Journal of Data-centric Machine Learning Research (2024) Submitted 01/24; Revised 03/24; Published 05/24

Learning Galaxy Intrinsic Alignment Correlations

Sneh Pandya PANDYA.SNE@NORTHEASTERN.EDU
Department of Physics, Northeastern University, Boston, MA 02115, USA
NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)

Yuanyuan Yang YANG.YUANYU@NORTHEASTERN.EDU
Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115, USA

Nicholas Van Alfen VANALFEN.N@NORTHEASTERN.EDU
Department of Physics, Northeastern University, Boston, MA 02115, USA

Jonathan Blazek J.BLAZEK@NORTHEASTERN.EDU
Department of Physics, Northeastern University, Boston, MA 02115, USA

Robin Walters R.WALTERS@NORTHEASTERN.EDU
Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115, USA

Reviewed on OpenReview: *https://openreview.net/forum?id=BTiIPYEXWY*

Abstract

The intrinsic alignments (IA) of galaxies, regarded as a contaminant in weak lensing analyses, represents the correlation of galaxy shapes due to gravitational tidal interactions and galaxy formation processes. As such, understanding IA is paramount for accurate cosmological inferences from weak lensing surveys; however, one limitation to our understanding and mitigation of IA is expensive simulation-based modeling. In this work, we present a deep learning approach to emulate galaxy position-position (ξ), position-orientation (ω), and orientation-orientation (η) correlation function measurements and uncertainties from halo occupation distribution-based mock galaxy catalogs. We find strong Pearson correlation values with the model across all three correlation functions and further find proper calibration of model-predicted aleatoric uncertainties at both 1σ and 2σ. $\xi(r)$ predictions are generally accurate to $\leq 10\%$. Our model also successfully captures the underlying signal of the noisier correlations $\omega(r)$ and $\eta(r)$, although with a lower average accuracy. We find that the model performance is inhibited by the stochasticity of the data, and will benefit from correlations averaged over multiple data realizations. Our code will be made open source upon journal publication.

Keywords: Cosmology, Weak Gravitational Lensing, Intrinsic Alignments

1 Introduction

Intrinsic Alignment (IA) describes the correlation between galaxy shapes themselves, as well as the correlation between galaxy shapes and the distribution of underlying dark matter, a hypothetical form of matter that does not emit, absorb, or reflect light, yet it is thought to constitute approximately 85% of the matter in the universe and accordingly influences the structure and behavior of galaxies and galaxy clusters. These correlations pose a significant challenge in cosmological analyses. While IA offers insights into the large-scale structure of the universe, it is also a contaminant for weak gravitational lensing signals (see Troxel and Ishak 2015 for a detailed review). Weak lensing, an effect where light is deflected by grav-

©2024 Pandya and Yang.

itational fields, serves as a critical tool for studying matter distributions and cosmological constraints. However, its subtlety makes it susceptible to contamination by IA, potentially leading to significant systematic errors in signal interpretation (e.g. Hirata and Seljak 2004, Krause et al. 2015, Blazek et al. 2019). IA analyses, traditionally modeled by analytic approaches which fail to capture alignments in the full non-linear regime, have recently turned to simulation models for more accurate descriptions. These methods, however, suffer from computational expense and would benefit from efficient modeling on GPUs.

Machine learning (ML) techniques, especially neural networks (NNs), have found wide success in the sciences with the advent of high performance computing and large datasets, particularly in astrophysics and cosmology (Dvorkin et al., 2022). Of particular interest is the potential for NNs to emulate expensive N-body (Jamieson et al., 2023) and magnetohydrodynamic simulations (Rosofsky and Huerta, 2023).

In this project, we present a novel deep learning method to model both IA correlation amplitudes and uncertainties for galaxy IA statistics. Our proposed solution is a NN-based encoder-decoder architecture, trained on a wide array of galaxy catalogs derived from N-body simulations and augmented with Halo Occupation Distribution (HOD) techniques. For a given set of HOD parameters, the model is capable of simultaneous inference for all three correlation functions constructed from galaxy positions and orientations. This emulator is intended to expedite and streamline the modeling process, thereby enabling comprehensive data analysis and efficient Monte Carlo based parameter inference.

1.1 Related Work

Several previous works have constructed simulation-based emulators for cosmological statistics, with a focus on matter and/or galaxy density. Zhai et al. (2019) constructed Gaussian process based emulators based on the AEMULUS Project's N-body simulations for nonlinear galaxy clustering. Kwan et al. (2023) similarly used a Gaussian process based emulator, HOD modeling, and the Mira-Titan Suite of N-body simulations to predict galaxy correlation functions, building on earlier work from the same group (Lawrence et al., 2010). The BACCO simulation project (Aricò et al. 2021b, Aricò et al. 2021a) built NN emulators to include nonlinear and baryonic effects from simulations. These projects emulate various cosmological statistics from simulations, but do not include IA. Jagvaral et al. (2022) and Jagvaral et al. (2023) developed generative models trained on the IllustrisTNG-100 simulation (Nelson et al., 2021) to emulate IA in hydrodynamic simulations, but these models do not emulate statistics. This work is the first attempt at emulating galaxy-IA correlation statistics using simulated galaxy catalogs.

2 Data & Method

2.1 Simulated Dataset

For training the emulator, we must generate extensive galaxy catalogs that reflect realistic universes, further extracting galaxy correlation statistics from them. These, in combination with HOD model input parameters, formulate the dataset used for supervised learning.

We generate catalogs of galaxies using the `halotools` Python package. This package was created to produce mock galaxy catalogs using a HOD model-based method by populating

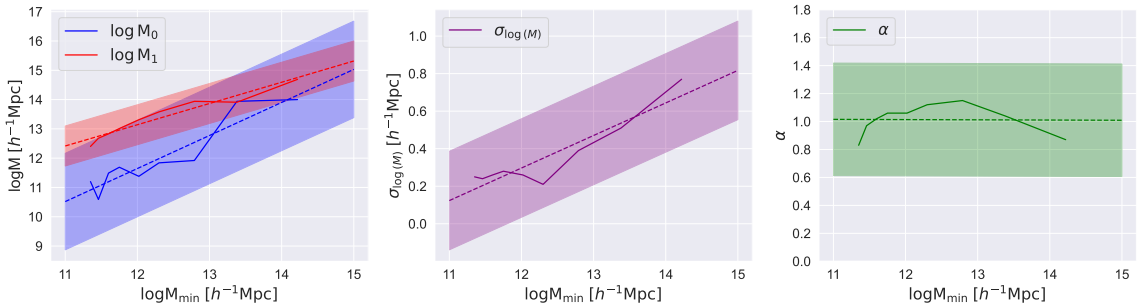

Figure 1: We generate uniform random values for the four occupation parameters, excluding $\log \mathrm{M}_{\mathrm{min}}$. These values are based on a linear relationship with $\log \mathrm{M}_{\mathrm{min}}$, serving as a central line. The range for random values extends $4 \times$ RMSE surrounding this line. To clarify the visualization, $\sigma_{\log(\mathrm{M})}$ is displayed separately from other mass variables. Each panel presents published data from Zheng et al. (2007) as a solid line, while the dotted line illustrates the linear fit to $\log \mathrm{M}_{\mathrm{min}}$, with the shaded area indicating the range for uniform random value selection for each parameter. Not shown here are the two IA parameters, μ_{cen} and μ_{sat}, which both vary uniformly on the range $[-1, 1]$ with no relation to these five occupation parameters.

existing catalogs of dark matter halos (Hearin et al., 2017). This model is extended in Van Alfen et al. 2023 to include a component for aligning galaxies to model IA; it is this version of `halotools` that we use to generate the training data for the emulator.

The HOD models encompass two populations of galaxies: central galaxies which are the most massive galaxies at the center of the larger parent halo (an isolated dark matter halo), and satellite galaxies which are less massive galaxies within that parent halo positioned elsewhere. Subhalos are defined as a dark matter halo existing within another dark matter halo. An HOD model is built using an occupation component which determines the number of galaxies to be populated within a given dark matter halo, and a phase space component which determines the galaxy positions and velocities. The extension to IA adds a galaxy alignment component, the strength of which can be set to depend on other galaxy properties.

For the purposes of this emulator, we use the `halotools` built-in occupation models `Zheng07Cens` and `Zheng07Sats` for central and satellite galaxies, respectively. These occupation model components populate dark matter halos following equations 2, 3, and 5 of Zheng et al. (2007). Central galaxies are placed at the center of the parent halo, and for simplicity, we use a subhalo phase space model for satellite galaxies which places satellites at the position of their respective subhalos. We adopt a central alignment model aligning central galaxies with their parent dark matter halo's major axis, and for satellites, a radial alignment model, aligning them along the vector from their central galaxy. Both alignment models are stochastic.

Correlation functions are measured for the position–position ($\xi(r)$), position–orientation ($\omega(r)$), and orientation–orientation ($\eta(r)$) correlations in 20 bins between galaxies in the simulation at a minimum separation of $0.1\,h^{-1}$Mpc up to a maximum separation of $16\,h^{-1}$Mpc.

The correlation function estimators are mathematically defined as

$$\xi(r) = \frac{DD(r)}{RR(r)} - 1, \qquad \omega(r) = \langle |\hat{e}(\boldsymbol{x}) \cdot \hat{r}|^2 \rangle - \frac{1}{3}, \qquad \eta(r) = \langle |\hat{e}(\boldsymbol{x}) \cdot \hat{e}(\boldsymbol{x} + \boldsymbol{r})|^2 \rangle - \frac{1}{3} \quad (1)$$

where $\boldsymbol{x}$ is the position vector of a galaxy, $DD(r)$ denotes the number of galaxy pairs separated by r, $RR(r)$ is the expected number of pairs for a random sample, and $\hat{e}(\boldsymbol{x})$ is a 3D orientation unit vector of a galaxy. In future work, we plan to extend the maximum range of this correlation, but for purposes of building this model we chose this maximum separation as correlations at this distance can still be measured quickly. In general, galaxies at $r \leq 1\,h^{-1}$Mpc are in the "1-halo regime"(galaxies within the same halo) and galaxies outside this range are in the "2-halo regime" (galaxies residing in separate halos).

The input model parameters to build the HOD model and generate the datasets are the central alignment strength, $\mu_{\rm cen}$, the satellite alignment strength, $\mu_{\rm sat}$, and five occupation components: $\log M_{min}$, $\sigma_{\log M}$, $\log M_0$, $\log M_1$, and α all described in Zheng et al. (2007). To get full coverage, we first attempted a Latin Hypercube on these parameters following Kwan et al. 2023, using the range of published values for the `Zheng07Cens` and `Zheng07Sats` occupation models in `halotools` (Zheng et al., 2007). While this provides an efficient way to generate unique combinations of parameters, we find that it can create mock galaxy catalogs with unphysical characteristics, such as $\xi(r)$ correlation amplitudes with much higher values than is typical. Specifically, the galaxy–galaxy $\xi(r)$ correlation functions were often more than 100 times greater than the dark matter $\xi(r)$ correlation function, suggesting unrealistic universes. Although the individual values chosen for each parameter fall within range of something physical, it is clear that the Latin Hypercube method can sample regions of parameter space that are not of interest. To avoid distorting the model with unphysical training data, we restrict ourselves to regions of parameter space produce realistic galaxy populations.

In selecting parameter values, we use the nine points from Table 1 in Zheng et al. (2007) to establish a linear relationship between $\log M_{min}$ and each parameter, calculating the RMSE for these fits. We then generate a sequence of evenly spaced $\log M_{min}$ values within the SDSS's range for each parameter. For each $\log M_{min}$ value, we randomly select values for the other parameters from a range centered on the linear fit, spanning $4 * \text{RMSE}$ as shown in Figure 1. $\mu_{\rm cen}$ and $\mu_{\rm sat}$ are sampled uniformly from the interval $[-1, 1]$. We populate a catalog for each set of parameters and obtain the correlations to compose the dataset of 116383 samples, further splitting it into a 70% train, 10% validation, and 20% test sets. The training data was generated using a combination of 2.4 GHz Intel E5-2680 CPUs and 2.1 GHz Intel Xeon Platinum 8176 CPUs. The simulations were parallelized across 150 cores, split evenly to allow simultaneous calculation of the correlation functions.

2.2 Model Architecture

Our objective is to construct an NN that embodies the HOD simulation and maps a 7-dimensional input vector of cosmological parameters outlined in § 2.1 to the correlation functions $\xi(r)$, $\omega(r)$, and $\eta(r)$, each comprising 20 bins. We further seek to predict the aleatoric uncertainties on the correlation amplitudes to capture the stochastic nature of HOD modeling. This is important for $\omega(r)$ and particularly $\eta(r)$, which are inherently

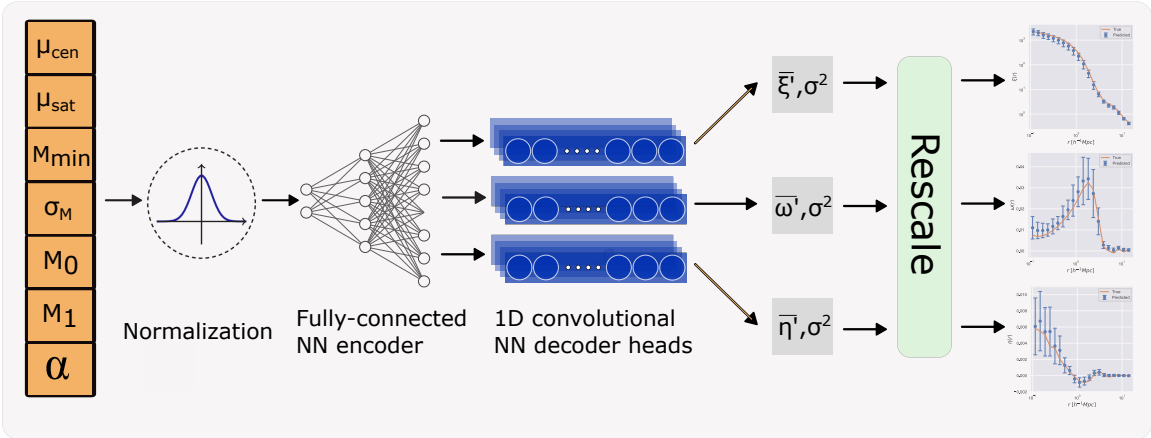

Figure 2: Model Pipeline. A logarithm factor is implied for all masses M. The HOD input model parameters are normalized before entering the 7-layer deep fully-connected encoder. The encoder expands the dimensionality of the input until the decoder stage, which features four 1D convolutional layers which learn the individual local correlations present in the output correlation functions, ξ', ω', and η'. These are then re-scaled back to their original values.

very noisy statistics due to the significant effects of galaxy orientation noise in correlations (Bernstein and Jarvis, 2002). Mathematically, this mapping can be represented by a function $f : \mathbb{R}^7 \to \mathbb{R}^{40} \times \mathbb{R}^{40} \times \mathbb{R}^{40}$. We utilize `PyTorch` to construct an architecture that encompasses a fully-connected NN shared encoder head and three 1D convolutional NN decoder heads that is trained with a multi-task learning approach.

The encoder contains seven fully connected linear layers, each accompanied with batch normalization and `LeakyReLU` activation (Xu et al., 2015). The 7D input vector undergoes a sequential expansion, reaching a width of 2048 neurons before entering the decoder stage. To mitigate overfitting, we implement dropout (Srivastava et al., 2014). Dropout later serves the purpose of isolating the epistemic uncertainty associated with the model's parameters, utilizing the Monte Carlo dropout technique (Gal and Ghahramani, 2016).

The encoder-decoder design serves a dual purpose: facilitating a vector-to-sequence conversion through the convolution of encoded representations, and, in a multi-task framework like ours, encouraging decoder heads to delineate features specific to the individual correlation function estimators, while the shared encoder captures features of the underlying HOD model. An initial bottleneck linear layer adjusts the encoded representation to a width of 200 neurons prior to entering the convolutional layers. The model features four convolution layers with batch normalization and `LeakyReLU` activation. Each decoder head outputs a $40D$ vector comprising at each bin and their accompanying variances. To ensure variances are strictly positive, they are passed through a `softplus` activation in the output layer. A diagram of the full model pipeline is shown in Figure 2.

2.3 Training

Since the decoder heads predict a distribution over the correlation function values, the model is trained with Gaussian negative log-likelihood loss (Nix and Weigend, 1994). To simultaneously optimize for predicting all three correlation functions, the losses are summed. It is thus critical that the scale of all loss terms are roughly equal to ensure uniform learning.

We apply normalization on the inputs and each output to scale them to have zero mean and unit variance. $\xi(r)$ can exhibit strong correlations at low r, reaching amplitudes of $O(1000)$ or higher. $\omega(r)$ and $\eta(r)$ are however significantly noisier and have amplitudes several orders of magnitude smaller than $\xi(r)$. Amplitudes are minuscule at high r for all correlation functions. Standard normalization is typical in deep learning but is especially important here due to the large variance in magnitude of the correlation amplitudes.

We train and validate our model with the AdamW optimizer (Loshchilov and Hutter, 2019), with hyperparameter tuning to optimize performance. Optimal parameters include a batch size of 128, a step learning rate scheduler (10% decay at 500-epoch intervals for a starting lr = 0.01), and for 1500 epochs with early stopping to discourage overfitting. We additionally employ L2-regularization via a weight decay factor of 10^{-4} in the optimizer. All training was conducted on one NVIDIA A100-80GB GPU.

In order to obtain epistemic uncertainties for the model predictions, we employ Monte Carlo dropout during inference. This method involves conducting several forward passes for a prediction with dropout layers left on, so that there is variation in the networks predictions. After 20 passes, we can obtain the model outputted mean and accordingly isolate the variance as the epistemic variance. A dropout rate of 0.2 is used throughout the encoder and decoder.

3 Results

Accuracy We evaluate the model performance on the 20% test set. It can be seen in Figure 3 that the median errors of $\xi(r)$ and $\omega(r)$ (blue line) are well-behaved and near the zero error benchmark (red-dashed line). Further, the 50% interquartile range (IQR) shows shows a uniform variability around the median error for $\omega(r)$ and $\eta(r)$ and includes an error of zero. For $\xi(r)$, the 50% IQR is skewed high at low r, indicating a bias. The median error of $\eta(r)$ is $\sim 25\%$ at low r and $\leq 10\%$ in the 2-halo regime. Outlier predictions are also shown in Figure 3, showcasing the variability in performance at low r for $\xi(r)$ and across all scales for $\eta(r)$ and $\omega(r)$.

In general, model accuracy and confidence increases with higher r, when most correlations are small. For $\xi(r)$, at low r model predictions are biased high and there is high aleatoric uncertainty. The fractional error is generally $\leq 15\%$, and the model achieves $\sim 2.5\%$ accuracy at high r. The statistics $\omega(r)$ and $\eta(r)$ exhibit significantly higher noise levels, resulting in larger aleatoric uncertainties. Preliminary testing reveals that the average uncertainties closely align with those extracted directly from the simulation. For $\omega(r)$, the fractional errors are generally $\leq 20\%$, and the model's predictions tend to be biased low. Moreover, for $\omega(r)$ and $\eta(r)$, the ground truth falls within the 50% IQR of the median prediction across all bins.

We find strong correlations between model predictions and the respective ground truths as characterized by the Pearson correlation coefficient (PCC), which quantifies agreement

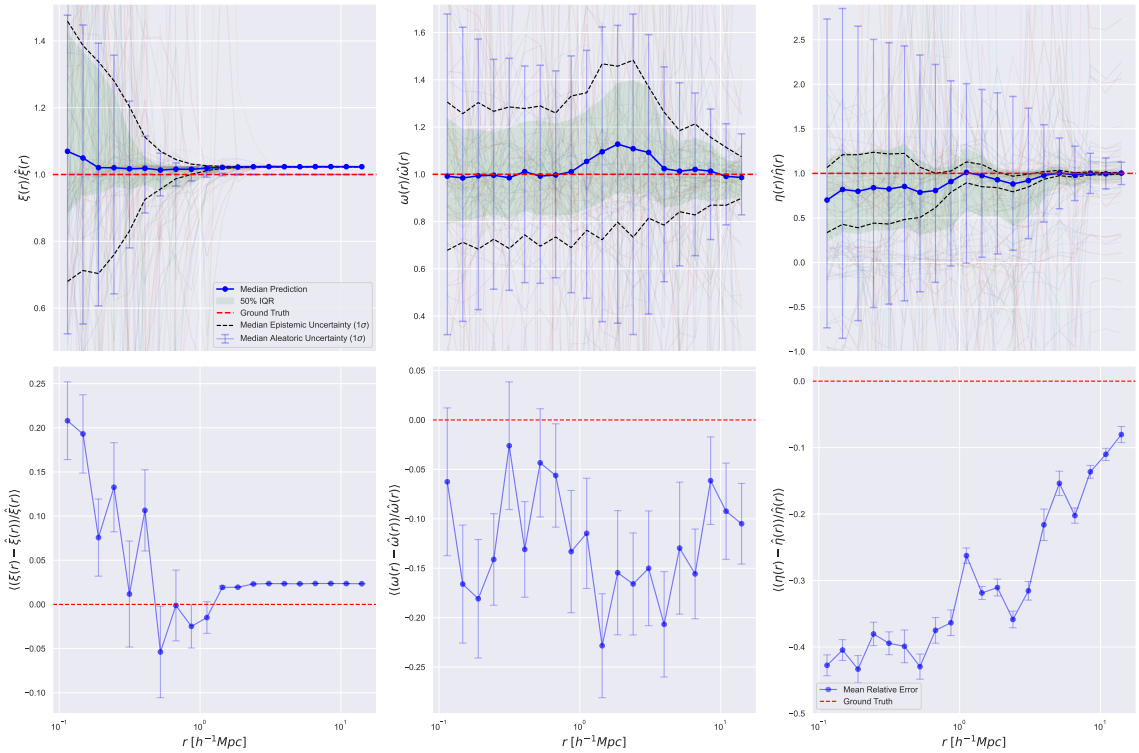

Figure 3: **Top:** Plot of median errors with aleatoric and epistemic 1σ regions of the median predictions for position-position ($\xi(r)$), position-orientation ($\omega(r)$), and orientation-orientation ($\eta(r)$) correlation functions in the scaled domain. The 50% IQR is also shown in green. The results for a random sample of 100 test-set predictions are shown in the background, showcasing the variability of performance. **Bottom:** Mean fractional errors with aleatoric uncertainties for $\xi(r)$, $\omega(r)$, and $\eta(r)$. Ratios are computed with respect to the running mean using a window size of 3 for $\omega(r)$ and $\eta(r)$ due to their increased noise.

between the overall shapes of the ground truth and model predictions. We find that the mean $PCC(\xi) = 0.98$, $PCC(\omega) = 0.88$, and $PCC(\eta) = 0.65$. The fractional errors of the model do not indicate a particularly strong performance in the case of $\omega(r)$ and $\eta(r)$; these metrics have potential to be inflated due to the small amplitudes and large stochasticity of these correlations, even when using methods such as computing the running mean to illustrate the errors as shown in Figure 3. Example model predictions for a variety of scenarios is further shown in Appendix A. We emphasize in this case that it was the intention for the model to capture this underlying signal of the correlations and not overfit to noise, which is supported by the PCC values.

Calibration We find that 45.18% of 1σ epistemic uncertainties are smaller than aleatoric uncertainties for $\xi(r)$, indicating that a larger training sample would be of benefit. 98.22% and 99.72% are smaller than the aleatoric uncertainty at 1σ for $\omega(r)$ and $\eta(r)$. For $\xi(r)$, we find that 40.32% and 47.23% of predictions fall within 1σ and 2σ of their ground truth values, respectively. This low percentage is largely due to the highly confident, but systemically biased predictions for correlations at large r as seen in the fractional error of $\sim 2.5\%$ in Figure 3. As seen in Figure 3, the high stochasticity in $\omega(r)$ and $\eta(r)$ has resulted in large model-predicted aleatoric uncertainties. For $\omega(r)$, the corresponding percentages are 85.80% within 1σ and 98.82% within 2σ. For $\eta(r)$, 77.43% of predictions fall within 1σ and 97.35% within 2σ. These confidence intervals indicate that the aleatoric uncertainty is slightly inflated; nevertheless, for a more robust conclusion, we plan to to conduct a thorough validation by directly extracting the uncertainty from the simulation data itself by running multiple realizations for each set of parameters.

Limitations We are largely limited by the sparse signal present in inherently noisy correlations such as $\omega(r)$ and $\eta(r)$. The model can capture the underlying signal for these two correlations but accordingly predicts a large uncertainty, particularly for low r. This was indeed the desired outcome – that the model would learn to predict the "cosmic mean" in the presence of significant noise. At high r, when correlations are small, the median predictions are largely correct and confident, aside from the $\sim 2.5\%$ systemic bias in $\xi(r)$. The PCC values indicate that despite considerable error in correlation amplitudes, the model successfully captures the long-range behavior of the correlations well.

Previous Considerations We previously tested other data normalization schemes such as min-max normalization and custom scaling, wherein we defined $\xi' = \xi/\xi_{DM}$ and scaled $\omega(r)$ and $\eta(r)$ as

$$\omega_i' = \frac{\mu_{\xi'}}{\mu_\omega} \cdot \omega_i \qquad \eta_i' = \frac{\mu_{\xi'}}{\mu_\eta} \cdot \eta_i \tag{2}$$

where μ denotes the mean. This normalization was extensively studied as it includes some information of the underlying dark matter distribution, which we believed would assist the learning. However, we found that it disadvantaged $\omega(r)$ in the 1-halo regime and did not work as well as standard normalization. Interestingly, the bias in $\xi(r)$ predictions at high r was absent with this normalization. Lastly, we studied other architectures, including fully-connected and U-Net based architectures, as well as variations to the current encoder-decoder design and found that the inclusion of 1D convolution aided performance. Single-task learning on individual correlations was also studied, in which we found that $\omega(r)$ and

$\eta(r)$ performance was slightly improved but more susceptible to overfitting. Thus, the information shared among correlations in the encoder stage yields benefits over single-task learning. We additionally conducted experiments in only predicting point-estimates, and found that the inclusion of mean-variance estimation was essential to understanding the degree of stochasticity in $\omega(r)$ and $\eta(r)$ and properly quantifying their performance.

4 Discussion

We have presented a model that efficiently predicts galaxy IA correlations in terms of HOD simulation parameters without costly simulation. The model can perform inference on a batch of $32,768$ input parameters in 1.02 seconds on one NVIDIA A100-80GB GPU, while the simulation when run in parallel on 150 CPU cores for the same parameters takes $\mathcal{O}(3 \text{ hours})$. The model simultaneously predicts point estimates and uncertainties for three correlations spanning 20 bins whose values span several orders of magnitude and are also significantly noisy, making it a notable data and model engineering task to isolate the relevant signal and not overfit to the noise when signal is sparse. The model effectively avoids overfitting to this noise and demonstrates conservative aleatoric uncertainties, which will be validated in future work. Furthermore, it accurately captures the underlying signal of correlations as shown by the PCC values. Point-predictions are most accurate for $\xi(r)$ ($\leq 15\%$, generally), with large fractional errors for $\omega(r)$ and $\eta(r)$ which can be inflated due to the large stochasticity of these correlations.

The epistemic uncertainty of the model is typically lower than the aleatoric uncertainty and is well-calibrated as noted by the confidence interval statistics shown in § 3. The epistemic uncertainty is something we aim to decrease in future work. This can be mitigated by conducting multiple realizations for parameters to curate a larger training sample, or alternatively by decreasing the number of parameters in the encoder stage of the model (Semenova et al., 2022), though typically at the expense of model expressivity. In doing so, we can improve the quality of model predictions, further calibrate its aleatoric uncertainty predictions, and simultaneously provide a benchmark for which to validate them with.

4.1 Future Work

Several improvements can be made at both a data engineering and architectural level. Of particular importance is understanding the true degree of stochasticity in $\omega(r)$ and $\eta(r)$ and addressing the systemic bias for $\xi(r)$ in the 2-halo regime. We also plan to study different parameter configurations for generating data and further investigate combinations which resulted in unphysical universes. We plan to perform parameter inference on μ_{sat} and μ_{cen}, as well as other galaxy occupation parameters, with the model utilizing Markov Chain Monte Carlo (MCMC) once the pipeline is refined. The simulation itself is very representative of real data; the eventual goal is to validate the model with the IllustrisTNG suite of simulations, similar to what was done in Van Alfen et al. (2023), as well as real data. We additionally plan to consider more complex HOD model dependence, such as distance-dependence alignment. The eventual goal is to generalize well beyond the cosmological parameters that are implicit in the models training data, such as those that determine the underlying cosmological simulation, with hopes of creating a unifying and efficient emu-

lator which will accelerate the study of IA correlations and their effects on cosmological measurements.

5 Reproducibility Statement

The entire procedure of this work, from data generation to modeling and evaluation, are able to reproduced. The data generation procedure can be reproduced using `halotools` (`https://halotools.readthedocs.io`) with the appropriate parameters and literature outlined in § 2.1. A description of the encoder-decoder architecture that was written in `PyTorch` is summarized in § 2.2 and § 2.3, with a detailed description of the architecture further provided in Appendix B. Our code will be made open source upon journal publication.

6 Acknowledgements

We thank the anonymous referees for their useful comments. S.P. acknowledges support from the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, `https://iaifi.org`). Y.Y. acknowledges support from Khoury Apprenticeship program. J.B. and N.V.A. are supported in this work by NSF award AST-2206563 and the Roman Research and Support Participation program under NASA grant 80NSSC24K0088. R.W. is supported by NSF award DMS-2134178. Data generation was conducted on the Discovery cluster, supported by Northeastern University's Research Computing team. The machine learning computations were run on the FASRC cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

References

Giovanni Aricò, Raul E. Angulo, Sergio Contreras, Lurdes Ondaro-Mallea, Marcos Pellejero-Ibañez, and Matteo Zennaro. The BACCO simulation project: a baryonification emulator with neural networks. *Monthly Notices of the Royal Astronomical Society*, 506(3):4070–4082, September 2021a. doi: 10.1093/mnras/stab1911.

Giovanni Aricò, Raul E. Angulo, and Matteo Zennaro. Accelerating Large-Scale-Structure data analyses by emulating Boltzmann solvers and Lagrangian Perturbation Theory. *arXiv e-prints*, art. arXiv:2104.14568, April 2021b. doi: 10.48550/arXiv.2104.14568.

G. M. Bernstein and M. Jarvis. Shapes and shears, stars and smears: Optimal measurements for weak lensing. *The Astronomical Journal*, 123(2):583, feb 2002. doi: 10.1086/338085. URL `https://dx.doi.org/10.1086/338085`.

Jonathan A. Blazek, Niall MacCrann, M. A. Troxel, and Xiao Fang. Beyond linear galaxy alignments. *Phys. Rev. D*, 100:103506, Nov 2019. doi: 10.1103/PhysRevD.100.103506. URL `https://link.aps.org/doi/10.1103/PhysRevD.100.103506`.

Cora Dvorkin, Siddharth Mishra-Sharma, Brian Nord, V. Ashley Villar, Camille Avestruz, Keith Bechtol, Aleksandra Ćiprijanović, Andrew J. Connolly, Lehman H. Garrison, Gautham Narayan, and Francisco Villaescusa-Navarro. Machine learning and cosmology, 2022.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, 2016.

Andrew P. Hearin, Duncan Campbell, Erik Tollerud, Peter Behroozi, Benedikt Diemer, Nathan J. Goldbaum, Elise Jennings, Alexie Leauthaud, Yao-Yuan Mao, Surhud More, John Parejko, Manodeep Sinha, Brigitta Sipöcz, and Andrew Zentner. Forward modeling of large-scale structure: An open-source approach with halotools. *The Astronomical Journal*, 154(5):190, oct 2017. doi: 10.3847/1538-3881/aa859f. URL `https://dx.doi.org/10.3847/1538-3881/aa859f`.

C. M. Hirata and U. Seljak. Intrinsic alignment-lensing interference as a contaminant of cosmic shear. *Physical Review D*, 70(6):063526–+, September 2004. doi: 10.1103/PhysRevD.70.063526.

Yesukhei Jagvaral, François Lanusse, Sukhdeep Singh, Rachel Mandelbaum, Siamak Ravanbakhsh, and Duncan Campbell. Galaxies and haloes on graph neural networks: Deep generative modelling scalar and vector quantities for intrinsic alignment. *Monthly Notices of the Royal Astronomical Society*, 516(2):2406–2419, August 2022. ISSN 1365-2966. doi: 10.1093/mnras/stac2083. URL `http://dx.doi.org/10.1093/mnras/stac2083`.

Yesukhei Jagvaral, Francois Lanusse, and Rachel Mandelbaum. Unified framework for diffusion generative models in SO(3): applications in computer vision and astrophysics. *arXiv e-prints*, art. arXiv:2312.11707, December 2023. doi: 10.48550/arXiv.2312.11707.

Drew Jamieson, Yin Li, Renan Alves de Oliveira, Francisco Villaescusa-Navarro, Shirley Ho, and David N. Spergel. Field-level neural network emulator for cosmological n-body simulations. *The Astrophysical Journal*, 952(2):145, July 2023. ISSN 1538-4357. doi: 10.3847/1538-4357/acdb6c. URL `http://dx.doi.org/10.3847/1538-4357/acdb6c`.

Elisabeth Krause, Tim Eifler, and Jonathan Blazek. The impact of intrinsic alignment on current and future cosmic shear surveys. *Monthly Notices of the Royal Astronomical Society*, 456(1):207–222, 12 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv2615. URL `https://doi.org/10.1093/mnras/stv2615`.

Juliana Kwan, Shun Saito, Alexie Leauthaud, Katrin Heitmann, Salman Habib, Nicholas Frontiere, Hong Guo, Song Huang, Adrian Pope, and Sergio Rodriguéz-Torres. Galaxy clustering in the mira-titan universe. i. emulators for the redshift space galaxy correlation function and galaxy–galaxy lensing. *The Astrophysical Journal*, 952(1):80, July 2023. ISSN 1538-4357. doi: 10.3847/1538-4357/acd92f. URL `http://dx.doi.org/10.3847/1538-4357/acd92f`.

Earl Lawrence, Katrin Heitmann, Martin White, David Higdon, Christian Wagner, Salman Habib, and Brian Williams. The Coyote Universe. III. Simulation Suite and Precision Emulator for the Nonlinear Matter Power Spectrum. *The Astrophysical Journal*, 713(2): 1322–1331, April 2010. doi: 10.1088/0004-637X/713/2/1322.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019.

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, 2021.

D.A. Nix and A.S. Weigend. Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, volume 1, pages 55–60 vol.1, 1994. doi: 10.1109/ICNN.1994.374138.

Shawn G Rosofsky and E A Huerta. Magnetohydrodynamics with physics informed neural operators. *Machine Learning: Science and Technology*, 4(3):035002, July 2023. ISSN 2632-2153. doi: 10.1088/2632-2153/ace30a. URL `http://dx.doi.org/10.1088/2632-2153/ace30a`.

Nadezhda Semenova, Laurent Larger, and Daniel Brunner. Understanding and mitigating noise in trained deep neural networks. *Neural Networks*, 146:151–160, February 2022. ISSN 0893-6080. doi: 10.1016/j.neunet.2021.11.008. URL `http://dx.doi.org/10.1016/j.neunet.2021.11.008`.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014. URL `http://jmlr.org/papers/v15/srivastava14a.html`.

M.A. Troxel and Mustapha Ishak. The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology. *Physics Reports*, 558:1–59, 2015. ISSN 0370-1573. doi: https://doi.org/10.1016/j.physrep.2014.11.001. URL https://www.sciencedirect.com/science/article/pii/S0370157314003974. The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology.

Nicholas Van Alfen, Duncan Campbell, Jonathan Blazek, C. Danielle Leonard, Francois Lanusse, Andrew Hearin, Rachel Mandelbaum, and The LSST Dark Energy Science Collaboration. An empirical model for intrinsic alignments: Insights from cosmological simulations, 2023.

Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network, 2015.

Zhongxu Zhai, Jeremy L. Tinker, Matthew R. Becker, Joseph DeRose, Yao-Yuan Mao, Thomas McClintock, Sean McLaughlin, Eduardo Rozo, and Risa H. Wechsler. The aemulus project. iii. emulation of the galaxy correlation function. *The Astrophysical Journal*, 874(1):95, March 2019. ISSN 1538-4357. doi: 10.3847/1538-4357/ab0d7b. URL http://dx.doi.org/10.3847/1538-4357/ab0d7b.

Zheng Zheng, Alison L. Coil, and Idit Zehavi. Galaxy evolution from halo occupation distribution modeling of DEEP2 and SDSS galaxy clustering. *The Astrophysical Journal*, 667 (2):760–779, oct 2007. doi: 10.1086/521074. URL https://doi.org/10.1086%2F521074.

A Model Predictions

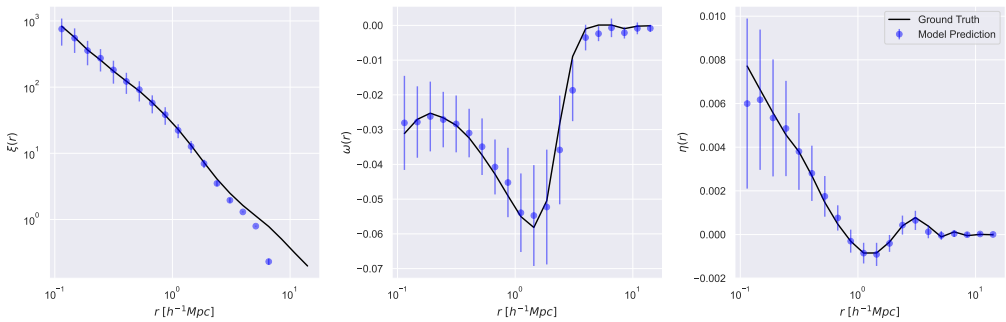

Figure 4: Good performance examples on test set in rescaled domain. On the left panel one can see the small bias in $\xi(r)$ predictions at high r.

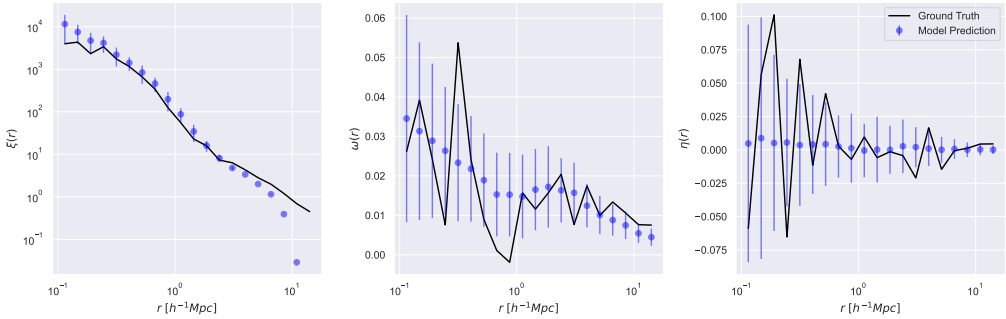

Figure 5: Noisy data examples on test set in rescaled domain. $\xi(r)$ exhibits a small bump at low r which is likely a noisy artifact. $\omega(r)$ and $\eta(r)$ are very noisy, but the model is able to capture the underlying signal as well as overall shape.

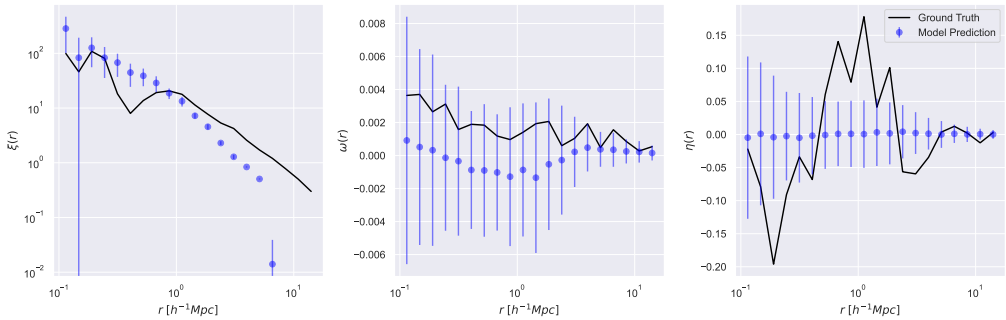

Figure 6: Poor performance examples on test set in rescaled domain.

B Model Architecture

Layers	Properties	Stride	Padding	Output Shape
	Encoder			
Input: 7				
Linear (w/ BatchNorm1D)	Width: 128 Dropout: 0.2 Activation: LeakyReLU	-	-	(, 128)
Linear (w/ BatchNorm1D)	Width: 256 Dropout: 0.2 Activation: LeakyReLU	-	-	(, 256)
Linear (w/ BatchNorm1D)	Width: 512 Dropout: 0.2 Activation: LeakyReLU	-	-	(, 512)
Linear (w/ BatchNorm1D)	Width: 1024 Dropout: 0.2 Activation: LeakyReLU	-	-	(, 1024)
Linear (w/ BatchNorm1D)	Width: 2048 Dropout: 0.2 Activation: LeakyReLU	-	-	(, 2048)
	Decoder (x3)			
Linear (Bottleneck)	Width: 200			(, 200)
Conv1D (w/ BatchNorm1D)	Filters: 20 Kernel: 3x3 Activation: LeakyReLU Dropout: 0.2	1	1	(20, 10)
Conv1D (w/ BatchNorm1D)	Filters: 40 Kernel: 3x3 Activation: LeakyReLU Dropout: 0.2	1	1	(40, 10)
Conv1D (w/ BatchNorm1D)	Filters: 80 Kernel: 5x5 Activation: LeakyReLU Dropout: 0.2	1	2	(80, 10)
Conv1D (w/ BatchNorm1D)	Filters: 20 Kernel: 3x3 Activation: LeakyReLU	5	1	(20, 2)
Output: (, 3, 40)				