FERMILAB-SLIDES-22-086-SCD

DeepBench: A simulation library for cosmology focused dataset generation

Ashia Lewis, Margaret Voetberg, Craig Jones, Renee Hlozek, Aleksandra Ćiprijanović, Gabriel Perdue, Brian Nord

DEEP SKIES
Bringing Artificial Intelligence to Astrophysics

Fermilab

U.S. DEPARTMENT OF ENERGY

Introduction

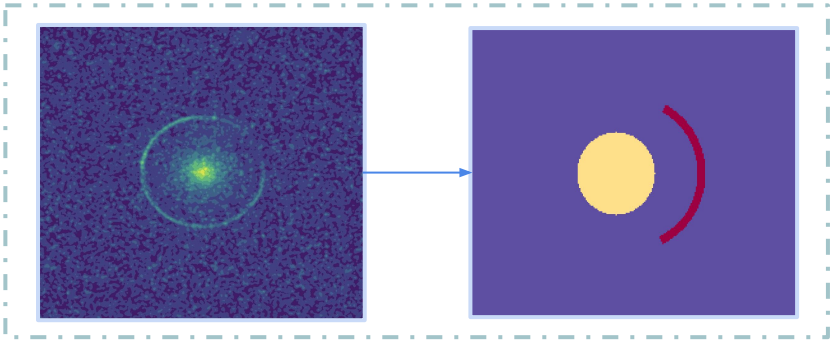
Motivation

Beginner-Friendly

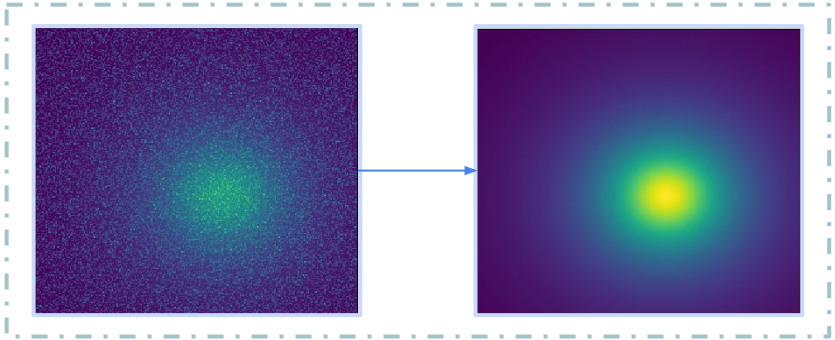
- Prioritize ease of use for new practitioners
- Finding a good dataset is hard
- Making your own is harder
- Hopefully this is less hard.

Diagnosis Tool

- Flexible nature of generation allows for “scaling back” a problem



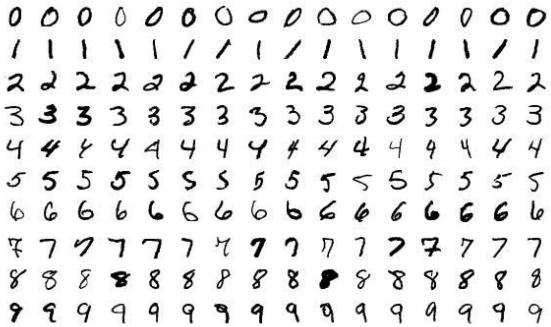
Higher to Lower Complexity



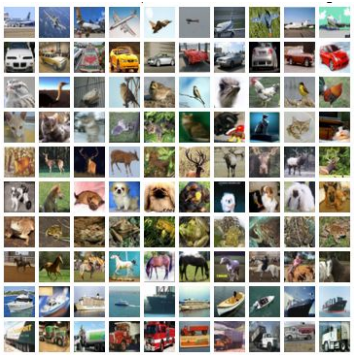
High to Low Noise

Transfer Learning and Network Benchmarking

- Introduce a varied dataset that prioritizes astronomy primitives
- Generate weights for popular architectures for use in transfer learning applications.
    - Generally decrease training time and increase accuracy
- Encourage use of astronomy and cosmology focused benchmarks for collaborative and comparative purposes across architectures



MNIST, a commonly used benchmark dataset

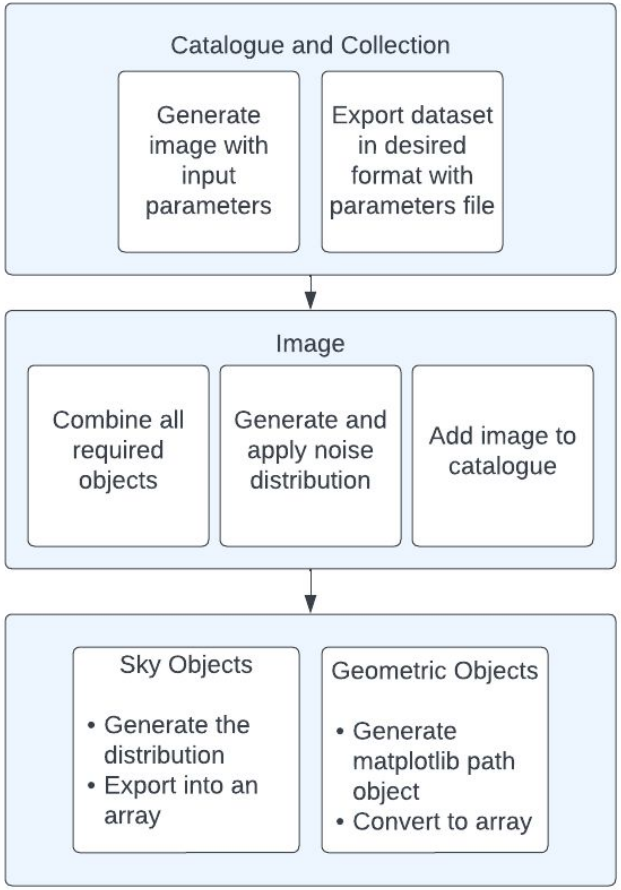


CIFAR-10, a ten class benchmark dataset

Methods

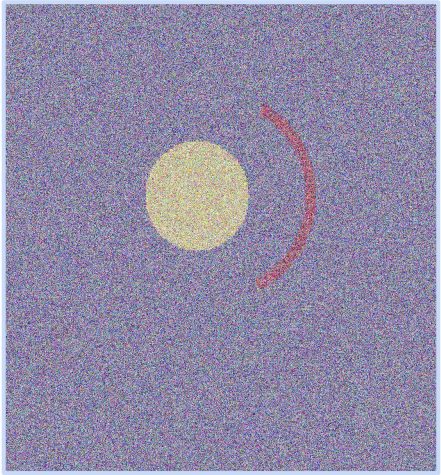
Overview

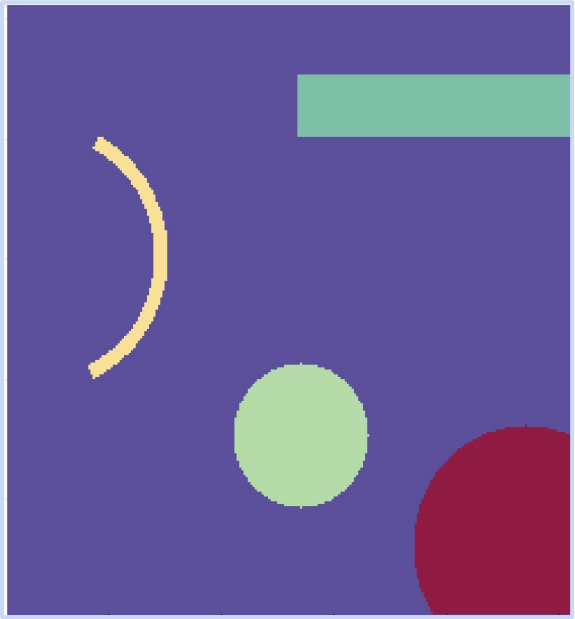
- Catalogue and Collection
  - Broker and export
- Image
  - Compose each image
  - Adds “image level” noise: Poisson or Gaussian
- Objects
  - Generate object used in the image
- Distributed with quickstart configuration files



Geometric Objects

- Geometric objects are generated via the library matplotlib
- Included shapes: n-sided, regular polygons, arcs, straight lines, ellipses, and circles



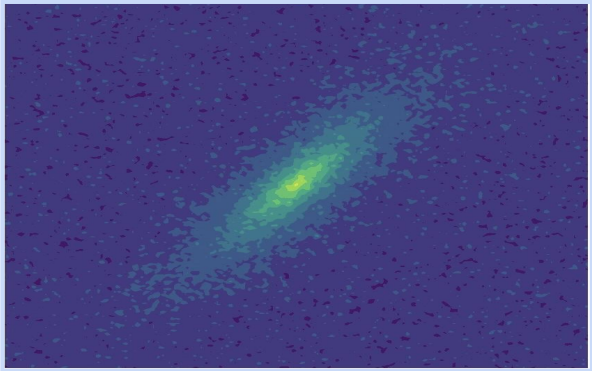


Astronomical Objects

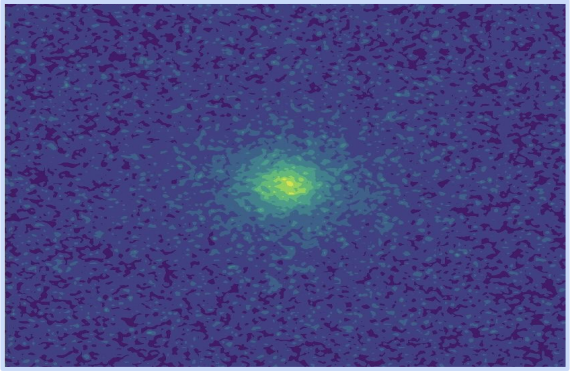
- Simplified renderings of common profiles found in astronomy datasets
- Flexible object parameters
- Currently focuses on 2D modelling

Astronomical Objects

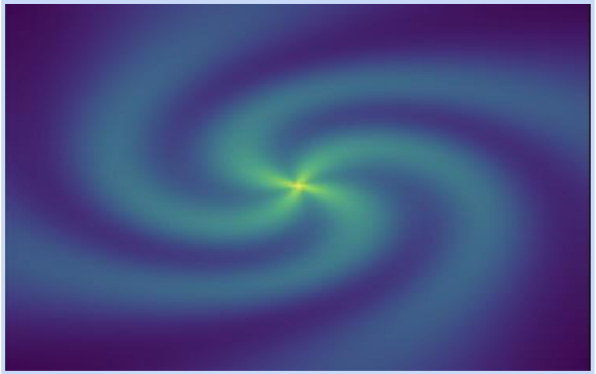
Specified Parameters:
None



Specified parameters:
Noise and Core Size

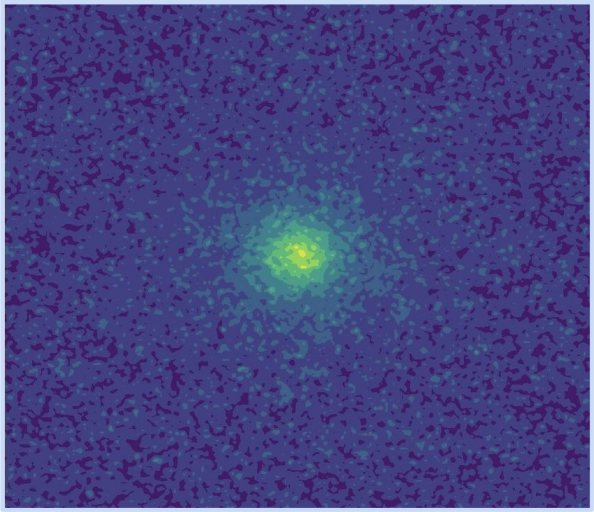


Specified parameters:
Number of Arms, Noise Level, Zoom

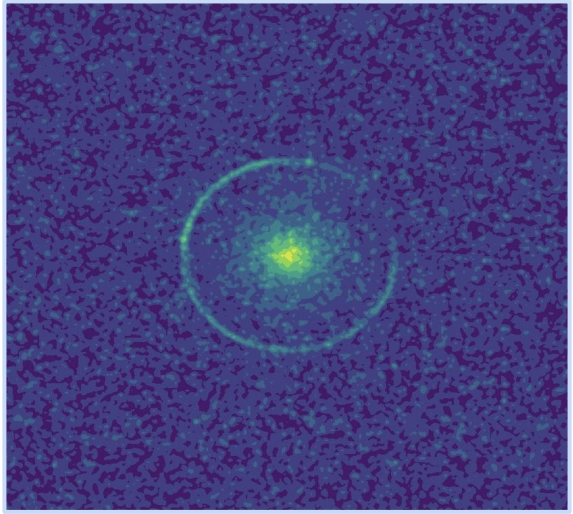


Images

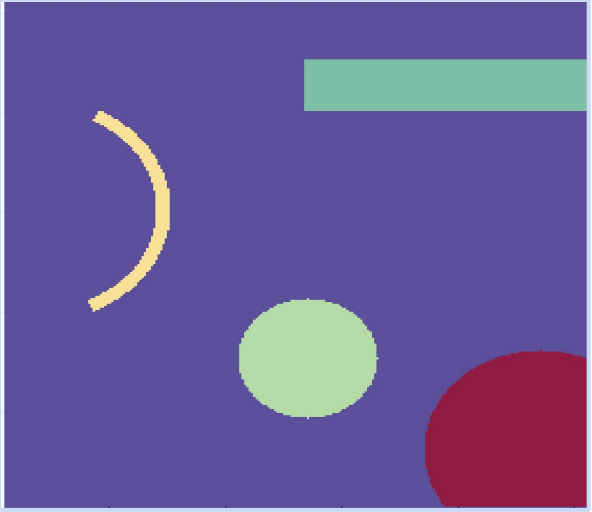
Sky Image



Lens Image

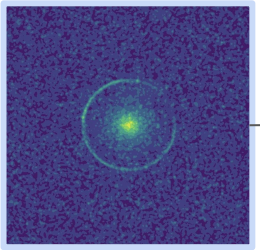


Geometric Image

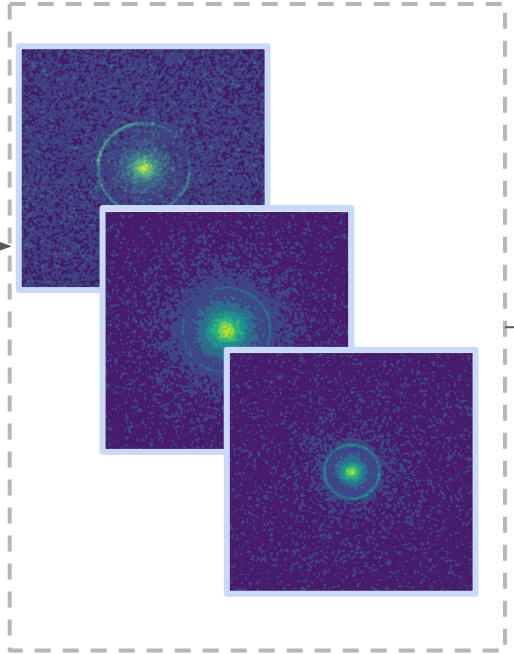


Images, Catalogues, and Collections

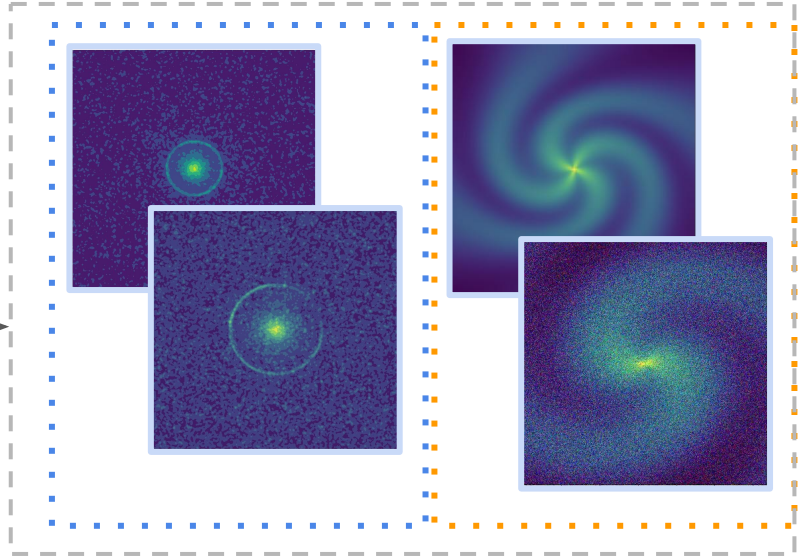
Lensing Image



Lensing Catalogue



Collection



Benchmark Performance Example

Table 1. Summary of results for the benchmark trained ResNet models.

Dataset	Accuracy
Sky Images	95.7
Geometric Images	97.3
Astronomy Objects	97.1

- To the left: Accuracy scores for training 10,000 image datasets generated with DeepBench using Resnet-50 models.

- Scores were recorded for three trials each using collections of different image types:
  - Trial 1: Sky Images
  - Trial 2: Geometric Images
  - Trial 3: Astronomy Objects

Summary

- Motivations for DeepBench:
  - beginner-friendly
  - Faster training convergence
  - Fills gap in benchmark dataset complexity

- Useful features:
  - Astronomical object profile simulations of varying complexity
  - Flexible parameter input requirements
  - Quick creation of benchmark dataset

Questions