ. Astron. Soc. Pac. 129, 045001 (2017); 1606.07473.
[14] P. E. Dewdney, P. J. Hall, R. T. Schlizzi and T. J. L. W. Lazio, IEEE Proc. 97, 1482 (2009).
[15] L. Koopmans et al., in Advancing Astrophysics with the Square Kilometre Array (AASKA14), SISSA Medialab (2015), p. 1; 1505.07568.
[16] S. R. Furlanetto, S. P. Oh and F. H. Briggs, Phys. Rep. 433, 181 (2006); astro-ph/0608032.
[17] M. F. Morales and J. S. B. Wyithe, Annu. Rev. Astron. Astrophys. 48, 127 (2010); 0910.3010.
[18] S. Zaroubi, The Epoch of Reionization (2013), vol. 396 of Astrophysics and Space Science Library, Springer-Verlag Berlin Heidelberg, p. 45.
[19] T. Di Matteo, R. Perna, T. Abel and M. J. Rees, Astrophys. J. 564, 576 (2002); astro-ph/0109241.
[20] M. Zaldarriaga, S. R. Furlanetto and L. Hernquist, Astrophys. J. 608, 622 (2004); astro-ph/0311514.
[21] F. H. Briggs and J. Kocz, Radio Sci. 40, RS5502 (2005).
[22] X. Wang, M. Tegmark, M. G. Santos and L. Knox, Astrophys. J. 650, 529 (2006); astro-ph/0501081.
[23] C. Schaef er, M. Geiger, T. Kuntzer and J. P. Kneib, Astron. Astrophys. 611, A2 (2018); 1705.07132.
[24] N. Gillet et al., Mon. Not. R. Astron. Soc. 484, 282 (2019); 1805.02699.
[25] J. Seiler, A. Hutter, M. Sinha and D. Croton, Mon. Not. R. Astron. Soc. 487, 5739 (2019); 1902.01611.
[26] J. Chardin et al., Mon. Not. R. Astron. Soc. 490, 1055 (2019); 1905.06958.
[27] F. List and G. F. Lewis, Mon. Not. R. Astron. Soc. 493, 5913 (2020); 2002.07940.
[28] H. J. Hortúa, L. Malago and R. Volpi, arXiv e-prints arXiv:2005.07694 (2020); 2005.07694.
[29] P. La Plante and M. Ntampaka, Astrophys. J. 880, 110 (2019); 1810.08211.
[30] H. Shimabukuro and B. Semelin, Mon. Not. R. Astron. Soc. 468, 3869 (2017); 1701.07026.
[31] S. Hassan, A. Liu, S. Kohn and P. La Plante, in The 34th Annual New Mexico Symposium, edited by A. D. Kapinska, National Radio Astronomy Observatory, 9 Nov, 2018; p. 7.
[32] A. Mesinger, S. Furlanetto and R. Cen, Mon. Not. R. Astron. Soc. 411, 955 (2011); 1003.3878.
[33] Planck Collaboration et al., arXiv e-prints arXiv:1807.06209 (2018); 1807.06209.
[34] J. R. Bond, S. Cole, G. Efstathiou and N. Kaiser, Astrophys. J. 379, 440 (1991).
[35] S. R. Furlanetto, M. Zaldarriaga and L. Hernquist, Astrophys. J. 613, 1 (2004); astro-ph/0403697.
[36] Y. B. Zel’Dovich, Astron. Astrophys. 500, 13 (1970).