Yefan Zhou

yefan0726@berkeley.edu | yefanzhou.github.io | Berkeley, CA, 94709 | 510-809-5378

Education

University of California, Berkeley — Berkeley, CA
M.Eng in Electrical Engineering and Computer Science; Major GPA: 4.0/4.0 — *Aug. 2021 – Dec. 2022*
Coursework: Convex Optimization (A), Deep Reinforcement Learning (A+), Parallel Computing (A)
Principles and Techniques of Data Science (A)

University of California, Berkeley — Berkeley, CA
Exchange Student; GPA: 4.0/4.0 — *Jan. 2019 – May. 2019*
Coursework: Data Structures (A+), Intro to Artificial Intelligence (A), Computational Structures in Data Science (A)
Robotic Manipulation and Interaction (A)

Southeast University — China
B.Eng in Information Engineering; GPA: 3.7/4 — *Aug. 2016 – Jun. 2020*
Coursework: Database System, Information Security, Computer Architecture, Microcomputer Systems
Analog Circuits, Electromagnetic Fields, Principle of Communication, Signals and Systems

Publication

- **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)
- A. Zhao, Y. Yang, E. Ye, **Y. Zhou**, Z. Liu, X. Yue, K. Keutzer, J. Gonzalez, R. Kannan, M. Mahoney "ME-Prune: Highly Compressed and Robust Neural Networks via Matrix Entropy Based Channel Pruning." Preprint
- X. Zhu, Y. Fan, C. Wang, **Y. Zhou**, S. Jin, M. Tomizuka "Multi-Fingered Grasp Pose Detection using Point Cloud." Preprint, submitted to IEEE Robotics and Automation Letters (RAL)

Research Experience

A Two-Regime model of Neural Network Pruning — Berkeley, CA
Graduate Research Assistant, Advised by **Prof. Michael Mahoney** — *Dec. 2021 –*

- Proposed a two-regime model on neural network pruning, in which early stopping in the pre-training can either benefit or hurt the final test-time performance of pruned model with different prune ratios;
- Leveraged the metric mode connectivity of loss landscape to quantify the transition between the two regimes (early stopping or not);
- Proposed an efficient approach based on mode connectivity to evaluate the performance of SOTA pruning algorithms (includes pruning criterions, retraining initializations/learning rate schedules) without training.
- First author of a paper under progress.

ME-Prune: Matrix Entropy Based Neural Network Pruning — Berkeley, CA
Graduate Research Assistant, Advised by **Prof. Michael Mahoney** — *Aug. 2021 –*

- Co-authored a paper on **ME-Prune**, a pruning algorithm for removing input channels in CNN;
- Analyzed the layer-wise importance of neural network based on the Empirical Spectral Density (ESD) of weight matrices. Utilized the entropy of ESD to determine the optimal pruning budget for each layer;
- Improved SOTA by **3%** of benign and out-of-distribution classification accuracy using DenseNet40 on CIFAR-100.

A Dataset-dispersion based Evaluation Metric on 3D Reconstruction — Berkeley, CA
Research Assistant, Advised by **Prof. Chen Feng, Prof. Yaoqing Yang** — *Dec. 2020 – Apr. 2021*

- First author of a paper on single-view 3D reconstruction (SVR) accepted by 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;
- Experimentally show that Dispersion Score can detect the bias of network towards recognition when training images are more dispersed, providing a novel perspective to evaluate models and diagnose training data.

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 accepted by ICRA 2022;
- Proposed a maximum likelihood grasp sampling loss (**MLGSL**) to predict grasps with a single-view depth image, focusing on tackling data label sparsity issue;
- MLGSL is **8**× more data-efficient than SOTA with a **91.8%** grasp success rate in real-world experiments.

Multi-Fingered Grasp Pose Detection using Point Cloud Berkeley, CA

Research Assistant, Mechanical Systems Control Lab, advised by **Prof. Masayoshi Tomizuka** *Jun. 2020 – Sep. 2020*

- Constructed a point-cloud based multi-fingered grasp dataset by rendering depth images from 3DNet object using Pyrender and Open3D library.
- Reproduced and implemented the SOTA point cloud grasp pose detection (GPD) algorithm as a baseline.

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;
- Applied Minimum Angle/Hand Jerk Algorithm to trajectory planner to realize Cartesian Space and Joint Space real-time humanoid trajectory planning and execution.

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.

World Exploration Engine Berkeley, CA

Data Structure Course Project *Nov. 2020 – Dec. 2020*

- Built a tile-based interactive game (like Zelda II);
- Implemented pseudo-randomly world generation algorithm, pathfinding AI player, and game status saving/loading.

Navigation Map of Berkeley Berkeley, CA

Data Structure Course Project *Oct. 2020 – Nov. 2020*

- Created a web mapping app for Berkeley area based on OpenStreetMap data;
- Implemented backend features: map rastering, auto-complete search, routing;
- Optimized nearest address positioning using **K-d Tree** and routes finding using **memory-optimizing A***.

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;
- Applied Point Cloud Library in C++, ROS and a RGBD camera to enable point cloud segmentation.
- Built a software to enable real-time 3D mapping using point cloud including stiffness distribution information.

Standard Tests:

TOEFL 100 (L: 26, R: 28, W: 23, S: 23) **GRE**: 332 (V:160, Q:172, AW 3.0)

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