# Perceive, Plan, Act and Learn: Towards Intelligent Robots in Human Environments

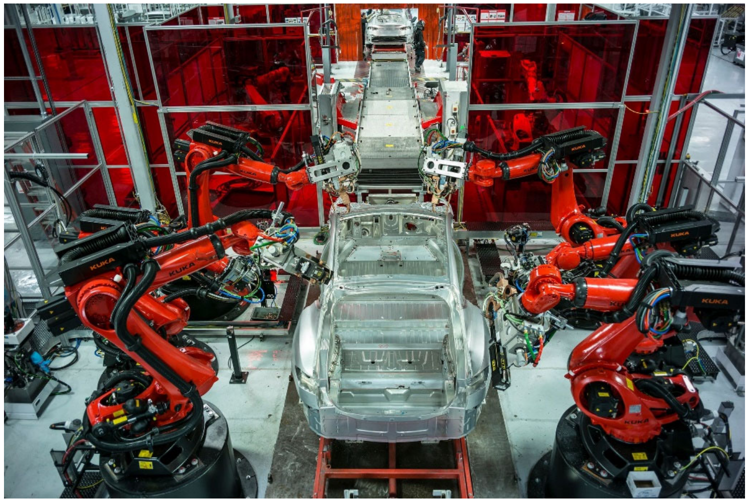
Yu Xiang

Senior Research Scientist

NVIDIA Research

NVIDIA

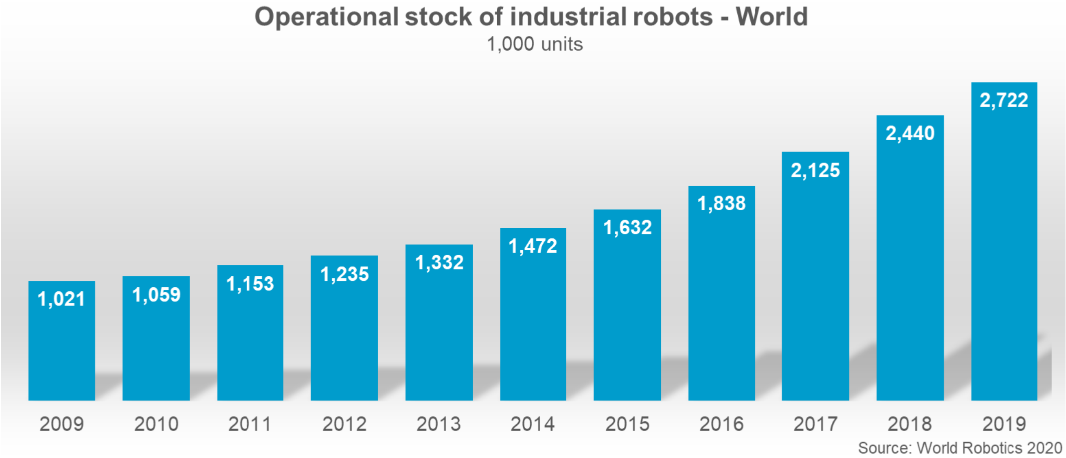
# Robots in Factories and Warehouses



Welding and Assembling

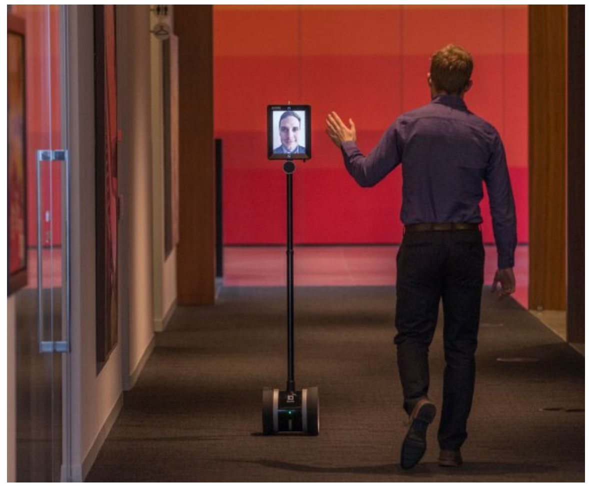
Material Handling

Delivering



# Current Robots in Human Environments

Cleaning Robots



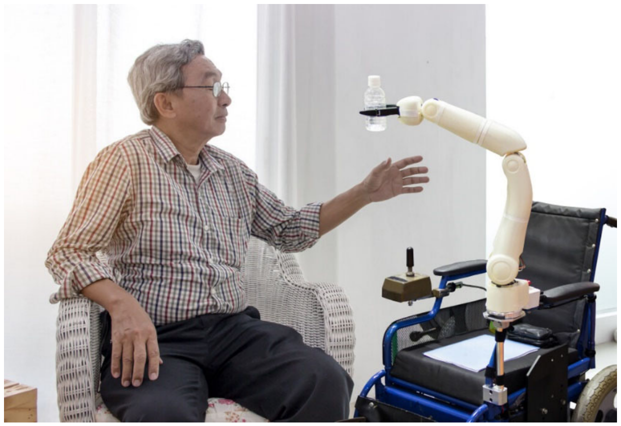
Telepresence Robots

Smart Speakers

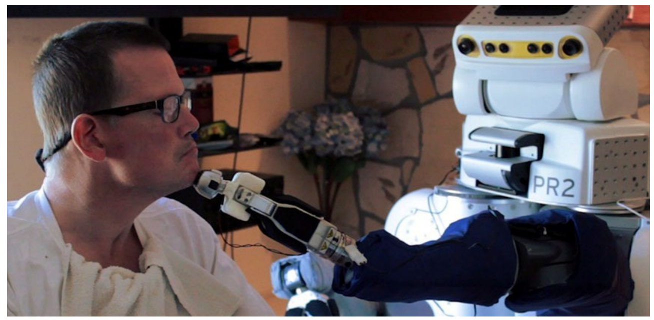
How can we have more powerful robots assisting people at homes or offices?

- Mobile manipulators
- Humanoids

# Future Intelligent Robots in Human Environments



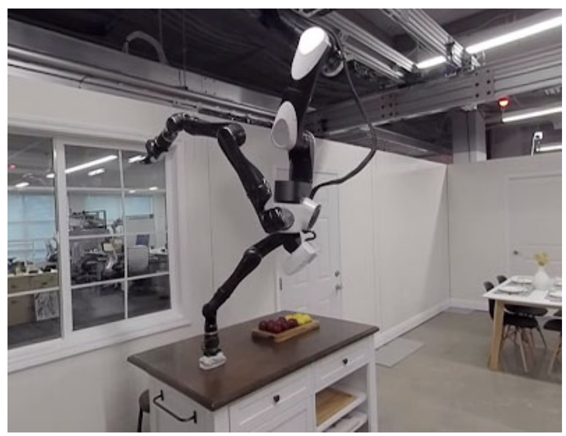
Senior Care




Assisting

Serving

Cooking



Cleaning

Dish washing

# Why Bringing Robots to Human Environments is Challenging?

Closed World: Factories & Warehouses

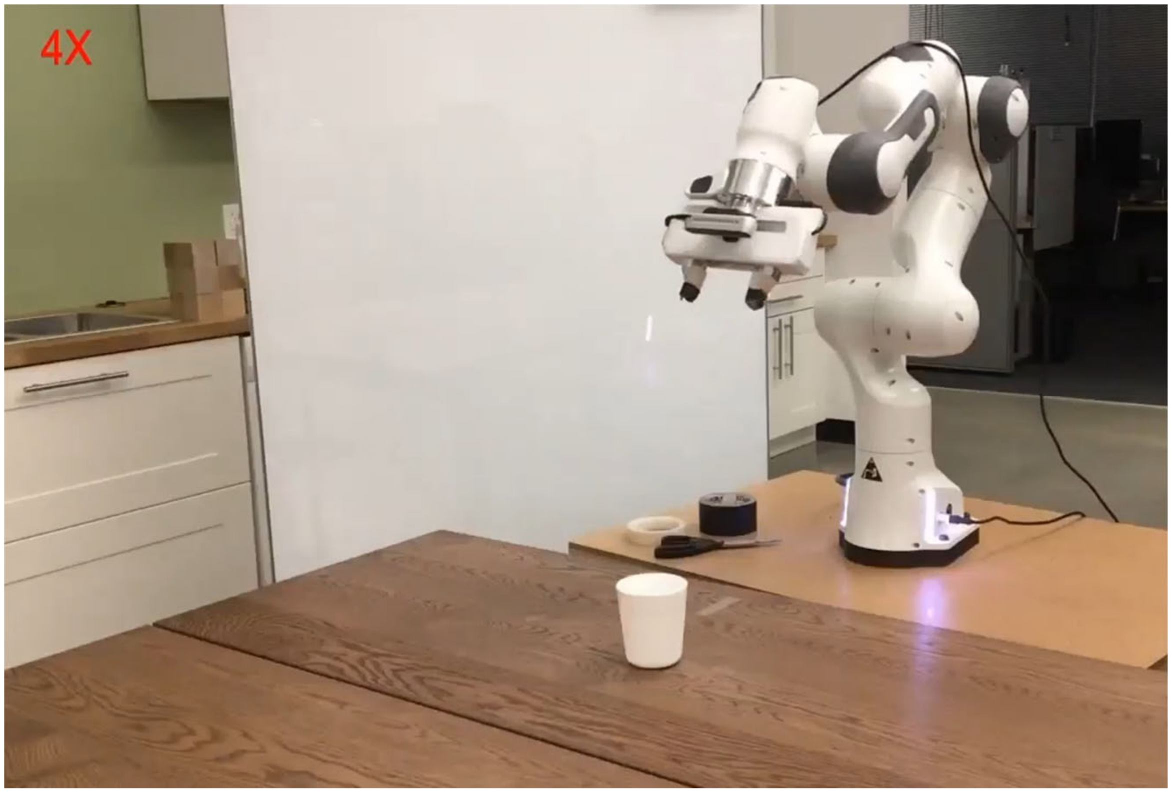
- Structured environments
- Single tasks

Open World: Human environments

- Unstructured and dynamic environments
- Various tasks

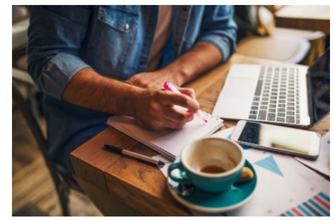
# Why Bringing Robots to Human Environments is Challenging?

Example: Picking up a mug



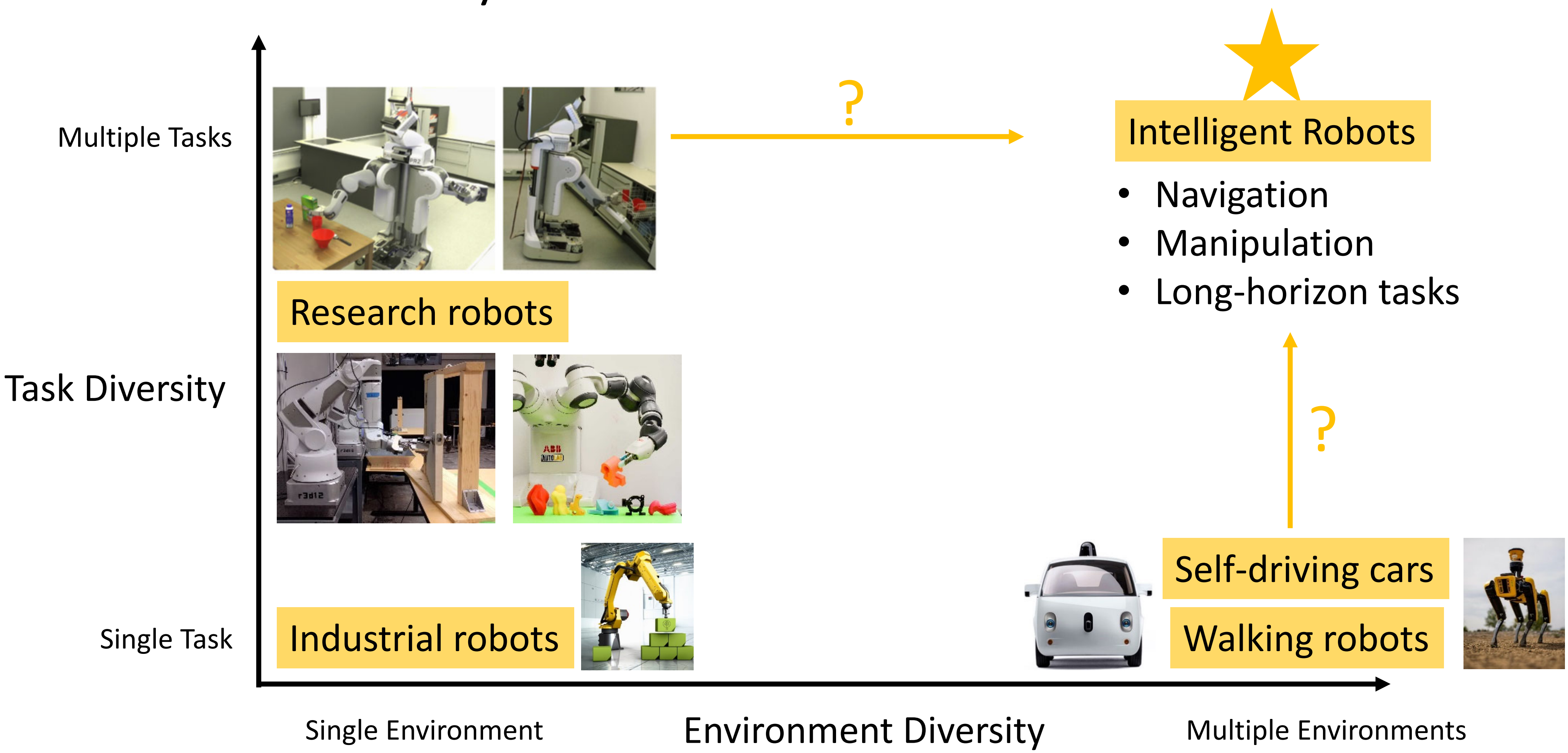

Our Lab

Environment Diversity

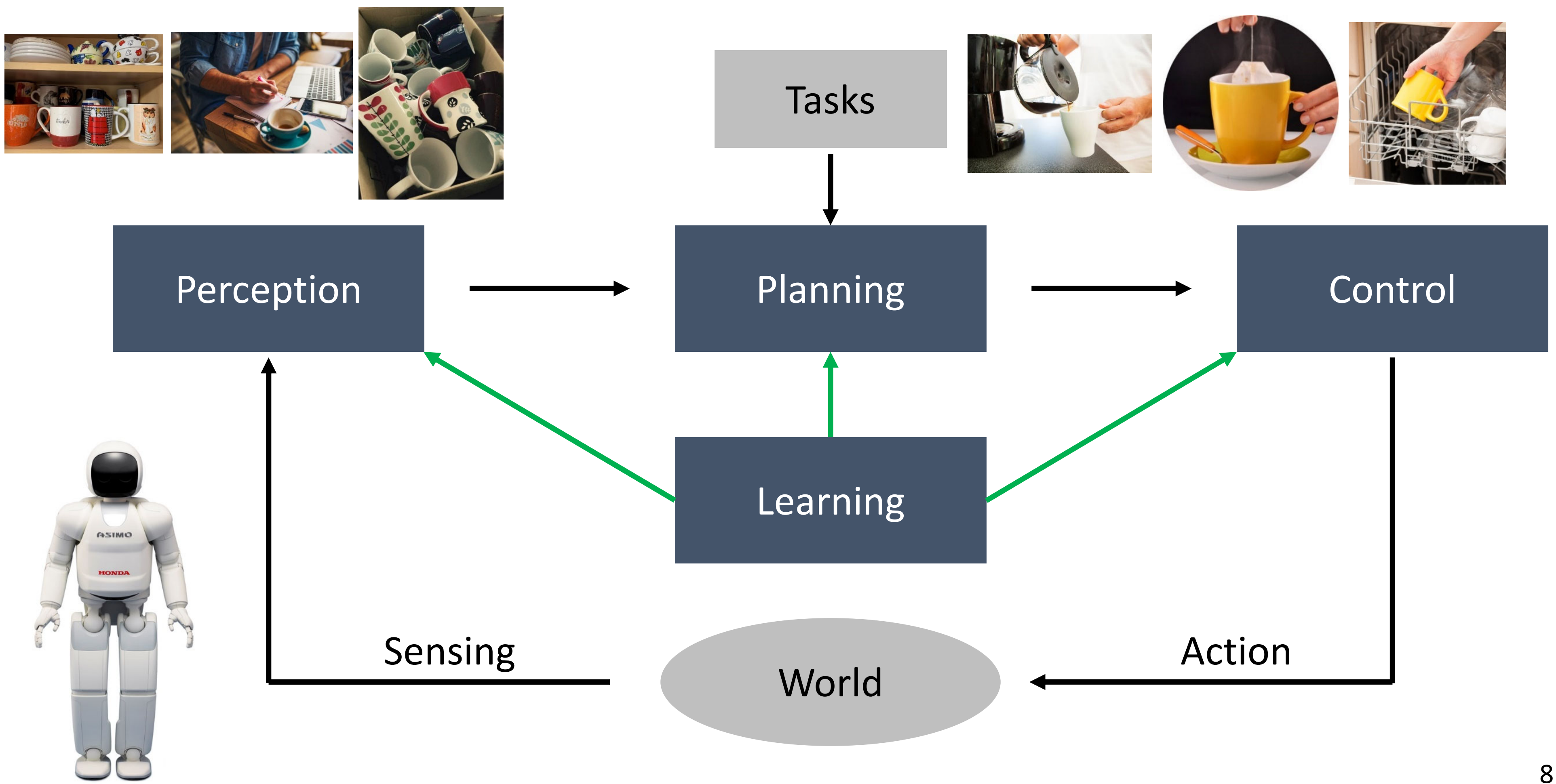


Task Diversity

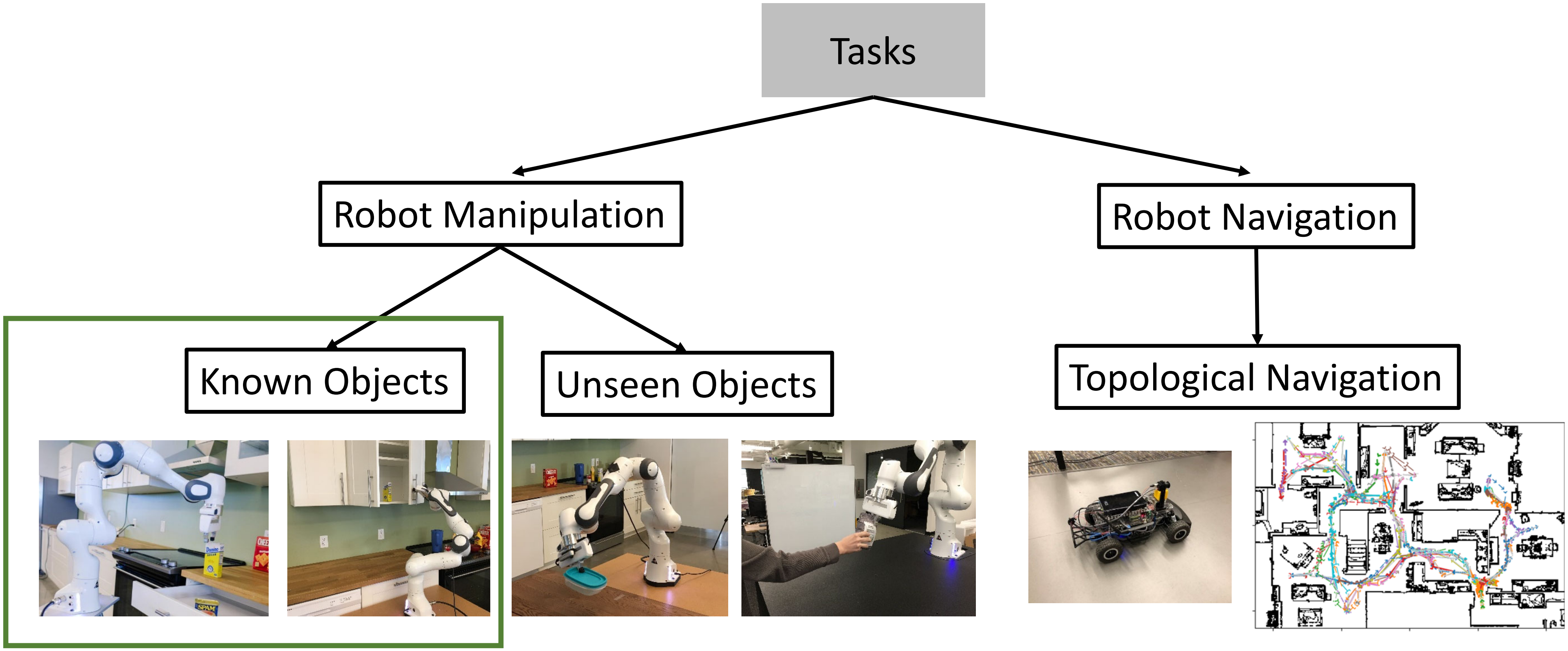
# Robot Autonomy



# The Perception, Planning and Control Loop



# Outline


Tasks
Robot Manipulation
Robot Navigation
Known Objects
Unseen Objects
Topological Navigation

# Robot Manipulation

Manipulation Tasks

↓

Perception → Planning → Control

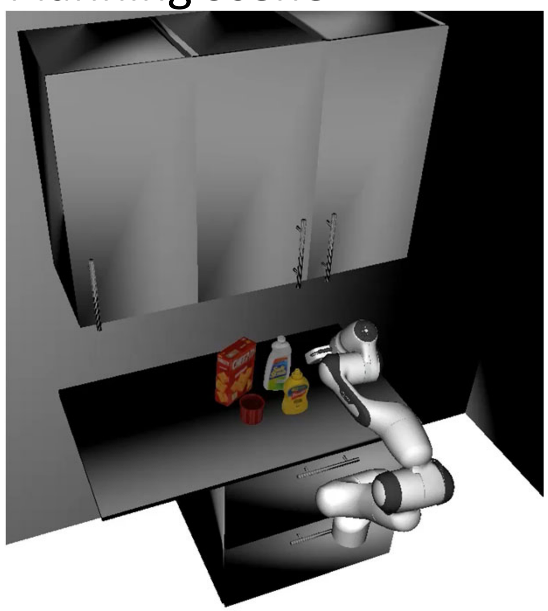
Robust and Accurate

High degree of freedom
Multi-modal grasping

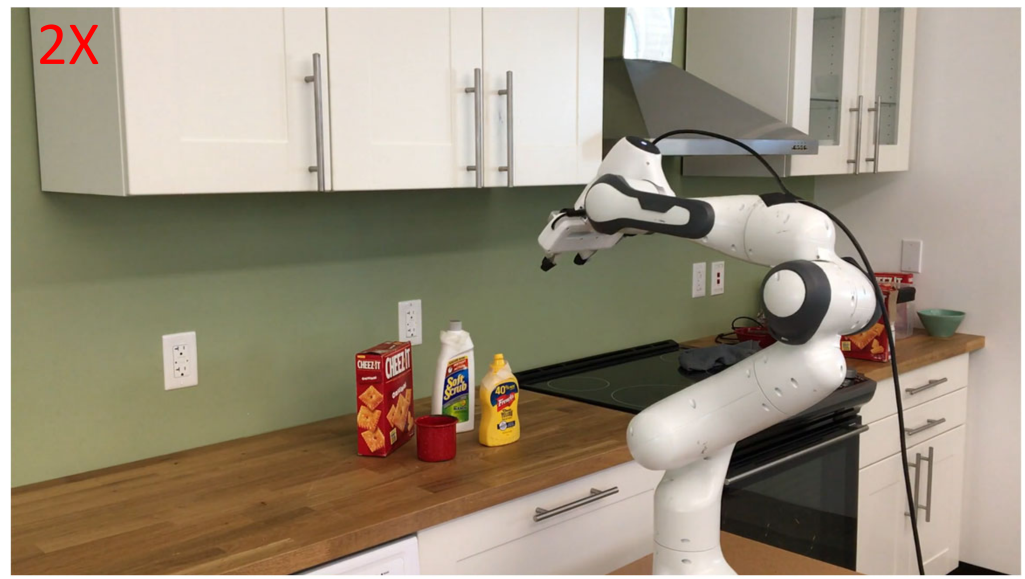
Contact with objects

Sensed image

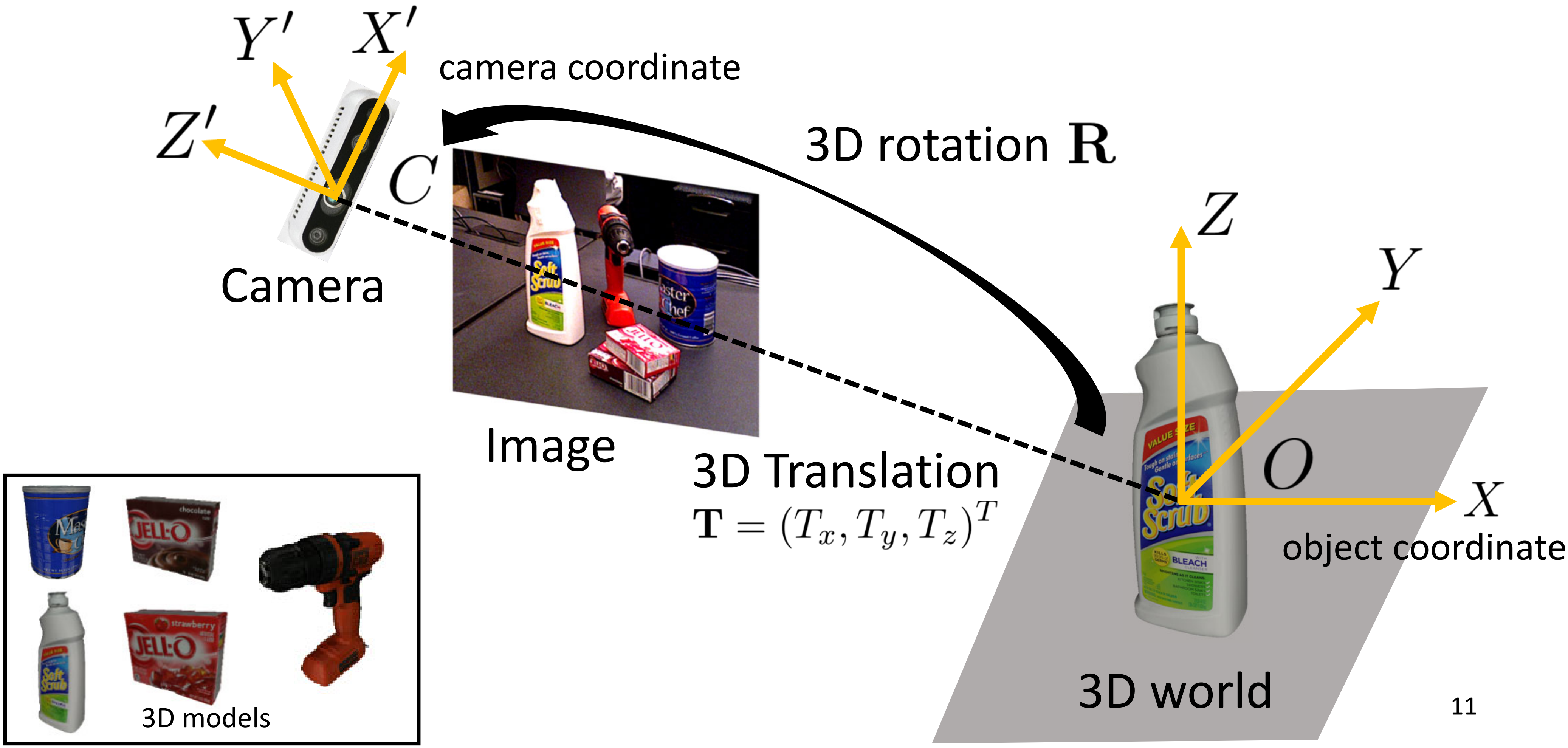
Planning scene



Real world execution



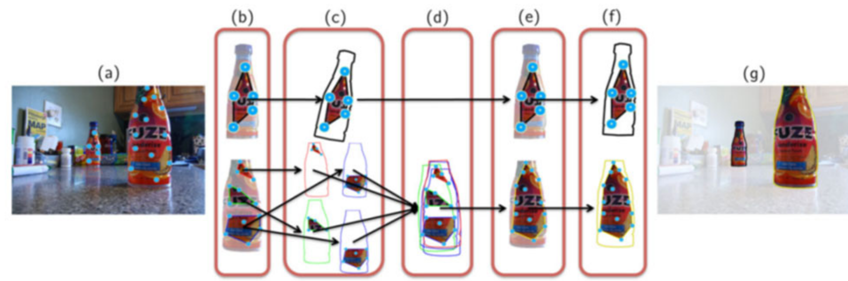
# Perception: Model-based 6D Object Pose Estimation



# Traditional Methods for 6D Object Pose Estimation

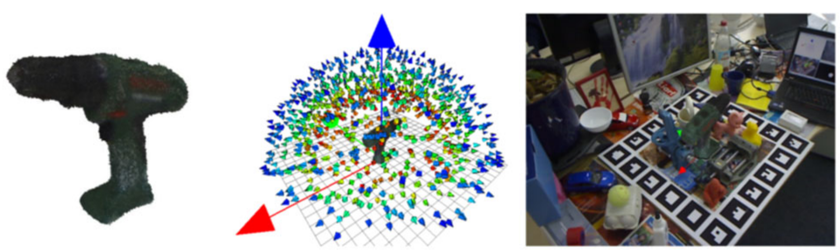
- Feature matching-based methods


Rothganger-Lazebnik-Schmid-Ponce, IJCV’06

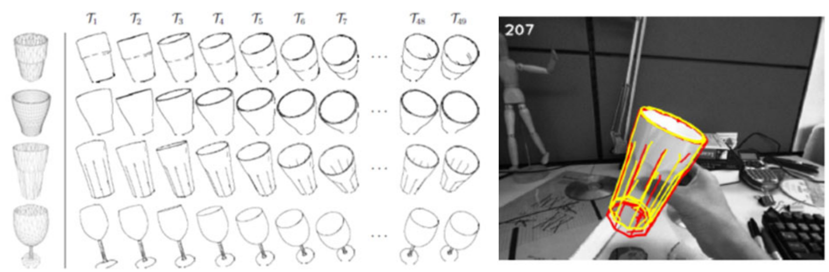



Collet-Martinez-Srinivasa, IJRR’11

- Template matching-based methods



Hinterstoisser-Lepetit-Ilic-Holzer-Bradski-Konolige-Navab, ACCV’12

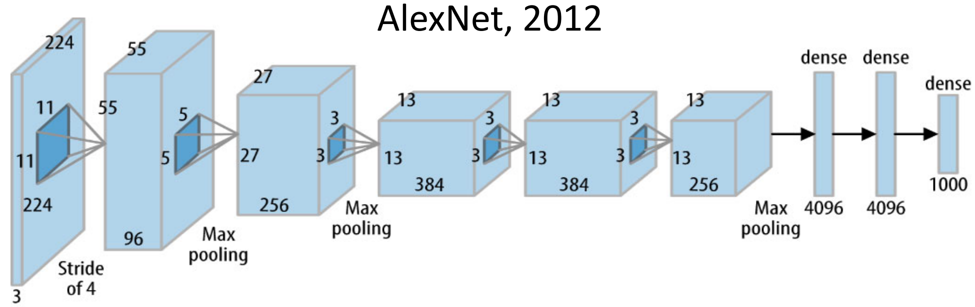



Choi-Christensen, IROS’12

# Deep Learning for Visual Recognition

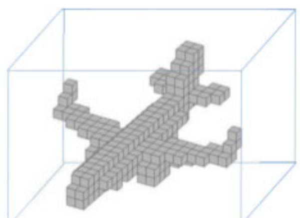
- Images

AlexNet, 2012




Class label

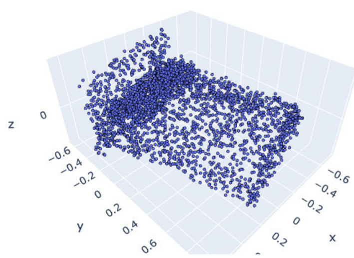
- Voxels



3D CNN

Class label

- Point Clouds




PointNet

Class label

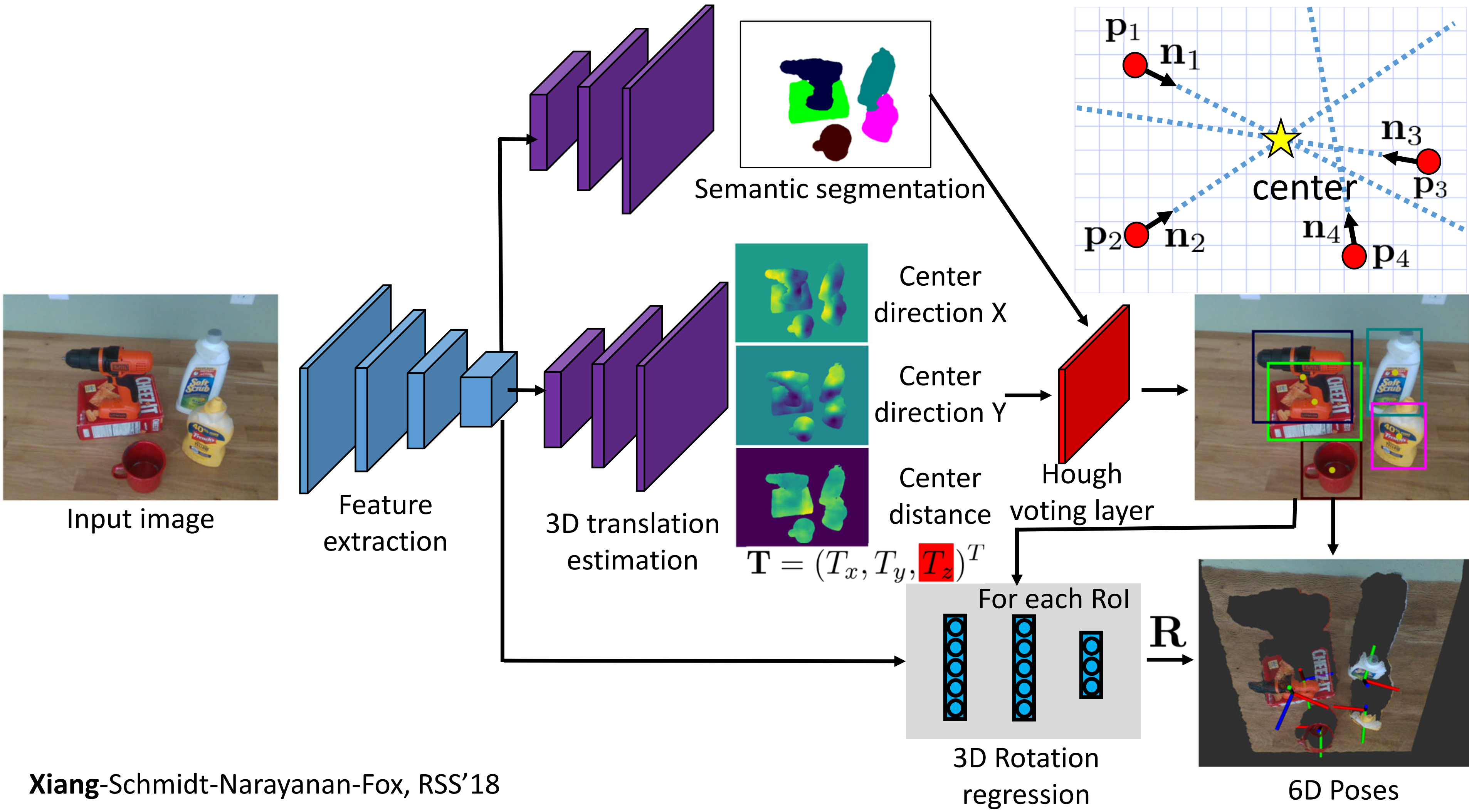
# PoseCNN: the First End-to-end 6D Pose Estimation Network

PoseCNN

- ✓ Texture-less objects
- ✓ Symmetric objects
- ✓ Occlusions

**Xiang**-Schmidt-Narayanan-Fox, RSS'18

# PoseCNN: the First End-to-end 6D Pose Estimation Network

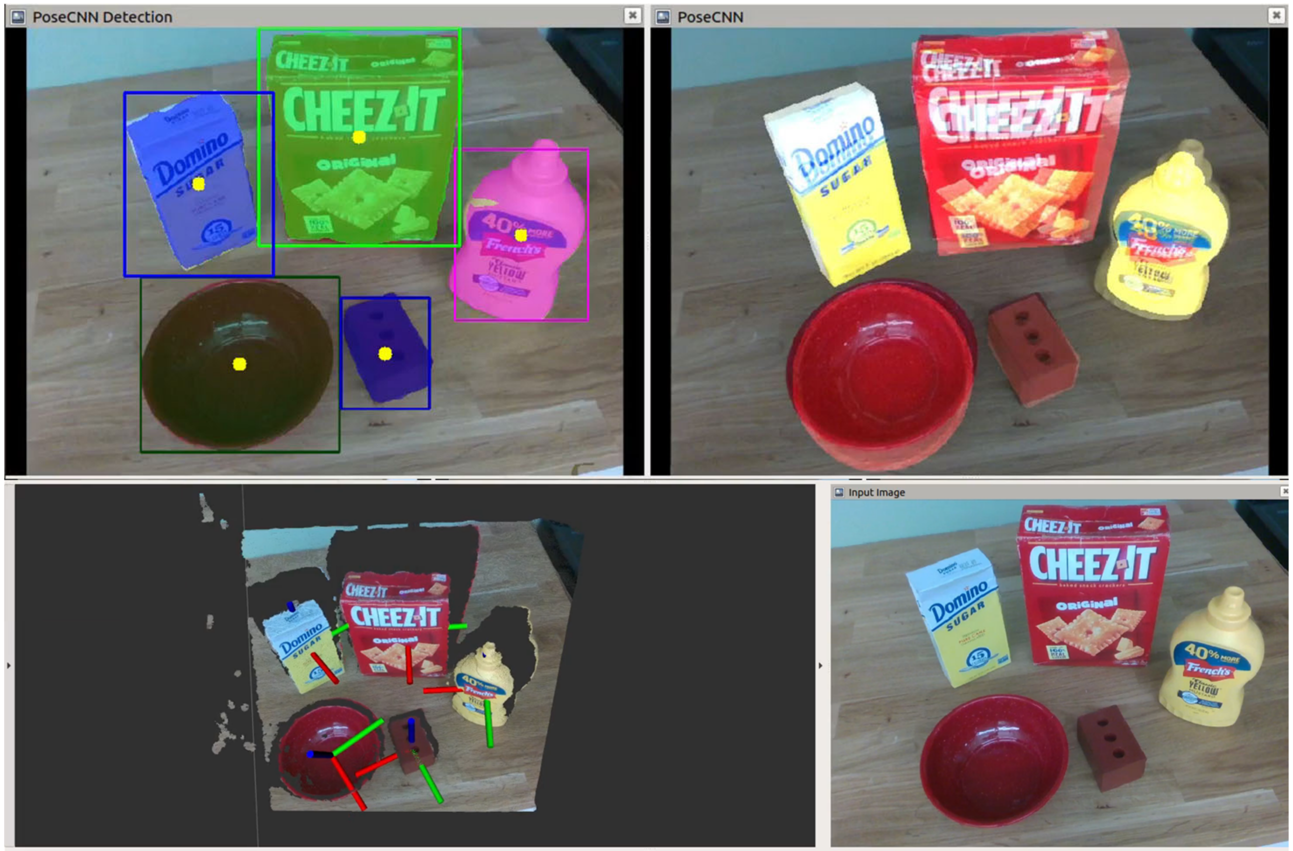



**Xiang**-Schmidt-Narayanan-Fox, RSS'18

# PoseCNN: the First End-to-end 6D Pose Estimation Network

Segmentation and Detection

Poses




3D World

Input image

**Xiang**-Schmidt-Narayanan-Fox, RSS’18

# The Sim-to-Real Gap

Synthetic images

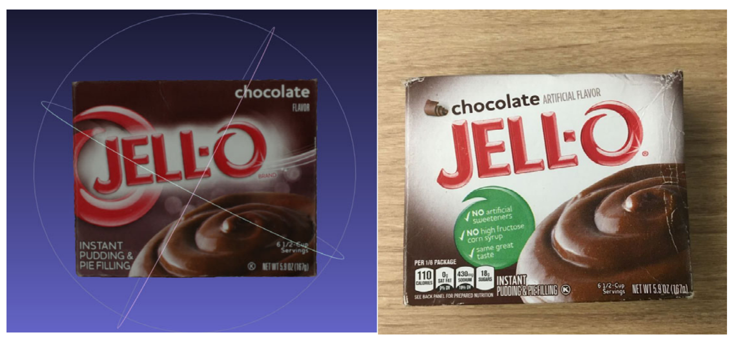

Training

PoseCNN

Domain randomization

Lighting and background

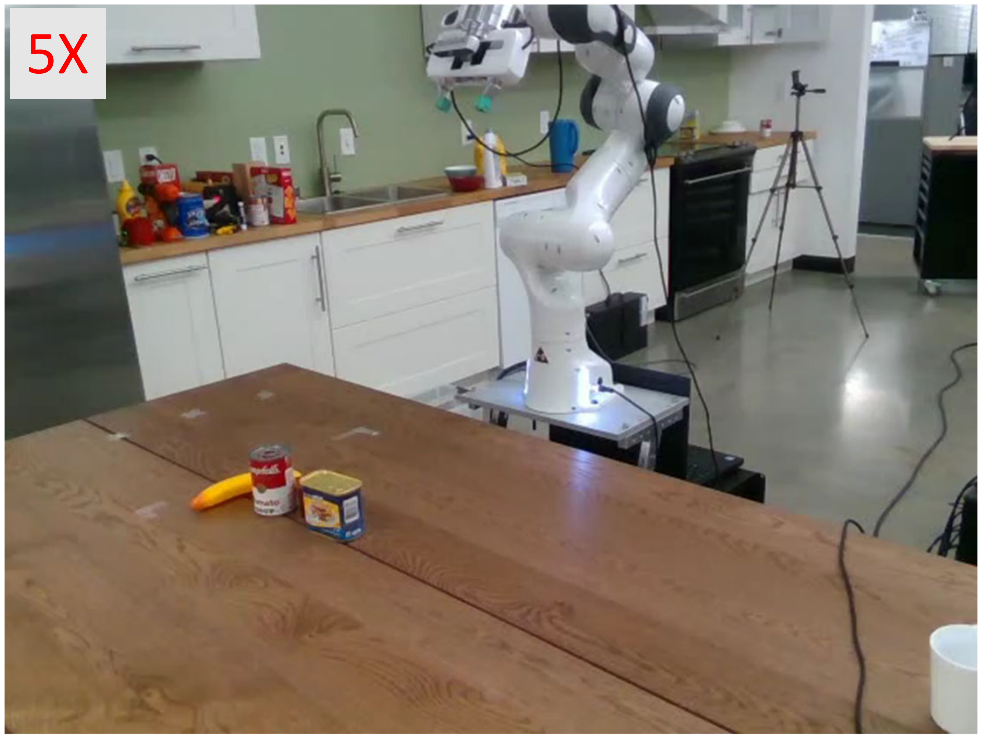
Texture




Moving Part

# Self-supervised 6D Object Pose Estimation

Interactive real-world data collection



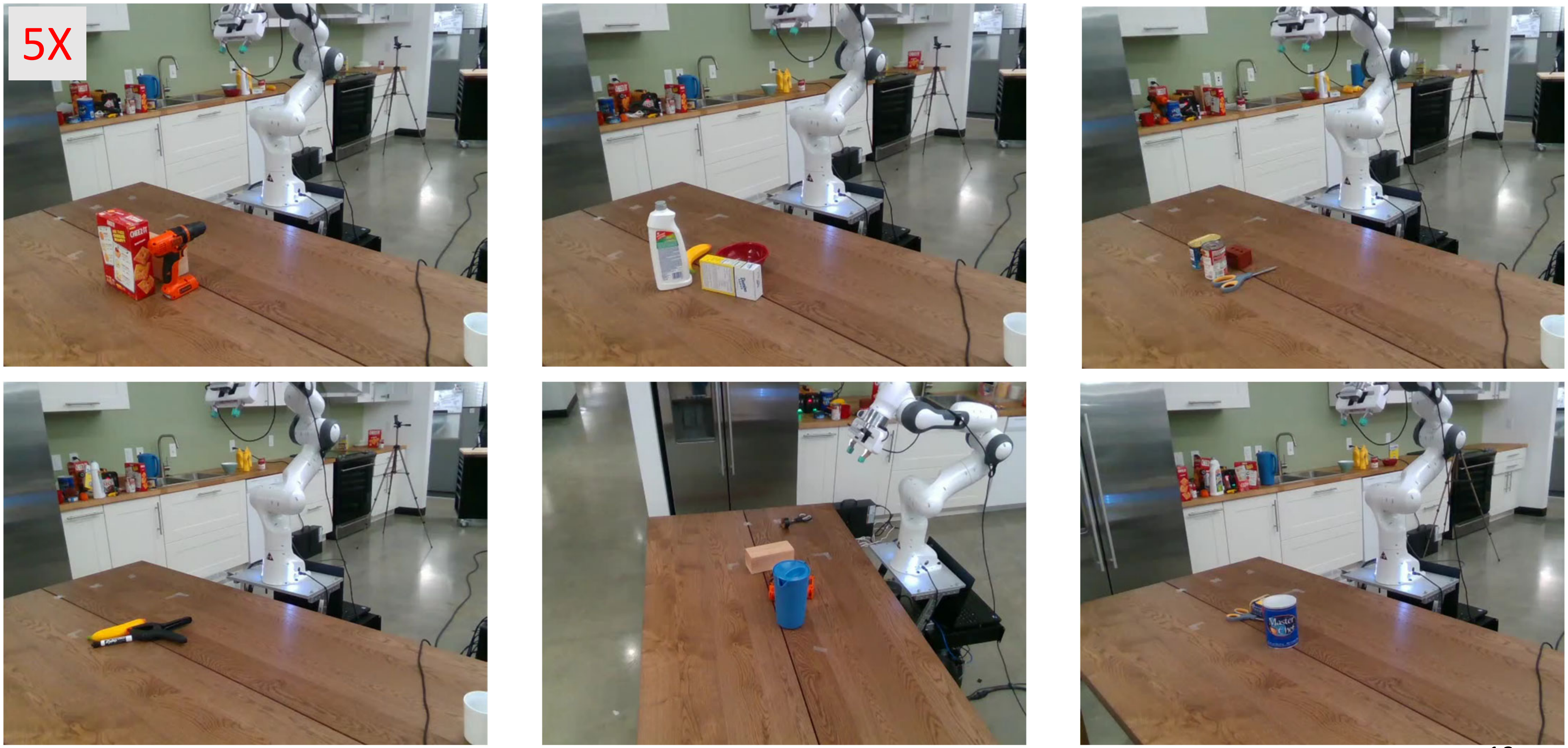

Generated pose annotations

Overlay of rendering onto image

Deng-**Xiang**-Mousavian-Eppner-Bretl-Fox, ICRA'20

# Self-supervised 6D Object Pose Estimation

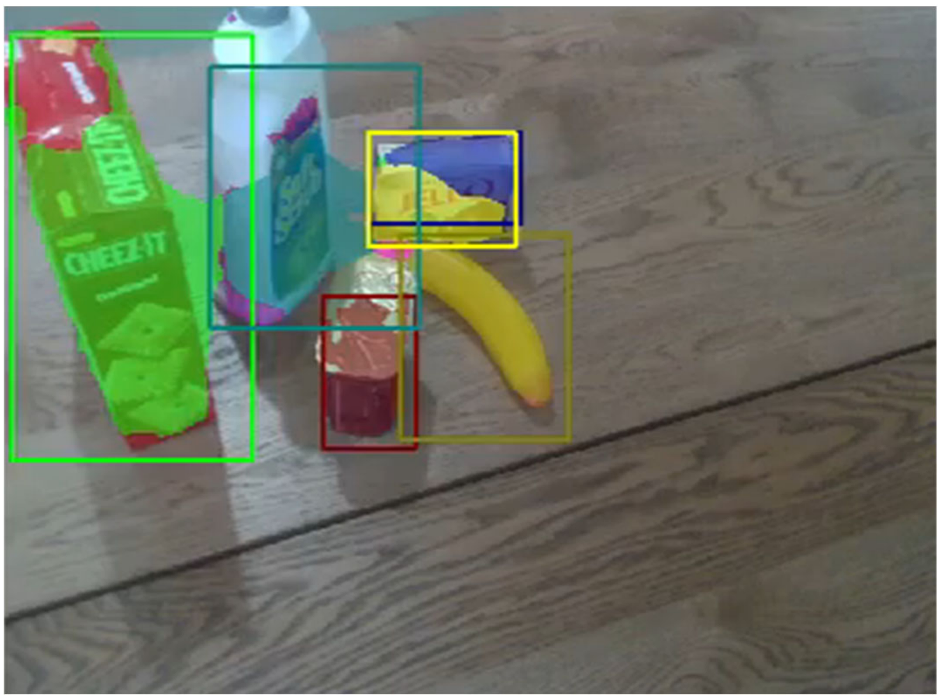
12 robot hours, 497 scenes
6,541 RGB-D images,
22,851 object instances



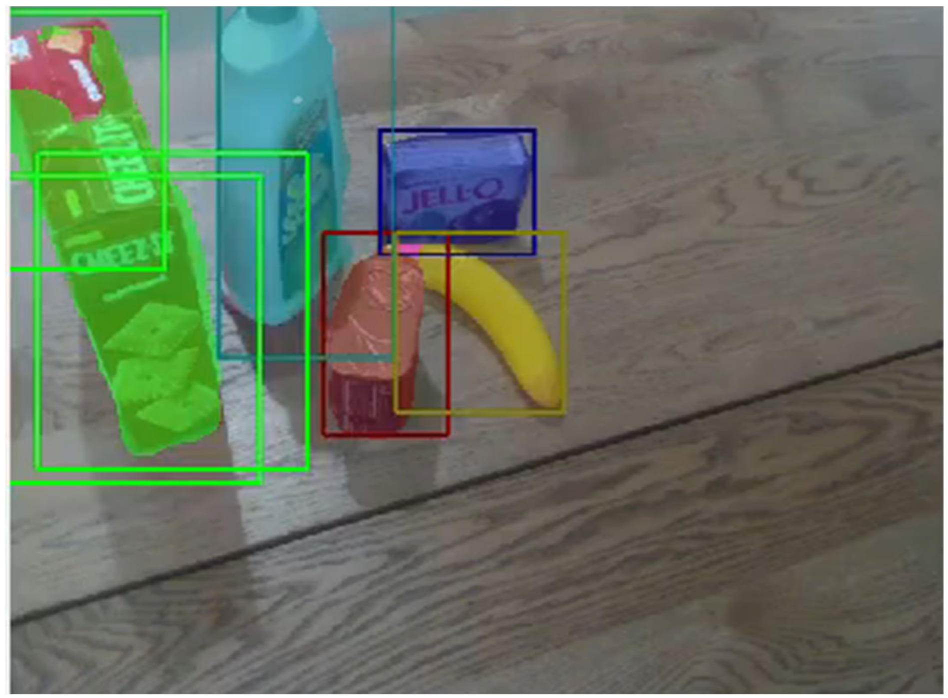

Deng-**Xiang**-Mousavian-Eppner-Bretl-Fox, ICRA'20

# Self-supervised 6D Object Pose Estimation

PoseCNN
trained with only synthetic data

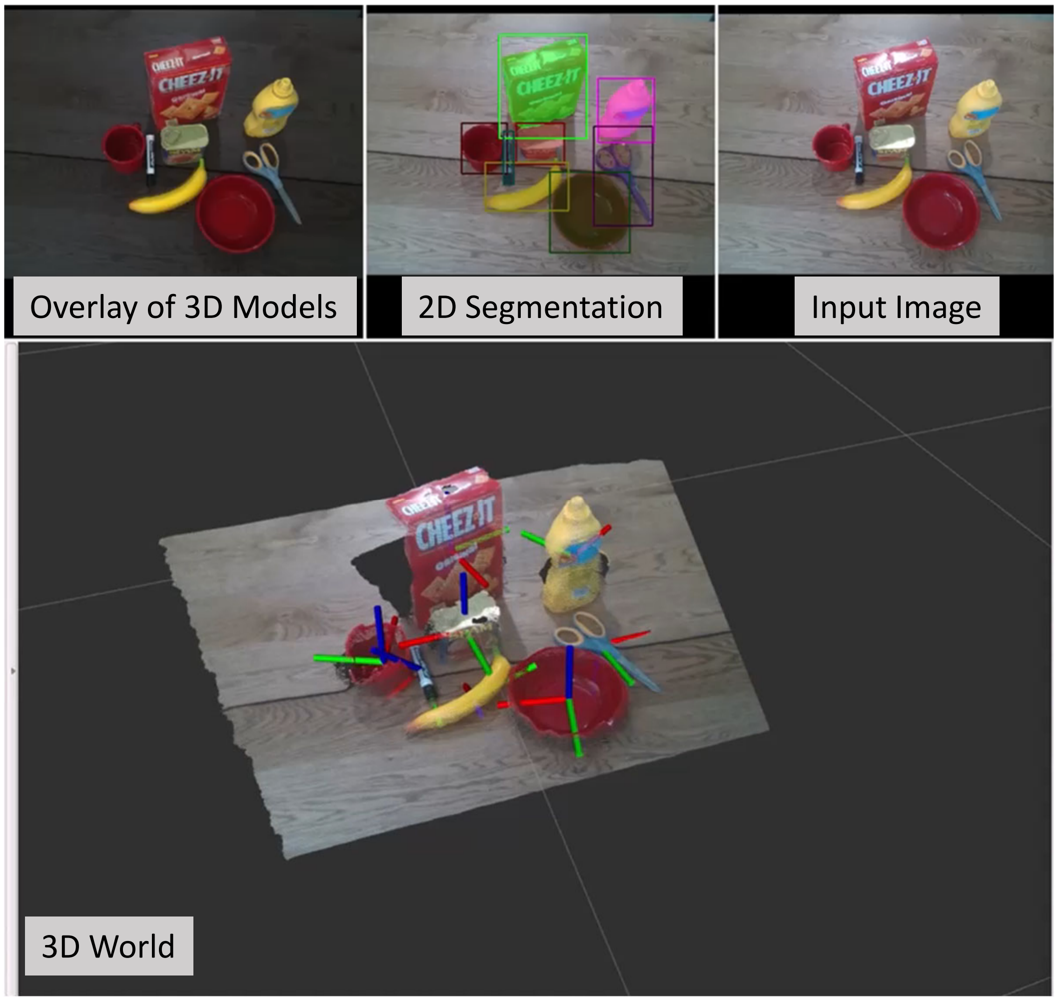


PoseCNN
fine-tuned with self-annotated data




Deng-**Xiang**-Mousavian-Eppner-Bretl-Fox, ICRA’20

# Perception: Model-based 6D Object Pose Estimation




3D models

PoseCNN: **Xiang**-Schmidt-Narayanan-Fox, RSS'18
DeepIM: Li-Wang-Ji-**Xiang**-Fox, ECCV'18 Oral, IJCV'19
PoseRBPF: Deng-Mousavian-**Xiang**-Xia-Bretl-Fox, RSS'19, T-RO'21
Self-supervision 6D Pose: Deng-**Xiang**-Mousavian-Eppner-Bretl-Fox, ICRA'20

Codes available online

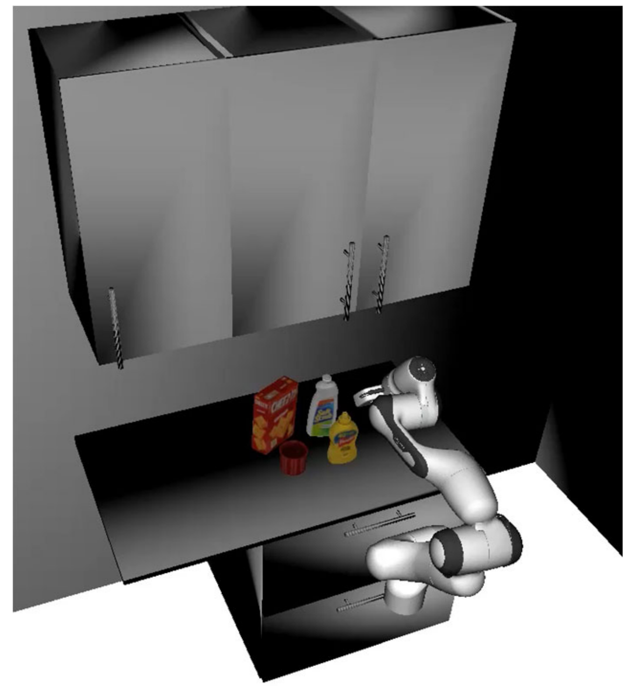
# Manipulation Planning

Input image


6D Object Pose Estimation

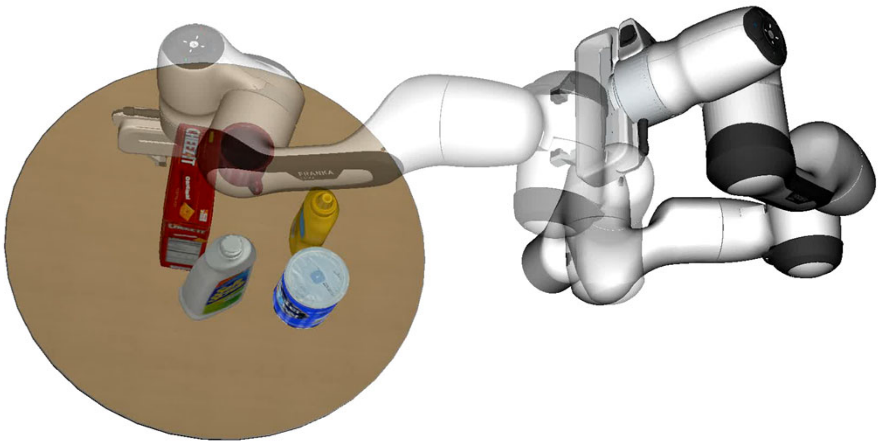

3D models

Planning scene



# Manipulation Planning

## Arm Motion Planning



We need to specify a goal configuration.

## Grasp Planning

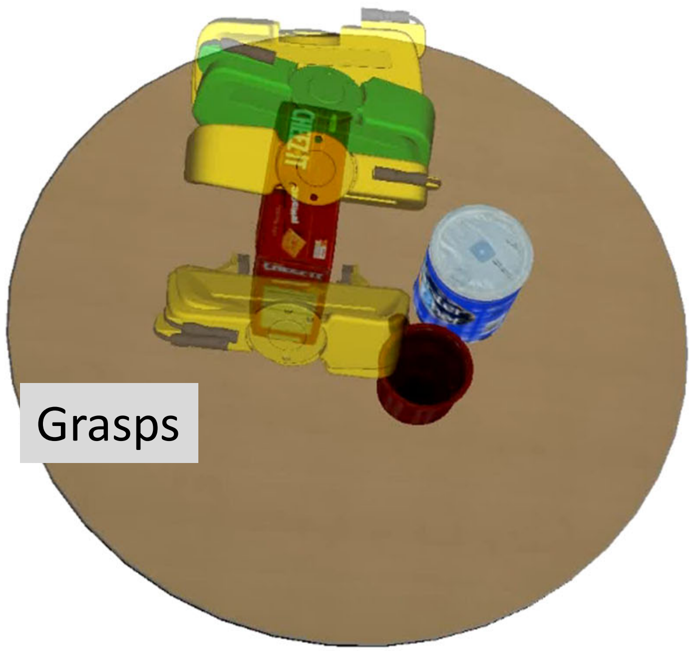
No arm motion is considered.

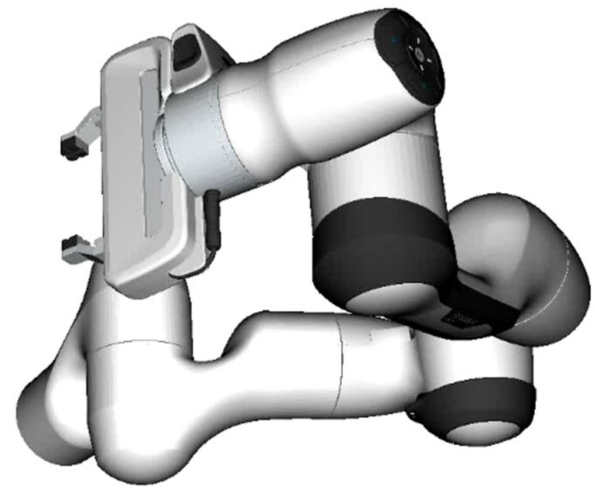
**Sampling-based methods**:
PRM: Kavraki-Svestka-Latombe-Overmars, T-RA’96
RRT: LaValle, Techincal Report’98
RRT-Connect: Kuffner-LaValle, ICRA’00
SMRM: Alterovitz-Simeon-Goldberg, RSS’07
RRT*: Karaman-Frazzoli, IJRR’11
FMT: Janson-Schmerling-Clark-Pavone, IJRR’15

**Trajectory optimization**:
CHOMP: Ratliff-Zucker-Bagnell-Srinivasa, ICRA’09
STOMP: Kalakrishnan-Chitta-Theodorou-Pastor-Schaal, ICRA’11
TrajOpt: Schulman-Duan-Ho-Lee-Awwal-Bradlow-Pan-Patil-Goldberg-Abbeel, IJRR’14
GPMP2: Mukadam-Dong-Yan-Dellaert-Boots, IJRR’18

Nguyen, IJRR’88
Ferrari-Canny, ICRA’92
Chen-Burdick, T-RA’93
Graspit!: Miller-Allen, RA Magazine’04
Ciocarlie-Goldfeder-Allen, RSS Workshop’07
ten Pas-Gualtieri-Saenko-Platt, IJRR’17
Fan-Lin-Tang-Tomizuka, CASE’18
Mousavian-Eppner-Fox, ICCV’19

# OMG Planner: An Optimization-based Motion and Grasp Planner



Arm motion

Joint Motion and Grasp Planning

Wang-**Xiang**-Fox, RSS'20

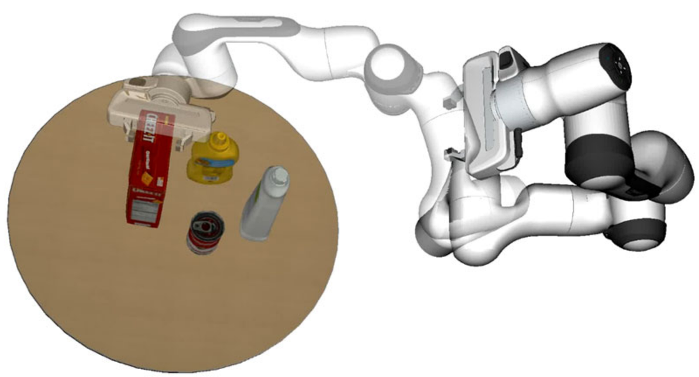
# Trajectory Optimization: CHOMP

$$f_{\text{motion}}(\xi) = f_{\text{obstacle}}(\xi) + \lambda f_{\text{smooth}}(\xi)$$

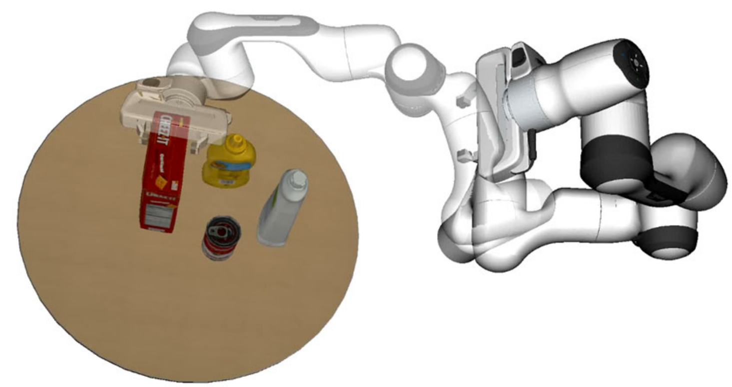
$$\xi = (q_1, \ldots, q_T)$$ A trajectory of robot joint configurations

N steps gradient descent

Initial trajectory with collision

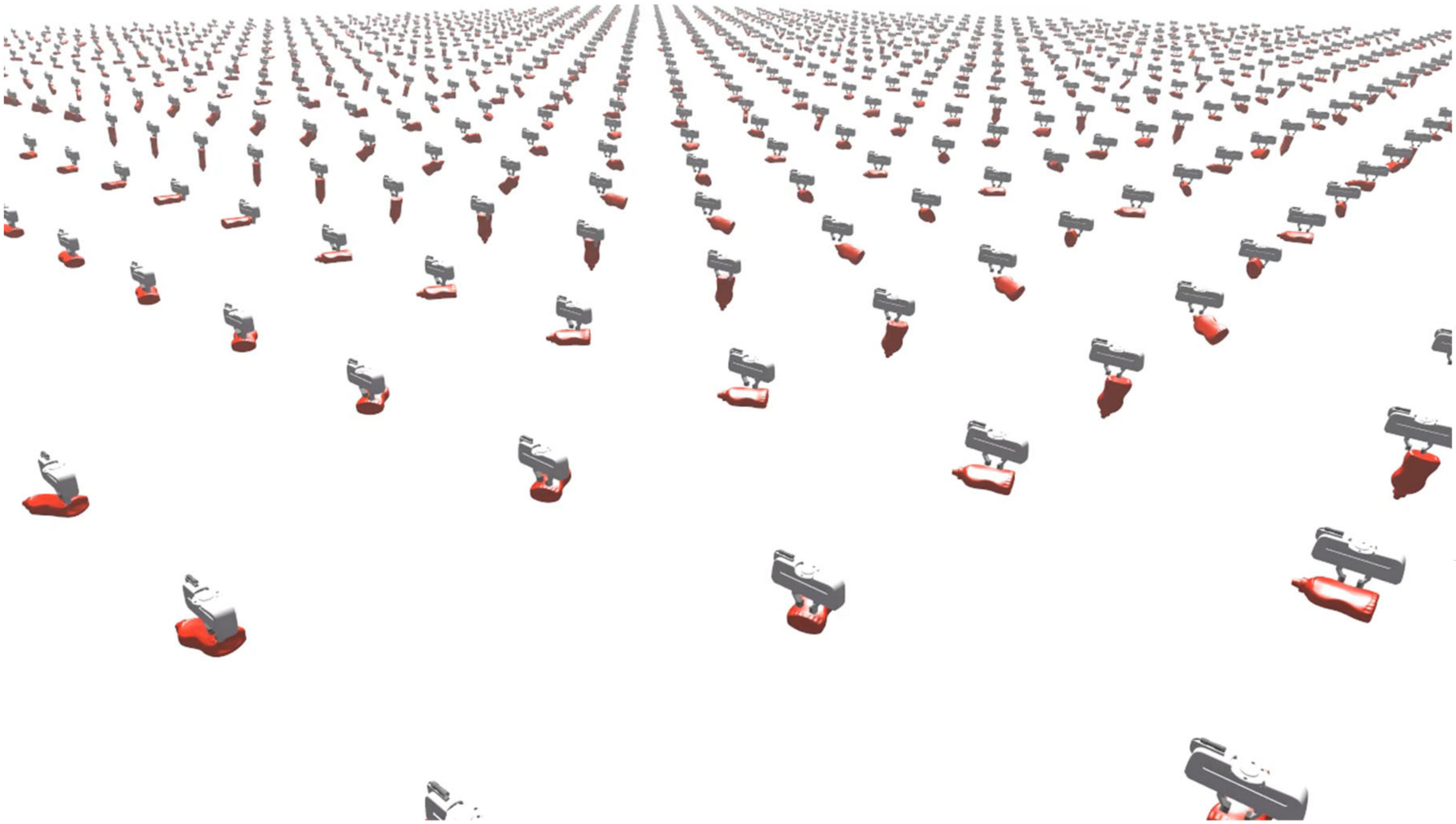


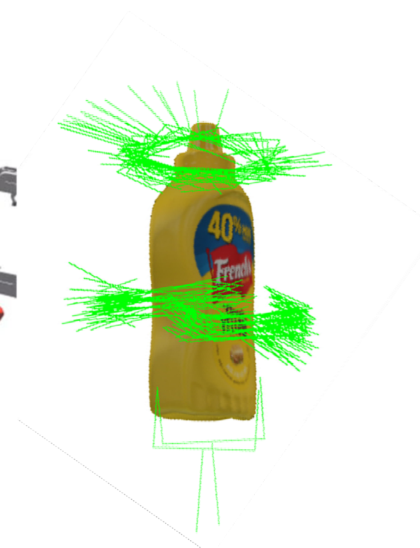
Final trajectory



Covariant Hamiltonian Optimization for Motion Planning (CHOMP): Ratliff-Zucker-Bagnell-Srinivasa, ICRA'09

# Grasp Planning: A Physics-based Approach



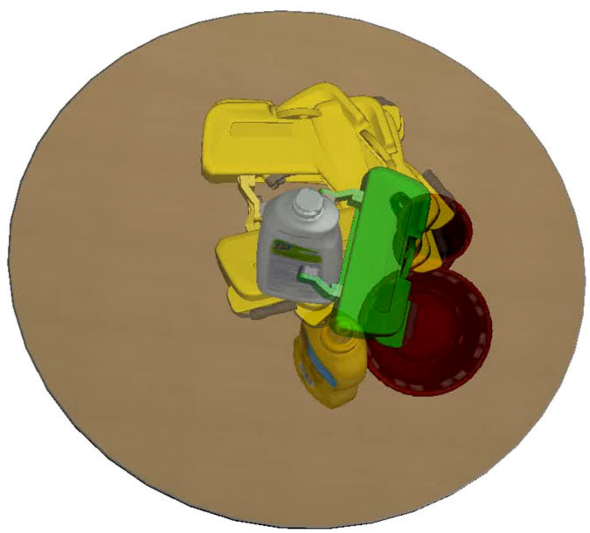


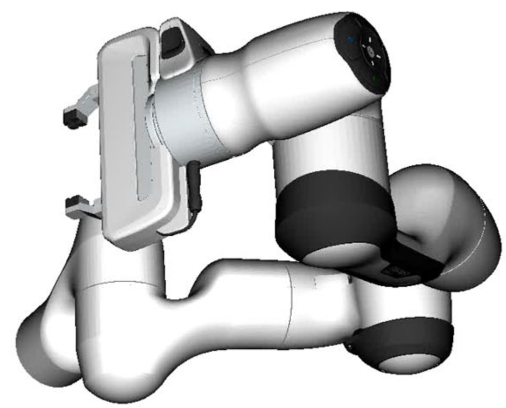
Eppner-Mousavian-Fox, ISRR’19

# OMG Planner: Trajectory Optimization and Grasp Selection

OMG Iter: 50

100 grasps



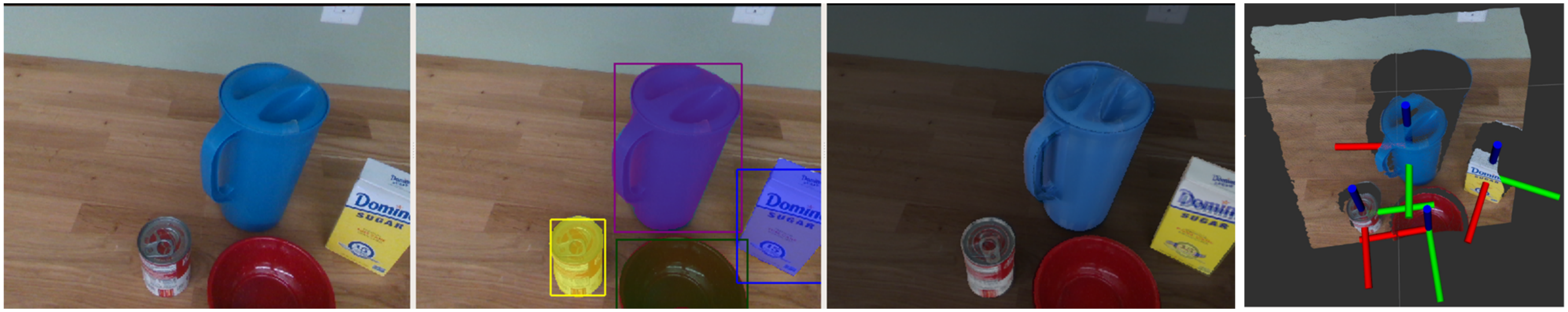


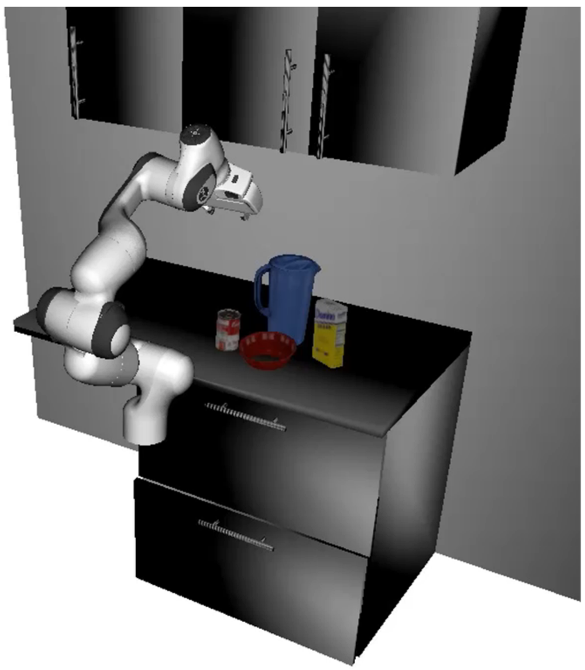
Modeling the goal set distribution

Code available online

Wang-**Xiang**-Fox, RSS’20

# Real-world Manipulation with 6D Pose Estimation and Planning

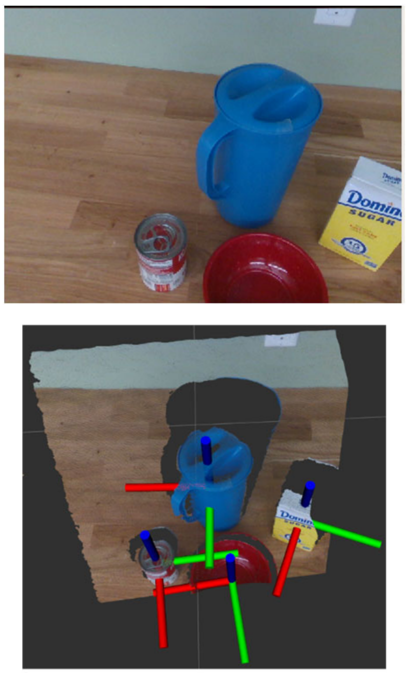




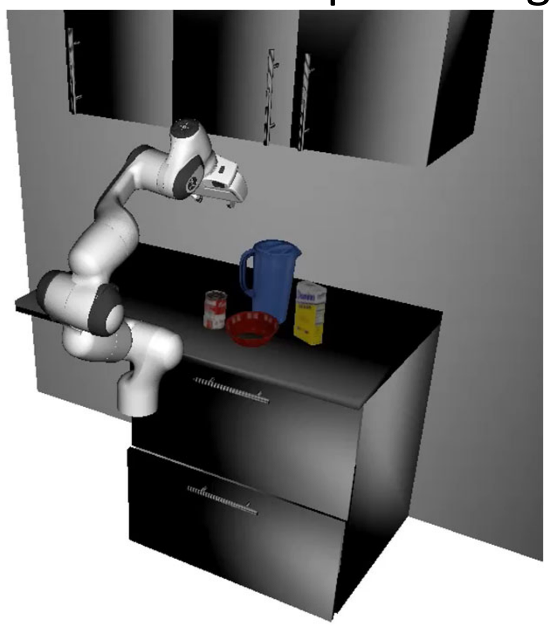

PoseCNN realsense
PoseRBPF realsense
PoseCNN realsense
PoseRBPF realsense
4X

# Model-based Robot Manipulation

6D Object Pose Estimation



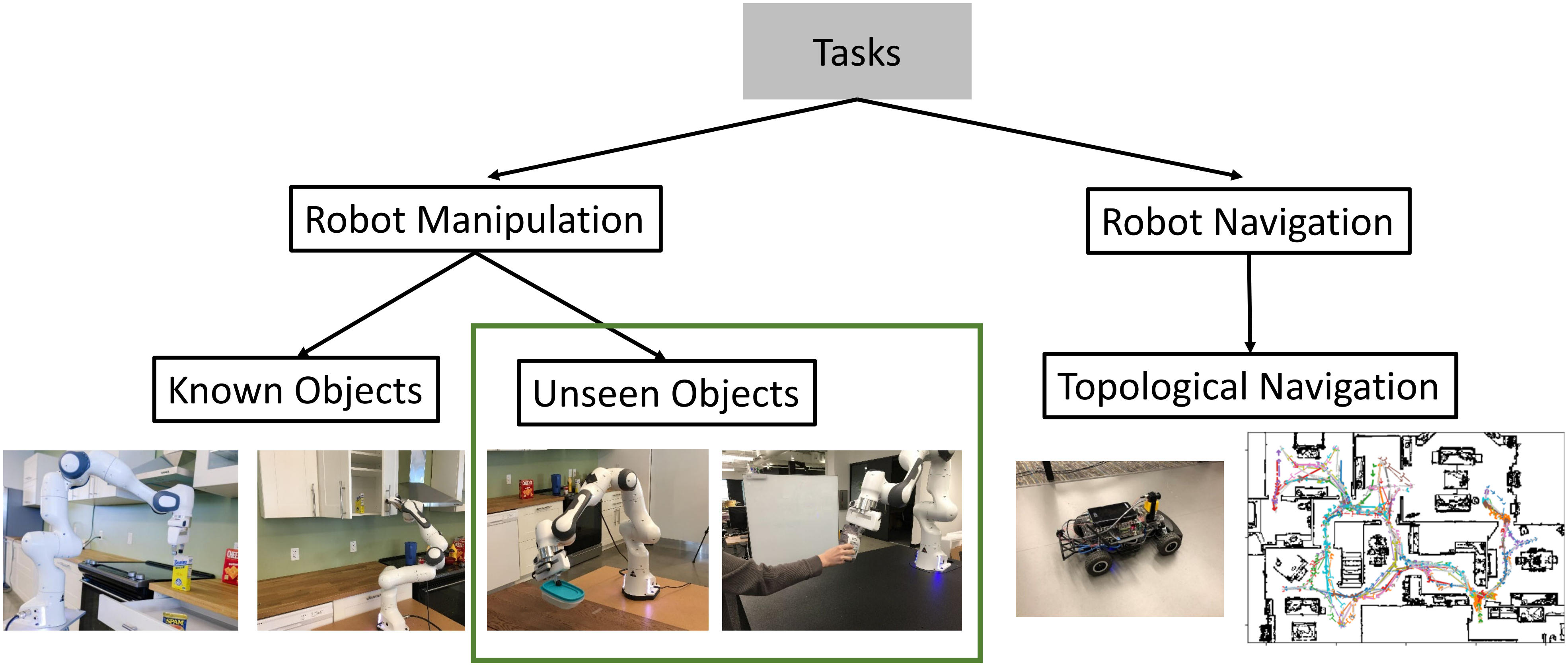
Motion and Grasp Planning



We need to have 3D models of objects

How can we enable robots to manipulate unseen objects?

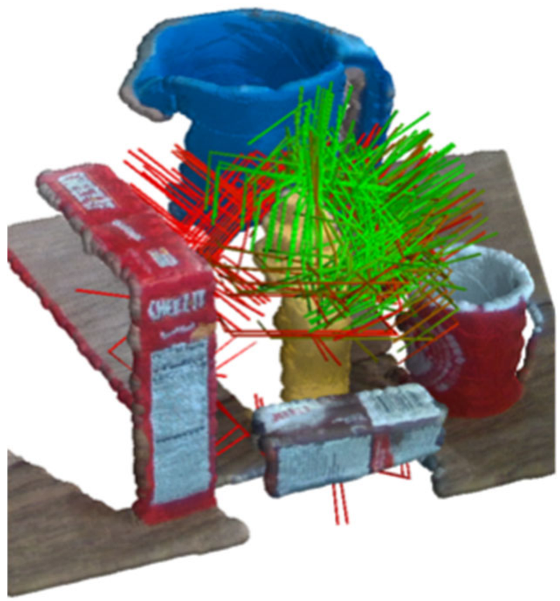
# Outline


Tasks
Robot Manipulation
Robot Navigation
Known Objects
Unseen Objects
Topological Navigation

# Model-free Robot Manipulation

Perception → Planning → Control


Unseen object instance segmentation



Grasp planning from point clouds


Position control to reach grasp

Figure Credit: Murali-Mousavian-Eppner-Paxton-Fox, ICRA'20

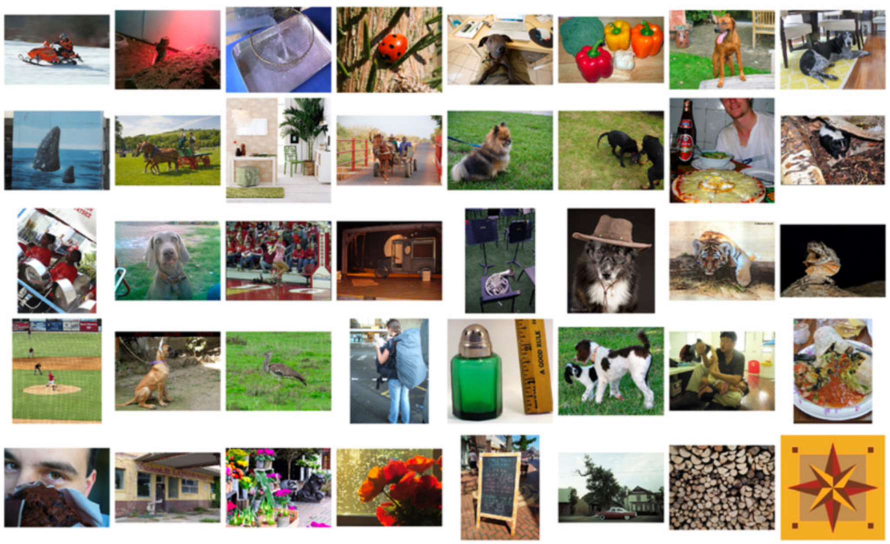
# Perception: Unseen Object Instance Segmentation

Xie-**Xiang**-Mousavian-Fox, CoRL'19, T-RO'21
**Xiang**-Xie-Mousavian-Fox, CoRL'20
Training on synthetic data, transferring well to the real images for segmenting unseen objects

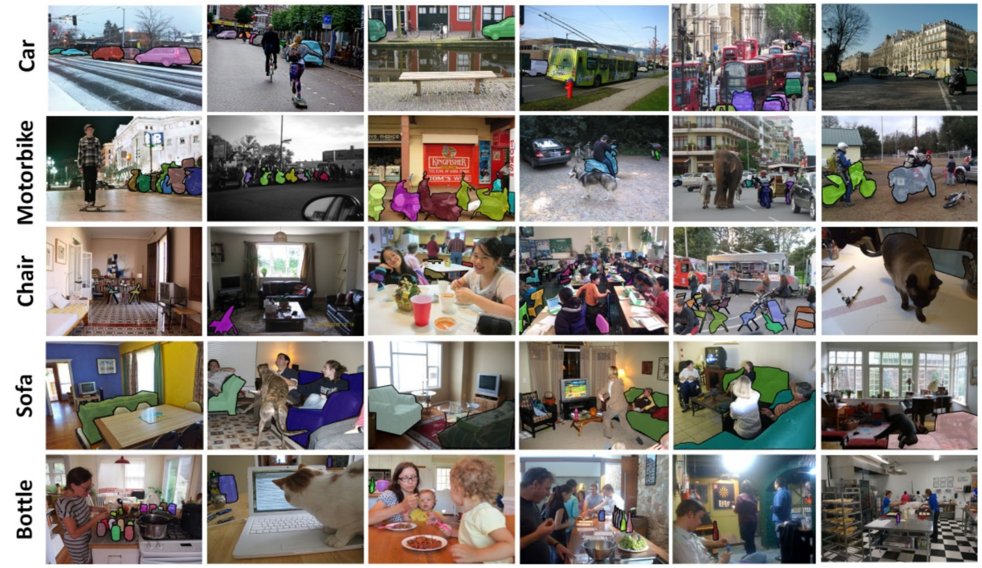
Codes available online

# Learning the Concept of "Objects"

- Learning from data



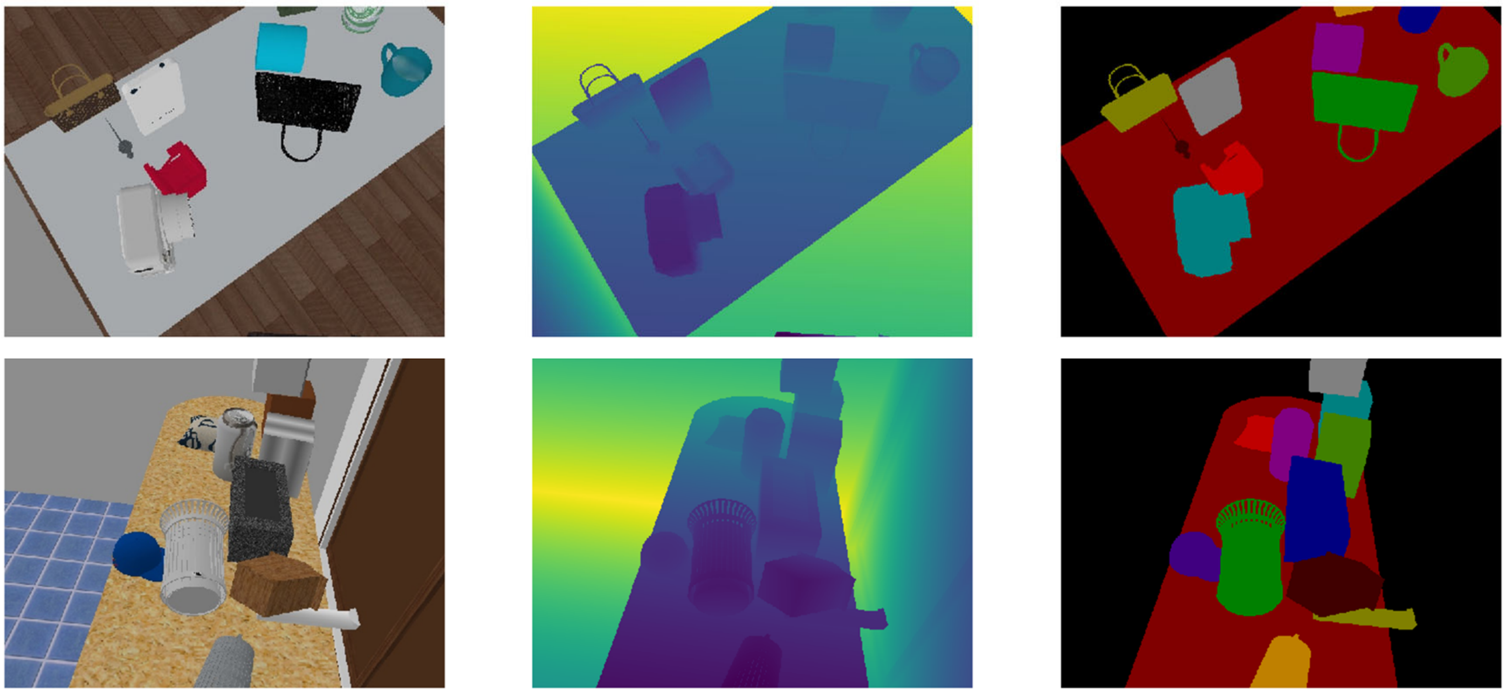
ImageNet: Deng-Dong-Socher-Li-Li-Fei-Fei, CVPR'09




COCO: Lin-Maire-Belongie-Bourdev-Girshick-Hays-Perona-Ramanan-Zitnick-Dollar, ECCV'14

Internet Images, not suitable for indoor robotic settings

# Learning from Synthetic Data



RGB

Depth

Instance Label

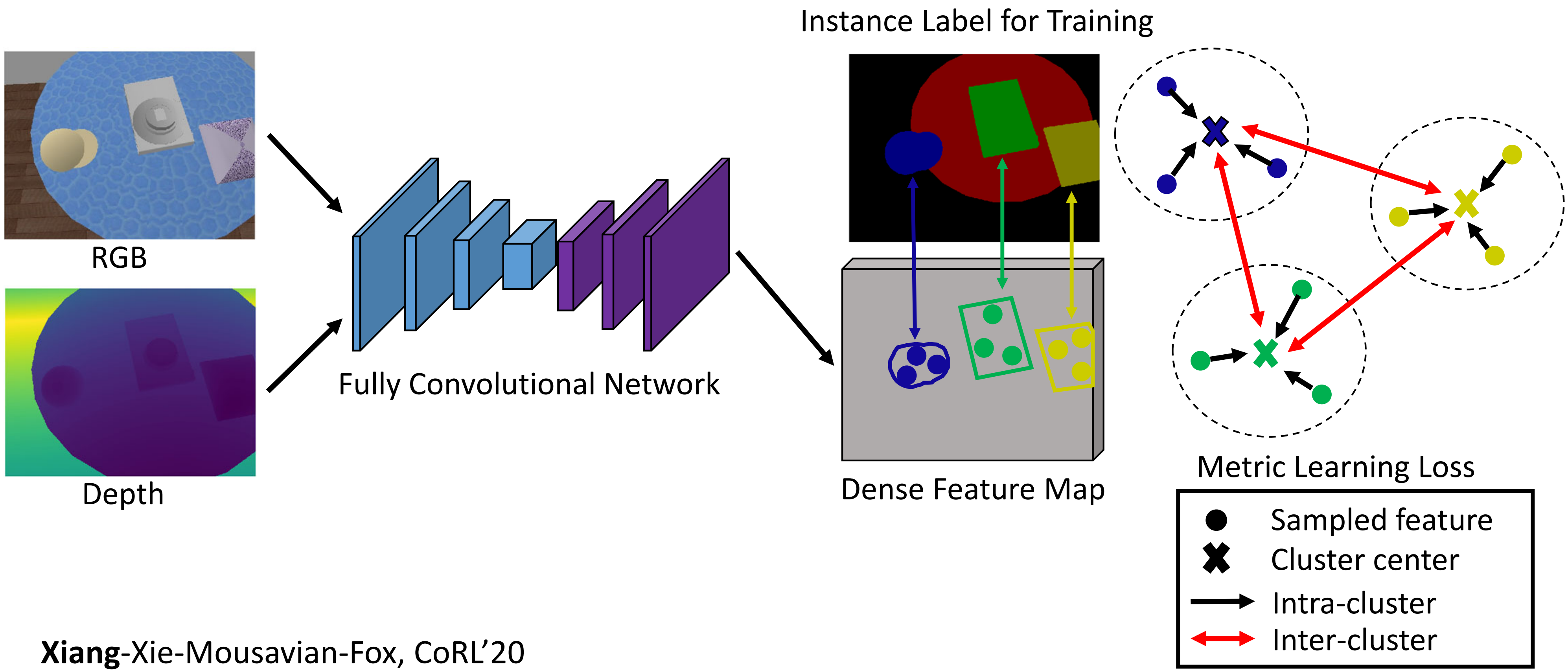
ShapeNet objects in the PyBullet simulator

40,000 scenes
7 RGB-D images per scene

Need to deal with the sim-to-real gap

Tabletop Object Dataset: Xie-**Xiang**-Mousavian-Fox, CoRL'19

# Unseen Object Instance Segmentation: Learning RGB-D Feature Embeddings




**Xiang**-Xie-Mousavian-Fox, CoRL'20

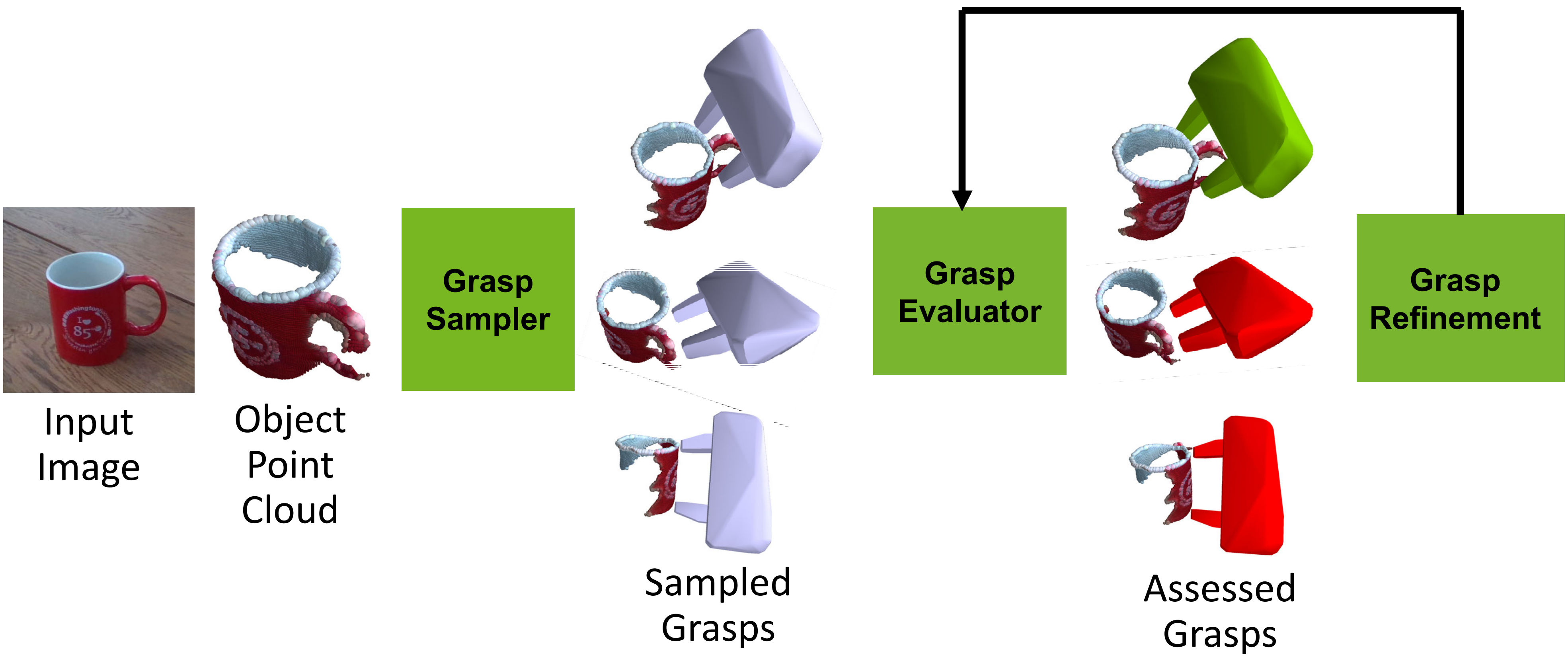
Input Image

Feature Map

Output Label


**Xiang**-Xie-Mousavian-Fox, CoRL'20

# Grasp Planning from Partially Observed Point Clouds




6-DOF GraspNet: Mousavian-Eppner-Fox, ICCV'19

# Grasping Unseen Objects



Unseen Object Instance Segmentation:
Xie-**Xiang**-Mousavian-Fox, CoRL'19, T-RO'21
**Xiang**-Xie-Mousavian-Fox, CoRL'20

6-DOF GraspNet:
Mousavian-Eppner-Fox, ICCV'19

# Open-Loop VS. Closed-Loop

Tasks

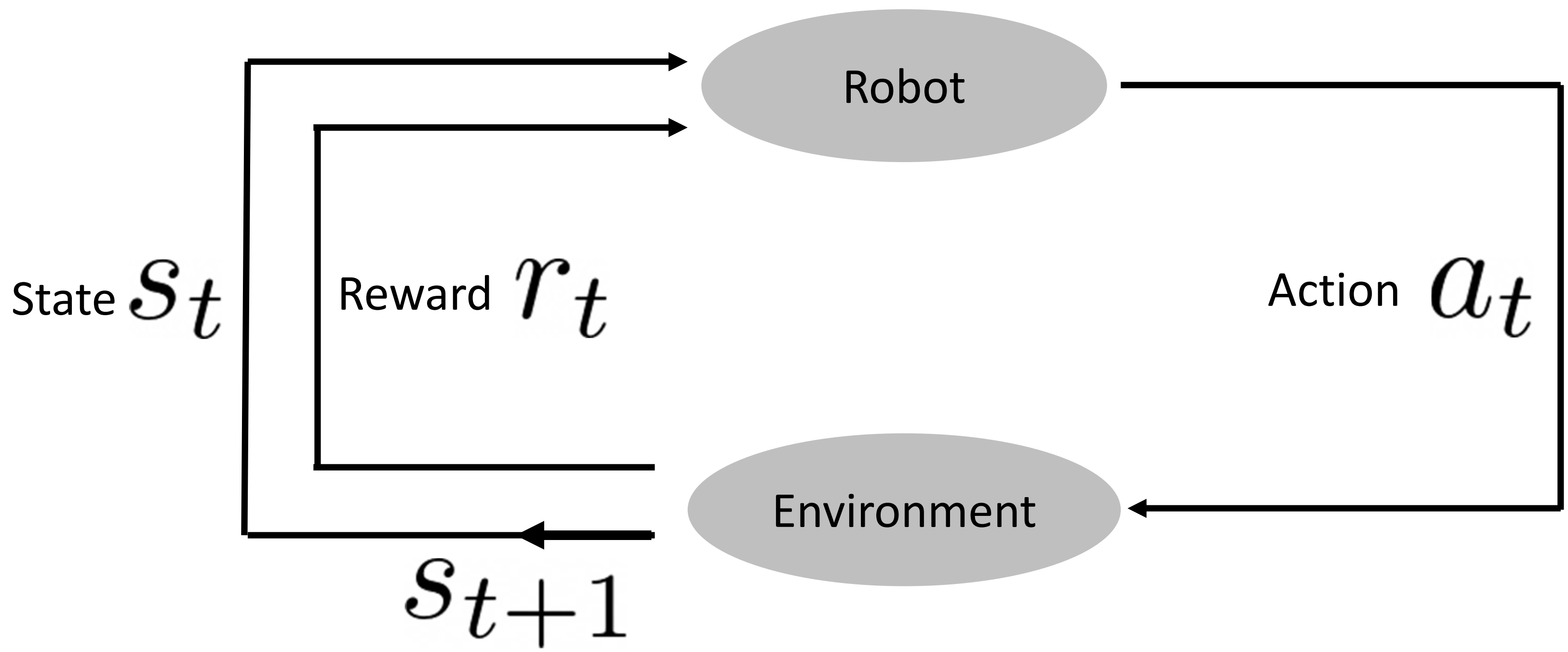
Perception → Planning → Control

Action

World

Sensing

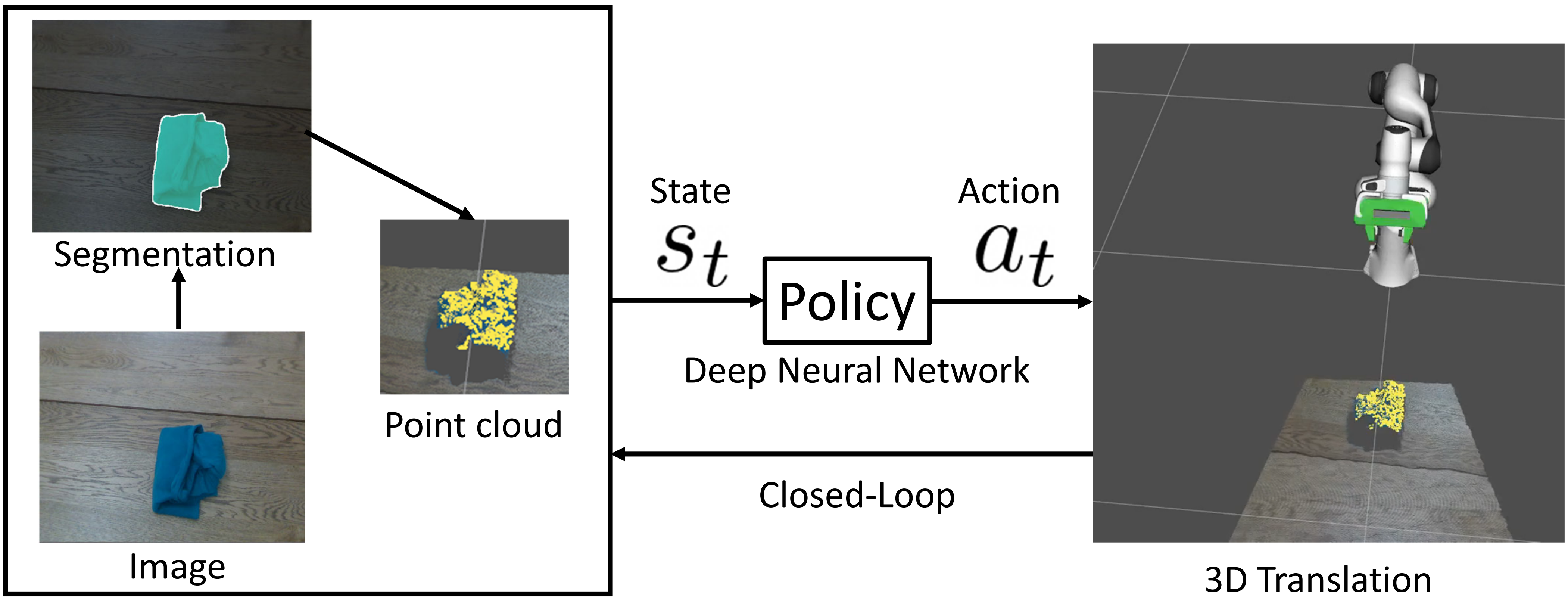
# Closed-loop Robot Control with Markov Decision Processes




Reinforcement Learning:
Imitation Learning:

$$a_t = \pi(s_t)$$

# Learning Closed-Loop Control Polices for 6D Grasping




3D Translation
3D Rotation

Perception

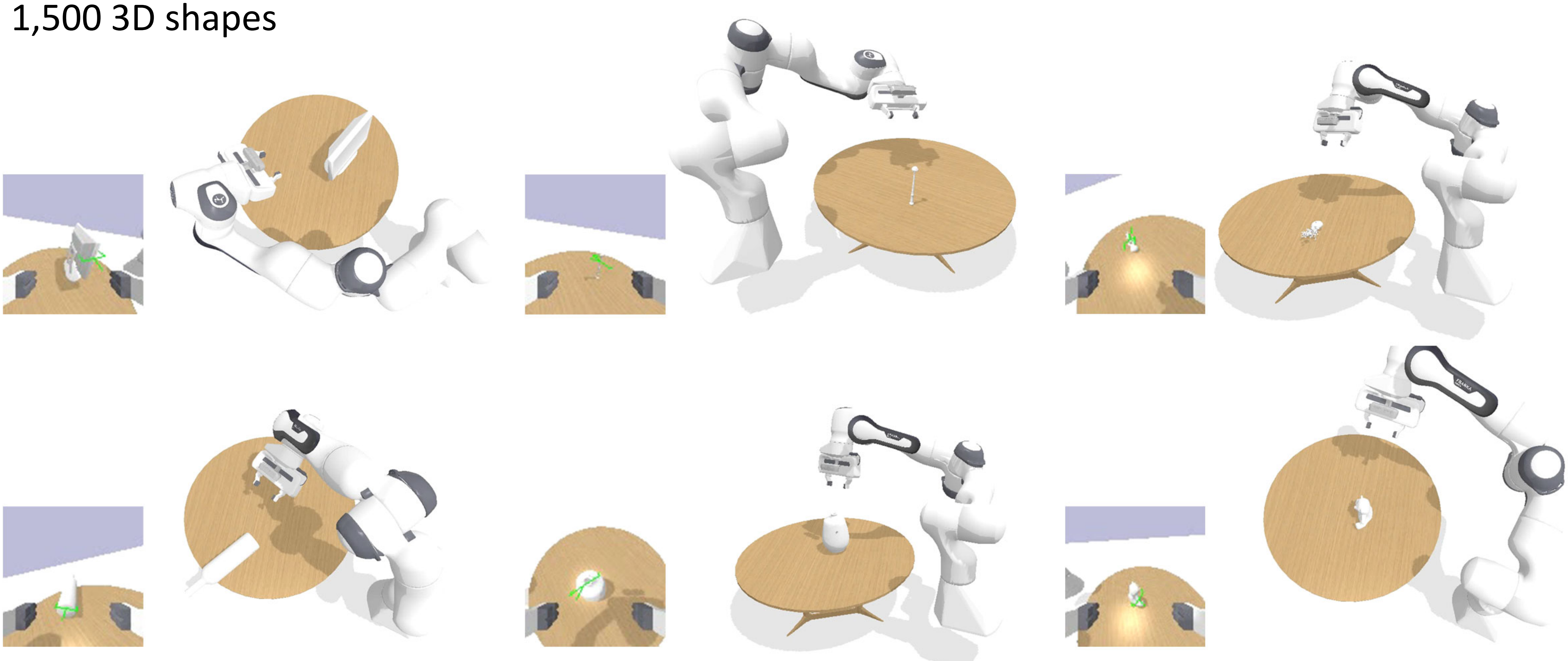
Control

Wang-**Xiang**-Fox, in arXiv'21

No planning?

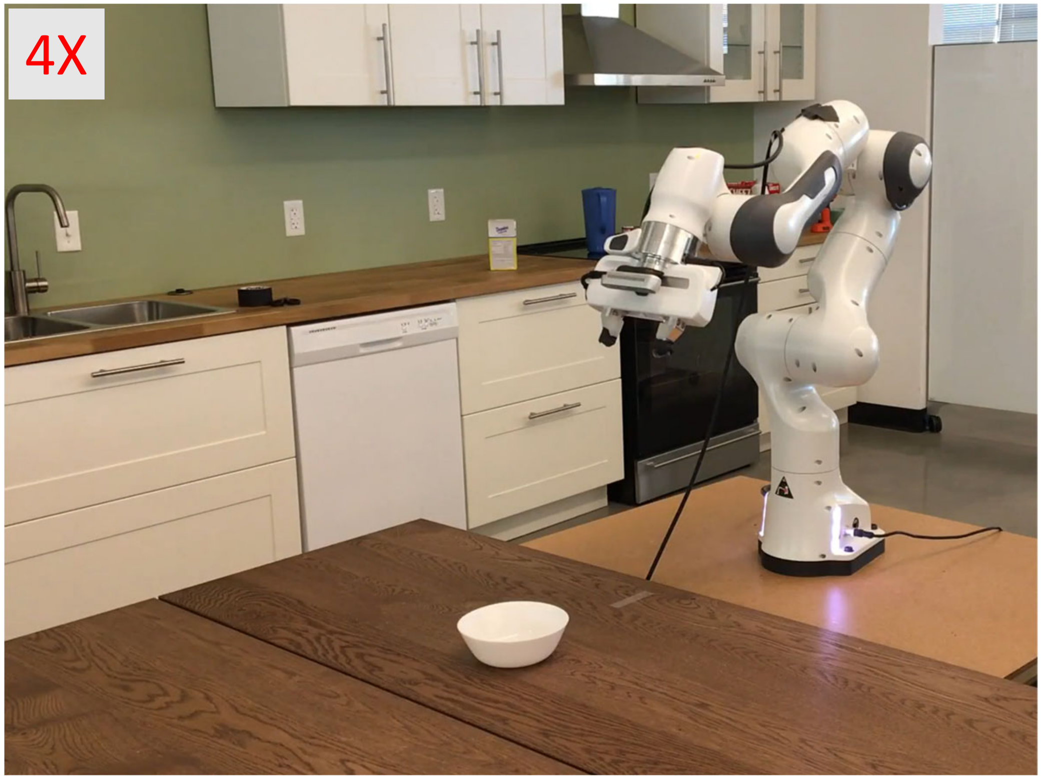
# Learning from Demonstration with the OMG-Planner

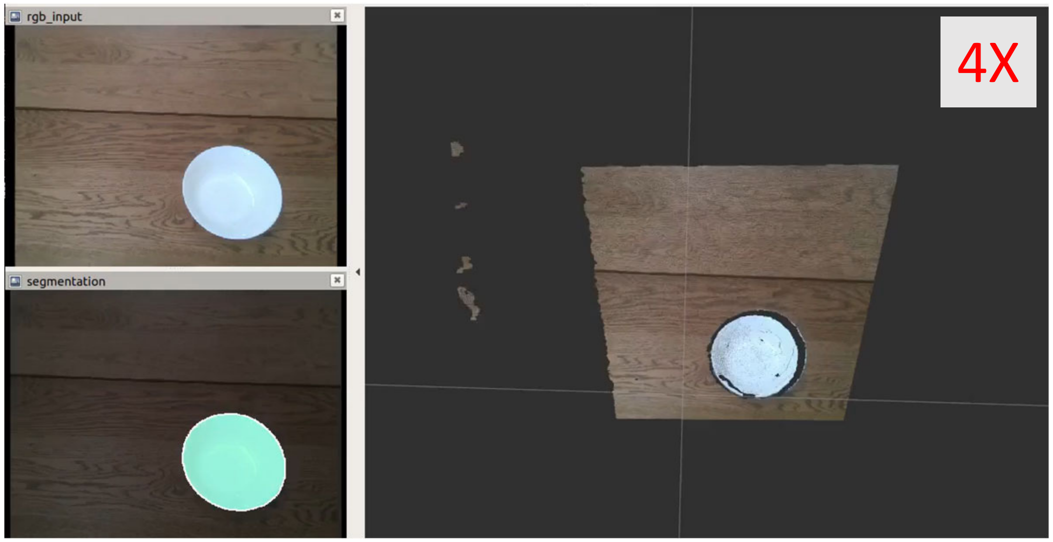
50,000 trajectories
1,500 3D shapes



Wang-**Xiang**-Fox, in arXiv'21

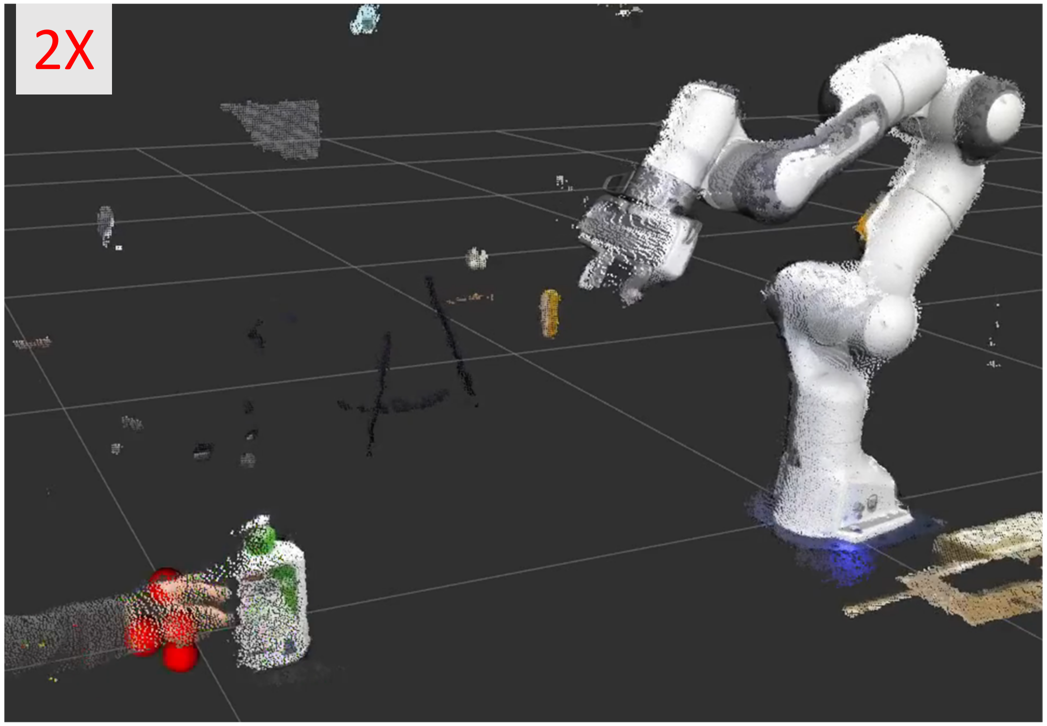
# Our Learned Policy in the Real World



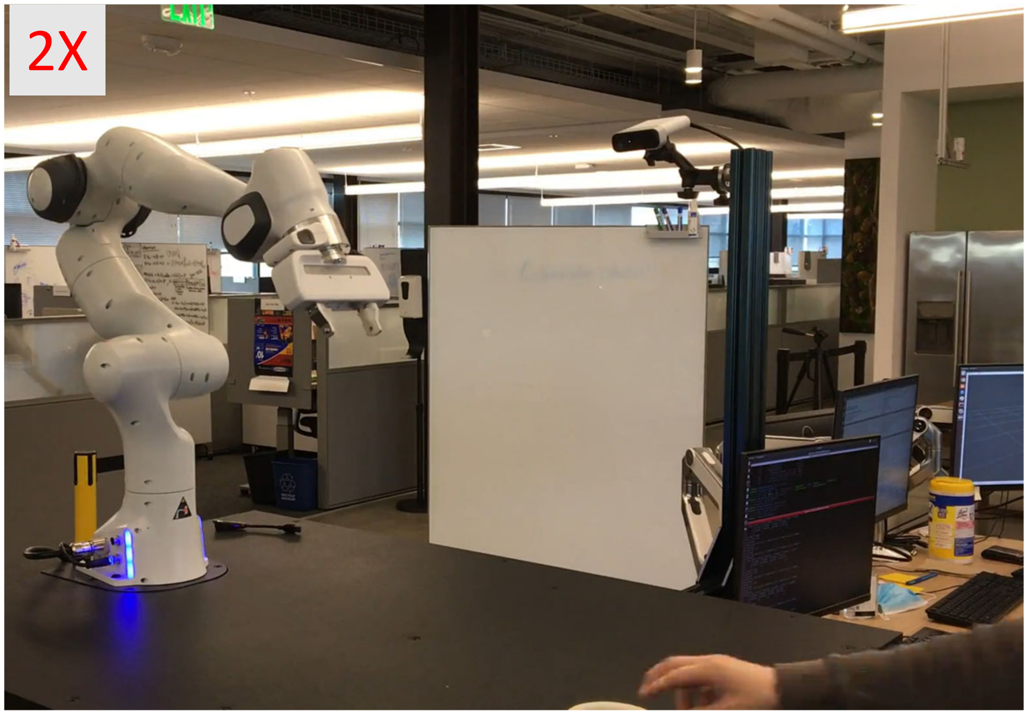

Wang-**Xiang**-Fox, in arXiv'21

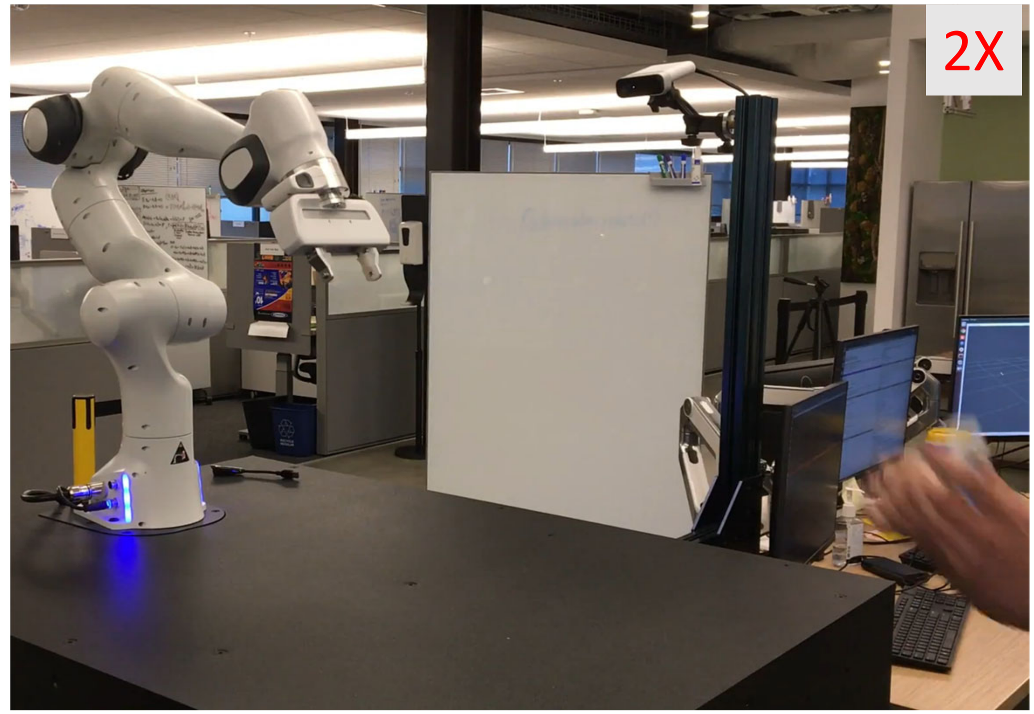
# Closed-Loop Human-Robot Handover


Yang-Paxton-Mousavian-Chao-Cakmak-Fox, in arXiv’20
Wang-**Xiang**-Fox, in arXiv’21

# Closed-Loop Human-Robot Handover



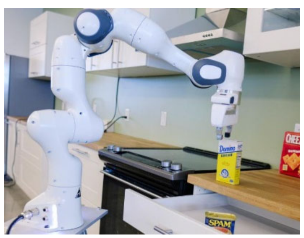

Yang-Paxton-Mousavian-Chao-Cakmak-Fox, in arXiv'20
Wang-**Xiang**-Fox, in arXiv'21

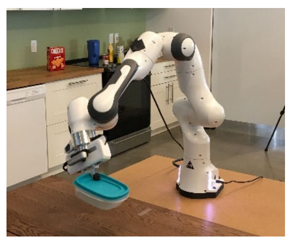
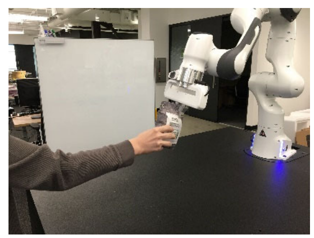
# Manipulation and Navigation

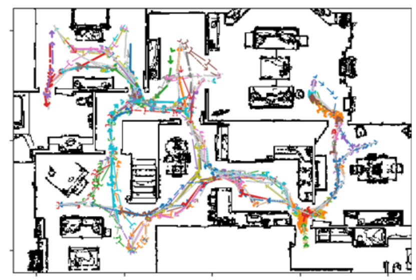
# Outline

Tasks

- Robot Manipulation
  - Known Objects
  - Unseen Objects
- Robot Navigation
  - Topological Navigation





# Traditional Robot Navigation

**Perception** → **Planning** → **Control**

Simultaneous localization and mapping (SLAM)

Path planning

Path following


Laser-based SLAM
2D occupancy grid map

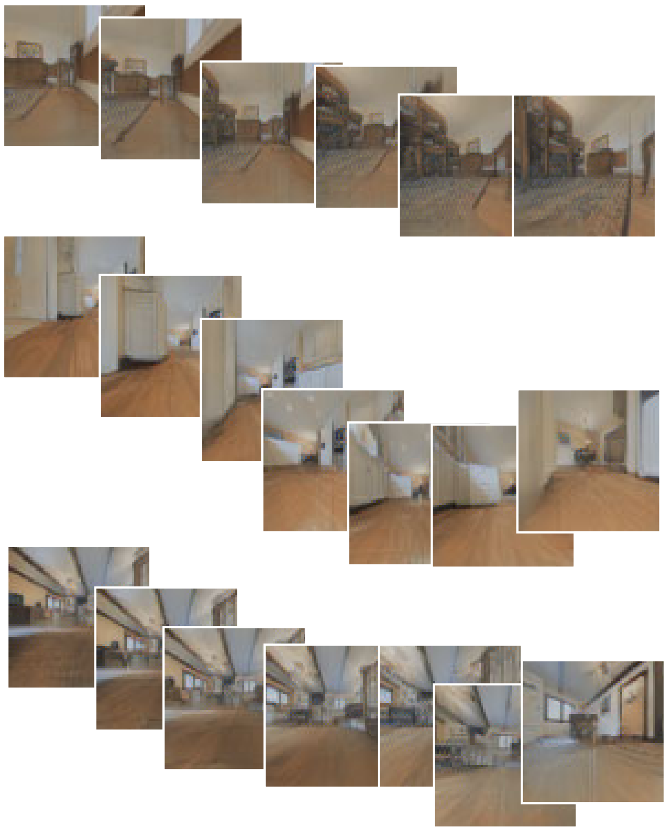
Limitations of SLAM-based navigation

- 3D reconstruction is expensive
- Detailed 3D geometry information may not be necessary

# Topological Navigation

Meng-Ratliff-**Xiang**-Fox, ICRA'19, '20
Meng-**Xiang**-Fox, RA-L'21

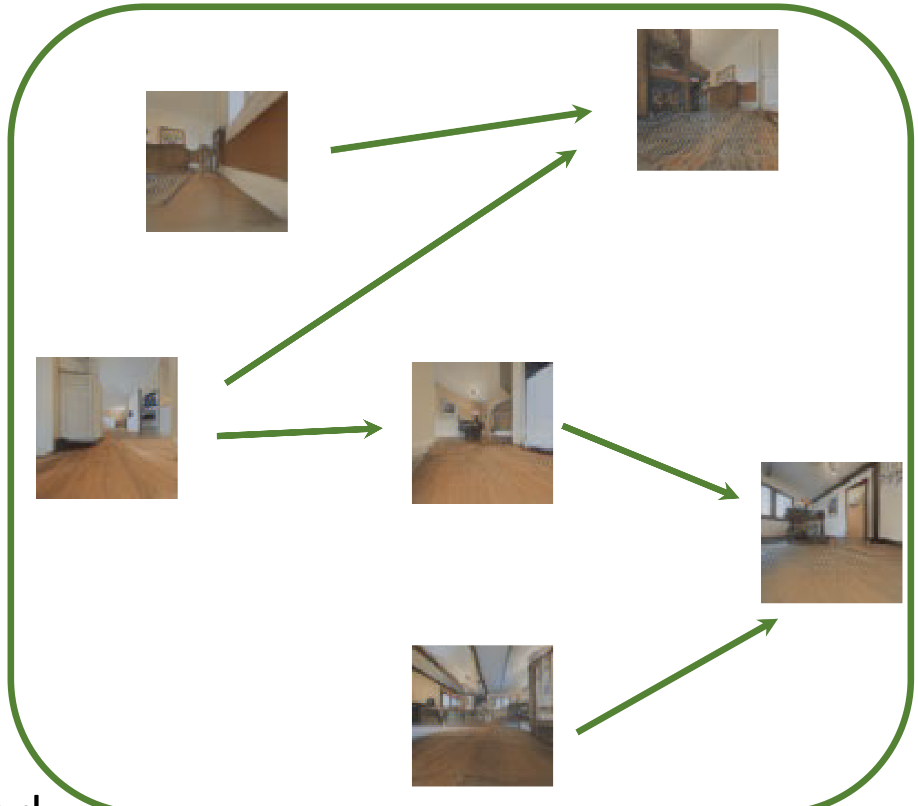
Dense Trajectories



Reachability
Estimator

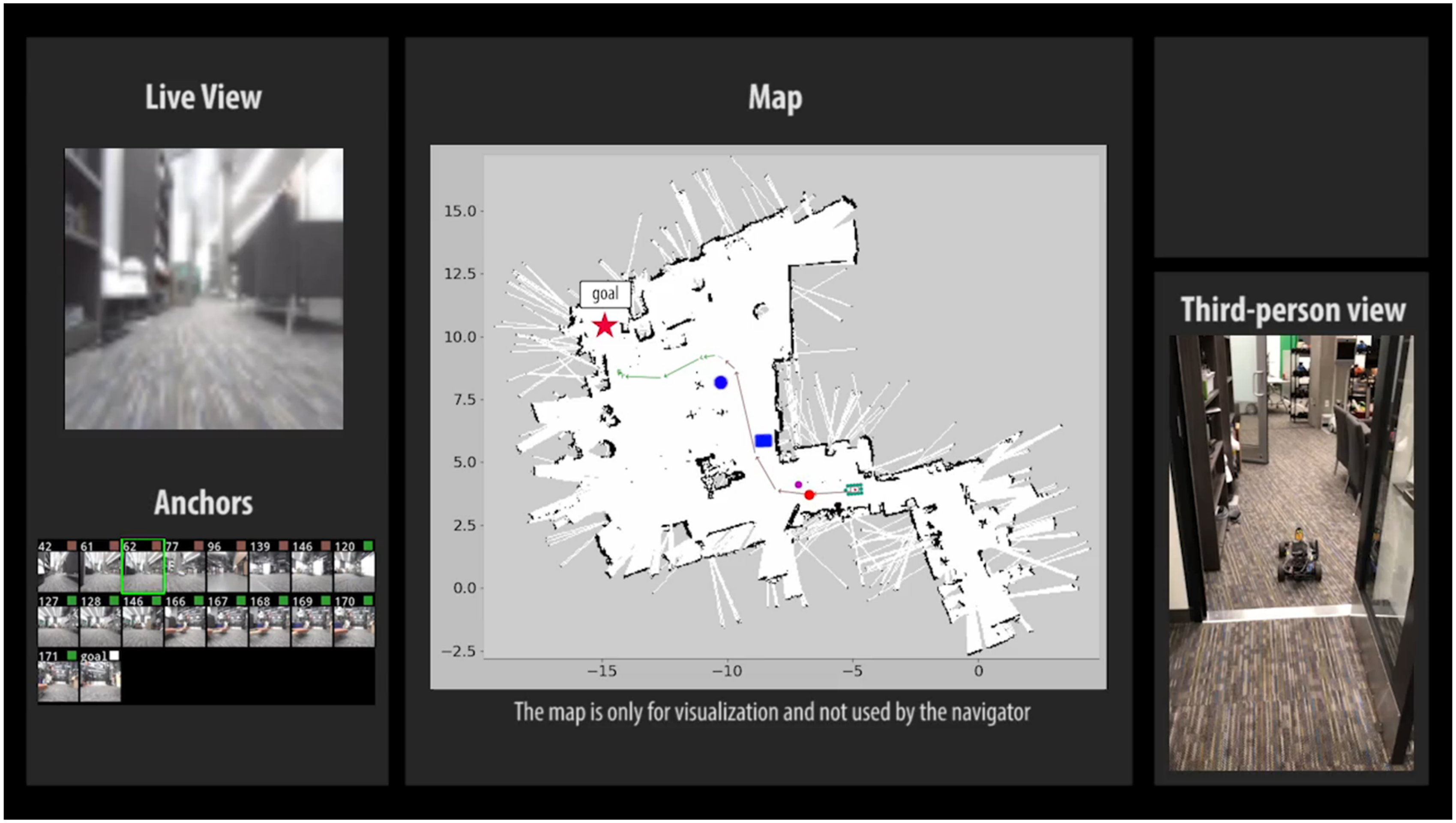
Local Controller
A learned neural network

Sparse Topological Map



# Topological Navigation

Meng-Ratliff-**Xiang**-Fox, ICRA'19, '20
Meng-**Xiang**-Fox, RA-L'21



# Summary

Tasks

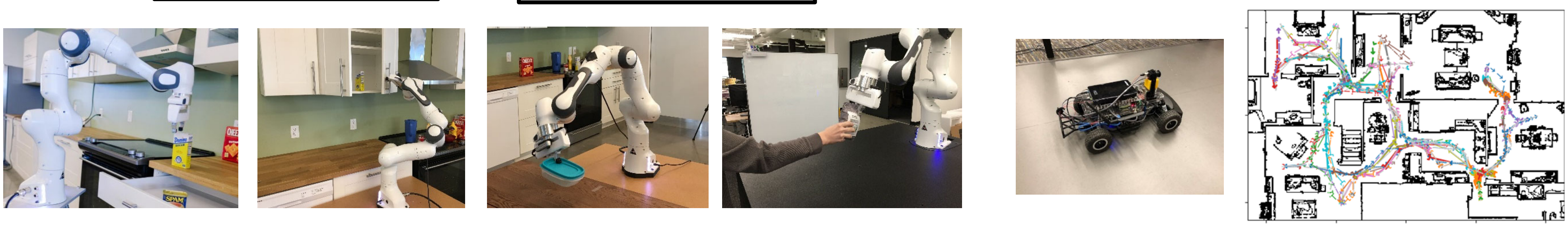
Robot Manipulation

Robot Navigation

Known Objects

Unseen Objects

Topological Navigation



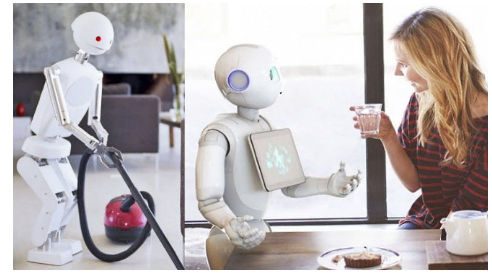
Perception

Planning

Control

Learning

# Future Work: Long-horizon Tasks in Human Environments



Intelligent Robots

- Make a cup of coffee
- Set a dining table
- Execute human instructions
  - “Bring a bottle of water”

Multiple Tasks

Task Diversity

Generalizability

Single Task

Industrial robots

- Manipulation
- Navigation

Single Environment

Environment Diversity

Multiple Environments

# Future Work: Learning Robot Skills and Building Robotic Systems

## Robot Skills Generalizable and Shareable

**Perception**
- Understand objects, scenes and space
- Understand humans and language

**Planning**
- Task planning
- Motion planning

**Control**
- Learning task-specific controllers

**Learning**
- Supervised Learning
- Imitation Learning
- Reinforcement Learning

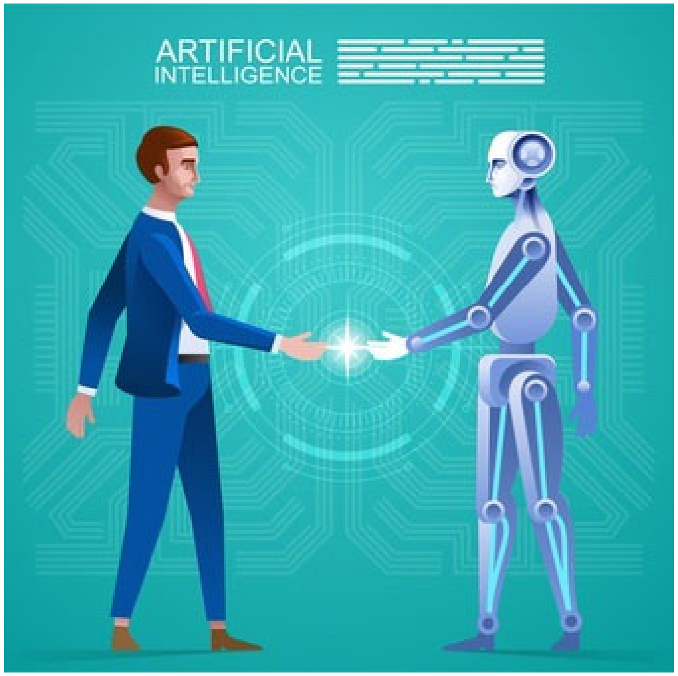
Deploy ↓ ↑ Improve

## Robotic Systems

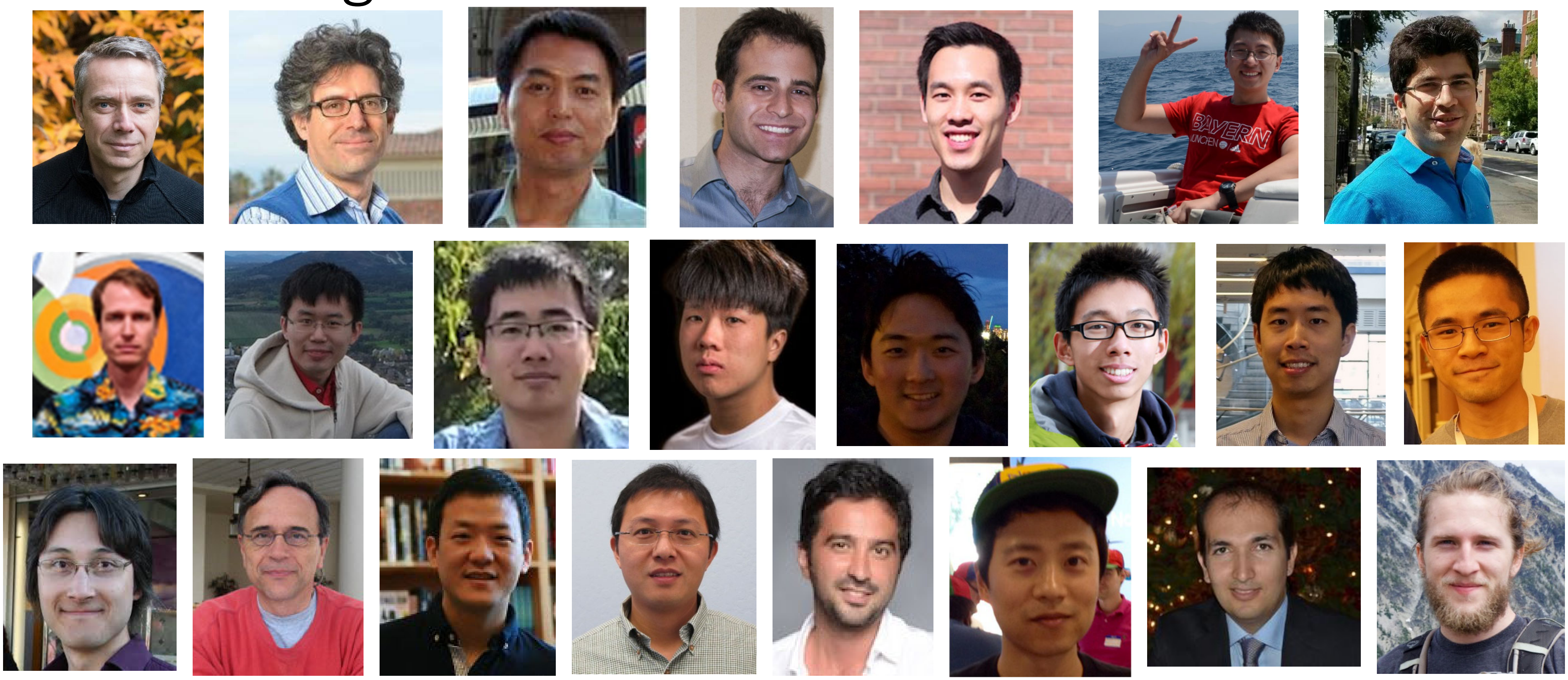
- Closing the perception, planning and control loop
- Self-supervised learning
- Life-long learning

# Our Missions of the Future Research Lab

- Advancing robot perception, planning and control
- Building intelligent robotic systems
- Open-sourcing and sharing
- Collaborating



# Acknowledgements




## Thank you!