Improving Our Understanding of Dark Matter Halos with Machine Learning

Ze'ev Vladimir[1], Calvin Osinga[1], Benedikt Diemer[1]

[1]University of Maryland, College Park

Motivation

- Dark matter constitutes 80% of all matter and underlies the universe's structure and directly impacts galaxy formation and evolution [1]
- Numerical simulations are needed to investigate the complexities of these structures and the connection between dark matter and galaxies
- In the simulations, we find roughly spherical regions with an overdensity of dark matter which we call halos. A halo is depicted in Figure 1

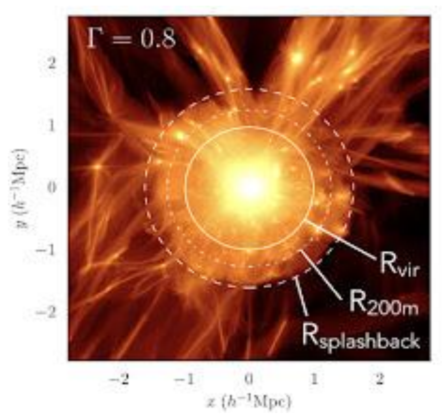

Figure 1: A dark matter halo with conventionally defined radii (R_{vir}, R_{200m}, $R_{splashback}$) for spherical overdensity models. However, it is visually clear that the halo is *not* spherically symmetrical. We see an population of orbiting particles concentrated in the center and filaments of particles infalling into the halo.

- However, determining which particles form a halo is most commonly oversimplified with the spherical over-density model [2]. Indicated by the particles contained within a sphere of the radii shown in Figure 1.
- A more physical definition for halos has recently been proposed that only the particles that orbit a halo's center of mass constitute the halo [3][4].
- Unfortunately, current methods that determine if a particle is orbiting either lack a high level of accuracy or are computationally expensive.
- **To fill this gap, we have developed a simple and robust machine-learning model that uses particle properties relative to the halo to determine if it is orbiting or infalling**

Methods

Dark Matter Simulation Output
Particle positions and velocities throughout the simulation

Halo Finder
Halo positions, velocities, and histories throughout the simulation

SPARTA: Halo and Simulation Analysis [5][6]
Halo properties
Infalling/Orbiting Particle Classification

Training Labels
True Particle Classifications

Training Dataset:
Particle Radii
Particle Radial Velocity
Particle Tangential Velocity

XGBoost Model Training [7]

Particle Classifications

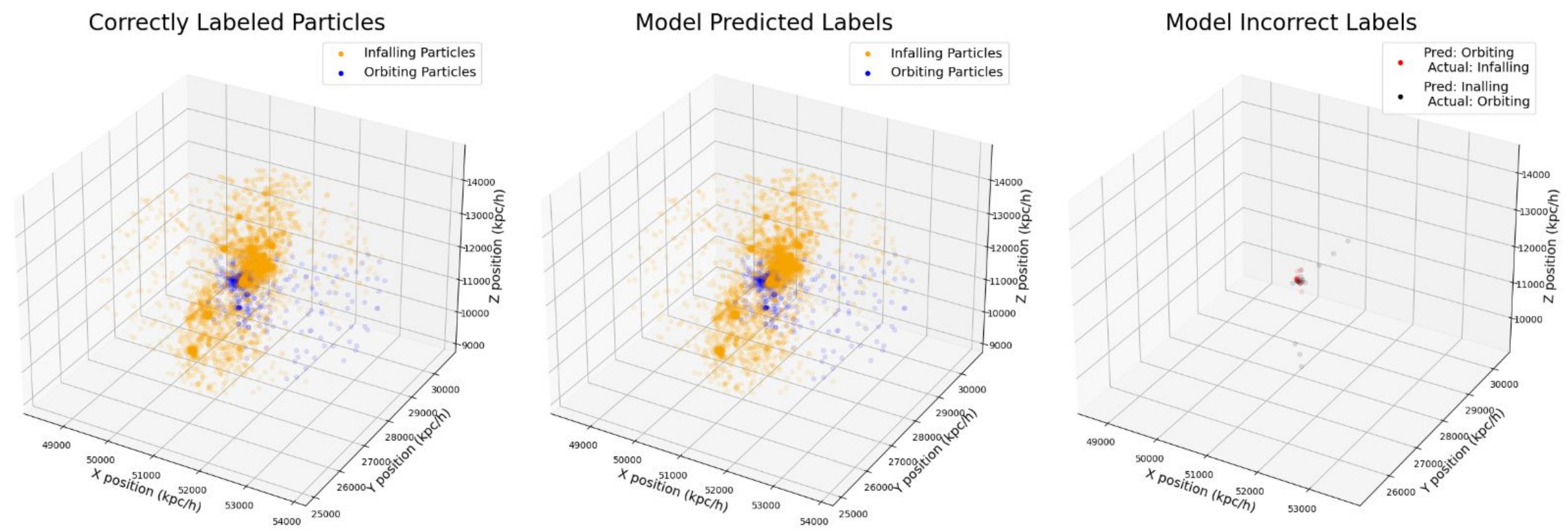

Figure 2: A single dark matter halo split by orbiting and infalling particles by the SPARTA algorithm which is defined as the truth (the left panel), the predictions from the XGBoost model (the middle panel), and which particles the XGBoost model incorrectly classifies (the right panel). We see that the model performs quite well with very few particles misclassified.

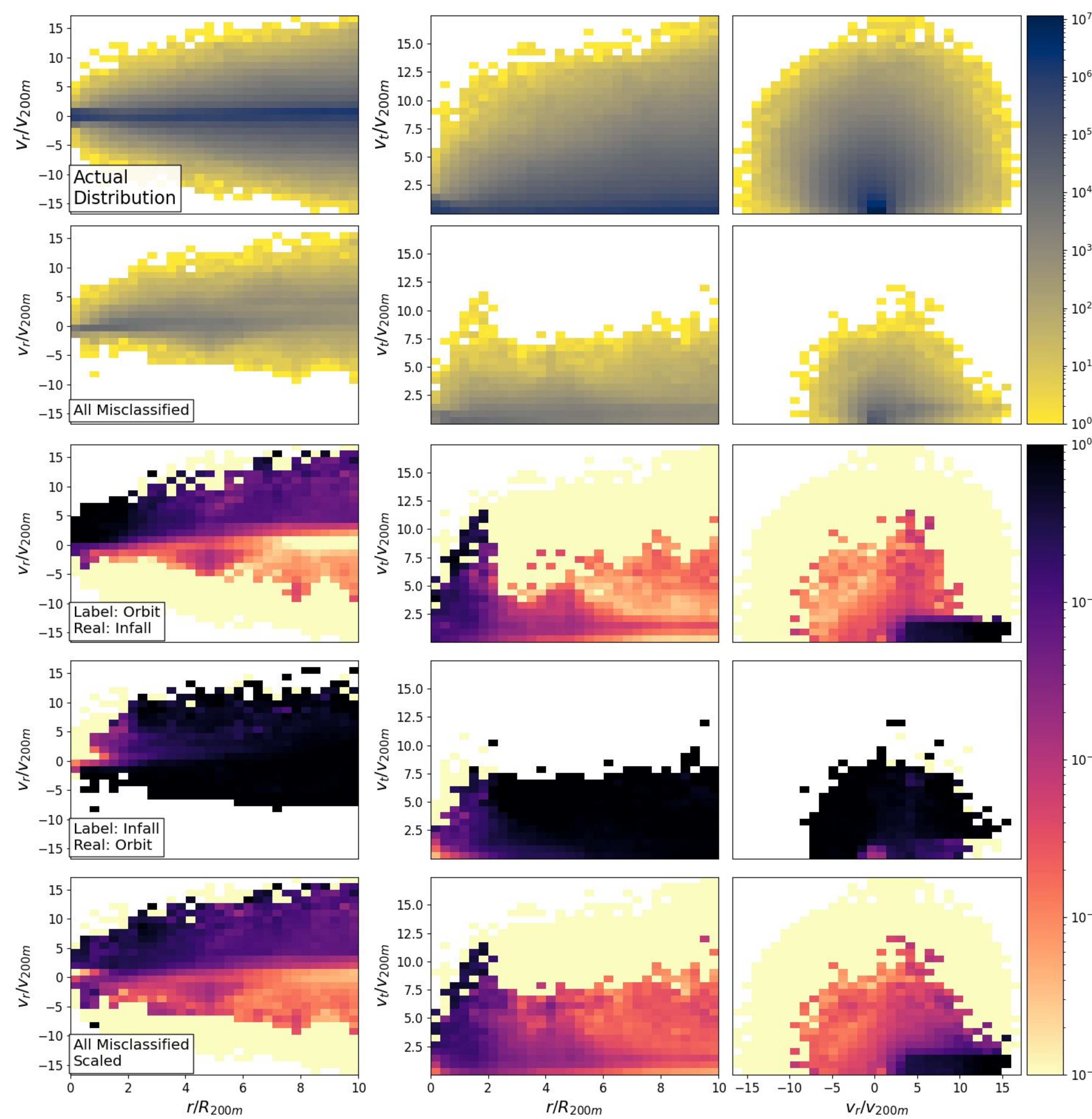

Figure 3: Phase space plots of the misclassified particles for all the halos from the validation set. We plot the distribution of the misclassified particles in the phase space of the used training parameters. This is then subdivided into particles that the Model classifies as orbiting but are actually infalling and particles that are classified as infalling but are actually orbiting scaled by the number of that type of particle in the bin.

Preliminary Results

Accuracy on the Validation Set

Depending on the simulation data used for the training set:
- The model misclassifies ~1-3% of total particles
- The model misclassifies ~3-5% of orbiting particles
- The model misclassifies ~0.5-3% of infalling particles

- We find that the model performs at the level of other current orbiting classifiers if not exceeds them.

Speed of Predictions

- After training the model is able to make predictions on all of the halos within a snapshot within minutes.
- This is a vast improvement over methods that track the particle throughout several snapshots and require large amounts of memory and computation time of up to hours.

Robustness and Applicability

- Preliminary results with a model trained on one simulation being used on another indicate high robustness with only a decrease in accuracy of 1-2%
- The model is trained on only parameters that are obtained from the positions and velocities of particles and halos from two snapshots in simulation time. These parameters are easily available from most simulation.

Next Steps

- **Further robustness tests**
  - Quantifying how the model performs on other dark matter only simulations
  - Applying the model to simulations with both baryonic matter and dark matter
- **Model accuracy improvement**
  - Using hyper-parameter tuning to best capture the information contained in the training data
  - Using models trained on specific ranges of radii to capture different groupings of particles
- **Direct model comparisons**
  - Comparing the different orbiting classification models on the same simulations
  - Comparing the speed and accuracy of these models

References

[1] White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
[2] Gunn J. E., Gott J. Richard I., 1972, APJ, 176
[3]García R., Rozo E., Becker M. R., More S., 2021, MNRAS, 505, 1195
[4] Diemer B., 2022, MNRAS, 513, 573
[5] Diemer B., 2017, APJS, 231, 5
[6] Diemer B., 2020, APJS, 251, 17
[7] Chen T., Guestrin C., 2016, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. ACM