Deep-learning Reconstruction of Three-dimensional Galaxy Distributions with Intensity Mapping Observations

Kana Moriwaki[1] and Naoki Yoshida[1,2,3,4]

[1] Department of Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan; kana.moriwaki@phys.s.u-tokyo.ac.jp
[2] 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
[3] Research Center for the Early Universe, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan
[4] Institute for Physics of Intelligence, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan

Abstract

Line-intensity mapping is emerging as a novel method that can measure the collective intensity fluctuations of atomic/molecular line emission from distant galaxies. Several observational programs with various wavelengths are ongoing and planned, but there remains a critical problem of line confusion; emission lines originating from galaxies at different redshifts are confused at the same observed wavelength. We devise a generative adversarial network that extracts designated emission-line signals from noisy three-dimensional data. Our novel network architecture allows two input data, in which the same underlying large-scale structure is traced by two emission lines of H α and [OIII], so that the network learns the relative contributions at each wavelength and is trained to decompose the respective signals. After being trained with a large number of realistic mock catalogs, the network is able to reconstruct the three-dimensional distribution of emission-line galaxies at $z = 1.3-2.4$. Bright galaxies are identified with a precision of 84%, and the cross correlation coefficients between the true and reconstructed intensity maps are as high as 0.8. Our deep-learning method can be readily applied to data from planned spaceborne and ground-based experiments.

1. Introduction

The large-scale distribution of galaxies carries rich information on the structure and the evolution of the universe and on how galaxies are formed from early days to the present day. Line-intensity mapping (LIM) is aimed at measuring large-scale intensity fluctuations of line emissions from galaxies and intergalactic medium. Complementary to traditional galaxy surveys, LIM covers a broad spectral range and detects signals from essentially *all* emission sources residing in a large cosmological volume (Kovetz et al. 2017). It is thus possible to make a structural "map" of the universe by a single observation. There have already been a few successful experiments that detect the hydrogen 21 cm line (Chang et al. 2010; Ali et al. 2015), and observations targeting other emission lines such as CO/[CII] and Lyα/Hα/[OIII] are also ongoing (e.g., Concerto Collaboration et al. 2020; Keating et al. 2020; Cleary et al. 2021) or planned (e.g., Doré et al. 2014, 2018). LIM can efficiently survey a large observational volume and the data from LIM are well suited, for instance, to studying the formation and evolution of galaxies (Breysse et al. 2016; Keating et al. 2016) as well as geometry and the matter content of the universe (see, e.g., Doré et al. 2014). LIM can also be used to study the reionization by combining with 21 cm observations of the intergalactic medium (Dumitru et al. 2019; Moriwaki et al. 2019).

A key process in the analysis of LIM data is to separate the contributions from different emission lines originating from sources at different redshifts. Let us consider two emission lines with rest-frame wavelengths λ_1 and λ_2. If they are emitted at redshifts z_1 and z_2 that satisfy $\lambda_1(1 + z_1) = \lambda_2(1 + z_2) = \lambda_o$, they are observed at the same wavelength λ_o, appearing as "interlopers" to each other. Cross correlation analyses are proposed to solve this line confusion problem (e.g., Visbal & Loeb 2010) and there are several other statistical methods (e.g., Gong et al. 2014; Cheng et al. 2016). It is technically challenging to isolate the contribution of a particular emission line and to infer the intensity distribution, but a successful direct reconstruction of the three-dimensional distribution of the emission sources would enhance the constraining power in cosmological studies as well as studies on the galaxy formation and evolution. If the contamination of interloper lines can be removed, we are able to analyze the large-scale structure accurately (see, e.g., Fonseca et al. 2017) and also constrain the galaxy population by using methods such as the voxel intensity distribution (Breysse et al. 2017).

Customized convolutional neural networks have been developed and applied to separate different emission-line signals and to effectively de-noise a map (Moriwaki et al. 2020, 2021), but such applications are limited to two-dimensional images without spectral information. Cheng et al. (2020) devise a reconstruction method that makes use of spectral analysis. Their algorithm effectively extracts the source galaxies with multiple emission lines brighter than a few times the noise level, but fainter signals still remain difficult to be detected. In this Letter, we propose utilizing the spectral information in an efficient manner so that a "machine" can learn the correlation of multiple emission lines at different wavelengths. It is possible to perform a full three-dimensional reconstruction by using a LIM observation with a broad
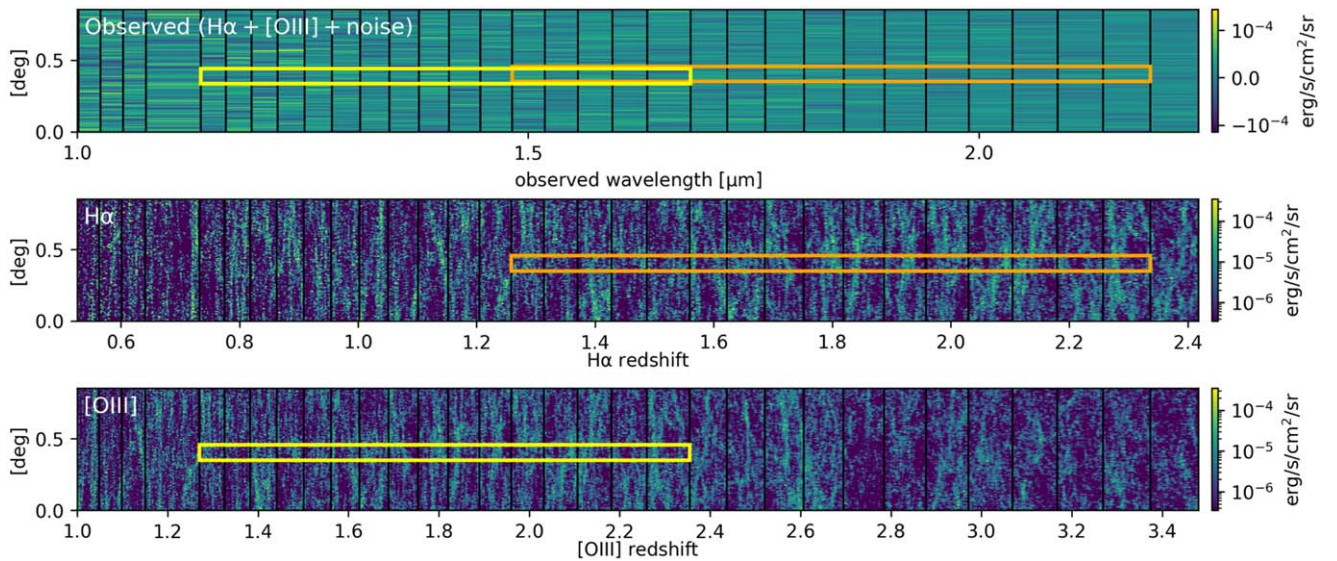
Figure 1. The intensity distribution on a past light cone of a hypothetical observer having a $0°.85$ field of view. The observed intensity (top), H α (middle), and [OIII] (bottom) contributions in units of erg s^{-1} cm^{-2} sr^{-1}. The black lines show the spectral binning of the SPHEREx detector. The yellow and orange boxes indicate the redshift ranges of H α and [OIII] emitters, whose signals originate from galaxies from $z = 1.3$ to 2.4. The data cubes within an angular size of $6'.4$ (the size of the yellow boxes) are used for training. Note that we have adopted larger angular resolution for visibility in this figure than the actual resolution of our training data.

wavelength coverage. This finally enables the reconstruction of the three-dimensional cosmic structure with LIM.

2. Method

We primarily consider NASA's SPHEREx[5] mission, to be launched in 2024, and identify the two brightest emission lines H α 6563 Å and [OIII] 5007 Å as our target to be detected by SPHEREx. We do not consider the other interlopers such as [OII] and Hβ. While the other lines' intensities are likely to be subdominant, they can also carry additional information in the spectral domain. Our method can easily be adjusted to deal with more than two emission lines, although the time needed for training may increase.

2.1. Training Data

To generate mock observation catalogs for training and test, we use a publicly available code PINOCCHIO (Monaco et al. 2013) that populates a large cosmological volume with dark matter halos.[6] We configure past light cones of a hypothetical observer by arranging several simulation outputs to fill the volume (Figure 1). We set the simulation box size to $690h^{-1}$ comoving Mpc and the aperture of the light cone to $1°.5$. The minimum halo mass considered is $2 \times 10^{11}h^{-1}M_\odot$. We have confirmed that the presence of smaller haloes does not affect the total intensity significantly or the intensity distribution. We carefully choose the line-of-sight direction of the light cone so that any galaxy does not appear more than once in the redshift range of our interest.

To assign line luminosities to the galaxies (haloes), we use halo mass-to-line luminosity relations computed in our previous study (Moriwaki et al. 2020) based on the results of cosmological hydrodynamics simulation IllustrisTNG (Nelson et al. 2019). We assign the luminosities by assuming that the line luminosities of haloes in a halo mass bin M_i follow an

asymmetric normal distribution with different variances on the larger and smaller side than the most frequent luminosity value L_i. This assigning process produces similar scatter in the halo mass-to-line luminosity relations as that of IllustrisTNG. Both the H α and [OIII] line luminosities are approximately proportional to the star formation rate, but the derived H α /[OIII] ratio varies over a factor of 10 because the [OIII] luminosity depends also on the properties of the interstellar medium such as metallicity and ionization parameter. We find that the line ratios of our catalog haloes are also scattered in a similar way as that computed with IllustrisTNG.

The middle and bottom panels of Figure 1 show the intensity distributions of H α and [OIII] on a past light cone. We adopt the spatial and spectral resolutions and the noise levels of SPHEREx.[7] The angular resolution is 0.1 arcmin and the spectral resolution is approximately constant ($R \sim 40$) over the wavelength range of our interest.[8] Note that the corresponding physical length to the angular resolution is much smaller than that of the spectral resolution. At $z = 1.5$, for instance, $0'.1$ corresponds to 52 kpc, while $R \sim 40$ corresponds to 47.2 Mpc. We add Gaussian noise to make realistic mock catalogs. The noise level is about two orders of magnitude larger than the mean intensities of line emissions (top panel of Figure 1), and thus detecting diffuse sources that are distributed over the entire intensity field is difficult even with our machine-learning method.

For training, we generate 500 independent light cones with $1°.5$ aperture over $\lambda_{obs} = 1.0\mu m$–$2.5\,\mu m$ using PINOCCHIO. The wavelength range corresponds to 32 spectral bins of SPHEREx as shown in Figure 1. To reduce computational cost, we generate input data with 64×64 angular pixels. This corresponds to a field of view of $6'.4 \times 6'.4$ with an angular resolution of 0.1 arcmin. From each light cone, we randomly extract 100 such small volumes. Then a total of 50,000 mock observational data cubes are generated. As discussed in the
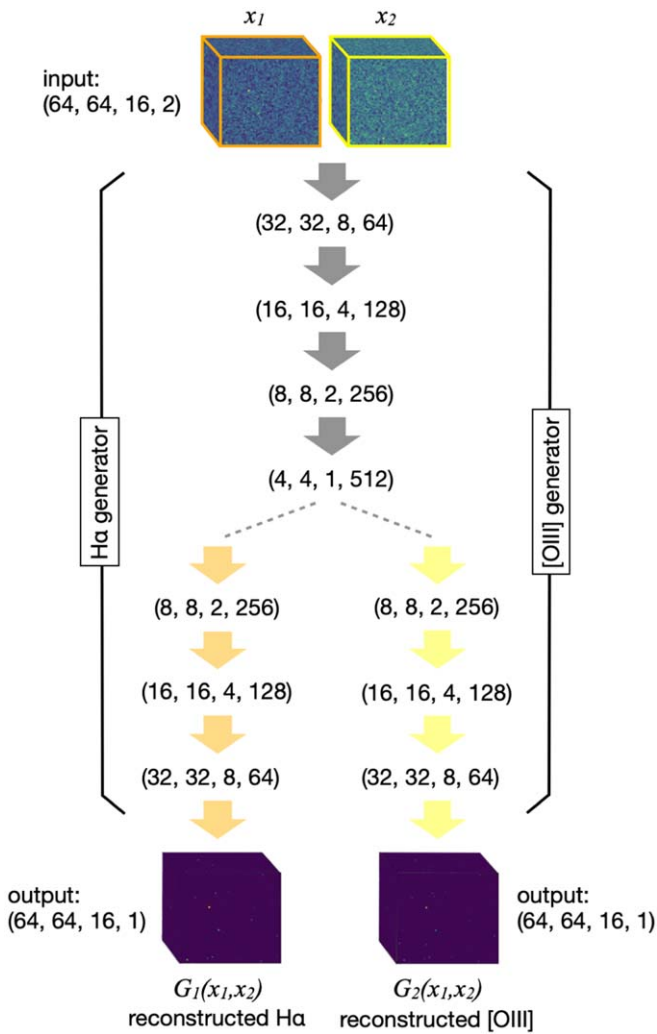
Figure 2. The architecture of the generator that takes two feature maps (data cubes) as an input and consists of four shared convolution layers, followed by four deconvolution layers.

following section, we use two portions of the mock observational data with different wavelength ranges (indicated by orange and yellow boxes in Figure 1) as input to the neural networks.

2.2. Network

We use a conditional generative adversarial network (cGAN; Isola et al. 2016) to perform the three-dimensional reconstruction. In particular, we adopt conditional Wasserstein GAN (WGAN; Arjovsky et al. 2017). WGAN is known to increase training stability and the diversity of generated data (Foster 2019). We have four 3D convolutional neural networks: two generators, G_1 and G_2, that reconstruct H α and [OIII] signals from observed data and two corresponding critics[9], D_1 and D_2, that distinguish true and reconstructed images. Each generator consists of four convolution layers followed by four deconvolution layers (Figure 2), whereas the critic consists of four deconvolution layers. The networks also include skip connections (Isola et al. 2016), dropout (Srivastava et al. 2014), and batch normalization (Ioffe & Szegedy 2015).

---

[9] In WGAN, a network that works as a discriminator in vanilla GANs is called a critic.

The most important information to be learned by the generators is the co-existence of multiple emission lines at different wavelengths. To make it easier for the generators to learn that the two emission lines are always observed with a separation of $\Delta\lambda_{obs} = (\lambda_{H\alpha} - \lambda_{[OIII]}) \times (1 + z)$, we arrange the architecture so that the generators receive a pair of observed data cubes as an input. The cubes are covered by 16 SPHEREx wavelength filters from 1.48μm to 2.19μm and 1.14μm to 1.68μm, which correspond to $1.25 \lesssim z \lesssim 2.4$ of Hα and [OIII] lines, respectively. The input cubes, denoted by x_1 and x_2, are indicated by the orange and yellow boxes in Figure 1. By giving the two data cubes arranged such that the two emission lines from the same source appear at the same pixel, we let the generators learn the consistent co-existence of the two lines.

The critics also receive two data cubes as an input: either a pair of the observed and reconstructed data, $(x_i, G_i(x_1, x_2))$, or a pair of observed and true data, (x_i, y_i), where y_i is the true data cubes of H α $(i = 1)$ or [OIII] $(i = 2)$ that cover the same wavelength range as x_i.

The networks are trained to optimize two loss functions defined by

$$L_i = D_i(x_i, y_i) - D_i(x_i, G_i(x_1, x_2)) + \lambda_i|y_i - G_i(x_1, x_2)|, \quad (1)$$

where the indices $i = 1, 2$ correspond to H α and [OIII], and λ_i is a hyperparameter which we set $\lambda_1 = \lambda_2 = 100$ after some experiments. The objective of the generators (critics) is to decrease (increase) the loss functions. Another important building block of WGAN is the Lipschitz constraint imposed on the critics, which prevents the outputs of the critics from changing abruptly. To enforce the constraint, we adopt the same approach as in the original proposal by Arjovsky et al. (2017) in which they clip the weights of the critic to lie within a small range of $[-0.01, 0.01]$.

We build our network using Tensorflow. We use Adam optimizer (Kingma & Ba 2014) with a learning rate of 0.0002 for training, set the batch size to be 50, and run 50 epochs on a single Nvidia Titan RTX GPU.

3. Result

To measure the performance of our WGAN, we generate an additional set of 1000 light cones that are independent of the training data. We randomly choose an area of $0°.85 \times 0°.85$ from each light cone and divide it into 8×8 cubes with the same size as the training data. The prepared test data are given to the generator of our WGAN. Finally, we reconstruct intensity cubes by combining 8×8 outputs.

In Figure 3, we show an example of the true and reconstructed Hα intensity distributions from $z = 1.3$ to 2.4. The large-scale galaxy distribution is reproduced accurately in 3D despite the large noise level (see Figure 1). Pixel-by-pixel comparison shows remarkably good agreement between the true and reconstructed maps (Figure 4). Our network reconstructs the brightest sources accurately, and thus the underlying large-scale distribution is also well reproduced. Diffuse sources are not well reproduced because of the large observational noise considered in our study. This can also be seen in the probability distribution function PDF (Figure 5). The bright ends are reproduced, but the WGAN appears to have learned that it is optimal to regard faint pixels just as noise-dominated. The vertical lines are the noise level of SPHEREx averaged over 16 wavelength bins of the input data
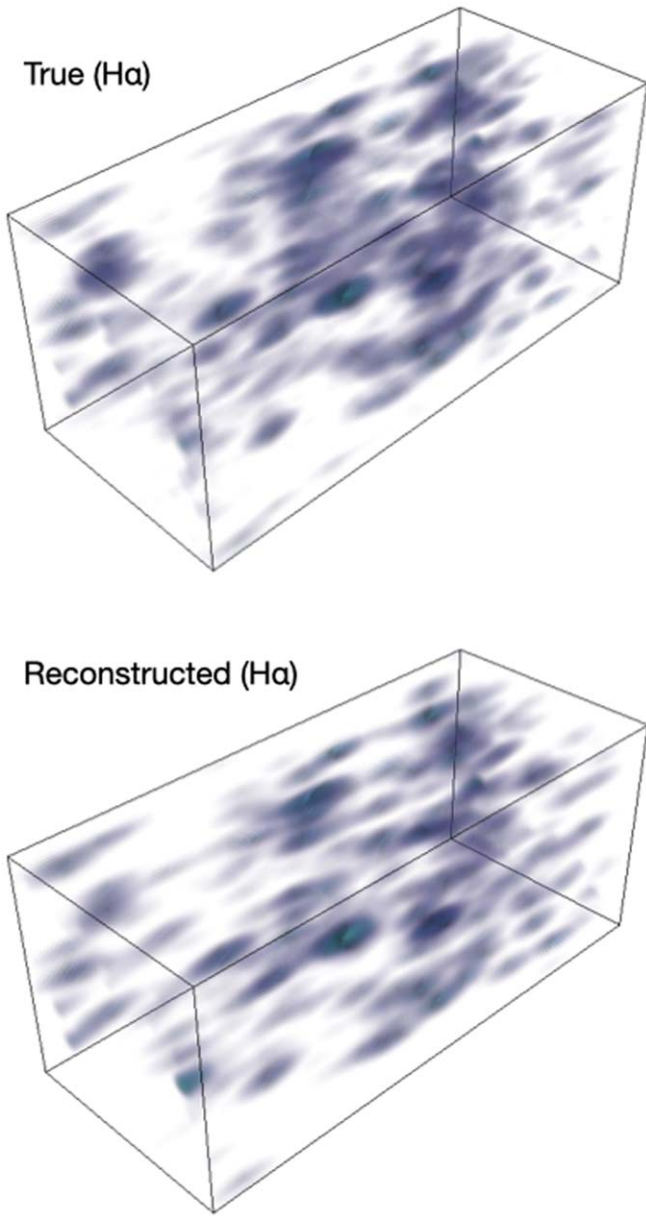
Figure 3. The true (top) and reconstructed (bottom) intensities of Hα line emission from $z = 1.3$ to 2.4. The angular size is $0°.43 \times 0°.43$. The intensities are smoothed for visibility with 6 and 0.5 times the pixel size for angular and spectral domains, respectively.
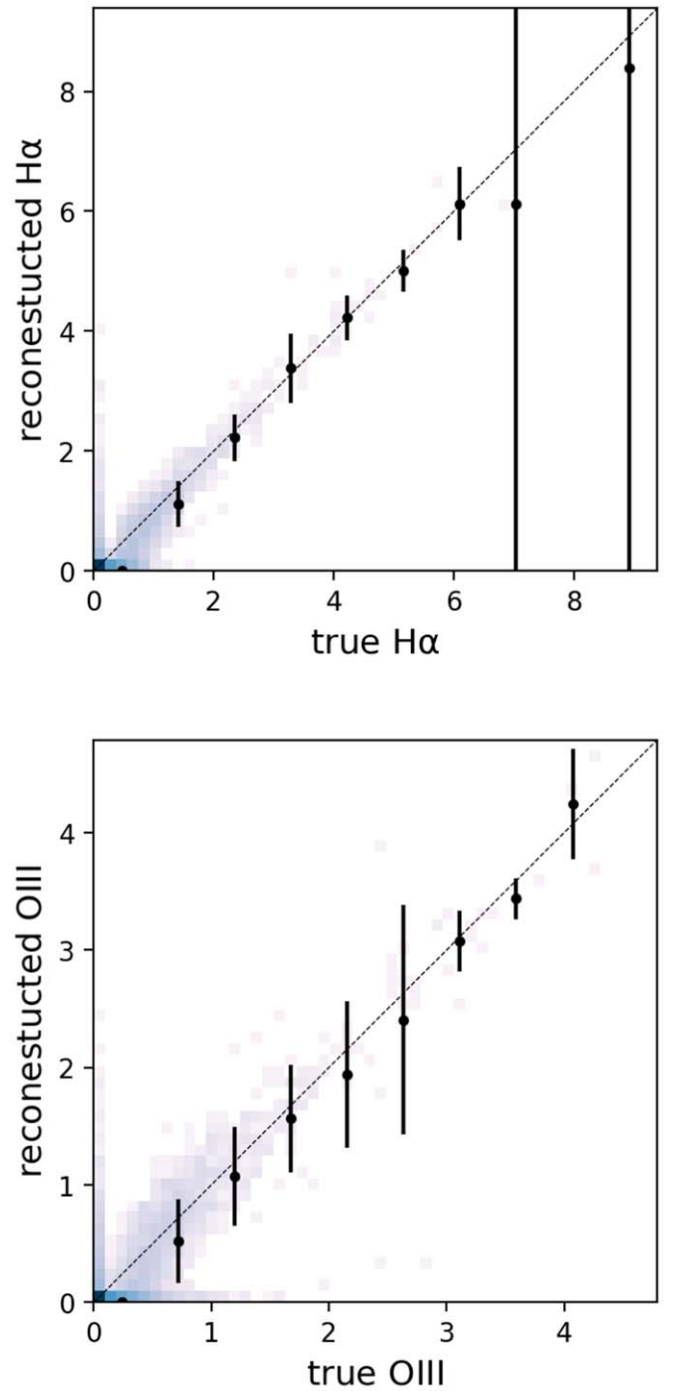

Figure 4. Pixel-by-pixel correspondence between the true and reconstructed intensities of Hα (top) and [OIII] (bottom). Intensities are normalized by 10^{-5} erg s^{-1} cm^{-2} sr^{-1}.

cubes, $\sigma_n = 2.25 \times 10^{-6}$ (upper), 3.06×10^{-6} erg s^{-1} cm^{-2} sr^{-1} (bottom). Figure 5 indicates that the effective limit of our machine-learning reconstruction is a few σ_n. This is similar to the result of Cheng et al. (2020), who show that the CO line signals from similar redshifts are reconstructed down to a few σ_n level. Detecting diffuse "clouds" would be extremely difficult unless the observational noise is significantly reduced in future experiments. It should be noted here that the weaker [OIII] signals are also accurately reconstructed even though the bright end of the observed PDF is dominated by foreground Hα intensities.

We count the numbers of the pixels with intensities larger than 3 σ_n in true (N_{true}) and reconstructed (N_{rec}) maps. We then compute the recall, N_X/N_{true}, and the precision, N_X/N_{rec}, where N_X is the number of pixels that are detected and matched in both the true and reconstructed maps. The recall and the precision are 0.67 and 0.84 for Hα, and the corresponding values for [OIII] are 0.78 and 0.68. We estimate that the typical intensities of [OIII] are roughly half of Hα at the same observed wavelength, and our previous study shows that the detection performance degrades for such weaker lines when only two-dimensional data are used for the machine-learning analysis (Moriwaki et al. 2020). The impressive reproducibility of the [OIII] distribution in the present study can be attributed to the inclusion of the spectral information, as we discuss in the following.
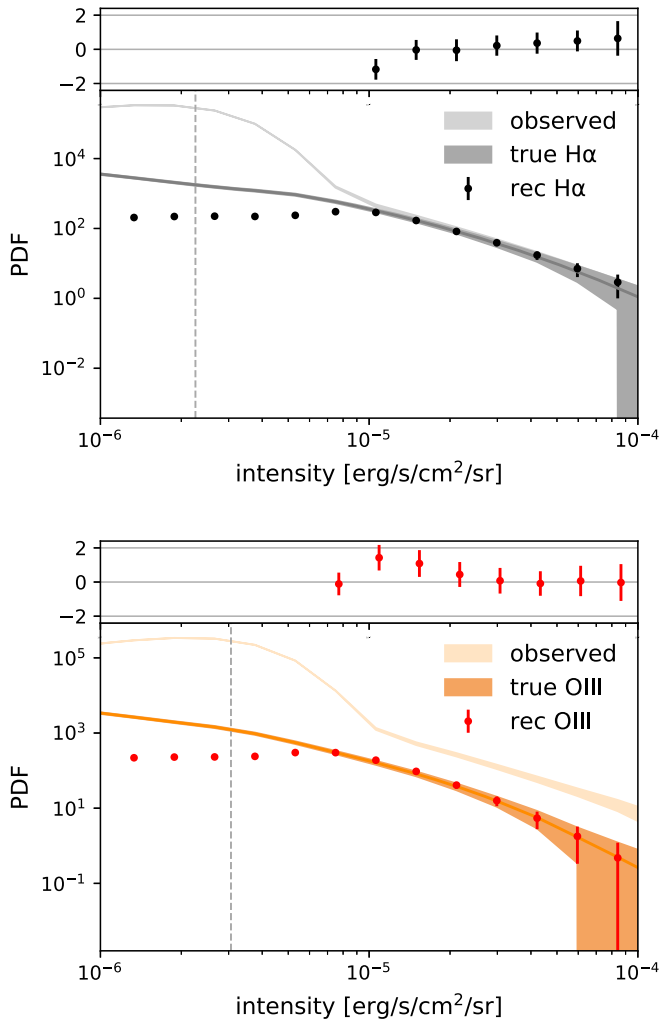
Figure 5. The one-point distribution function (PDF) of the true and reconstructed maps of Hα (left) and [OIII] (right). The 1σ variation of the observed (light shades), true (dark shades), and reconstructed (error bars) PDFs over 1000 test data are shown. The dashed vertical lines are the noise level of SPHEREx, σ_n, averaged over 16 wavelength bins of the input data cubes.

To quantify the reconstruction accuracy of the large-scale distribution, we compute the cross correlation coefficient

$$r(k) = \frac{P_X(k)}{\sqrt{P_{\mathrm{true}}(k) P_{\mathrm{rec}}(k)}}, \qquad (2)$$

where P_X is the cross power spectrum and P_{true} and P_{rec} are the auto-power spectra of the true and reconstructed maps. We find that a high reconstruction performance with $r \sim 0.8$ at $k = 0.3$ arcmin^{-1} for both Hα and [OIII] has been achieved over the wide redshift range. This is consistent with the point-source detection accuracy discussed above.

The high reproducibility of weaker [OIII] signals suggests that the [OIII] generator refers to the Hα intensities that are more easily reconstructed from the two inputs. This is exactly what we expect the machine to learn, and it is important to understand how much it depends on the Hα intensity. To investigate the learning process further, we generate test data with different, uncorrelated realizations for Hα and [OIII] and feed to the [OIII] generator. The result shows that the reconstructed [OIII] map is biased toward the true Hα map, indicating that the [OIII] generator strongly relies on the input

x_1 that includes the Hα signals rather than the input x_2. However, the test case yields the cross correlation coefficients between the reconstructed Hα and [OIII] maps that are smaller than the real case with actual Hα–[OIII] pairs by 0.2. This indicates that the information on the weak [OIII] line in the observed maps is still used to reconstruct accurately [OIII] intensity distributions.

To examine if the spatial clustering information is used along with the spectral information, we perform an additional test. We randomly shuffle the pixels of the test data and get rid of the angular correlation in the signals while preserving the spectral correlation. We then input the shuffled data into our network. The test result shows that the network still achieves high reproducibility; the bright pixels ($> 10^{-5}$ erg s^{-1} cm^{-2} sr^{-1}) are reproduced with similar precision of ~ 0.6–0.8 for both the lines. This implies that our network emphasizes the spectral information (emission-line features) more than the spatial correlation information. We note that we consider a small area of 6.4×6.4 arcmin^2 for the reconstruction in this study. With the finest resolution achievable for our available computational resources, we are able to represent point sources but the particular configuration does not allow incorporating large-scale clustering features. In our previous study (Moriwaki et al. 2021), we showed that the information on the large-scale clustering is more properly used when the training data are generated with a sufficiently large area. Clearly, there is room for improvement in our method. In our future work, we will use data set with larger dimensions so that a machine can learn both the spectral information and the large-scale clustering of galaxies.

4. Summary

We have developed, for the first time, neural networks that extract signals of two emission lines from noisy data obtained in LIM observations. Our 3D WGAN makes use of the information on the co-existence of two emission lines in a given pair of data cubes. It is able to reconstruct the bright sources when trained with a large number of mock observational maps that are closely configured for the SPHEREx experiment. Our method can be extended and applied to LIM observations at any other wavelengths. Once we can extract the individual signals, the reconstructed data can be used for cosmological/astrophysical parameter estimate, cross correlation analysis, and planning follow-up observations.

ORCID iDs

Kana Moriwaki ⓘ https://orcid.org/0000-0003-3349-4070
Naoki Yoshida ⓘ https://orcid.org/0000-0001-7925-238X

References

Ali, Z. S., Parsons, A. R., Zheng, H., et al. 2015, ApJ, 809, 61
Arjovsky, M., Chintala, S., & Bottou, L. 2017, arXiv:1701.07875
Breysse, P. C., Kovetz, E. D., Behroozi, P. S., Dai, L., & Kamionkowski, M. 2017, MNRAS, 467, 2996
Breysse, P. C., Kovetz, E. D., & Kamionkowski, M. 2016, MNRAS, 457, L127
Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Natur, 466, 463

Cheng, Y.-T., Chang, T.-C., Bock, J., Bradford, C. M., & Cooray, A. 2016, ApJ, 832, 165
Cheng, Y.-T., Chang, T.-C., & Bock, J. J. 2020, ApJ, 901, 142
Cleary, K. A., Borowska, J., Breysse, P. C., et al. 2021, arXiv:2111.05927
Concerto Collaboration, Ade, P., Aravena, M., et al. 2020, A&A, 642, A60
Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv:1412.4872
Doré, O., Werner, M. W., Ashby, M. L. N., et al. 2018, arXiv:1805.05489
Dumitru, S., Kulkarni, G., Lagache, G., & Haehnelt, M. G. 2019, MNRAS, 485, 3486
Fonseca, J., Silva, M. B., Santos, M. G., & Cooray, A. 2017, MNRAS, 464, 1948
Foster, D. 2019, Generative Deep Learning (Newton, MA: O'Reilly Media)
Gong, Y., Silva, M., Cooray, A., & Santos, M. G. 2014, ApJ, 785, 72
Ioffe, S., & Szegedy, C. 2015, arXiv:1502.03167
Isola, P., Zhu, J., Zhou, T., & Efros, A. A. 2016, arXiv:1611.07004
Keating, G. K., Marrone, D. P., Bower, G. C., et al. 2016, ApJ, 830, 34

Keating, G. K., Marrone, D. P., Bower, G. C., & Keenan, R. P. 2020, ApJ, 901, 141
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Kovetz, E. D., Viero, M. P., Lidz, A., et al. 2017, arXiv:1709.09066
Monaco, P., Sefusatti, E., Borgani, S., et al. 2013, MNRAS, 433, 2389
Moriwaki, K., Filippova, N., Shirasaki, M., & Yoshida, N. 2020, MNRAS, 496, L54
Moriwaki, K., Shirasaki, M., & Yoshida, N. 2021, ApJL, 906, L1
Moriwaki, K., Yoshida, N., Eide, M. B., & Ciardi, B. 2019, MNRAS, 489, 2471
Nelson, D., Springel, V., Pillepich, A., et al. 2019, ComAC, 6, 2
Planck Collaboration VI, Aghanim, N., & Akrami, M. 2018, A&A, 641, A6
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, J Mach Learn Res, 15, 1929
Visbal, E., & Loeb, A. 2010, JCAP, 11, 016