Shijia Yang

shijiay@stanford.edu • (626) 861-1312 • Menlo Park, California

EDUCATION

Stanford University   **May 2023 - Present**

- Master of Science in Computer Science with GPA: 3.8/4.0

University of California, Berkeley   **September 2019 - May 2023**

- Bachelor of Art in Computer Science with GPA: 3.9/4.0
- Relevant Coursework:
  - <u>Applied Theory</u> - 3D Vision (Graduate) [A+], Machine Learning, DNN, Computer Graphics
  - <u>Theory</u> - Optimization, Linear Algebra, Probability, Discrete Mathematics, Multivariate Calculus

RESEARCH EXPERIENCE

Stanford Vision and Learning Lab   **April 2023 - December 2023**
Graduate Researcher, advised by Prof. Jiajun Wu   Stanford, CA

- Researching on large language model spatial knowledge for robotic vision

Mechanical Systems Control Lab   **March 2021 - May 2023**
Undergraduate Researcher, advised by Prof. Masayoshi Tomizuka   Berkeley, CA

- Researched on multi-modal multi-task learning for autonomous vehicles

Berkeley Artificial Intelligence Lab   **March 2021 - May 2023**
Undergraduate Researcher, advised by Prof. Kurt Keutzer & Prof. Joseph E. Gonzalez   Berkeley, CA

- Researched on data and training efficiency for 3D vision and language-vision models

PUBLICATIONS

[1] Image2Point: 3D Point-Cloud Understanding with 2D Image Pretrained Models
Shijia Yang*, Chenfeng Xu*, Tomer Galanti, Bichen Wu, Xiangyu Yue, Bohan Zhai, Wei Zhan, Peter Vajda, Kurt Keutzer, Masayoshi Tomizuka
European Conference on Computer Vision (ECCV) 2022

- Proposed a simple pretraining scheme of using 2D model weights for 3D point-cloud understanding
- Brought consistent improvement for various baselines, including 3D Sparse Conv and PointNet++, and 10.0% improvement for few-shot classification on the ModelNet40 dataset
- Explained feasibility of image-point-cloud transfer from the aspect of Neural Collapse

[2] Time Will Tell: New Outlooks and a Baseline for Temporal Multi-View 3D Object Detection
Jinhyung Park*, Chenfeng Xu*, Shijia Yang, Kurt Keutzer, Kris Kitani, Masayoshi Tomizuka, Wei Zhan
International Conference on Learning Representations (ICLR) 2023 Top-5%

- Proposed SOLOFusion for temporal camera-only 3D detection using a long history of image observations with long-term but coarse resolution, then augment with short-term, fine-grained resolution
- SOLOFusion sets new state-of-the-art on nuScenes, achieving first place on the test set and outperforming previous best art by 5.2% mAP and 3.7% NDS on the validation set

[3] Multitask Vision-Language Prompt Tuning

Shijia Yang*, Sheng Shen*, Tianjun Zhang*, Bohan Zhai, Joseph E. Gonzalez, Kurt Keutzer, Trevor Darrell

Winter Conference on Applications of Computer Vision (WACV) 2023

- Proposed the multitask vision-language prompt tuning (MVLPT) framework, including multitask prompt initialization and multitask prompt adaptation, and demonstrate the efficacy for each component
- Evaluate MVLPT on the few-shot ELEVATER and cross-task generalization benchmarks, which sets the new state-of-the-art on 20-shot ELEVATER benchmark

[4] HallE-Switch: Controlling Hallucinations in Large Vision Language Models

Shijia Yang*, Bohan Zhai*, Chenfeng Xu, Sheng Shen, Kurt Keutzer, Chunyuan Li, Manling Li

Conference on Language Modeling (COLM) 2024 Under Review

- Proposed a GPT-4 assisted evaluation for detailed captioning and a method, HallE-Switch that reduces hallucination in large vision-langauge models by 44% compared to LLaVA7B
- HallE-Switch conditions the captioning to shift between (i) exclusively depicting contextual knowledge for grounded objects and (ii) blending it with parametric knowledge to imagine inferred objects

INDUSTRY EXPERIENCE

Microsoft - DeepSpeed Team | **January 2023 - August 2023**
Research Scientist Intern | Bellevue, Washington

- Worked on kv cache pruning for efficient inference on long text generation of LLMs, such as Llama2, MPT, and OPT

TikTok - Large Language Models Team | **February 2024 - Present**
Research Scientist Intern | San Jose, California

- Working on visual feature alignment of Large Multimodal Models

PERSONAL PROJECTS

Multi-modal Multi-task Learning for 3D Perceptive via Closed-loop Representation *advised by Prof. Yi Ma*

- Proposed a paradigm, inspired by closed-loop feedback from control theory, for a multi-modal multi-task 3D model
- Performed five common tasks for point-cloud perception, including foreground detection and object detection
- Achieved at least 4% mAP improvement for object detection on Argoverse dataset compared with UNet3D baseline

Snow Simulation *advised by Prof. Ren Ng*

- Used Taichi language designed for physical simulation and implemented the moving-least-square algorithm (MPM) to simulate realistic snow particles in 2D and 3D
- Studied snow particles' update rule for positions, affine velocity field, deformation gradient, etc.
- Implemented snowballs, frictional surface, and rigid body by adjusting particle distribution and parameters

Data Free Bert Quantization *advised by Prof. Sergey Levine*

- Alleviating the problem that quantized models may not have original data to perform post-quantization finetuning
- Proposed token generation algorithm to generate pseudo data for low precision Bert model quantization
- Improved accuracy of quantized Bert by 12.2%-61.7% on GLUE tasks compared to quantized Bert baseline