

Stanley Z. Hua

Data Enthusiast ♦ Toronto, Canada

437-986-3444 | stanley.hua@mail.utoronto.ca | [linkedin](#) | [github](#) | [website](#)

EDUCATION

University of Toronto

Toronto, CA

Honours BS Computer Science Specialist, Statistics Minor (GPA: 3.86/4)

Sept. 2019 – May 2024 (Expected)

TECHNICAL SKILLS

Languages: Python, SQL, Shell Script, Git, Javascript, HTML/CSS, C/C++, Java, R, Assembly

Software Tools: GitHub, AWS, Perforce, Azure ML

Data Skills: Machine Learning, Computer Vision, Time-Series, Data Visualization, Clustering, Dim. Reduction

Python Libraries: Pandas, NumPy, Scikit-Learn, PyTorch, Tensorflow, Matplotlib, Dask, psycopg2, OpenCV, Selenium

EXPERIENCE

Software Engineer Intern

May 2022 – Present

Intel Corporation

Toronto, CA

- Optimized SQL (and Python Pandas) code for an existing dashboard, speeding up page loading time by 400%
- Developed a tool using Python, SQL and shell scripting to estimate runtime of queued benchmarking jobs, to prevent users from abusing the job queue.
- Refactored a core tool to represent a benchmarking experiment, interfacing with the file system and database.
- Created a dashboard using ChartJS, HTML/CSS and Python, to monitor the volume of CRON jobs every hour.

Junior ML Specialist

May. 2022 – Present

The Hospital for Sick Children

Toronto, CA

- Exploring image/video-based self-supervised methods (SimCLR, MoCo) for modeling ultrasound image sequences
- Productionized MLOps code for data validation (quality checks, unit-testing) and model deployment, forecasting the number of patients in the ER hourly.
- Proposed and implemented a CNN in PyTorch to forecast the number of hourly patients in the ER.
- Explored and implemented Bayesian methods for prediction interval estimation (GP, Bayesian NNs)
- Implemented saving/loading of a complex nested model (scikit-learn and PyTorch).

AI Research Student

Sept. 2021 – May. 2022

The Hospital for Sick Children (Goldenberg Lab)

Toronto, CA

- Adapted deep learning methods from video modeling (Conv.Pooling, CNN-LSTM, TSM) to improve prediction of a kidney disease from medical images (ultrasound) taken over multiple hospital visits using PyTorch.
- Assessed model performance based on key metrics (AUROC, AUPRC) with bootstrapped confidence intervals.
- Demonstrated that multi-visit image models do not outperform single-visit models in predicting obstruction.

AI Research Student

May 2021 – Sept. 2021

University of Toronto (Moses Lab)

Toronto, CA

- Created a large-scale dataset **CytoImageNet** (890K images, 894 classes) from 20 TB of open-source microscopy images (and tabular metadata) using Python (pandas, numpy, opencv) and shell scripting.
- Pretrained deep convolutional models (EfficientNetB0) on CytoImageNet using Tensorflow Keras, providing biologists a new means to extract information from microscopy images.
- The CytoImageNet dataset has attracted attention on Kaggle (4879 views, 454 downloads).

PUBLICATIONS

CytoImageNet: A large-scale pretraining dataset for bioimage transfer learning

NeurIPS 2021 (*Learning Meaningful Representations of Life Workshop*)

From Single-Visit to Multi-Visit Image-Based Models: Single-Visit Models are Enough to Predict Obstructive Hydronephrosis

SIPAIM 2022 (*International Symposium on Medical Information Processing and Analysis*)