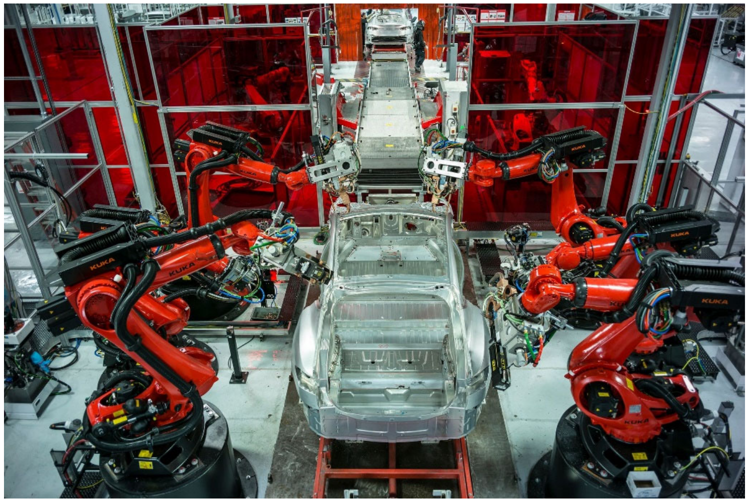
6D Robotic Grasping of Unseen Objects

Yu Xiang

Assistant Professor

The University of Texas at Dallas

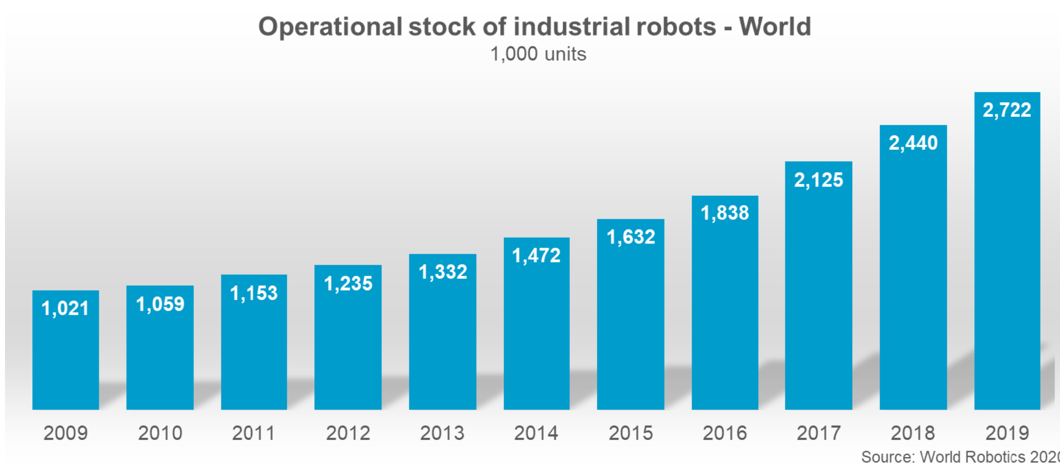
Robots in Factories and Warehouses



Welding and Assembling

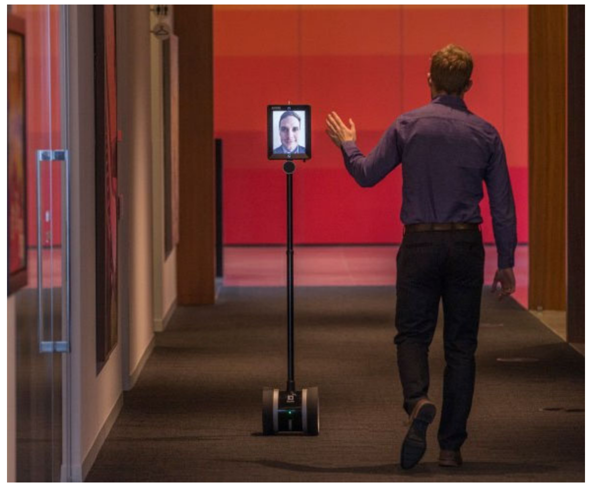
Material Handling

Delivering



Current Robots in Human Environments

Cleaning Robots



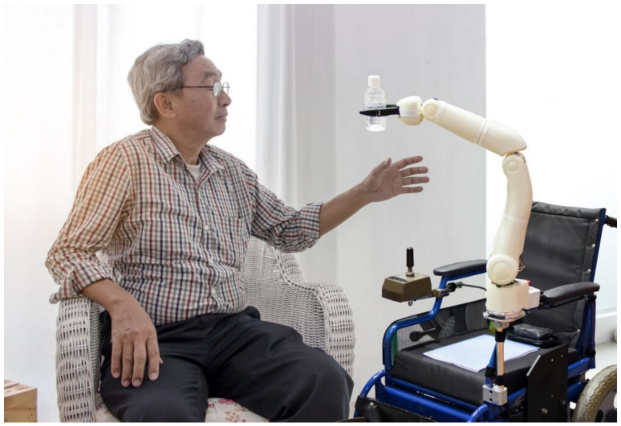
Telepresence Robots

Smart Speakers

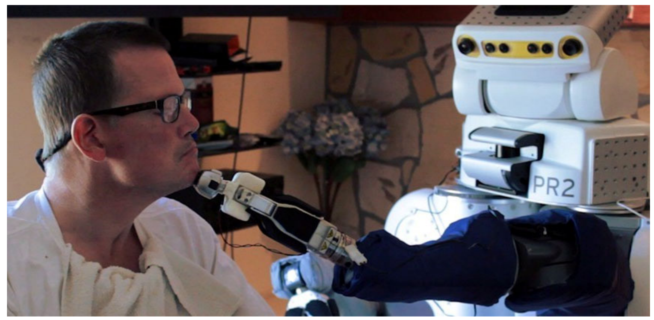
How can we have more powerful robots assisting people at homes or offices?

- Mobile manipulators
- Humanoids

Future Intelligent Robots in Human Environments



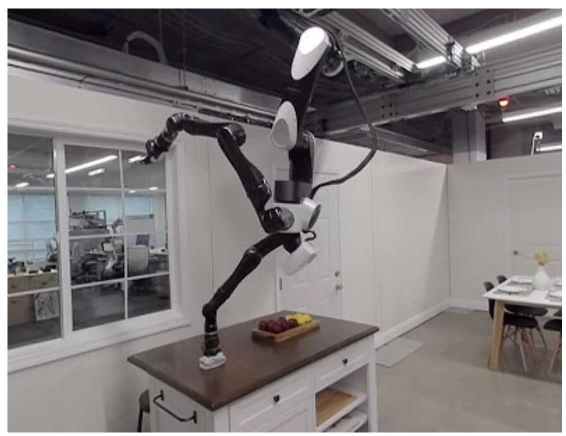
Senior Care



Assisting

Serving

Cooking



Cleaning

Dish washing

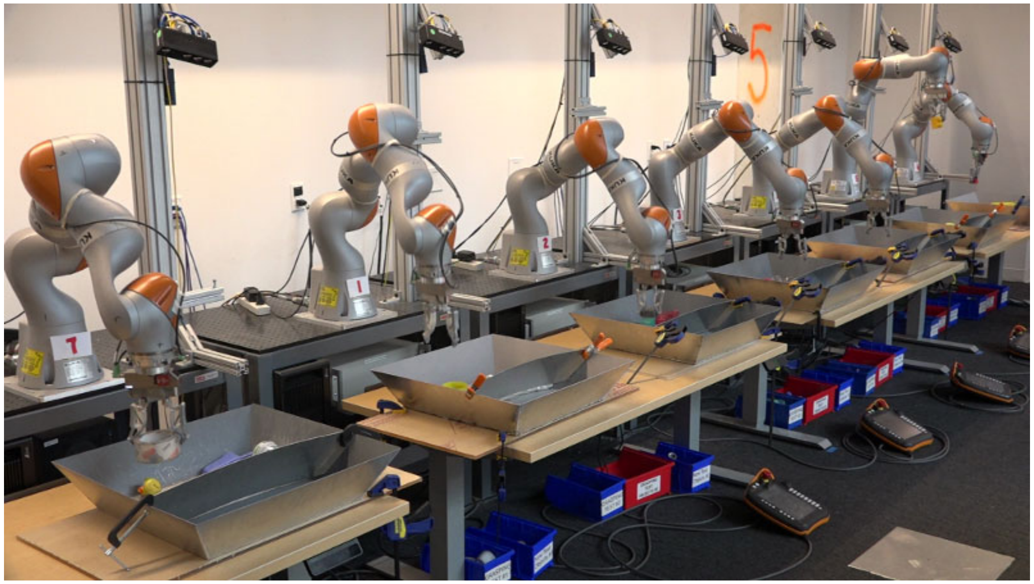
Robot Manipulation

Assembling

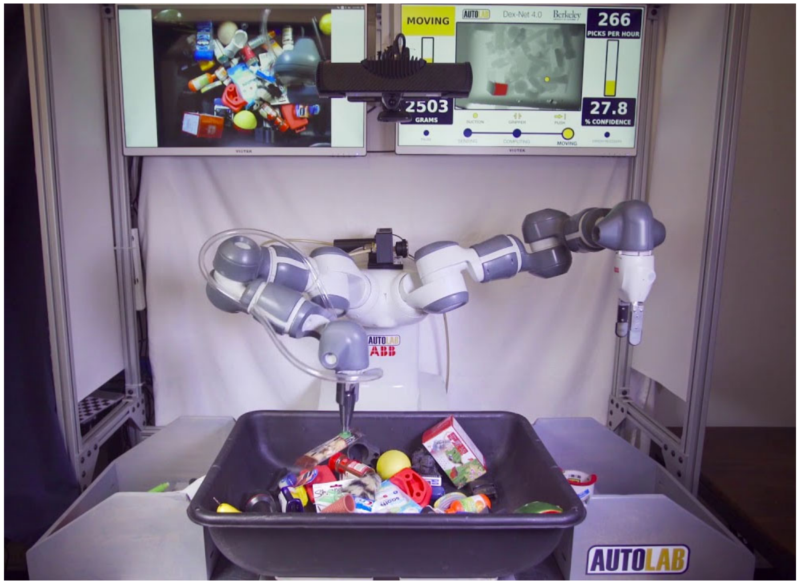
Cooking

Top-Down Grasping

- 3 degrees of freedom



Google



Berkeley: Dex-Net

6D Grasping: 3D Location and 3D Orientation

Perception → Planning → Control

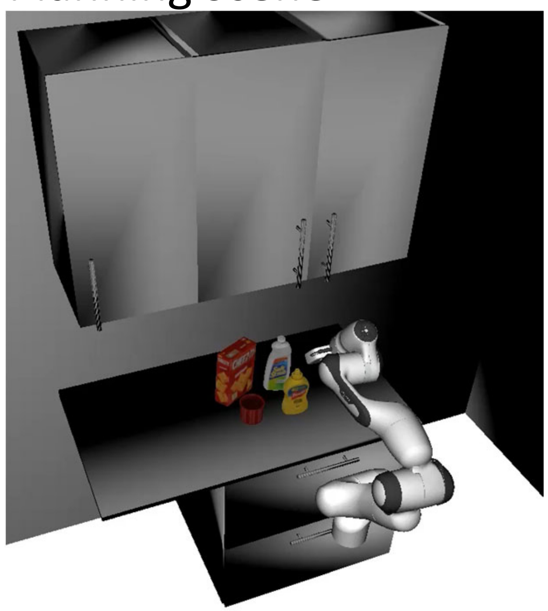
Perception: Robust and Accurate

Planning: High degree of freedom, Multi-modal grasping

Control: Contact with objects

Sensed image

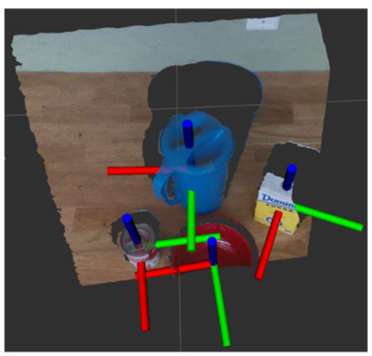
Planning scene



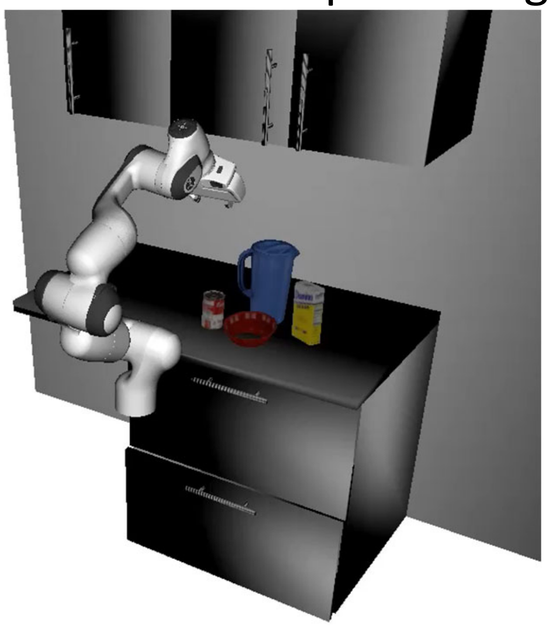
Real world execution

Model-based 6D Grasping

6D Object Pose Estimation



Motion and Grasp Planning



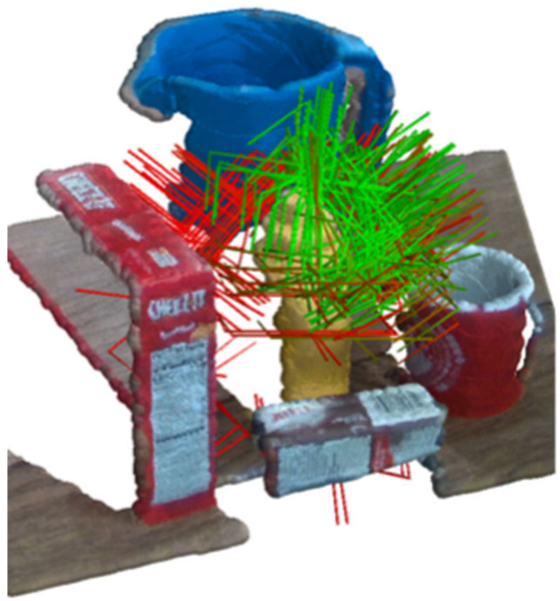
We need to have 3D models of objects

How can we enable robots to manipulate unseen objects?

Model-free 6D Grasping

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

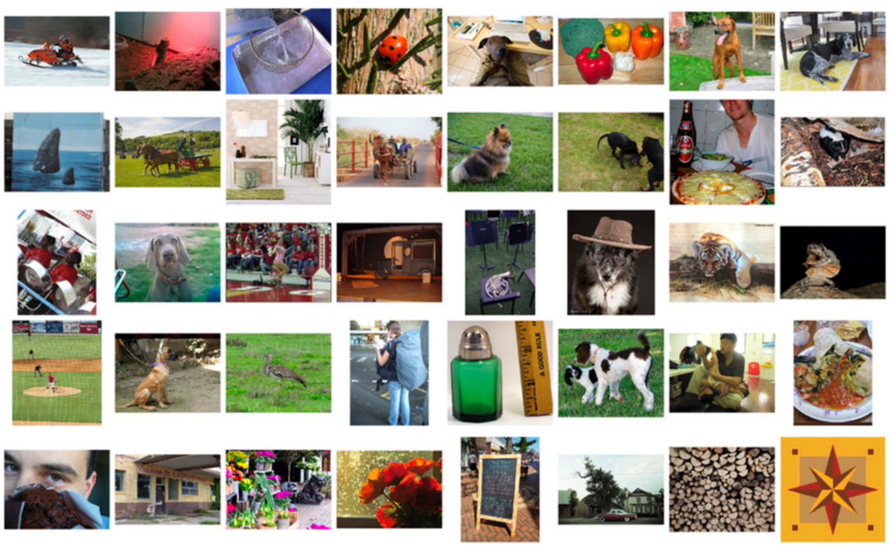
Perception: Unseen Object Instance Segmentation

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

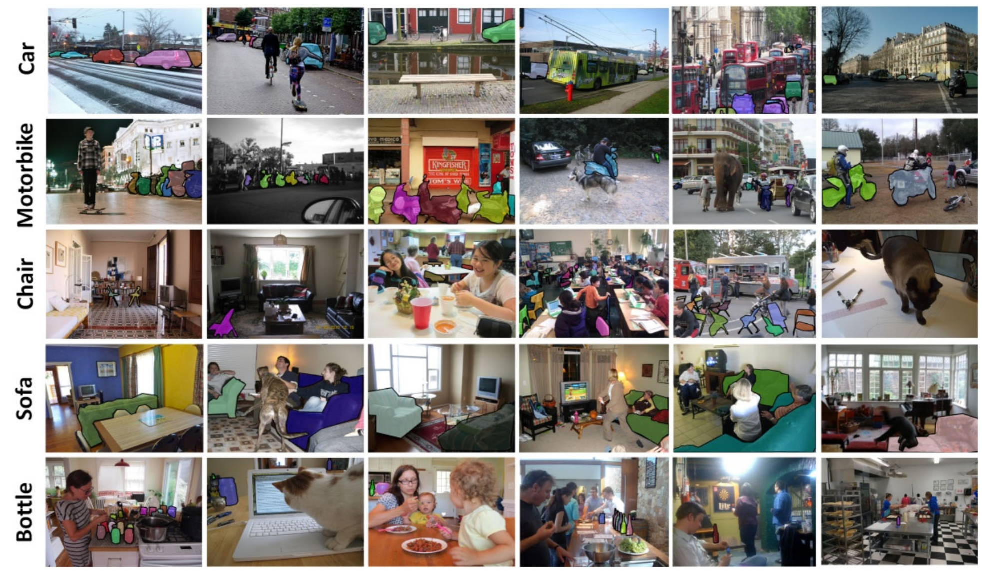
Codes available online

Learning the Concept of “Objects”

- Learning from data



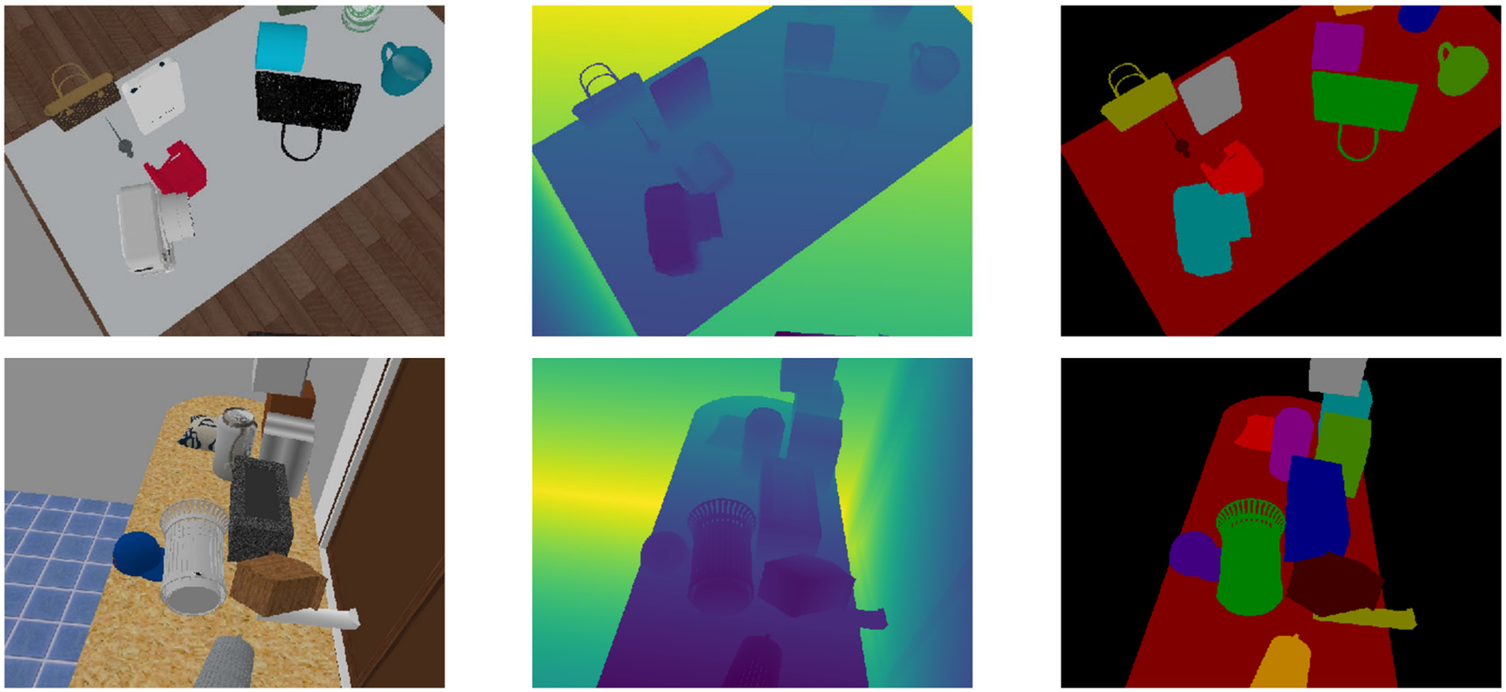
ImageNet: Deng-Dong-Socher-Li-Li-Fei-Fei, CVPR’09



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

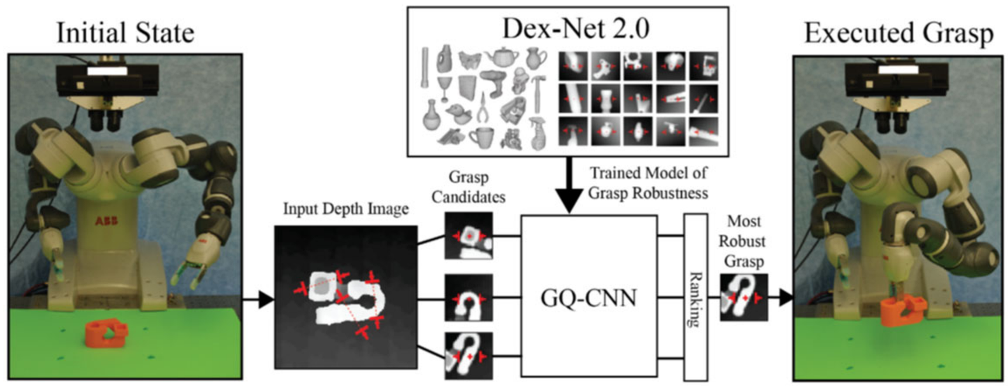
Tabletop Object Dataset: Xie-**Xiang**-Mousavian-Fox, CoRL'19

Previous Works: Learning from Depth

- Synthetic depth generalizes better to the real depth images

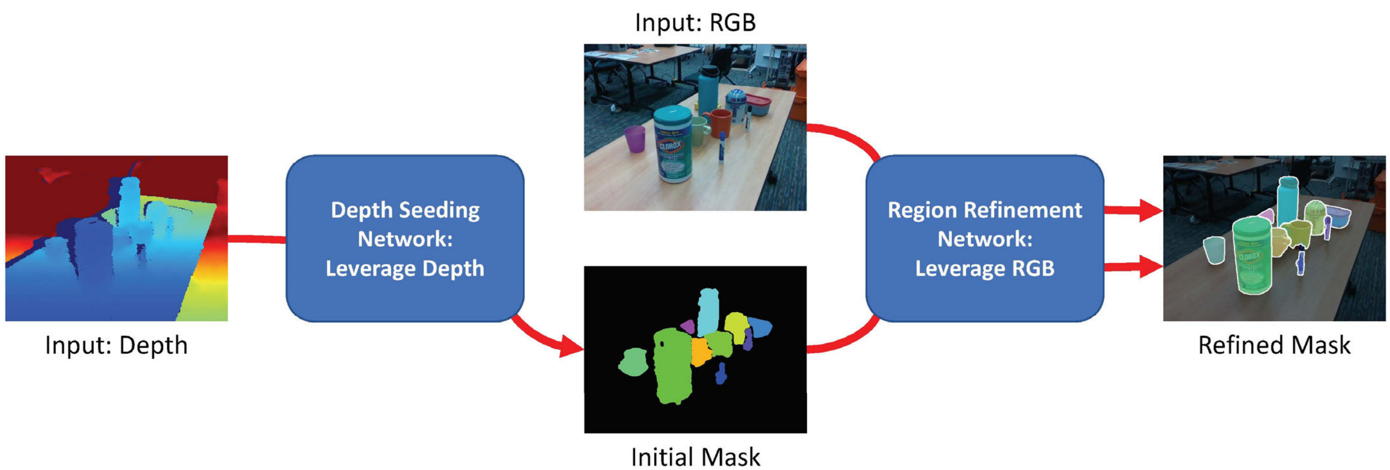
Dex-Net 2.0

Mahler et al. RSS’17



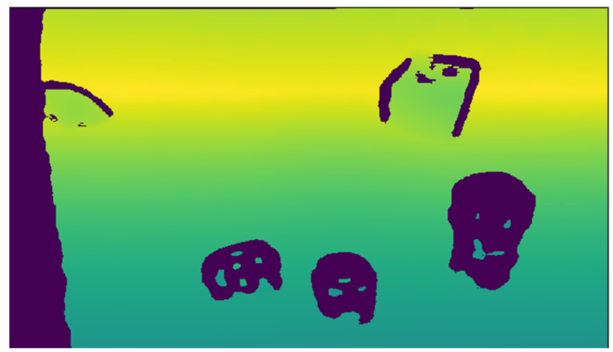
UOIS-Net

Xie et al. CoRL’19

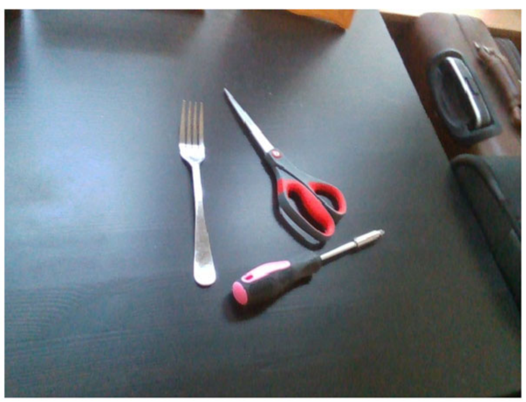


Can We Utilize Non-photorealistic Synthetic RGB images?

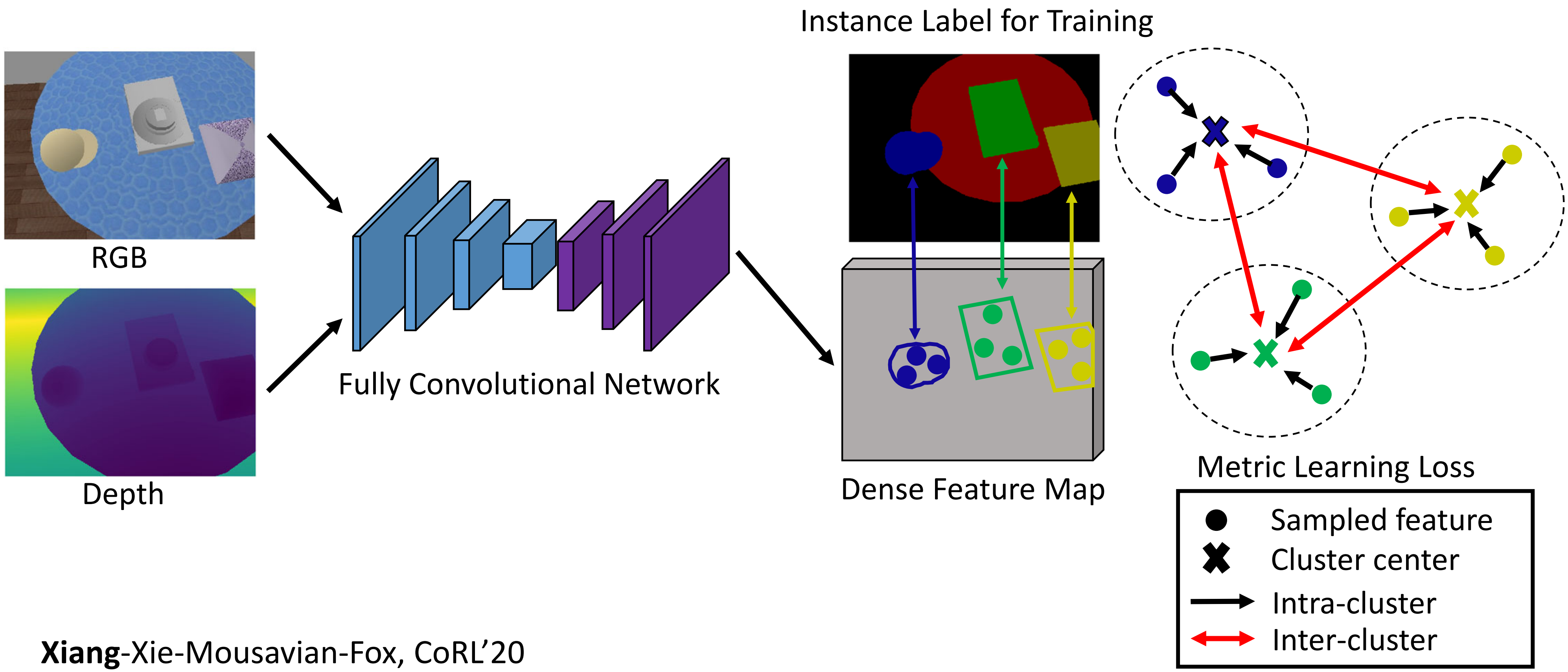
- Depth is not good for transparent objects or thin objects



ClearGrasp
Sajjan et al. ICRA'20

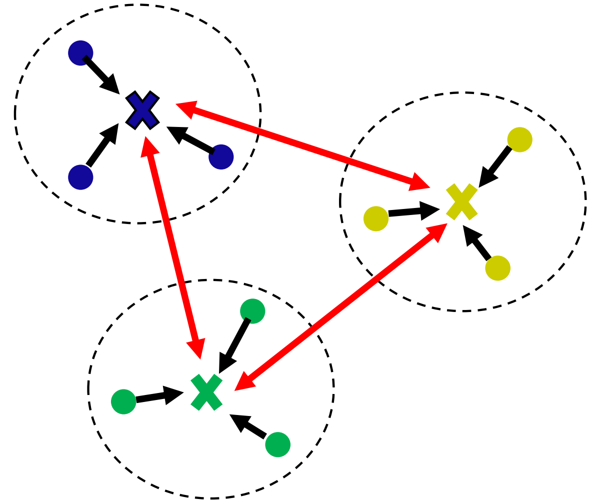


Unseen Object Instance Segmentation: Learning RGB-D Feature Embeddings



Xiang-Xie-Mousavian-Fox, CoRL’20

Metric Learning Loss Function



- Intra-cluster loss function

$$\mu^k = \frac{\sum_{i=1}^N \mathbf{x}_i^k}{\| \sum_{i=1}^N \mathbf{x}_i^k \|}$$

Spherical mean

$$d(\mu^k, \mathbf{x}_i^k) = \frac{1}{2}(1 - \mu^k \cdot \mathbf{x}_i^k)$$

Cosine distance

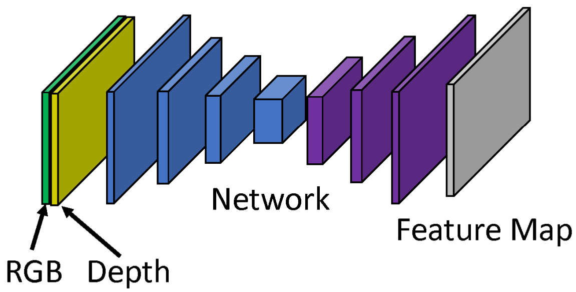
$$\ell_{\text{intra}} = \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{N} \frac{\mathbb{1}\left\{d(\mu^k, \mathbf{x}_i^k) - \alpha \geq 0\right\} \; d^2(\mu^k, \mathbf{x}_i^k)}{\sum_{i=1}^N \mathbb{1}\left\{d(\mu^k, \mathbf{x}_i^k) - \alpha \geq 0\right\}}$$

- Inter-cluster loss function

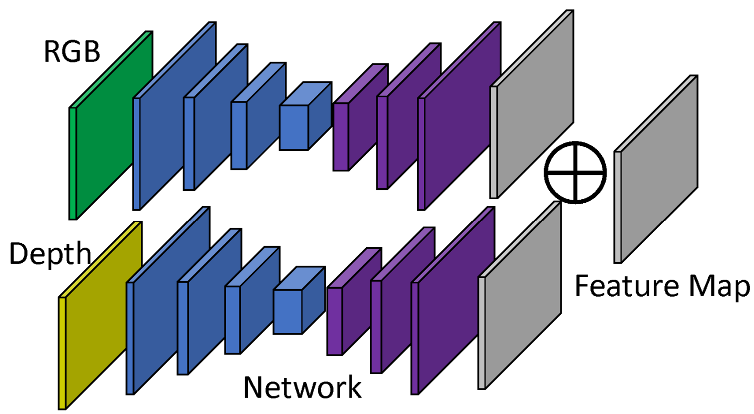
$$\ell_{\text{inter}} = \frac{2}{K(K-1)} \sum_{k<k'} \left[\delta - d(\mu^k, \mu^{k'})\right]_+^2$$

Xiang-Xie-Mousavian-Fox, CoRL'20

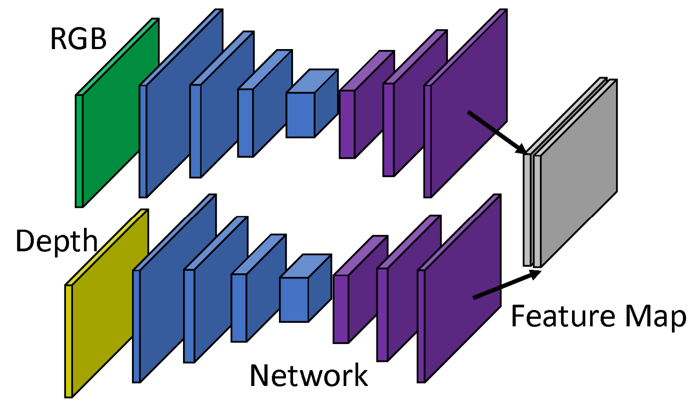
Fusing RGB and Depth



(a) Early Fusion

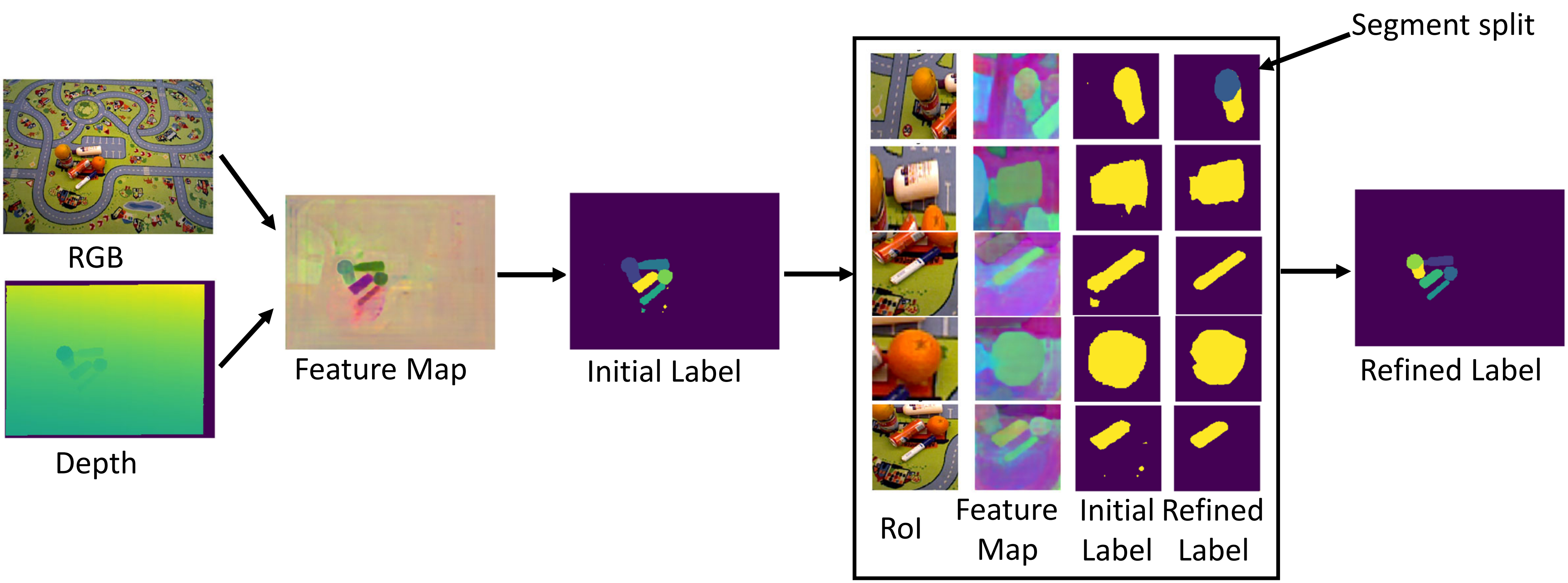


(b) Late Fusion Addition



(c) Late Fusion Concatenation

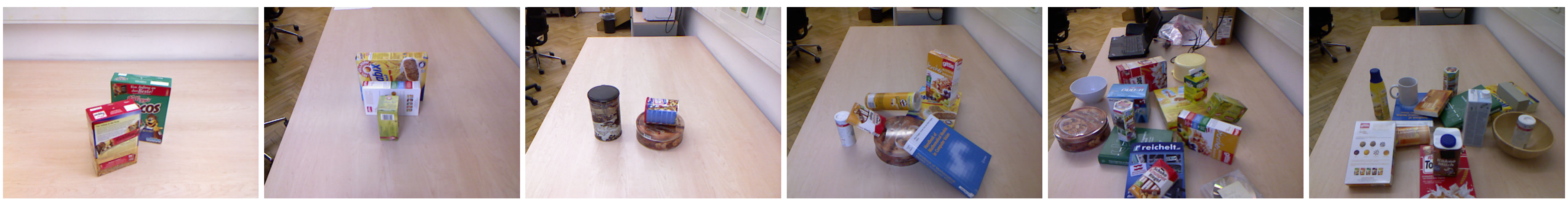
Two-stage Clustering



Experiments: Datasets

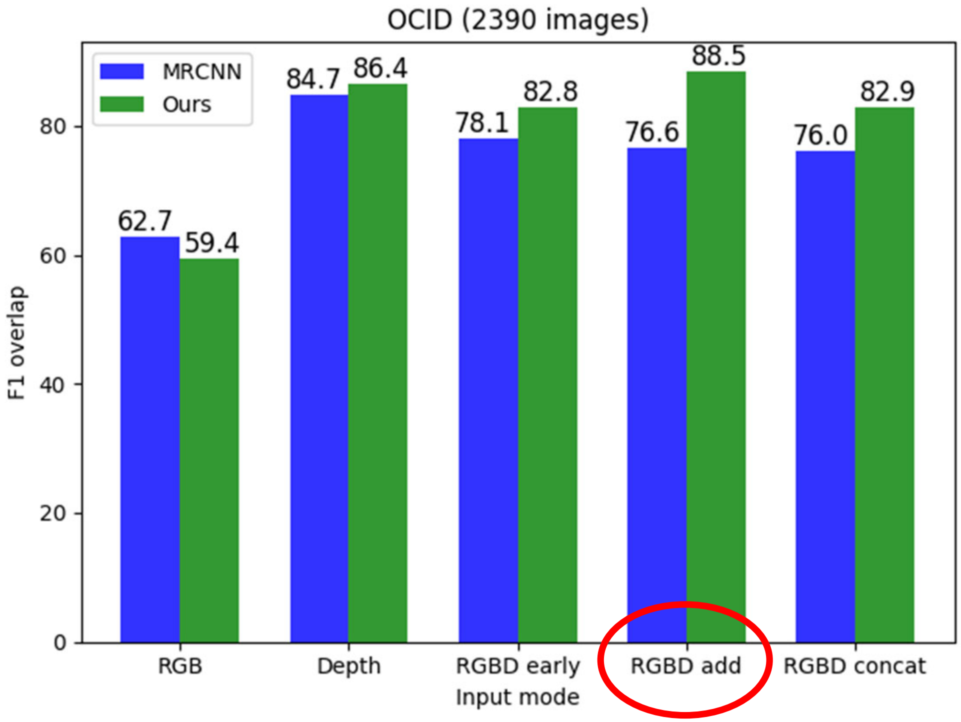
- Object Cluster Indoor Dataste (OCID), 2,390 RGB-D images   Sushi et al. ICRA'19

- Object Segmentation Database (OSD), 111 RGB-D images   Richtsfeld et al. IROS'12

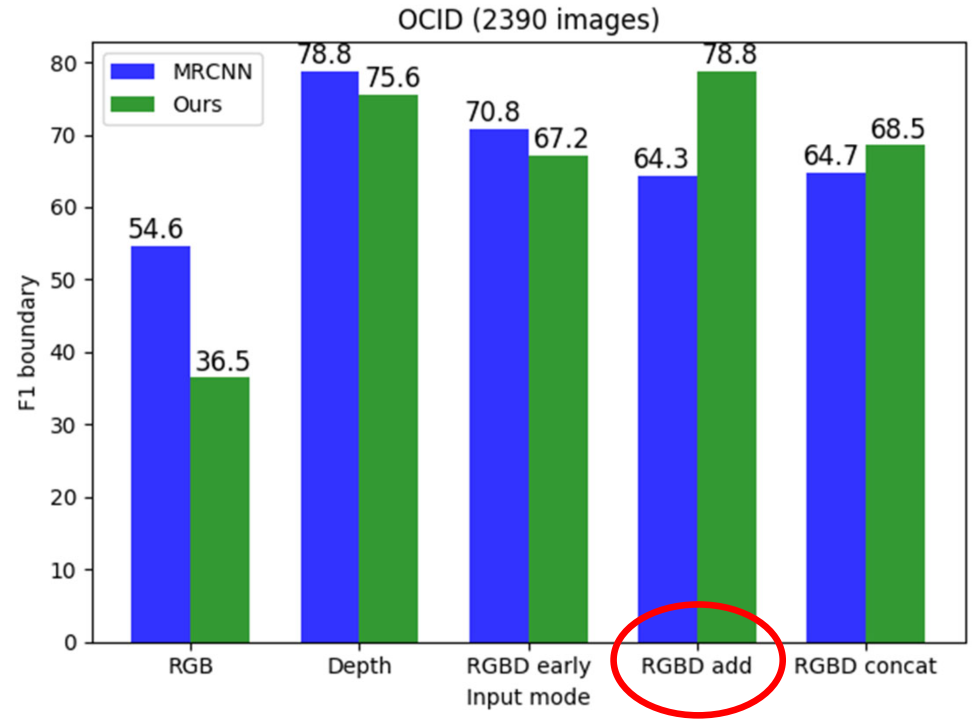


Effect of the Input Mode

Mask R-CNN. He et al. CVPR’17

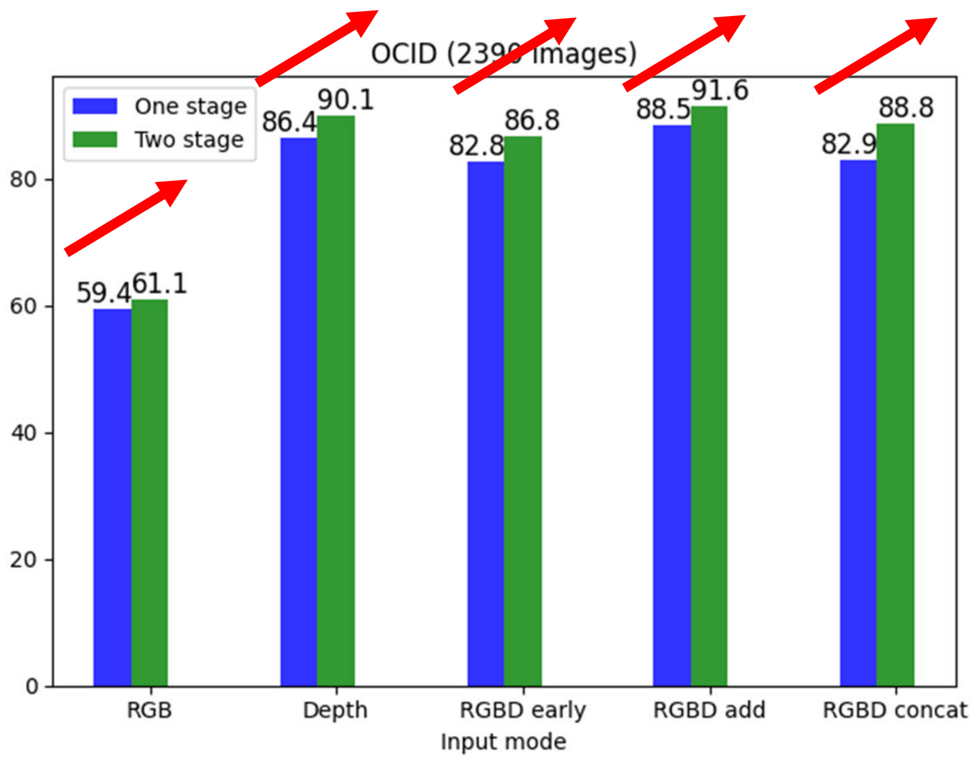


F-score overlap

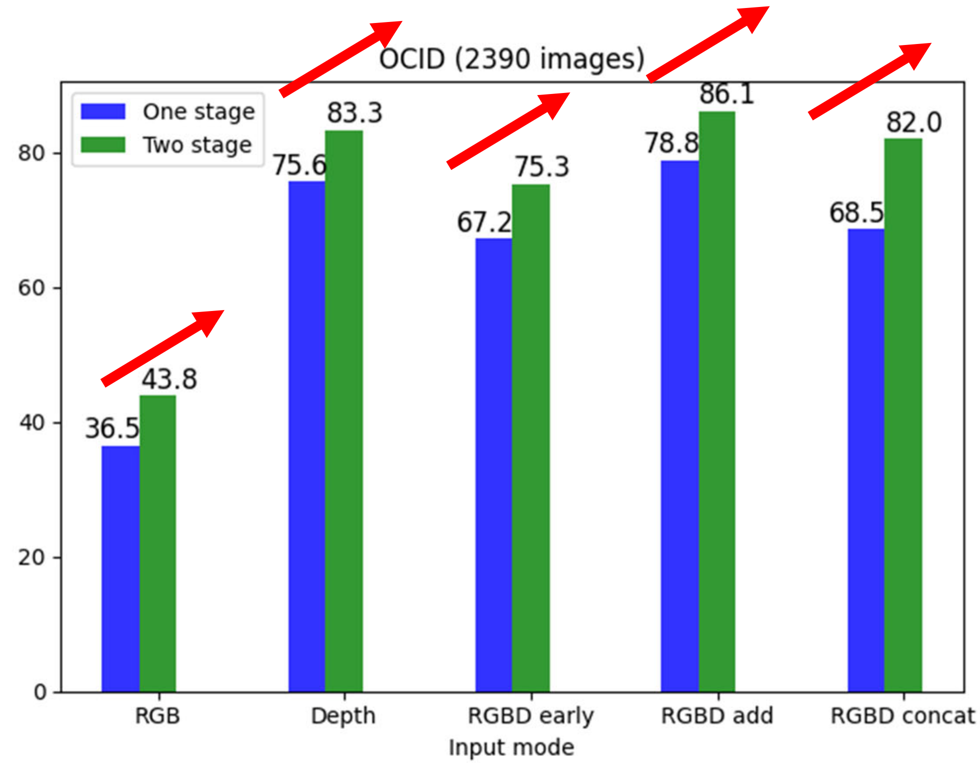


F-score boundary

Effect of the Two-stage Clustering

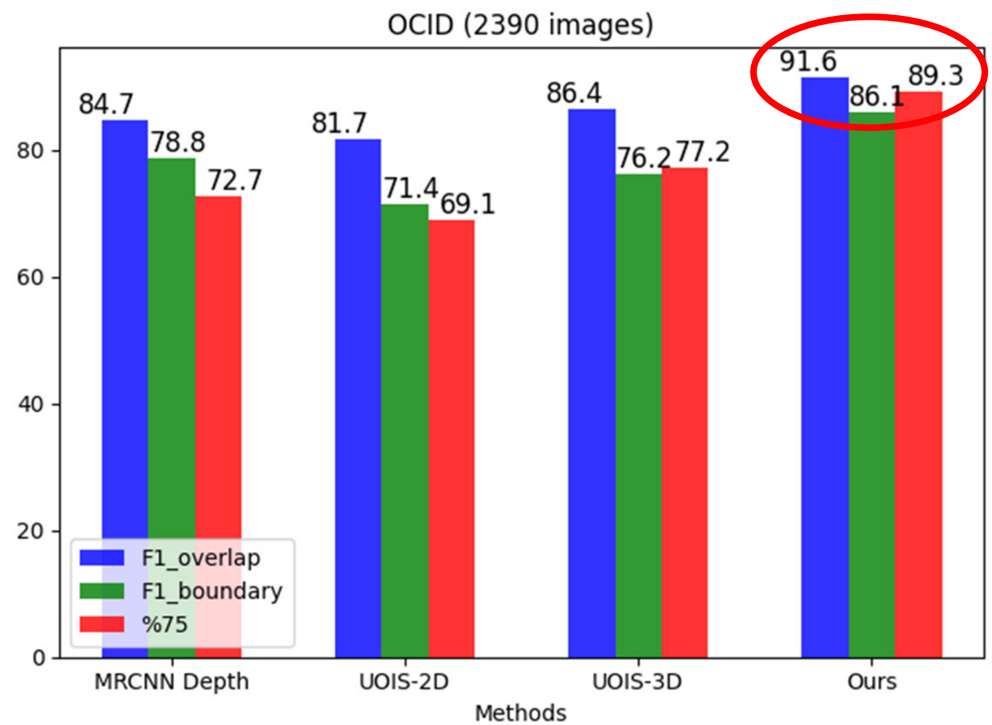


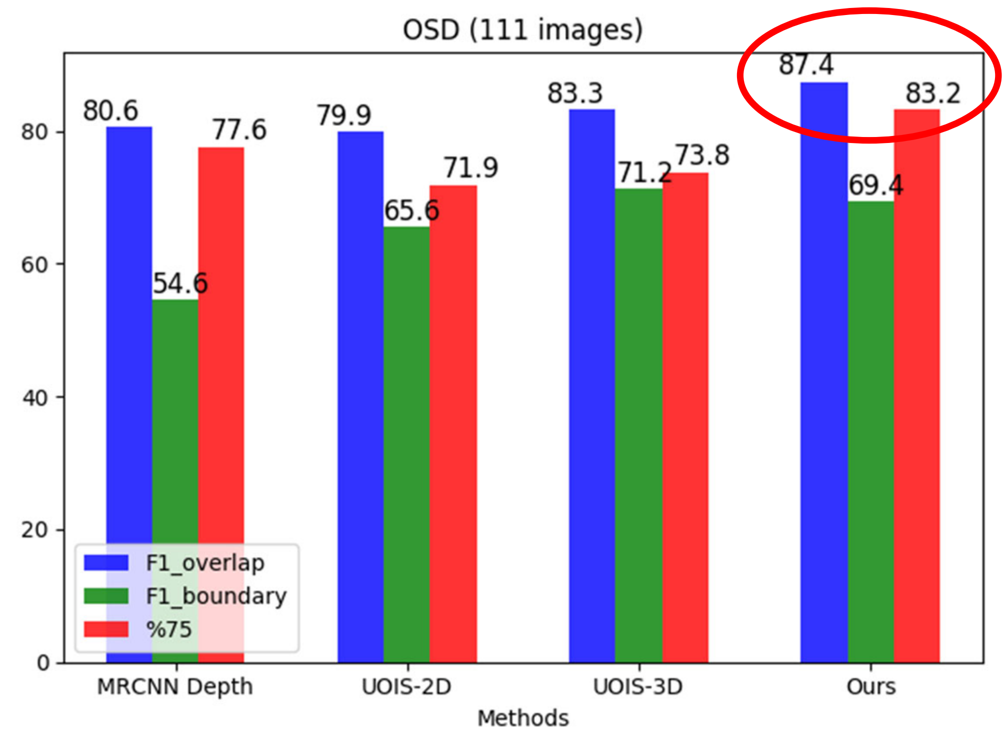
F-score overlap



F-score boundary

Comparison to State-of-the-arts





Mask R-CNN. He et al. CVPR’17
UOIS-2D. Xie et al. CoRL’19
UOIS-3D. Xie et al. T-RO’21

Input Image

Feature Map

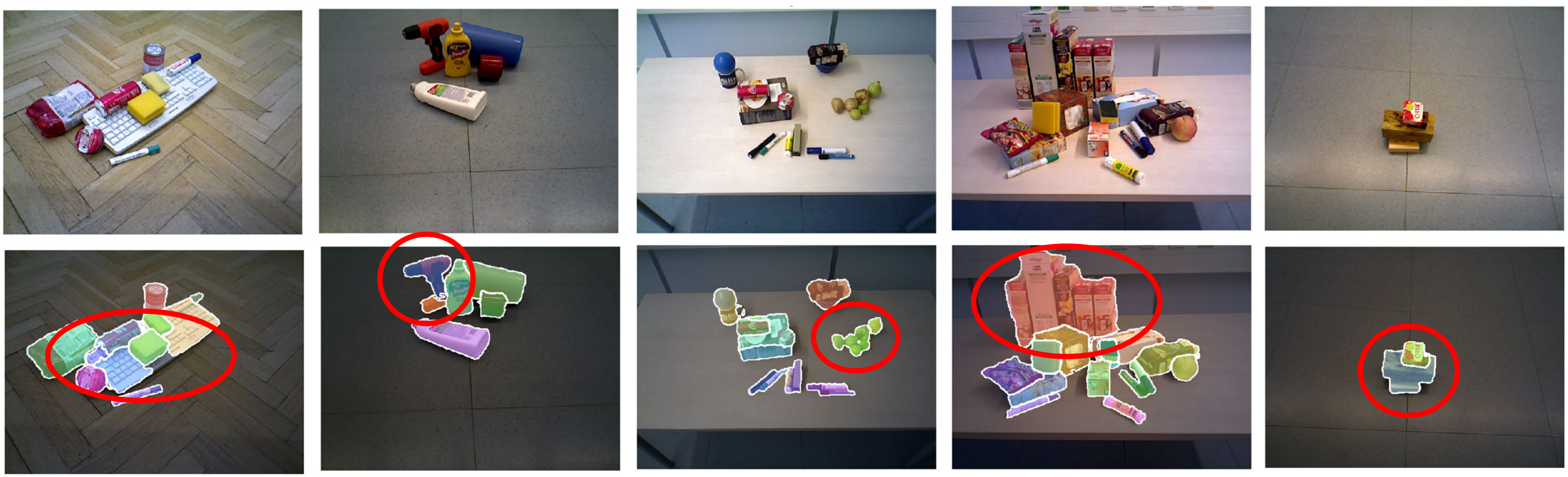
Output Label

Xiang-Xie-Mousavian-Fox, CoRL'20

Failure Cases

Input Image

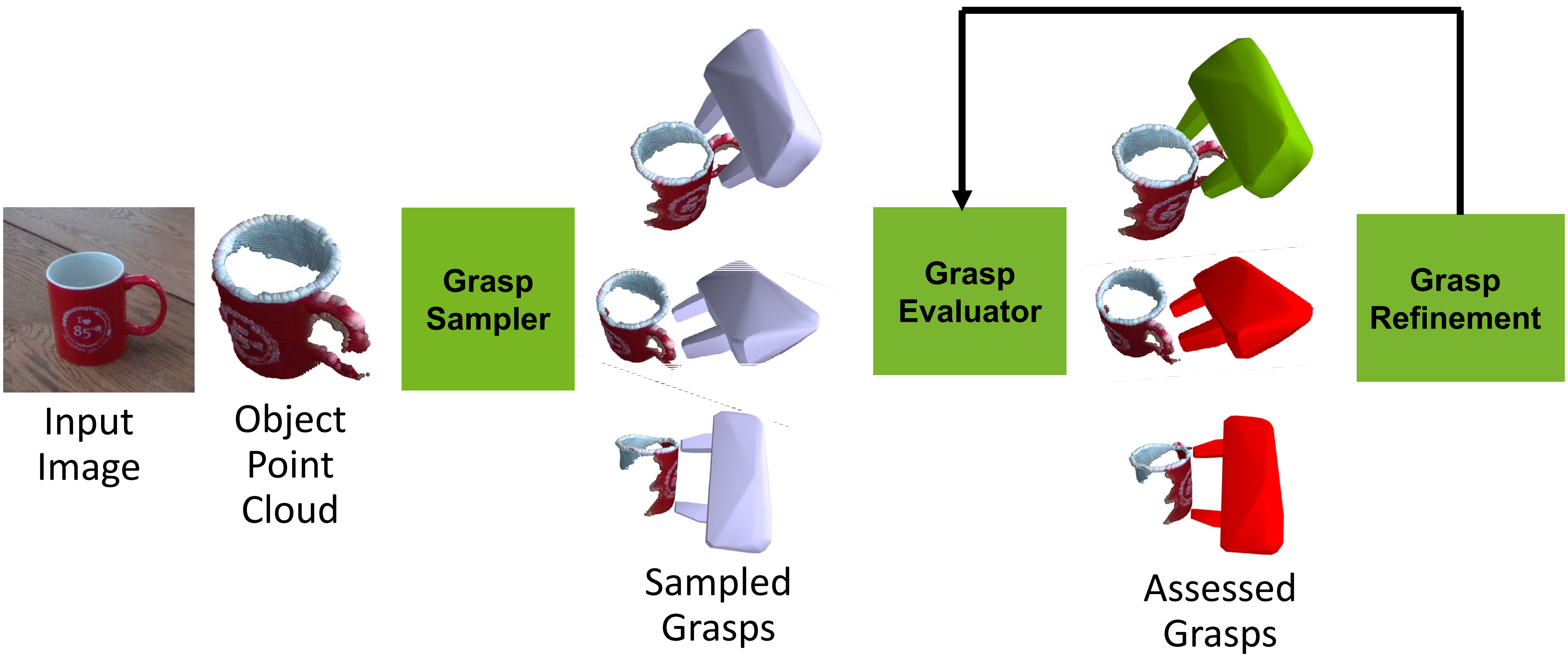
Final Label



Over-segmentation

Under-segmentation

Grasp Planning from Partially Observed Point Clouds



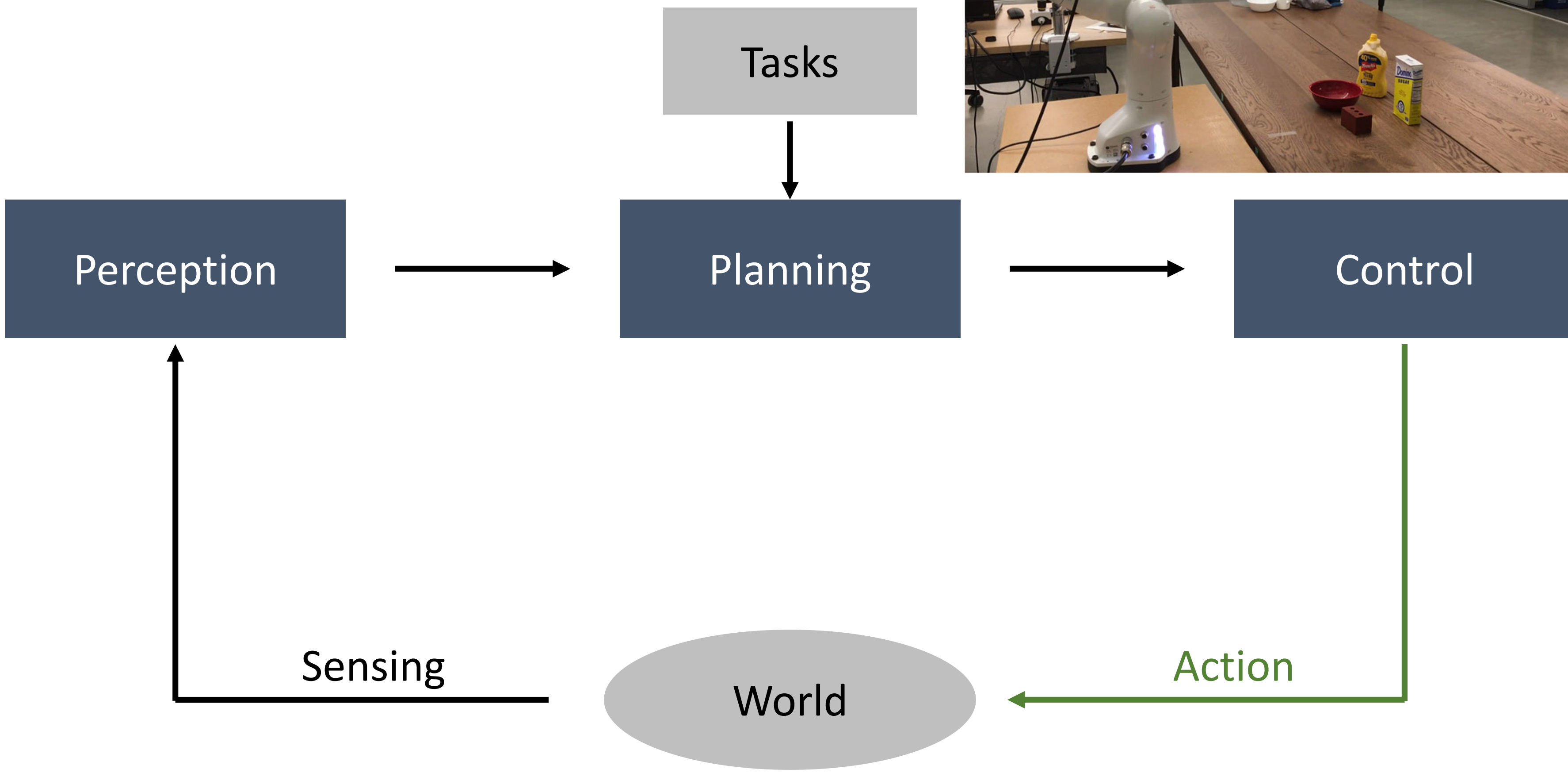
6-DOF GraspNet: Mousavian-Eppner-Fox, ICCV'19

6D Grasping of Unseen Objects

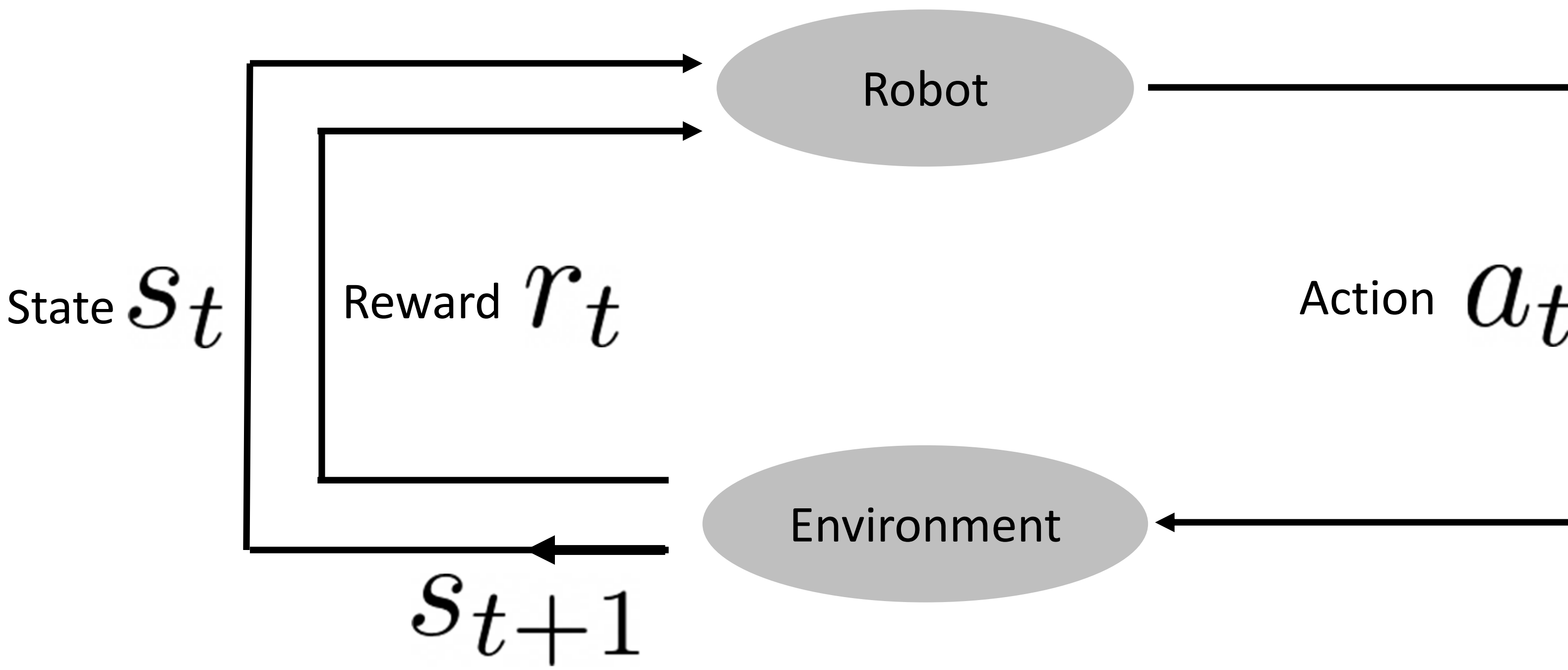
Unseen Object Instance Segmentation:
Xie-**Xiang**-Mousavian-Fox, CoRL'19, T-RO'21
Xiang-Xie-Mousavian-Fox, CoRL'20

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

Open-Loop VS. Closed-Loop



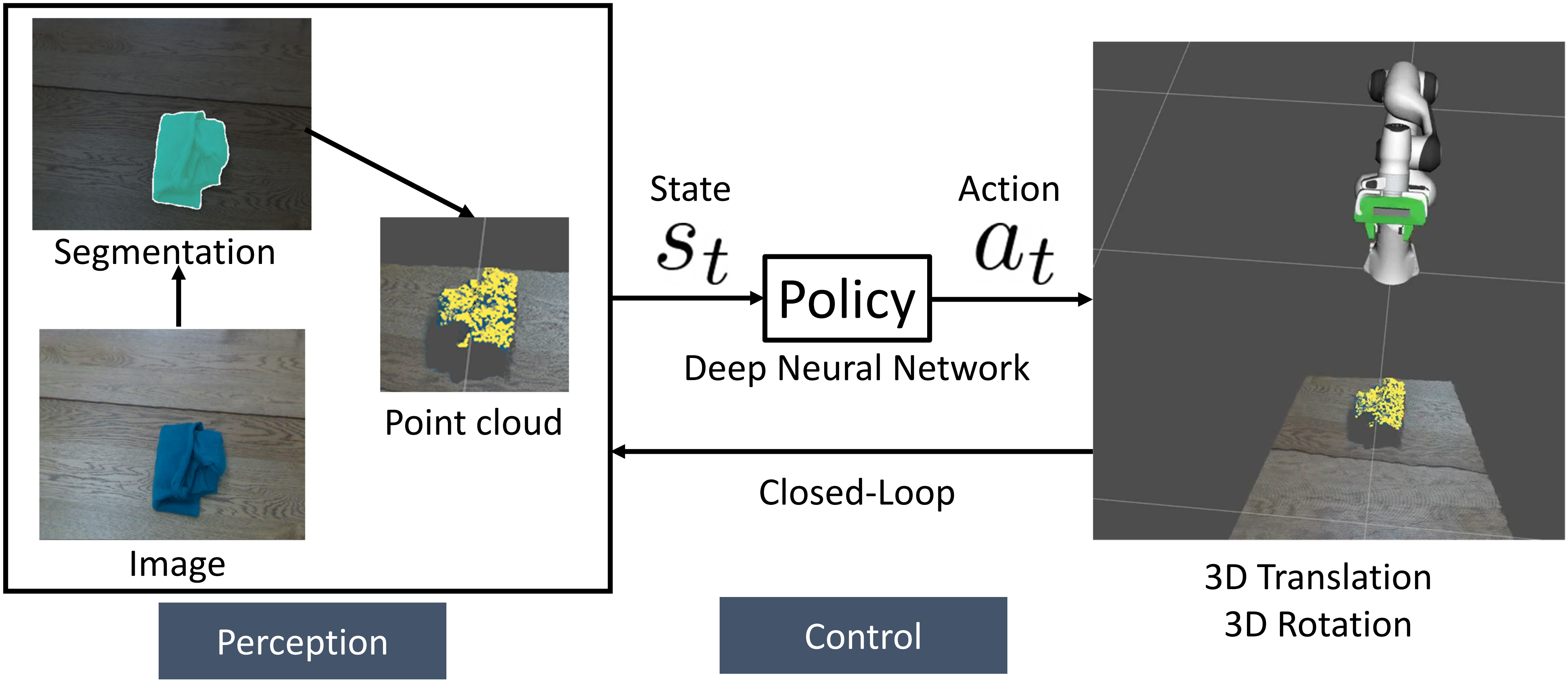
Closed-loop Robot Control with Markov Decision Processes



Reinforcement Learning:
Imitation Learning:

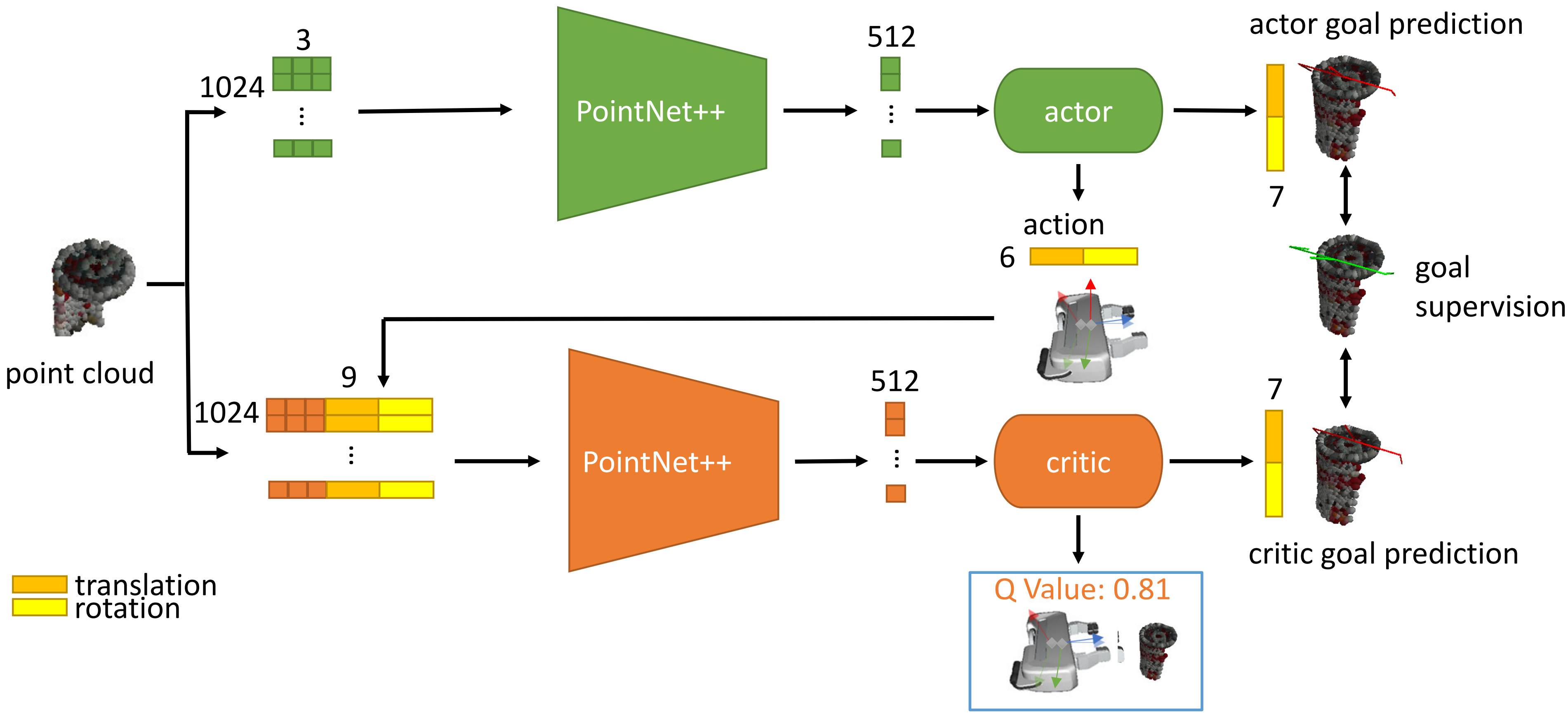
$$a_t = \pi(s_t)$$

Learning Closed-Loop Control Polices for 6D Grasping



Wang-**Xiang**-Yang-Mousavian-Fox, in arXiv’21

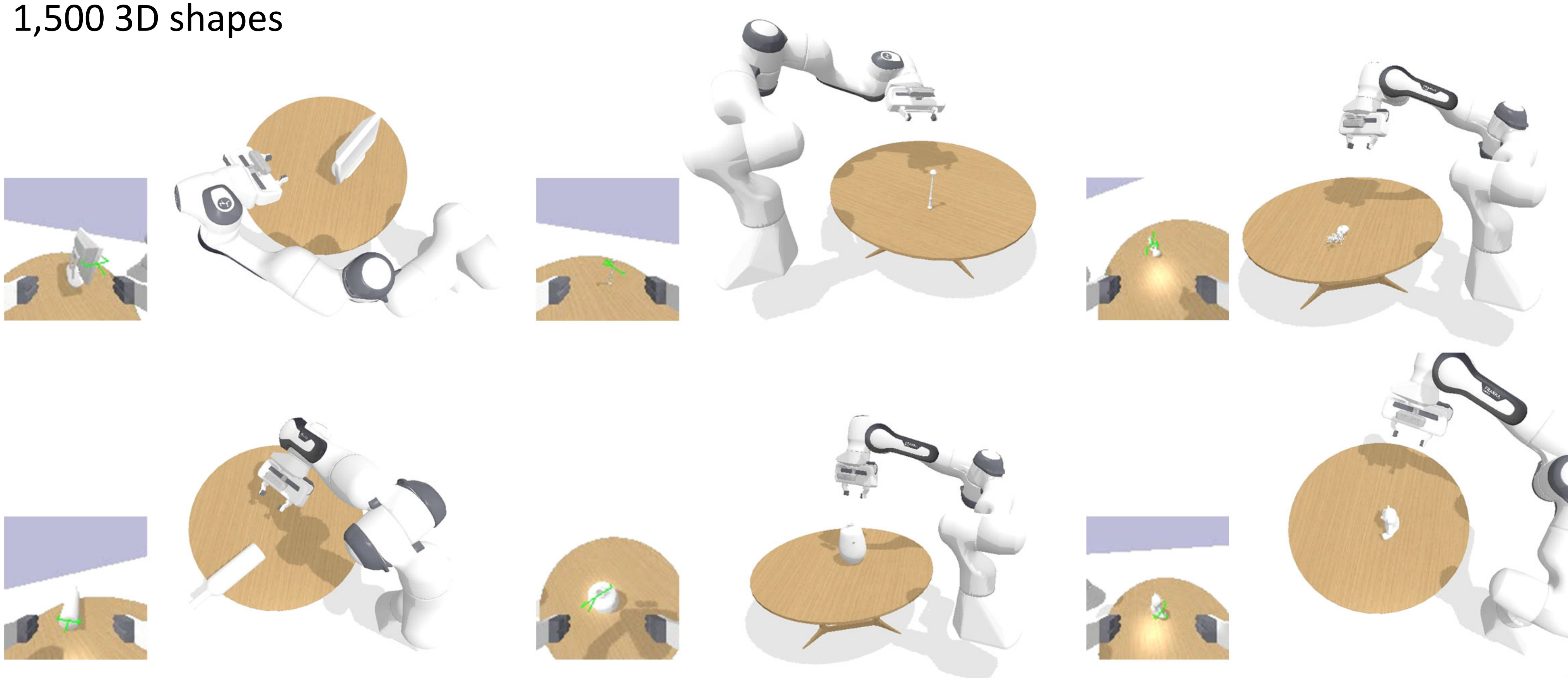
Goal-Auxiliary Actor-Critic Network



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

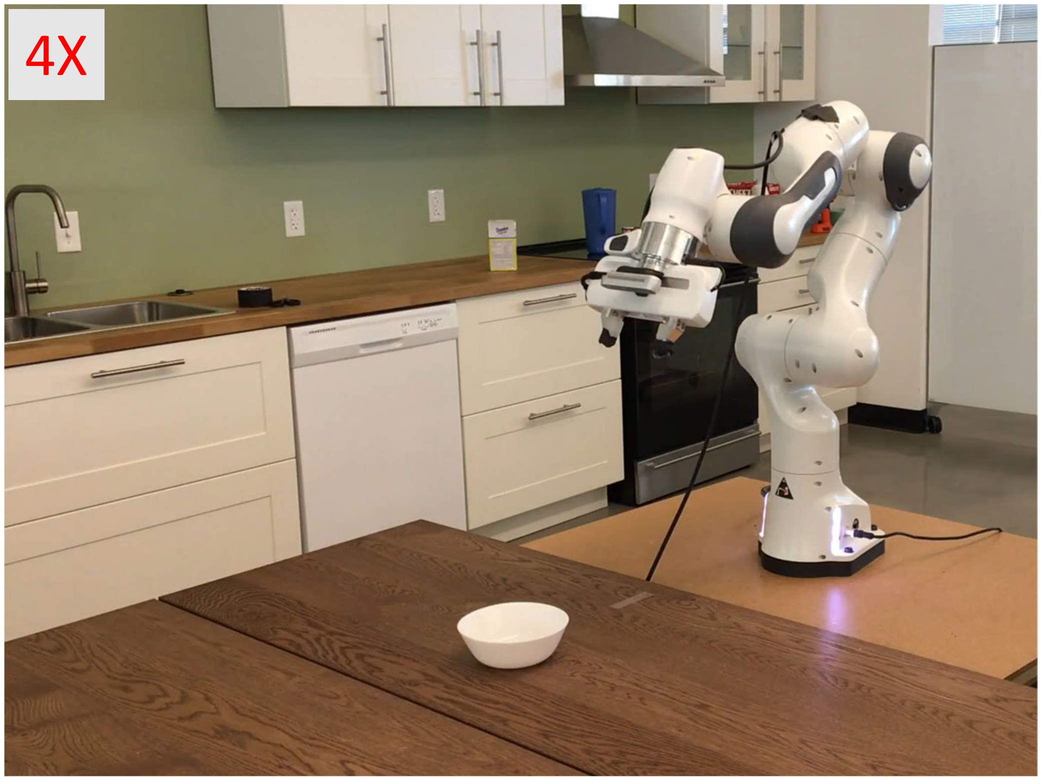
Learning from Demonstration with the OMG-Planner

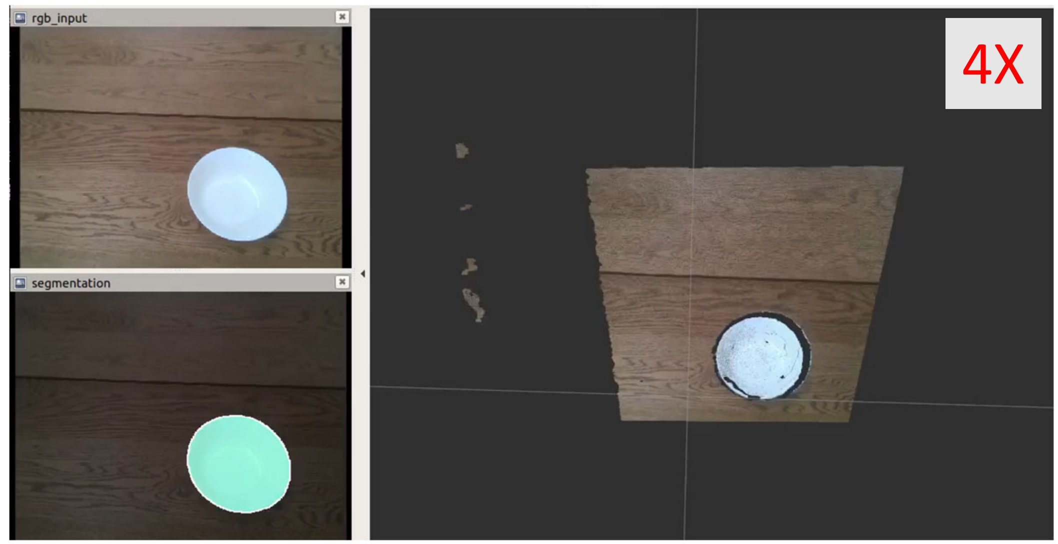
50,000 trajectories
1,500 3D shapes



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

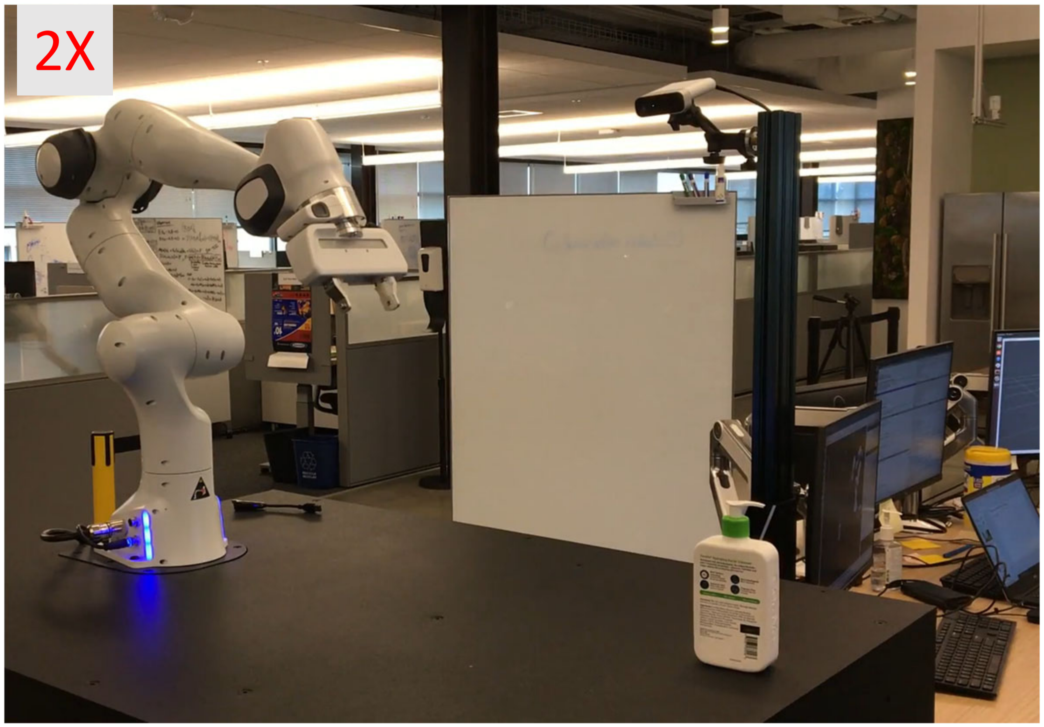
Our Learned Policy in the Real World

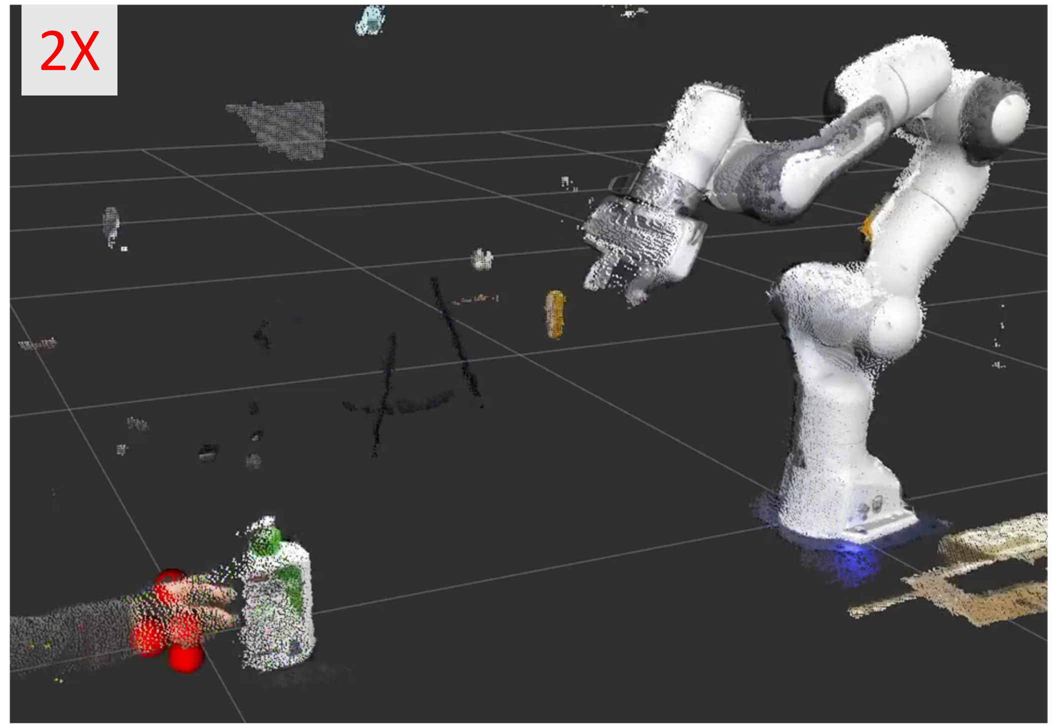




Wang-**Xiang**-Yang-Mousavian-Fox, in arXiv'21

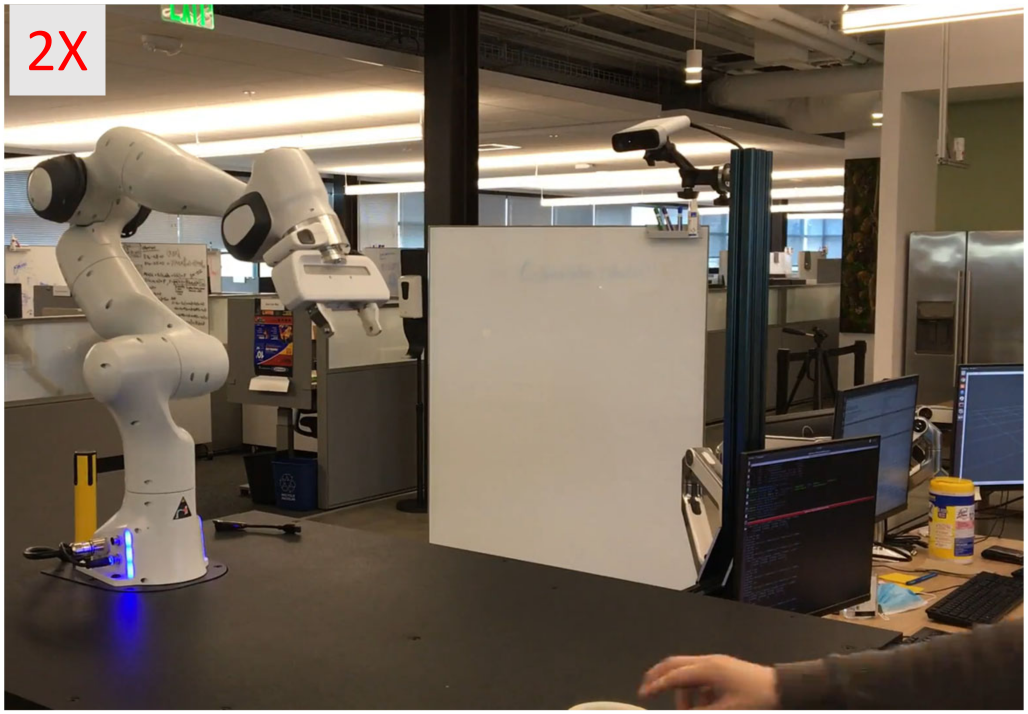
Closed-Loop Human-to-Robot Handover

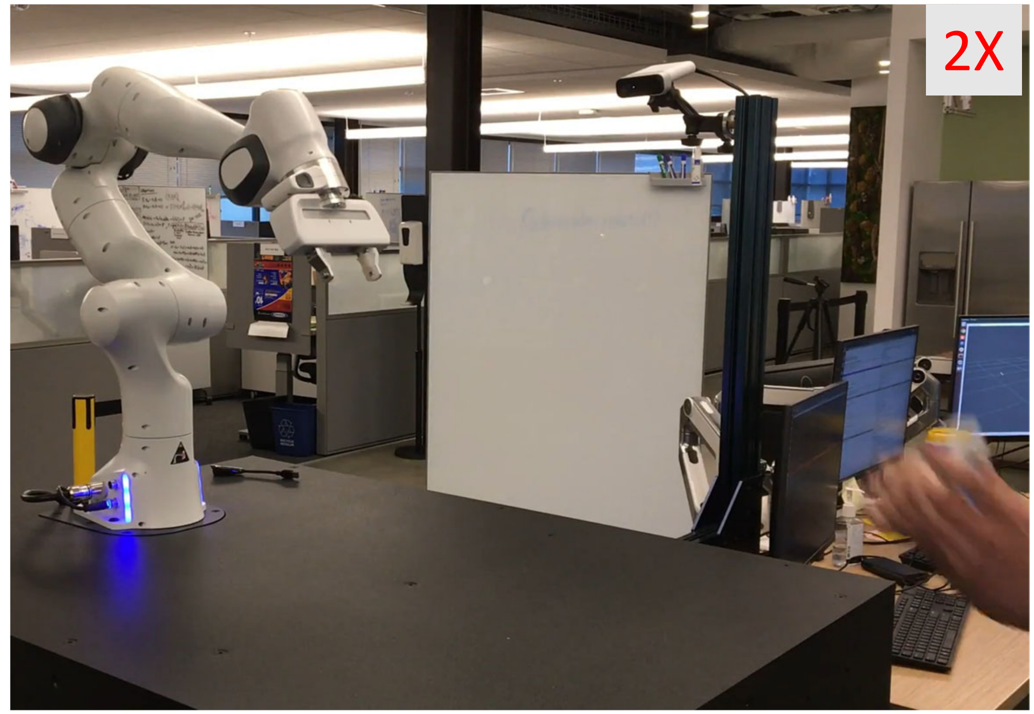




Yang-Paxton-Mousavian-Chao-Cakmak-Fox, in arXiv'20
Wang-**Xiang**-Yang-Mousavian-Fox, in arXiv'21

Closed-Loop Human-to-Robot Handover

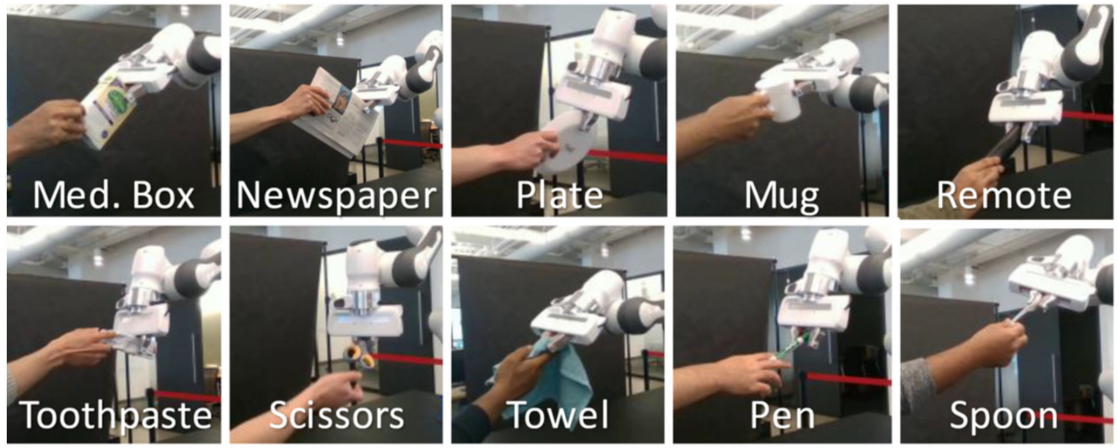




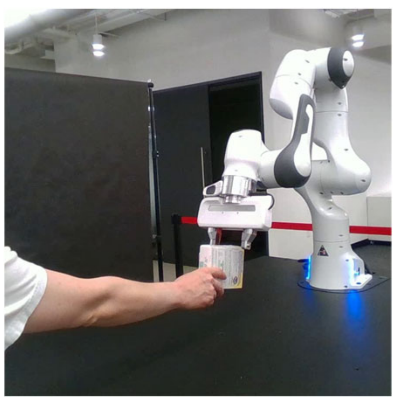
Yang-Paxton-Mousavian-Chao-Cakmak-Fox, in arXiv’20
Wang-**Xiang**-Yang-Mousavian-Fox, in arXiv’21

Closed-Loop Human-to-Robot Handover

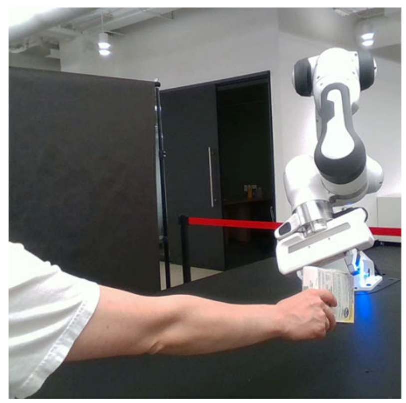
10 objects



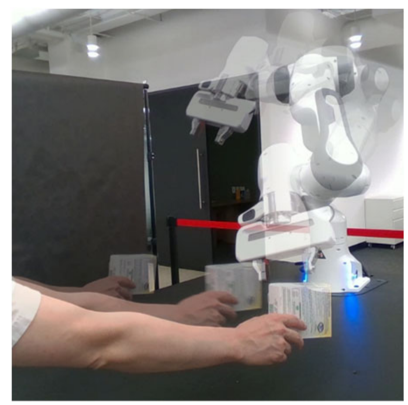
Left: 90%



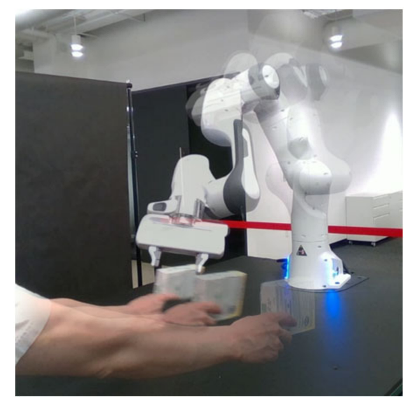
Middle: 100%



Right: 100%



Left→Right: 60%



Right→Left: 40%

Conclusion

- Unseen Object Instance Segmentation
  - Train on synthetic data, test on real-world images
  - Learning RGB-D feature embeddings for clustering

- Learning closed-loop control policies for 6D robotic grasping
  - Learning from demonstrations
  - Using point clouds as input for generalization
  - Polices trained in simulation work in the real world
  - Tabletop 6D grasping and human-to-robot handover

yu.xiang@utdallas.edu

Thank you!