Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **505,** 3982–3997 (2021) https://doi.org/10.1093/mnras/stab1518
Advance Access publication 2021 May 27

Deep learning approach for identification of H II regions during reionization in 21-cm observations

Michele Bianco,[1,2]★ Sambit K. Giri [3,2] Ilian T. Iliev [1] and Garrelt Mellema [2]

[1]*Astronomy Centre, Department of Physics & Astronomy, Pevensey III Building, University of Sussex, Falmer, Brighton BN1 9QH, UK*
[2]*The Oskar Klein Centre, Department of Astronomy, Stockholm University, AlbaNova, SE-10691 Stockholm, Sweden*
[3]*Institute for Computational Science, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*

Accepted 2021 May 21. Received 2021 May 18; in original form 2021 February 12

ABSTRACT

The upcoming Square Kilometre Array (SKA-Low) will map the distribution of neutral hydrogen during reionization and produce a tremendous amount of three-dimensional tomographic data. These image cubes will be subject to instrumental limitations, such as noise and limited resolution. Here, we present `SegU-Net`, a stable and reliable method for identifying neutral and ionized regions in these images. `SegU-Net` is a U-Net architecture-based convolutional neural network for image segmentation. It is capable of segmenting our image data into meaningful features (ionized and neutral regions) with greater accuracy compared to previous methods. We can estimate the ionization history from our mock observation of SKA with an observation time of 1000 h with more than 87 per cent accuracy. We also show that `SegU-Net` can be used to recover the size distributions and Betti numbers, with a relative difference of only a few per cent from the values derived from the original smoothed and then binarized neutral fraction field. These summary statistics characterize the non-Gaussian nature of the reionization process.

Key words: image processing – interferometric – dark ages, reionization, first stars – early Universe.

1 INTRODUCTION

The epoch of reionization (EoR) is a period of great importance in the study of structure formation and evolution in the Universe. During this period, the predominately cold and neutral intergalactic medium (IGM) transitioned to a hot and ionized state due to the appearance of the first luminous cosmic sources. These sources, which may have been star-forming galaxies and quasi-stellar objects, produced the ionizing photons, which over a period of approximately 500 million years completed the reionization of the Universe (Furlanetto, Oh & Briggs 2006; Zaroubi 2012; Ferrara & Pandolfi 2014).

This period is one of the least understood epochs in the history of the Universe, due to the lack of direct observations. Indirect constraints have been put on the reionization process based on observations of the Lyman α forest (e.g. Fan et al. 2006; McGreer, Mesinger & Fan 2011; McGreer, Mesinger & D'Odorico 2014), the number density of Lyman α emitters (e.g. Ota et al. 2008; Ouchi et al. 2010; Robertson et al. 2015), high-z quasar spectra (e.g. Schroeder, Mesinger & Haiman 2013; Totani et al. 2016; Davies et al. 2018; Greig, Mesinger & Bañados 2019), and the measurement of the Thomson scattering optical depth towards the cosmic microwave background (CMB; e.g. Komatsu et al. 2011; Aghanim et al. 2020).

The ground state of neutral hydrogen atom can produce a signal through a spin-flip transition, which is known as the 21-cm signal. This signal will be a unique signature of EoR (e.g. Madau, Meiksin & Rees 1997; Furlanetto et al. 2006). When observed, this 21-cm signal would have redshifted to radio band of the electromagnetic spectrum. Various radio experiments, such as Low Frequency Array[1] (e.g. van Haarlem et al. 2013), Murchison Widefield Array[2] (e.g. Tingay et al. 2013), and the Hydrogen Epoch of Reionization Array[3] (HERA; e.g. DeBoer et al. 2017), have been trying to detect this signal. Recently, these facilities have provided useful upper limits on the 21-cm power spectrum (e.g. Mertens et al. 2020; Trott et al. 2020) that have been used to derive constraints on the properties of reionization (e.g. Ghara et al. 2020, 2021; Greig et al. 2020a, b; Mondal et al. 2020).

The 21-cm signal during the EoR will be highly non-Gaussian and therefore the power spectrum will not give a full statistical characterization of it (e.g. Mellema et al. 2006; Ichikawa et al. 2010; Watkinson & Pritchard 2015; Majumdar et al. 2018; Giri et al. 2019c). In the coming years, the Square Kilometre Array[4] (SKA) will be built. The low-frequency component of the SKA will be sensitive enough to detect the 21-cm signal produced during EoR and create images of its distribution on the sky (Mellema et al. 2013; Koopmans et al. 2015; Wyithe, Geil & Kim 2015). These images contain more information about our Universe as the detection of the signal at different observed frequencies depicts the distribution of neutral hydrogen at a given time during the EoR.

SKA-Low will observe a sequence of such 21-cm images from different redshifts that will constitute a three-dimensional (3D) set of data known as a tomographic data set. The evolution of the 21-cm signal can be seen along the redshift axis. See for example Giri (2019) for more description about tomographic 21-cm images. The

★ E-mail: m.bianco@sussex.ac.uk

[1]https://www.astron.nl/telescopes/lofar/
[2]https://www.mwatelescope.org/
[3]http://reionization.org/
[4]https://skatelescope.org

reionization process is driven by growing H II regions, often referred to as bubbles (e.g. Furlanetto, Zaldarriaga & Hernquist 2004). As the sources of ionizing photons reside inside them, observing these bubbles and their evolution will be interesting. Numerous studies have provided various methods to detect and study properties of H II bubbles (e.g. Datta, Bharadwaj & Choudhury 2007; Mason & Gronke 2020; Zackrisson et al. 2020). We can also study the properties of reionization with 21-cm images (Giri et al. 2018, 2019a). However, tomographic images from SKA-Low will be prone to instrumental limitations, such as noise, limited resolution, and foreground contamination (e.g. Koopmans et al. 2015; Ghara et al. 2016). In the field of image processing, methods that can classify objects or features in images into meaningful segments are known as 'image segmentation' methods. Giri et al. (2018) implemented an image segmentation method to classify neutral and ionized regions in 21-cm images in the presence of instrumental limitations and demonstrated that key properties of reionization can be derived from such observations.

Artificial intelligence (AI) and deep learning methods are capable of learning patterns in image data and identifying interesting regions. Image segmentation based on AI is quite popular in the field of data analysis and has been applied to study objects with complex visual form contained in big data (Long, Shelhamer & Darrell 2014). In recent years, several papers made use of machine learning techniques for a range of problems in astrophysics (e.g. Giri et al. 2019b; Lee 2019; Chen et al. 2020; Yoshiura et al. 2020) and cosmology (e.g. Jeffrey et al. 2020; Sadr & Farsian 2020; Guzman & Meyers 2021). In the case of reionization, several of these methods are aimed to remove foreground emission (Li et al. 2019; Makinen et al. 2021; Villanueva-Domingo & Villaescusa-Navarro 2021), emulate reionization simulations (e.g. Kern et al. 2017; Jennings et al. 2018; Schmit & Pritchard 2018; Cohen et al. 2020; Ghara et al. 2020), or constrain reionization history (e.g. Shimabukuro & Semelin 2017; Chardin et al. 2019; Mangena, Hassan & Santos 2020; Shimabukuro, Mao & Tan 2020) and its astrophysical inputs (e.g. Sullivan, Iliev & Dixon 2017; Gillet et al. 2019; Hassan, Andrianomena & Doughty 2020).

In this work, we present a new approach for the identification of the distribution of H II regions in 21-cm images using a deep learning method named U-shaped convolutional neural network (CNN; U-Net), which is specially designed for image segmentation and feature extraction (Ronneberger, Fischer & Brox 2015). In our case, we adapt this network for processing our image data, which are mock observations of the 21-cm signal during the EoR. The method will segment the images into ionized and neutral regions. We call this framework `SegU-Net`.

This paper is organized as follows: In Section 2, we present how we generate the simulated data sets used for this work. In Section 3, we describe the design of our neural network, including the error estimation. In Section 4, we discuss its application to our simulated SKA-Low data sets, considering a range of summary statistics such as the mean ionization fraction, power spectra, and topological quantities such as size distributions and Betti numbers. In Section 5, we test our framework on various instrumental noise levels, and in Section 6 we test it on a data set produced from a fully numerical reionization simulation. We discuss and summarize our conclusions in Section 7.

2 21-CM SIGNAL

For any deep learning-based method, we need a data set containing a sample of all the possible scenarios, known as the training set. In Section 2.1, we describe the reionization simulation code that we use to create the training set. The observable for radio telescopes observing the 21-cm signal is defined in Section 2.2. Finally, in Section 2.3 we give the methodology we use to mimic the observations expected with SKA-Low.

2.1 Reionization simulation

To train our network, we require a large set of simulations that represent the 21-cm radio signal for a wide range of redshifts during reionization and different assumptions about the astrophysical sources of ionizing radiation. To do so, we employ `py21cmFAST`, the `python` wrapped version of the seminumerical cosmological simulation code `21cmFAST` (Mesinger, Furlanetto & Cen 2011; Murray et al. 2020). The code computes the evolution of the matter density field using the Zel'dovich approximation (Zel'Dovich 1970). The ionization field and the corresponding 21-cm differential brightness temperature are then calculated from the matter density distribution based on the excursion set formalism (Furlanetto et al. 2004; Mesinger & Furlanetto 2007), which considers a region to be ionized when the fraction of collapsed matter fluctuation exceeds a mass threshold. The ionization fraction $x_{\rm H\,II}(\boldsymbol{r})$ at a position $\boldsymbol{r}$ is given as

$$x_{\rm H\,II}(\boldsymbol{r}) = \begin{cases} 1 & \text{if } f_{\rm coll} \geq 1/\zeta \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where ζ is the ionizing efficiency of high-redshift galaxies and $f_{\rm coll}(R_{\rm s},\, M_{\rm min})$ is the fraction of collapsed matter within radius $R_{\rm s}$ that can form haloes with mass greater than $M_{\rm min}$. $f_{\rm coll}$ is calculated at every pixel varying $R_{\rm s}$ within 0 and $R_{\rm mfp}$. The maximum value of $f_{\rm coll}$ is used in equation (1). $R_{\rm mfp}$ implements the effect of a finite mean free path for ionizing photons in the ionized IGM.

The cosmological parameters considered in this work are based on WMAP 5 yr data observation (Komatsu et al. 2009) and consistent with Aghanim et al. (2020) results. We assume a flat ΛCDM cosmology with the following parameters: $\Omega_\Lambda = 0.73$, $\Omega_{\rm m} = 0.27$, $\Omega_{\rm b} = 0.046$, $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$, $\sigma_8 = 0.82$, and $n_{\rm s} = 0.96$.

2.2 Differential brightness temperature

Radio interferometry-based telescopes record the differential brightness temperature $\delta T_{\rm b}$ while observing the redshifted 21-cm signal. $\delta T_{\rm b}$ depends on position on the sky $\boldsymbol{r}$ and redshift z and can be given as (e.g. Mellema et al. 2013)

$$\delta T_{\rm b}(\boldsymbol{r}, z) \approx 27 x_{\rm H\,I}(\boldsymbol{x}, z)\big(1 + \delta_{\rm b}(\boldsymbol{r}, z)\big) \left(\frac{1+z}{10}\right)^{\frac{1}{2}} \left(1 - \frac{T_{\rm CMB}(z)}{T_{\rm s}(\boldsymbol{r}, z)}\right) \left(\frac{\Omega_{\rm b}}{0.044}\frac{h}{0.7}\right) \left(\frac{\Omega_{\rm m}}{0.27}\right)^{-\frac{1}{2}} {\rm mK}, \tag{2}$$

where $x_{\rm H\,I}$, $\delta_{\rm b}$, $T_{\rm CMB}$, and $T_{\rm s}$ are neutral fraction, baryon density contrast, CMB temperature, and spin temperature, respectively.

Previous studies have shown that our Universe will be heated before reionization begins (e.g. Pritchard & Furlanetto 2007; Ross et al. 2017, 2019). Therefore, we assume $T_{\rm s} \gg T_{\rm CMB}$ throughout this work, which is known as the spin saturated approximation and is relevant at lower redshift $z \lesssim 12$ (e.g. Furlanetto et al. 2004; Furlanetto 2006). In the spin saturated approximation scenario, the differential brightness signal is always in emission ($\delta T_{\rm b} \geq 0$ mK) and locations with $\delta T_{\rm b} = 0$ mK correspond to H II regions.

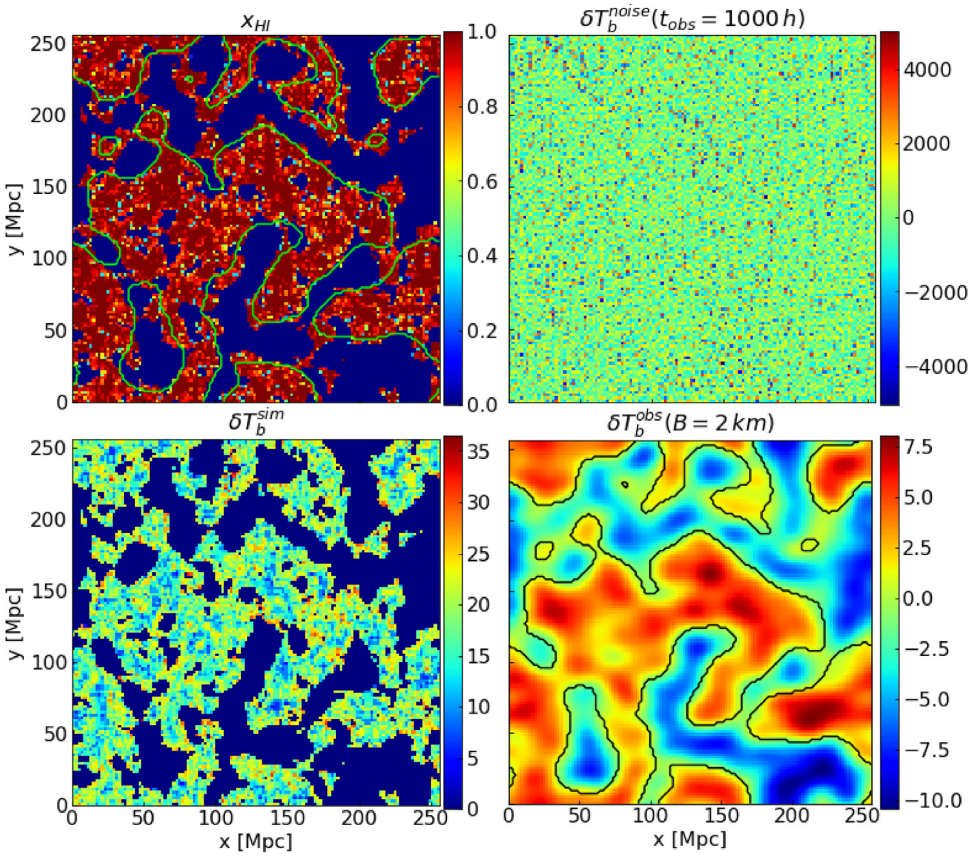

Figure 1. *Top left*: the neutral hydrogen fraction at simulation resolution. Green contours indicate the boundary between neutral and ionized regions after reducing the resolution to an observation with a maximum baseline of $B = 2$ km and matching frequency resolution. *Bottom left*: the 21-cm signal at simulation resolution. *Top right*: the 21-cm signal plus noise realization at simulation resolution for an observing time of 1000 h. To mimic the effect of the lack of a zero baseline, the mean signal has been subtracted. *Bottom right*: the noisy 21-cm image after smoothing to the resolution to an observation with a maximum baseline of $B = 2$ km and matching frequency resolution. This is an example of a smoothed box slice used during the network training. The solid black line shows the same contour as in the top left panel.

Table 1. The parameters used in this study to model the telescope properties.

Parameters	Values
System temperature	$60(\frac{\nu}{300\,\mathrm{MHz}})^{-2.55}$ K
Effective collecting area	962 m^2
Declination	-30°
Observation hours per day	6 h
Signal integration time	10 s

2.3 Mock 21-cm observation

In order to train `SegU-Net` for application to actual observations, we need a training set of mock observations. We create these mock observations by simulating δT_b using the methods described in previous sections and adding instrumental effects, such as the absence of zero baselines, limited resolution, and noise. We follow the methods in Ghara et al. (2016) and Giri et al. (2018) for mimicking the expected effects of SKA1-Low.

We consider a simulation volume of $(256\,\mathrm{Mpc})^3$ and an intrinsic resolution of $\Delta x = 2$ Mpc for simulating the signal. This intrinsic resolution corresponds to an angular aperture of $\Delta\theta = 0.777$ arcmin and a frequency depth of $\Delta\nu = 0.124$ MHz along the line of sight at $z = 7$. As an example, in Fig. 1, we show a coeval cube slice of the neutral fraction field and δT_b field in the top left and bottom left panels, respectively. These slices are taken from the epoch when the universe was about 50 per cent ionized. For each δT_b coeval cube, we assume one axis as the line of sight or frequency direction and subtract the mean signal from each frequency channel, such that this could be considered as a sub-volume from the 3D tomographic data set. We consider this simulation as our reference throughout the result analysis in Section 4; its astrophysical parameters are given in Table 2.

We simulate the instrumental noise using the method given in Giri et al. (2018) and implemented in `Tools21cm`[5] (Giri, Mellema & Jensen 2020). We change the noise seed for each new member of the training set so that the network is trained on different noise realizations and we list our assumed parameters for the telescope set-up in Table 1. In the top right panel of Fig. 1, we show a slice from the simulated noise cube produced from 1000 h of observation with SKA1-Low at simulation resolution. When we add this noise to our simulated signal at the simulation resolution, we cannot discern any feature of the signal as the noise is several orders of magnitude higher than the signal. Therefore, we reduce the resolution of the noisy signal in the field-of-view direction by smoothing with a Gaussian kernel with a full width at half-maximum (FWHM) of $\lambda_0(1 + z)/B$, where B is the maximum baseline. For example, $B = 2$ km corresponds to a resolution of 2.905 arcmin at redshift $z \approx 7$ and 3.631 arcmin at redshift $z \approx 9$, respectively. In the frequency direction, we reduce

[5] A PYTHON package for EoR simulations analysis. https://github.com/sambit-giri/tools21cm

Table 2. Astrophysical parameters used for our fiducial simulation.

Parameters	Values
ζ	39.204
$R_{\rm mfp}$	12.861 Mpc
$T_{\rm vir}^{\rm min}$	3.46×10^4 K

the resolution by convolving with a top-hat bandwidth filter of a width matching the FWHM of the angular smoothing in comoving units. This width corresponds to 0.462 MHz at redshift $z \approx 7$ and 0.551 MHz at redshift $z \approx 9$, respectively. In the bottom right panel of Fig. 1, we show a slice from our noisy signal at this reduced resolution. At this resolution, the smallest H II regions seen in the top left panel of Fig. 1 can no longer be discerned. However, we can still identify the larger H II regions.

To illustrate what we can achieve with these images, we apply the same smoothing to the neutral fraction field and apply a threshold of $x_{\rm th} = 0.5$ to label neutral/ionized regions. We refer to the smoothed and then binarized neutral fraction field as the *ground truth*. We use this field to compare the accuracy of the recovered binary field throughout our paper. We want to point out that this is different from the ground truth of the original reionization simulation as the limited resolution of the radio telescope will limit the observation of small-scale features. Then, we overplot the boundaries of these ionized regions, the neutral fraction slice and signal slice in top left and bottom right panels of Fig. 1, respectively.

2.3.1 Training and testing set

For our training set, we randomly sample the astrophysical simulation parameters by a normal distribution, such that the ionizing efficiency of high-redshift galaxies ζ is sampled with $\mathcal{N} \sim (52.5, 20)$, the mean free path of ionizing photons $R_{\rm mfp}$ with $\mathcal{N} \sim (12.5\,{\rm Mpc}, 5\,{\rm Mpc})$, and the (logarithmically spaced) minimum virial temperature for haloes to host star-forming galaxies $T_{\rm vir}^{\rm min}$ with $\mathcal{N} \sim (4.65, 0.5)$. The choice of these values is such that for a majority of the samples most of the reionization history ($x_{\rm H\,I}^{\rm V}$ from 0.9 to 0.1) falls within the redshift interval of 9–7. The redshift is randomly sampled with a uniform distribution $\mathcal{U} \sim [7,\ 9]$. The initial conditions of the cosmological density field are changed for each simulation. This helps us avoid the impact of cosmic variance on our trained model. With the list of all the parameter values, we produce 10 000 mock observations of the 21-cm signal. Out of these mock observations, we use 15 per cent as the so-called network validation set. This validation set is used during the training method to provide an unbiased evaluation of the network model fit.

Eventually, we will use `SegU-Net` on actual 21-cm image observations. Here, we rely on an additional 300 mock observations as the testing or prediction set. Just as for the training set, the parameter values are randomly chosen. We call this the 'random' testing set. The training process is blind to the prediction set. Apart from the above testing set, we create an additional simulation with fixed values of astrophysical parameters (given in Table 2). We have chosen these values such that between $z = 9$ and 7 reionization proceeds from $x_{\rm H\,I}^{\rm V} \approx 0.9$ to 0.1. We call this set the 'fiducial' testing set. Since the signal evolves as reionization progresses in this testing set, it better mimics the upcoming 21-cm observations. With this testing set, we will test `SegU-Net`'s capability to capture the evolution of structures and recover the binary field from untrained data in Section 4.

2.3.2 Fully numerical simulations testing set

To train `SegU-Net`, we relied on `21cmFAST` for creating the training set. However, our Universe may not exactly be described by this seminumerical code. If our neural network has learnt to find structures in `21cmFAST` simulations only, then we cannot use it for SKA observations. To ensure that the neural network is not overfitted, we consider a different reionization simulation code to build the mock observations.

We first simulate the matter density field and track the evolution of cosmic structures by using the `CUBEP`^{3}`M` *N*-body code (Harnois-Déraps et al. 2013). The simulation is carried out in a volume of $(349{\rm Mpc})^3$ with 64 billion particles. Dark matter haloes down to a mass of 10^9 M$_\odot$ are found at various redshifts using the spherical average halo finder (Watson et al. 2013); meanwhile, haloes with masses between 10^8 and 10^9 M$_\odot$ are implemented with a sub-grid method (Ahn et al. 2015). We use the same cosmology that is given in Section 2.1.

We then employ the `C`^{2}`RAY` radiative transfer (RT) code to simulate the cosmic reionization. `C`^{2}`RAY` requires the matter density field in a 3D grid. Therefore, the distribution *N*-body particles are put in 3D grids with a smoothed particle hydrodynamic method (e.g. Shapiro et al. 1996; Mao et al. 2019). This grid has a spatial resolution of $\Delta x = 2.1$ Mpc and a 166^3 mesh grid. Source ionizing photon production rate per unit time is proportional to the mass of the hosting halo $M_{\rm halo}$ such that

$$\dot{N}_\gamma = f_\gamma \frac{M_{\rm halo}\,\Omega_{\rm b}}{\Delta t_{\rm s}\, m_{\rm p}\, \Omega_{\rm m}}, \tag{3}$$

where $m_{\rm p}$ is the proton mas and $\Delta t_{\rm s} = 11.53$ Myr is the star lifetime. The efficiency factor of sources is defined as $f_\gamma = f_\star\, f_{\rm esc}\, N_i$, where $f_\star$ is the star formation efficiency, $f_{\rm esc}$ is the photons escape fraction, and N_i is the star ionizing photon production efficiency per stellar atom. The efficiency factor for haloes with masses $M_{\rm halo} < 10^9$ M$_\odot$ is set to $f_\gamma = 2$. For the lower mass haloes, it is initially set to $f_\gamma = 8.2$. When their environment becomes ionized (above 10 per cent), their efficiency is reduced to $f_\gamma = 2$ to account for radiative feedback. `C`^{2}`RAY` outputs the hydrogen ionization field at a time interval of 11.5 million years. For more details on the RT and *N*-body simulations methods, see Iliev et al. (2012) and Bianco et al. (2021).

We derive the differential brightness temperature $\delta T_{\rm b}$ from the ionization field and the density using equation (2). We select four outputs, which are at redshifts $z = 7.96$, 8.06, 8.17, 8.28, 8.40, 8.52, and 8.64, corresponding to a volume-averaged neutral fraction of $x_{\rm H\,I}^{\rm V} = 0.17$, 0.29, 0.42, 0.57, 0.70, 0.81, and 0.90, respectively. The simulated $\delta T_{\rm b}$ from these epochs are converted into mock observations using the procedure outlined in Section 2.3. We use these mock observations as a testing set.

The right-hand panel of Fig. 2 shows a slice of the calculated $\delta T_{\rm b}$ for redshift $z = 8.06$ ($x_{\rm H\,I}^{\rm V} = 0.38$ at simulation resolution). Similar to the bottom right panel of Fig. 1, we add the instrumental noise corresponding to a 1000 h observation and smooth the signal to a resolution corresponding to a maximum baseline $B = 2$ km. The black contours correspond to the boundary between neutral and ionized regions. These boundaries are derived from the simulated neutral fraction field at the same resolution as the $\delta T_{\rm b}$ data set.

3 U-NET FOR 21-CM IMAGE SEGMENTATION

Here, we describe our machine learning method for identifying ionized and neutral regions in noisy 21-cm images and our approach

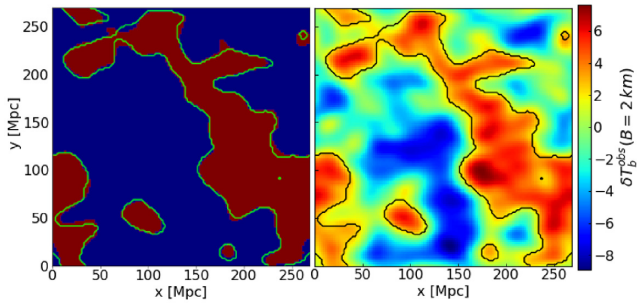

Figure 2. *Right-hand panel*: An example of a smoothed cube slice from the C^{2}RAY simulation on the right employed to test the stability and reliability of the network. This slice is for $z = 8.06$ and corresponds to a volume-averaged neutral fraction of $x_{\rm H\,I}^{\rm V} = 0.38$. As for the training set, at simulation resolution, we subtracted the mean signal in the frequency direction from the differential brightness temperature. We then added simulated instrumental noise for the observed time of 1000 h and smoothed the signal with the same baseline as SKA1-Low. *Left-hand panel*: The binary field recovered with our neural network. In red/blue, the prediction performed with our network and the green contour shows the boundary between neutral and ionized regions. The same contour is shown with a solid black line on the right-hand panel for comparison.

to estimate the uncertainty of its results in Sections 3.1 and 3.2 respectively.

3.1 Our network, SegU-Net

Our segmentation network[6] is based on the U-Net framework first introduced by Ronneberger et al. (2015). U-Net consists of two likewise symmetric paths, an encoder operator that contracts the image and a decoder operator that expands the extracted features. The encoder corresponds to a classical CNN. This CNN aims to reduce the size of the input image in such a way that only information of the most interesting features remains. A series of concatenated convolution operations (layers) returns a low-dimensional latent space (or latent vector) that contains information about these extracted features. In Appendix A, we provide a visual representation of the low-dimensional latent space for the example case of a sphere. We show a schematic representation of the U-Net in Fig. 3. The left part of the U-shape in the diagram and the bottom layer represent the encoder and the low-dimensional latent space, respectively. For a detailed discussion of CNNs, we refer the reader to Mehta et al. (2019), and for examples of employing CNNs to infer cosmological and astrophysical parameters in the context of reionization to Gillet et al. (2019) and Hassan et al. (2020). In our case, the information in the latent space (or latent vector) of U-Net (bottom layer) is expanded by a decoder into a binary map of the same size as the input image. The right part of the U-shape of the diagram in Fig. 3 represents the decoder. The decoder gradually increases the spatial resolution of the latent vector with an up-sampling operation (transposed convolution) until we obtain the same dimension of the input image. After each up-sampling step, the output is combined with the corresponding encoder layer with the same dimension. We illustrate this further in Appendix B with an example.

Even though each of our image data sets is 3D, SegU-Net is trained on two-dimensional (2D) slices. We identify structures in 3D image data by running on every slice along the third axis. Tests show that the method is not sensitive to the choice of the third axis.

[6]https://github.com/micbia/SegU-Net

When compared to a neural network trained on 3D data, we found that our approach is computationally less expensive without loss of accuracy. Therefore, the U-Net architecture described in this work is only applied to 2D image data.

The structure of the encoder layers consists of two convolutional blocks followed by a 2D max-pooling layer (MaxPool) of size 2×2 and a 5 per cent rate dropout layer (Drop). This regularization technique randomly shuts down a portion of the layer neurons to avoid overfitting (Hinton et al. 2012; Srivastava et al. 2014). The convolutional block (ConvBlock) consists of a 2D convolution layer (Conv2D) with 3×3 kernel size. We add a layer that normalizes the previous input layer over the batch sample to avoid overfitting (BN; Ioffe & Szegedy 2015) and as an activation function we employ a Rectified Linear Unit (ReLU) activator (Jarrett et al. 2009; Glorot, Bordes & Bengio 2011), ConvBlock = Conv2D+BN + ReLU. This layer structure is repeated for a total of four levels (Encoder-Level). At each step, the dimension of the input image is halved by the max-pooling operation. The number of feature channels is doubled by the convolutional layer, Encoder-Level = 2∗ConvBlock+MaxPool + Drop. The decoder structure is somewhat similar to the encoder. We replace the pooling operation with a transposed 2D convolution (TConv2D; Zeiler & Fergus 2013; Dumoulin & Visin 2016), which has an opposite scaling effect on the resolution and channel size. This layer output is then concatenated (CC) with the corresponding encoder level to preserve the features extracted in the contracting path. This step is followed by a dropout layer and two convolutional blocks, Decoder-Level = TConv2D+CC+Drop + 2∗ConvBlock. The final output consists of a 2D convolutional layer followed by a sigmoid activation. Our network has a total of 23 2D convolutional layers distributed on four down- and up-sampling scaling levels and a bottom layer, for a total of approximately 2.5 million trainable parameters. In Fig. 3, we show our best-performing network and label the shape of the output from each intermediate hidden layer of this network. More details are provided in Appendices A and B.

During our training process, the hyperparameters of the network are learnt by minimizing a loss function. We employ the balanced cross-entropy (BCE; Salehi, Erdogmus & Gholipour 2017):

$$\mathcal{L}(\mathrm{y}, \hat{\mathrm{y}}) = -\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=0}^{\mathrm{N}} (\beta\, \mathrm{y_i} \log_{10}(\hat{\mathrm{y}}_\mathrm{i}) + (1-\beta)(1-\mathrm{y_i}) \log_{10}(1-\hat{\mathrm{y}}_\mathrm{i}))\,, \tag{4}$$

where $y_i \in \{0; 1\}$ is the pixel-wise ground truth, $\hat{y}_i$ is the predicted value, N is the batch size, which is our case is of size 32, and the parameter $\beta = \sum_{i=0}^{N} y_i$ is the average volume neutral fraction of the batch. In our context, at early/late stage of reionization the statistical weight of the ionized/neutral pixels is underrepresented. This situation is known in data science as a problem affected by *'class unbalanced' data*. To deal with this, we use the above loss function that has been shown to be well suited for segmentation problems that are affected by class unbalanced data (Cui et al. 2019). We further used the Adaptive Moment Estimator Adam (Kingma & Ba 2014), an optimized stochastic gradient descent algorithm for error minimization. The initial learning rate, the step size of the rate of convergence that minimizes the loss function, is set to 10^{-3}. We trained the network using two GPUs, and it took approximately 1500 wall clock hours.

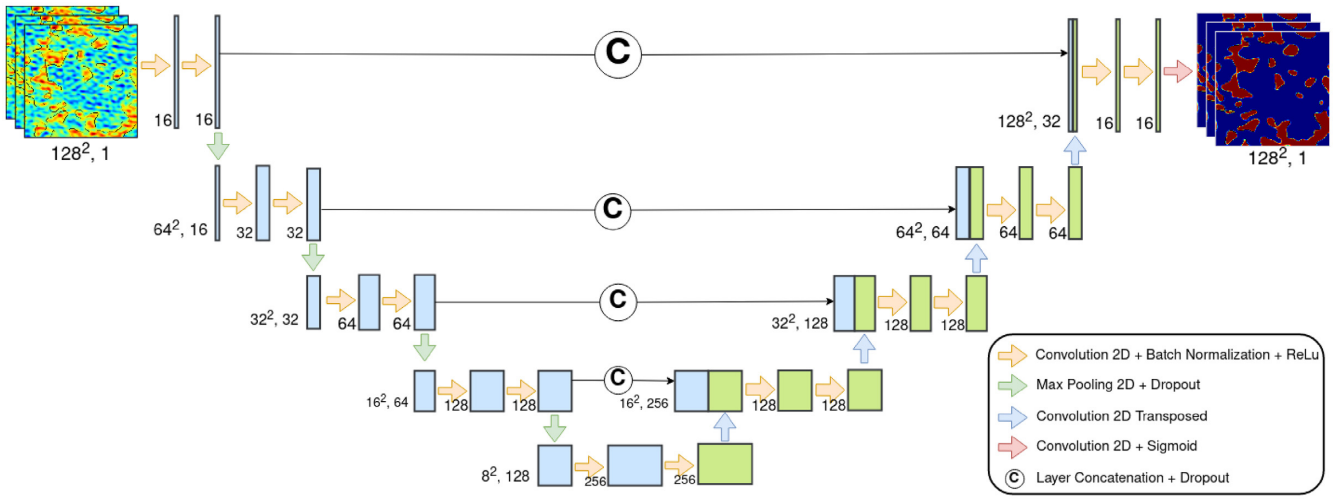

Figure 3. Schematic representation of `SegU-Net` network architecture. The orange arrow indicates a 2D convolutional layer, followed by batch normalization and ReLU activation. Pooling operations followed by dropout layer are indicated with green arrows. The blue arrow indicates an up-sampling layer by transposed 2D convolutional layer and with a red arrow the closing layer, a 2D convolution followed by a sigmoid activation function. The descending path on the left-hand side divides the resolution of the image after each pooling operation and doubles the channel dimension after each convolution. On the other hand, the expansion path doubles the spatial dimension at each up-sampling operation and decreases the channel dimension after concatenation with its counterpart layer in the descending path.

3.2 Uncertainty estimation on `SegU-Net`

One of the main drawbacks of machine learning is that it is unable to quantify uncertainties and confidence intervals for its predictions, and only recently attempts have been made (Charnock, Perreault-Levasseur & Lanusse 2020; Hortúa, Volpi & Malagò 2020) to include error estimation. However, this has not yet been generally implemented for U-Nets. Additionally, if not well optimized, neural networks are prone to overfitting and tend to be biased. Therefore, we have developed an error estimation procedure to be used during the prediction process. This procedure gives our network additional power by providing a pixel-by-pixel error map.

Image manipulations, such as zooming, shifting along an axis, flipping axes, and rotation along an axis, are commonly performed on 2D or 3D image training data to increase the number of samples (Simonyan & Zisserman 2015; Szegedy et al. 2015). This technique is known as time-test augmentation (TTA) of data (Perez & Wang 2017; Wang et al. 2020). Here, we use this approach to estimate the error on the final result. We perform several copies ($\sim$100) of the same sample during the prediction process through image manipulations. These manipulated copies are then independently processed by `SegU-Net`. Each of the recovered binary fields is transformed back. We calculate the final result as the average of these fields and the per-pixel standard deviation to estimate the error for each pixel.

An example of the pixel per pixel error map can be seen in Fig. 4 (rightmost panel). We will discuss this figure further in Section 4.1. This simple method provides our neural network with an uncertainty estimation for each labelled pixel.

4 RESULTS

Once the network is trained, we want to estimate how well it recovers the binary field from noisy 21-cm images. To do so, we include in our analysis the state-of-the-art Super-Pixel method presented in Giri et al. (2018). The Super-Pixel method is based on an advanced image processing technique called the Simple Linear Iterative Clustering (SLIC; Achanta et al. 2012). SLIC groups similar pixels in images into *'superpixels'*. These Super-Pixels are then classified into neutral and ionized ones to get the final map containing the identified features. In previous studies, this method has been shown to be superior compared to other methods, such as putting a simple threshold to the mean signal (e.g. used in Kakiichi et al. 2017), the k-means method (e.g. used in Giri et al. 2018), or the maximum deviation method (e.g. used in Gazagnes, Koopmans & Wilkinson 2021). The Super-Pixel method proves to be quite efficient in recovering the binary fields from noisy 21-cm images. The summary statistics extracted from those are accurately reproducing the ones obtained using the simulation data sets. As shown by Giri et al. (2018), the choice for the number of superpixels depends on the simulation box size and resolution. In our case, we tested for a few values between 500 and 7000. We noticed that above the value of 5000, the algorithm becomes more computationally expensive without yielding a substantial increase in the segmentation accuracy. Hence, we employ 5000 superpixels.

4.1 Visual comparison

To start, we show a visual comparison of slices in Fig. 4. We compare the predicted binary field recovered by the Super-Pixel method (leftmost panel) and `SegU-Net` (central panel) with the ground truth (green contours in both panels). As explained in Section 2.3, the ground truth is the boundary of ionized regions extracted from the simulation neutral fraction field at the same resolution by putting a threshold of 0.5. The red and blue pixels represent neutral and ionized pixels, respectively. In the rightmost panel, we show the pixel error estimated from `SegU-Net` with a colour bar. The error is determined by calculating the standard deviation of the same pixel from the different version of the same mock observation produced with TTA (see Section 3.2).

`SegU-Net` shows better precision in recovering shapes of the ionized regions compared to the Super-Pixel method. As expected, most of the network uncertainty is located at the boundaries of neutral regions or between two large ionized bubbles when these are percolating, and the gap is getting narrower. This uncertainty has

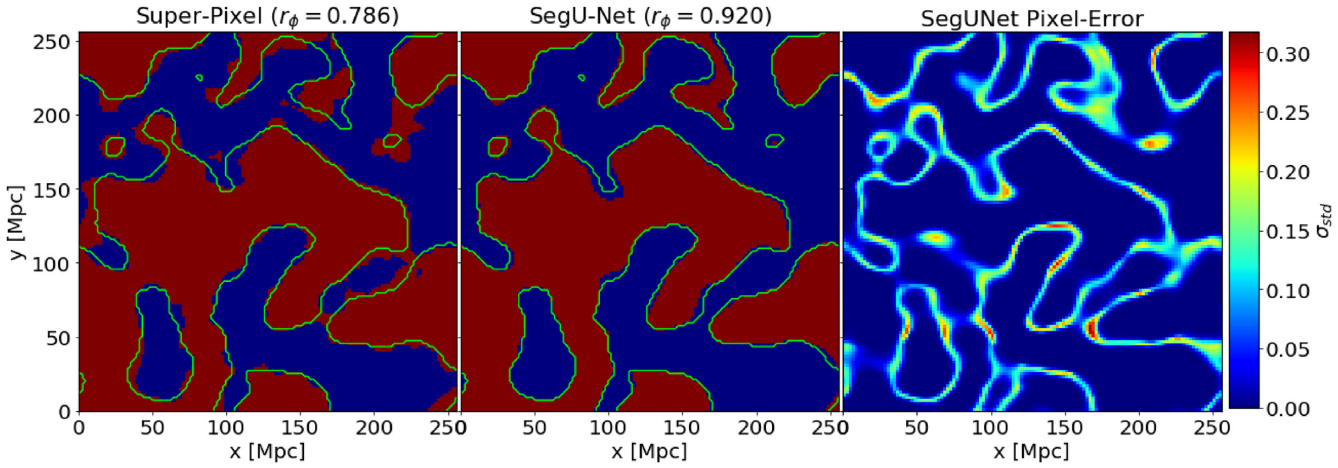

Figure 4. Slice comparison of the binary field, in blue ionized regions and in red neutral. *Left-hand panel*: binary field recovered by the Super-Pixel method. *Central panel*: binary field recovered by our neural network. Green lines indicate the true separation between ionized/neutral regions, derived from a smoothed version of the simulated neutral hydrogen distribution. *Right-hand panel*: the per-pixel error as calculated by `SegU-Net`. The colour bar indicates the intensity of the network uncertainty.

a direct bearing on small neutral islands of a few Mpc scale, residing in vast ionized regions. Moreover, larger uncertainties, $\sigma_{\rm std} \geq 0.25$ are located around narrow ionized regions protruding into large neutral regions (e.g. in Fig. 4 rightmost panel, at coordinates $x \sim 140$ Mpc and $y \sim 125$ Mpc). This behaviour suggests that the uncertainty mainly depends on the contrast between the local neutral and ionized regions. The network selects regions in the image based on the largest gradient in the 21-cm signal intensities to recover the binary field. Therefore, we expect larger uncertainties for reionization scenarios in which the contrast in the 21-cm intensities is relatively small.

4.2 Correlation coefficient

To compare the predicted ionized fields from the 21-cm images mathematically, we use the Matthews correlation coefficient (MCC; also known as r_ϕ coefficient), defined as

$$r_\phi = \frac{N_{\rm TP} \cdot N_{\rm TN} - N_{\rm FP} \cdot N_{\rm FN}}{\sqrt{(N_{\rm TP} + N_{\rm FP})(N_{\rm TP} + N_{\rm FN})(N_{\rm TN} + N_{\rm FP})(N_{\rm TN} + N_{\rm FN})}}\,, \quad (5)$$

where $N_{\rm TP}$ and $N_{\rm TN}$ are the total numbers of neutral and ionized pixels recovered correctly, respectively. $N_{\rm FP}$ is the total numbers of pixels incorrectly guessed as neutral and $N_{\rm FN}$ is the total numbers of pixels incorrectly guessed as ionized. In our case, a positive/negative result corresponds to the neutral/ionized case since the quantity 1 in our binary fields indicates the neutral condition and 0 the ionized. Thus, MCC is a useful metric to correlate binary fields.

In Fig. 5, we show the MCC estimated from the fields segmented into ionized and neutral regions in our testing sets. In the left-hand panel, we provide a scatter plot of MCC values against the reionization history ($x^{\rm v}_{\rm H\,I}$) for the 'random' testing set. We indicate the redshift of the realization by the colour of the points and respective confidence interval with an error bar. We show the number of samples in our training set at a different neutral fraction in an inset panel. After a first attempt, we realized that to overcome the unbalanced class problem requires a better representation of the early ($x^{\rm v}_{\rm H\,I} \approx 1$) and late stages of reionization ($x^{\rm v}_{\rm H\,I} \approx 0$). For this reason, we increased the number of training samples for these stages. Therefore, the distribution of samples against neutral fraction has a bimodal shape with peaks at approximately $x^{\rm v}_{\rm H\,I} \approx 0.1$ and 0.9.

As a result, the r_ϕ value for the overall prediction data set (Fig. 5, left-hand panel) is about 87 per cent for `SegU-Net` (blue dashed line) and 62 per cent in the case of the Super-Pixel method (orange dashed line). The noise level increases with redshift. Therefore, the score is slightly less accurate for redshift $z \geq 8.25$ with an 85 per cent accuracy, meanwhile higher for lower redshift $z \leq 7.75$ with 88 per cent. In the future, we consider increasing the proportion of the training data with high redshift to decrease this performance dissimilarity. The same trend is present in the case of the Super-Pixel method, with an accuracy of 60 per cent and 63 per cent, respectively.

In the right-hand panel of Fig. 5, we compare the MCC values from `SegU-Net` (blue line with circles) with that from the Super-Pixel method (orange line with squares) for our 'fiducial' simulation. As we already know from Giri et al. (2018), the Super-Pixel method performs best for $x^{\rm V}_{\rm H\,I} \approx 0.5$ and deteriorates towards earlier and later stages of reionization. The reason for this behaviour is that during these stages, structures are usually smaller and, therefore, more difficult to identify. With `SegU-Net`, we are able to overcome this problem by employing a specifically designed BCE loss function (equation 4) during the validation process after each training epoch. Therefore, the average r_ϕ value for the 'fiducial' simulation is about 91 per cent for `SegU-Net` (blue dashed line) and 70 per cent in the case of the Super-Pixel method (orange dashed line). In Table 3, we summarize the r_ϕ score for the two test sets.

4.3 Average neutral fraction

After identifying the ionized regions, we can determine the volume-averaged neutral fraction $x^{\rm v}_{\rm H\,I}$, which quantifies the reionization history. In Fig. 6, we show the volume-averaged neutral fraction $x^{\rm v}_{\rm H\,I,\,predicted}$ as calculated from the recovered binary fields extracted by the two methods. In the left-hand panel, we show the $x^{\rm v}_{\rm H\,I,\,predicted}$ from the `SegU-Net` outputs against the true volume-averaged neutral fraction $x^{\rm v}_{\rm H\,I,\,true}$ for our 'random' testing set. The colour of the points indicates the redshifts. The black dashed line indicates $x^{\rm v}_{\rm H\,I,\,predicted} = x^{\rm v}_{\rm H\,I,\,true}$. Except for a few points, all the points lie on or near the black dashed line.

In the right-hand panel of Fig. 6, we compare the results of $x^{\rm V}_{\rm H\,I,\,predicted}$ derived with the Super-Pixel method (orange line with squares) and `SegU-Net` (blue line with circles) for our 'fiducial'

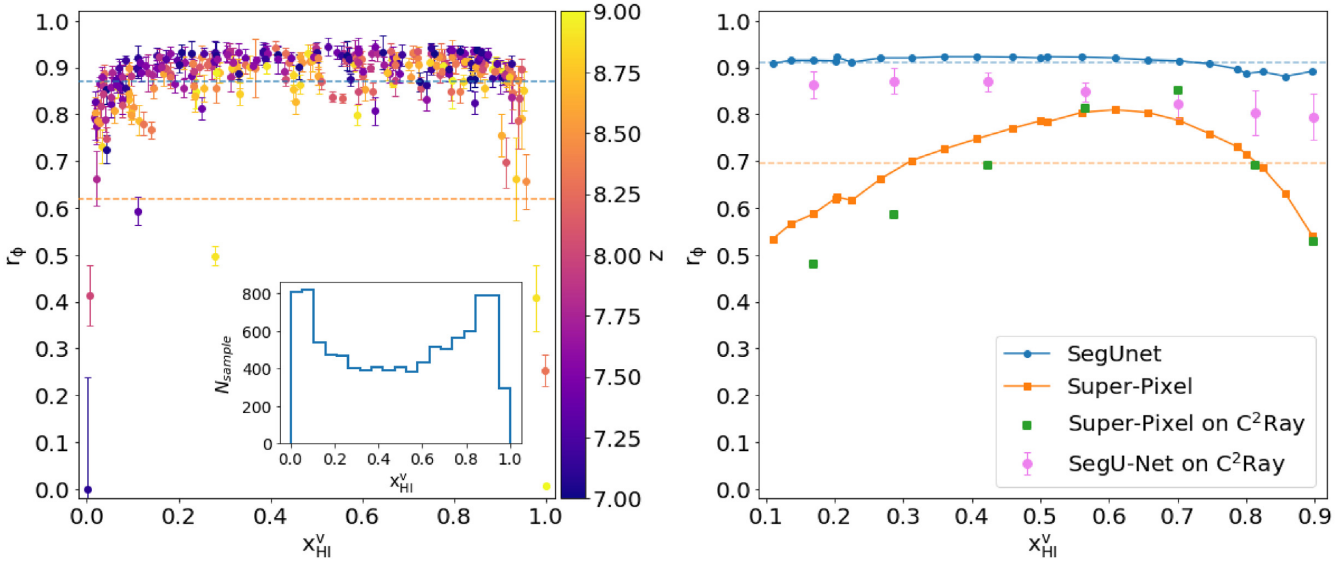

Figure 5. *Left-hand panel*: the MCC r_ϕ of the recovered binary field for the prediction set, against its volume-averaged neutral fraction. Error bar indicates the network confidence interval, and colours indicate the redshift of the simulated coeval cube. On the inset panel, we show the distribution of the training set (blue histogram) against the volume-averaged neutral fraction. *Right-hand panel*: comparison of the same correlation coefficient for recovery performed on the fiducial simulation with our neural network (blue circle line) and the Super-Pixel method (orange square line). We also include the result from the test on the C^{2}RAY simulation, from left to right, redshift $z = 7.96$, 8.06, 8.17, 8.28, 8.40, 8.52, and 8.64 corresponding to a volume-averaged neutral fraction of $x^{\rm V}_{\rm HI} = 0.17$, 0.29, 0.42, 0.57, 0.70, 0.81, and 0.90. The violet dots with relative confidence intervals are predictions performed with SegU-Net for these cases and the green squares are the corresponding results from the Super-Pixel method. Horizontal dashed lines in both panels indicate the overall average r_ϕ coefficient for the entire data set and the fiducial simulation, respectively, in blue for SegU-Net and orange for Super-Pixel method.

Table 3. Summary of the MCC score (in per cent) of our two test sets for the two feature identification methods.

	SegU-Net		Super-Pixel	
Redshift	Random set (per cent)	Fiducial (per cent)	Random set (per cent)	Fiducial (per cent)
$z \leq 7.75$	88.9	91.7	63.7	62.6
$z \geq 8.25$	85.3	90.1	60.7	71.8
$7 \leq z \leq 9$	87.1	91.2	62.0	69.5

simulation. Again, the black dashed line represents $x^{\rm V}_{\rm HI,\,predicted} = x^{\rm V}_{\rm HI,\,true}$. In the case of our neural network, all results lie within the half standard deviation (0.5σ) of the true value (grey dashed lines). With the Super-Pixel method, this is true only from $x^{\rm V}_{\rm HI} \approx 0.5$ to 0.85. The recovered neutral fraction is either underestimated at $x^{\rm V}_{\rm HI} > 0.6$ or largely overestimate for $x^{\rm V}_{\rm HI} < 0.4$.

4.4 Size distributions

From the 3D tomographic data that will be produced with the upcoming SKA experiment, we will be able to study the size distribution of neutral or ionized region during the EoR. Ionized regions are often called bubbles and neutral regions are referred to as islands. The Bubble and Island size distributions (BSDs and ISDs) are useful to derive the properties of reionization and its evolution (Xu, Yue & Chen 2017; Giri et al. 2019a). Several approaches were presented to calculate this distribution (Lin et al. 2016; Kakiichi et al. 2017; Giri et al. 2018). In this work, we employ the Mean-Free-Path method (Mesinger & Furlanetto 2007; Giri et al. 2018) to calculate the size distribution ($R\frac{{\rm d}N}{{\rm d}R}$) of recovered neutral (ISD) and ionized fields (BSD). Previous works have demonstrated that this method should be preferred since the calculated size distributions are almost unbiased (Lin et al. 2016; Giri et al. 2018).

In the left and right columns of Fig. 7, we show the ISDs and BSDs, respectively, of the binary fields recovered with SegU-Net (blue line) and Super-Pixel method (orange line) compared to the ground truth (black dashed line). The Super-Pixel method performs best when the simulation is halfway through the reionization process $x^{\rm v}_{\rm HI} = 0.5$ (central panel). However, it is considerably less accurate compared to SegU-Net. We show the relative difference with the ground truth in the plots below the ISDs and BSDs. The blue shaded region shows the error on each of the size distributions determined by SegU-Net. In both the ISD and BSD cases, the main difference between the two recovered distributions occurs at the earlier $x^{\rm v}_{\rm HI} = 0.8$ (top) and later $x^{\rm v}_{\rm HI} = 0.2$ (bottom) stages of reionization. SegU-Net shows a relative difference of a few per cent while the distributions determined from the Super-Pixel segmentations show relative differences of up to 10 per cent for large sizes.

4.5 Dimensionless power spectra

The dimensionless power spectrum of the neutral field is defined as $\Delta^2_{\rm xx} = k^3 P_{\rm xx}(k)/2\pi^2$, where $P_{\rm xx}$ is the autopower spectrum that quantifies the fluctuations due to the distribution of neutral regions. These fluctuations contribute to the 21-cm power spectrum that is observed with radio interferometric telescopes. See, for example, Furlanetto et al. (2006) and Lidz et al. (2007) for descriptions of the fluctuations of the 21-cm signal. In this section, we study the $\Delta^2_{\rm xx}$ estimated from the neutral fields recovered from various methods.

In Fig. 8, we consider the 'fiducial' simulation at three stages of reionization, which are $x^{\rm V}_{\rm HI} = 0.8$ (top panel), $x^{\rm V}_{\rm HI} = 0.5$ (central

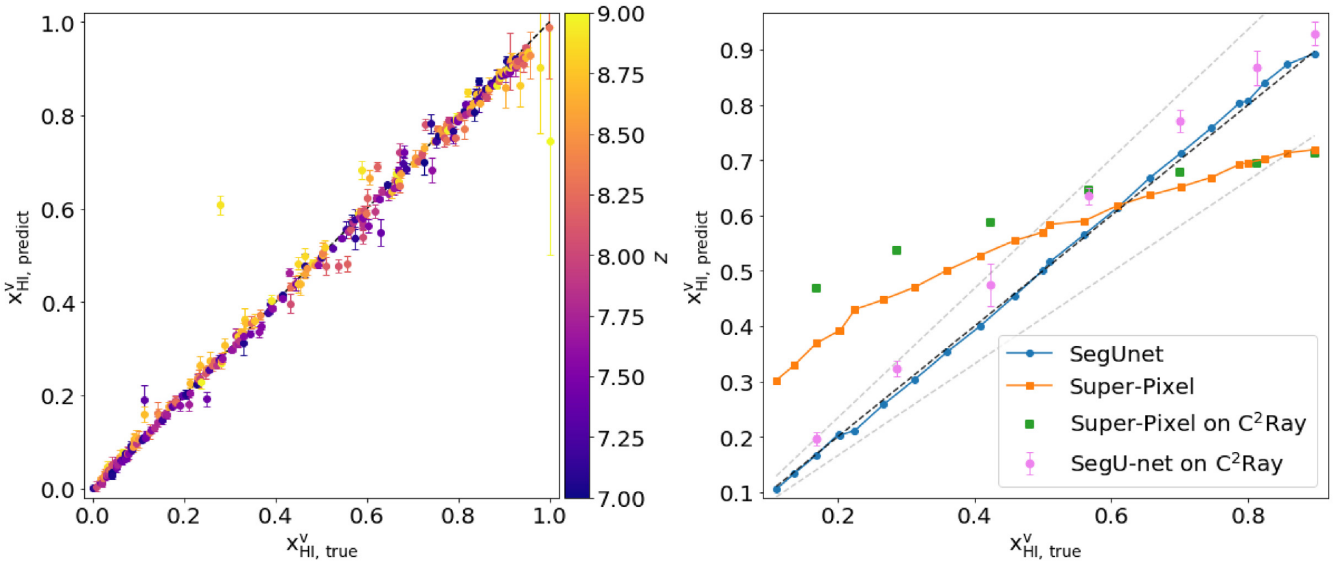

Figure 6. *Left-hand panel*: Comparison of the simulated neutral fraction against the recovered one. Error bar and colour bar are the same as Fig. 5. *Right-hand panel*: The same comparison for the 'fiducial' simulation. We also include the results from the C^{2}RAY simulation. The redshifts of the C^{2}RAY simulation are $z = 7.96$, 8.06, 8.17, 8.28, 8.40, 8.52, and 8.64 corresponding to a volume-averaged neutral fraction of $x^{\rm v}_{\rm H\,I} = 0.17$, 0.29, 0.42, 0.57, 0.70, 0.81, and 0.90. The violet dots with relative confidence interval are predictions performed with `SegU-Net` and green squares are with the Super-Pixel method.

panel), and $x^{\rm V}_{\rm H\,I} = 0.2$ (bottom panel). At the mid-point of reionization (central panel), the Super-Pixel method performs well at large scales $k < 0.2\,{\rm Mpc}^{-1}$ with a relative difference within 25 per cent for lower k-values. The $\Delta^2_{\rm x\,x}$ values of the neutral field recovered by the Super-Pixel method at early and late times have the correct shape but differ in magnitude. The $\Delta^2_{\rm x\,x}$ values of the neutral field recovered by `SegU-Net` match the ground truth well at all three stages of reionization. The network maintains a maximum difference compared with the ground truth, of a few tens of per cent at all scales. For $k \lesssim 0.5\,{\rm Mpc}^{-1}$, the network uncertainty interval grows to 25–50 per cent relative difference.

4.6 Betti numbers

During reionization, ionized bubbles form, grow, and connect with each other to form a complex topology (Furlanetto & Oh 2016). Various studies have proposed topological descriptors for this distribution, such as Euler characteristics (e.g. Friedrich et al. 2011) and Betti numbers (Elbers & van de Weygaert 2019; Giri & Mellema 2020; Kapahtia et al. 2021). Giri & Mellema (2020) pointed out that Betti numbers contain more information compared to the Euler characteristics. Therefore, in this section we study the zeroth β_0, first β_1, and second β_2 Betti number (Betti 1870) of the binary 3D maps recovered by the two feature identification methods.

β_0, β_1, and β_2 describe the number of isolated ionizing regions, tunnels, and isolated neutral regions, respectively. In the top, middle, and bottom panels of Fig. 9, we show the β_0, β_1, and β_2 values estimated from the recovered binary fields of our 'fiducial' model at $x^{\rm V}_{\rm H\,I}$ between 0.1 and 0.9. The black, blue, and orange curves represent the Betti numbers calculated from the ground truth, recovered field with `SegU-Net`, and Super-Pixel method, respectively. In line with the results for the other quantities discussed earlier, we find that the topology recovered with `SegU-Net` is much closer to the ground truth than the one recovered by the Super-Pixel method.

5 TESTS ON DIFFERENT INSTRUMENTAL NOISE LEVELS

We have trained and tested `SegU-Net` for one specific noise level, corresponding to the theoretically expected noise for $t_{\rm obs} = 1000$ h with the current design of SKA-Low. However, in practice, the noise level may differ from this, either because the observing time or telescope design is different from our assumptions or simply because the theoretical noise level is not achieved due to complications with the calibration. Therefore, it is important to test to which extent the performance of our network is sensitive to the noise level in the actual data. To change the noise level, we choose different observing times, one shorter ($t_{\rm obs} = 500$ h) and one longer ($t_{\rm obs} = 1500$ h). The former case corresponds to a noise level $\sqrt{2}$ higher than that used in the training set and the latter to a noise level that is $\sqrt{2/3}$ lower.

In the left-hand panel of Fig. 10, we show the r_ϕ coefficient of the recovered binary field against the volume-averaged neutral fraction $x^{\rm V}_{\rm H\,I}$. We compare the prediction on the reference simulation for the higher ($t_{\rm obs} = 500$ h, green line with squares) and lower noise cases ($t_{\rm obs} = 1500$ h, red line with triangles) with the one using the noise level employed during the training and validation process ($t_{\rm obs} = 1000$ h, blue line with circles). It is evident from the plot that although the noise level does impact the accuracy of the results, we still achieve approximately the same level of precision as in our test case, as commented in Section 4.2. In fact, the overall average accuracy, indicated with horizontal dashed lines in Fig. 10, on the simulation of reference is 89 per cent for the higher noise case (green dashed line) and slightly better, 92 per cent, for the lower noise case (red dashed line). In both cases, there is a drop in performance down to 88 per cent accuracy during the early stages of reionization $x^{\rm V}_{\rm H\,I} > 0.7$, due to the redshift dependence of the simulated noise.

We also want to test how far we can push our `SegU-Net` trained on data with $t_{\rm obs} = 1000$ h instrumental noise to identify structures in the presence of a higher or lower noise level. In the right-hand panel of Fig. 10, we plot the r_ϕ coefficient at different observation times $t_{\rm obs}$; for three different stages of reionization in our reference simulation, namely for volume-averaged neutral fractions, $x^{\rm V}_{\rm H\,I}$ is 0.2 (blue line

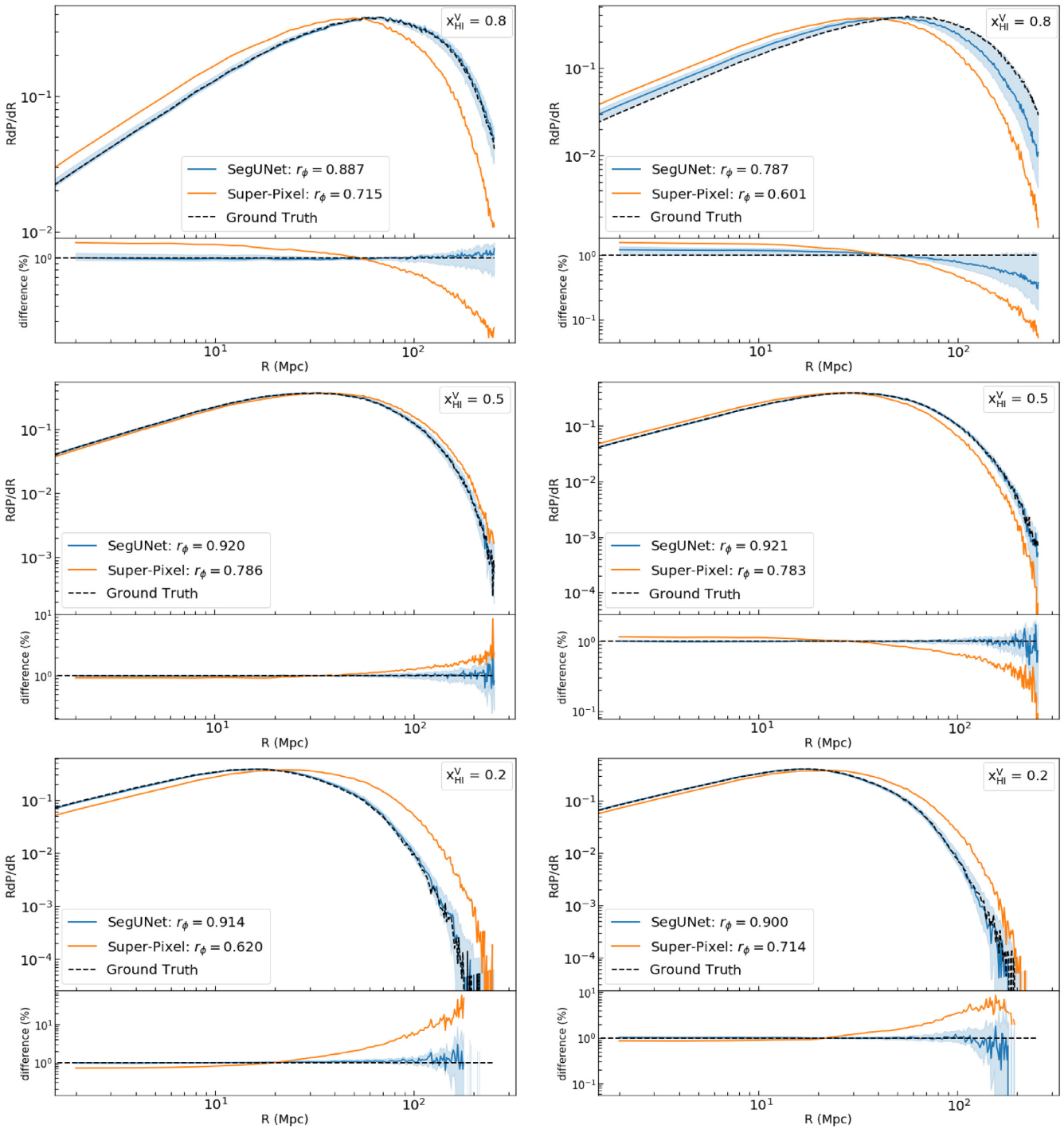

Figure 7. *Left column*: the size distribution of neutral regions (ISD). *Right column*: the size distribution of ionized region (BSD). Rows from top to bottom represent early ($x_{\rm HI}^{\rm V} = 0.8$), middle ($x_{\rm HI}^{\rm V} = 0.5$), and late ($x_{\rm HI}^{\rm V} = 0.2$) stages of reionization, respectively. On each panel, we show the size distributions from the binary fields of the 'fiducial' simulation recovered by `SegU-Net` (blue line) and its respective confidence interval (blue shadow). Black dashed lines and orange lines give the size distributions of the ground truth and binary field recovered by the Super-Pixel method. At the bottom of each size distribution panel, we show the relative deviation from the true binary field distribution.

with squares), 0.5 (orange line with circles), and 0.8 (green line with triangles), corresponding to redshifts $z = 7.310$, 8.032, and 8.720. This plot shows that our network performs well for $t_{\rm obs} \gtrsim 500$ h, where $r_\phi \gtrsim 0.85$. The spike in the curve for $x_{\rm HI}^{\rm V} = 0.8$ at $t_{\rm obs} = 1000$ h is due to the fact that this is the noise level for which the network was trained.

To put our noise level into perspective, the inset plot in the right-hand panel of Fig. 10 shows the signal-to-noise ratio (SNR) achieved for different observation times. The SNR is defined as $\sigma_{21}/\sigma_{\rm noise}$ (e.g. Kakiichi et al. 2017), where σ_{21} and $\sigma_{\rm noise}$ are the standard deviations of the 21-cm signal and noise, respectively, at the resolution corresponding to a maximum baseline of 2 km.

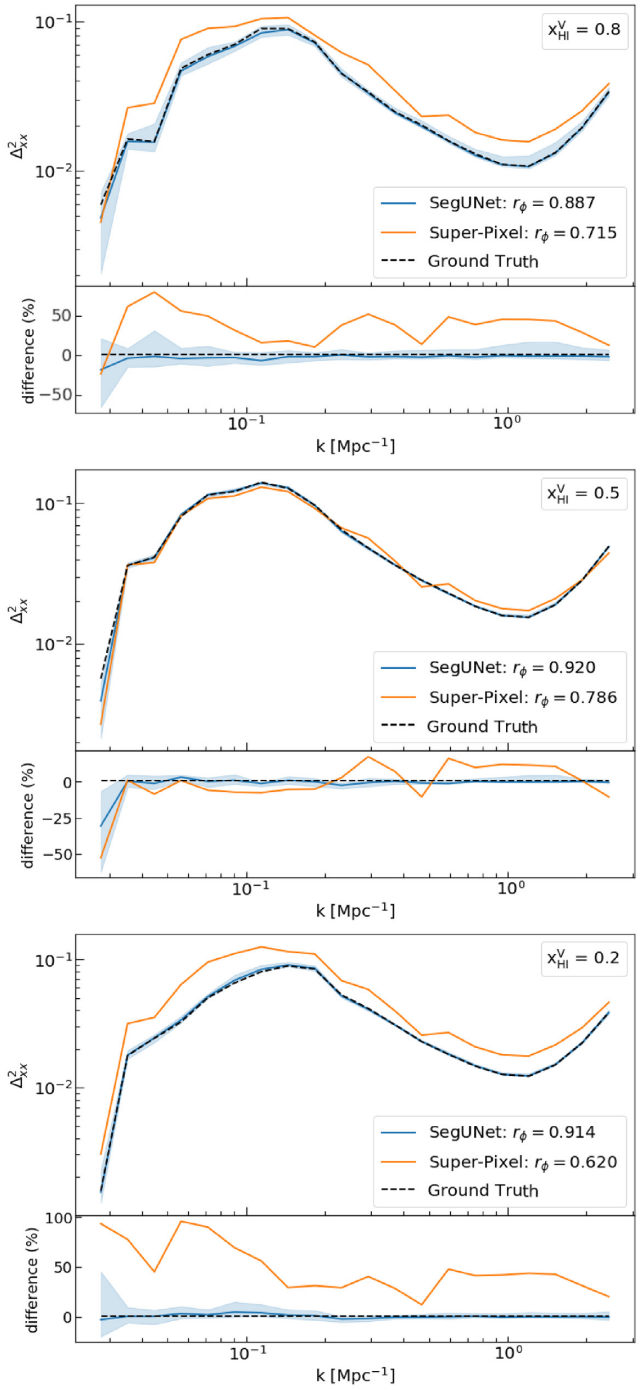

Figure 8. Dimensionless power spectra of the neutral field from the fiducial simulation as recovered by our network (blue line) and its respective confidence interval (blue shadow). Compared at early, middle, and late stages of reionization (from top to bottom $x_{\rm HI}^{\rm V} = 0.8,\ 0.5,$ and 0.2) with the same quantity derived from the ground truth (black dashed line) and the Super-Pixel method (orange line). At the bottom of each panel, we show the relative difference compared to the ground truth for both cases: the network and Super-Pixel methods.

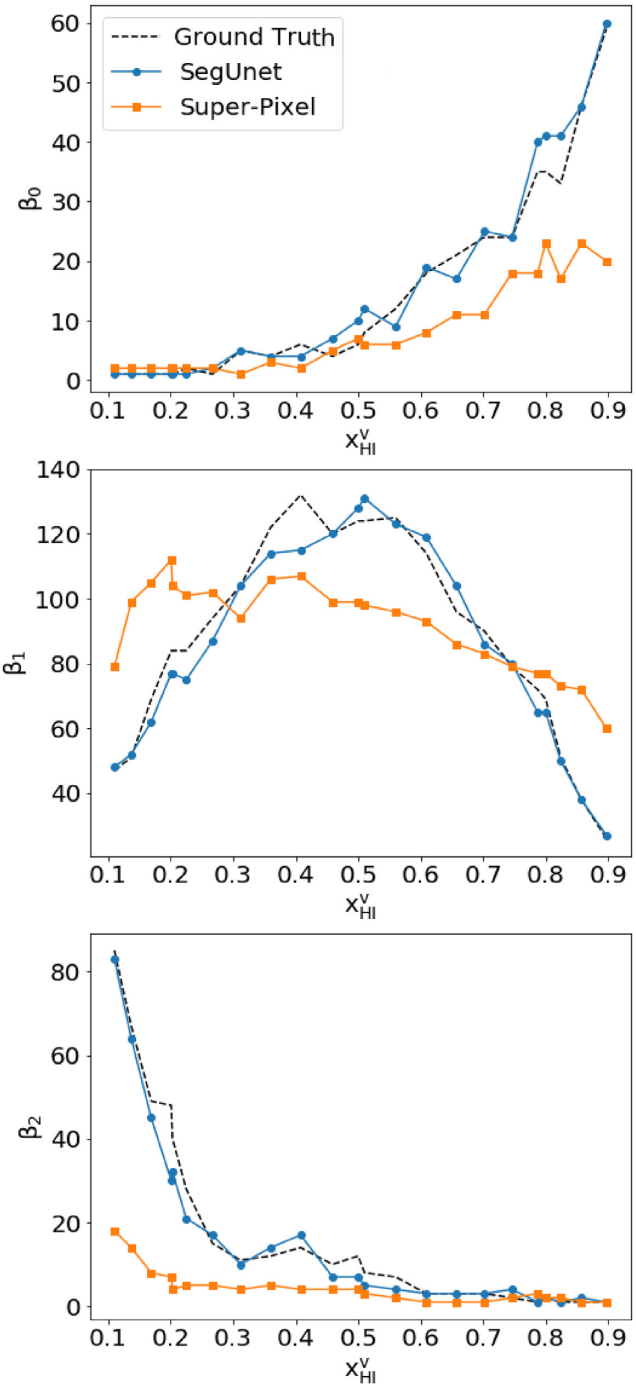

Figure 9. Comparison of the topology of the identified regions with Betti numbers estimated from the original neutral field (black dashed line), the `SegU-Net` (blue line with circles), and the Super-Pixel method (orange line with squares), for the case of our 'fiducial' simulation. The top, middle, and bottom panels give β_0, β_1, and β_2, respectively. The Betti numbers recovered with `SegU-Net` match the ground truth better than those recovered with the Super-Pixel method.

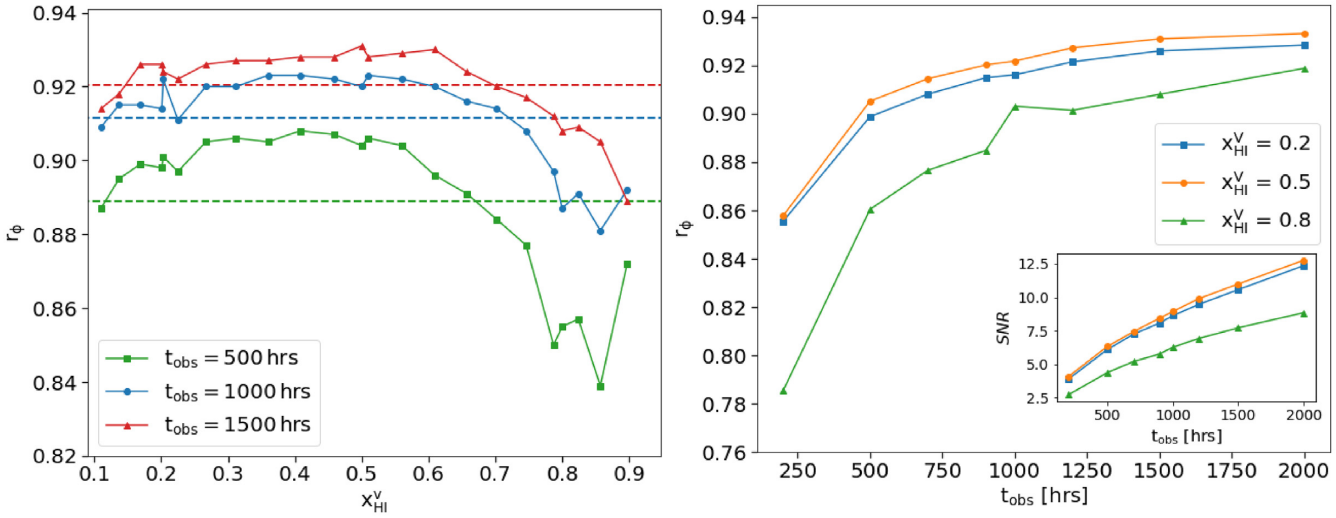

Figure 10. *Left-hand panel*: the MCC r_ϕ of the recovered binary field against its volume-averaged neutral fraction. We compare the prediction set for a high noise level ($t_{obs} = 500$ h, green line with squares) and low noise level ($t_{obs} = 1500$ h, red line with triangles) against the noise level employed during the training ($t_{obs} = 1000$ h, blue line with squares). Horizontal dashed lines of the respective colour represent the MCC average score of the reference simulation. *Right-hand panel*: the evolution of the MCC score for increasing observation time for a set of mock observations with volume-averaged neutral fractions of $x_{H\,I}^V = 0.2$ ($z = 7.310$), 0.5 ($z = 8.032$), and 0.8 ($z = 8.720$), respectively, in blue, orange, and green colours. In the same panel, an inset plot shows the signal-to-noise ratio (SNR $= \sigma_{21}/\sigma_{noise}$) of 21-cm images at a resolution corresponding to a maximum baseline of 2 km we achieve for different observation times.

From this, we conclude that a good performance, with the same accuracy as the 'random' testing set ($r_\phi \gtrsim 0.85$), requires an SNR $\gtrsim$3.

6 TESTS ON A FULLY NUMERICAL SIMULATION

We applied our network to mock δT_b cubes calculated with the C^{2}RAY code, presented in Section 2.3.2, with a spatial resolution close to the 21cmFAST simulations employed in the training process, 2 Mpc, in order to obtain the same level of noise per pixel. A visual comparison of the recovered binary field, similar to the results in Section 4.1, is shown in Fig. 2. In the left-hand panel, the red/blue colour indicates the network prediction.

We go through the same process presented in Section 4. The r_ϕ score with SegU-Net is represented by the violet dots with error bars on the right-hand panel of Fig. 5; from left to right we have redshifts $z = 7.96$, 8.06, 8.17, 8.28, 8.40, 8.52, and 8.64 corresponding to a universe with volume-averaged neutral fractions of $x_{H\,I}^V = 0.17$, 0.29, 0.42, 0.57, 0.70, 0.81, and 0.90; green squares represent the score obtained with the Super-Pixel method. As we can see, our neural network is performing with similar accuracy as for the prediction set of seminumerical simulations as discussed in Section 4.1. For $x_{H\,I}^V \approx 0.55$, SegU-Net performs slightly better than the Super-Pixel method. The Super-Pixel method shows a drop in accuracy at the late ($x_{H\,I}^V < 0.5$) and early ($x_{H\,I}^V > 0.8$) stages of reionization. We do the same comparison with the recovered volume-averaged neutral fraction $x_{H\,I}^V$; in Fig. 6 right-hand panel, the green error bar points are the same data as mentioned earlier. As we can see, also for the C^{2}RAY simulation, SegU-Net recovered quantity resides within the 0.5σ confidence interval (violet dots with error bars). For the Super-Pixel results, this is true only for $x_{H\,I}^V = 0.57$, 0.70, and 0.81, with approximately the same precision as SegU-Net in the case of $x_{H\,I}^V = 0.57$ and slightly better results for $x_{H\,I}^V = 0.70$.

7 DISCUSSION AND CONCLUSIONS

This work has developed a CNN based on the U-Net architecture, which can be used to segment redshifted 21-cm image observations into neutral and ionized regions. We have shown that this application of deep learning provides a fast and stable method that significantly improves the identification of ionized/neutral regions during the EoR over previously proposed methods. To train our network, we employ a large set of simulated mock observations of the 21-cm signal.

Our image segmentation network, SegU-Net, also contains an uncertainty estimator. This uncertainty estimator is a simple but efficient application of the test-time augmented (TTA) technique. With this uncertainty estimator, our network can create a pixel-by-pixel error map during the prediction process. The pixel-by-pixel error map can later be used to determine the error in any quantity derived from the segmentation.

Once the network has been trained, the binary field's extraction is swift. In our case for simulations of volume $(256\,\mathrm{Mpc})^3$ and mesh grid of 128^3, a run in serial on an Intel® Core™i7-6500U CPU @ 2.5 GHz processor and using 16 Gigabytes of RAM takes between 5 to 10 s. Including the pixel-error map calculation increases the computing time to approximately 10 min. For comparison, the Super-Pixel method typically requires several minutes to extract the binary field, where the actual time depends on the image pixel resolution and the number of Super-Pixels employed. The computational speed of our network thus makes it particularly useful as a component in a Bayesian statistical inference framework to perform EoR parameter estimation using tomographic statistics (e.g. Gazagnes et al. 2021).

We compare the accuracy of our approach with a feature finding method from the field of image processing, the so-called Super-Pixel method, which Giri et al. (2018) showed to be superior over simple thresholding methods. The results show that our neural network can identify neutral regions in the mock observations at least as well and often much better than the Super-Pixel method. We show a visual comparison and the resulting pixel-per-pixel error map tested on our 'fiducial' model. This error map gives valuable insight into the

parts of the image that are hard to recover and helps us check for overfitting.

We studied the accuracy of a range of derived quantities from the recovered binary fields, comparing the performance of `SegU-Net` with the Super-Pixel method. These quantities are the volume-averaged ionization fraction – the evolution of which provides the reionization history, the size distribution of the ionized (BSD) and neutral (ISD) regions, the dimensionless power spectra of the recovered binary fields, and the three Betti numbers, which quantify the topology of the segmented data sets. For all quantities, we find that the `SegU-Net` results are more accurate than the Super-Pixel results, especially for the early and late stages of reionization, where the Super-Pixel method often struggles to produce accurate results.

Machine learning methods generally are sensitive to the properties of the training set. Therefore, we tested `SegU-Net` on input data with different properties than the training set. First, we analysed the performance on data sets with different noise levels than the network was trained. We found that `SegU-Net` yields accurate results for data sets in which the noise level is characterized by an observing time of $t_{\rm obs} > 500\,{\rm h}$, which approximately corresponds to an SNR $\gtrsim 3$. Secondly, we applied the network to mock observations calculated from the results of a fully numerical reionization simulation, rather than the seminumerical simulations used to train the network. We find that `SegU-Net` achieves the same level of accuracy when applied to this data set and therefore is not sensitive to the type of simulation employed during the training process.

We want to point out that similar efforts are being made by Gagnon-Hartman et al. (2021). They focus on reconstructing the segmented maps of ionized and neutral regions in the context of foreground mitigation using the foreground avoidance method (e.g. Liu & Tegmark 2011; Pober et al. 2014), and also consider the possibility of doing so with instruments that are not optimized for imaging such as HERA. We include the effect of instrumental noise and study in great detail the summary statistics of the reconstructed binary maps and the dependence of the results on the noise level. In the future, we will extend our study to include the impact of foreground mitigation strategies while recovering the summary statistics.

Here, we assumed the spin temperature to be saturated ($T_{\rm S} \gg T_{\rm CMB}$). However, it is possible to have such a scenario where this assumption fails, especially during the time when reionization starts. In the future, we will evolve our `SegU-Net` to identify H II regions in such scenarios. Even though our network is built to identify H II regions, U-Net architecture can be trained to identify any interesting feature. Before reionization started, the luminous sources heated the IGM and left its impact on the 21-cm signal (e.g. Ross et al. 2017, 2019). The U-Net architecture can also be trained to identify these heated regions.

ACKNOWLEDGEMENTS

The authors would like to thank Leon Koopmans for useful discussions and comments. We also acknowledge helpful discussion with Adrian Liu and collaborators. MB is supported by PhD Studentship from the Science and Technology Facilities Council (STFC) and appreciates the Oskar Klein Center at Stockholm University for hospitality during the completion of this work. This work was possible thanks to the STFC Long Term Attachment (LTA) travel grant. ITI is supported by the Science and Technology Facilities Council (grant numbers ST/I000976/1 and ST/T000473/1) and the Southeast Physics Network (SEPNet). GM is supported in part by the Swedish Research Council grant 2020-04691. We acknowledge PRACE for awarding us access to the KAY facility hosted by the Irish Centre for High-end Computing (ICHEC) and the GALILEO hosted by the Super Computer Application and Innovation (SCAI) in collaboration with the Consortium Interuniversitario del Nord est Italiano per il Calcolo Automatico (CINECA). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for partly funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCSSupercomputer Juelich Wizard for European Leadership Science (JUWELS) at Juelich Supercomputing Centre (JSC). The deep learning implementation was possible thanks to the application programming interface of `Tensorflow` (Abadi et al. 2015) and `Keras` (Chollet et al. 2017). The algorithms and image processing tools operated on our data were performed with the help of `NumPy` (Harris et al. 2020), `SciPy` (Virtanen et al. 2020), and `scikit-image` packages (van der Walt et al. 2014). All figures were created with `mathplotlib` (Hunter 2007).

DATA AVAILABILITY

The data underlying this article are available upon request, and can also be regenerated from scratch using the publicly available `21cmFAST`, `CUBEP`^{3}`M`, `C`^{2}`RAY`, and `Tools21cm` codes. The `SegU-Net` code and its trained network weights are available on the author's `GitHub` page: https://github.com/micbia/SegU-Net.

REFERENCES

Abadi M. et al., 2015, Tensorflow. Available at: http://tensorflow.org/
Achanta R., Shaji A., Smith K., Lucchi A., Fua P., Süsstrunk S., 2012, IEEE Trans. Pattern Anal. Mach. Intell., 34, 2274
Aghanim N. et al., 2020, A&A, 641, A6
Ahn K., Iliev I. T., Shapiro P. R., Srisawat C., 2015, MNRAS, 450, 1486
Betti E., 1870, Ann. Mat. Pura Appl., 4, 140
Bianco M., Iliev I. T., Ahn K., Giri S. K., Mao Y., Park H., Shapiro P. R., 2021, MNRAS, 504, 2443
Chardin J., Uhlrich G., Aubert D., Deparis N., Gillet N., Ocvirk P., Lewis J., 2019, MNRAS, 490, 1055
Charnock T., Perreault-Levasseur L., Lanusse F., 2020, preprint (arXiv:2006.01490)
Chen B. H. et al., 2020, MNRAS, 501, 3951
Chollet F. et al., 2017, Keras. Available at: https://github.com/rstudio/keras
Cohen A., Fialkov A., Barkana R., Monsalve R. A., 2020, MNRAS, 495, 4845
Cui Y., Jia M., Lin T.-Y., Song Y., Belongie S., 2019, preprint (arXiv:1901.05555)
Datta K. K., Bharadwaj S., Choudhury T. R., 2007, MNRAS, 382, 809
Davies F. B. et al., 2018, ApJ, 864, 142
DeBoer D. R. et al., 2017, PASP, 129, 045001
Dumoulin V., Visin F., 2016, preprint (arXiv:1603.07285)
Elbers W., van de Weygaert R., 2019, MNRAS, 486, 1523
Fan X. et al., 2006, ApJ, 132, 117
Ferrara A., Pandolfi S., 2014, Proc. Int. Sch. Phys. Fermi, 186, 1
Friedrich M. M., Mellema G., Alvarez M. A., Shapiro P. R., Iliev I. T., 2011, MNRAS, 413, 1353
Furlanetto S. R., 2006, MNRAS, 371, 867
Furlanetto S. R., Oh S. P., 2016, MNRAS, 457, 1813
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181
Gagnon-Hartman S., Cui Y., Liu A., Ravanbakhsh S., 2021, MNRAS, 504, 4716
Gazagnes S., Koopmans L. V. E., Wilkinson M. H. F., 2021, MNRAS, 502, 1816
Ghara R., Choudhury T. R., Datta K. K., Choudhuri S., 2016, MNRAS, 464, 2234

Ghara R. et al., 2020, MNRAS, 493, 4728
Ghara R., Giri S. K., Ciardi B., Mellema G., Zaroubi S., 2021, MNRAS, 503, 4551
Gillet N., Mesinger A., Greig B., Liu A., Ucci G., 2019, MNRAS, 484, 282
Giri S. K., 2019, PhD thesis. Stockholm Univ., Dep. Astron
Giri S. K., Mellema G., 2020, preprint (arXiv:2012.12908)
Giri S. K., Mellema G., Dixon K. L., Iliev I. T., 2018, MNRAS, 473, 2949
Giri S. K., Mellema G., Aldheimer T., Dixon K. L., Iliev I. T., 2019a, MNRAS, 489, 1590
Giri S. K., Zackrisson E., Binggeli C., Pelckmans K., Cubo R., 2019b, MNRAS, 491, 5277
Giri S. K., D'Aloisio A., Mellema G., Komatsu E., Ghara R., Majumdar S., 2019c, J. Cosmol. Astropart. Phys., 2019, 058
Giri S. K., Mellema G., Jensen H., 2020, J. Open Source Softw., 5, 2363
Glorot X., Bordes A., Bengio Y., 2011, in Gordon G., Dunson D., Dudík M., eds, Vol. 15, Deep Sparse Rectifier Neural Networks. Proceedings of Machine Learning Research, Fort Lauderdale, FL, p. 315
Greig B., Mesinger A., Bañados E., 2019, MNRAS, 484, 5094
Greig B., Trott C. M., Barry N., Mutch S. J., Pindor B., Webster R. L., Wyithe J. S. B., 2020a, MNRAS, 500, 5322
Greig B. et al., 2020b, MNRAS, 501, 1
Guzman E., Meyers J., 2021, preprint (arXiv:2101.01214)
Harnois-Déraps J., Pen U. L., Iliev I. T., Merz H., Emberson J. D., Desjacques V., 2013, MNRAS, 436, 540
Harris C. R. et al., 2020, Nature, 585, 357
Hassan S., Andrianomena S., Doughty C., 2020, MNRAS, 494, 5761
Hinton G. E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R. R., 2012, preprint (arXiv:1207.0580)
Hortúa H. J., Volpi R., Malagò L., 2020, preprint (arXiv:2005.02299)
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Ichikawa K., Barkana R., Iliev I. T., Mellema G., Shapiro P. R., 2010, MNRAS, 406, 2521
Iliev I. T., Mellema G., Shapiro P. R., Pen U.-L., Mao Y., Koda J., Ahn K., 2012, MNRAS, 423, 2222
Ioffe S., Szegedy C., 2015, preprint (arXiv:1502.03167)
Jarrett K., Kavukcuoglu K., Ranzato M., LeCun Y., 2009, 2009 IEEE 12th international conference on computer vision. IEEE, p. 2146
Jeffrey N., Lanusse F., Lahav O., Starck J.-L., 2020, MNRAS, 492, 5023
Jennings W. D., Watkinson C. A., Abdalla F. B., McEwen J. D., 2018, MNRAS, 483, 2907
Kakiichi K. et al., 2017, MNRAS, 471, 1936
Kapahtia A., Chingangbam P., Ghara R., Appleby S., Choudhury T. R., 2021, J. Cosmol. Astropart. Phys., 2021, 26
Kern N. S., Liu A., Parsons A. R., Mesinger A., Greig B., 2017, ApJ, 848, 23
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Komatsu E. et al., 2009, ApJS, 180, 330
Komatsu E. et al., 2011, ApJS, 192, 18
Koopmans L. V. E. et al., 2015, Proc. Sci., The Cosmic Dawn and Epoch of Reionization with the Square Kilometre Array. SISSA, Trieste, PoS# 001
Lee C., 2019, Planet. Space Sci., 170, 16
Li W. et al., 2019, MNRAS, 485, 2628
Lidz A., Zahn O., McQuinn M., Zaldarriaga M., Dutta S., Hernquist L., 2007, ApJ, 659, 865
Lin Y., Oh S. P., Furlanetto S. R., Sutter P. M., 2016, MNRAS, 461, 3361
Liu A., Tegmark M., 2011, Phys. Rev. D, 83, 103006
Long J., Shelhamer E., Darrell T., 2014, preprint (arXiv:1411.4038)
Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429
Majumdar S., Pritchard J. R., Mondal R., Watkinson C. A., Bharadwaj S., Mellema G., 2018, MNRAS, 476, 4007
Makinen T. L., Lancaster L., Villaescusa-Navarro F., Melchior P., Ho S., Perreault-Levasseur L., Spergel D. N., 2021, J. Cosmol. Astropart. Phys., 04, 081
Mangena T., Hassan S., Santos M. G., 2020, MNRAS, 494, 600
Mao Y., Koda J., Shapiro P. R., Iliev I. T., Mellema G., Park H., Ahn K., Bianco M., 2019, MNRAS, 491, 1600
Mason C. A., Gronke M., 2020, MNRAS, 499, 1395
McGreer I. D., Mesinger A., Fan X., 2011, MNRAS, 415, 3237
McGreer I. D., Mesinger A., D'Odorico V., 2014, MNRAS, 447, 499
Mehta P., Bukov M., Wang C.-H., Day A. G. R., Richardson C., Fisher C. K., Schwab D. J., 2019, Phys. Rep., 810, 1
Mellema G., Iliev I. T., Pen U.-L., Shapiro P. R., 2006, MNRAS, 372, 679
Mellema G. et al., 2013, Exp. Astron., 36, 235
Mertens F. G. et al., 2020, MNRAS, 493, 1662
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Mondal R. et al., 2020, MNRAS, 498, 4178
Murray S. G., Greig B., Mesinger A., Muñoz J. B., Qin Y., Park J., Watkinson C. A., 2020, J. Open Source Softw., 5, 2582
Ota K. et al., 2008, ApJ, 722, 803
Ouchi M. et al., 2010, ApJ, 723, 869
Perez L., Wang J., 2017, preprint (arXiv:1712.04621)
Pober J. C. et al., 2014, ApJ, 782, 66
Pritchard J. R., Furlanetto S. R., 2007, MNRAS, 376, 1680
Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJ, 802, L19
Ronneberger O., Fischer P., Brox T., 2015, preprint (arXiv:1505.04597)
Ross H. E., Dixon K. L., Iliev I. T., Mellema G., 2017, MNRAS, 468, 3785
Ross H. E., Dixon K. L., Ghara R., Iliev I. T., Mellema G., 2019, MNRAS, 487, 1101
Sadr A. V., Farsian F., 2020, preprint (arXiv:2004.04177)
Salehi S. S. M., Erdogmus D., Gholipour A., 2017, preprint (arXiv:1706.05721)
Schmit C. J., Pritchard J. R., 2018, MNRAS, 475, 1213
Schroeder J., Mesinger A., Haiman Z., 2013, MNRAS, 428, 3058
Shapiro P. R., Martel H., Villumsen J. V., Owen J. M., 25 1996, ApJ, 270
Shimabukuro H., Semelin B., 2017, MNRAS, 468, 3869
Shimabukuro H., Mao Y., Tan J., 2020, preprint (arXiv:2002.08238)
Simonyan K., Zisserman A., 2015, preprint (arXiv:1409.1556)
Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, J. Mach. Learn. Res., 15, 1929
Sullivan D., Iliev I. T., Dixon K. L., 2017, MNRAS, 473, 38
Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z., 2015, preprint (arXiv:1409.1556)
Tingay S. J. et al., 2013, Publ. Astron. Soc. Aust., 30, e007
Totani T., Aoki K., Hattori T., Kawai N., 2016, PASJ, 68, 15
Trott C. M. et al., 2020, MNRAS, 493, 4711
van Haarlem M. P. et al., 2013, A&A, 556, A2
van der Walt S. et al., 2014, Scikit-Image: Image Processing in Python, PeerJ 2. p. e453
Villanueva-Domingo P., Villaescusa-Navarro F., 2021, ApJ, 907, 44
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Wang Y., Huang G., Song S., Pan X., Xia Y., Wu C., 2020, preprint (arXiv:2007.10538)
Watkinson C. A., Pritchard J. R., 2015, MNRAS, 454, 1416
Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230
Wyithe S., Geil P. M., Kim H., 2015, Proc. Sci., Imaging HII Regions from Galaxies and Quasars During Reionisation with SKA. SISSA, Trieste, PoS#015
Xu Y., Yue B., Chen X., 2017, ApJ, 844, 117
Yoshiura S., Shimabukuro H., Hasegawa K., Takahashi K., 2020, preprint (arXiv:2004.09206)
Zackrisson E. et al., 2020, MNRAS, 493, 855
Zaroubi S., 2012, in Wiklind T., Mobasher B., Bromm V., eds, Astrophysics and Space Science Library, Vol. 396, The First Galaxies. Springer-Verlag, Berlin, p. 45
Zeiler M. D., Fergus R., 2013, preprint (arXiv:1311.2901)
Zel'Dovich Y. B., 1970, A&A, 500, 13

APPENDIX A: `SegU-Net` HIDDEN LAYER OUTPUTS

We test `SegU-Net` to see if it can recover the binary field for a simple case, namely a single spherical neutral region. We assume a uniform density field at $z = 8.032$ and calculate the differential brightness

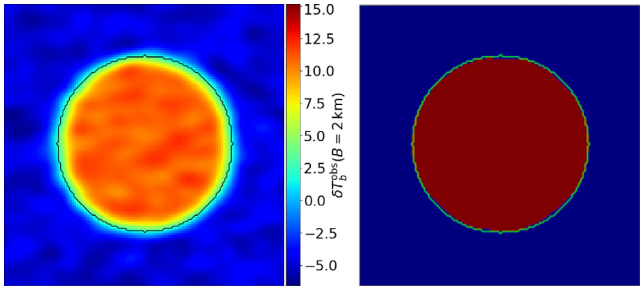

Figure A1. Test `SegU-Net` on a spherical ionized region. *Left-hand panel*: slice through the input image. The colour map shows the differential brightness temperature, and the black contour shows the boundary between neutral and ionized regions. *Right-hand panel*: the recovered binary field with `SegU-Net`. The green contour represents the same boundary again. The identified neutral and ionized regions are indicated in red and blue, respectively.

temperature with equation (2), adding noise corresponding to $t_{\rm obs} = 1000\,{\rm h}$ and reducing the resolution to correspond to a maximum baseline of $B = 2\,{\rm km}$. In Fig. A1, we show the input image of the sphere (left-hand panel) and the corresponding recovered binary field by `SegU-Net` (right-hand panel). The black contour in the left-hand panel and the green contour in the right-hand panel show the true boundary of the sphere. For this test, `SegU-Net` achieves an accuracy of 98 per cent.

In Fig. A2, we show the output of the bottom hidden layer of `SegU-Net`, which is the last layer of the left part of the U-shape in Fig. 3. The colour coding is such that blue corresponds to negative, red to positive, and white to zero values. This output gives a visual representation of the low-dimensional latent space of our network encoder. In our case, this consists of 256 images, where

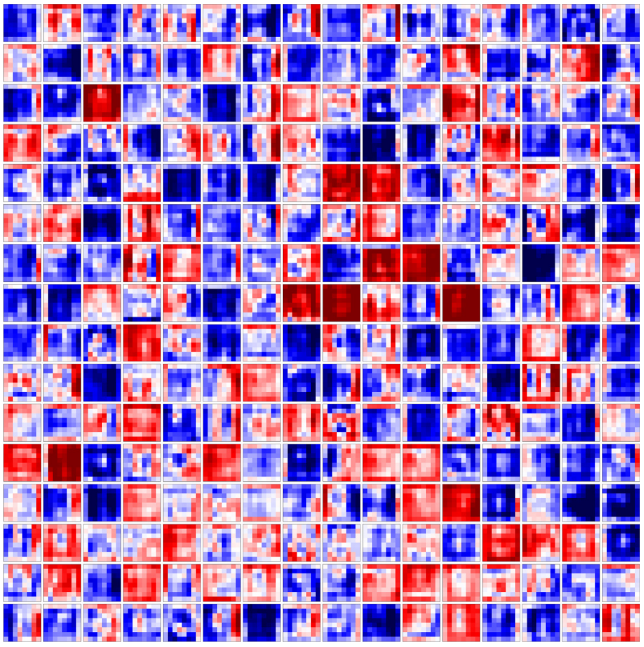

Figure A2. Visual representation of `SegU-Net`'s low-dimensional latent space (bottom layer), which contains information about the extracted features of our test input image.

each corresponds to a convolutional kernel and contains information about the image's extracted features. The encoder path contracts the input image from an original mesh size of 128^2 down to 8^2. The information contained in the latent space is then expanded by `SegU-Net` decoder into a binary map of the same size as the input image (see right-hand panel of Fig. A1).

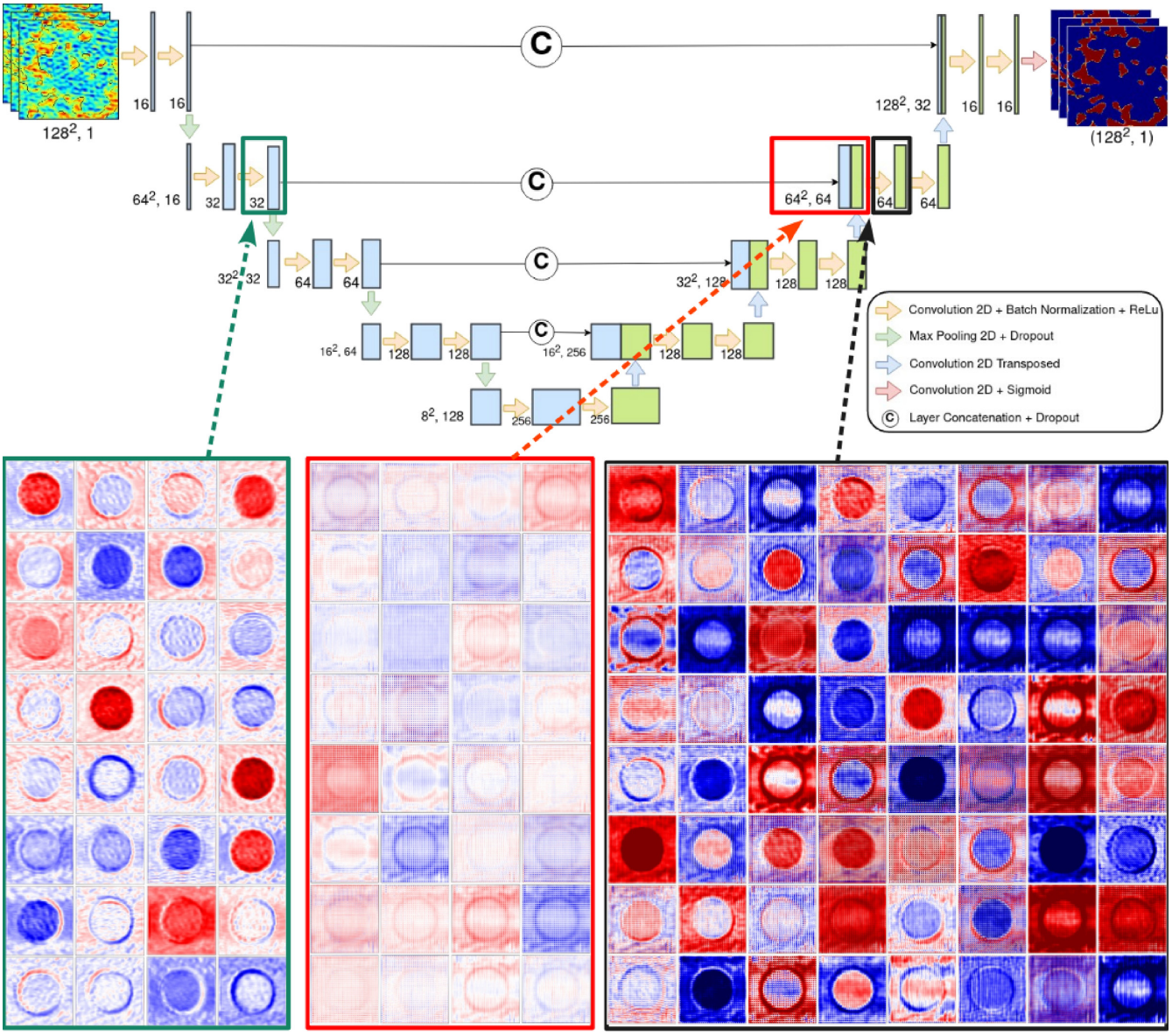

Figure A3. Example of skip connection between encoder and decoder levels. The top panel shows the architecture of our network. The bottom panels display the output of three hidden layers. On the left (green dashed line), a convolutional block (`ConvBlock`) output is interconnected with the output of the second to last up-sampling operation (central panel, red dashed line). The rightmost panel shows the result of the merge after a convolution block (black dashed line).

APPENDIX B: SKIP CONNECTION BETWEEN ENCODER AND DECODER LEVELS

The main advantage of a U-shaped network is that it overcomes the bottleneck limitation encountered by auto-encoder networks (a classical encoder/decoder architecture) by adding interconnections between the descending (encoder) and ascending (decoder) paths (Long et al. 2014; Ronneberger et al. 2015). These connections allow feature representations to pass through the bottleneck (bottom layer) and avoid loss of information due to contraction.

In Fig. A3, we show a visual example of interconnections between the encoder (left part of the U-shape) and the decoder (right part). The top panel shows a schematic representation of our network architecture, and the bottom part displays a visual output of three hidden layers for the test case of a sphere. The leftmost panel (connected by a green dashed line) shows the output of the second convolutional block in the encoder's second level. This block consists of 32 kernels with a mesh size of 64^2. At this level, the shape and form of the input image are still visible. The centre panel (connected by a red dashed line) shows the result of the second to last up-sampling operation of the decoder. The number of kernels and mesh size match with the corresponding encoder layers. The skip connection merges the encoder- and decoder-level output for a total of 64 images with mesh 64^2. The rightmost panel (connected with a black dashed line) shows the concatenation after a convolutional block. The effect of the up-sampling operation is still visible.

This paper has been typeset from a TEX/LATEX file prepared by the author.