

# Separating the EoR Signal with a Convolutional Denoising Autoencoder: a Deep-learning-based Method

Weitian Li,<sup>1</sup>★ Haiguang Xu,<sup>1,2</sup>★ Zhixian Ma,<sup>3</sup> Ruimin Zhu,<sup>4</sup> Dan Hu,<sup>1</sup> Zhenghao Zhu,<sup>1</sup> Chenxi Shan,<sup>1</sup> Jie Zhu<sup>3</sup> and Xiang-Ping Wu<sup>5</sup>

<sup>1</sup>*School of Physics and Astronomy, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China*

<sup>2</sup>*Tsung-Dao Lee Institute / IFSA Collaborative Innovation Center, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China*

<sup>3</sup>*Department of Electronic Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China*

<sup>4</sup>*Department of Statistics, Northwestern University, 2006 Sheridan Road, Evanston, IL 60208, US*

<sup>5</sup>*National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing 100012, China*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

When applying the foreground removal methods to uncover the faint EoR signal, the foreground spectra are assumed to be smooth. However, this assumption can be seriously violated in practice since the unresolved or mis-subtracted foreground sources, when entangled with the frequency-dependent beam effects of the interferometers, leave significant rapid fluctuations along the frequency dimension. To address this issue, we propose a novel deep-learning-based method that uses a 9-layer convolutional denoising autoencoder (CDAE) to separate the EoR signal. After being trained on the SKA images simulated with realistic beam effects, the CDAE achieves excellent performance as the correlation coefficient ( $\rho$ ) between the reconstructed and input EoR signals reaches  $0.969 \pm 0.020$ . In comparison, the traditional polynomial fitting method fails to uncover the EoR signal ( $\rho = 0.241 \pm 0.103$ ). We conclude that the CDAE can overcome the frequency-dependent beam effects and accurately separate the EoR signal, which can be mainly contributed to the powerful technique of hierarchically learning sophisticated features by stacking multiple convolutional layers. Our method also exhibits the great potential of deep-learning-based methods in future EoR experiments.

**Key words:** methods: data analysis – techniques: interferometric – dark ages, reionization, first stars – radio continuum: general

## 1 INTRODUCTION

The 21 cm line emission of neutral hydrogen from the epoch of reionization (EoR) is regarded as a decisive probe to directly explore this stage (see Furlanetto 2016 for a review). To detect the 21 cm signal, which is believed to have been redshifted to the frequencies below 200 MHz, low-frequency radio interferometers such as the SKA (Koopmans et al. 2015) and its pathfinders have been built or under construction. The observational challenges, however, are immense due to complicated instrumental effects, ionospheric distortions, radio frequency interference, and the strong foreground contamination that overwhelms the EoR signal by about 4–5 orders of magnitude (see Morales & Wyithe 2010 for a review). Fortunately, in the frequency dimension the foreground contamination is expected to be intrinsically smooth, while the EoR signal fluctuates rapidly on  $\lesssim$  MHz scales. This difference is the key characteristic exploited by many foreground removal methods in order to uncover the faint EoR signal, including parametric fitting approaches (e.g.,

Wang et al. 2006; Liu et al. 2009b; Wang et al. 2013) and non-parametric approaches (e.g., Harker et al. 2009; Chapman et al. 2013; Mertens et al. 2018).

However, the smoothness of the foreground spectra can be destroyed by the frequency-dependent beam effects, i.e., the variation of the point spread function (PSF) with frequencies that cannot be perfectly calibrated (Liu et al. 2009a). Because of the incomplete  $uv$  coverage, the PSF has a complicated profile consisting of a narrow peaky main lobe and a multitude of jagged side lobes with relative amplitudes of about 0.1–1 per cent that extend beyond the field of view (FoV) (e.g., Liu et al. 2009a, figures 1 and 3). A source that is unresolved or mis-subtracted (e.g., due to the limited FoV) during the CLEAN process leaves catastrophic residuals, the locations of which vary with the frequency since the angular position of a PSF side lobe is inversely proportional to the frequency. These effects lead to complicated residuals fluctuating along the frequency dimension, which cannot be correctly separated from the EoR signal by the traditional foreground removal methods that rely on the smoothness of the foreground spectra.

★ E-mail: liweitianux@sjtu.edu.cn (WL); hgxu@sjtu.edu.cn (HX)

Given the complicated profiles and frequency-dependent vari-

ations of the PSF, it would be difficult to craft a practicable model for most, if not all, existing foreground removal methods to overcome the beam effects, even at the cost of extensive computation burden (e.g., Lochner et al. 2015). Therefore deep-learning-based methods, which can distill knowledge from the data to automatically refine the model, seem more feasible and appealing (e.g., Herbel et al. 2018; Vafaei Sadr et al. 2018). In recent years, the deep learning algorithms have seen prosperous developments and have brought breakthroughs into many fields (see LeCun et al. 2015 for a recent review). Among various deep learning algorithms, the convolutional denoising autoencoder (CDAE) and its variants are flexible and powerful in learning subtle and complicated features from the data and have been successfully applied to weak gravitational wave signal denoising (e.g., Shen et al. 2017), monaural audio source separation (e.g., Grais & Plumbley 2017), and so on. These applications have demonstrated the outstanding abilities of the CDAE in extracting weak signals from highly temporal-variable data, thus it is worth trying to apply the CDAE to separate the EoR signal. Although the signal-to-noise ratio in the EoR detection is much lower than in existing applications, the EoR signal and foreground emission as well as the beam effects are stationary or semi-stationary.

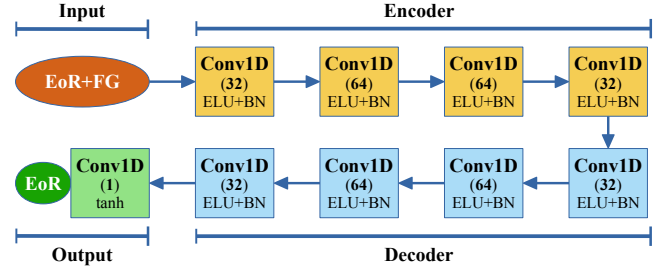
In this paper, a novel deep-learning-based method that uses a CDAE is proposed to tackle the intricate frequency-dependent beam effects and to separate the EoR signal along the frequency dimension. In Section 2, we briefly introduce the CDAE and elaborate the proposed method. In Section 3, we demonstrate the performance of the CDAE by applying it to the simulated SKA images. We discuss the method and carry out a comparison to the traditional polynomial fitting method in Section 4. Finally, we summarize our work in Section 5. The implementation code and data are made public at <https://github.com/lwieitianux/cdae-eor>.

## 2 METHODOLOGY

### 2.1 Convolutional denoising autoencoder

An autoencoder is composed of an encoder and a decoder, which can be characterized by the functions  $f(\cdot)$  and  $g(\cdot)$ , respectively. The encoder maps the input  $\mathbf{x}$  to an internal code  $\mathbf{h}$ , i.e.,  $\mathbf{h} = f(\mathbf{x})$ , and the decoder tries to reconstruct the wanted signal from the code  $\mathbf{h}$ , i.e.,  $\mathbf{r} = g(\mathbf{h})$ , where  $\mathbf{x}$ ,  $\mathbf{h}$ , and  $\mathbf{r}$  are all vectors for this work. By placing constraints (e.g., dimensionality, sparsity) on the internal code  $\mathbf{h}$  and training the autoencoder to minimise the loss  $L(\mathbf{r}, \mathbf{x})$ , which quantifies the difference between the reconstruction  $\mathbf{r}$  and the input  $\mathbf{x}$ , the autoencoder tries to learn the codes that effectively represent the input data (e.g., Goodfellow et al. 2016, chapter 14). When being trained to reconstruct the noise-free signal from the noisy input data, the autoencoder is able to learn robust features that can be used to denoise the input data and to recover the noise-free signal as much as possible (i.e., denoising autoencoder; Vincent et al. 2008, 2010). Furthermore, by introducing the architecture of stacking multiple convolutional layers, one of the major innovations of the convolutional neural network, into the denoising autoencoder, the CDAE gains the ability of learning more and more sophisticated features from the data in a hierarchical way (Du et al. 2017) and can be employed to distinguish the subtle differences between the EoR signal and the foreground emission.

In the EoR detection task, the foreground emission  $\mathbf{x}_{fg}$  can be regarded as the strong noise that suppresses the EoR signal  $\mathbf{x}_{EoR}$ . The CDAE is then employed to learn the features of the EoR signal from



**Figure 1.** The network architecture of the proposed CDAE, consisting of a 4-layer encoder (the orange boxes), a 4-layer decoder (the blue boxes), and one output layer (the green box). The layers in the encoder and decoder use the ELU activation function and batch normalisation (BN), while the output layer uses the 'tanh' activation function. All filters are vectors of length 3 and the numbers of filters in each layer are marked in braces.

the input ( $\mathbf{x} = \mathbf{x}_{EoR} + \mathbf{x}_{fg}$ ) and to obtain the reconstructed signal  $\mathbf{r}_{EoR}$  that is expected to be as close to the input  $\mathbf{x}_{EoR}$  as possible.

### 2.2 Network architecture

We follow common practices to build the CDAE architecture (e.g., Suganuma et al. 2018; Géron 2017). The encoder part is symmetric to the decoder part, and both of them use the exponential linear unit (ELU) as the activation function (Clevert et al. 2016), while the output layer uses the hyperbolic tangent function (i.e., tanh; see also Section 3.2). The batch normalisation is applied to all layers except for the output layer to improve the training process as well as to act as a regularizer to prevent overfitting (Ioffe & Szegedy 2015). The filters are vectors of length 3 in all layers since each of the CDAE's input  $\mathbf{x}$  is the emission of one sky pixel, i.e., a vector of length  $n_f$  with  $n_f$  being the number of frequency channels in the image cube (see also Section 3.1).

We have evaluated the performance of multiple CDAE architectures, each containing a different number of layers and filters. The simplest one with sufficiently good performance is selected, which consists of a 4-layer encoder with (32, 64, 64, 32) filters, a 4-layer decoder with (32, 64, 64, 32) filters, and one output layer, as illustrated in Fig. 1.

### 2.3 Training and evaluation

The CDAE is initialised by the He uniform initialiser (He et al. 2015) and is trained using the Adam optimisation method (Kingma & Ba 2015). The loss, which describes the difference between the reconstructed EoR signal  $\mathbf{r}_{EoR}$  and the input EoR signal  $\mathbf{x}_{EoR}$ , is calculated as the mean squared error (MSE), i.e.,

$$L = \frac{1}{N} \sum_{i=1}^N \left[ \mathbf{r}_{EoR}^{(i)} - \mathbf{x}_{EoR}^{(i)} \right]^T \left[ \mathbf{r}_{EoR}^{(i)} - \mathbf{x}_{EoR}^{(i)} \right], \quad (1)$$

where  $N$  is the number of data points in the training dataset. By being trained to minimise the loss, the CDAE gains the ability to reconstruct the EoR signal from the input data  $\mathbf{x}$ .

To evaluate the performance of the CDAE in separating the EoR signal, the commonly used Pearson's correlation coefficient (e.g., Harker et al. 2009; Chapman et al. 2013) is adopted to measure the similarity between the reconstructed EoR signal  $\mathbf{r}_{EoR}$  and the

input EoR signal  $\mathbf{x}_{\text{EoR}}$ :

$$\rho = \frac{\sum_{j=1}^n (r_{\text{EoR},j} - \bar{r}_{\text{EoR}})(x_{\text{EoR},j} - \bar{x}_{\text{EoR}})}{\sqrt{\sum_{j=1}^n (r_{\text{EoR},j} - \bar{r}_{\text{EoR}})^2 \sum_{j=1}^n (x_{\text{EoR},j} - \bar{x}_{\text{EoR}})^2}}, \quad (2)$$

where  $n$  is the length of the signals, and  $\bar{r}_{\text{EoR}}$  and  $\bar{x}_{\text{EoR}}$  represent the mean values. The closer the correlation coefficient is to one, the better the performance of separation.

### 3 EXPERIMENTS

#### 3.1 Simulation of the SKA images

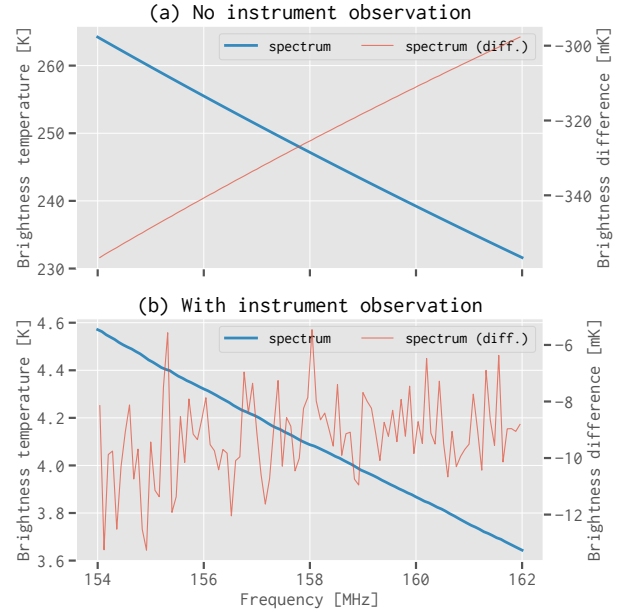
We carry out end-to-end simulations to generate the SKA images to train the proposed CDAE and evaluate its performance. We choose the 8 MHz frequency band centred at 158 MHz as an example, which is divided into  $n_f = 101$  channels with a resolution of 80 kHz. Following the approach described in detail in our previous works (Wang et al. 2010, 2013), we simulate the sky maps of the foreground emission contributed by the Galactic synchrotron and free-free emissions, extragalactic point sources (the bright ones with a 158 MHz flux density  $S_{158} > 10$  mJy are assumed to have been removed; e.g., Liu et al. 2009a), and radio haloes. The sky maps of the EoR signal are created using the ‘faint galaxies’ data released by the Evolution Of 21 cm Structure project (EOS; Mesinger et al. 2016). The simulated sky maps cover a sky area of  $10^\circ \times 10^\circ$  and are pixelised into  $1800 \times 1800$  with a pixel size of  $20''$ .

To fully take into account the frequency-dependent beam effects, the SKA1-Low layout configuration<sup>1</sup> is employed in the OSKAR<sup>2</sup> (Mort et al. 2010) to simulate observations for the sky maps, yielding the visibility data, from which the observed images are created by the WSCLEAN<sup>3</sup> imager (Offringa et al. 2014) with the natural weighting and baseline range of 30–1000 wavelengths in order to emphasise the faint diffuse EoR signal. The created images are cropped to keep only the central  $2^\circ \times 2^\circ$  regions (i.e.,  $360 \times 360$  pixels) for the sake of the best imaging quality. Therefore, we obtain two image cubes of the size  $360 \times 360 \times 101$  for the EoR signal ( $C_{\text{EoR}}$ ) and the foreground emission ( $C_{\text{fg}}$ ), respectively. Example spectra and the corresponding differential spectra for the foreground emission with and without the beam effects are shown in Fig. 2. As clearly shown, when the beam effects are considered, the smoothness of the foreground spectrum is seriously damaged by the jagged fluctuations with strength of about 10 mK (the bottom panel).

#### 3.2 Data preprocessing

The dataset  $\mathcal{S} = \{(\mathbf{x}, \mathbf{x}_{\text{EoR}})\}$  for the CDAE is derived from the simulated image cubes  $C_{\text{EoR}}$  and  $C_{\text{fg}}$ , each data point ( $\mathbf{x} = \mathbf{x}_{\text{EoR}} + \mathbf{x}_{\text{fg}}$ ,  $\mathbf{x}_{\text{EoR}}$ ) representing the total emission and the EoR signal of one sky pixel, respectively. The dataset thus consists of  $N = 360 \times 360 = 129\,600$  data points.

For the input data  $X = \{\mathbf{x}\}$ , we propose to apply the Fourier Transform (FT) along the frequency dimension, which makes the EoR signal more distinguishable from the foreground emission and thus easier to be learned by the CDAE (a comparison with the results



**Figure 2.** Example spectra (the bold blue lines) and the corresponding differential spectra (the thin red lines) for the foreground emission. The top and bottom panels show the cases without and with instrument observation, respectively.

derived without applying the FT is presented in Section 4.1). The Blackman-Nuttall window function is applied to suppress the FT side-lobes caused by the sharp discontinuities at both ends of the finite frequency band (e.g., Chapman et al. 2016). It is sufficient to keep only half the Fourier coefficients because  $\mathbf{x}$  is real, thus  $\mathbf{x}$  of length  $n_f = 101$  is transformed to be 51 complex Fourier coefficients, among which the  $n_{\text{ex}}$  coefficients of the lowest Fourier frequencies are excised since they are mostly contributed by the spectral-smooth foreground emission. We adopt  $n_{\text{ex}} = 6$  to achieve a balance between the foreground suppression and the loss of the EoR signal. The real and imaginary parts of the remaining 45 complex coefficients are then concatenated into a new real vector of length 90, since the CDAE requires real data. Finally, the data are zero-centred and normalised to have unit variance.

The preprocessing steps for the input EoR signal  $X_{\text{EoR}} = \{\mathbf{x}_{\text{EoR}}\}$  are basically the same except for minor adjustments. After applying the FT, excising the  $n_{\text{ex}}$  lowest Fourier components, and concatenating the real and imaginary parts, the data elements that have a value less than the 1<sup>st</sup> percentile or greater than the 99<sup>th</sup> percentile are truncated, in order to prevent the possible outliers hindering the training of the CDAE. Finally, the value range of the data is scaled to be  $[-1, 1]$  by dividing by the maximum absolute value, which allows to use the ‘tanh’ activation function whose value range is also  $[-1, 1]$  in the output layer of the CDAE (Section 2.2).

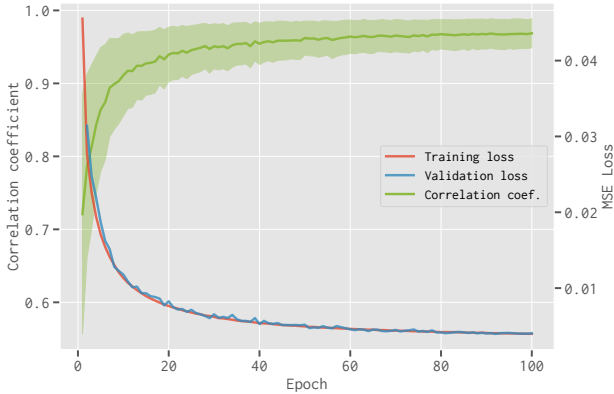
#### 3.3 Training and results

The preprocessed dataset is randomly partitioned into the training set ( $S_{\text{tr}}$ ; 60 per cent) to train the parameters of the CDAE by minimising the loss, the validation set ( $S_{\text{val}}$ ; 20 per cent) to determine the hyperparameters (e.g., the number of layers and filters, the choice of loss function), and the test set ( $S_{\text{test}}$ ; 20 per cent) that is employed to evaluate the performance of the trained CDAE.

<sup>1</sup> SKA1-Low layout: [https://astronomers.skatelescope.org/wp-content/uploads/2016/09/SKA-TEL-SK0-0000422\\_02\\_SKA1\\_LowConfigurationCoordinates-1.pdf](https://astronomers.skatelescope.org/wp-content/uploads/2016/09/SKA-TEL-SK0-0000422_02_SKA1_LowConfigurationCoordinates-1.pdf)

<sup>2</sup> OSKAR: <https://github.com/OxfordSKA/OSKAR> (version 2.7.0)

<sup>3</sup> WSClean: <https://sourceforge.net/p/wsclean> (version 2.5)



**Figure 3.** The training loss (the red line), validation loss (the blue line), and correlation coefficient ( $\rho$ ) calculated on the validation set  $S_{\text{val}}$  (the green line with the shaded region representing the standard deviation) along the training of the CDAE.

We implement the proposed CDAE using the popular KERAS<sup>4</sup> framework (Chollet et al. 2015) with the TENSORFLOW<sup>5</sup> back end (Abadi et al. 2015). The parameters of the Adam optimisation method are set to the default values (learning rate  $\alpha = 0.001$ , exponential decay rates for the first moment estimates  $\beta_1 = 0.9$  and for the second moment estimates  $\beta_2 = 0.999$ ; Kingma & Ba 2015). The CDAE is trained on the training set ( $S_{\text{tr}}$ ) with a batch size of 100 for 100 epochs until the training loss converges.

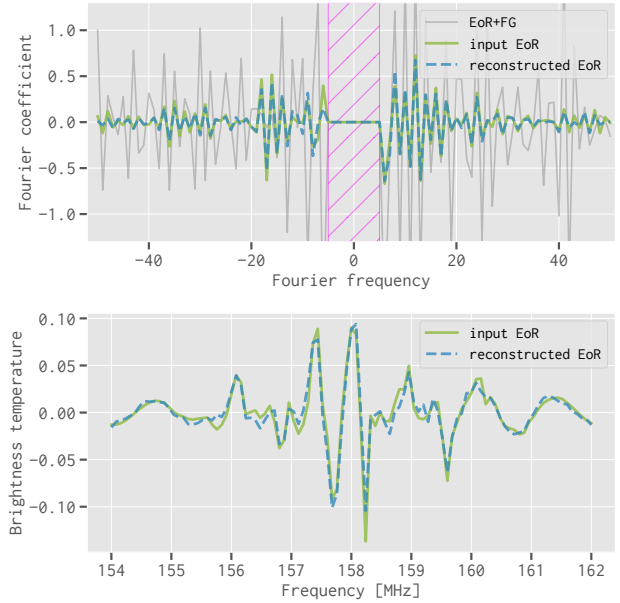
The training and validation losses together with the evaluation index (i.e., the correlation coefficient) calculated on the validation set ( $S_{\text{val}}$ ) during the training phase are shown in Fig. 3. The steadily decreasing losses and increasing correlation coefficient suggest that the CDAE is well trained without overfitting. By evaluating on the test set ( $S_{\text{test}}$ ), the trained CDAE achieves excellent performance with a correlation coefficient of  $\rho_{\text{CDAE}} = 0.969 \pm 0.020$ . As an example, Fig. 4 illustrates the reconstructed EoR signal ( $\rho = 0.965$ ) for one random pixel. In addition, we calculate the one-dimensional power spectra for each pixel in  $S_{\text{test}}$  (e.g., Chapman et al. 2013) and find that the reconstructed EoR signal can recover the total power to a fraction of  $R_{\text{CDAE}} = 90.1 \pm 8.9$  per cent.

The achieved excellent performance of the CDAE can be mainly attributed to the architecture of stacking multiple convolutional layers, which implements a powerful feature extraction technique by hierarchically combining the simpler features learned in each layer to build more and more sophisticated features (LeCun et al. 2015). Combined with its flexibility provided by the 54 337 trainable parameters, the CDAE, after being well trained, can learn a model that is optimised to accurately separate the faint EoR signal (e.g., Domingos 2012).

## 4 DISCUSSIONS

### 4.1 Why preprocess the dataset with Fourier Transform?

We have performed another experiment using the same CDAE architecture, dataset, and data preprocessing steps, except for applying



**Figure 4.** An example of the EoR signal reconstructed by the trained CDAE for one random pixel. **(top)** The input EoR signal  $x_{\text{EoR}}$  (the solid green line) and the reconstructed EoR signal  $r_{\text{EoR}}$  (the dashed blue line) in the Fourier domain. The grey line represents the input total emission  $x = x_{\text{fg}} + x_{\text{EoR}}$ . The magenta hatched region marks the excised Fourier coefficients in data preprocessing. **(bottom)** The input EoR signal  $x_{\text{EoR}}$  (the solid green line) and the reconstructed EoR signal  $r_{\text{EoR}}$  (the dashed blue line) transformed back to the observing frequency domain.

the FT as depicted in Section 3.2. After training the CDAE in the same way as described in Section 3.3, the achieved performance is  $\rho_{\text{noFT}} = 0.927 \pm 0.051$ , which is smaller and has a larger uncertainty than the case with FT applied, indicating a worse performance. We also find that the training process is slightly unstable given the small spikes on the curves of both the training loss and the correlation coefficient, and the training loss converges more slowly, as presented in Fig. 5. These indicate that it is beneficial to preprocess the dataset by applying the FT along the frequency dimension, because the EoR signal and the foreground emission become more distinguishable in the Fourier domain, where the fluctuating EoR signal concentrates on larger Fourier modes while the spectral-smooth foreground emission distributes mainly on smaller Fourier modes (e.g., Parsons et al. 2012).

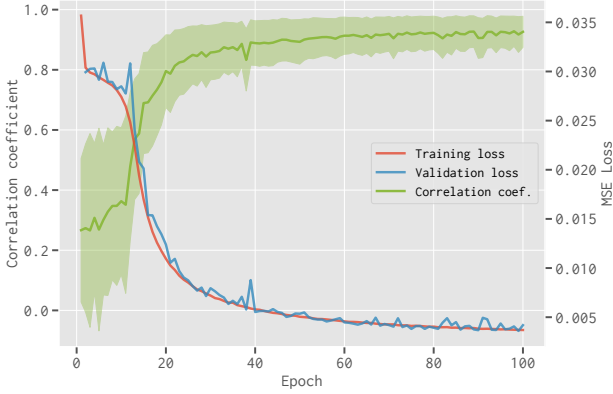
### 4.2 Comparing with the polynomial fitting method

In order to further demonstrate the performance of our method, we have carried out a comparison between our method and the traditional polynomial fitting method (e.g., Wang et al. 2006; Liu et al. 2009a). Using the same image cubes simulated in Section 3.1, a low-degree polynomial is fitted along the frequency dimension for each sky pixel in the image cube of the total emission (i.e.,  $C_{\text{tot}} = C_{\text{EoR}} + C_{\text{fg}}$ ). Then by subtracting the fitted smooth component, which is regarded as the foreground emission, the EoR signal is uncovered. We have tested polynomials of the degree from 2 (quadratic) to 5 (quintic), and find that the quartic polynomial (degree of 4) can give the best result. However, the correlation coefficient calculated for the separated EoR signal in such a case is only

<sup>4</sup> Keras: <https://keras.io> (version 2.2.0)

<sup>5</sup> TensorFlow: <https://www.tensorflow.org> (version 1.4.1)





**Figure 5.** Same as Fig. 3 but for the case that the data are preprocessed without applying the FT.

$\rho_{\text{poly}} = 0.241 \pm 0.103$ , which indicates that the polynomial fitting method performs poorly in removing the foreground emission. As illustrated in Fig. 2(b), the frequency-dependent beam effects cause rapid fluctuations, which can be even stronger than the EoR signal, on the originally smooth foreground spectrum. The polynomial fitting method, which can only model the smooth spectrum, is unable to distinguish these rapid fluctuations from the EoR signal, hence leading to inferior results. On the contrary, given its flexibility and data-driven nature, the CDAE can distill knowledge from the data to optimise itself for the EoR signal separation and hence achieve superior performance.

## 5 SUMMARY

The frequency-dependent beam effects of interferometers can cause rapid fluctuations along the frequency dimension, which destroy the smoothness of the foreground spectra and prevent traditional foreground removal methods from uncovering the EoR signal. Given the difficulty in crafting practicable models to overcome the complicated beam effects, methods that can learn tailored models from the data seem more feasible and appealing. To this end, we have proposed a deep-learning-based method that uses a 9-layer CDAE to separate the EoR signal. The CDAE has been trained on the simulated SKA images and has achieved excellent performance. We conclude that the CDAE has outstanding ability to overcome the intricate beam effects and accurately separate the EoR signal, exhibiting the great potential of deep-learning-based methods to play an important role in the forthcoming EoR experiments.

## ACKNOWLEDGEMENTS

We would like to thank Jeffrey Hsu for reading the manuscript and providing suggestions. This work is supported by the Ministry of Science and Technology of China (grant Nos. 2018YFA0404601, 2017YFF0210903), and the National Natural Science Foundation of China (grant Nos. 11433002, 11621303, 11835009, 61371147).

## REFERENCES

- Abadi M., et al., 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, <https://www.tensorflow.org/>
- Chapman E., et al., 2013, *MNRAS*, **429**, 165
- Chapman E., Zaroubi S., Abdalla F. B., Dulwich F., Jelić V., Mort B., 2016, *MNRAS*, **458**, 2928
- Chollet F., et al., 2015, Keras, <https://keras.io>
- Clevert D.-A., Unterthiner T., Hochreiter S., 2016, in The International Conference on Learning Representations (ICLR 2016). ([arXiv:1511.07289](https://arxiv.org/abs/1511.07289))
- Domingos P., 2012, *Communications of the ACM*, **55**, 78
- Du B., Xiong W., Wu J., Zhang L., Zhang L., Tao D., 2017, *IEEE Transactions on Cybernetics*, **47**, 1017
- Furlanetto S. R., 2016, *Understanding the Epoch of Cosmic Reionization: Challenges and Progress*, **423**, 247
- Géron A., 2017, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 1st edn. O'Reilly Media, Inc.
- Goodfellow I., Bengio Y., Courville A., 2016, Deep Learning. MIT Press, <http://www.deeplearningbook.org>
- Grais E. M., Plumbley M. D., 2017, in 5th IEEE Global Conference on Signal and Information Processing (GlobalSIP 2017). IEEE, pp 1265–1269 ([arXiv:1703.08019](https://arxiv.org/abs/1703.08019)), [doi:10.1109/GlobalSIP.2017.8309164](https://doi.org/10.1109/GlobalSIP.2017.8309164)
- Harker G., et al., 2009, *MNRAS*, **397**, 1138
- He K., Zhang X., Ren S., Sun J., 2015, in Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV 2015). IEEE Computer Society, Washington, DC, USA, pp 1026–1034, [doi:10.1109/ICCV.2015.123](https://doi.org/10.1109/ICCV.2015.123)
- Herbel J., Kacprzak T., Amara A., Refregier A., Lucchi A., 2018, *Journal of Cosmology and Astro-Particle Physics*, **2018**, 054
- Ioffe S., Szegedy C., 2015, in Proceedings of the 32nd International Conference on Machine Learning (ICML 2015). JMLR.org, pp 448–456
- Kingma D. P., Ba J., 2015, in The International Conference on Learning Representations (ICLR 2015). ([arXiv:1412.6980](https://arxiv.org/abs/1412.6980))
- Koopmans L., et al., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), **p. 1**
- LeCun Y., Bengio Y., Hinton G., 2015, *Nature*, **521**, 436
- Liu A., Tegmark M., Zaldarriaga M., 2009a, *MNRAS*, **394**, 1575
- Liu A., Tegmark M., Bowman J., Hewitt J., Zaldarriaga M., 2009b, *MNRAS*, **398**, 401
- Lochner M., Natarajan I., Zwart J. T. L., Smirnov O., Bassett B. A., Oozeer N., Kunz M., 2015, *MNRAS*, **450**, 1308
- Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, *MNRAS*, **478**, 3640
- Mesinger A., Greig B., Sobacchi E., 2016, *MNRAS*, **459**, 2342
- Morales M. F., Wyithe J. S. B., 2010, *ARA&A*, **48**, 127
- Mort B. J., Dulwich F., Salvini S., Adami K. Z., Jones M. E., 2010, in 2010 IEEE International Symposium on Phased Array Systems and Technology. pp 690–694, [doi:10.1109/ARRAY.2010.5613289](https://doi.org/10.1109/ARRAY.2010.5613289)
- Offringa A. R., et al., 2014, *MNRAS*, **444**, 606
- Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012, *ApJ*, **756**, 165
- Shen H., George D., Huerta E. A., Zhao Z., 2017, preprint, ([arXiv:1711.09919](https://arxiv.org/abs/1711.09919))
- Suganuma M., Ozay M., Okatani T., 2018, preprint, ([arXiv:1803.00370](https://arxiv.org/abs/1803.00370))
- Vafaei Sadr A., Vos E. E., Bassett B. A., Hosenie Z., Oozeer N., Lochner M., 2018, preprint, ([arXiv:1807.02701](https://arxiv.org/abs/1807.02701))
- Vincent P., Larochelle H., Bengio Y., Manzagol P.-A., 2008, in Proceedings of the 25th International Conference on Machine Learning (ICML 2008). ACM, pp 1096–1103, [doi:10.1145/1390156.1390294](https://doi.org/10.1145/1390156.1390294)
- Vincent P., Larochelle H., Lajoie I., Bengio Y., Manzagol P.-A., 2010, The Journal of Machine Learning Research, **11**, 3371
- Wang X., Tegmark M., Santos M. G., Knox L., 2006, *ApJ*, **650**, 529
- Wang J., et al., 2010, *ApJ*, **723**, 620
- Wang J., et al., 2013, *ApJ*, **763**, 90

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