Nr. 61

VoidNet: Void Galaxy Selection from g-dropout Catalog by Deep Learning

*Yoshihiro Takeda[1], Nobunari Kashikawa[1], Kei Ito[2], Jun Toshikawa[3], Yongming Liang[2], Rikako Ishimoto[1], Takehiro Yoshioka[1], and HSC project 353 members

[1] The University of Tokyo
[2] The Graduate University for Advanced Studies, SOKENDAI
[3] University of Bath
*y.takeda@astron.s.u-tokyo.ac.jp

ABSTRACT

High-z void galaxies, whose evolution has been driven almost completely free from galaxy mergers, are ideal targets to provide valuable insights into the role of environment in galaxy evolution. However, a very wide galaxy survey with spectroscopic redshifts are required to find void regions, and there have been no studies beyond $z > 3$. In this work, we develop a new deep learning method to select $z \sim 4$ void galaxies from the g-dropout catalog produced by the HSC-SSP survey; called **VoidNet**. The **VoidNet** uses the sky distribution of galaxies and their $(g - r)$ colors as a proxy for redshift despite of large uncertainty to characterize the three-dimensional spatial distribution of galaxies. We train the **VoidNet** by using Millennium simulation, and when setting a conservative threshold (recall = 0.1%), the **VoidNet** achieves 90% precision, which is about 20% better than 2D selection in detecting void galaxies. This result shows that deep learning can provide better estimate of the large-scale structure of the universe even when using the photometric data. We are applying the same method to the identification of high-z protoclusters as well as voids.

INTRODUCTION

- Cosmic voids are low-density environments: galaxies in voids are isolated from other galaxies.
- These galaxies experienced almost no galaxy mergers, and their evolution may be different from those in high-density regions, clusters.
- Some void finders (Pan et al. 2012, Sutter et al. 2012, Sánchez et al. 2017, Krolewski et al. 2018) were developed to reveal the void features at relatively low redshift.
- It is important to observe at high redshift because the higher redshift, the more purely the role of environment in galaxy evolution can be studied.
- In this study, we developed a new deep learning method, **VoidNet** (Figure 1), to detect galaxies in voids at $z \sim 4$ from a g-dropout catalog.

DATA

Simulation data (Figure 2, for training and evaluation)

- N-body Simulation (Millennium Simulation: Springel et al. 2005)
  +light-cone model (Henriques et al. 2012)
  +semi-analytic model (Guo et al. 2011)
  →g-dropout catalog($3.25 < z < 4.6$)(Toshikawa et al. 2016)
- Magnitude limit: $23.00 < i < 26.07$
- Redshift range of target galaxies: $3.4 < z < 3.7$

Observation data (Toshikawa et al. 2018)

- g-dropout catalog from HSC-SSP Wide survey S16A
- Region: W-WIDE12H (area: 17deg^2)
- Magnitude limit: $23.0 < i < 25.5$
- g dropout criterion:

$$\begin{cases} 1.0 < g - r \\ -1.0 < r - i < 1.0 \\ 1.5(r - i) < g - r - 0.8 \end{cases}$$

- Target galaxies (10521 galaxies):
  - Mask occupies less than 10% of a circle of radius 10' of the target galaxy
  - $1 < g - r < 1.5$ (To match redshift range with simulation)

METHOD

Underdensity:

- Defined as $\delta = \frac{\rho - \langle\rho\rangle}{\langle\rho\rangle}$, where ρ is the galaxy number density and $\langle\rho\rangle$ is ensemble average of the density.
- $\delta^{(3)}$ denotes the 3D underdensity and $\delta^{(2)}$ denotes the 2D underdensity.

Void galaxy:

- Defined as a galaxy with $\delta^{(3)} < 0$ in a sphere of radius 4 pMpc.

PointNet (Qi et al. 2017)

- It was developed to handle 3D point cloud data and to solve 3D shape classification and segmentation.
- Two main features:
  1. Output is invariable regardless of the order of input
  2. The effect of rotation is eliminated by a predicted affine transformation.
- We call a modified PointNet for void galaxy detection as **VoidNet** (Figure 1).

VoidNet Input

- First input layer: the relative sky coordinates and $(g - r)$ color, which is a rough proxy for redshift, for the 350 galaxies surrounding the target galaxy
- Middle input layer: the absolute color $(g - r)$

VoidNet Output

- Probability of being a void galaxy

Baseline model (for comparison)

- It determines whether a void galaxy or not using only $\delta^{(2)}$ without any information of the LOS distances.

RESULT

Evaluation

- Over the whole range of recall, the **VoidNet** keeps higher precision than the baseline model (figure 3).
- **At a conservative threshold (recall = 0.1%), the VoidNet archives 90% precision, which corresponds to an improvement of about 20% compared to the Baseline model.**

Selection from observation data

- We succeeded to predict **6 galaxies as void galaxies with high probability, $P > 70\%$** (Figure 4).

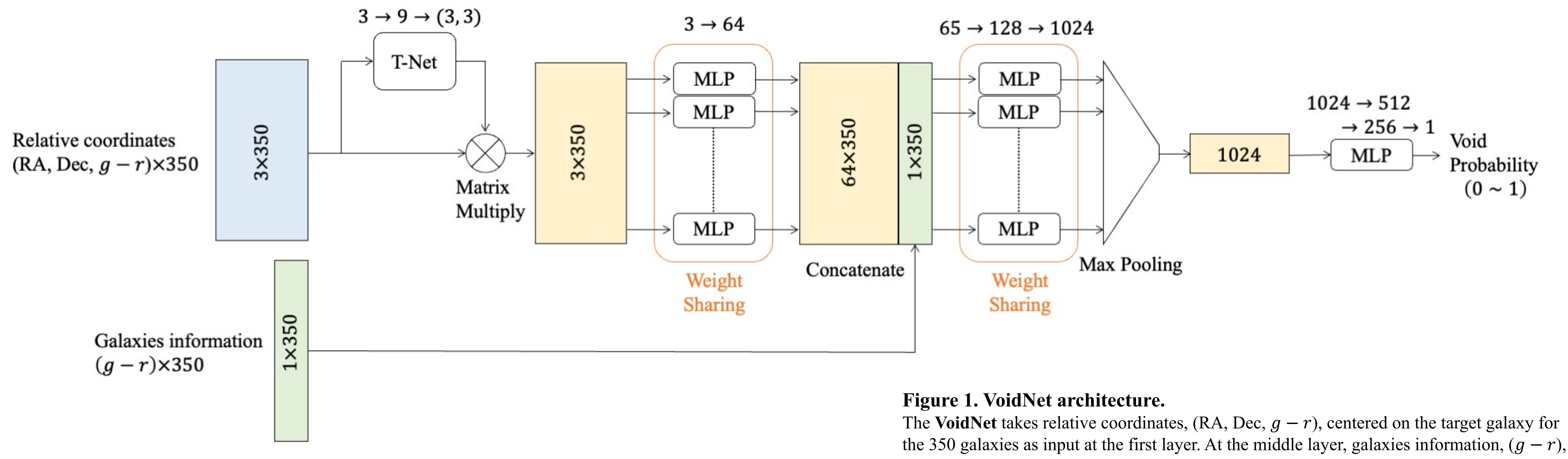

Figure 1. VoidNet architecture.
The **VoidNet** takes relative coordinates, (RA, Dec, $g - r$), centered on the target galaxy for the 350 galaxies as input at the first layer. At the middle layer, galaxies information, $(g - r)$, is inputted. Note that $(g - r)$ on this second input is an absolute value, not a relative one. The output is the probability of the target galaxy being a void galaxy. MLP in the figure denotes Multi-Layer Perceptron and the numbers over each layers are layer channel sizes.

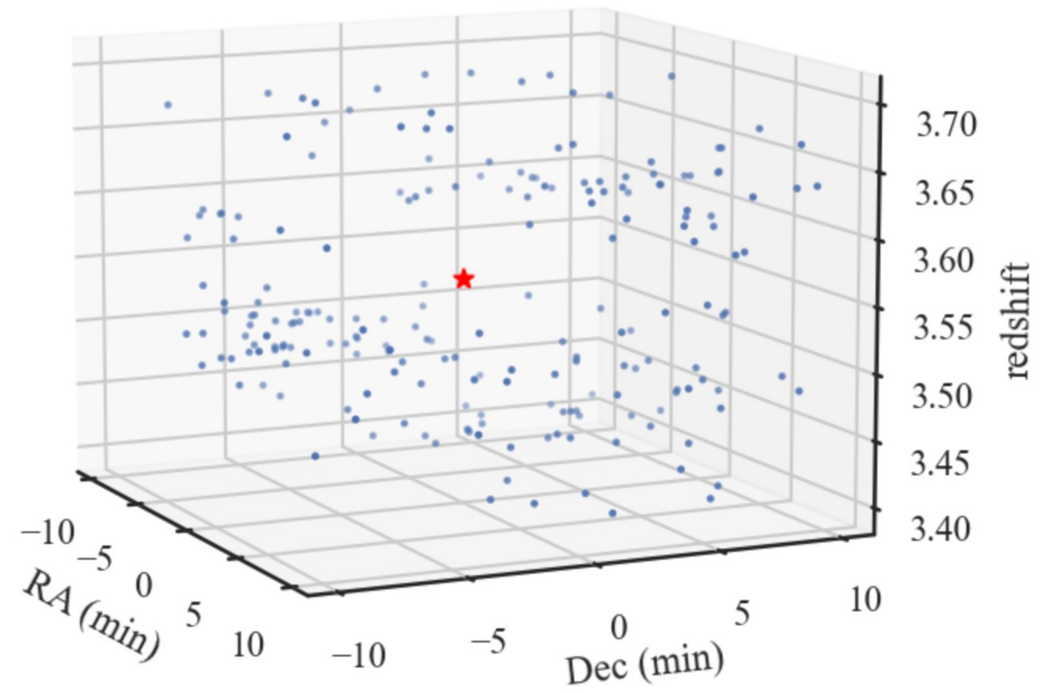

Figure 2. 3D distribution of galaxies surrounding a void galaxy
The red star shows the position of a void galaxy, and the blue points show that of other galaxies in the simulation data. The units of RA and Dec are minutes. The void galaxy is isolated form other galaxies and settles in the large-scale void structure.

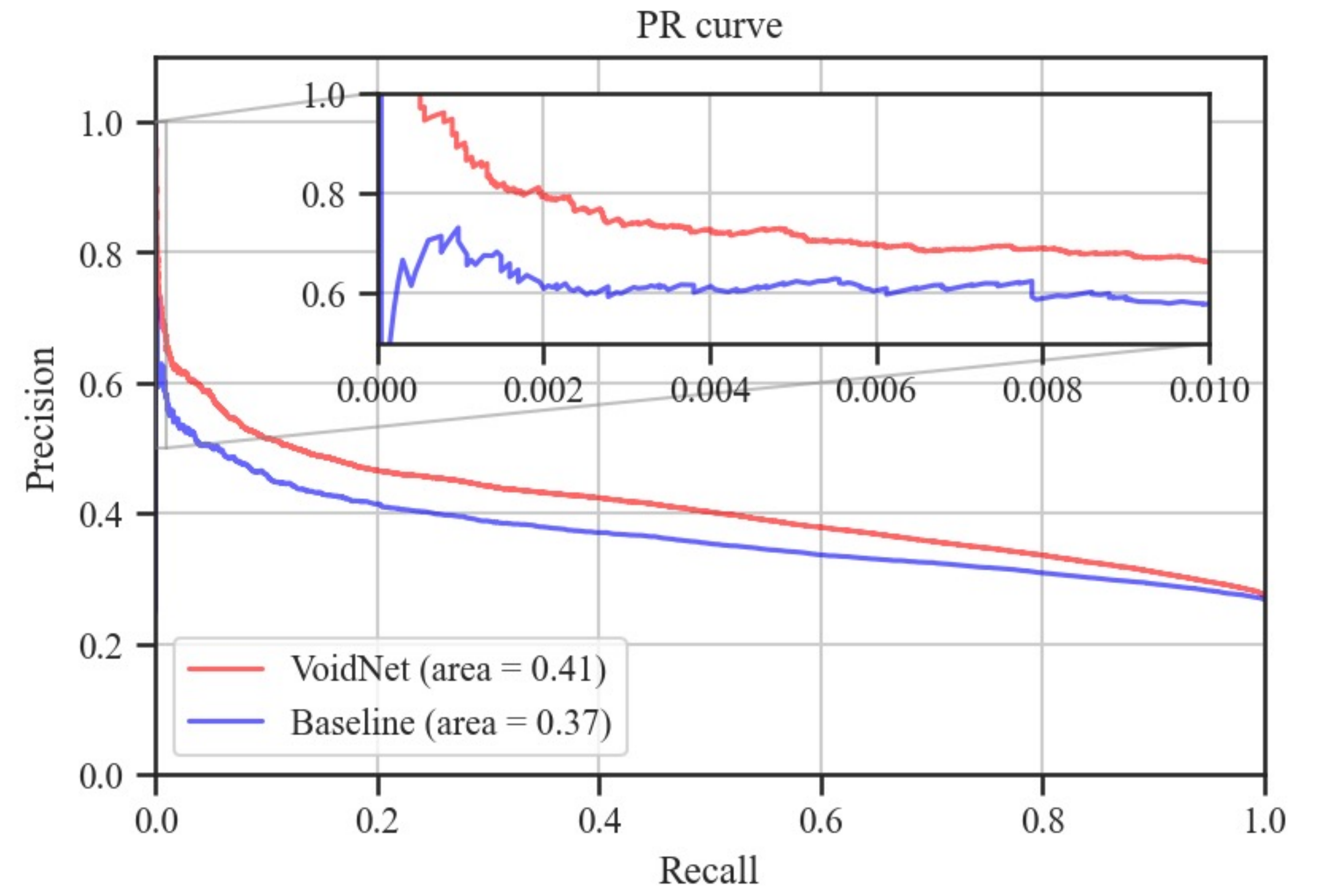

Figure 3. Precision-Recall curve.
The blue line shows a curve of the baseline model, and the red line shows a **VoidNet** curve. The **VoidNet** is constantly better than the baseline model over the whole range. Especially, at an ultimate low recall, the **VoidNet** overwhelms the baseline model by about 20%.

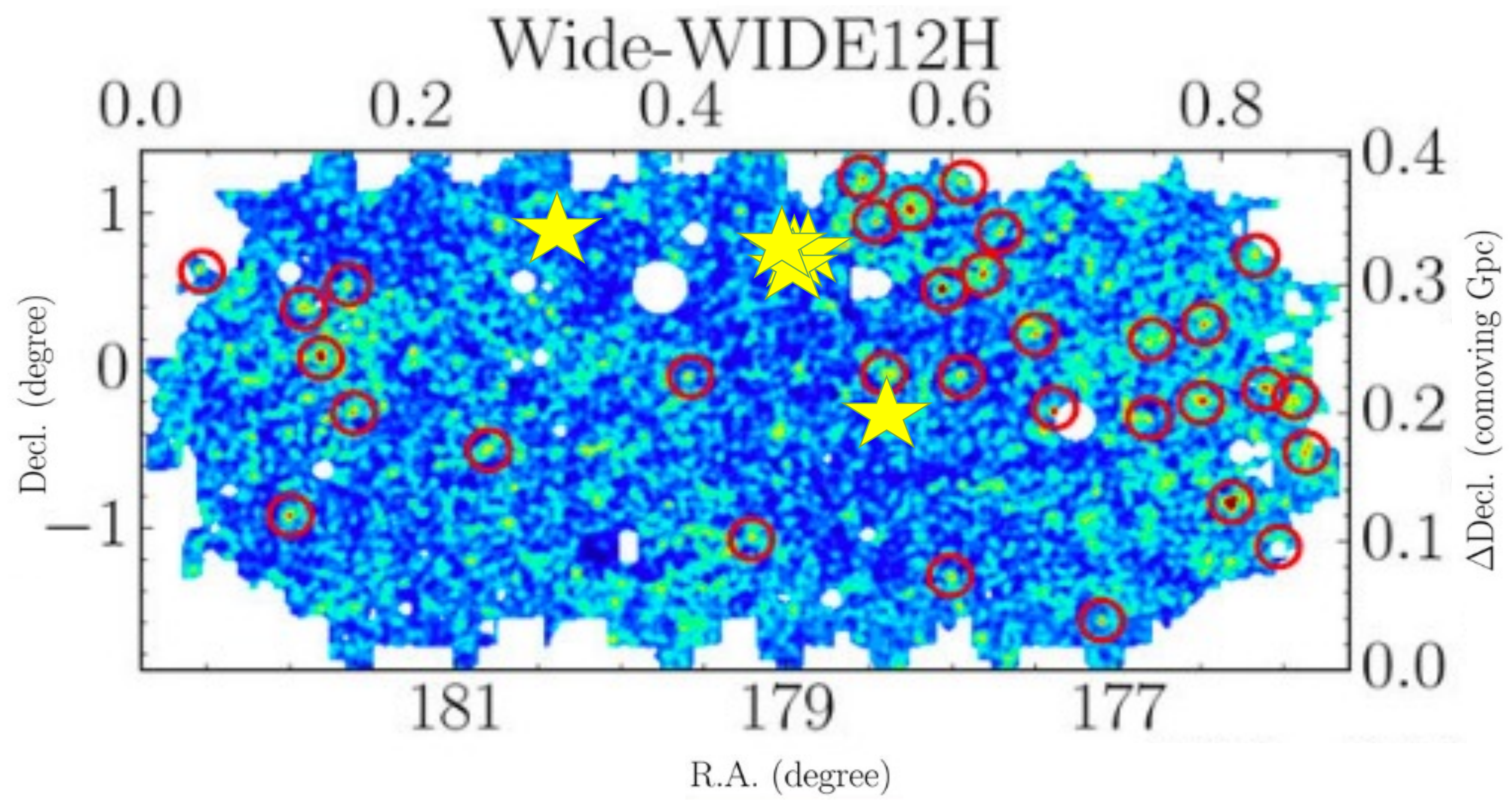

Figure 4. Distribution of 6 void galaxy candidates
The yellow star shows the position of 6 void galaxy candidates and the red circle shows the positions of the protocluster candidates (Toshikawa et al. 2018). The background map (cited from Toshikawa et al. 2018 Fig.2) is underdensity contours, which lower-density regions are indicated by bluer colors.

CONCLUSION

We showed **VoidNet** is effective to select void galaxy from the g-dropout catalog, which has large uncertainties in the LOS, and succeeded to select void galaxy candidates. We will improve more precision and make a catalog of void galaxies in $z \sim 4$ to investigate with statistical analysis. In the future, we will further improve the accuracy of the prediction, create a catalog of void galaxies at z~4 to perform statistical analysis. Follow-up spectroscopic observation will confirm the existence and clarify the properties of void galaxies.