George Stein, Ph.D.

Machine Learning Scientist

⌂ georgestein.github.io   ⊙ georgestein   in george-stein   ✉ george.f.stein@gmail.com

Summary

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

Skills

PROGRAMMING

Proficient:
Python

Experienced:
Fortran

Familiar:
C • SQL • CSS • HTML

LIBRARIES/FRAMEWORKS

PyTorch • TensorFlow • PyTorch-Lightning • OpenMP • MPI • Pandas • ++

TOOLS/PLATFORMS

Git • Docker • Streamlit

Education

PHD. 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 7 publications in machine learning and cosmology, and have presented research at 50+ venues. Publication list @ inspirehep.net/authors/1712647

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. | PyTorch

→ Applied the latest self-supervised representation learning techniques to more than double the efficiency of human labeling efforts looking for extremely rare objects.

→ Created and deployed an interactive similarity search web app to facilitate rapid investigations of prohibitively-large datasets. | Streamlit

→ Lead a proposal awarded 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 by detecting and characterizing an anomaly occurring in only 0.08% of 1 million events.

→ Improved the modelling of spectral timeseries to achieve tighter constraints on cosmological measurements by solving the inverse problem in the presence of noisy and incomplete data with a probabilistic autoencoder. | TensorFlow

CANADIAN INSTITUTE FOR THEORETICAL ASTROPHYSICS | Graduate Researcher

Sept. 2014 - Aug. 2019 | Toronto, Canada

→ Constructed a pipeline for generating simulated observations of our universe using high performance computing systems, achieving 100x speed-up. | Fortran

→ Published ready-to-use simulation products and integrated them into the workflows of a number of large telescope collaborations.

→ Teaching assistant for 12 undergraduate-level courses, with course materials ranging from computational astrophysics to the history of astronomy.

Projects

SEGMENTATION OF SATTELLITE IMAGERY

2021 - Current

→ Trained a variety of segmentation models to identify cloud cover in satellite imagery as part of DrivenData's cloud detection challenge, achieving an IoU only 0.0037 behind the winning submission. | Pytorch-Lightning

→ Leveraged public API to increase provided dataset size 10 fold, and designed a custom CloudMix data augmentation that nearly eliminated model overfitting.

ML-IN-COSMOLOGY

2018 - Current

→ Curator of comprehensive archive of machine learning applications to the study of galaxies and cosmology, facilitating cross-disciplinary and academic/industry projects. | Git