George Stein, Ph.D.

Machine Learning Researcher

✉ george.f.stein@gmail.com
🏠 georgestein.github.io
⭘ georgestein
in george-stein

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   **Berkeley, CA**
Postdoctoral Scholar, Machine Learning   *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 learning to utilize 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   **Berkeley, CA**
Postdoctoral Scholar, Machine Learning   *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 a costly calculation, achieving 10,000x speed-up. | *JAX*
- Developed a deep generative model for high-dimensional time series data and performed probabilistic data reconstruction to improve the constraints on parameters of interest. | *TensorFlow*

Canadian Institute for Theoretical Astrophysics   **Toronto, ON**
Graduate Researcher, High Performance Computing   *Sept. 2014–Aug. 2019*
- Constructed a pipeline for generating simulated observations of our universe on high performance computing systems, achieving 100x speed-up from previous methods, adopted by numerous forward modeling pipelines. | *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   **Toronto, ON**
Ph.D., Astrophysics   *Sept. 2014–Aug. 2019*

University of British Columbia   **Vancouver, BC**
B.Sc. (Honours), Physics and Astronomy, with distinction   *Sept. 2010–May 2014*

SKILLS

Programming Languages: Python (advanced), Fortran (experienced).
Frameworks/Packages: Proficient in PyTorch, TensorFlow, PyTorch-Lightning, Scikit-Learn, NumPy, Pandas, Jupyter Notebook, Git, Bash, and Streamlit. Familiar with JAX, SQL, CSS, HTML, TensorFlow Probability, Pyro, and NumPyro.
Machine Learning Experience: Computer Vision, Convolutional Neural Networks, Self-supervised learning, Generative modeling, Density estimation, MCMC 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