# XIAORUI HUANG

## Always Fascinated


Availability: **From May 2024**
Preferred Name: **Richard**
richardxr.huang@mail.utoronto.ca
+1 (289) 772–8682 Toronto, Canada
xiaorui-richard-huang Xiaorui-Huang


## EXPERIENCE

### eAI Machine Learning Engineer

**Qualcomm**

May 2023 — August 2023 Markham, ON

- Led efforts on **Neural Architecture Search (NAS)** and model compression within the **Edge AI (eAI)** R&D team.
- Developed a NAS framework, leveraging Qualcomm's patented NAS techniques, to optimize **arbitrary models**[1] for **any profiled hardware**, harnessing Pytorch's **torch.fx** extensively.
- Streamlined the NAS workflow for incoming client models, slashing **engineering time** by **80%**.
- Achieved a **50%** reduction in **model size** and a **60%** drop in **inference latency** without compromising accuracy across benchmark models[2].
- Engaged in lab meetings focused on cutting-edge model compression research, particularly **Quantization**.
- Delivered a comprehensive presentation on the NAS framework to the broader eAI team.

NAS | Quantization | torch.fx | Pytorch | ONNX | R&D

### RPA Backend Developer

**IBM**

May 2022 — April 2023 Markham, ON

- Worked on backend development for IBM's Robotics Process Automation (RPA) platform.
- Augmented IBM RPA's *WAL* programming language, introducing a reflection feature resembling Java and **C#**.
- Collaborated with cross-functional teams, achieving a **15%** reduction in customer issues and defects per release.
- Employed **agile** methodologies, showed both independent and collaborative competencies in a hybrid environment.
- Articulated and presented solution strategies to RPA's senior architects and product teams.

C# | Pragramming Language Design | Agile | Visual Studio

## EDUCATION

### University of Toronto

**Candidate for B.Sc. in Computer Science**

2019 — Now (Exp 2024) Toronto, ON

**Relevant Courses**

- **CSC317 Computer Graphics (97%)** Ray Tracing, Mass Spring Systems, Bounding Volume Hierarchy, Meshes, Kinematics, OpenGL Shaders in **C++** using **Eigen** and **libigl**
- **CSC413 Deep Learning (96%)** Transformers, CNN, RNN, GAN, VAE, GNN, RL. **original research** on optimization strategy as final course project. RolandGao/pycls

C++ | Pytorch | Linear Algebra | Algorithms | Stats & Probablity

[1] NAS support is required for NN layers *E.g. nn.Conv2d is supported*
[2] Results vary; models include MobileNetV2, ResNet50, BERT

## RESEARCH

### Linearly Explored Learning Rate Scheduler (LES)

Apr 2022 RolandGao/pycls

- We introduced the LES method to automate and refine the resource-intensive task of **learning rate tuning**.
- LES achieves a final error rate of 8% on par with other commonly used optimizer and schedulers on pycls code base **without the need for learning rate tuning**.
- Developed a custom **SGD with momentum** algorithm to facilitate exploration of various backpropagation strategies during LES creation.

## PROJECTS

### CUDA Ray Tracing

**Almost Real Time Ray Tracing**

November 2023 Xiaorui-Huang/cuda-ray-tracing

- Implemented a **CUDA** ray tracer with **BVH** acceleration structure, with **Blinn-Phong** shading.
- Achieved **almost real-time** rendering of **1000+** triangles with **30** FPS.
- Incorporated dynamically loaded Scene generation to allow for future interactivity.

CUDA | C/C++ | CMake

### Woodoku Learn

**Reinforcement Learning Model**

Jul 2022 EdwardHaoranLee/WoodokuLearn

- Replicated the mobile game Woodoku for CLI using **Python**, enabling both human and AI gameplay through dedicated environment APIs.
- Employed Q-Learning, a **Reinforcement Learning** approach with Pytorch, targeting top scores on the Woodoku leaderboard.
- Adhered to **agile** methodologies; integrated CI testing, static type checks, and employed tools like GitHub Actions, pytest, and mypy for efficient code reviews and development.

Pytorch | OOP | Agile

## SKILLS

**Programming Languages**

Python | CUDA | C/C++ | Rust | C# | Java | LaTeX | PowerShell | R | TypeScript | HTML&CSS | SQL

**Other Frameworks & Development Environments**

Pytorch | torch.fx | Docker | WSL | git | Vim | VSCode

*Idiomatic in English and in Mandarin Chinese*