Deep learning for intensity mapping observations: Component extraction

Kana Moriwaki\textsuperscript{1}, Nina Filippova\textsuperscript{1,2}, Masato Shirasaki\textsuperscript{3,4}, Naoki Yoshida\textsuperscript{1,5,6,7}

\textsuperscript{1}Department of Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan
\textsuperscript{2}Department of Physics, Princeton University, Princeton, NJ 08544, USA
\textsuperscript{3}National Astronomical Observatory of Japan (NAOJ), Mitaka, Tokyo 181-8588, Japan
\textsuperscript{4}Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
\textsuperscript{5}Kavli Institute for the Physics and Mathematics of the Universe (WPI), UT Institutes for Advanced Study, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan
\textsuperscript{6}Research Center for the Early Universe, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan
\textsuperscript{7}Institute for Physics of Intelligence, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan

\section*{ABSTRACT}

Line intensity mapping (LIM) is an emerging observational method to study the large-scale structure of the universe and its evolution. LIM does not resolve individual sources but probes the fluctuations of integrated line emissions. A serious limitation with LIM is that contributions of different emission lines from sources at different redshifts are all confused at an observed wavelength. We propose a deep learning application to solve this problem. We use conditional generative adversarial networks to extract designated information from LIM. We consider a simple case with two populations of emission line galaxies; H\textsc{\textalpha} emitting galaxies at \(z = 1.3\) are confused with [O\textsc{\textthree}] emitters at \(z = 2.0\) in a single observed waveband at 1.5 \(\mu\text{m}\). Our networks trained with 30,000 mock observation maps are able to extract the total intensity and the spatial distribution of H\textsc{\textalpha} emitting galaxies at \(z = 1.3\). The intensity peaks, i.e. galaxy clusters, are successfully located with 61\% precision. The precision increases to 83\% when we combine the results of 5 networks. The mean intensity and the power spectrum are reconstructed with an accuracy of \(\sim 10\%\). The extracted galaxy distributions at a wider range of redshift can be used for studies on cosmology and on galaxy formation and evolution.

\textbf{Key words:} galaxies: high-redshift — large-scale structure of Universe — cosmology: observations

\section{INTRODUCTION}

Line intensity mapping (LIM) is a promising observational technique for next-generation cosmology. LIM probes the large-scale structure of the Universe at a wide range of redshift, and thus enable us to study cosmology as well as galaxy formation and evolution (Kovetz et al. 2017). Fluctuations of the integrated intensity of emission lines such as Lyman-\(\alpha\), H\textsc{\textalpha}, [C\textsc{\textii}], and CO lines, trace the distributions of the underlying galaxies, while hydrogen 21-cm line can be used to study the distribution and ionization state of the intergalactic medium in the early universe (Pritchard & Loeb 2012).

A number of LIM observation programmes have been proposed and are planned (see Kovetz et al. 2017). LIM measures the integrated emission from all the sources redshifted to a wavelength bin. While it can provide rich information on the sources and their large-scale distribution in principle, the confusion of sources or contamination from foreground/background emission is an inevitable problem in practice. Fonseca et al. (2017) show that multiple emission lines from galaxies often contribute roughly equally to the total intensity at a certain observed wavelength. There are a few methods to infer the contribution from a designated redshift. One is to perform cross-correlation analysis with other known tracers of galaxies or of the matter distribution at the same redshift (e.g. Visbal & Loeb 2010). More practical methods such as masking brightest pixels allow to detect subdominant signals (Gong et al. 2014). It is also possible to distinguish signals from different redshifts using the anisotropic power spectrum shapes (e.g. Cheng et al. 2016). These methods are aimed at esti-
2 K. Moriwaki et al.

mating the statistical quantities, but do not generate direct images of the intensity distribution. It would be more informative and useful if contaminants are removed from an image to show explicitly the intensity distribution at an arbitrary redshift. Here, we propose to use deep learning to separate/extract information from intensity maps.

Convolutional neural networks (CNNs) are a popular and promising tool for image processing including problems related to LIM. Recent studies propose to use CNNs to analyze hydrogen 21-cm line signals from the epoch of reionization (Hassan et al. 2019; Hassan, Andrianomena & Doughty 2019; Gillet et al. 2019; Zamudio-Fernandez et al. 2019) or to estimate the line luminosity function from a CO intensity map (Pfeffer, Breysse & Stein 2019; Shirasaki, Yoshida & Ikeda 2019) use conditional generative adversarial networks (cGANs, e.g., Isola et al. 2016) to de-noise observed weak-lensing mass maps. A cGAN consists of a pair of CNNs which learn an image-to-image translation in an adversarial way, and can generate fine and complicated images.

In this Letter, we apply cGANs to intensity maps to reconstruct the intensity distribution and basic statistics of galaxy distribution. We aim at decoding cosmological information from future intensity map observations using ground-based and space-borne telescopes. We show that our networks, after appropriately trained with a large set of mock observations, can generate accurately the intensity distribution from a single source population. Throughout this Letter, we adopt $\Lambda$CDM cosmology with $\Omega_M = 0.316, \Omega_L = 0.684, H_0 = 67.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration VI 2018).

2 METHODS

We consider the line intensity observed at wavelength of 1.5 $\mu$m. Planned or proposed near-infrared LIM projects include the SpectroPhotometer for the History of the Universe, Epoch of Reionization, and Ice Explorer (SPHEREx, Dore et al. 2016) and the Cosmic Dawn Intensity Mapper (CDIM, Cooray et al. 2019). Emission lines from galaxies at $z \sim 0.5$ are considered to be the dominant sources in this spectral regime. As a simple but realistic case, we assume that the observed line intensity map consists of two most dominant emission lines: H$\alpha$ line from $z = 1.3$ and [OII] 5007Å line from $z = 2.0$. Observational noises and other contaminants such as [OIII] 3727Å are to be considered in a forthcoming paper (Moriwaki et al. in preparation).

2.1 Mock intensity maps

We generate a number of mock intensity maps for training and testing in the following manner. We use the second-order Lagrangian perturbation theory (Crocco, Pueblas & Scoccimarro 2006) to generate matter density fluctuations with $512^3$ particles in a cubic volume of $300 h^{-1}$ cMpc. For a matter density field $\delta(x)$, we calculate the line intensity

$$I_{\text{line}}(x) = \bar{I}_{\text{line}}(1 + b_{\text{line}} \delta(x)),$$

where $\bar{I}_{\text{line}}$ and $b_{\text{line}}$ are the mean intensity and the luminosity-weighted bias, respectively.

We compute $\bar{I}_{\text{line}}$ and $b_{\text{line}}$ from the outputs of the cosmological hydrodynamics simulation Illustris-TNG (Nelson et al. 2019). We use the dataset of TNG300-1 which has a simulated volume of $V_{\text{box}} = (302.6 \text{ cMpc})^3$. The line luminosity from a simulated galaxy is computed as

$$L_{\text{line}} = 10^{-A_{\text{line}}/2.5} C_{\text{line}}(Z) \text{ SFR},$$

where $A_{\text{line}}$ accounts for attenuation by dust, which we adopt $A_{\text{H}$ = $1.0 \text{ mag and } A_{[\text{OIII}]} = 1.35 \text{ mag.}$ Z is the mean metallicity of the galaxy, and $C_{\text{line}}(Z)$ is a coefficient computed with the photoionization calculation code CLOUDY (Ferland et al. 2017). The CLOUDY computation is done in the same manner as in Moriwaki et al. (2018) except that we adopt constant values of the electron density $n = 100 \text{ cm}^{-3}$ and the ionization parameter $U = 0.01$.

The mean line intensity is calculated as

$$\bar{I}_{\text{line}} = \frac{L_{\text{line, tot}} \chi^2 d\chi}{4 \pi D_L^2 V_{\text{box}}}$$

where $L_{\text{line, tot}}$ is the total luminosity within the simulation box, $D_L$ is the luminosity distance, and $\chi$ is the comoving distance. We also compute the luminosity-weighted bias as

$$b_{\text{line}} = \frac{\int dM b(M) L_{\text{line}}(M) \frac{dn}{dM}}{\int dM L_{\text{line}}(M) \frac{dn}{dM}},$$

where $b(M)$ is the halo bias, $L_{\text{line}}(M)$ and $dn/dM$ are the halo mass-luminosity relation and the mass function of the DM halo in the simulation. The computed mean intensities and the biases are summarized in Table 2.

Table 1. The mean intensities in units of [erg s$^{-1}$ cm$^{-2}$ sr$^{-1}$] and the luminosity weighted bias.

| redshift | line   | $v_{\text{obs}} I_{\text{line}}$ | $b_{\text{line}}$ |
|----------|--------|-------------------------------|-----------------|
| 1.3      | H$\alpha$ | $1.5 \times 10^{-8}$ | 1.2 |
| 2.0      | [OII]  | $8.7 \times 10^{-9}$ | 1.6 |

2.2 Convolutional networks

We construct a cGAN based on pix2pix by Isola et al. 2016. We train the network so that they output both H$\alpha$ and [OII] images from an observed image. This kind of

1 https://github.com/yenchenlin/pix2pix-tensorflow
one-to-many image translation network is studied by, for instance, Lee, Yang & Oh (2018) for separating transparent and reflection scenes.

We have two pairs of adversarial convolutional networks called generator and discriminator. They are denoted by \((G_1, D_1)\) for H\(\alpha\) map and by \((G_2, D_2)\) for [O\(\text{iii}\)] map. The generator tries to reconstruct the true map \(X_{\text{true}}\) from the observed map \(X_{\text{obs}}\), whereas the discriminator tries to distinguish the true map \(X_{\text{true}}\) and the fake map \(G(X_{\text{obs}})\). In other words, for an input \((X_{\text{obs}}, X)\) with \(X\) being either \(X_{\text{true}}\) or \(G(X_{\text{obs}})\), the discriminator returns a probability that \(X\) is \(X_{\text{true}}\). The generator consists of 8 convolution layers followed by 8 de-convolution layers. The discriminator consists of 4 convolution layers. Two generators \(G_1\) and \(G_2\) share the first 8 convolution layers. The kernel size of the convolutions is \(5 \times 5\). In each layer, batch normalization\(^2\), dropout, and skip connection are performed (see Isola et al. 2016 for more details). We normalize the input images by \(1 \times 10^{-7}\) erg s\(^{-1}\) cm\(^{-2}\) sr\(^{-1}\) before given to the network.

During the training phase, the performance of the generators and the discriminators are evaluated by a linear combination of the cross-entropy losses and the mean \(L_1\) norms:

\[
\mathcal{L} = \sum_{i=1,2} \left[ \mathcal{L}_{\text{GAN}}(G_i, D_i) + \lambda_i \mathcal{L}_{L_1}(G_i) \right] + \lambda_{\text{tot}} \mathcal{L}_{L_1,\text{tot}}(G_1, G_2),
\]

where

\[
\mathcal{L}_{\text{GAN}}(G_i, D_i) = \log D_i(X_{\text{obs}}, X_{\text{true},i}) + \log[1 - D_i(X_{\text{obs}}, G_i(X_{\text{obs}}))],
\]

\[
\mathcal{L}_{L_1}(G_i) = \frac{1}{N_{\text{pix}}} \sum_{\text{map}} |X_{\text{true},i} - G_i(X_{\text{obs}})|,
\]

\[
\mathcal{L}_{L_1,\text{tot}}(G_1, G_2) = \frac{1}{N_{\text{pix}}} \sum_{\text{map}} |X_{\text{obs}} - G_1(X_{\text{obs}}) - G_2(X_{\text{obs}})|.
\]

\(X_{\text{true},i}(i = 1, 2)\) denote the true maps of H\(\alpha\) and [O\(\text{iii}\)]. In each training set, the generator (discriminator) is updated to decrease (increase) the loss function \(\mathcal{L}\) averaged over a mini batch\(^3\). We adopt \(\lambda_1 = \lambda_2 = \lambda_{\text{tot}} = 50\) and a batch size of 4. The network is trained for 8 epochs. We use the Adam optimizer (Kingma & Ba 2014) with learning rate 0.0002, and decay rate parameters \(\beta_1 = 0.5\) and \(\beta_2 = 0.999\).

3 RESULTS

3.1 Intensity reconstruction

We use 1000 validation data to study the performance of our networks. Fig. 1 shows an example of true and reconstructed maps and their pixel-by-pixel correspondence. In our fiducial case of \(\lambda_{\text{obs}} = 1.5\mu\text{m}\), the contribution from H\(\alpha\) is larger than [O\(\text{iii}\)] (see Table 1), and then outstanding peaks and structures in the observed map mostly originate from the H\(\alpha\) emission at \(z = 1.3\). It is thus remarkable that

\[X_{\text{true},i}(i = 1, 2)\] denote the true maps of H\(\alpha\) and [O\(\text{iii}\)]. In each training set, the generator (discriminator) is updated to decrease (increase) the loss function \(\mathcal{L}\) averaged over a mini batch. We adopt \(\lambda_1 = \lambda_2 = \lambda_{\text{tot}} = 50\) and a batch size of 4. The network is trained for 8 epochs. We use the Adam optimizer (Kingma & Ba 2014) with learning rate 0.0002, and decay rate parameters \(\beta_1 = 0.5\) and \(\beta_2 = 0.999\).

\(^3\) Mini batch is a randomly selected set of training data, \(\{X_{\text{obs},i}, X_{\text{true},i}\}_{i=0}^{n_d}\) where \(n_d\) is batch size. In training phase, the network passes through all the training data without duplication. When we set the number of epochs \(n_e > 1\), this passing through is repeated for \(n_e\) times. For \(n_d\) training data, updates are performed for \(n_d n_e / n_b\) times in total.

Figure 1. An observed map (top-left) is contributed by H\(\alpha\) (top-center) and [O\(\text{iii}\)] (top-right) emission. The reconstructed H\(\alpha\) and [O\(\text{iii}\)] maps and the sum of them are shown in bottom. The shown area is 1.7 deg on a side, and the intensities are in units of erg s\(^{-1}\) cm\(^{-2}\) sr\(^{-1}\). Note the relative difference in intensity for H\(\alpha\) and [O\(\text{iii}\)] (colorbars). Our networks can reconstruct the fainter [O\(\text{iii}\)] component.
not only the Hα distribution but also the weaker [Oiii] intensity is reproduced well. Note that the brightest peak in the reconstructed [Oiii] map in Fig. 1 is not clearly visible in the observed map.

It is important to study if statistical quantities are also reproduced. We first examine the peaks in our intensity maps. We select local maxima with heights greater than 3σ. We find 20101 (28391) and 20724 (31839) peaks in true and reconstructed Hα ([Oiii]) maps over our 1000 validation data sets. Among them, 12531 (11899) peaks are matched correctly. This means that 62% (42%) of the true peaks are reproduced, and 61% (37%) of the reconstructed peaks are true. We note that, the intensity of a peak is not always accurately reproduced. There are underestimated peaks that are not identified as 3σ peaks even though the positions of the peaks are accurately reproduced. If we aim at locating peaks (galaxy groups and clusters) but with less accurate peak-height estimates, the overall completeness/precision would be higher than quoted above.

In regions where one component (e.g., [Oiii] line) dominates, the local structures is also well reconstructed. If our purpose is to reconstruct accurately a part of the entire image, we can use a method adopted by Lee, Yang & Oh (2018), in which the networks also learn easy and difficult parts. Another promising idea is to combine multiple networks. To test this, we use 5 networks that have an identical set of convolutional layers but are trained with different sets of data. Intensity maps reconstructed by the 5 networks are similar, but have some differences. We find that it is generally difficult to reproduce the true intensity in portions where these networks commonly fail. For Hα ([Oiii]) maps, the number of peaks detected by all the 5 networks is 6404 (4393). Among them, 5308 (3167) peaks are true, which means a 83% (73%) confidence level for our peak detection. We note that if we take the average or the median of the reconstructed maps by multiple networks on a pixel-by-pixel basis, dark structures in void regions and small-scale structures tend to be smoothed out.

### 3.2 Statistical information

Summary statistics such as the mean intensity and the power spectrum are primary tools to study the distribution and the properties of the emission line galaxies. These can then be used for galaxy population studies or for cosmological parameter inference. In this section, we examine how well the mean intensity and the power spectrum are reconstructed. We have checked that the reconstructed quantities do not vary significantly between multiple networks, and thus use a single network in the following study.

Fig. 2 shows the correspondence of the true and the reconstructed mean intensities. The true mean intensities are widely distributed because of the cosmological variance of the underlying density field. We see clear correlations between true and reconstructed mean intensities. Fig. 2 shows that the mean Hα ([Oiii]) intensity of each (1.7 deg)² map can be estimated with ≲ 10% (~ 30%) accuracy. We note that the residual uncertainty is comparable to those resulting from the luminosity function estimates by recent galaxy surveys (Sobral et al. 2013; Khostovan et al. 2015). Planned LIM observations would have a much larger observational area. For instance, SPHEREx (Doré et al. 2016) will perform a deep survey over 200 deg² and thus the estimated mean intensities would have a much smaller statistical uncertainty.

We have tested if our networks generate an accurate image (intensity map) if the input observation map is significantly different from the training data. To this end, we input intensity maps with T_line (Eq. 1) differing as much as 20 percent. Some of these samples have mean intensities below or above the range plotted in Fig. 2. We have confirmed that the networks reconstruct the Hα and [Oiii] intensities with accuracy similar to those shown in Fig. 2.

Another important statistic is two-dimensional power spectrum. Fig. 3 shows the variation of the true (shaded regions) and the reconstructed (error bars) power spectra. The top panel of Fig. 3 shows the difference between the true and the reconstructed power spectra normalized by the square root of the variance of the true power spectra, σtrue. Since our training and test data are generated in the same manner, the variance of the training data is also σtrue. For Hα map, the difference is typically less than σtrue at large scales; our network is able to recover the power spectrum of Hα at z = 1.3 from a confused map. At k ≲ 1 arcmin⁻¹, which corresponds to k ≲ 1 cMpc⁻¹ at z = 1.3, the accuracy is ~ 10%. Clearly, our network learns the clustering of galaxies even though we do not explicitly teach that galaxies at different redshifts have different clustering amplitudes.
4 DISCUSSION

We have shown, for the first time, that cGANs can separate desired signals confused in an intensity map. We can also locate intensity peaks where emission line galaxies are clustered at the target redshift. Combining the distribution of peaks and follow-up observations of individual galaxies would allow us to study the environmental dependence of the galaxy formation. A promising approach is to combine our deep learning method with other conventional method such as cross-correlation analysis. From the statistical information such as the power spectrum and the mean intensity of the reconstructed intensity map (galaxy distribution) at a wide range of redshift, we can infer cosmological parameters and can also learn about the evolution of galaxy populations.

In this Letter, we have presented the results from our first attempt, and there is much room for improvement. In order for our method to be applied to real LIM observations, the networks need to be trained with observational noises and other contaminants. We can utilize the three-dimensional information or train the networks with a larger survey area/volume. For cosmology studies, it would be important to train the networks with a variety of astrophysical/cosmological models and parameters to improve robustness, or it may even be necessary to have a set of independent networks that are trained with different cosmological models. We continue exploring the deep learning approach to, for instance, de-noise intensity maps or to extract designated information from a map with more than two components.

ACKNOWLEDGEMENTS

We thank Yasuhiro Imoto for useful discussion. KM is supported by JSPS KAKENHI Grant Number 19J21379. NF’s visit at the University of Tokyo was supported by the Princeton-U-Tokyo strategic partnership grant. MS is supported by JSPS Overseas Research Fellowships. NY acknowledges financial support from JST CREST (JPMHCR1414). A part of our computations in this study is carried out on Cray XC50 at Center for Computational Astrophysics, National Astronomical Observatory of Japan.

REFERENCES

Cheng Y.-T., Chang T.-C., Bock J., Bradford C. M., Cooray A., 2016, ApJ, 832, 165
Cooray A. et al., 2019, arXiv e-prints, arXiv:1903.03144
Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369
Doré O. et al., 2016, arXiv e-prints, arXiv:1606.07039
Ferland G. J. et al., 2017, Rev. Mexicana Astron. Astrofis., 53, 385
Fonseca J., Silva M. B., Santos M. G., Cooray A., 2017, MNRAS, 464, 1948
Gillet N., Mesinger A., Greig B., Liu A., Ucci G., 2019, MNRAS, 484, 282
Gong Y., Silva M., Cooray A., Santos M. G., 2014, ApJ, 785, 72
Hassan S., Andrianomena S., Doughty C., 2019, arXiv e-prints, arXiv:1907.07787
Hassan S., Liu A., Kohn S., La Plante P., 2019, MNRAS, 483, 2524
Isola P., Zhu J., Zhou T., Efros A. A., 2016, CoRR, abs/1611.07004
Khostovan A. A., Sobral D., Mobasher B., Best P. N., Smail I., Stott J. P., Hemmati S., Nayyeri H., 2015, MNRAS, 452, 3948
Kingma D. P., Ba J., 2014, arXiv e-prints, arXiv:1412.6980
Kovetz E. D. et al., 2017, arXiv e-prints, arXiv:1709.09066
Lee D., Yang M.-H., Oh S., 2018, arXiv e-prints, arXiv:1801.04102
Moriwaki K. et al., 2018, MNRAS, 481, L84
Nelson D. et al., 2019, Computational Astrophysics and Cosmology, 6, 2
Pfeffer D. N., Breysse P. C., Stein G., 2019, arXiv e-prints, arXiv:1905.10376
Planck Collaboration VI, 2018, arXiv e-prints, arXiv:1807.06209
Pritchard J. R., Loeb A., 2012, Reports on Progress in Physics, 75, 086901
Shirasaki M., Yoshida N., Ikeda S., 2019, Phys. Rev. D, 100, 043527
Silva B. M., Zaroubi S., Kooistra R., Cooray A., 2018, MNRAS, 475, 1587
Sobral D., Smail I., Best P. N., Geach J. E., Matsuda Y., Stott J. P., Cirasuolo M., Kurk J., 2013, MNRAS, 428, 1128
Visbal E., Loeb A., 2010, J. Cosmology Astropart. Phys., 11, 016
Zamudio-Fernandez J., Olan A., Villaescusa-Navarro F., Bilaloglu S., Derin Cengiz A., He S., Perreault Levasseur L., Ho S., 2019, arXiv e-prints, arXiv:1904.12846