YUNAN ZHANG

yzhan828@ucr.edu | (760)583-6982 | www.linkedin.com/in/yunandrewzhang
Programming: Python, C/C++, CUDA C/C++, Java, SQL, Verilog, R
OS/software/tool: Unix/Linux, Windows, VS Code, Git/GitHub, Jupyter, Latex, AWS, Nsight Compute, Quartus/ModelSim

---

EDUCATION

University of California, Riverside   *Sep 2020 - present*
Ph.D. | Electrical and Computer Engineering | GPA: 4.0/4.0

University of California, San Diego   *Sep 2018 - Apr 2020*
Bachelor of Science | Computer Science | GPA: 3.86/4.00

PROFESSIONAL EXPERIENCES

Software Engineering Intern | Google - Mountain View, CA   *June 2022 - Sep 2022*

- Collaborated cross team with (1) infrastructure team to set up experimental workflow on search result evaluation and (2) machine learning team to improve document filtering/ranking using pre-trained curves to best meet MVP requirement.
- Improved search indexing for semi-structured web data using both horizontal and vertical-specific enrichment and extraction using C++, customized result evaluation process that aims at collecting user feedback on search experience.
- Optimized ranking of an emerging search vertical in Google Search by deploying Generalized Additive Models using Python via Collab, targeted training objectives including AUROC, MSE, information satisfaction scores, etc.
- Sanitized and filtered training queries using an internal tool that annotates language-independent POS on natural language input.
- Analyzed and filtered thousands of pre-computed document features using Spearman's rank correlation, variance thresholding, etc.

Embedded Software Engineer Intern | Sleepnumber Labs - San Jose, CA   *July 2021- Sep 2021*

- Investigated existing product architecture of an embedded IoT system, contributed to optimizing features developed cross-team.
- Improved device-cloud communication interface by integrating a more efficient protocol onto the product's existing features.
- Deployed Mosquitto MQTT broker on an AWS EC2 instance with SSL/TLS enabled.
- Extended and built the OS image that supports MQTT device C++ SDK using Yocto project and Bitbake.
- Integrated AWS-IoT core MQTT to existing embedded software that is able to send Redis data to the cloud and responses to RPC using C++ and python, which potentially saves 10X cost on cloud services.

Research Assistant | Extreme Storage and Computer Architecture Lab - Riverside, CA   *Jun 2020 - present*

- Designed and developed algorithms for matrix-based problems to efficiently use modern AI/ML accelerators.
- Revisited and re-designed a set of algorithms working on high-dimensional input datasets to map compute kernels into matrix algebras and enable the potential acceleration from modern AI/ML accelerators.
- Developed a framework that emulates the performance gain of algorithms by matrix algebra on NVIDIA's GPUs with Tensor Cores using CUDA.
- Developed a modern GPUs benchmark, which enables efficient use of less-precise floating point formats to accelerate compute kernels by 1.2X using C/CUDA.
- Published paper *$SIMD^2$: a generalized matrix instruction set for accelerating tensor computation beyond GEMM* on ISCA'22 as first author.
- Developed a python interface with Cython, supported native CUDA libraries including cuBlas, Thrust. Provided an easy-to-use GPU programming framework that reduces supported GPU API calls to less than 5 lines of python code. Achieved an average of 2X speedup over C-python APIs.

SELECTED PROJECTS

Resource container - operating system project   Jan 2021 - Apr 2021

- Designed and developed a Linux kernel module that implements resource containers to isolate and schedule threads in C.
- Implemented features to support concurrent thread execution including data sharing, locking/unlocking in the resource container module and verified the correctness of the module with up to 256 threads in a shared 1GB memory space.

MIPS ISA Processor - Competition Project   *Jun 2019 - Aug 2019*

- Designed and developed a 5-stages pipelined processor that supports MIPS32 ISA using Verilog on Quartus and ModelSim.
- Achieved 3X over the baseline and ranked 3rd in the class-wide competition by Implementing data forwarding and branch predictors in the pipelined processor.

JAVA web server - network projects   *Jan 2020 - Apr 2020*

- Designed and implemented a web HTML server using Java socket that supports up to 500 requests using JAVA.
- Implemented an apache xmlrpc server that supports video streaming with self-designed ABR that eliminates re-buffering, and Hash-based remote file systems that support time-based version synchronization.

Hacking experiments - a serial of hacking tasks   *Jan 2020 - Apr 2020*

- Experienced well-designed tasks that involved buffer overwrites, side-channel attacks, SQL injection, network attacks, and cryptography.

Kaggle competition on recommendation system   *Sept 2019 - Oct 2019*

- Predicted whether a user would read a book using features including Jaccard similarity, cosine similarity, and the average rating for books on a collaborative filter using python. Achieved 0.75 accuracy and placed 26th out of 847 on the leaderboard.
- Predicted the category of a book based on its review by applying a mixed vectorizer using TF-IDF and hashing with feature union. Classified data with Linear SVM which achieved 0.8 accuracy and placed 13th out of 335 on the leaderboard.

Business closure prediction model - recommender system research project   *Sept 2019 - Dec 2019*

- Applied inverse weights justification to resolve an imbalance dataset that reduces the false-positive rate.
- Developed a linear-SVM-based predictor using python that automatically extracts features from customer reviews with TF-IDF vectorizer, and achieved 0.85 accuracy.