XIAORUI HUANG

Solve all that fascinate me

Availability: From July 2023
richardxr.huang@mail.utoronto.ca
+1 (289) 772–8682 Toronto, Canada
xiaorui-richard-huang Xiaorui-Huang

EDUCATION

University of Toronto

Candidate for B.Sc. in Computer Science

2019 — Now (Exp 2024) Toronto, ON

Relevant Courses

- **CSC413 Deep Learning — 96%** Pytorch, Language Models, CNN, Interpretability, Optimization, RNN & Attetion, Transformers, GAN, VAE, GNN, Q-Learning. Conducted **original research** on optimization strategy as final course project. RolandGao/pycls
- **CSC317 Computer Graphics — 97%** Ray Tracing, Mass Spring Systems, Bounding Volume Hierarchy, Meshes processing, Forward & Inverse Kinematics, Shader Pipeline with OpenGL Shading Language. Implemented graphics algorithms using **C++** assisted by **Eigen** and **libigl** libraries.

Other Courses

Linear Algebra, Statistic and Probability, Multivariate Calculus, Algorithm Design and Complexity

EXPERIENCE

Machine Learning Engineer

Qualcomm

May 2023 — August 2023 Markham, ON

- Machine Learning Engineer working on **Neural Architecture Search (NAS)** and as apart of the **edge AI (eAI)** R&D team.
- Designed and implemented a NAS framework[1] for optimizing **Any generic model**, against **Any generic hardware**.
- Employed PyTorch and torch.fx extensively to design and optimize various machine learning models.
- Facilitated weekly lab meetings, aligning the team on research directions, findings, and breakthroughs.
- Acquired a robust understanding of model quantization, ensuring optimized model deployment on edge devices.
- Collaborated closely with interdisciplinary teams, ensuring alignment of machine learning initiatives with overarching company goals.
- Assisted in the mentoring and onboarding of junior engineers, fostering a culture of continuous learning and knowledge sharing.
- Demonstrated proactive problem-solving by identifying potential areas of improvement in the NAS workflow and proposing actionable solutions.
- Kept abreast of the latest developments in the machine learning space, ensuring Qualcomm remains at the forefront of technological advancements.
- Presented research findings and model improvements to senior management, influencing the strategic direction of machine learning projects.

[1] NAS framework for automating internal researched NAS method

PROJECTS

Woodoku Learn

Reinforced Learning Model

Jul 2022 EdwardHaoranLee/WoodokuLearn

- Designed and implemented the mobile game Woodoku in the CLI using python.
- Woodoku could be played by human users and by a game-playing algorithm through game environment APIs.
- Actively implementing the game-playing algorithm using Q-Learning, a **Reinforce Learning** algorithm, with Pytorch to beat the Woodoku leader board.
- Practiced **agile** software development, CI testing and static type checking, achieved through GitHub Action, GitHub Pull Request and code review process, pytest, mypy.

python Pytorch mypy Agile

Eedimator

Online course performance predictor

Nov 2021 Xiaorui-Huang/Eedimator

- A **predictor of students' ability to answer questions**, based on previous answers and other students' answers allowing online education platforms to provide tailored assistance.
- Used Machine Learning algorithms such as **Neural Networks, Matrix Factorization, Item Response Theory, and K-NN** to create an **ensembled** prediction model.
- Based on **NeurIPS** 2020 Education Challenge and uses real-world data collected on eedi.com

Python Pytorch numpy scikit-learn

Boomba — Run-away Alarm

New Hacks 2020 — *Hackathon 2nd Place Overall*

March 2020 Boomba on devpost.com

- A moving alarm secured with a puzzle to snooze, built with **Arduino** and **Raspberry Pi**.
- The user is required to chase down the alarm, solve puzzles, then give the correct voice commands to snooze the alarm.
- Configured **Google Speech-to-Text API**, developed voice command feature in python and implemented motor and puzzle logic in C++.

C++ Python Arduino Raspberry Pi Google Cloud API

SKILLS

Programming Languages

Python C/C++ C# Java TypeScript HTML&CSS PowerShell Bash Scripts SQL

Other Frameworks & Development Environments

Pytorch React Vue.js Django MongoDB Express.js tailwindcss Vim VSCode WSL

Idiomatic in English and in Mandarin Chinese

RPA Backend Developer Intern

IBM

May 2022 — April 2023 Markham, ON

- Backend software development working on IBM Robotics Process Automation (RPA).
- Increased IBM RPA's *WAL* programming language usability by developing reflection feature similar to Java and C#.
- Reduced 10% customer issues and product defects per release through collaboration with multi-disciplined teams.
- Conducted **agile** development process, demonstrated independent and teamwork ability through both remote and onsite work environment.
- Presented and communicated novel solutions to RPA senior architect and product teams.

C# Visual Studio Agile

RESEARCH

Linearly Explored Learning Rate Scheduler

Apr 2022 RolandGao/pycls

- We proposed the Linearly Explored Learning Rate Scheduler (LES) to help automate and optimize the resource and time consuming task of **learning rate tuning**.
- LES achieves a final error rate of 8% compared to other commonly used optimizer and schedulers on pycls code base **without the need for learning rate tuning**.
- Implemented and tested custom **SGD with momentum** algorithm for LES, allowing the team to explore different backpropagation approaches when developing LES.
- Implemented visualization tools and authored parts of experiment and discussion section.