ETH zürich

Shreyas Kalvankar

Galaxy Formation Modelling using deep learning

Thesis Pre-proposal

15th of December, 2023

Recent advancements in astronomical surveys have resulted in an exponential growth of data. Many upcoming surveys, such as Euclid and Sloan Digital Sky Survery (SDSS) are projected to yield more data than ever before. With this influx, machine learning techniques are becoming increasingly useful for expediting astronomical discovery by creating various classification and predictive models. Several machine learning approaches have been employed in tackling the data-driven problems in astronomy - e.g. semi-analytic model tuning for evolution of disc galaxies [1], galaxy morphology classification [2, 3, 4, 5], star-galaxy classification [6], and many others. The process of galaxy formation is characterized by non-linear physics constituting wide range of physical processes, making it challenging to model it analytically. Two major approaches, semi-analytic modeling, and numerical N-body simulations, are utilized to circumvent this challenge. N-body simulations utilize the fundamental equations of gravitation and hydrodynamics and solve them for a large number of particles to directly simulate the process of galaxy formation. In large scale simulations, e.g. for Milky Way mass galaxies, the number of gas and star particles exceeds 10^{10} and the physical properties for these particles, after integration of short-ranged and long-ranged forces, are calculated in short timesteps. As discussed in [7], computation of these short integration timesteps are serious bottlenecks in high-resolution simulations of galaxy formation.

In this thesis, I would like to exploit the ability of machine learning and deep learning techniques to capture the non-linear and interconnected nature of these processes to significantly reduce this bottleneck, providing faster simulations. The main objective of this thesis is to develop a deep learning algorithm for predicting the spatiotemporal properties of multiple particles in hydrodynamic processes involved in galaxy formation. Furthermore, the effective scalability of deep learning techniques can also enable us to perform complex simulations on larger datasets. [7] has shown that utilizing convolutional neural networks and LSTMs for 3D spatiotemporal forecasting has been highly effective in accurately predicting shell expansions of supernovae explosions. I would like to further this work and explore how these techniques can be efficiently applied to large-scale galaxy simulations.

The first step in this research would be to familiarize myself with numerical methods, specifically Smoothed Particle Hydrodynamics (SPH), the theory behind galaxy evolution, λCDM model of the universe & current machine learning methods employed in astrophysics for simulating complex interactions between particles. Second, to address the problem of data, different databases (e.g. [8, 9]) need to be explored, and a proper dataset must be curated consisting of data used for hydrodynamic processes from previous galaxy simulations. The deep learning algorithm 3D-MIM [7] serves as a foundation for potential architectures aimed at predicting physical properties such as velocity and gravitational fields. Each particle can be represented as a graph node where the physical properties of the particle like velocity, mass, position, etc. are the attributes of that node. I am interested in utilizing Graph Neural Networks, in addition to CNNs and LSTMs, for learning the patterns and interactions of these particles as they evolve over time. The training of such an algorithm would be computationally intensive, so I would make efficient use of hardware acceleration methods and distributed training. To assess the efficiency of this algorithm, we can benchmark its performance by comparing the predictions against the simulation results obtained using SPH.

In the final phases of the project, I would like to integrate the algorithms into a deep-learning framework that can be easily used by astrophysicists and astronomers in their simulation studies.

References

[1] John C Forbes, Mark R Krumholz, and Joshua S Speagle. Towards a radially resolved semi-analytic model for the evolution of disc galaxies tuned with machine learning. *Monthly Notices of the Royal Astronomical Society*, 487(3):3581–3606, 06 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz1473. URL https://doi.org/10.1093/mnras/stz1473.

[2] Manda Banerji, Ofer Lahav, Chris J. Lintott, Filipe B. Abdalla, Kevin Schawinski, Steven P. Bamford, Dan Andreescu, Phil Murray, M. Jordan Raddick, Anze Slosar, Alex Szalay, Daniel Thomas, and Jan Vandenberg. Galaxy Zoo: reproducing galaxy morphologies via machine learning*. *Monthly Notices of the Royal Astronomical Society*, 406(1):342–353, 07 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.16713.x. URL https://doi.org/10.1111/j.1365-2966.2010.16713.x.

[3] Andrew Schutter and Lior Shamir. Galaxy morphology—an unsupervised machine learning approach. *Astronomy and Computing*, 12:60–66, 2015.

[4] Xiao-Pan Zhu, Jia-Ming Dai, Chun-Jiang Bian, Yu Chen, Shi Chen, and Chen Hu. Galaxy morphology classification with deep convolutional neural networks. *Astrophysics and Space Science*, 364:1–15, 2019.

[5] Shreyas Kalvankar, Hrushikesh Pandit, and Pranav Parwate. Galaxy morphology classification using efficientnet architectures. *arXiv preprint arXiv:2008.13611*, 2020.

[6] Philip, N. S., Wadadekar, Y., Kembhavi, A., and Joseph, K. B. A difference boosting neural network for automated star-galaxy classification. *AA*, 385(3):1119–1126, 2002. doi: 10.1051/0004-6361:20020219. URL https://doi.org/10.1051/0004-6361:20020219.

[7] Keiya Hirashima, Kana Moriwaki, Michiko S. Fujii, Yutaka Hirai, Takayuki R. Saitoh, and Junichiro Makino. 3d-spatiotemporal forecasting the expansion of supernova shells using deep learning toward high-resolution galaxy simulations, 2023.

[8] Igor Chilingarian, Sviatoslav Borisov, Vladimir Goradzhanov, Kirill Grishin, Anastasia Kasparova, Ivan Katkov, Vladislav Klochkov, Evgenii Rubtsov, and Victoria Toptun. Rcsedv2: the largest database of galaxy properties from a homogeneously processed multi-wavelength dataset, 2021.

[9] S. McAlpine, J.C. Helly, M. Schaller, J.W. Trayford, Y. Qu, M. Furlong, R.G. Bower, R.A. Crain, J. Schaye, T. Theuns, C. Dalla Vecchia, C.S. Frenk, I.G. McCarthy, A. Jenkins, Y. Rosas-Guevara, S.D.M. White, M. Baes, P. Camps, and G. Lemson. The eagle simulations of galaxy formation: Public release of halo and galaxy catalogues. *Astronomy and Computing*, 15:72–89, April 2016. ISSN 2213-1337. doi: 10.1016/j.ascom.2016.02.004. URL http://dx.doi.org/10.1016/j.ascom.2016.02.004.

Shreyas Kalvankar