# 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, psycopg2
**ML Libraries**: PyTorch, Tensorflow, Scikit-Learn, MLFlow
**CV/NLP Libraries**: Transformers, Langchain, Guardrails, nltk, OpenCV

## Research Experience

**ML Research Assistant** — Jan. 2024 – Present
*University of Toronto* (Rahul Krishnan Lab) — *Toronto, CA*

- Investigating multi-agent prompting methods to improve zero-shot performance of large multi-modal models for rare diseases.

**Junior ML Specialist** — May 2022 – Aug. 2023
*The Hospital for Sick Children* — *Toronto, CA*

- Demonstrated that (MoCo) **supervised contrastive image pre-training** can improve **in-domain** and **out-of-distribution generalization** for automatic 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.

**ML Research Assistant** — Sept. 2021 – May 2022
*The Hospital for Sick Children* (Anna 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* (Alan 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* (Pascal 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.

## Industry Experience

**Data Science Intern** June 2023 – Sept. 2023
*Wealthsimple* *Toronto, CA*

- Created a **multi-purpose labeling engine for text data**; to perform **zero-shot classification** with LLMs, resulting in significantly lower labeling errors (5%) versus human-annotated labels (20%) for ticket routing. Communicated with stakeholders to define requirements, implemented asynchronous batched calls and automated 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.

**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

## 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, Velear 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** (with Alex Lu and Alexander Lin)
Models, Inference & Algorithms Seminar, Broad Institute
Boston, USA, October 25th, 2023

## Honors & Awards

2023 **Samuel Beatty Fund Travel Grant**, $600

2021 **University of Toronto CSB Undergraduate Research Award**, $4000

2020-22 **Dean's List Award**

2021 **The F. M. Hill Scholarship in Biology**, $1100