

## Revealing the Local Cosmic Web by Deep Learning

SUNGWOOK E. HONG,<sup>1</sup> DONGHUI JEONG,<sup>2</sup> HO SEONG HWANG,<sup>3</sup> AND JUHAN KIM<sup>4</sup>

<sup>1</sup>*Natural Science Research Institute, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul, 02504, Republic of Korea*

<sup>2</sup>*Department of Astronomy and Astrophysics, and Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*

<sup>3</sup>*Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro, Yuseong-gu, Daejeon 34055, Republic of Korea*

<sup>4</sup>*Center for Advanced Computation, Korea Institute for Advanced Study, 85 Heogiro, Dongdaemun-gu, Seoul, 02455, Republic of Korea*

(Revised August 6, 2020)

### ABSTRACT

The eighty percent of the matter in the Universe is in the form of dark matter that comprises the skeleton of the large-scale structure called the Cosmic Web. As the Cosmic Web dictates the motion of all matters in galaxies and inter-galactic media through gravity, knowing the distribution of dark matter is essential for studying the large-scale structure. However, as dominated by dark matter and warm-hot inter-galactic media, both of which are hard to trace, the detailed structure of the Cosmic Web is unknown. Here we show that we can reconstruct the Cosmic Web from the galaxy distribution using the convolutional-neural-network based deep-learning algorithm. We find the mapping between the position and velocity of galaxies and the Cosmic Web using the results of the state-of-the-art cosmological galaxy simulations, Illustris-TNG. We confirm the mapping by applying it to the EAGLE simulation. Finally, using the local galaxy sample from Cosmicflows-3, we find the dark-matter map in the local Universe. We anticipate that the local dark-matter map will illuminate the studies of nature of dark matter and the formation and evolution of the Local Group. High-resolution simulations and precise distance measurements to local galaxies will improve the accuracy of the dark-matter map.

**Keywords:** Local Group; dark matter; large-scale structure of universe

### 1. INTRODUCTION

Since Fritz Zwicky inferred its existence from the large velocity dispersion of the Coma cluster (Zwicky 1933) and Vera Rubin confirmed it with the flat rotation curve of galaxies (Rubin & Ford 1970), astronomers have been only strengthening the necessity of the non-baryonic matter providing excess gravity. We call that dark matter. The most substantial pieces of evidence include an excessive mass-to-light ratio in the dwarf galaxies (Aaronsen 1983), the mismatch between the X-ray map (gas distribution) and the weak gravitational lensing map (mass distribution; Clowe et al. 2006), as well as the disparity between the heights of even- and odd-acoustic peaks in the temperature power spectrum of the cosmic microwave background (Larson et al. 2011). Dark matter is also an indispensable component of the con-

dance cosmological model. Accounting for the measured expansion rate of the Universe (Planck Collaboration et al. 2018) requires the matter component whose energy density is over five times larger than that of atoms for which the robust upper limit comes from big-bang nucleosynthesis (Cooke et al. 2014). The observed large-scale distribution of galaxies (Anderson et al. 2014) and the map of weak-gravitational lensing potential (Abbott et al. 2018) also require the dark matter providing the skeleton of the large-scale structure within which atoms collapse to form stars and galaxies (Davis et al. 1985).

With the essential role that dark matter plays in modern astronomy and cosmology, in the past few decades, there have been continuous efforts for searching for the nature of dark-matter *particles* in the particle accelerators (Atlas Collaboration 2019; Vannerom 2019), cosmic rays (Giesen et al. 2015), gamma-rays (Ackermann et al. 2015), and high-energy neutrinos (Aartsen et al. 2018). Beyond the Milky-way halo, there have also been recent studies focusing on the dark-matter signals from

the extra-galactic sources by cross-correlating the high-energy cosmic rays with the distribution of galaxies (For-nasa et al. 2016; Fang et al. 2020) and dark matters traced by weak-gravitational lensing (Tröster et al. 2017; Ammazalorso et al. 2020). All searches for the dark matter particles thus far, however, have not concluded with a firm detection. They have been only narrowing down the possible dark-matter masses and the interaction strengths among dark matter particles as well as between dark matter and atoms (Akerib et al. 2017; Arcadi et al. 2018). For these efforts of searching for the nature of dark matter, the most basic information currently lacking is the distribution of the dark matters, or Cosmic Web, in the local large-scale structure beyond the Milky-way halo. Of course, we have a good reason to believe that dark-matter halos surround each galaxy in the Universe. It is, however, also well known that the galaxies are biased, rather than faithful, tracers of the large-scale structure (Desjacques et al. 2018).

In this article, we shall present a novel method of unveiling the Cosmic Web in the local Universe. As dark matters are dark, of course, we cannot observe them directly from the telescope. The only guaranteed way of searching for the dark matters is the same method for their discovery, through their gravitational influence on visible objects. On the inter-galactic scales, it is dark matter that dominates the gravitational interaction and determines the cosmic velocity flow. We can, therefore, infer the distribution of dark matters by carefully studying the distribution and motion of galaxies. Taking the observed distribution of galaxies and their peculiar velocity flow, in what follows, we shall decipher the dark matter distribution, or Cosmic Web, within local  $\sim 20 \text{ Mpc}/h$ .

When reconstructing the local dark-matter distribution directly from observed galaxy distributions, we face the following challenges. First, the local galaxy distribution at the low Galactic latitudes is hidden behind the intense radiation from the Galactic disk and contaminated by the interstellar gas and dust, which makes it hard to obtain the complete map of the galaxy distribution. Second, even if we had the complete map of galaxies, they are biased tracers of the large-scale structure; that is, the distribution of galaxies does not necessarily reflect the distribution of dark matters. Previous attempts (Gottloeber et al. 2010; Libeskind et al. 2010; Carrick et al. 2015) of making the local dark-matter map, therefore, have relied on the cosmological simulations constrained by the smoothed (with smoothing radius  $\geq 4 \text{ Mpc}/h$ ) density field at high-Galactic latitudes. This observational constraint for the fully evolved galaxy distribution is non-trivial to implement in the simula-

tion because what the simulation needs is the density distribution at the initial time.

Here we overcome the challenges by taking a novel approach based on the deep learning (DL) that allows us to incorporate *all* information in the observed galaxy data: the spatial distribution and the radial peculiar velocity of galaxies. That is, we use the convolutional neural network (CNN)-based DL algorithm to find the mapping between the local dark-matter distribution and the observed positions and the radial peculiar velocities of local galaxies.

The structure of this paper is as follows. In Section 2, we describe the simulation and observational data used for DL training and prediction, respectively. In Section 3, we will briefly describe our DL architecture and the evaluation of our DL model. In Section 4, we will show the reconstructed local dark matter map and its statistical robustness. We will summarize our result in Section 5.

Throughout the paper, we assume a standard  $\Lambda$ CDM cosmology in concordance with the Planck 2018 analysis (Planck Collaboration et al. 2018):  $(\Omega_m^0, \Omega_\Lambda^0, h) = (0.31, 0.69, 0.6777)$ . It is similar to the standard cosmologies adopted in Illustris-TNG and EAGLE simulations:  $(\Omega_m^0, \Omega_\Lambda^0, h) = (0.3089, 0.6911, 0.6774)$  and  $(0.307, 0.693, 0.6777)$ , respectively. (Springel et al. 2018; Schaye et al. 2015)

## 2. DATA

### 2.1. Observational Data: Cosmicflows-3

We use the Cosmicflows-3 galaxy catalogue (Tully et al. 2016, CF3 hereafter), one of the most comprehensive galaxy catalogue that provides distance, radial peculiar velocity, and luminosity of 17,647 galaxies up to 200 Mpc. To produce a fair galaxy sample over the given region, we make the volume-limited sub-sample of the CF3 as follows. First, as the number density of the CF3 galaxies close to the Galactic plane (Galactic latitude  $|b| < 10^\circ$ ) is lower than average, we only use the galaxies at  $|b| > 10^\circ$ . Also, we use the  $B$ -band absolute magnitude ( $M_B$ ) compiled from Lyon Extragalactic Database (LEDA; Paturel et al. 2003) as a proxy of the stellar mass ( $M_*$ ; Wilman & Erwin 2012). We set the  $B$ -band magnitude  $-15$  as the selection criterion, which is sufficient for covering the  $20 \text{ Mpc}/h$ - and  $40 \text{ Mpc}/h$ -cubic volume around the Milky-way galaxy. We have also tested the cases with  $M_B < -16$  and  $-17$  and find no noticeable difference to the predictions from the fiducial choice. Note that we have not used the  $K_S$ -band absolute magnitude, one of the best-known tracer of the stellar mass (Bell et al. 2003), because that information

is missing for about 30% of the galaxies in our sample (Lavaux & Hudson 2011; Huchra et al. 2012).

We calculate the radial peculiar velocity by subtracting the Hubble flow from the velocity in the Galactic standard of rest ( $V_{\text{GSR}}$ ; Kourkchi et al. 2020). Note that we do not use the velocity in the CMB standard of rest ( $V_{\text{CMB}}$ ), to reduce any bias that might be introduced in the conversion. Instead, when generating training and test samples from simulation data, we include the peculiar motion of the Milky-way corresponding galaxy in each simulation. There exists a difference on the Hubble constant between recent CMB observations ( $H_0 = 67.77 \text{ km/s/Mpc}$ ; Planck Collaboration et al. 2018) and the best-fit from the CF3 ( $H_0 = 75 \text{ km/s/Mpc}$ ; Tully et al. 2016). In this study, we have tested both values and find that the effect from the different Hubble constants stays within the uncertainty of the dark-matter map.

## 2.2. Simulation Data: Illustris-TNG & EAGLE

We use the TNG100-1, a simulation with  $V = (75 \text{ Mpc}/h)^3$  and  $1820^3$  dark-matter and gas particles from the Illustris-TNG simulation suite (Springel et al. 2018), as our high-resolution simulation data (TNG100 hereafter). To mimic the observation from the Milky-way galaxy, we select 988 galaxies with stellar mass

$$4 \times 10^{10} \text{ M}_\odot < M_\star < 10^{11} \text{ M}_\odot \quad (1)$$

(“center galaxies” hereafter), by adopting that the Galactic stellar mass is about  $5.2 \times 10^{10} \text{ M}_\odot$  (Licquia & Newman 2015). Around each center galaxy, we make a sub-cube with  $20 \text{ Mpc}/h$  box-size and calculate the dark-matter density field within the  $64^3$  uniform grid. We also calculate the relative position of galaxies with  $M_B < -15$  (“target galaxies” hereafter), as well as the difference of peculiar velocity between the target galaxy and centre galaxy.

For the low-resolution dark-matter map with  $V = (40 \text{ Mpc}/h)^3$ , we use the TNG300-1 from the Illustris-TNG simulations, whose volume and number of particles are  $V = (205 \text{ Mpc}/h)^3$  and  $2500^3$ , respectively (TNG300 hereafter). Note that the amplitude of the luminosity function of TNG300 is lower than the observation and TNG100, mainly due to the lower spatial resolution of the simulation (Pillepich et al. 2018). Therefore, we additionally apply the “resolution correction” to find the center and target galaxies by using the number density obtained from TNG100 rather than directly using the face values of  $M_\star$  or  $M_B$ . We also use the TNG300-1-Dark, a dark-matter-only counterpart of the TNG300, to test how baryonic physics affects our result. We select the center and target galaxies by finding the mass cut of

dark matter halos with the same number density. The result from the TNG300-1-Dark is similar to or slightly worse than TNG300.

Also, we use the RefL0100N1504, a reference simulation with  $V = (67.77 \text{ Mpc}/h)^3$  and  $1504^3$  dark-matter and gas particles from the EAGLE simulation suite (Schaye et al. 2015; Crain et al. 2015, EAGLE hereafter), to check the generality of our result. For the center galaxies, we use the same selection criterion to TNG100 and find 478 center galaxies. For the target galaxies, however, we do not directly use  $M_B$ . It is because the luminosity function of EAGLE is reliable only for bright galaxies ( $M_B \lesssim -18$ ), since the EAGLE simulations calculate the luminosity only to massive galaxies ( $M_\star \geq 10^{8.5} \text{ M}_\odot$ ; Camps et al. 2018). Instead, similar to TNG300, we use the galaxy number density obtained from TNG100 to find the stellar mass cut of target galaxies.

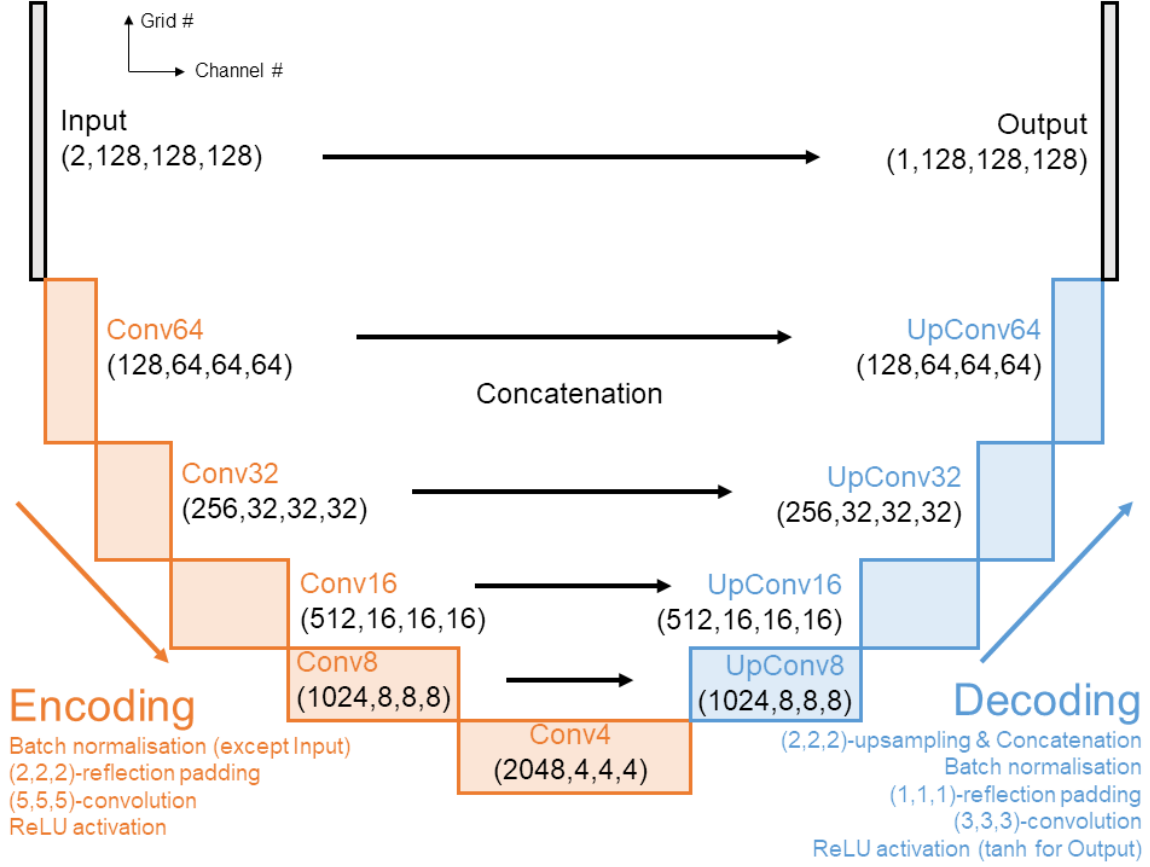
## 3. METHODS

### 3.1. Deep Learning Architecture

We use the convolutional neural network (CNN) similar to the U-Net (Ronneberger et al. 2015) or V-Net (Milletari et al. 2016) to predict the dark-matter density field from the galaxy position and radial peculiar velocity (see Figure 1). Our CNN architecture consists of the following two stages. The encoding stage (Input-Conv#) extracts features by increasing the number of filters and decreasing the layer size. The decoding stage (UpConv#-Output), on the other hand, constructs a new 3-dimensional map by increasing the layer size and decreasing the number of filters. The decoding stage uses the layer with the same size from the encoding stage by concatenation to increase the output resolution.

In practice, the encoding stage consists of multiple Conv# layers, where # denotes the layer size. Conv# consists of (1) 2-pixel reflection padding; (2) 3-dimensional convolution with 5-pixel filters and 2-pixel strides; (3) rectified linear unit (ReLU) activation (Hahnloser et al. 2000; Glorot et al. 2011); and (4) batch normalization (Ioffe & Szegedy 2015). On the other hand, UpConv# layers in the decoding stage contain (1) up-sampling; (2) concatenation with a layer from the encoding stage; (3) batch normalization; (4) 1-pixel reflection padding; (5) 3-dimensional convolution with 3-pixel filters; and (6) ReLU activation. For the output layer, however, we use hyperbolic tangent so that the values stay between  $-1$  and  $+1$ .

We apply two different types of layer size for TNG100 and TNG300. In the case of TNG100, we perform the encoding stage to the  $64^3$ -grid input layer to produce 2,048 channels of the  $2^3$ -grid layer (Conv2) and apply it



**Figure 1.** The convolutional neural network (CNN) architecture used for TNG300. The size (except the number of filter) of each layer for TNG100 is half of TNG300.

to the decoding stage to produce  $64^3$ -grid output layer. In the case of TNG300, on the other hand, we start with the  $128^3$ -grid input layer to 2,048 channels of the  $4^3$ -grid layer, then to the  $128^3$ -grid output layer. We have also tested other CNN architectures with smaller values of the maximum number of channel, and find that the CNN architecture shown in Figure 1 performs the best.

### 3.2. Training

We divide the training and validation samples from TNG100 so that all sub-cubes from the validation sample do not overlap with those from the training sample. As a result, we only use 525 sub-cubes — 432 for training and 93 for validation. For each sub-cube, we make two  $64^3$  uniform grids as a two-channel input layer: the number of target galaxies and the averaged radial peculiar velocity in units of km/s. For the input layer, we apply the same Galactic-latitude mask to the CF3 data (masking out  $|b| < 10^\circ$ ). For the output layer, we normalize the logarithm of dark-matter density to be

$$y = \frac{1}{4.5} \log_{10}(\rho/\rho_0) \quad (2)$$

so that all values in the output layer would be between  $-1$  and  $+1$ .

For data augmentation, we allow swapping the  $(x, y, z)$ -axes of each sub-cube so that the number of samples becomes three times larger. We also make more data augmentation by flipping the axis direction, which in return increases the number of samples eight times. Note that, unlike U-Net or V-Net, we have not split a single cube into multiple smaller cubes for data augmentation, to prevent the change of Galactic-latitude mask and the radial peculiar velocity. In the end, we obtain the total 10,368 and 2,232 samples, respectively, in training and validation sets.

We build our CNN architecture by using the Keras (Chollet et al. 2015) with the Tensorflow backend (Abadi et al. 2015). We use the Adam optimizer (Kingma & Ba 2014) with learning rate  $10^{-3}$  and adopt the mean squared error as the loss function. We use the NVIDIA Tesla V100-PCIE 16GB GPU card for training, and we set mini-batch size as 6 to be able to run within the given GPU memory. For each training, we run 200 epochs by using 1,728 mini-batches per epoch, and it takes about 73 hours for a single run.



We perform a similar training for the TNG300 outcome, except the following differences. First, we have 10,629 training sub-cubes and 1,256 validation sub-cubes, with each sub-cube having  $128^3$ -grids. Unlike TNG100, we do not apply further data augmentation, mainly due to the expensive computational cost from large CNN architecture size. Second, since the dynamic range of dark-matter density of TNG300 is wider than TNG100, we use

$$y = \frac{1}{5} \log_{10}(\rho/\rho_0) \quad (3)$$

for the output layer instead. Third, for the predictions of logarithmic dark-matter density, we use the cyclic learning rate between  $3 \times 10^{-8}$  and  $4 \times 10^{-5}$ , which we find by additional learning rate tests (Smith 2015). Finally, due to the large CNN architecture size, we use four NVIDIA Tesla V100-SXM2 32GB GPU cards with mini-batch size as 8. For each training, we run 400 epochs by using only 157 mini-batches per epoch, and it takes about 90 hours for a single run.

For TNG300, we perform two more alternative trainings. One alternative training uses the logarithm of the total stellar mass-weighted galaxy number density rather than the simple galaxy number density as an input layer, and its result is similar to our fiducial choice. The other one, on the other hand, does not use the radial peculiar velocity, and its result does not show any small-scale structure.

In both cases, the validation loss function reaches its global minimum, and it does not increase more than 1.1 times its global minimum until the last epoch. From each run, we select three models from three different epochs for the following performance test: at the minimum validation loss, at the minimum training loss, and the last epoch.

### 3.3. Performance Test

For both TNG100 and TNG300 training outcomes, we test the resulting models by applying them to the validation samples, so that we can compare the resulting dark-matter density cube with the ground truth. Specifically, we use the following four methods for the performance test—visual comparison, joint probability distribution, histogram, and two-point correlation function (2pCF) (see Figures 2 & 3). We find that 2pCF is the most efficient method to test the performance of each model. For both TNG100 and TNG300, the models at the minimum training loss provide the closest distribution of the 2pCF predictions to their truth, and we adopt them as our optimal models.

After choosing the optimal models, we perform the convergence test between models with different simula-

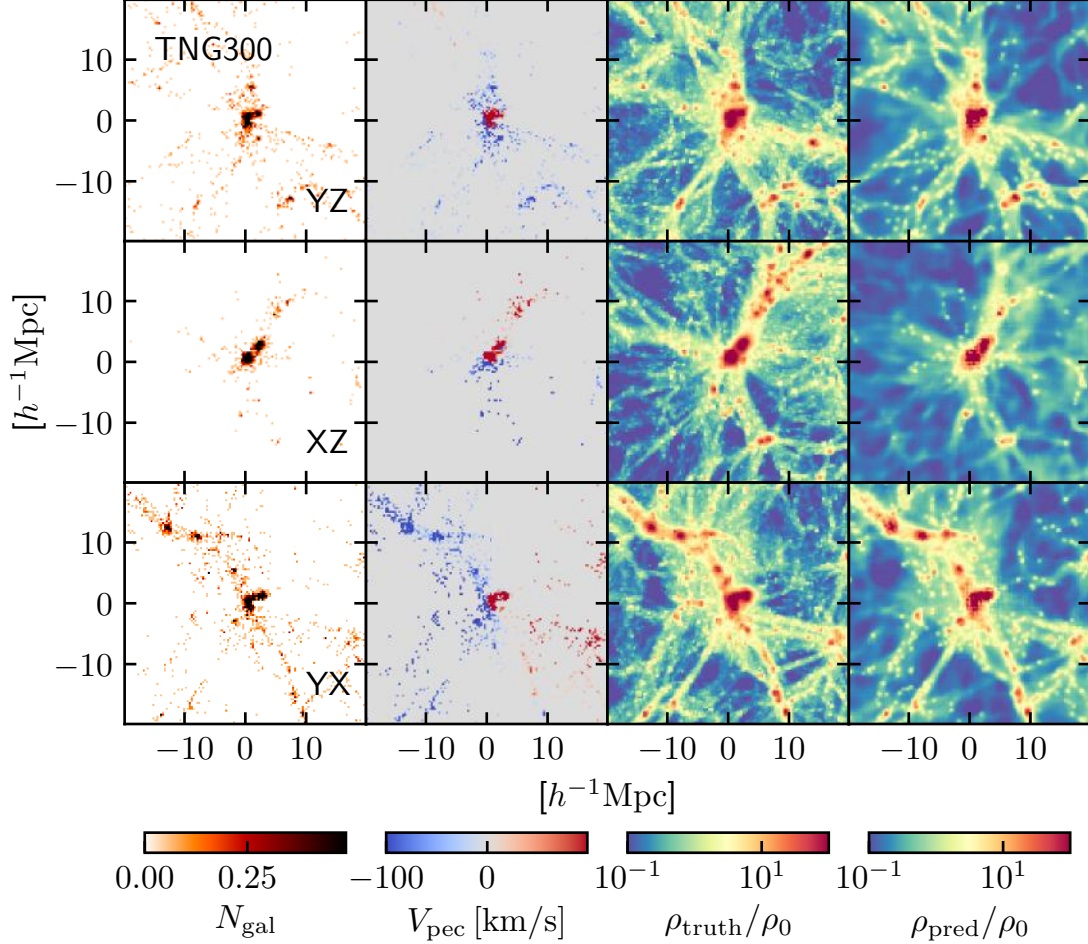
tion resolutions and setups. First, we compare the local dark-matter density field predictions from TNG100 and TNG300 within the radius  $r = 10 \text{ Mpc}/h$ . We find that they show similar distribution up to  $r \sim 4 \text{ Mpc}/h$ , while the dark-matter map from TNG100 shows finer small-scale structures than TNG300 (see Figure 3). Also, we apply the CNN model from TNG100 to the test sample of EAGLE and find that its performance test result is similar to the TNG100 validation sample (see Figure 3). Note that we do not apply the CNN model from TNG300 to EAGLE because the volume of EAGLE is not sufficiently larger than the volume of TNG300 sub-cubes.

Furthermore, to estimate the uncertainties of the dark-matter map, we perform a stress test to our CNN models by incorporating the uncertainties of distance measurement in the CF3. We first use the  $1\sigma$  uncertainty of the distance modulus of each galaxy and generate 1,000 sets of random distance moduli that follow the normal distribution. Then, we re-calculate the radial peculiar velocity by subtracting the Hubble flow corresponding to the random distances from the  $V_{\text{GSR}}$ . Since the distance measurement error exists only along the radial direction, we have generated the two-dimensional column density map of the dark matter that is less affected by the error than the three-dimensional dark-matter density field (see Figure 6). Also, we find that the dark-matter column density map driven from TNG300 shows significantly less deviation than that of TNG100, which suffers from some spurious structure consistently appearing near the Galactic plane.

## 4. RESULTS

### 4.1. Three-dimensional view of the local Cosmic Web

Figure 4 shows a sliced view of the reconstructed Cosmic Web integrated over  $4 \text{ Mpc}/h$  thickness. Each panel shows the Cosmic Web on the plane of the Supergalactic Cartesian coordinates (SGX, SGY, and SGZ), extended to the full cube with the side length  $40 \text{ Mpc}/h$ . Figure 4 clearly shows known local objects that we designated by their common name. The figure also recovers known local large-scale structures. For example, we find a  $10 \text{ Mpc}/h$ -spread along +SGY-direction in the SGZ-SGY (upper left panel) and SGY-SGX (lower-right panel) planes. This structure is known as Local Sheet that connects the Local Group and Virgo cluster and containing M81, NGC5194, Canes II, and Coma I groups (Tully et al. 2008; Courtois et al. 2013). We also find that, around the Local Group, the Local Sheet is connected to the Fornax Wall (Fairall et al. 1994), which is a  $20 \text{ Mpc}/h$ -sized spread along  $(-\text{SGY}, -\text{SGZ})$ -direction, containing Fornax cluster, Eridanus cluster,



**Figure 2.** 3-way projections of a single TNG300 validation sample with  $5 \text{ Mpc}/h$ -thickness. From left to right: galaxy number, radial peculiar velocity, true dark-matter density, and reconstructed dark-matter density.

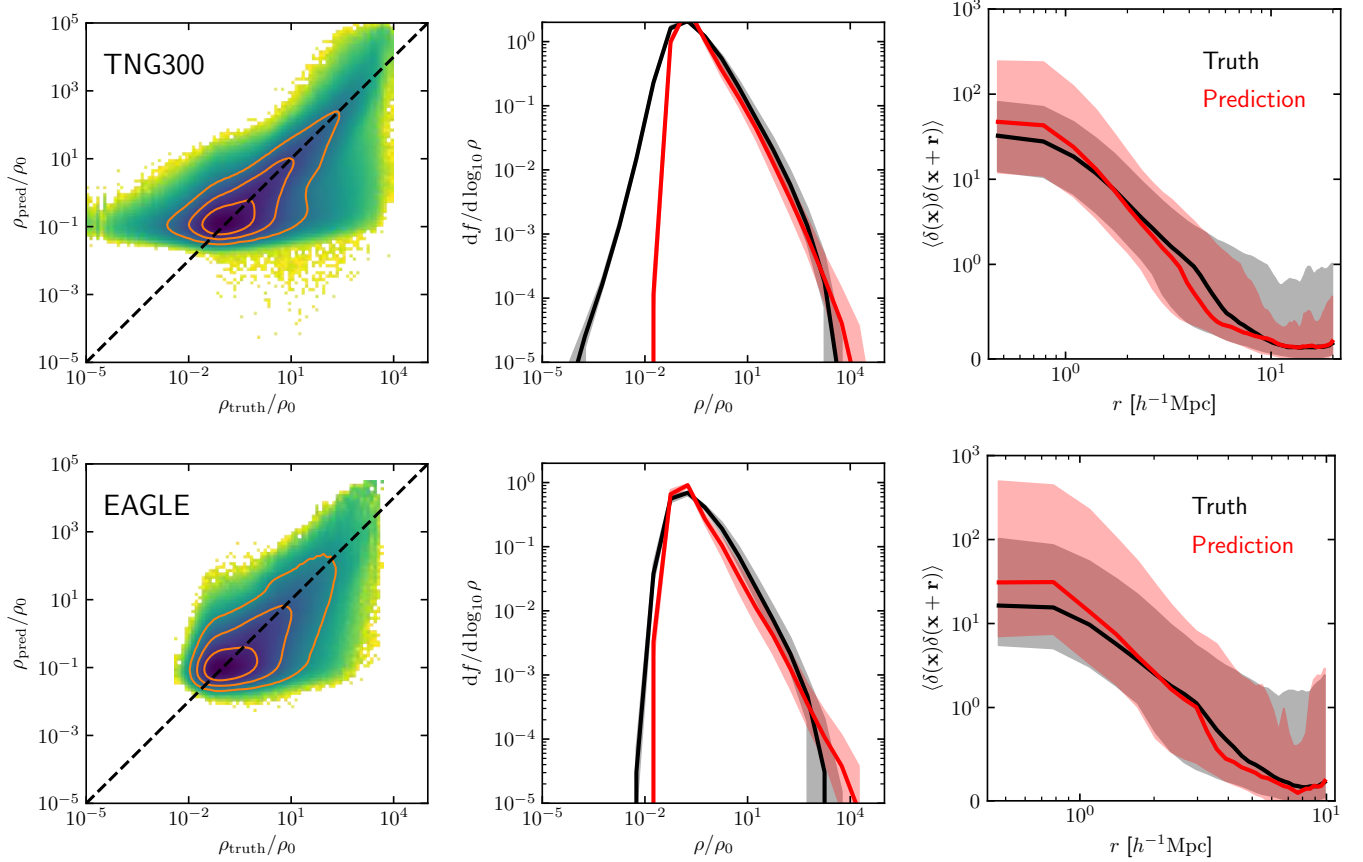
and Dorado group as members (upper-left panel). At the opposite direction to the Fornax Wall on the SGZ-SGY plane, the Local Void (Tully & Fisher 1987) is also apparent (also shown on the SGZ-SGX plane), which might extend beyond the boundary of our local universe sample. In Figure 4, we also present the velocity flow lines derived from the gradient of the reconstructed gravitational potential with arrows and black lines. The velocity flow shows the motion of material from the Local Void to nearby filamentary structures and clusters such as the Local Sheet, Fornax Wall, and Virgo cluster. It should be noted that we cannot reproduce the velocity flow from the Virgo cluster to the Great Attractor (+SGX-direction), because of the limited extension of the volume that we analyze here. However, we would like to emphasize that the recovered dark-matter map provides us detailed density and velocity fields around these known local large-scale structures.

The recovered Cosmic Web also shows a hint of new structures that require further investigation. For exam-

ple, the direction of the Local Sheet is similar to the direction of the so-called vast polar structure (VPOS), which consists of satellite galaxies, globular clusters and stellar streams around the Milky-way galaxy (Pawlowski et al. 2012). As shown in Figure 4, the Local Sheet, being the strongest filamentary structure around the Local Group, is a source of velocity flow; that might cause a connection between the two. Also, there are a couple of small filaments visible in our maps, which could be good targets for systematic examination with deep imaging surveys.

#### 4.2. Sky map of the Local Cosmic Web

The left panels of Figure 5 (labeled as TNG300) show the recovered local dark-matter map (gray map) on the sky (with the Healpix (Zonca et al. 2019; Górski et al. 2005) parameter of  $N_{\text{side}} = 128$ , which roughly corresponds to the angular resolution of 27 arc minutes) along with the location and radial velocities (color-coded dots) of galaxies that we use for the reconstruction. The fig-



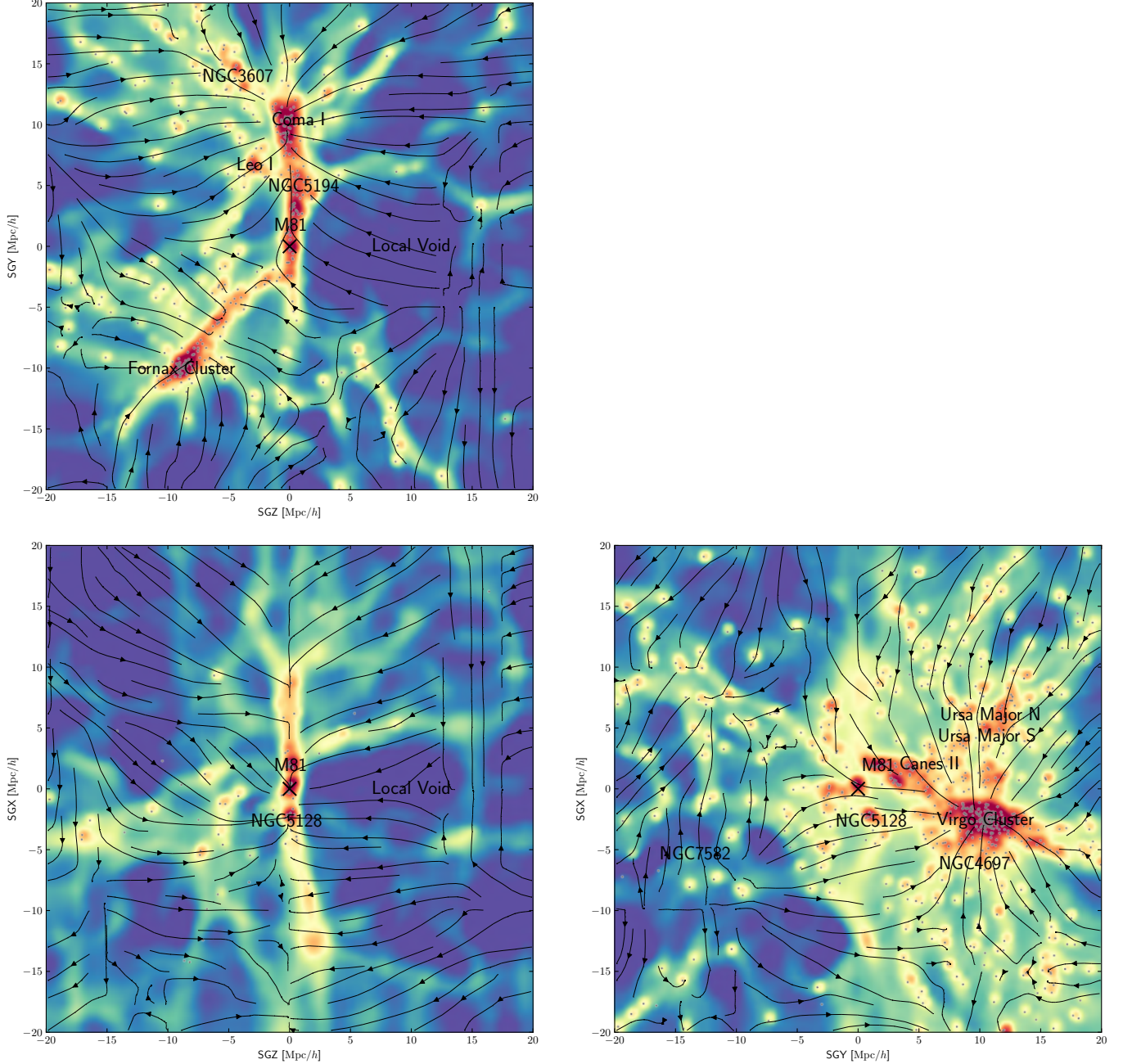
**Figure 3.** Result of the performance tests for the deep learning result using the three-dimensional dark-matter density field of simulations. Top panel: statistical comparison between the ground truth and the predicted dark-matter density from the entire TNG300 validation sample. From left to right: joint probability distribution (colors) with 1, 2, 3- $\sigma$  certainty level contours (lines), median (lines) and 1- $\sigma$  deviation (shades) of histograms, and median (lines) and 1- $\sigma$  deviation (shades) of the two-point correlation functions. Bottom panel: similar to the top panel, but by applying the TNG100 training to the entire EAGLE test sample.

ure also shows the locations of some well-known galaxy groups and clusters. The map in Figure 5 uses the radial distance and radial peculiar velocities as reported in the Cosmicflows-3 catalogue (Tully et al. 2016). We have mitigated the 10 ~ 30% uncertainties of distance measurement in the catalogue by adopting the radial binning of  $\Delta r = 4 \text{ Mpc}/h$ . We further analyze the statistical uncertainties of the recovered dark-matter map by generating 1,000 realizations incorporating the uncertainties of the distance measurement. From the high angular resolution map ( $N_{\text{side}} = 128$ ), we find that the angular covariance function indicates a strong pixel-to-pixel correlation below angular scale  $\theta_0 = 20.7^\circ, 9.71^\circ, 6.53^\circ, 5.04^\circ$ , and  $4.24^\circ$ , respectively, from the nearest ( $r < 4 \text{ Mpc}/h$ ) to the farthest ( $16 \text{ Mpc}/h < r < 20 \text{ Mpc}/h$ ) radial bins. For the statistical analysis, we degrade the angular resolution of each map to  $\theta_0$  and assume that each pixel in the degraded map is statistically independent. Figure 6 shows the mean and the 1- $\sigma$  (68% C.L.)

uncertainty maps of the recovered local dark-matter distribution, adopting the distance uncertainties reported in the Cosmicflows-3 galaxy catalogue. We find that the standard deviation per pixel stays in the range of  $\sigma(\log_{10} \Sigma/\Sigma_0) \simeq 0.1 \sim 0.4$ , regardless of the density contrast, and the signal-to-noise ratio scales almost linearly as the density contrast, reaching up to signal-to-noise ratio 10 for the density peaks. On average, the signal-to-noise ratios for dark-matter distribution per pixel at higher Galactic latitudes ( $|b| > 10^\circ$ ) are 4.25, 3.76, 3.94, 4.19, and 4.52, respectively, from the nearest to the farthest radial bin.

Note that, when estimating the uncertainty map in Figure 6, we have not included the systematic uncertainties in DL mapping itself into the error budget. For example, the galaxy simulations with different resolutions or different sub-grid prescriptions can lead to different DL mapping. First, we check the systematic effect of the resolution by comparing the local dark-matter



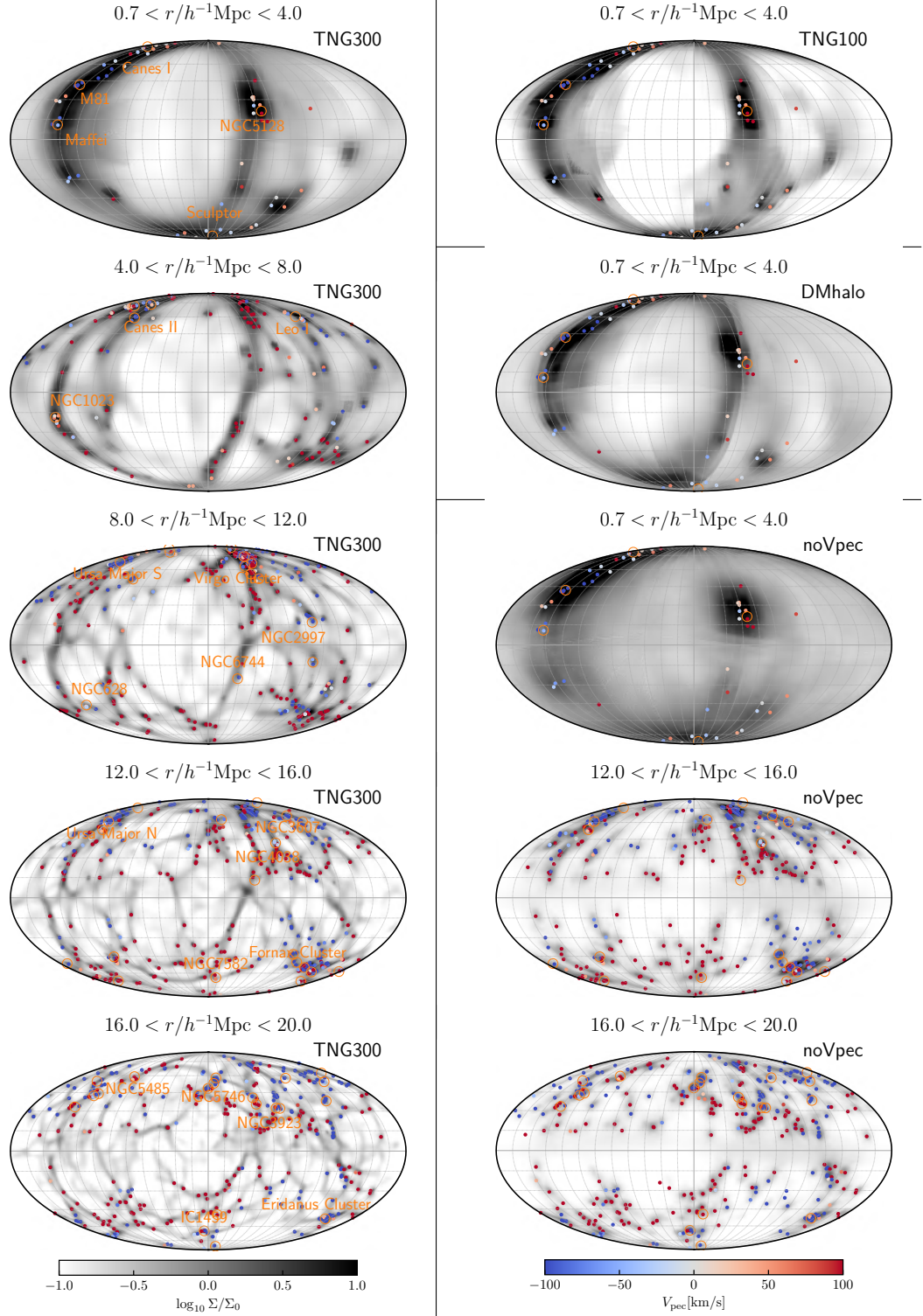


**Figure 4.** Three-dimensional density maps of the local dark matter with  $40 \text{ Mpc}/h$ -boxsize and  $4 \text{ Mpc}/h$ -thickness. ‘X’-mark at the center: Milky-way galaxy. Dots: galaxies with  $M_B < -15$ . Texts: galaxy groups, clusters, and local structures. Arrows: estimated directions of motion derived from the gradient of the reconstructed gravitational potential.

map estimated from TNG100 and TNG300. The top-right panel of Figure 5 shows the  $r < 4 \text{ Mpc}/h$  bin dark-matter map driven from the TNG100 simulation. The high-resolution result (TNG100) systematically underestimate the density contrast by  $2.3\text{-}\sigma$  on average. To estimate the systematic effect from different sub-grid prescriptions, we have repeated the deep-learning procedure by using the dark-matter halo samples (matching number density) from the dark-matter only TNG simu-

lation. The right panels of Figure 6 show the difference between the two dark-matter maps in units of standard deviation at each pixel. Even with this extreme comparison (full hydrodynamic simulation against pure  $N$ -body simulation), we find that systematic effects lead to  $1.7, 1.4, 1.2, 1.1, 1.0\text{-}\sigma$  deviations per pixel from the top (nearest) to the bottom (furthest) maps. We further test the systematic effect due to different Hubble parameters ( $H_0 = 75 \text{ km/s/Mpc}$ ) and find only  $0.15\text{-}\sigma$  devia-





**Figure 5.** Two-dimensional full-sky map of the local dark-matter column density with  $4 \text{ Mpc}/h$  width. Left panels: predictions from TNG300 training, from the nearest to the farthest radial bin. Right panels: comparison predictions from TNG100 training (‘TNG100’), training with dark matter halos from the dark matter-only simulation (‘DMhalo’), and training without using the radial peculiar velocity (‘noVpec’). Small dots: positions and peculiar velocity (color) of known local galaxies. Large dots: galaxy groups and clusters with their names.

tion. Different  $B$ -band magnitude cuts ( $M_B < -16$ , and  $M_B < -17$ ) lead to  $\simeq 1\text{-}\sigma$  deviation. Most importantly, none of the systematic maps shows a significant correlation with the derived cosmic web structure, ensuring the robustness of the derived dark-matter distribution, or the Cosmic Web.

The most striking feature that we have recovered in this study is the filamentary Cosmic Web that is apparent in both Figures 5 and 6. We find that the radial peculiar velocity information is vital to reconstructing the cosmic web, without which the same DL algorithm can not reproduce the Cosmic Web structure at all. For example, the right panels in Figure 5, indicated by ‘noVpec’, show the deep-learning result only using galaxy distributions. Note the absence of the filamentary structure in those maps. We note that the ‘noVpec’ maps simply resemble the smoothed version of the galaxy distribution. The deep-learning algorithm with stellar-mass weighted galaxy distribution, without peculiar velocity information, leads to the similarly poor quality map. Another interesting feature in the map is the dark-matter distribution at lower Galactic latitudes ( $|b| < 10^\circ$ ) where we do not have any input galaxy data. To our surprise, we find that the signal-to-noise ratios per pixel for this region are 4.18, 4.73, 5.31, 5.80, and 6.21, respectively, from the nearest to the farthest radial bin. We, however, anticipate that the theoretical uncertainties for the DL mapping would be most substantial for this region. For example, from the aforementioned studies on systematic uncertainties, we find that on average lower Galactic latitudes ( $|b| < 10^\circ$ ) map suffers about  $0.5\text{-}\sigma$  more systematical shifts than higher Galactic latitudes ( $|b| > 10^\circ$ ) map. This is indicated in the top two panels of Figure 5, as well as the systematic shifts shown in the right panels of Figure 6.

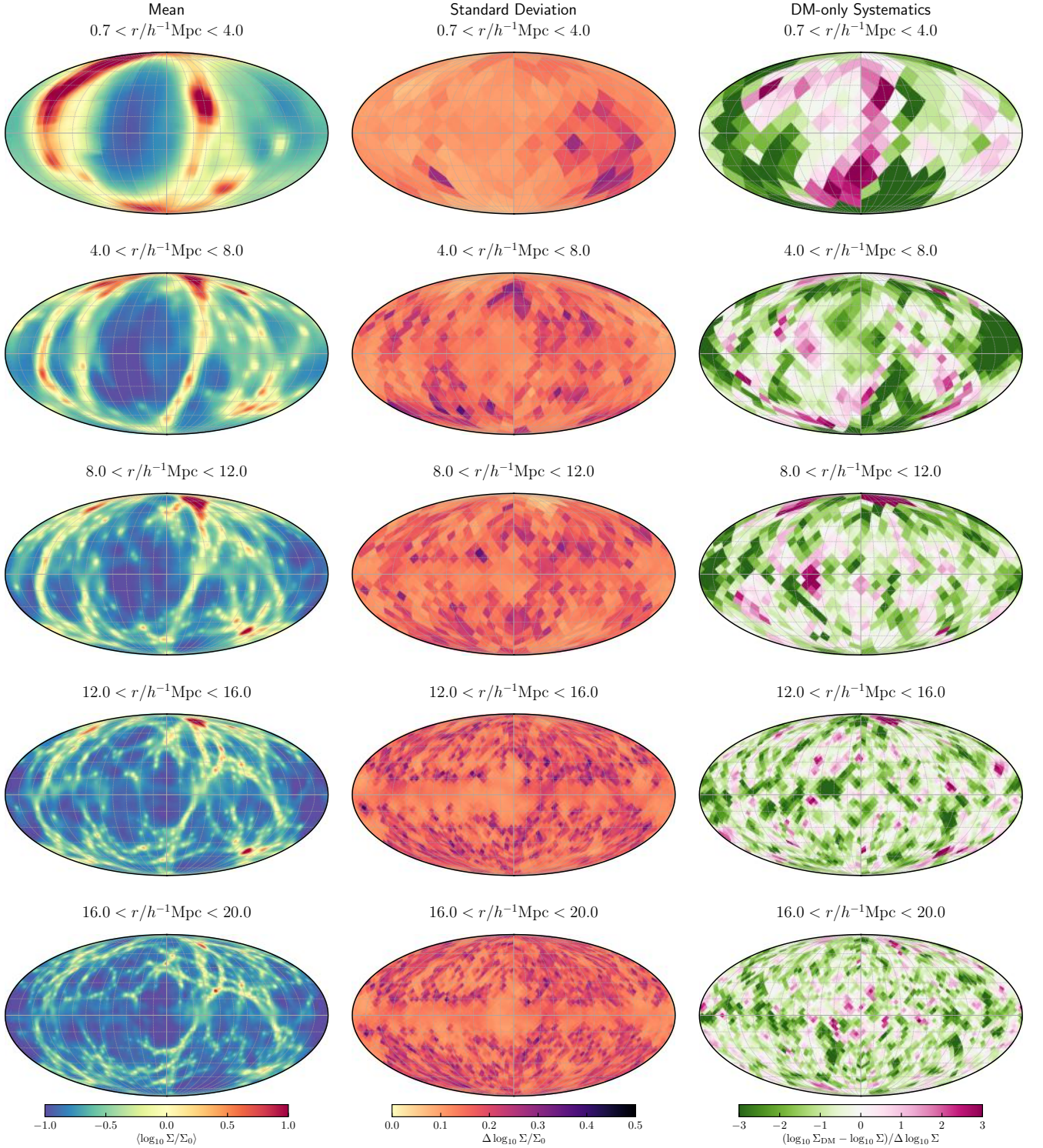
## 5. DISCUSSION

In this paper, we present a novel CNN-based Deep Learning method of reconstructing the local dark-matter distribution map and discover the local Cosmic-Web structure traced by the positions and radial peculiar velocities of Cosmicflow-3 galaxies. We find that including the radial peculiar velocity field is the key to recover the dark matter distribution in the Cosmic Web. Incorporating the observational uncertainties in the galaxy distance measurements, the average detection significance of the dark-matter map exceeds  $4.1\text{-}\sigma$  for each Healpix pixel at higher Galactic latitudes ( $|b| > 10^\circ$ ). The quoted statistical significance, however, does not include the uncertainties in the galaxy-to-dark-matter mapping itself. Here, we have tested that the DL results stay robustly for three different simulations: TNG100,

TNG300, and EAGLE but future studies must quantify the theoretical uncertainties by applying the same method to the large-scale structure simulations with different baryonic prescriptions. The comparison of the DL results between TNG300 and  $N$ -body simulations, however, indicates that the filamentary Cosmic-Web structure may not suffer from the systematic effects.

The main statistical uncertainty in the galaxy data comes from the uncertainty in the distance measurement. As the observed shift in the galaxy spectra constrains the sum of the distance (Hubble flow) and the radial peculiar velocity, the uncertainty affects both the galaxy distribution and the radial peculiar velocity field. Therefore, to obtain a dark-matter map with higher significance, it is necessary to explore the ways to reduce the uncertainties of the current distance estimators such as the Tip of the Red Giant Branch, the Type Ia supernova, and the Fundamental Plane through continuous cross calibration (Tully et al. 2016), and to increase the number of galaxies with measured distances through systematic surveys (e.g., 6dFGS (Springob et al. 2014), James Webb Space Telescope (Gardner et al. 2006)).

We anticipate that the reconstructed three-dimensional dark-matter map and peculiar velocity field will open an entirely new chapter of cosmological study. For example, the dark-matter map can make it possible to run the cosmological galaxy simulations with the *precise* initial condition of the Local Group for studying the past and future of our cosmic neighborhood. It will also allow the in-depth study of the nature of dark matter by cross-correlating the reconstructed dark matter map with the full-sky diffuse emission maps constructed from the radio-to-gamma-ray electromagnetic spectra as well as the full-sky map of gravitational wave binaries. The latter can test the models where black holes in binaries have formed out of dark matter (Shandera et al. 2018).



**Figure 6.** Same as Figure 5, but showing statistical maps. Left panels: mean of the logarithm of dark-matter column-density estimated from 1000 random realizations incorporating the uncertainties in distance estimate to the local galaxies. Middle panels: standard deviation from 1,000 random realizations ( $N_{\text{side}} = 4, 8, 8, 16, 16$  from top to bottom). Right panels: systematic bias from different simulation input for the deep-learning (TNG300 galaxies vs. dark-matter halos).



## ACKNOWLEDGMENTS

Authors acknowledge Christophe Pichon, Changbom Park, Sungryong Hong, Inkyu Park, Dongsu Bak, Graziano Rossi, Yung-Kyun Noh for discussion. The full-sky maps in this paper have been derived using the healpy and HEALPix package. All figures in this paper have been produced with helps of NumPy/SciPy, pandas, astropy, and Matplotlib packages. The list of nearby galaxy groups and clusters are derived from [www.atlasoftheuniverse.com](http://www.atlasoftheuniverse.com). Authors acknowledge the Korea Institute for Advanced Study for providing computing resources (KIAS Center for Advanced Computation Linux Cluster System). Computational data were transferred through a high-speed network provided by the Korea Research Environment Open Network (KREONET). SEH was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (2018R1A6A1A06024977). DJ was supported at Pennsylvania State University by NASA ATP program (80NSSC18K1103). JK was supported by a KIAS Individual Grant (KG039603) via the Center for Advanced Computation at Korea Institute for Advanced Study.

## REFERENCES

- Aaronson, M. 1983, *The Astrophysical Journal Letter*, 266, L11, doi: [10.1086/183969](https://doi.org/10.1086/183969)
- Aartsen, M. G., Ackermann, M., Adams, J., et al. 2018, *Eur. Phys. J. C*, 78, 831, doi: [10.1140/epjc/s10052-018-6273-3](https://doi.org/10.1140/epjc/s10052-018-6273-3)
- Abadi, M., Agarwal, A., Barham, P., et al. 2015, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. <https://www.tensorflow.org/>
- Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, *Physical Review D*, 98, 043526, doi: [10.1103/PhysRevD.98.043526](https://doi.org/10.1103/PhysRevD.98.043526)
- Ackermann, M., Albert, A., Anderson, B., et al. 2015, *Physical Review Letter*, 115, 231301, doi: [10.1103/PhysRevLett.115.231301](https://doi.org/10.1103/PhysRevLett.115.231301)
- Akerib, D. S., Alsum, S., Araújo, H. M., et al. 2017, *Physical Review Letter*, 118, 021303, doi: [10.1103/PhysRevLett.118.021303](https://doi.org/10.1103/PhysRevLett.118.021303)
- Ammazzalorso, S., Gruen, D., Regis, M., et al. 2020, *Physical Review Letter*, 124, 101102, doi: [10.1103/PhysRevLett.124.101102](https://doi.org/10.1103/PhysRevLett.124.101102)
- Anderson, L., Aubourg, É., Bailey, S., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 441, 24, doi: [10.1093/mnras/stu523](https://doi.org/10.1093/mnras/stu523)
- Arcadi, G., Dutra, M., Ghosh, P., et al. 2018, *Eur. Phys. J. C*, 78, 203, doi: [10.1140/epjc/s10052-018-5662-y](https://doi.org/10.1140/epjc/s10052-018-5662-y)
- Atlas Collaboration. 2019, *Journal of High Energy Physics*, 2019, 142, doi: [10.1007/JHEP05\(2019\)142](https://doi.org/10.1007/JHEP05(2019)142)
- Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, *The Astrophysical Journal Supplement Series*, 149, 289, doi: [10.1086/378847](https://doi.org/10.1086/378847)
- Camps, P., Trčka, A., Trayford, J., et al. 2018, *The Astrophysical Journal Supplement Series*, 234, 20, doi: [10.3847/1538-4365/aaa24c](https://doi.org/10.3847/1538-4365/aaa24c)
- Carrick, J., Turnbull, S. J., Lavaux, G., & Hudson, M. J. 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 317, doi: [10.1093/mnras/stv547](https://doi.org/10.1093/mnras/stv547)
- Chollet, F., et al. 2015, *Keras*, <https://keras.io>
- Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, *The Astrophysical Journal Letter*, 648, L109, doi: [10.1086/508162](https://doi.org/10.1086/508162)
- Cooke, R. J., Pettini, M., Jorgenson, R. A., Murphy, M. T., & Steidel, C. C. 2014, *The Astrophysical Journal*, 781, 31, doi: [10.1088/0004-637x/781/1/31](https://doi.org/10.1088/0004-637x/781/1/31)
- Courtois, H. M., Pomarède, D., Tully, R. B., Hoffman, Y., & Courtois, D. 2013, *The Astronomical Journal*, 146, 69, doi: [10.1088/0004-6256/146/3/69](https://doi.org/10.1088/0004-6256/146/3/69)



- Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 1937, doi: [10.1093/mnras/stv725](https://doi.org/10.1093/mnras/stv725)
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *The Astrophysical Journal*, 292, 371, doi: [10.1086/163168](https://doi.org/10.1086/163168)
- Desjacques, V., Jeong, D., & Schmidt, F. 2018, *Physics Reports*, 733, 1, doi: [10.1016/j.physrep.2017.12.002](https://doi.org/10.1016/j.physrep.2017.12.002)
- Fairall, A. P., Paverd, W. R., & Ashley, R. P. 1994, in *Astronomical Society of the Pacific Conference Series*, Vol. 67, *Unveiling Large-Scale Structures Behind the Milky Way*, ed. C. Balkowski & R. C. Kraan-Korteweg, 21
- Fang, K., Banerjee, A., Charles, E., & Omori, Y. 2020, arXiv e-prints, arXiv:2002.06234. <https://arxiv.org/abs/2002.06234>
- Fornasa, M., Cuoco, A., Zavala, J., et al. 2016, *Physical Review D*, 94, 123005, doi: [10.1103/PhysRevD.94.123005](https://doi.org/10.1103/PhysRevD.94.123005)
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, *Space Science Reviews*, 123, 485, doi: [10.1007/s11214-006-8315-7](https://doi.org/10.1007/s11214-006-8315-7)
- Giesen, G., Boudaud, M., Génolini, Y., et al. 2015, *Journal of Cosmology and Astroparticle Physics*, 2015, 023, doi: [10.1088/1475-7516/2015/09/023](https://doi.org/10.1088/1475-7516/2015/09/023)
- Glorot, X., Bordes, A., & Bengio, Y. 2011, in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 315–323
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *The Astrophysical Journal*, 622, 759, doi: [10.1086/427976](https://doi.org/10.1086/427976)
- Gottloeber, S., Hoffman, Y., & Yepes, G. 2010, arXiv e-prints, arXiv:1005.2687. <https://arxiv.org/abs/1005.2687>
- Hahnloser, R. H. R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. 2000, *Nature*, 405, 947, doi: [10.1038/35016072](https://doi.org/10.1038/35016072)
- Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, *The Astrophysical Journal Supplement Series*, 199, 26, doi: [10.1088/0067-0049/199/2/26](https://doi.org/10.1088/0067-0049/199/2/26)
- Ioffe, S., & Szegedy, C. 2015, arXiv e-prints, arXiv:1502.03167. <https://arxiv.org/abs/1502.03167>
- Kingma, D. P., & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980. <https://arxiv.org/abs/1412.6980>
- Kourkchi, E., Courtois, H. M., Graziani, R., et al. 2020, *The Astronomical Journal*, 159, 67, doi: [10.3847/1538-3881/ab620e](https://doi.org/10.3847/1538-3881/ab620e)
- Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, *The Astrophysical Journal Supplement Series*, 192, 16, doi: [10.1088/0067-0049/192/2/16](https://doi.org/10.1088/0067-0049/192/2/16)
- Lavaux, G., & Hudson, M. J. 2011, *Monthly Notices of the Royal Astronomical Society*, 416, 2840, doi: [10.1111/j.1365-2966.2011.19233.x](https://doi.org/10.1111/j.1365-2966.2011.19233.x)
- Libeskind, N. I., Yepes, G., Knebe, A., et al. 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 1889, doi: [10.1111/j.1365-2966.2009.15766.x](https://doi.org/10.1111/j.1365-2966.2009.15766.x)
- Licquia, T. C., & Newman, J. A. 2015, *The Astrophysical Journal*, 806, 96, doi: [10.1088/0004-637X/806/1/96](https://doi.org/10.1088/0004-637X/806/1/96)
- Milletari, F., Navab, N., & Ahmadi, S.-A. 2016, arXiv e-prints, arXiv:1606.04797. <https://arxiv.org/abs/1606.04797>
- Paturel, G., Petit, C., Prugniel, P., et al. 2003, *Astronomy & Astrophysics*, 412, 45, doi: [10.1051/0004-6361:20031411](https://doi.org/10.1051/0004-6361:20031411)
- Pawlowski, M. S., Pflamm-Altenburg, J., & Kroupa, P. 2012, *Monthly Notices of the Royal Astronomical Society*, 423, 1109, doi: [10.1111/j.1365-2966.2012.20937.x](https://doi.org/10.1111/j.1365-2966.2012.20937.x)
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, *Monthly Notices of the Royal Astronomical Society*, 475, 648, doi: [10.1093/mnras/stx3112](https://doi.org/10.1093/mnras/stx3112)
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, arXiv e-prints, arXiv:1807.06209. <https://arxiv.org/abs/1807.06209>
- Ronneberger, O., Fischer, P., & Brox, T. 2015, arXiv e-prints, arXiv:1505.04597. <https://arxiv.org/abs/1505.04597>
- Rubin, V. C., & Ford, W. Kent, J. 1970, *The Astrophysical Journal*, 159, 379, doi: [10.1086/150317](https://doi.org/10.1086/150317)
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 521, doi: [10.1093/mnras/stu2058](https://doi.org/10.1093/mnras/stu2058)
- Shandera, S., Jeong, D., & Grasshorn Gebhardt, H. S. 2018, *Physical Review Letter*, 120, 241102, doi: [10.1103/PhysRevLett.120.241102](https://doi.org/10.1103/PhysRevLett.120.241102)
- Smith, L. N. 2015, arXiv e-prints, arXiv:1506.01186. <https://arxiv.org/abs/1506.01186>
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *Monthly Notices of the Royal Astronomical Society*, 475, 676, doi: [10.1093/mnras/stx3304](https://doi.org/10.1093/mnras/stx3304)
- Springob, C. M., Magoulas, C., Colless, M., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 445, 2677, doi: [10.1093/mnras/stu1743](https://doi.org/10.1093/mnras/stu1743)
- Tröster, T., Camera, S., Fornasa, M., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 467, 2706, doi: [10.1093/mnras/stx365](https://doi.org/10.1093/mnras/stx365)
- Tully, R. B., Courtois, H. M., & Sorce, J. G. 2016, *The Astronomical Journal*, 152, 50, doi: [10.3847/0004-6256/152/2/50](https://doi.org/10.3847/0004-6256/152/2/50)
- Tully, R. B., & Fisher, J. R. 1987, *Atlas of Nearby Galaxies*
- Tully, R. B., Shaya, E. J., Karachentsev, I. D., et al. 2008, *The Astrophysical Journal*, 676, 184, doi: [10.1086/527428](https://doi.org/10.1086/527428)

Vannerom, D. 2019, Proc. Sci., DIS2019, 111,

doi: [10.22323/1.352.0111](https://doi.org/10.22323/1.352.0111)

Wilman, D. J., & Erwin, P. 2012, The Astrophysical  
Journal, 746, 160, doi: [10.1088/0004-637X/746/2/160](https://doi.org/10.1088/0004-637X/746/2/160)

Zonca, A., Singer, L., Lenz, D., et al. 2019, J. Open Sour.  
Soft., 4, 1298, doi: [10.21105/joss.01298](https://doi.org/10.21105/joss.01298)

Zwicky, F. 1933, Helv. Phys. Acta, 6, 110