

George Stein, Ph.D.

Machine Learning Researcher

 georgestein.github.io  georgestein  george-stein  george.f.stein@gmail.com

SUMMARY

Research scientist with extensive experience developing and applying machine learning methods on high performance computing systems to extract information from large datasets.

SKILLS

PROGRAMMING

Proficient:
Python • $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$

Experienced:
Fortran

Familiar:
C • SQL • CSS • HTML

LIBRARIES/Frameworks

PyTorch • TensorFlow •
PyTorch-Lightning • OpenMP •
MPI • scikit-learn • Pandas

TOOLS/Platforms

Git • Docker • Streamlit

EDUCATION

PH.D. IN ASTROPHYSICS

UNIVERSITY OF TORONTO
Sept. 2014 - Aug. 2019 | Toronto, Canada

B.SC. (HONS.) IN PHYSICS AND ASTRONOMY

UNIVERSITY OF BRITISH COLUMBIA
Sept. 2010 - May 2014 | Vancouver Canada

PUBLICATIONS

Lead author of a variety of publications in machine learning and cosmology which can be found through my ORCID: [0000-0002-5193-516X](https://orcid.org/0000-0002-5193-516X)

EXPERIENCE

LAWRENCE BERKELEY NATIONAL LABORATORY | POSTDOCTORAL SCHOLAR

Sept. 2019 – Current | Berkeley, California

- Compiled a dataset of 72 million galaxy images and utilized multi-node GPU-accelerated systems to develop and train convolutional neural networks for a variety of science targets.
- Applied the latest self-supervised representation learning techniques to improve classification performance of extremely rare objects, more than doubling the efficiency of human labeling efforts compared to previous methods.
- Constructed an interactive similarity search web app to facilitate rapid investigations of prohibitively-large datasets.
- PI of machine learning for sky surveys proposal awarded an allocation of 12k DGX A100 node-hours and 200TB storage on Argonne National Laboratory's flagship compute system.

UNIVERSITY OF CALIFORNIA, BERKELEY | POSTDOCTORAL SCHOLAR

Sept. 2019 – Current | Berkeley, California

- Designed a new anomaly detection method – in-distribution anomaly detection through conditional density estimation – and applied it to win a blind anomaly detection challenge, detecting an anomaly occurring in only 0.08% of 1 million events.
- Created models for spectral timeseries to generate new observations consistent with the data distribution and to solve the inverse problem in the presence of noisy and incomplete data to provide tighter constraints on cosmological measurements.

CANADIAN INSTITUTE FOR THEORETICAL ASTROPHYSICS | GRADUATE RESEARCHER

Sept. 2014 - Aug. 2019 | Toronto, Canada

- Utilized high performance computing systems to construct a pipeline for generating simulated observations of our universe, achieving 100x speed-up compared to previous method.
- Packaged simulation products and integrated them into the workflows of a number of large telescope collaborations for data analysis and testing purposes.

PROJECTS

ML IN COSMOLOGY | GIT

2018 - Current

- Curator of comprehensive archive of machine learning applications to the study of galaxies and cosmology, facilitating a number of cross-disciplinary projects.