

Stanley Z. Hua

Toronto, Canada

(+1) 4379863444 | 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. 2022 – June 2024 (Expected)

University of Toronto

Toronto, CA

Honours BS Bioinformatics & Computational Biology Specialist

Sept. 2019 – Sept. 2022 (Switched Majors)

TECHNICAL SKILLS

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

Software Tools: Airflow, AWS, Docker, GitHub, Redis, Figma

Data Skills: Data Engineering, Data Visualization, Machine Learning, Computer Vision, NLP

Python Libraries: Pandas, NumPy, Matplotlib, Dask, Ray, psycogp2

ML Libraries: PyTorch, Tensorflow, Scikit-Learn, MLFlow

CV/NLP Libraries: Transformers, Langchain, Guardrails, nltk, OpenCV

EXPERIENCE

Data Science Intern

June 2023 – Sept. 2023

Wealthsimple

Toronto, CA

- Created a **multi-purpose text labeling pipeline** leveraging LLMs for **training data acquisition**, resulting in significantly lower labeling errors (5%) versus human-annotated labels (20%). Defined requirements, tested various prompting strategies, implemented asynchronous batched calls and added output validation with guardrails.
- Designed, implemented and deployed model to an API endpoint that detects if a user entered invalid account details when initiating an account transfer, potentially saving \$12M in flagged transfers annually.

Junior ML Specialist

May 2022 – Aug. 2023

The Hospital for Sick Children

Toronto, CA

- Showed the potential for (MoCo) **self-supervised image pretraining** to improve **in-domain** and **out-of-distribution generalization** for the task of view labeling on renal **ultrasound videos**.
- Trained and deployed CNN to forecast ER patient volumes, and explored Bayesian methods (GP, Bayesian NNs) for confidence interval estimation.

Software Engineer Intern

May 2022 – May 2023

Intel Corporation

Toronto, CA

- Refactored a **data extraction tool** that allows users to easily retrieve data for benchmarking experiments from the database and file system, resulting in improved efficiency, code architecture and test coverage (99%).
- Optimized **SQL** and Pandas code for existing dashboards, speeding up loading time by 400%.
- Developed dashboards and automated checks, to ensure CRON jobs are spaced apart and benchmarking jobs are not abusing high-priority job queues, increasing stability of jobs and saving cloud compute by +1K hours per month

ML Research Assistant

Sept. 2021 – May 2022

The Hospital for Sick Children (Goldenberg Lab)

Toronto, CA

- Adapted video-based deep learning methods for **feature aggregation** (Conv.Pooling, CNN-LSTM, TSM) to predict if a child needs kidney surgery from ultrasound images taken over multiple hospital visits.
- Demonstrated that ultrasounds from the first hospital visit alone is enough to predict the need for surgery.

ML Research Assistant

May 2021 – Sept. 2021

University of Toronto (Moses Lab)

Toronto, CA

- Created the **CytoImageNet dataset** (890K images, 894 classes) from 20 TB of open-source microscopy images.
- Showed that CytoImageNet-pretrained models are competitive with ImageNet-pretrained models on downstream microscopy datasets, despite weak labels and minimal hyperparameter tuning (compared to ImageNet).
- The CytoImageNet dataset attracted attention on Kaggle (10696 views, 526 downloads).

ML Research Assistant

Jul. 2020 – Jul. 2021

University of Toronto (Tyrrell Lab)

Toronto, CA

- Showed that choice of dimensionality following dim. reduction (PCA, autoencoder) is important for clustering (K-Means, DBSCAN, Agglomerative) of medical images under small sample sizes using Tensorflow.

PUBLICATIONS

Hua SBZ, Lu AX, Moses AM. CytoImageNet: A large-scale pretraining dataset for bioimage transfer learning. NeurIPS Workshop on Learning Meaningful Representations for Life. 2021 Dec.

Hua SBZ, Rickard M, Weaver J, Xiang A, Alvarez D, Veleaz KN, Sheth K, Tasian GE, Lorenzo AJ, Goldenberg A, Erdman L. From Single-Visit to Multi-Visit Image-Based Models: Single-Visit Models are Enough to Predict Obstructive Hydronephrosis. 18th Symposium on Medical Information Processing and Analysis (SIPAIM). 2022 Nov.

CONFERENCE POSTERS

Supervised Contrastive Learning for Improved View Labeling in Pediatric Renal Ultrasound Videos
ISBI 2023 (Cartagena, Colombia)

***Longitudinal Image-Based Prediction of Surgical Intervention in Hydronephrosis Patients: Perhaps Earlier Decision-Making Is Possible!**

ESPU 2023 (Lisbon, Portugal)

*Not presenter

INVITED TALKS

Towards Meaningful Pretraining Data
Models, Inference & Algorithms Seminar, Broad Institute
Boston, USA, October 25th, 2023

HONORS & AWARDS

- | | |
|---------|--|
| 2023 | Samuel Beatty Fund Travel Grant , \$600 (Declined) |
| 2021 | University of Toronto CSB Undergraduate Research Award , \$4000 |
| 2020-22 | Dean's List Award |
| 2021 | The F. M. Hill Scholarship in Biology , \$1100 |

EXTRACURRICULAR ACTIVITIES

- | | |
|---|---|
| Teaching Assistant
<i>University of Toronto – Independent Summer Statistics Community</i> <ul style="list-style-type: none">Organized data science - related workshops for undergraduate statistics students.Presented hands-on workshop introducing the Pandas library. | June 2021 – Aug. 2021
<i>Toronto, CA</i> |
| Club President & Founder
<i>University of Toronto – Offline</i> <ul style="list-style-type: none">Led a team to build an online community of 200+ undergraduate students on Discord during the pandemic.Hosted multiple game nights and movie nights. | Sept. 2021 – Sept. 2022
<i>Toronto, CA</i> |
| Events Coordinator
<i>University of Toronto – Table Tennis Club</i> <ul style="list-style-type: none">Led new initiatives to increase social media presence (sharing table tennis matches, player highlights, infographics) | May 2020 – May 2021
<i>Toronto, CA</i> |
| General Council Member
<i>University of Toronto – Bioinformatics & Computational Biology Student Union</i> <ul style="list-style-type: none">Assisted in planning of online social activities for bioinformatics undergraduate students. | Sept. 2020 – May 2021
<i>Toronto, CA</i> |
| Volunteer Chess Teacher
<i>Kensington Health – Second Mile Club</i> <ul style="list-style-type: none">Organized, promoted and hosted beginner chess workshops for seniors. | Aug. 2020 – Aug. 2020
<i>Toronto, CA</i> |
| Volunteer Photographer
<i>West Neighborhood House</i> <ul style="list-style-type: none">Took photographs and videos to help promote the not-for-profit's community services. | Feb. 2020 – Feb. 2020
<i>Toronto, CA</i> |