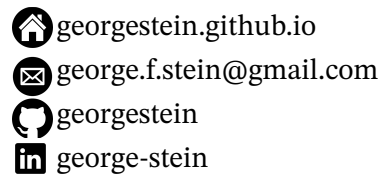


George Stein, Ph.D.

Machine Learning Specialist



SUMMARY

5+ years experience adapting and advancing deep learning techniques to solve applied problems at scale. 10+ years in high performance computing, data analysis, and visualization. Highly adept at extracting insights from large datasets.

PROFESSIONAL EXPERIENCE

Lawrence Berkeley National Laboratory,
Postdoctoral Scholar, Machine Learning

Berkeley, CA
Sept. 2019–Present

- Compiled a dataset of 100 million galaxy images and utilized multi-node GPU-accelerated systems to develop and train convolutional neural networks. | *PyTorch*
- Implemented self-supervised pre-training to learn from unlabelled data and increase the efficiency of human labeling efforts by over 160%.
- Created and deployed an image retrieval [web app](#) built upon 50 million image embeddings. | *Streamlit*
- Principal investigator of data science proposal awarded 12k NVIDIA DGX A100 node-hours and 200TB storage.

University of California, Berkeley,
Postdoctoral Scholar, Machine Learning

Berkeley, CA
Sept. 2019–Present

- Designed a new anomaly detection method 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.
- Built a Neural Network emulator to replace the calculation of costly physical models, achieving 1000x speed-up. | *JAX*
- Developed a deep generative model for high-dimensional time series data and performed probabilistic data reconstruction to constrain physical parameters. | *TensorFlow*

Canadian Institute for Theoretical Astrophysics,
Graduate Researcher, High Performance Computing

Toronto, Canada
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*

SELECTED INDEPENDENT PROJECTS

Segmentation of Satellite Imagery

Nov. 2021–March 2022

- Trained an ensemble of segmentation models to identify cloud cover in satellite imagery as part of a popular data science competition, finishing in the top 3% of 850 participants. | *Pytorch-Lightning*
- Leveraged public APIs to increase public dataset size 10-fold, and designed a custom set of physically motivated augmentations that nearly eliminated overfitting.

ML-in-cosmology

Dec. 2018–Present

- Curator of [a comprehensive archive](#) of ML applications to the study of cosmology, with over 250 stars on GitHub.

EDUCATION

University of Toronto
Ph.D., Astrophysics

Toronto, Canada
Sept. 2014–Aug. 2019

University of British Columbia
B.Sc. (Honours), Physics and Astronomy, with distinction

Vancouver, Canada
Sept. 2010–May 2014

SKILLS

Programming Languages: Python (advanced), Fortran (experienced), C (familiar).

Frameworks/Packages: Proficient in PyTorch, TensorFlow, PyTorch-Lightning, Scikit-Learn, NumPy, Pandas, Jupyter Notebook, Git, and Streamlit. Familiar with JAX, SQL, CSS, HTML, TensorFlow Probability, Pyro, and NumPyro.

Machine Learning Experience: Advanced: Convolutional Neural Networks, Self-supervised learning, Generative modeling, Probabilistic sampling, Dimensionality reduction & Clustering (Autoencoders, TSNE, UMAP, k-NN), etc.

PUBLICATIONS

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

Full publication list and additional projects available at [georgestein.github.io](#)