Equivariant and Modular DeepSets with Applications in Cluster Cosmology

Leander Thiele*
Department of Physics
Princeton University
Princeton, NJ 08544

Miles Cranmer
Department of Astrophysical Sciences
Princeton University
Princeton, NJ 08544

William Coulton, Shirley Ho, David N. Spergel
Center for Computational Astrophysics
Flatiron Institute
162 5th Ave, New York, NY 10010

Abstract

We design modular and rotationally equivariant DeepSets for predicting a continuous background quantity from a set of known foreground particles. Using this architecture, we address a crucial problem in Cosmology: modelling the continuous electron pressure field inside massive structures known as "clusters." Given a simulation of pressureless, dark matter particles, our networks can directly and accurately predict the background electron pressure field. The modular design of our architecture makes it possible to physically interpret the individual components. Our most powerful deterministic model improves by $70\,\%$ on the benchmark. A conditional-VAE extension yields further improvement by $7\,\%$, being limited by our small training set however. We envision use cases beyond theoretical cosmology, for example in soft condensed matter physics, or meteorology and climate science.[2]

1 Introduction

A pressing problem in cosmology is the accurate modeling of observables sourced or influenced by physics beyond gravity, in short baryonic effects. Hydrodynamic simulations are the canonical forward model for such fields; however, their computational cost is too high for them to be a viable contender in generating the vast number of realizations necessary to sample distributions. Thus, an approach that has recently emerged is the use of neural networks to map cheaper gravity-only simulations to their full-physics counterparts. Not only does this idea enable a substantial speed-up in generating realizations, but it could also improve our physical understanding; to this aim interpretable models are required.

The problem that we tackle in this work is the prediction of the electron pressure $P_e(\vec{x})$ given a gravity-only simulation. Since this is a translationally equivariant spatial problem, the seemingly natural approach chosen for similar problems, e.g. by Refs. [1–6], is a convolutional neural net (CNN), taking as input the density field of a gravity-only simulation. However, in this work we argue that existing domain knowledge on $P_e(\vec{x})$ and similar fields renders the CNN approach inferior to a set-based architecture. In fact, electron pressure values high enough to affect observables are

*`lthiele@princeton.edu`

[2]We make our code publicly available at this URL. Data products, trained models, and hyperparameter SQL databases will be shared upon reasonable request.

Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021).

predominantly found in massive gravitationally collapsed structures, called clusters. A neural network should naturally take this property into account. From this point of view, translational invariance is in fact broken, negating the main advantage of CNNs.

To a first approximation, clusters are described by their mass M_{200} and radius R_{200}. These quantities determine a characteristic pressure scale $P_{200} \propto M_{200}/R_{200}$. The electron pressure[3] $P_e(\vec{r})$ is to leading order a spherically symmetric function, commonly approximated as a generalized Navarro-Frenk-White (GNFW) profile [7–9],

$$P_e(\vec{r}) \approx \text{GNFW}(|\vec{r}|; M_{200}, R_{200})\,, \tag{1}$$

which we will use as the benchmark (with the parameterization as chosen in Ref. [9], fitted to our data). There is an inherent random element in the electron pressure field if viewed as a function of a gravity-only simulation's snapshot at a given time. The reason is that chaos washes out some of the history; in particular the time-integrated activity of the black holes at the cluster centers is difficult to infer.

We propose to learn a probabilistic mapping directly from the simulation representation, i.e. from a set of dark matter particles with associated positions $\vec{q}_i$ and velocities $\vec{v}_i$. For a cluster α our most general model can be written as

$$\hat{P}_e(\vec{r}) = F(\{(\vec{q}_i^{(\alpha)}, \vec{v}_i^{(\alpha)})\}_{i\in\alpha}; \{(\vec{q}_i^{(\vec{r})}, \vec{v}_i^{(\vec{r})})\}_{|\vec{q}_i-\vec{r}|<R}; \mathbf{s}_\alpha, \mathbf{e}_\alpha; \mathbf{a}; \vec{r})\,, \tag{2}$$

where $\mathbf{s}_\alpha$ are scalar properties describing the cluster, $\mathbf{e}_\alpha$ are normed vector properties, $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbb{I})$ is drawn from a standard normal, and we distinguish between feature tuples and SO(3) vectors using the given notation. The first argument to F is the set of dark matter particles comprising the cluster, positions and velocities are evaluated relative to the cluster position and bulk motion respectively. Conversely, the second argument is the set of particles in the vicinity of the target position $\vec{r}$, where the positions are relative to $\vec{r}$ and the velocities relative to the local bulk motion; R is a hyperparameter.

DeepSets [10] are a class of architectures that naturally operate on such sets. Given a tuple of scalars $\mathbf{f}_i$ associated with the ith dark matter particle, a DeepSet first computes another tuple $\mathbf{g}_i$ using a multi-layer perceptron (MLP). Then a pooling operation (in our case the mean) over the i-direction produces a feature tuple that is invariant under the ordering of the input particles. We denote such an architecture as a scalar DeepSet. We construct the input features $\mathbf{f}_i$ so as to make its elements SO(3) scalars [11]. This can be achieved by using properties such as $|\vec{q}_i|$, $|\vec{v}_i|$, and contractions between $\vec{q}_i$, $\vec{v}_i$ and the elements of $\mathbf{e}_\alpha$. A simple extension multiplies the $\mathbf{g}_i$ with the $\vec{q}_i$ before pooling, thus leading to an output feature tuple in which each element is an SO(3) vector. We denote such an architecture as a vector DeepSet. It is easy to see that the described vector DeepSet is rotationally equivariant, since its output is a linear combination of SO(3) vectors with SO(3) scalar coefficients. Thus we obtain a rotationally equivariant class of architectures operating directly on the particle representation instead of gridded fields.

2 Architecture

Fig. 1 schematically illustrates the various architecture components. We emphasize that most modules can be trained and evaluated independently. This modular design makes the architecture amenable to interpretation. At the end of Sec. 4 we will briefly mention several modules we have experimentally added to the architecture. At various points the cluster-scale properties $\mathbf{s}_\alpha$, $\mathbf{e}_\alpha$ are passed, which we omit for conciseness.

The function f produces the final output $\hat{P}_e(\vec{r})$ using two components, namely a (modified) GNFW prediction and the output of the 'Aggregator' MLP.

The GNFW model takes as input the target radial position $|\vec{r}|$, which is corrected for mis-centering by the 'Origin' module (the cluster finder estimates cluster positions that are not necessarily best to center the GNFW profile at).

The 'Aggregator' MLP combines multiple inputs. The 'Local' module produces scalar features from the set of dark matter particles in the vicinity of the target position, where the cutoff R is a hyperparameter. After passing the local particles through the DeepSet, we concatenate the resulting

[3] We use $\vec{r}$ for coordinates relative to a cluster's position.

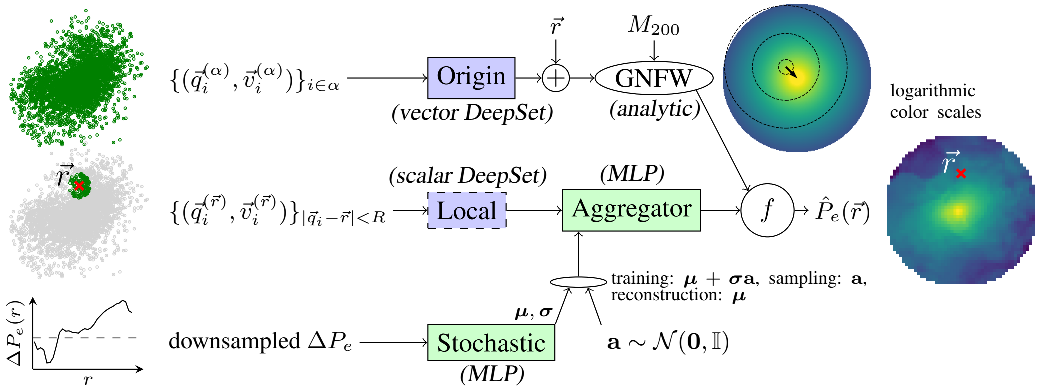

Figure 1: Schematic illustration of the architecture used in this work. Modules can be independently removed from and added to the architecture. See Sec. 2 for details.

tuple with the number of particles within R to set the scale. If other cluster-scale information, besides the Local module, is passed to f, we also pass information on $\vec{r}$ to the Aggregator. It should be noted that the use of the Local module is a form of expanding our training set (one could imagine compressing the set of all particles comprising the cluster into some code which is then evaluated at different positions; in this case, the training set would be very small however). The other input to the Aggregator models the probabilistic nature of the mapping, through a conditional VAE [12–15] architecture. The 'Stochastic' module is the standard VAE encoder, taking as input the residuals of the electron pressure field with respect to a deterministic model. Since the simulation we are using implements feedback by the central black holes in a spherically symmetric fashion, we average these residuals in spherical shells around the cluster position (another instance of domain knowledge).

3 Data and training

We use the IllustrisTNG 300-1 simulation [16–21] for training and testing. This simulation provides a gravity-only and a full-physics run with the same initial conditions. In this work, we restrict ourselves to the present-day (redshift $z = 0$) snapshot; a generalization to earlier times is naturally possible.

We use the state-of-the-art code Rockstar [22, 23] to identify clusters with masses $M_{200} > 5 \times 10^{13}\,M_\odot/h$ in the gravity-only snapshot.[4] The resulting 463 clusters are randomly assigned to training (70 %), validation (20 %), and testing (10 %) sets. They have radii R_{200} ranging from 600 to $1600\,\mathrm{kpc}/h$ and contain between 1.5 and 47 million dark matter particles within $2.5R_{200}$. The units are customary in cosmology, with $M_\odot$ the sun's mass, $1\,\mathrm{kpc} \sim 3200$ light-years, and $h \sim 0.7$.

We produce electron pressure fields from the full-physics simulation using Voxelize [24], with a voxel sidelength of $5R_{200}/64$. Our reconstruction loss $\mathcal{L}_{\text{recon}}$ for a given cluster is the mean-squared error on $P_e(\vec{r})/P_{200}$, where $|\vec{r}| < 2R_{200}$ and the normalization with P_{200} mitigates our dearth of clusters at the high-mass end.[5] The target positions $\vec{r}$ are randomly sampled during training for efficiency, while testing is of course performed on all available voxels.

For hyperparameter searches we use the Optuna package [25], solving the problem

$$\theta_{\text{opt}} = \text{argmin}\mathcal{L}_{\text{opt}}(\theta) \text{ with } \mathcal{L}_{\text{opt}}(\theta) \equiv \text{median}(\mathcal{L}_{\text{recon}}[\text{network}_\theta]/\mathcal{L}_{\text{recon}}[\text{GNFW benchmark}])\,, \quad (3)$$

where the median is over the validation set at the end of training. During training runs on architectures in which the stochastic module is included, we take as the training loss the sum of reconstruction loss and negative KL divergence of the VAE code with respect to a standard normal, the latter multiplied with a generally epoch-dependent hyperparameter. For such architectures, we perform multi-objective optimization on both $\mathcal{L}_{\text{opt}}$ and the mean of the KL divergence over the validation set.[6]

[4]The reason for this choice of mass cutoff is that in the IllustrisTNG astrophysics model at lower masses the gas physics changes qualitatively as AGN feedback is more effective in driving gas out of the cluster.

[5]For practical applications the scaling with P_{200} should be omitted, in which case a somewhat larger training set will likely be required.

[6]Total compute cost is 13.4 (Tesla P100+9CPU) khr (1.09t CO_2e [26]) with a PyTorch [27] implementation.

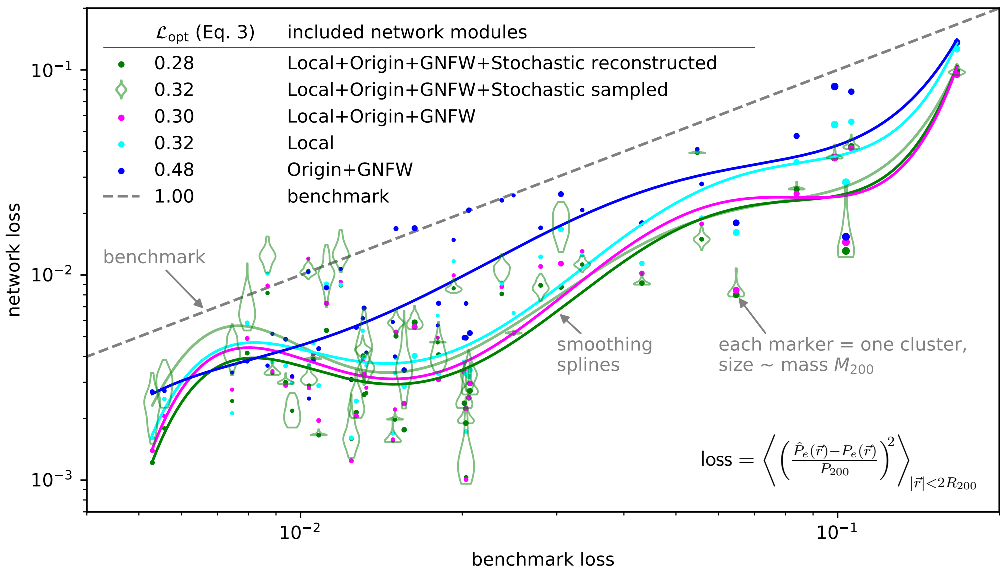

Figure 2: Network losses evaluated on testing set and compared against the GNFW benchmark model. Each data point is an individual cluster, the marker size indicating mass. The violins indicate the distribution of losses when the VAE code is randomly sampled. The lines are simple smoothing splines and only meant to guide the eye. The numbers in the legend's first column are the performance metric $\mathcal{L}_{\rm opt}$ introduced in Eq. 3 (lower is better, benchmark $\equiv 1$).

4 Results and Discussion

In Fig. 2 we plot several network losses compared against the GNFW benchmark. Only correcting for mis-centering (blue) already gives a factor ~ 2 improvement over the use of cluster centers as identified by Rockstar. Likewise, only using the dark matter matter particles in the vicinity of the evaluation point[7] (cyan) yields a further improvement. Combining the local information with the shifted GNFW profiles (magenta) performs better than Local-only by a few percent, the improvement being most pronounced in the high-loss regime. We conjecture that this could be because the addition of the simpler GNFW model helps the network generalize in these relatively rare situations. Expectedly, the model including the Stochastic module (green) generally obtains lower reconstruction losses than the other models. The corresponding losses with random VAE samples are not much worse in most cases, although a larger training set would certainly help the network learn a more robust representation of the probabilistic component.

Naturally, we should ask whether our models are learning something trivial. We have checked that a more general spherically symmetric model, implemented as an MLP that takes as input $|\vec{r}|$ and the cluster scalars s_α, does not perform more than a few percent better than the GNFW benchmark. Similarly, we find that a network using only the local density achieves more than twice the loss $\mathcal{L}_{\rm opt}$ compared to the Local network, demonstrating that the DeepSet is providing substantial information.

We have also experimented with adding further modules to the network. First, between Origin and GNFW we have inserted an MLP that uses the cluster s_α, e_α to account for deviations from spherical symmetry. We find no improvement from this modification. Second, we have constructed vector and scalar DeepSets operating on the cluster set $\{(\vec{q}_i^{(\alpha)}, \vec{v}_i^{(\alpha)})\}_{i\in\alpha}$ whose outputs were then passed to the Aggregator. Since these additional modules also do not yield any improvements, we conclude that the relatively large local regions contain enough information to infer the global properties of the cluster. It is important to appreciate that even these null results can tell us something physical, again a consequence of the interpretable, modular design.

[7]We find that $R \sim 300\,\mathrm{kpc}/h$ is a good choice.

5 Future directions

We have developed a general method to construct interpretable models that predict continuous fields from a set of points while enforcing the underlying symmetries. The application to cosmological structures demonstrates the power of our approach and opens up several directions of further investigation, e.g., symbolic regression of individual modules. Beyond cosmology, we see potential use cases in irregular structures in condensed matter or in super-resolution atmosphere models from scattered meteorological measurements.

Acknowledgements

LT thanks Kaze Wong for useful discussions. The authors acknowledge support by the Simons Foundation. All computations have been performed on the Tiger high-performance computing cluster at Princeton University.

References

[1] Tilman Tröster, Cameron Ferguson, Joachim Harnois-Déraps, and Ian G. McCarthy. Painting with baryons: augmenting N-body simulations with gas using deep generative models. *Mon. Not. R. Astron. Soc.*, 487(1): L24–L29, July 2019. doi: 10.1093/mnrasl/slz075.

[2] Jacky H. T. Yip, Xinyue Zhang, Yanfang Wang, Wei Zhang, Yueqiu Sun, Gabriella Contardo, Francisco Villaescusa-Navarro, Siyu He, Shy Genel, and Shirley Ho. From Dark Matter to Galaxies with Convolutional Neural Networks. *arXiv e-prints*, art. arXiv:1910.07813, October 2019.

[3] Xinyue Zhang, Yanfang Wang, Wei Zhang, Yueqiu Sun, Siyu He, Gabriella Contardo, Francisco Villaescusa-Navarro, and Shirley Ho. From Dark Matter to Galaxies with Convolutional Networks. *arXiv e-prints*, art. arXiv:1902.05965, February 2019.

[4] Noah Kasmanoff, Francisco Villaescusa-Navarro, Jeremy Tinker, and Shirley Ho. dm2gal: Mapping Dark Matter to Galaxies with Neural Networks. *arXiv e-prints*, art. arXiv:2012.00186, November 2020.

[5] Leander Thiele, Francisco Villaescusa-Navarro, David N. Spergel, Dylan Nelson, and Annalisa Pillepich. Teaching Neural Networks to Generate Fast Sunyaev-Zel'dovich Maps. *Astrophys. J.*, 902(2):129, October 2020. doi: 10.3847/1538-4357/abb80f.

[6] Digvijay Wadekar, Francisco Villaescusa-Navarro, Shirley Ho, and Laurence Perreault-Levasseur. HInet: Generating Neutral Hydrogen from Dark Matter with Neural Networks. *Astrophys. J.*, 916(1):42, July 2021. doi: 10.3847/1538-4357/ac033a.

[7] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. A Universal Density Profile from Hierarchical Clustering. *Astrophys. J.*, 490:493–508, December 1997. doi: 10.1086/304888.

[8] Daisuke Nagai, Andrey V. Kravtsov, and Alexey Vikhlinin. Effects of Galaxy Formation on Thermodynamics of the Intracluster Medium. *Astrophys. J.*, 668(1):1–14, October 2007. doi: 10.1086/521328.

[9] Nicholas Battaglia, J. Richard Bond, Christoph Pfrommer, and J. L. Sievers. On the Cluster Physics of Sunyaev-Zel'dovich and X-Ray Surveys. II. Deconstructing the Thermal SZ Power Spectrum. *Astrophys. J.*, 758(2):75, October 2012. doi: 10.1088/0004-637X/758/2/75.

[10] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander Smola. Deep Sets. *arXiv e-prints*, art. arXiv:1703.06114, March 2017.

[11] Soledad Villar, David W. Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Gauge-equivariant machine learning, structured like classical physics. *arXiv e-prints*, art. arXiv:2106.06610, June 2021.

[12] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *arXiv e-prints*, art. arXiv:1603.08155, 2016.

[13] Patrick Esser, Ekaterina Sutter, and Björn Ommer. A variational u-net for conditional appearance and shape generation. *arXiv e-prints*, art. arXiv:1804.04694, 2018.

[14] François Lanusse, Rachel Mandelbaum, Siamak Ravanbakhsh, Chun-Liang Li, Peter Freeman, and Barnabás Póczos. Deep generative models for galaxy image simulations. *Mon. Not. R. Astron. Soc.*, 504 (4):5543–5555, July 2021. doi: 10.1093/mnras/stab1214.

[15] Benjamin Horowitz, Max Dornfest, Zarija Lukić, and Peter Harrington. HyPhy: Deep Generative Conditional Posterior Mapping of Hydrodynamical Physics. *arXiv e-prints*, art. arXiv:2106.12675, June 2021.

[16] Volker Springel, Rüdiger Pakmor, Annalisa Pillepich, Rainer Weinberger, Dylan Nelson, Lars Hernquist, Mark Vogelsberger, Shy Genel, Paul Torrey, Federico Marinacci, and Jill Naiman. First results from the IllustrisTNG simulations: matter and galaxy clustering. *Mon. Not. R. Astron. Soc.*, 475(1):676–698, March 2018. doi: 10.1093/mnras/stx3304.

[17] Jill P. Naiman, Annalisa Pillepich, Volker Springel, Enrico Ramirez-Ruiz, Paul Torrey, Mark Vogelsberger, Rüdiger Pakmor, Dylan Nelson, Federico Marinacci, Lars Hernquist, Rainer Weinberger, and Shy Genel. First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium. *Mon. Not. R. Astron. Soc.*, 477(1):1206–1224, June 2018. doi: 10.1093/mnras/sty618.

[18] Dylan Nelson, Annalisa Pillepich, Volker Springel, Rainer Weinberger, Lars Hernquist, Rüdiger Pakmor, Shy Genel, Paul Torrey, Mark Vogelsberger, Guinevere Kauffmann, Federico Marinacci, and Jill Naiman. First results from the IllustrisTNG simulations: the galaxy colour bimodality. *Mon. Not. R. Astron. Soc.*, 475(1):624–647, March 2018. doi: 10.1093/mnras/stx3040.

[19] Federico Marinacci, Mark Vogelsberger, Rüdiger Pakmor, Paul Torrey, Volker Springel, Lars Hernquist, Dylan Nelson, Rainer Weinberger, Annalisa Pillepich, Jill Naiman, and Shy Genel. First results from the IllustrisTNG simulations: radio haloes and magnetic fields. *Mon. Not. R. Astron. Soc.*, 480(4):5113–5139, November 2018. doi: 10.1093/mnras/sty2206.

[20] Annalisa Pillepich, Dylan Nelson, Lars Hernquist, Volker Springel, Rüdiger Pakmor, Paul Torrey, Rainer Weinberger, Shy Genel, Jill P. Naiman, Federico Marinacci, and Mark Vogelsberger. First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. *Mon. Not. R. Astron. Soc.*, 475(1):648–675, March 2018. doi: 10.1093/mnras/stx3112.

[21] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, Luke Kelley, Mark Lovell, Benedikt Diemer, and Lars Hernquist. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019. doi: 10.1186/s40668-019-0028-x.

[22] Peter Behroozi, Risa Wechsler, and Hao-Yi Wu. Rockstar: Phase-space halo finder, October 2012.

[23] Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *Astrophys. J.*, 762(2):109, January 2013. doi: 10.1088/0004-637X/762/2/109.

[24] Francisco Villaescusa-Navarro, Shy Genel, Daniel Angles-Alcazar, Leander Thiele, Romeel Dave, Desika Narayanan, Andrina Nicola, Yin Li, Pablo Villanueva-Domingo, Benjamin Wandelt, David N. Spergel, Rachel S. Somerville, Jose Manuel Zorrilla Matilla, Faizan G. Mohammad, Sultan Hassan, Helen Shao, Digvijay Wadekar, Michael Eickenberg, Kaze W. K. Wong, Gabriella Contardo, Yongseok Jo, Emily Moser, Erwin T. Lau, Luis Fernando Machado Poletti Valle, Lucia A. Perez, Daisuke Nagai, Nicholas Battaglia, and Mark Vogelsberger. The CAMELS Multifield Dataset: Learning the Universe's Fundamental Parameters with Artificial Intelligence. *arXiv e-prints*, art. arXiv:2109.10915, September 2021.

[25] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. *arXiv e-prints*, art. arXiv:1907.10902, July 2019.

[26] Loïc Lannelongue, Jason Grealey, and Michael Inouye. Green algorithms: Quantifying the carbon footprint of computation. *Advanced Science*, 8(12):2100707, 2021. doi: https://doi.org/10.1002/advs.202100707. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/advs.202100707`.

[27] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] We discuss how extensions to non-zero redshifts may be performed. We also mention that a production-level model should be trained without the normalization of electron pressure by P_{200}. There is also some discussion of the limitations due to the small training set.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A] It is difficult to imagine a malicious use of the techniques presented here. In particular, although information about a group of humans could be treated as a set, our work specifically deals with the situation in which the set members have a spatial interpretation (so that a continuous function can be predicted in that space).
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] A link to an anonymized repository is included on page 1.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] We have explained our data splits and the way we optimized hyperparamters. The hyperparameter vector is quite complex and impossible to explain in the limited space without going into the details of our code. We will share hyperparameters upon request, as stated on page 1.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] There is some variance with the weight initialization, but we focus on the best models we were able to train. Marginalizing over the data splitting in this way would be very expensive, but we believe that our testing set is representative enough that a single split suffices. In particular, we observe only very minor overfitting on the validation set.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] page 3.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] We have cited IllustrisTNG, PyTorch, Optuna, Rockstar, and Voxelize.
   (b) Did you mention the license of the assets? [No] External code is only imported, not copied/modified.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] We have linked to an anonymized repository containing the entire pipeline used in this work. The processed simulation data is quite large, we would share it upon request as stated on page 1.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]