George Stein, Ph.D.

Machine Learning Scientist

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

EXPERIENCE

Postdoctoral Scholar, Machine Learning — *Lawrence Berkeley National Laboratory* | Sept. 2019 - Present

- Compiled a dataset of 72 million galaxy images and utilized multi-node GPU-accelerated systems to develop and train convolutional neural networks for a diverse set of science targets. | *PyTorch*
- Utilized self-supervised pre-training to learn from unlabelled data and more than double the efficiency of human labeling efforts for 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.

Postdoctoral Scholar, Machine Learning — *University of California, Berkeley* | Sept. 2019 - Present

- 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.
- Performed data-driven modeling of spectral timeseries with a probabilistic autoencoder to identify anomalous samples and constrain physical parameters of the system. | *TensorFlow*

Graduate Researcher, High Performance Computing — *Canadian Institute for Theoretical Astrophysics* | Sept. 2014 - Aug. 2019

- Constructed a pipeline for generating simulated observations of our universe on high performance computing systems, achieving 100x speed-up, and integrated them into production workflows. | *Fortran*

PROJECTS

Segmentation of Satellite Imagery

- Trained an ensemble 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 APIs to increase provided dataset size 10-fold, and designed a custom set of physically motivated augmentations that nearly eliminated overfitting.

ML-in-cosmology

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

CAREER OBJECTIVE

To adapt and advance state-of-the-art deep learning techniques to solve applied problems at scale.

EDUCATION

University of Toronto
PhD., Astrophysics
Sept. 2014 - Aug. 2019

University of British Columbia
B.Sc. (Hons.), Physics and Astronomy
Sept. 2010 - May. 2014

SKILLS

Proficient:

Python: PyTorch, TensorFlow, PyTorch-Lightning, Scikit-Learn, NumPy, Pandas

OpenMP, MPI, Git

Experienced:

Fortran

Familiar:

C, SQL, CSS, HTML, Streamlit

PUBLICATIONS

Lead author of 7 publications in machine learning and cosmology, and have presented research at 50+ venues.

Publication list and additional projects available at georgestein.github.io.