Ge Shi

Davis, CA (413) 425-3452 geshi@ucdavis.edu
Linkedin **Github** **Personal Page**

Education

University of California, Davis Davis, CA
Ph.D. candidate in Computer Science; GPA: 3.8/4.0 *Expected: June. 2024*

University of Massachusetts, Amherst Amherst, MA
Master of Science in Computer Science; GPA: 3.9/4.0 *May. 2019*

Zhejiang University Hangzhou, China
Bachelor of Engineering in Automation; GPA: 3.7/4.0 *July. 2017*

Programming Skills

Coding: Python, C/C++, Java, Matlab, SQL, Javascript, R, D3.js, Node.js, Flask, TensorFlow, Pytorch, Hugging Face.
Tools: LangChain, DGL, Gym, OpenCV, Robotics(ROS), AWS EC2, Latex, Git, DataGrip, Docker.

Projects

LLM-related Projects Davis, CA
Self-driven *Feb. 2024 - present*

- **LangChain Agent Development**: Building a service that assists job hunting using **LangChain**. The agent is able to extract structured output from job description and adaptively generate customized cover letter.
- **Parameter-Efficient-Finetuning (PEFT)**: Finetuning LLaMA2 with LoRA method on alpaca from llama-recipes which is able to answer questions following instructions.

Research Projects Davis, CA
Supervisor: Prof. Ian Davidson *June. 2020 - present*

- **Data Mining (KDD)**: Proposed a model-agnostic framework to aggregate post-hoc local explanations in a hierarchical way on various types of data including text (BERT, NLP), images (ResNet, CV), and tabular data (MLP).
- **Scientific ML (IJCAI)**: Benchmarked the robustness of a wide range of popular deep learning architectures in identifying irrelevant features on both time-series data and image data and improve their performances through bivariate optimization.
- **Visual Analytics (TVCG)**: Created a visualization tool (Losslens) for ML model diagnostics. Primarily worked on the back-end implementation of pyHessian loss landscapes and mode-connectivity on ViT and UNet.
- **Computer Vision (KDD)**: Studied the problem of auto-prognosis with deep learning. Proposed a novel multi-view multi-instance learning scheme to solve the data imbalance problem and reach SOTA acc 75.6%.
- **Generative AI**: Proposed a weighting method that leverages synthetic data generated by **GANs** for data augmentation to improve the fairness by 2.7% on CelebA dataset while maintaining the generalization error.

Software Projects Amherst, MA
Selected Internship and Academic Projects *Sept. 2017 - May. 2019*

- **Distributed System**: Designed and implemented a two-tier online book store called "Pygmy" possessing FrontEnd, Order, and Catalog servers with the **Flask** which achieves concurrency of access to database through Python threads.
- **Model Inference Processes**: Collected log traces of human-driven processes and managed them into **Node.js** network type using **JavaScript**; Customized Synoptic to infer an FSM model that captures all the variations in the traces.
- **D3 Visualization (*gist repo*)** : Create a set of interactive visualizations including point cloud, networks, heat maps, radial chart with Data-Driven Documents (**D3**), **Javascript** and **HTML**.
- **Web Scraping**: Developed python **web crawler** to extract over 600k user feedbacks on Wildfire Game from the public online game forums. Analyzed the demographic information and interests of potential customers with **SQL** and **R**.

Robotics Projects Hangzhou, China
ZJU Robotics Lab *Jan. 2016 - June. 2017*

- **Simultaneous Localization and Mapping (SLAM)**: Implemented the fusion of stereo camera and infrared sensor to create 3D points cloud in **C++**; Improved the performance of DSO of in the in-door setting with a smaller accumulated error curve.
- **Humanoid RoboCup Soccer Competetion**: Took charge of the simulation of actuator controlling with Matlab, and migrate the code base to **ROS** system for standarization. We won the 2nd Place in Robocup Kidsize Humanoid League 2016.

Internships

Lawrence Berkeley National Laboratory Berkeley, USA
Research Assistant in Michael Mahoney's Lab *Apr. 2023 - Mar. 2024*

- **Loss Landscapes Visualization**: Implemented the backend of a loss landscape visualization tool with OOP principals that aids ML practitioners in explaining, diagnosing, and validating deep learning models.
- **CRF-transplant on Image Segmentation**: Proposed and validated an training and concatenation algorithm using CRF that's able to stably improve 98% of all backbone networks with hundreds of different hyperparameters.