Xiang Li

765-637-5476 li2068@purdue.edu xiangli2410 mrxiangli.github.io

Education

Purdue University, West Lafayette, IN

Ph.D in Computer Engineering (GPA: 3.80 / 4.00) — *Jan 2027*

B.S in Computer Engineering (GPA: 3.89 / 4.00) — *May 2019*

Technical skills

Programming Languages: Python • C • Bash • Assembly • Embedded C • Julia • Javascript
Framework/Library: Pytorch • Tensorflow • Simulink • WebRTC
Related Skills: Nvidia Jetson • LLM inference optimization • Distributed System• CAN • Model-based software

Work Experience

Houston Methodist Research Institute — May 2023 - Aug 2023

Machine Learning Research Intern — *Houston, TX*

- Designed a deep-learning network for brain tissue segmentation of non-contrast CT scans with 87% accuracy to help medical professionals quickly isolate the pathogen area.
- Integrated group equivariant convolution into nnUNet to further improve the segmentation accuracy.
- Optimized segmentation pipeline into a self-contained IoT device and deployed for clinical trials.

Cummins.Inc — Jun 2019 – Aug 2021

Software Engineer / Control Engineer — *Columbus, IN*

- Designed control software interface APIs for Cummins ECM and Bosch fuel systems running on RAM trucks.
- Enhanced cybersecurity software of ECM by adding 2 different encryptions to the ECM bootloader to prevent unauthorized tempering of Cummins products.
- Created model-based control software with Simulink for Cummins owned components equipped on around 250,000 newly manufactured vehicles annually such as glow plug, grill shutter, and grid heater.
- Tuned more than 30 calibration parameters of OBD control software in the RAM truck to ensure proper functionality of the OBD system.
- Integrated transmission software with ECM control software from multiple teams and performed regression tests on software. Analyze test results pertain to FMEA procedure to diagnose root causes of software failure.

Research Experience

Graduate Research Assistant | *Oct 2021 - Present — Dependable Computing System Lab*

- Developing latency sensitive, memory efficient, high throughput serving system for LLM models on large scale GPU clusters.
- Developed deep reinforcement learning based bandwidth prediction framework to achieve low-latency video analytics for real-time communication system, i.e. WebRTC.
- Deployed Cisco Wifi 6E testbed (switch, controller, AP) and performed measurement based experiment to compare Wifi6/5G in smart manufacturing setting.
- Proposed a novel multi-object, real-time tracking heuristic on resource-constrained devices that achieve tracking accuracy up to 84%. The tracker can be easily adapted into state of the arts detectors and achieve state of the art accuracy on MOT test sequences on embedded devices (Jetson TX2/AGX)
- Developed AI-based image processing pipeline for computational tomography and MRI scans, which are dedicated for clinical use.

Publications

X. Li, Chen, C., Lou, Y.Y., Abdallah, M., Kim, K.T. and Bagchi, S., 2024. HopTrack: A Real-time Multi-Object Tracking System for Embedded Devices. arXiv preprint arXiv:2411.00608

X. Li, M. Abdallah, S. Suryavansh, M. Chiang, K. T. Kim and S. Bagchi, "DAG-based Task Orchestration for Edge Computing," 2022 41st International Symposium on Reliable Distributed Systems (SRDS), Vienna, Austria, 2022, pp. 23-34.