

# Guoyao Li

Pittsburgh, PA | +1-412-980-3874 | guoyaol@cs.cmu.edu | linkedin.com/in/guoyaoli

## EDUCATION

**Carnegie Mellon University** – School of Computer Science

**Pittsburgh, PA**

*Master of Computational Data Science*

*Dec. 2023*

- **Current GPA** 4.33/4

**Zhejiang University** – Chu Kochen Honors College

**Hangzhou, China**

*Bachelor of Engineering in Computer Science and Technology*

*June 2022*

- **Overall GPA** 3.95/4     **Rank:** top 5% among 269     |     **Minor in Entrepreneurship Management**

- **Core Courses:** Advanced Data Structures & Algorithm Analysis(92/100), Probability and Mathematical Statistics(97/100), Computer Networks(93/100), Computer Architecture(97/100), Operating System(93/100), etc.

## PUBLICATION

**Guoyao Li**, Shahbaz Rezaci, Xin Liu “User-level Membership Inference Attack against Metric Embedding Learning”

ICLR 2022 workshop

## EXPERIENCE

**Catalyst Group at CMU**

*Research Software Development Engineer*     *Supervisor: Prof. Tianqi Chen*

*Feb. 2023 – Present*

- Learning about Apache TVM, the pioneering machine learning compiler

**Intel Corporation**

*Deep Learning Software Intern*

*Apr. 2022 – June 2022*

- Contributed code to BigDL, a distributed machine learning system; Developed two new features for BigDL Nano AutoML
- Implemented experiment on Nano AutoML; Re-implemented a recommender system (Wide&Deep network) with Nano AutoML TensorFlow API
- Optimized the experiment and achieved SOTA performance in company; Achieved ~0.93 AUC and ~2.5x acceleration/core, outperforming SigOpt's ~0.89 AUC
- Developed Nano HPO demo notebooks with high performance and user-friendly visualization

**University of California, Davis**

*Research Intern (Machine Learning)*

*Supervisor: Prof. Xin Liu*

*July 2021 – Feb. 2022*

- Proposed a novel user-level membership inference attack (MIA) on metric embedding learning; Published as 1st author
- Addressed the limitation: existing MIAs unrealistically assume exact training samples are available at inference time
- Utilized compactness of cluster in latent space to solve existing MIAs' inability in metric embedding learning
- Implemented experiment with ~10k lines of code in Python, with PyTorch, sklearn, NumPy, pandas, torchvision
- Achieved SOTA performance (up to 21.60% accuracy improvement) on person re-identification datasets, including Market-1501 and PRID 2011, where user-level MI attack is of paramount importance

## PROJECTS (SELECTED)

**Web Development: IoT Device Platform** | Java, JavaScript, React, SpringBoot, MySQL

*Apr. 2021 – July 2021*

- Developed an IoT device management website based on React, SpringBoot, and MySQL, supporting MQTT messages
- Implemented the registration and management of devices, user registration, login, API encryption, and visual dashboard

**Compiler: TinyC Compiler** | C, llvm, Flex, Bison

*Apr. 2021 – June 2021*

- Designed and developed a compiler supporting a subset of C language, with Flex+Bison frontend and the LLVM backend
- Implemented efficient Lexical analysis, syntax analysis, semantic analysis, and intermediate code generation

**Computer Graphics: Carrier & Aircraft Combat Game** | C++, OpenGL

*Mar. 2021 – June 2021*

- Developed the game of manipulating 3D aircraft firing and combating against carriers and other aircrafts based on OpenGL
- Implemented particle system, collision detection, Phong-model light, zoom in/out/pan/orbit, material, texture with C++

**Database System: PySQL** | Python, SQL

*Apr. 2020 – July 2020*

- Designed and developed a mini database engine in Python, supporting SQL language and all basic database operations including create/drop tables, create/drop indexes, insert/delete/select table records

## SKILLS

**Languages:** Python, C/C++, Java, SQL, JavaScript, Verilog

**Libraries and Frameworks:** PyTorch, sklearn, TensorFlow, NumPy, pandas, QT, OpenGL, React, Spring Boot

**Tools:** Git, LaTeX, Shell Scripting, Linux/Unix Command-line, MATLAB