

New Perspectives



Contribution ID: 38

Type: **not specified**

DeepBench: A simulation library for cosmology focused dataset generation

Wednesday, June 22, 2022 2:15 PM (15 minutes)

The physics community lacks user-friendly computational tools for constructing simple simulated datasets for benchmarking and education in machine learning and computer vision. We introduce the python library DeepBench, which generates highly reproducible datasets at varying levels of complexity, size, and content focused on a cosmological context. DeepBench produces both highly simplified and more complex models of astronomical objects. For instance, basic geometric shapes, such as a disc and multiple arcs, could be used to simulate a strong gravitational lens. For more realistic models of astronomical objects, such as stars or elliptical galaxies, DeepBench simulates each of their well-recorded profile distribution functions. Beyond 2D images, we can also produce 1D representations of quasar light curves and galaxy spectra. We also include tools to collect and store the dataset for consumption by a machine learning algorithm. Finally, we present a trained ResNet50 model as an illustration of the expected use of the software as a benchmarking tool for testing the suitability of various architectures for a scientifically motivated problem.

We envision this tool to be useful in a suite of contexts at the intersection of cosmology and machine learning. The simplistic nature of the simulated data permits us to rapidly generate arbitrarily large data sets, from single-object fields to multi-object fields. The data can have both categorical and floating point labels so that a variety of tasks can be tested simultaneously or in a progression on the same data set – e.g., both classification and regression. We expect the tool to be of significant interest and utility both for a wide range of users. For those new to machine learning, it can produce toy-model datasets that behave similarly to astronomical data. For ML experts, it can be used to carefully and systematically test models.

Primary authors: LEWIS, Ashia (Fermilab); VOETBERG, Margaret (Fermi National Accelerator Laboratory)

Co-authors: JONES, Craig; HLOZEK, Renee (Dunlap Institute & Department of Astronomy and Astrophysics, University of Toronto); CIPRIJANOVIC, Aleksandra (Fermi National Accelerator Laboratory); PERDUE, Gabriel (Fermilab); NORD, Brian (Fermilab)

Presenters: LEWIS, Ashia (Fermilab); VOETBERG, Margaret (Fermi National Accelerator Laboratory)

Session Classification: Cosmic Physics