

# YI LI

<https://yili.vision> [Google scholar](#) [LinkedIn](#)  
yili18@cs.washington.edu

## RESEARCH INTERESTS

---

Vision-Language Models, Foundational Models, Vision-Language-Action Learning, Open-World Manipulation, Imitation Learning, Generalization to Unseen Scenarios, Robotic Perception.

## EDUCATION

---

### University of Washington

Ph.D. student, advised by Prof. Dieter Fox

*Sep. 2018 -*  
Computer Science and Engineering

### Tsinghua University

Master of Science (Summa Cum Lauda in Beijing)

*Aug. 2014 - Jun. 2017*  
Department of Automation

### Tsinghua University

Bachelor of Engineering

*Aug. 2010 - Jun. 2014*  
Department of Automation

## PUBLICATIONS

---

(\* indicates co-first author, **red** denotes representative papers)

Yi Li\*, Yuquan Deng\*, Jesse Zhang\*, Joel Jang, Marius Memmel, Caelan Garrett, Fabio Ramos, Dieter Fox, Anqi Li, Abhishek Gupta, Ankit Goyal

### **HAMSTER: Hierarchical Action Models for Open-World Robot Manipulation**

*International Conference on Learning Representations (ICLR)*, 2025, *under review*, project website

Keywords: VLM, Hierarchical VLA, Open-World Manipulation, Imitation Learning, 3D Reasoning

Zijian Zhang\*, Kaiyuan Zheng\*, Zhaorun Chen, Joel Jang, Yi Li, Chaoqi Wang, Mingyu Ding, Dieter Fox, Huaxiu Yao

### **GRAPE: Generalizing Robot Policy via Preference Alignment**

*Computer Vision and Pattern Recognition (CVPR)*, 2025, *under review*, project website

Keywords: Reinforcement Learning, DPO, Manipulation, Safety

Soofiyan Atar, Yi Li, Markus Grotz, Michael Wolf, Dieter Fox, Joshua Smith

### **OptiGrasp: Optimized Grasp Pose Detection Using RGB Images for Warehouse Picking Robots**

*International Conference on Robotics and Automation (ICRA)*, 2025, *under review*, project website

Keywords: RGB-only, Grasp Pose Prediction, Warehouse Robot

Yi Li, Muru Zhang, Markus Grotz, Kaichun Mo, Dieter Fox

### **STOW: Discrete-Frame Segmentation and Tracking of Unseen Objects for Warehouse Picking Robots**

In *Conference on Robot Learning (CoRL)*, 2023, project website

Keywords: Unseen Object, Identification, Tracking, Warehouse Robot

Xingyu Liu, Gu Wang, Yi Li, Xiangyang Ji

### **CATRE: Iterative Point Clouds Alignment for Category-level Object Pose Refinement**

In *European Conference on Computer Vision (ECCV)*, 2022, arxiv

Keywords: Category-Level Object 6D Pose Refinement, Point Cloud

Yi Li, Gu Wang, Xiangyang Ji, Yu Xiang, Dieter Fox

### **DeepIM: Deep Iterative Matching for Object Pose Estimation**

In *International Journal of Computer Vision (IJCV)*, 2020, arxiv

In *European Conference on Computer Vision (ECCV)*, 2018 (oral), top 12 papers

Keywords: Object 6D Pose Refinement, RGB-only

Jifeng Dai\*, Haozhi Qi\*, Yuwen Xiong\*, Yi Li\*, Guodong Zhang\*, Han Hu, Yichen Wei

**Deformable Convolutional Networks**

In *International Conference on Computer Vision (ICCV)*, 2017 (oral). arxiv

Keywords: Convolution Layer, Attention

Yi Li\*, Haozhi Qi\*, Jifeng Dai, Xiangyang Ji, Yichen Wei

**Fully Convolutional Instance-aware Semantic Segmentation**

In *Computer Vision and Pattern Recognition (CVPR)*, 2017 (spotlight). arxiv

Keywords: Instance Segmentation, End-to-end

Jifeng Dai, Yi Li, Kaiming He, Jian Sun

**R-fcn: Object detection via region-based fully convolutional networks**

In *Advances in Neural Information Processing Systems(NeurIPS)*, 2016. arxiv

Keywords: Object Detection, Efficiency, Fully Convolution Network

Jifeng Dai, Kaiming He, Yi Li, Shaoqing Ren, Jian Sun

**Instance-sensitive fully convolutional networks**

In *European Conference on Computer Vision (ECCV)*, 2016. arxiv

Keywords: Regional Proposal Network, Object Detection

**RESEARCH EXPERIENCE**

---

**NVIDIA AI Robotics Research Lab**

*Research Intern*

Dec. 2023 - Present

*supervised by Dr. Ankit Goyal*

- HAMSTER: Hierarchical Action Models for Open-World Robot Manipulation.

Developed hierarchical Vision-Language-Action (VLA) models for robotic generalization. Designed VLA models with high-level VLMs trained on scalable off-domain data to produce semantically meaningful intermediate predictions guiding 3D-aware control policies. Enabled broad visual, semantic, and geometric generalization across domain gaps, improving manipulation in both simulation and real-world environments.appearance. Submitted to ICLR 2025.

**Robotics and State Estimate Lab, University of Washington**

*Research Assistant*

Sep. 2018 - Present

*supervised by Prof. Dieter Fox*

- Leading the perception team in Amazon-UW-Robotics-Manipulation-Research

The project aims to have the robot arm to pick products from amazon pods automatically.

STOW identify and track the segmentation of each object given a sequence of images which are captured every time a human operator put an object into the pod. Accepted by CoRL 2023.

OptiGrasp use depth estimation network to get 3D understanding objects and predict the gripper pose to pick objects from bins using only rgb images. Submitted to ICRA 2024.

- Research on object pose estimation and tracking with only RGB images

DeepIM, a novel approach to provide high-accuracy 6D pose estimation, accepted by ECCV 2018 (oral, top 12), IJCV 2020

**NVIDIA AI Robotics Research Lab**

*Research Intern*

Sep. 2021 - March. 2022

*supervised by Dr. Arsalan Mousavian and Dr. Lucas Manuelli*

- General object embedding for multiple robotic manipulation tasks like grasping and pushing etc.

Introduce the idea of bipartite matching into grasp pose prediction to solve the problem that annotation only cover a subset of the whole solution space.

**Visual Computing Group, Microsoft Research Asia**

*Research Intern*

Nov. 2015 - Jun. 2017

*supervised by Dr. Jifeng Dai and Dr. Yichen Wei*

- Developed Deformable Convolution Network accepted in ICCV 2017 (oral)

Propose a novel way to do conv and roi-pooling method which can help the network better deal with the variance of scale and rotation of objects in images

- Developed instance-aware segmentation framework FCIS accepted in CVPR 2017 (spotlight)  
End-to-end instance segmentation framework.  
Won the first prize in the MS COCO Object Detection(SEGM) Challenge 2016 by a large margin
- Developed fast and accurate object detection method R-FCN accepted in NIPS 2016  
A novel method to generate instance-level segment candidates

## HONORS

---

Reviewer of AAAI, ICCV, CVPR, ECCV, ICRA, IROS, ICLR, CoRL, RA-L

Outstanding 2017 Master Thesis by Chinese Institute of Electronics (10 in China)

1st Prize in MSCOCO 2016 Object Detection Challenge

Outstanding 2016 Intern in MSRA

2017 Summa Cum Lauda in Beijing

2013 National Scholarship

## SKILLS

---

VLM, VLA, Imitation Learning, Large-Scale Training, Computer Vision, Robotics, Python, Pytorch, Diffusion Model, Diffusion Policy, CUDA, RoS