Xiao Song

xiaosong9905@gmail.com | +1 530 302 6564 (US) | +86 186 0059 6037 (CHN)
https://www.linkedin.com/in/xiaosx/ | https://xiaosong9905.github.io

Education:

University of California - Berkeley 09/2021 - 05/2022
GPA : 3.85 / 4.0 | Master of Engineering in Electronic Engineering and Computer Science
Courses: Parallel Computing, ML-Sys, AR/VR, Computer Graphic, ML Model Analysis.
Honors: Fung Excellent Scholar

University of Michigan - Ann Arbor 09/2018 - 05/2021
GPA : 3.984 / 4.0 | Bachelor of Science in Computer Science, Minor in Statistics
Courses: Data Structure & Algorithms, Operating System, Search Engine, Database Management System, Web System, Computer Organization, Machine Learning, Data Mining, App Dev, Computer Vision.
Honors: University Honors (Fall 2018, Winter&Fall 2019, Winter 2020), James B. Angell Scholar (2020, 2021)

Skills:

Proficient: C++, CUDA, Arm Intrinsic (NEON), Intel Intrinsic(AVX2, AVX512), OpenMP, MPI
Experienced: Python, PyTorch

Internship:

YITU | High Performance Computation Intern 05/2021 - 08/2021
- Use arm intrinsic (NEON) and NVIDIA CUDA to improve the speed and accuracy of general & customized operators (functions).
- Some common operators are 20-50% faster. Some uint8_t operators are nearly as accurate as double, with only 0.8% of values differing by 1.

Project:

DGEMM on Cori KNL Node (C++, AVX2, AVX512) 05/2022
https://github.com/XiaoSong9905/dgemm-knl
- Implement DGEMM on Intel Cori Node. Achieve average of 75% MKL performance on single core.
- Use AVX512 inline assembly for embedded broadcast, increase theoratical peak from 22.4 to 44.8 GFLOPs.
- Mix use of AVX2 & AVX512 for 8x8 matrix transpose, reduce the pressure on load port. Incr 0.4% peak perf.

Search Engine from Scratch (C++) 02/2021 - 04/2021
- Mainly responsible for search engine back-end: webpage crawler and HTML parser.
- Implement our STL string, vector, map, priority queue.
- Use pthread, OpenSSL, and socket to implement multi-machine multi-thread crawler that supports download prioritized web pages, remove duplicated web pages, handling URL redirection, and support IPv4 & IPv6 download at the same time.
- Implement HTML parser to extract URL links, anchor text, title, and body from HTML. Our parser can handle more corner cases than Python 3 html.parser.

PicassoXS: IOS Photo Editing APP that Change Photo to Painting (Python, Tensorflow) 01/2020 - 04/2020
https://github.com/DynamicProgrammingEECS441/PicassoXS
- Mainly responsible for the back-end CNN model building and deployment, backend server, algorithm research and implementation.
- Algorithm: compared various style transfer algorithms. Combined ideas from multiple algorithms to deal with the trade-offs between algorithm visual effect, computing time, and hardware consumption. The final model can process 512^2 image on 8 CPU instance within 4s.
- Back end: Use tensorflow to build CNN model, train the model, and deploy the model service to Google Cloud through TF server, Docker and K8S. Use Python, Flash to implement RESTful API to receive IOS requests, call corresponding model services, and return processed photos to IOS.