

YEFAN ZHOU

yefan.zhou.gr@dartmouth.edu | Homepage [↗](#) | Google Scholar [↗](#) | LinkedIn [↗](#) | Hanover, NH, 03755 | 510-809-5378

EDUCATION

Dartmouth College

Ph.D. student in Computer Science

Advisor: Prof. Yaoqing Yang

Research Area: Trustworthy Machine Learning

Hanover, NH

Sep. 2023 – present

University of California, Berkeley

M.Eng in Electrical Engineering and Computer Science; Major GPA: 4.0/4.0

Advisor: Prof. Michael Mahoney

Research Area: Efficient Deep Learning

Berkeley, CA

Aug. 2021 – Dec. 2022

University of California, Berkeley

Exchange Student; GPA: 4.0/4.0

Berkeley, CA

Jan. 2019 – May. 2019

Southeast University

B.Eng in Information Engineering; GPA: 3.7/4

China

Aug. 2016 – Jun. 2020

SERVICES

Reviewers: ICLR 2024, CPAL 2024, NeurIPS 2023, IROS 2022

PUBLICATION

- **Y. Zhou***, T. Pang*, K. Liu, C. H. Martin, M. W. Mahoney, Y. Yang “Temperature Balancing, Layer-wise Weight Analysis, and Neural Network Training” *Thirty-seventh Conference on Neural Information Processing Systems* (NeurIPS 2023 Spotlight)
- **Y. Zhou**, Y. Yang, A. Chang, M. W. Mahoney “A Three-regime model of Network Pruning” *2023 International Conference on Machine Learning (ICML 2023)* [↗](#)
- **Y. Zhou**, Y. Shen, Y. Yan, C. Feng, Y. Yang “A Dataset-Dispersion Perspective on Reconstruction Versus Recognition in Single-View 3D Reconstruction Networks.” *2021 International Conference on 3D Vision (3DV 2021)* [↗](#)
- X. Zhu, **Y. Zhou**, Y. Fan, J. Chen, M. Tomizuka “Learn to Grasp with Less Supervision: A Data-Efficient Maximum Likelihood Grasp Sampling Loss.” *2022 International Conference on Robotics and Automation (ICRA 2022)* [↗](#)

PROFESSIONAL EXPERIENCE

Research Engineer, International Computer Science Institute

Big Data Group, supervised by Prof. Michael Mahoney

Berkeley, CA

Jan. 2023 – Jun. 2023

- Research on model compression and deep learning generalization.

RESEARCH EXPERIENCE

A Three-Regime Model of Neural Network Pruning

Graduate Research Assistant, Advised by Prof. Michael Mahoney

Berkeley, CA

Dec. 2021 – Dec. 2022

- Proposed a three-regime model based on the mode connectivity of the sparse models obtained from network pruning;
- Leveraged the three-regime model to predict the optimal early stopping in the pre-training stage of pruning pipeline.

A Dataset-dispersion based Evaluation Metric on 3D Reconstruction [↗](#)

Research Assistant, Advised by Prof. Chen Feng, Prof. Yaoqing Yang

Berkeley, CA

Dec. 2020 – Apr. 2021

- First author of a paper on single-view 3D reconstruction (SVR) in 3DV 2021;
- Proposed an evaluation metric **Dispersion Score**, a data-driven metric used to measure the tendency of SVR models to perform recognition or reconstruction;

Learn to Grasp with Less Supervision ☑

Berkeley, CA

Research Assistant, Mechanical Systems Control Lab, advised by **Prof. Masayoshi Tomizuka**

Sep. 2020 – Dec. 2020

- Co-authored a paper on supervised learning based robotic grasping in ICRA 2022;
- Proposed a maximum likelihood grasp sampling loss (**MLGSL**) and **attention-based** model architectures to predict grasps with a single-view depth image, focusing on tackling data label sparsity issue;

Robotics Exoskeleton and Humanoid Trajectory Planning ☑

Berkeley, CA

Research Assistant, Human-Assistive Robotic Tech Lab, advised by **Prof. Ruzena Bajcsy**

Jun. 2019 – Sep. 2019

- Developed trajectory planner and low-level controller for wearable robotic upper limb exoskeleton to assist patient's arm movement in rehabilitation training;

PROJECT EXPERIENCE

Hybrid Policy Gradient with Robotic Grasp Planning ☑

Berkeley, CA

Deep RL Course Project, Co-advised by **Prof. Sergey Levine**

Sep. 2021 – Dec. 2021

- Proposed Hybrid Policy Gradient (**H-PG**), a deep RL framework for predicting robotic grasping in warehouse;
- Proposed to define the RL problem in continuous-discrete action space, and solved it using deep policy gradient.
- H-PG improves baseline by **7.4%** of grasp success rate on YCB dataset in **PyBullet** simulator.

Blink+: Increase GPU Communication Bandwidth using Idle Links across Tenants ☑

Berkeley, CA

Parallel Computing Course Project, Co-advised by **Prof. James Demmel**

Feb. 2022 – May 2022

- Optimized the NVIDIA GPU communication library **NCCL** for large-scale model parallel training;
- Increased the GPU tensor communication bandwidth by utilizing the idle NVLink across user groups;
- Improved the bandwidth of communication operators like broadcast and allreduce up to **200%** on 2 GPU subset.

Robotic Tactile Sensor for Stiffness Estimation ☑

Berkeley, CA

Robotic Manipulation Course Project, Co-advised by **Prof. Ruzena Bajcsy**

Feb. 2019 – May. 2019

- Built a compliance-modulating tactile sensor using pneumatic sealed elastic membrane and a depth-sensing camera for scanning a 3D terrain and measuring the distribution of stiffness;
- Applied Point Cloud Library in C++, ROS and a RGBD camera to enable point cloud segmentation, geometry recognition, and Baxter robot arm control.
- Built a software to enable real-time 3D visualization of point cloud registration and integration including stiffness distribution information in mapping.

SKILLS

Language: Python, Java, C/C++, CUDA, SQL, MATLAB

Learning: Linear/Logistic Regression, Decision Tree, Random Forest, PCA, Clustering (K-means), Deep Models (Transformers, CNN), RL Algorithms (Q-Learning, Offline RL. etc), Model Compression (Pruning)

Developer Tools: PyTorch, Ubuntu, MuJoCo, ROS, PyBullet, Slurm, PyRender, Open3D