DeepHalo - Finding Dark Matter Halo Substructure Using Deep Learning

Jyotirmaya Shivottam
National Institute of Science Education and Research
Khurdha, Odisha, India
jyotirmaya.shivottam@niser.ac.in

Subhankar Mishra
National Institute of Science Education and Research
Khurdha, Odisha, India
smishra@niser.ac.in

ABSTRACT

This work focuses on the application of point-cloud segmentation techniques to the problem of substructure-finding in dark matter halos from gravity-only N-body simulations. To find substructure or subhalos, simulations use percolation algorithms, which typically require simulation history and therefore, repeated invocations, inflating the simulation cost. In contrast, our neural network does not demand simulation history, alleviating a portion of the simulation cost, while producing good results. We expect that our approach can act as a drop-in replacement to the traditional substructure-finding algorithms in gravity-only N-body simulations.

ACM Reference Format:
Jyotirmaya Shivottam and Subhankar Mishra. 2023. DeepHalo - Finding Dark Matter Halo Substructure Using Deep Learning. In *6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD) (CODS-COMAD 2023), January 4–7, 2023, Mumbai, India.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 MOTIVATION

Simulation time directly affects the rate at which adjustments can be made to the cosmological model under study, making lower simulation times desirable. In this work, we outline a neural network, called *DeepHalo*, that does not need simulation history unlike traditional substructure-finders, therefore reducing the computational cost and simulation time (refer to Table 1 for runtime comparison).

2 METHODOLOGY

We consider halos as point-clouds with a global gravitational potential field (ρ), i.e. $(3+1)D$ points, that are to be segmented into *background*, comprising particles gravitationally bound to the halo & central subhalo ($\geq$ 90%), and *foreground*, constituted by the remaining particles ($\leq$ 10%). Clearly, there is a class imbalance. We choose RandLA [2] as a base architecture to address this, as it has attentive pooling & dilation modules that jointly allow for better local structure capture. Here, the key idea is to utilize the local variation in ρ, along with particle positions. Another design choice is to use *Sigmoidal Focal Loss* [3] instead of BinaryCrossEntropy, so that overlapping components can be better distinguished. We train on halos from the $z = 0$ snapshot of the TNG50-3-Dark catalog from IllustrisTNG simulation [4] for 2 hours on 1 RTX 2080Ti, after applying usual point-cloud data augmentation techniques. After training, the network is used to segment arbitrary halos. This inferencing step does not need simulation history.

CODS-COMAD 2023, January 4–7, 2023, Mumbai, India

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Table 1: Inference / runtime for substructure finders at redshift, $z = 0$ [1, 4]. Note that our method does not require repeated calls.

Substructure Finder	Computation Time per call
SubFind (used in TNG)	35 hours $/10^9$ particles
HBT+	100 *seconds* $/10^6$ particles
DeepHalo (RTX 2080Ti)	**40 *seconds* $/10^6$ particles**

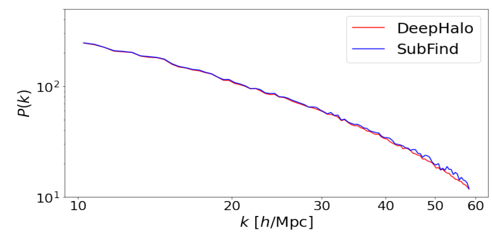

Figure 1: Power spectrum comparison between our model and SubFind. These are in good agreement at various scales.

3 ANALYSIS & CONCLUSION

We quantify segmentation performance using *accuracy* and *Dice Coefficient*, which are 92.3% and 0.831 respectively. To check, whether our substructure-finder leads to cosmologically consistent statistical results, we compute the power spectrum, as shown in Fig. 1. This along with inference times in Table 1 shows that DeepHalo is sufficiently accurate and performant to act as a drop-in replacement to traditional algorithms. In future work, we will test integration with low-cost N-body solvers and analyze segmentation output at smaller scales.

REFERENCES

[1] Jiaxin Han et al. 2017. HBT+: an improved code for finding subhaloes and building merger trees in cosmological simulations. *Monthly Notices of the Royal Astronomical Society* 474, 1 (October 2017), 604–617.
[2] Qingyong Hu et al. 2020. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE, New York, 11108–11117.
[3] FangYuan Huang et al. 2019. Weight Loss for Point Clouds Classification. *Journal of Physics: Conference Series* 1229, 1 (May 2019), 012045.

[4] Annalisa Pillepich et al. 2018. Simulating galaxy formation with the IllustrisTNG model. *MNRAS* 473, 3 (Jan. 2018), 4077–4106.