Connecting 6D Object Pose Estimation with Robot Manipulation

Yu Xiang

Assistant Professor

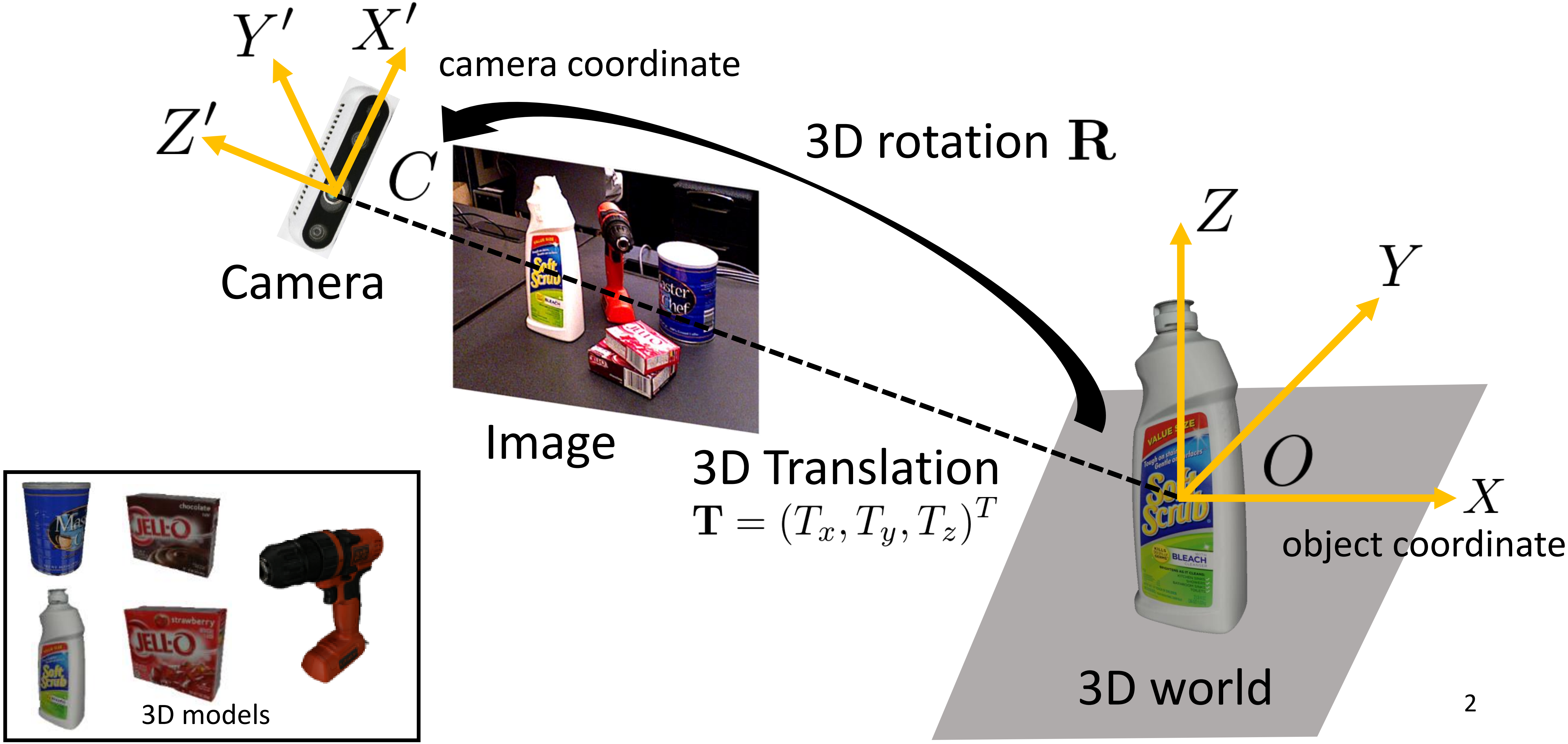
Computer Science

The University of Texas at Dallas

10/3/2023

8th International Workshop on Recovering 6D Object Pose @ ICCV 2023

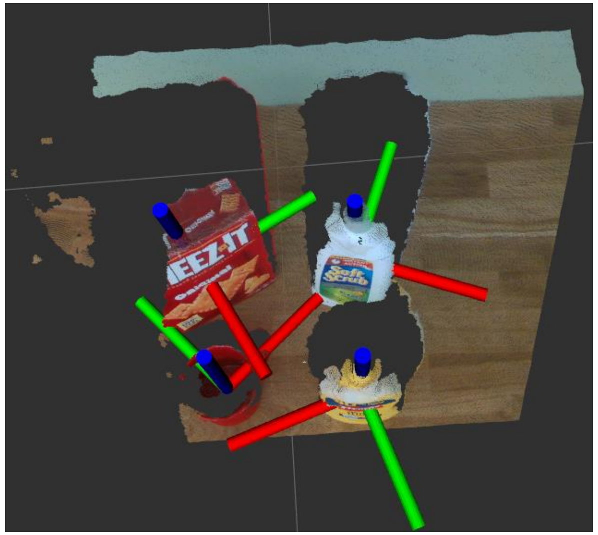
Model-based 6D Object Pose Estimation



Model-based 6D Object Pose Estimation

- What information can be obtained from 6D object pose estimation?
  - Object position in camera frame $\mathbf{T} = (T_x, T_y, T_z)^T$
  - Object orientation in camera frame $\mathbf{R}$

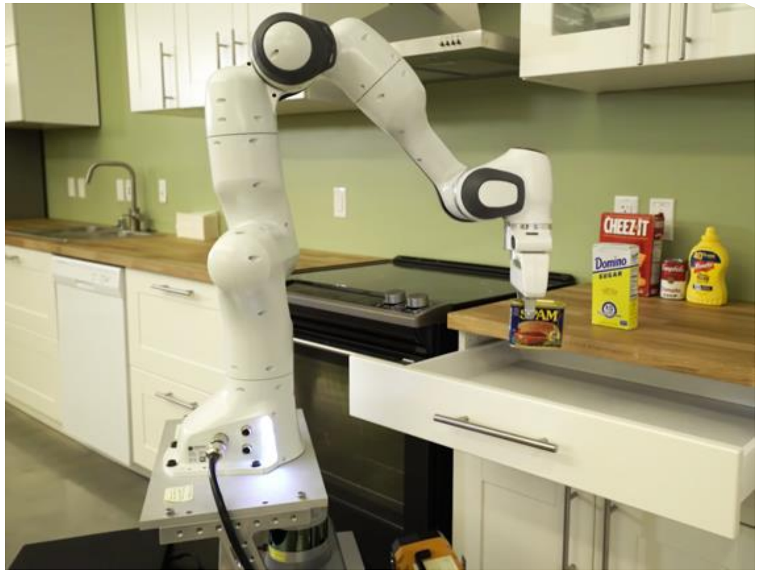
Input image



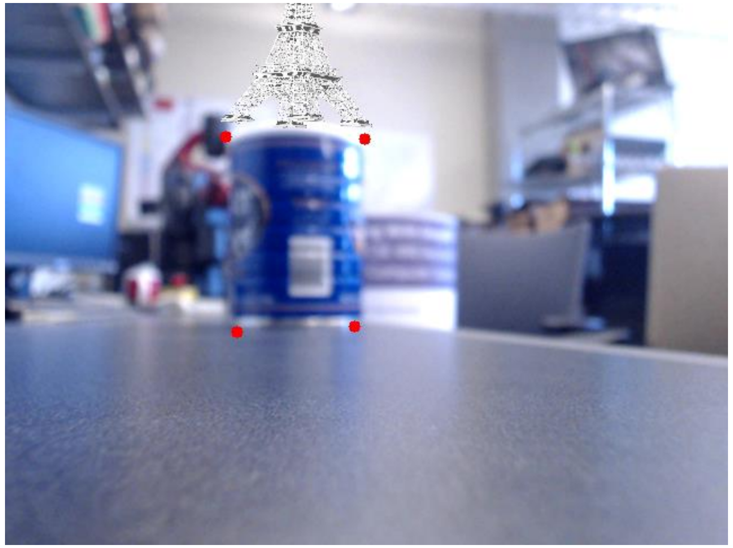
Point cloud and object axes in camera frame

Projection of 3D models onto the input image

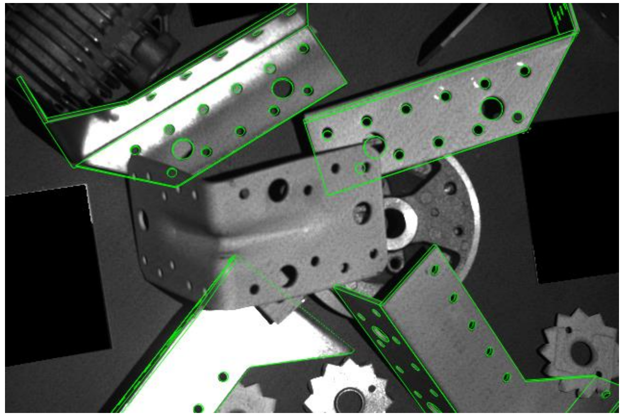
Applications



Robot Manipulation



Augmented Reality

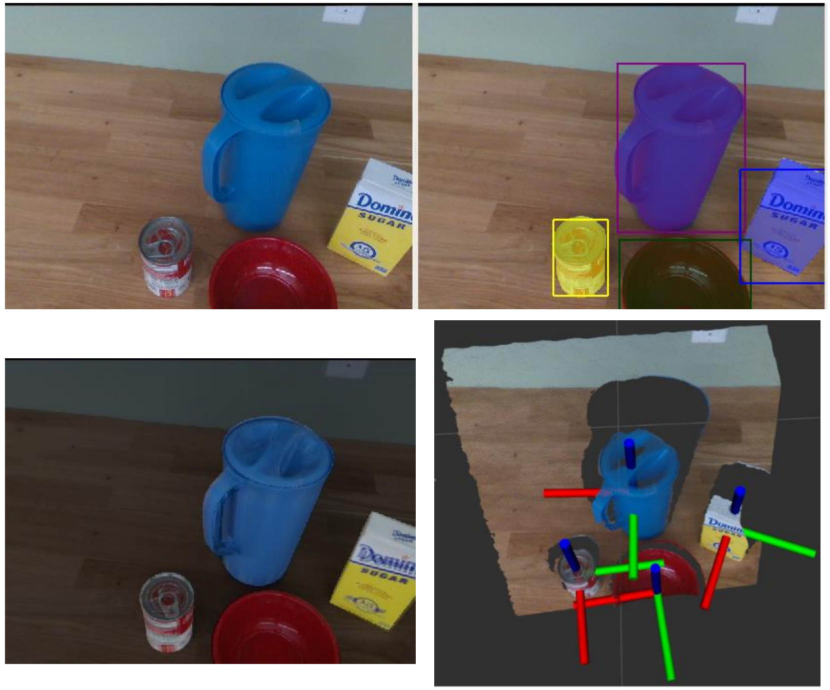


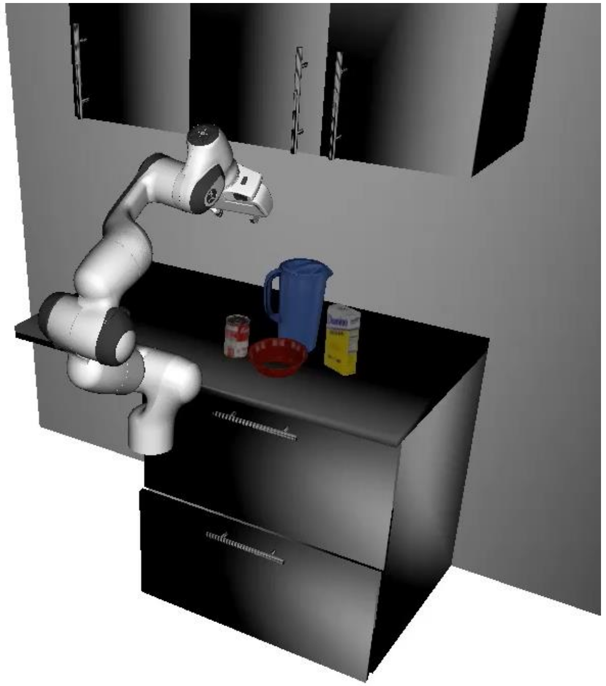
https://www.mvtec.com/company/research/datasets/mvtec-itodd

Industrial Object Inspection

Model-Based Robot Manipulation

Perception	→	Planning	→	Control
6D object pose estimation		Grasp planning and motion planning		Manipulation trajectory following

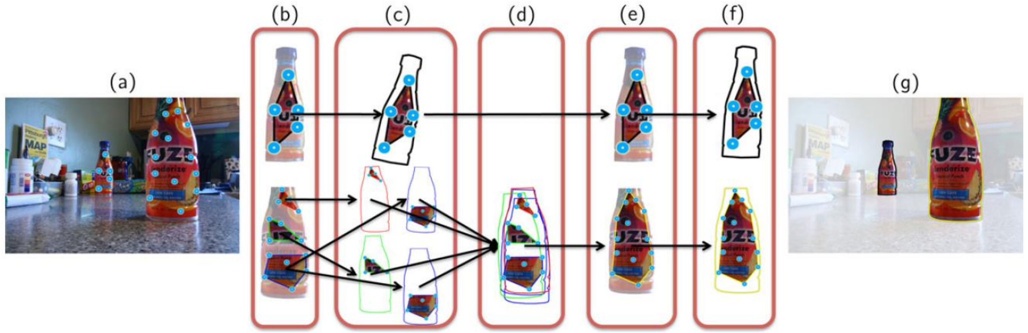




Traditional Methods for 6D Object Pose Estimation

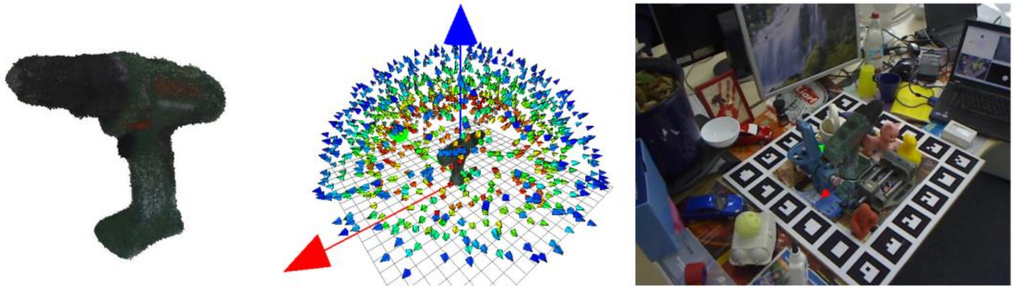
- Feature matching-based methods

Rothganger-Lazebnik-Schmid-Ponce, IJCV'06

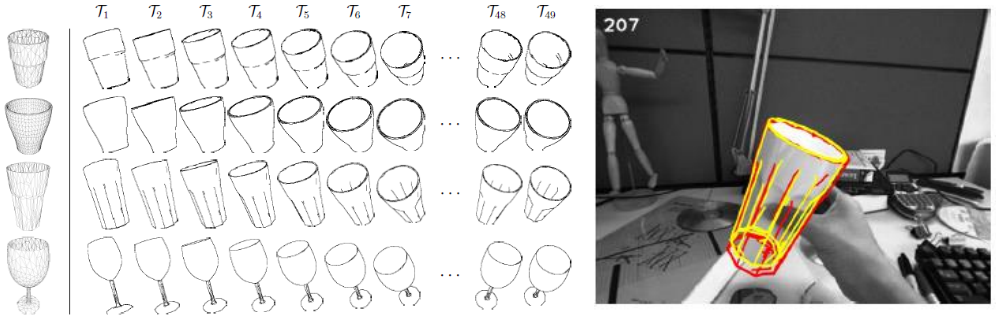


Collet-Martinez-Srinivasa, IJRR'11

- Template matching-based methods



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



Choi-Christensen, IROS'12

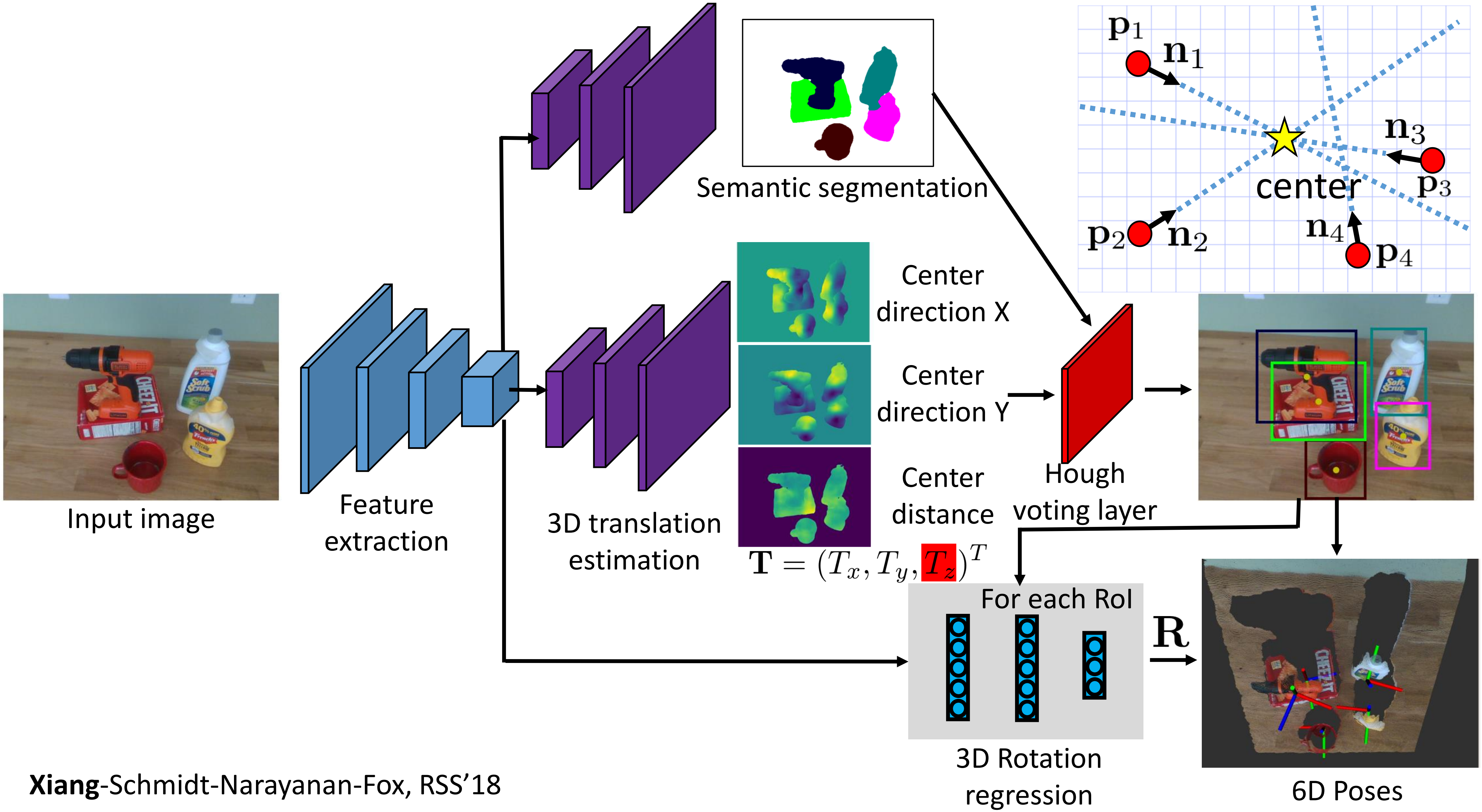
PoseCNN: the First End-to-end 6D Pose Estimation Network

PoseCNN

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

Xiang-Schmidt-Narayanan-Fox, RSS'18

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

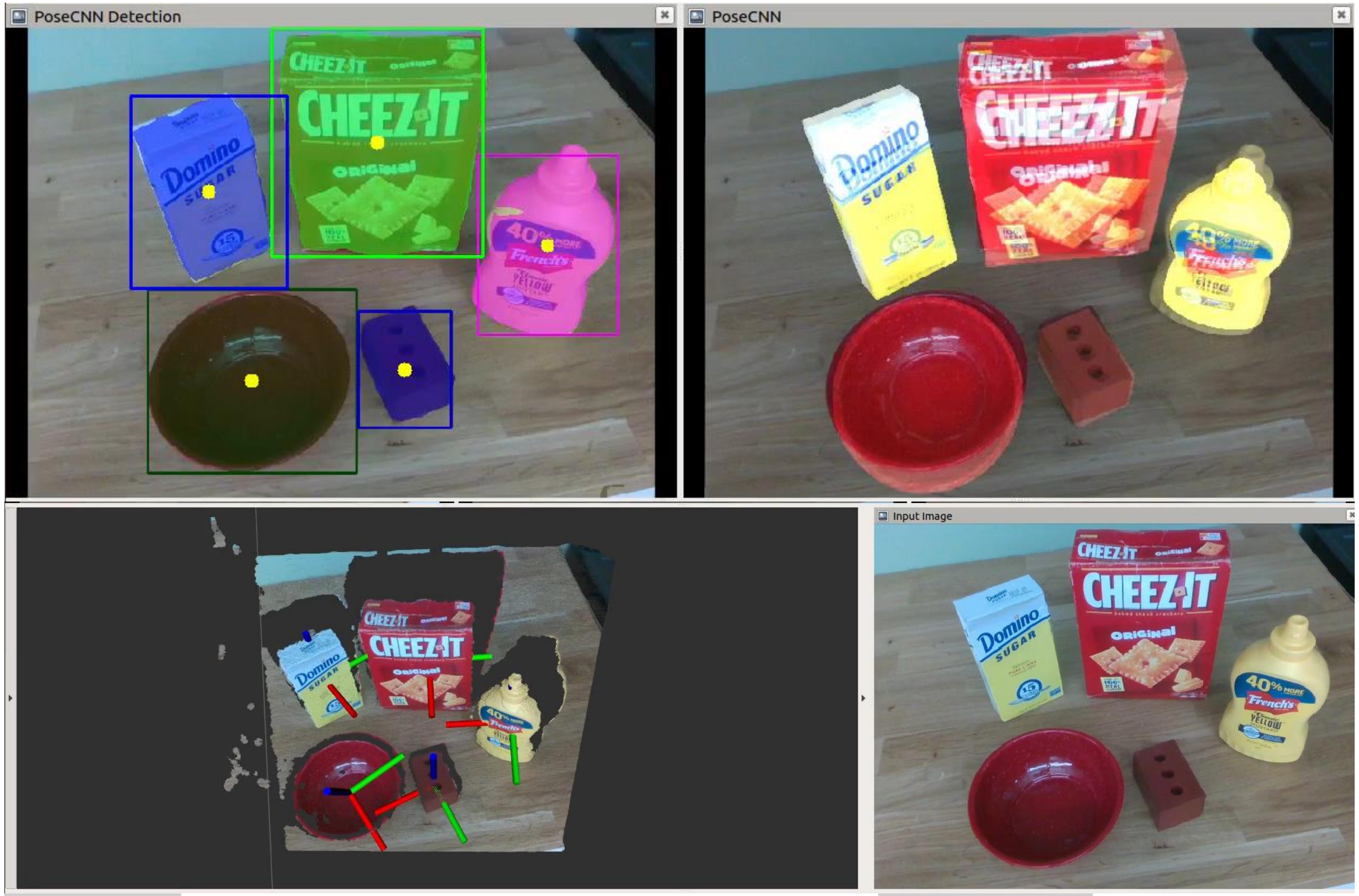


Xiang-Schmidt-Narayanan-Fox, RSS'18

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

Segmentation and Detection

Poses

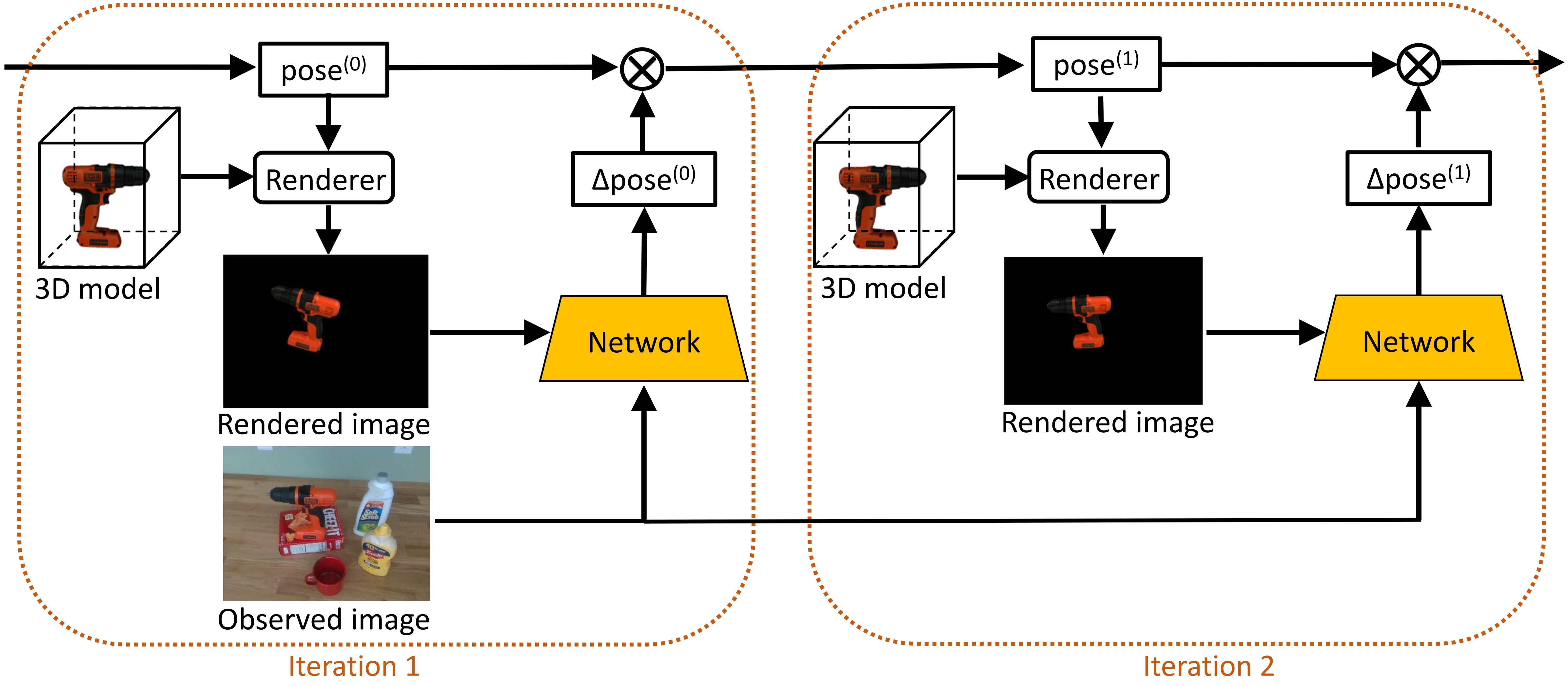


3D World

Input image

Xiang-Schmidt-Narayanan-Fox, RSS’18

DeepIM: Deep Iterative Matching for 6D Pose Estimation



Li-Wang-Ji-**Xiang**-Fox, ECCV'18 Oral, IJCV'19

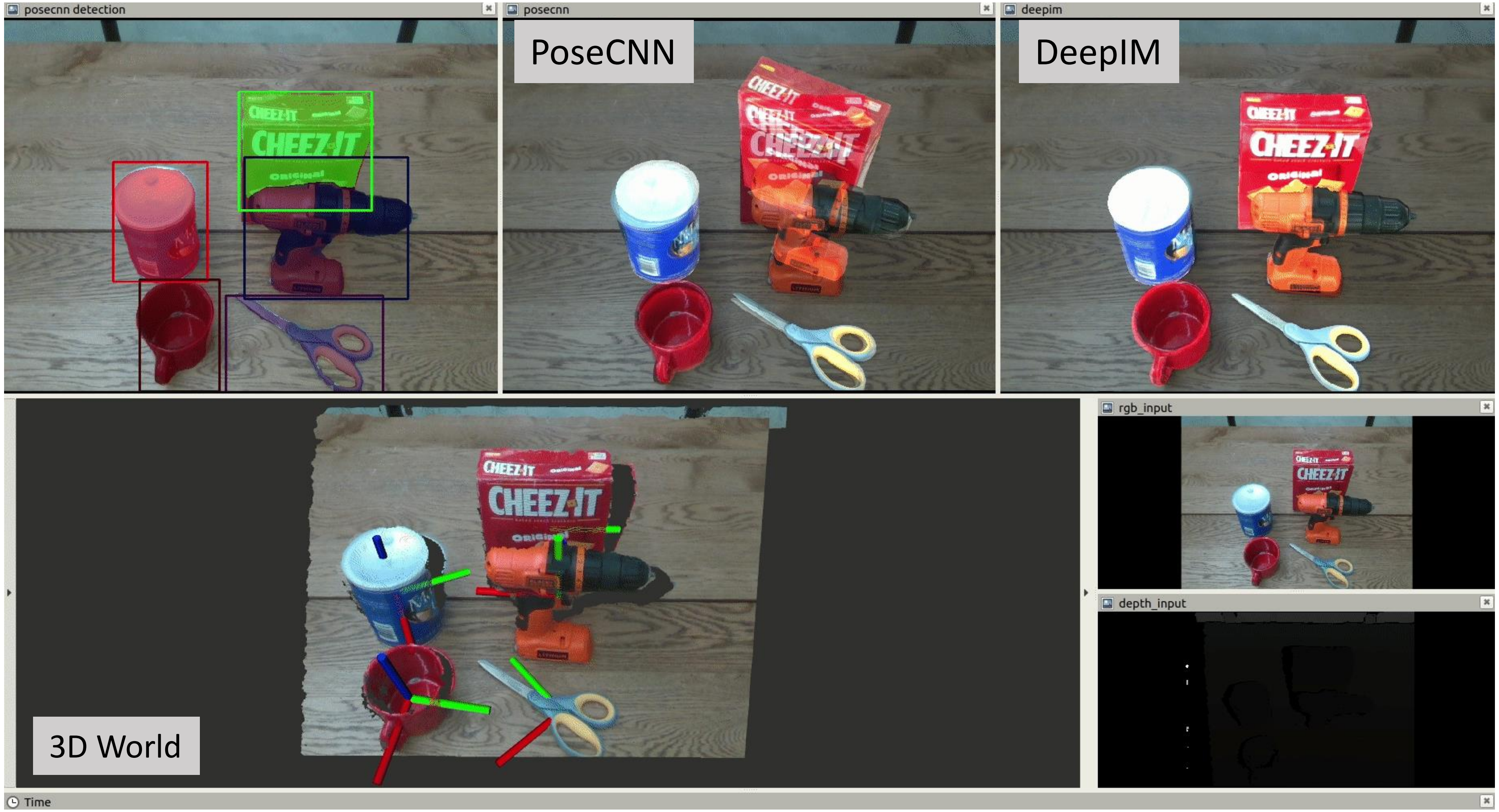
CosyPose: Labbe-Carpentier-Aubry-Sivic, ECCV'20

DeepIM: Deep Iterative Matching for 6D Pose Estimation

	RGB	
YCB Video	PoseCNN	PoseCNN+ DeepIM
Accuracy	75.9	**88.1**

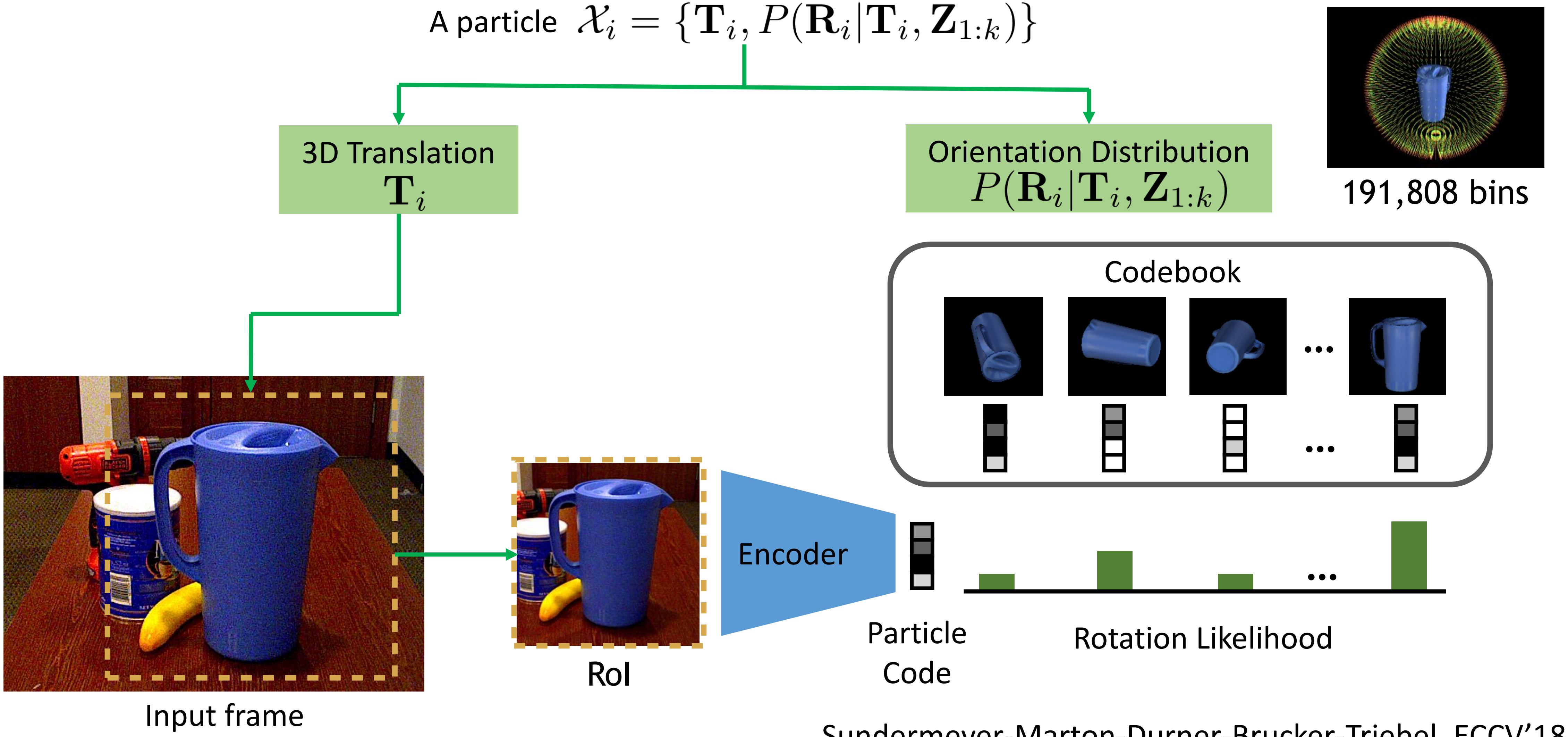
	RGB-D	
YCB Video	PoseCNN+ ICP	PoseCNN+ DeepIM
Accuracy	93.0	**94.0**

Real World



Li-Wang-Ji-**Xiang**-Fox, ECCV’18 Oral, IJCV’19

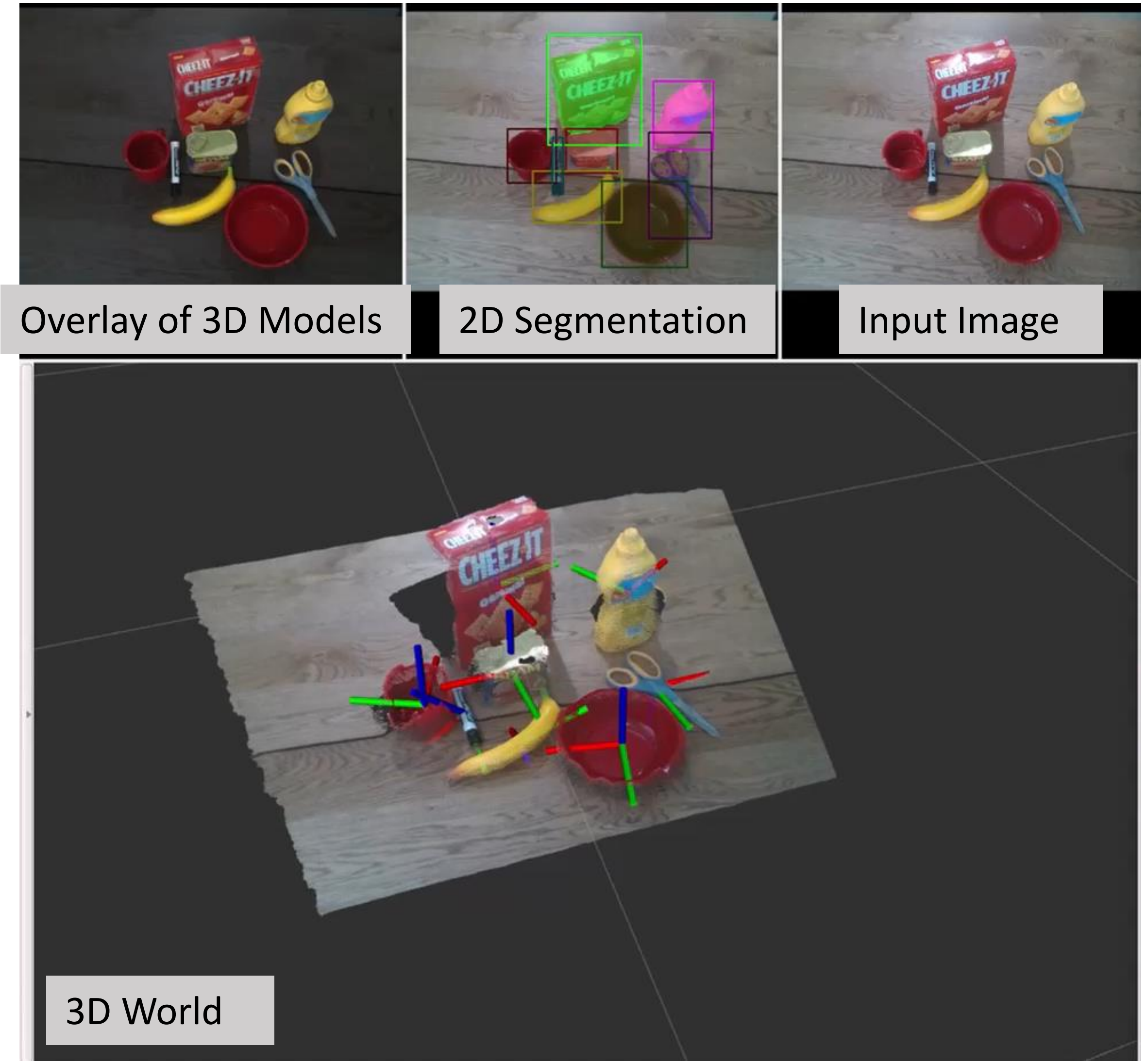
PoseRBPF: A Rao-Blackwellized Particle Filter for 6D Object Pose Tracking



Sundermeyer-Marton-Durner-Brucker-Triebel, ECCV'18

Deng-Mousavian-**Xiang**-Xia-Bretl-Fox, RSS'19, T-RO'21

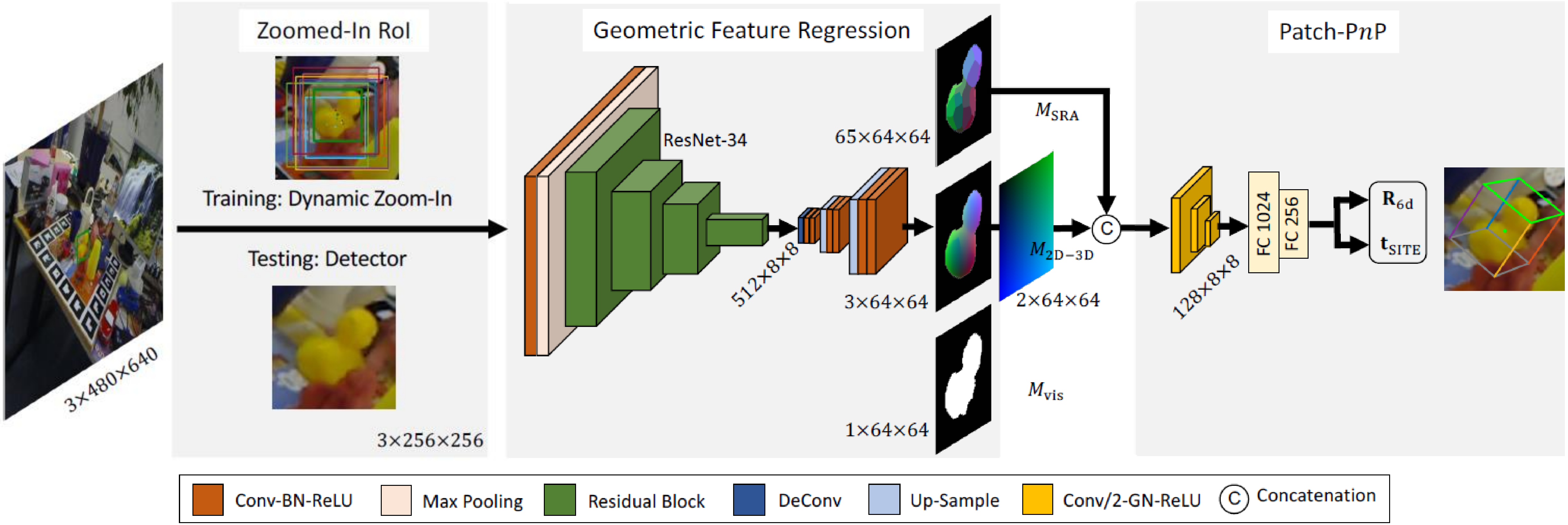
PoseRBPF: A Rao-Blackwellized Particle Filter for 6D Object Pose Tracking



With SDF-based depth refinement

Deng-Mousavian-**Xiang**-Xia-Bretl-Fox, RSS'19, T-RO'21

GDR-Net: Geometry-guided Direct Regression Network



Overall best method in the BOP Challenge 2022

G. Wang, F. Manhardt, F. Tombari, and X. Ji, GDR-Net: Geometry-guided direct regression network for monocular 6d object pose estimation, in CVPR, 2021.

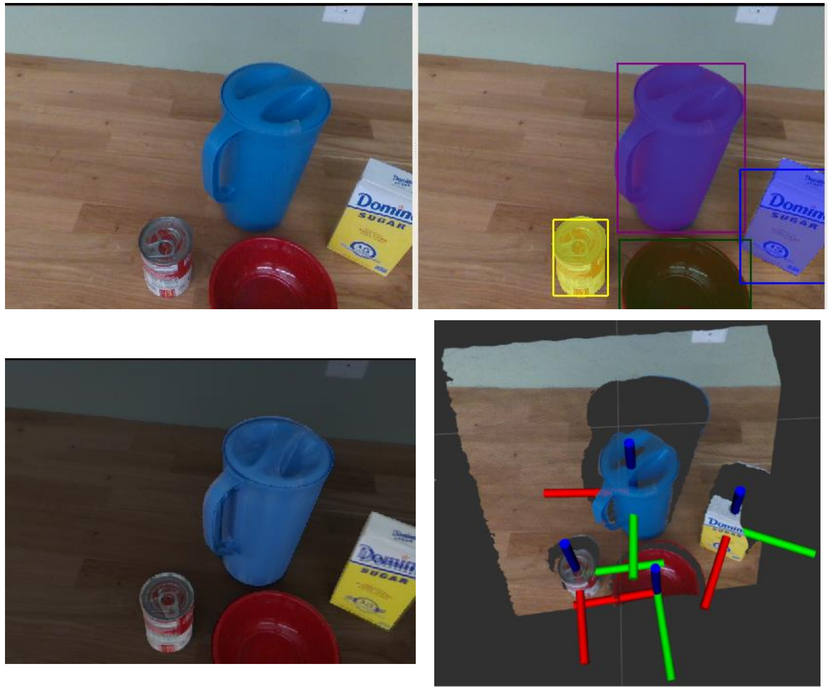
Model-Based Robot Manipulation

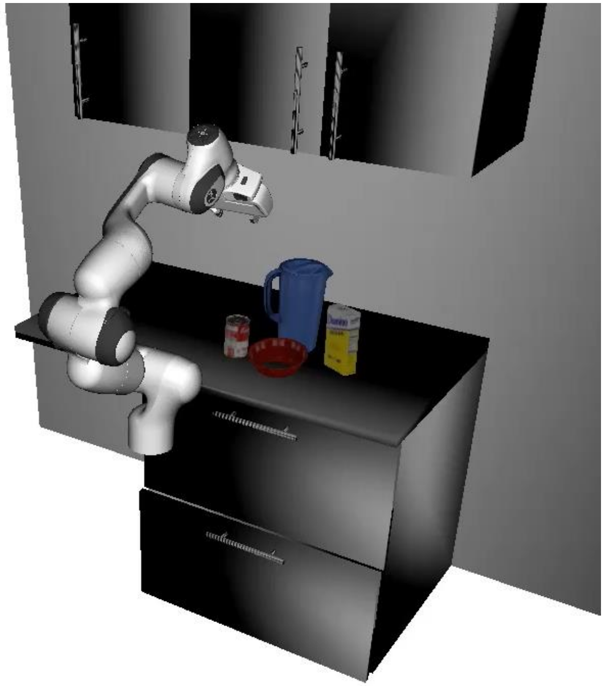
Perception → Planning → Control

6D object pose estimation

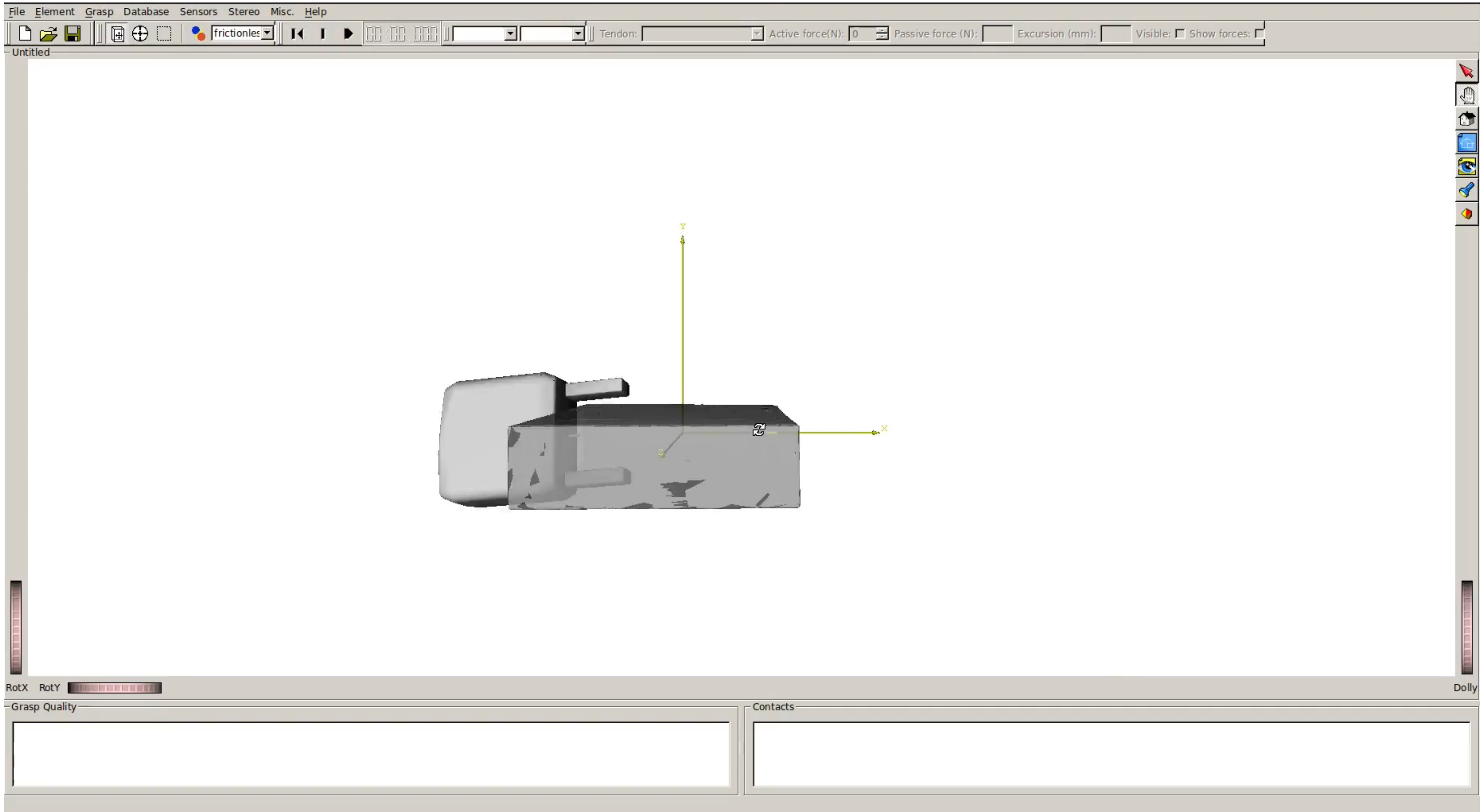
Grasp planning and motion planning

Manipulation trajectory following



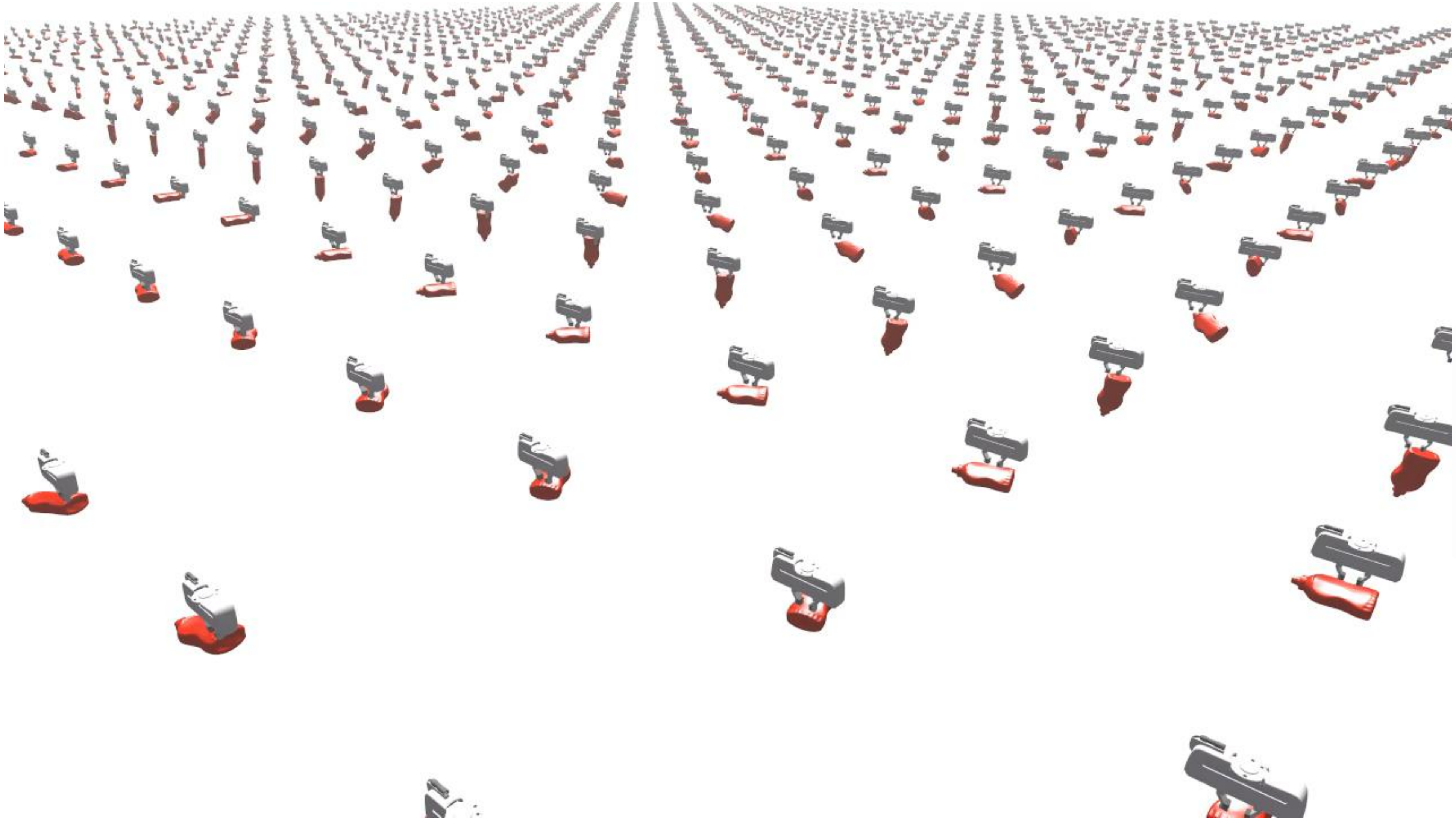


Grasp Planning: GraspIt!



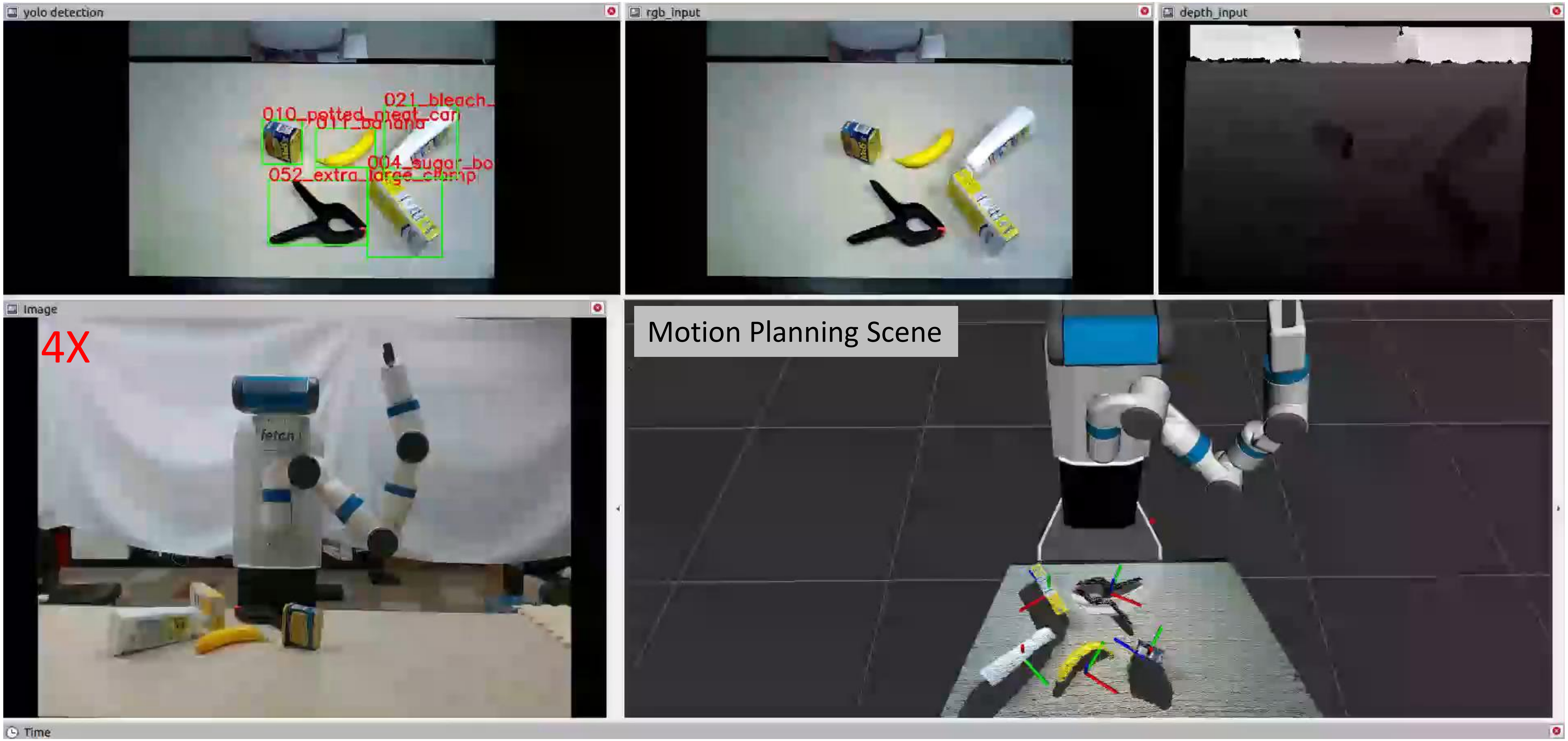
GraspIt! https://graspit-simulator.github.io/

Grasp Planning: A Physics-based Approach



Eppner-Mousavian-Fox, ISRR'19

Motion Planning

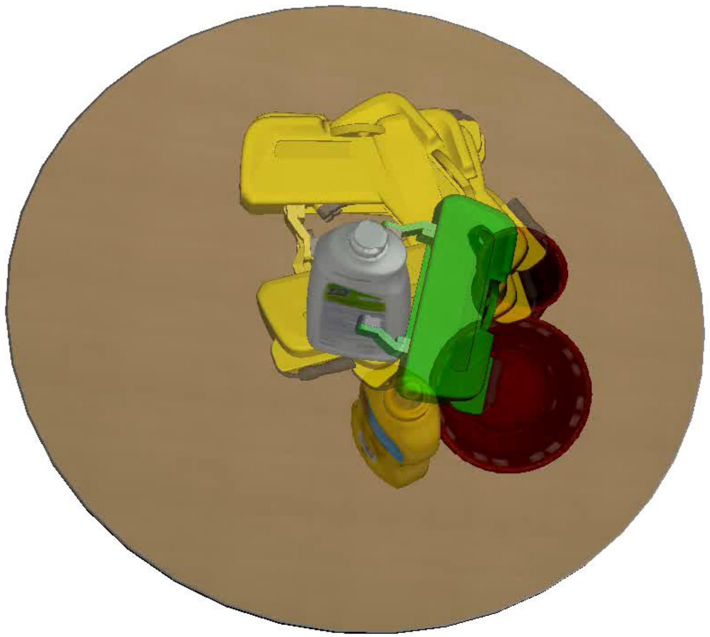


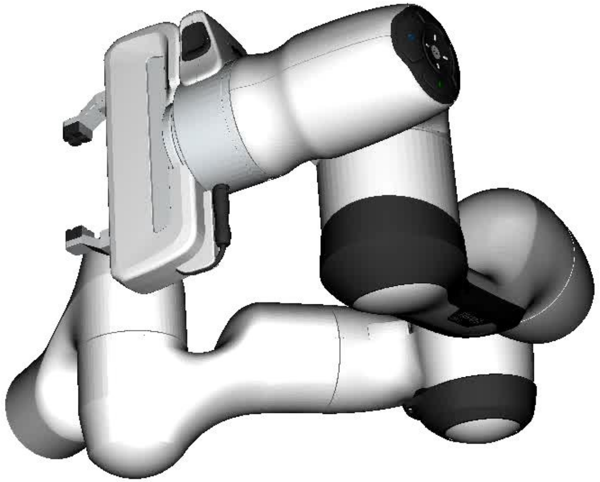
The Open Motion Planning Library in MoveIt https://ompl.kavrakilab.org/index.html

OMG Planner: Trajectory Optimization and Grasp Selection

OMG Iter: 50

100 grasps





Modeling the goal set distribution

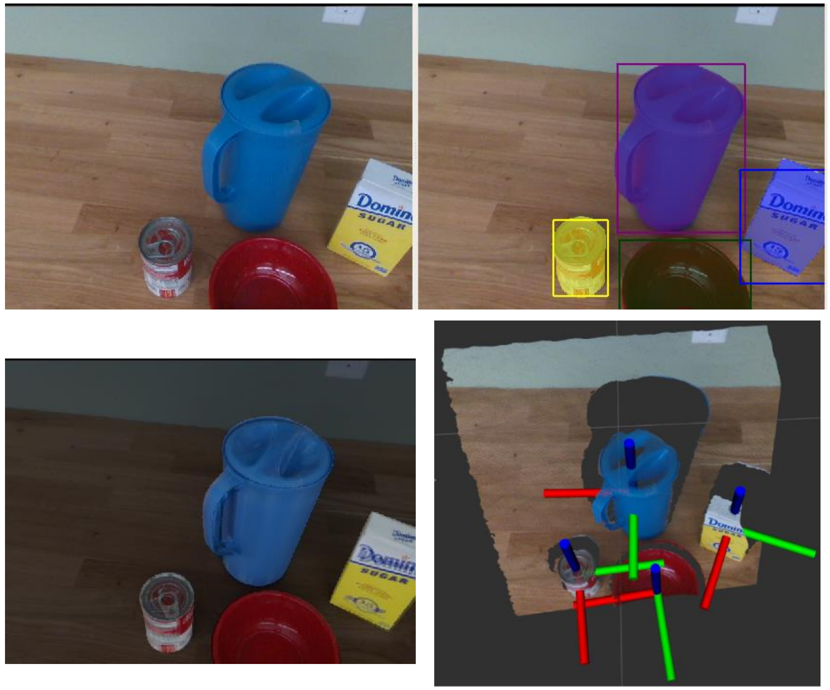
Code available online

Model-Based Robot Manipulation

Perception → Planning → Control

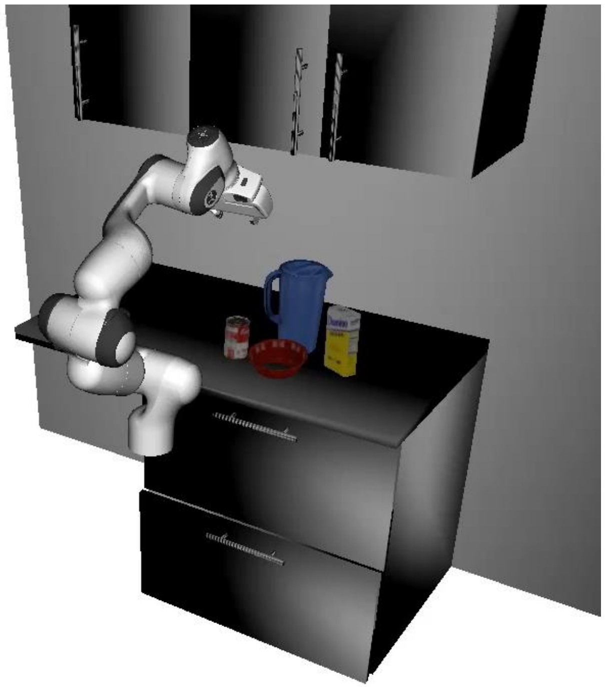
Perception

6D object pose estimation



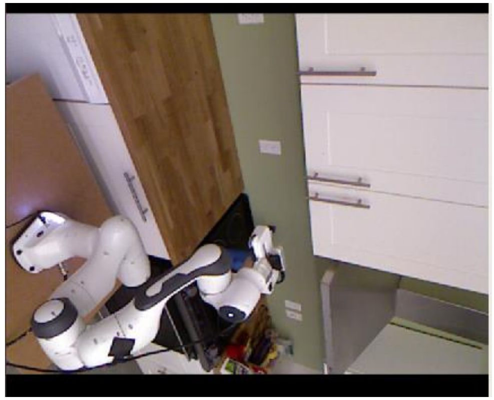
Planning

Grasp planning and motion planning

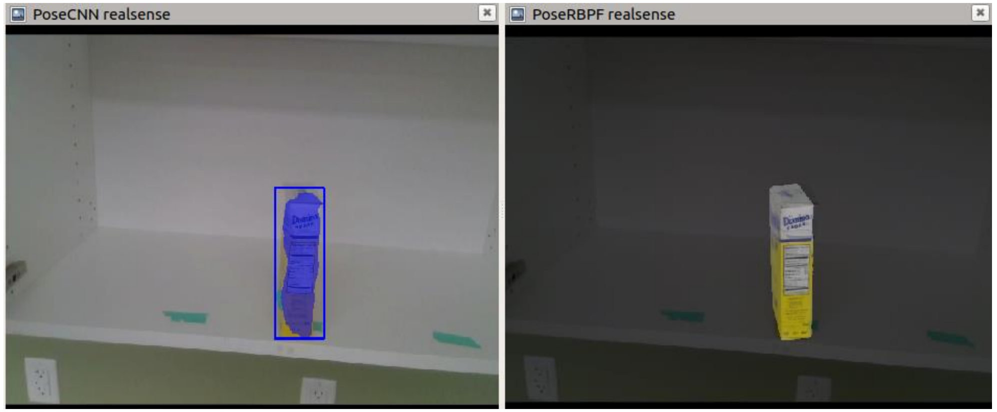


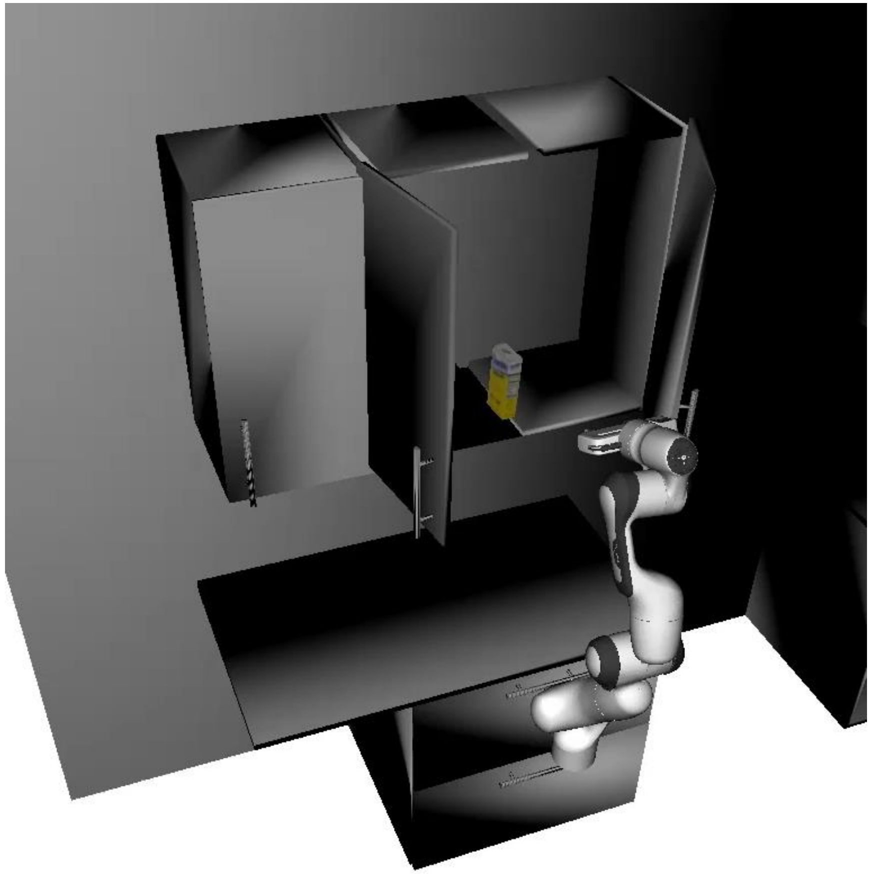
Control

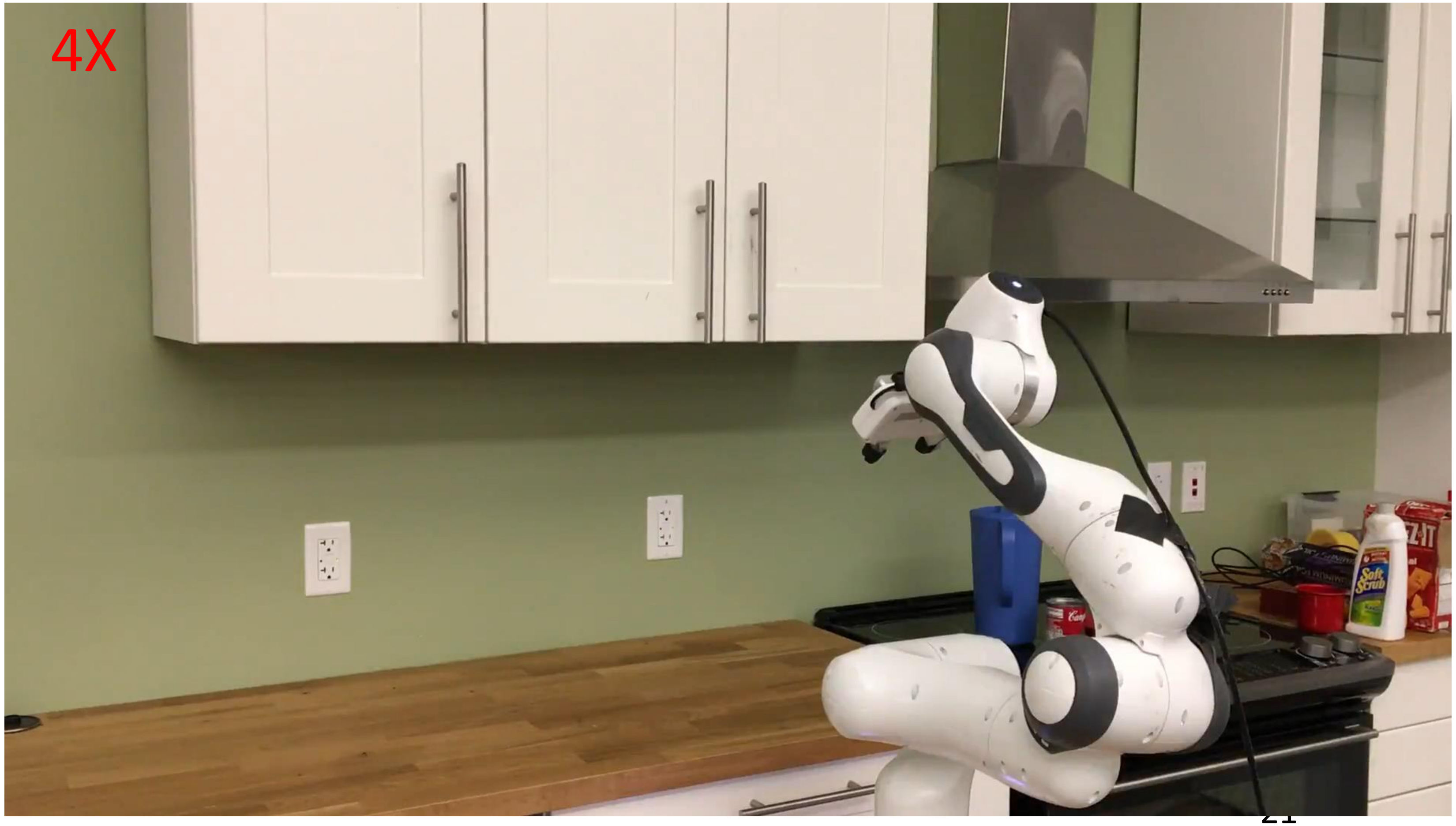
Manipulation trajectory following




PoseCNN realsense
PoseRBPF realsense

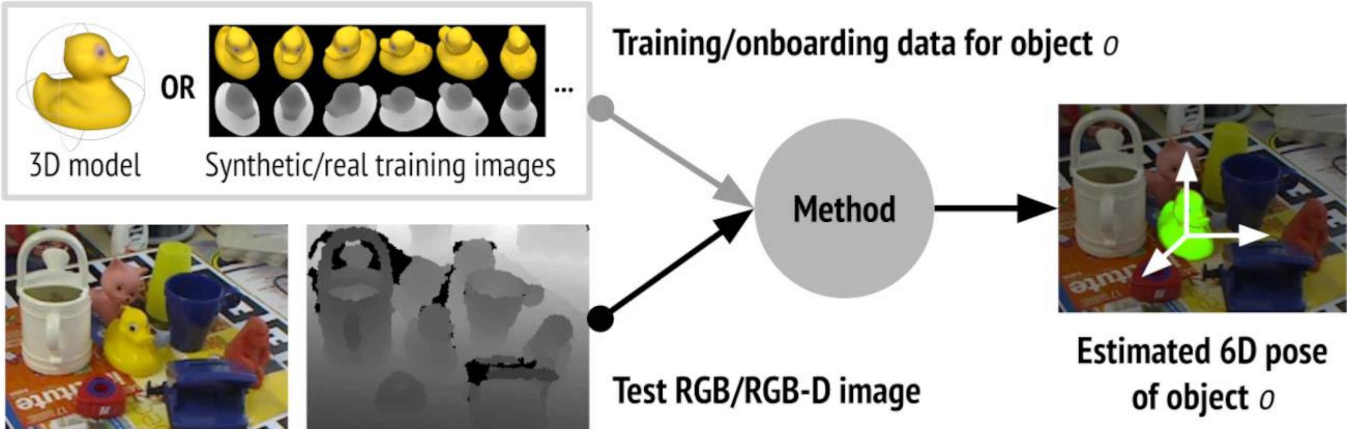

PoseCNN realsense
PoseRBPF realsense




4X

Benchmarking

- 6D object pose
  - BOP: Benchmark for 6D object pose estimation https://bop.felk.cvut.cz/home/



Datasets: Core datasets LM LM-O T-LESS ITODD HB HOPE YCB-V RU-APC IC-BIN IC-MI TUD-L TYO-L

6D localization of seen objects – Core datasets

This leaderbord shows the overall ranking for Task 1 on the core datasets (LM-O, T-LESS, TUD-L, IC-BIN, ITODD, HB, YCB-V). For each method, the date of the latest considered submission is reported. If more submissions of a method are available for a dataset, the submission with the highest AR_{Core} score is considered. The reported time is the average image processing time averaged over the core datasets.

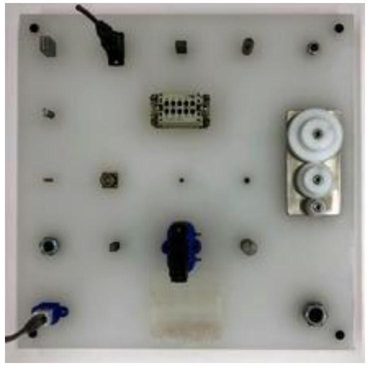
Show 50 entries Search:

	Date (UTC)	Method	Test image	AR_{Core}	AR_{LM-O}	AR_{T-LESS}	AR_{TUD-L}	AR_{IC-BIN}	AR_{ITODD}	AR_{HB}	AR_{YCB-V}	Time (s)
1	2023-09-24	GPose2023-OfficialDet	RGB-D	0.851	0.805	0.895	0.966	0.734	0.687	0.944	0.929	4.575
2	2022-10-15	GDRNPP-PBRReal-RGBD-MModel	RGB-D	0.837	0.775	0.874	0.966	0.722	0.679	0.926	0.921	6.263

Benchmarking

- Robot manipulation

Benchmark	Type	Task	Objects	AR Tag-Free	Scene Reproducibility
Meta-World [11]	Simulation	50 tasks	Synthetic	✓	✓
RLBench [12]	Simulation	100 Tasks	Synthetic	✓	✓
robosuite [13]	Simulation	9 Tasks	Synthetic	✓	✓
Grasp Planning Protocol [10]	Real	Grasp Planning	YCB (single)	✓	✗
NIST Assembly [7]	Real	Assembly	Task Boards	✓	✓
FurnitureBench [14]	Real	Assembly	3D Printing	✗	✓
GRASPA [8]	Real	Grasping	YCB (clutter)	✗	✓
OCRTOC [15]	Real	Rearrangement	YCB + Others	✓	✗
RB2 [16]	Real	Pouring, Scooping, Zipping, Insertion	Others	✓	✗
Box and Blocks Test [17]	Real	Pick-and-Place	Blocks	✓	✗
SceneReplica (Ours)	Real	Pick-and-Place	YCB (clutter)	✓	✓



NIST Assembly board

FurnitureBench

GRASPA

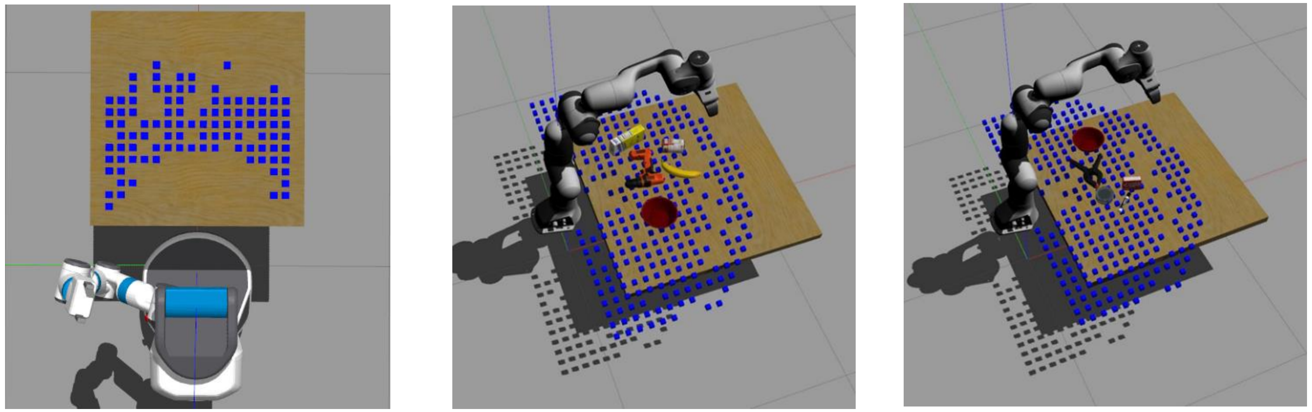
SceneReplica

SceneReplica for Real-World Robot Manipulation

- 16 YCB objects

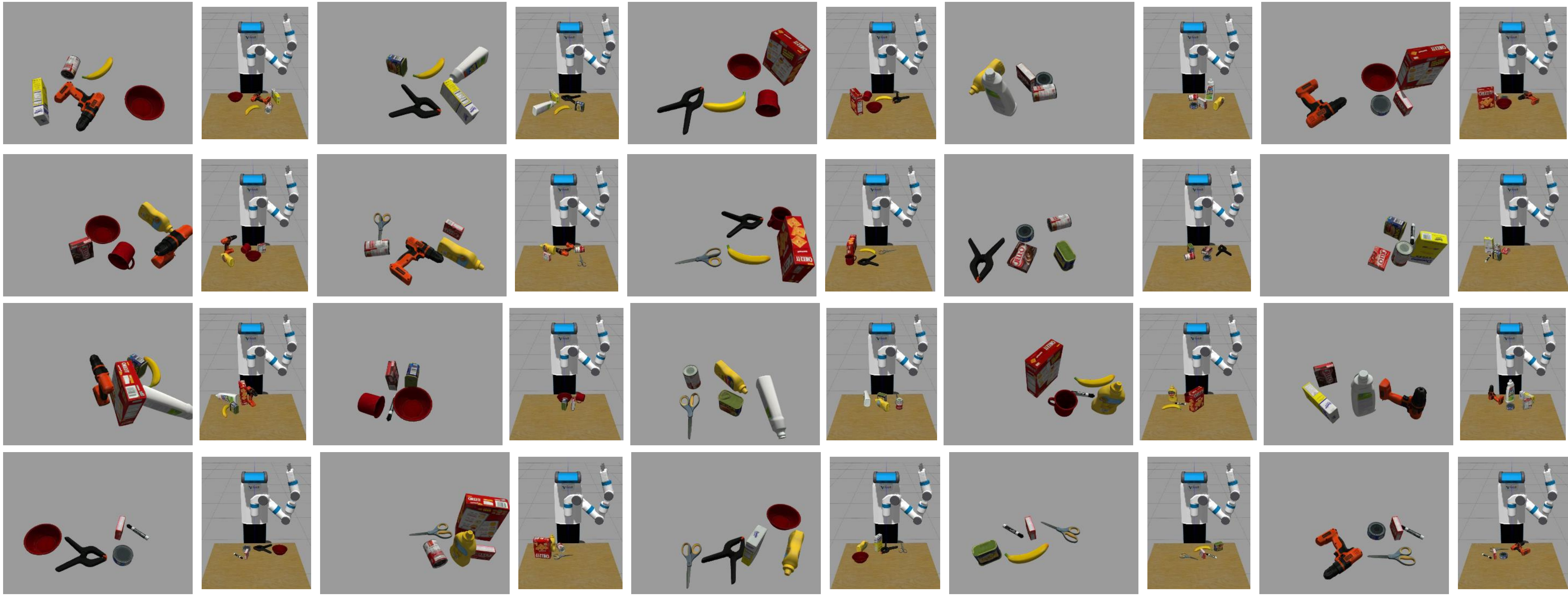
- Stable poses

- Reachability testing in simulation



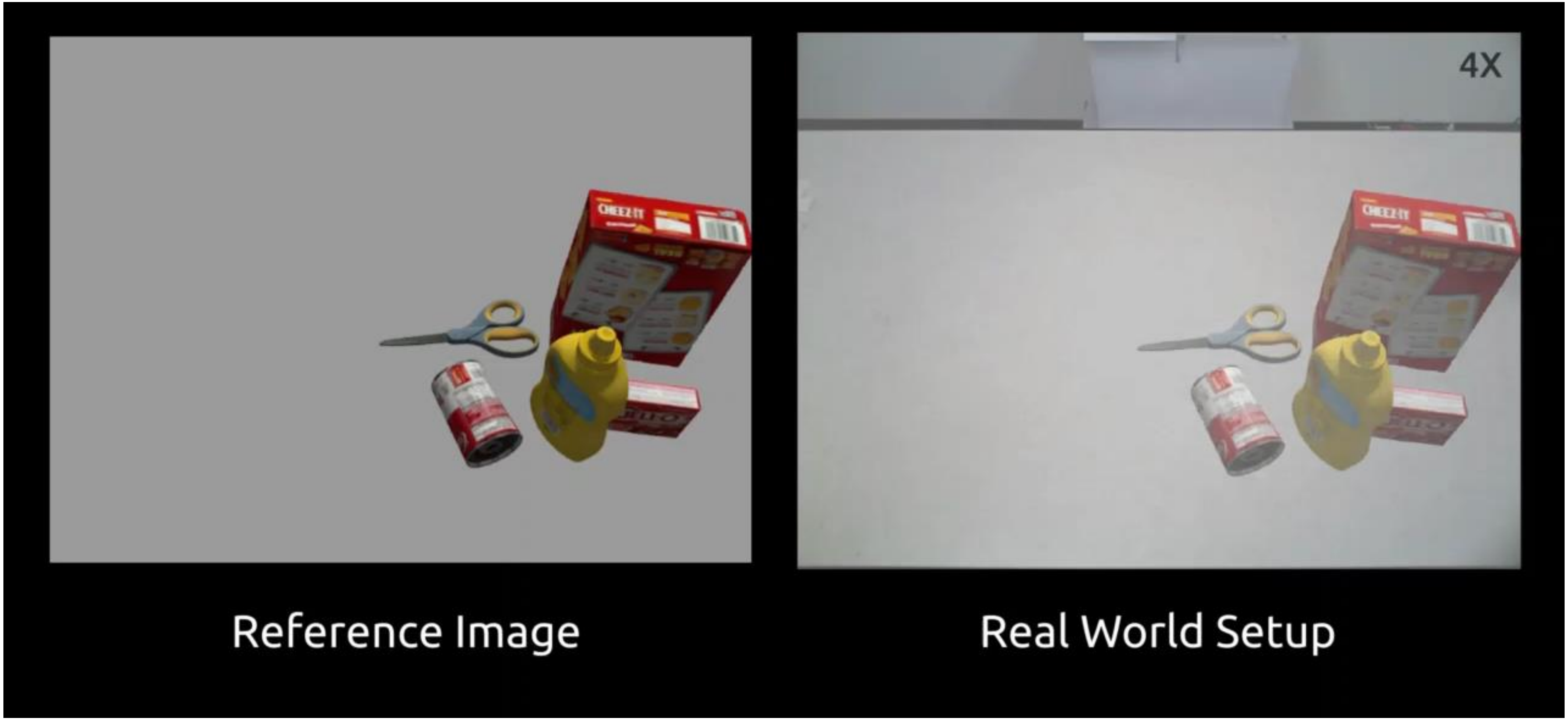
SceneReplica for Real-World Robot Manipulation

20 Scenes



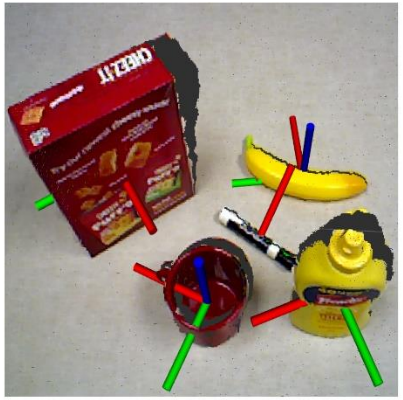
SceneReplica: https://irvlutd.github.io/SceneReplica/

Real-World Scene Setup



Model-based Grasping vs Model-free Grasping

Input real world scene

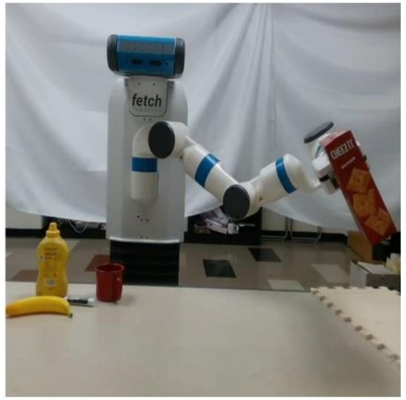


6D Pose Estimation

Offline Grasp Database

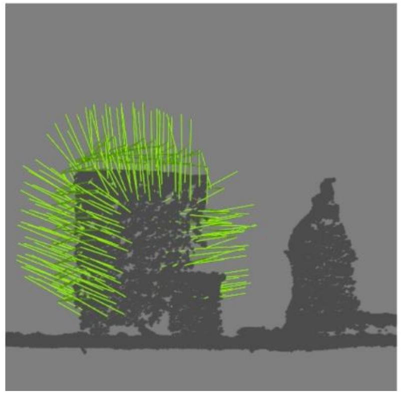
Motion Planning Setup

Grasping & Lifting

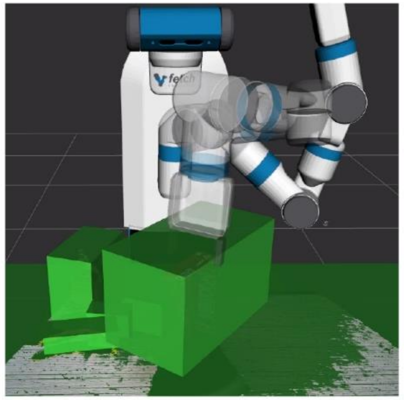


Moving arm for Dropoff

Unseen Object Segmentation

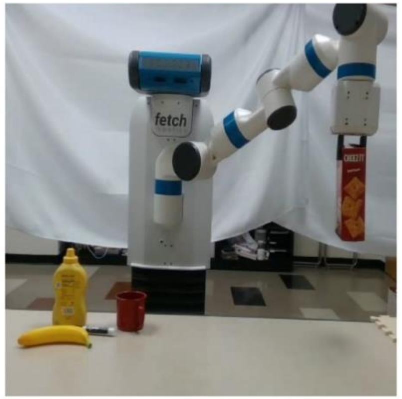


Model-free Grasp Planning



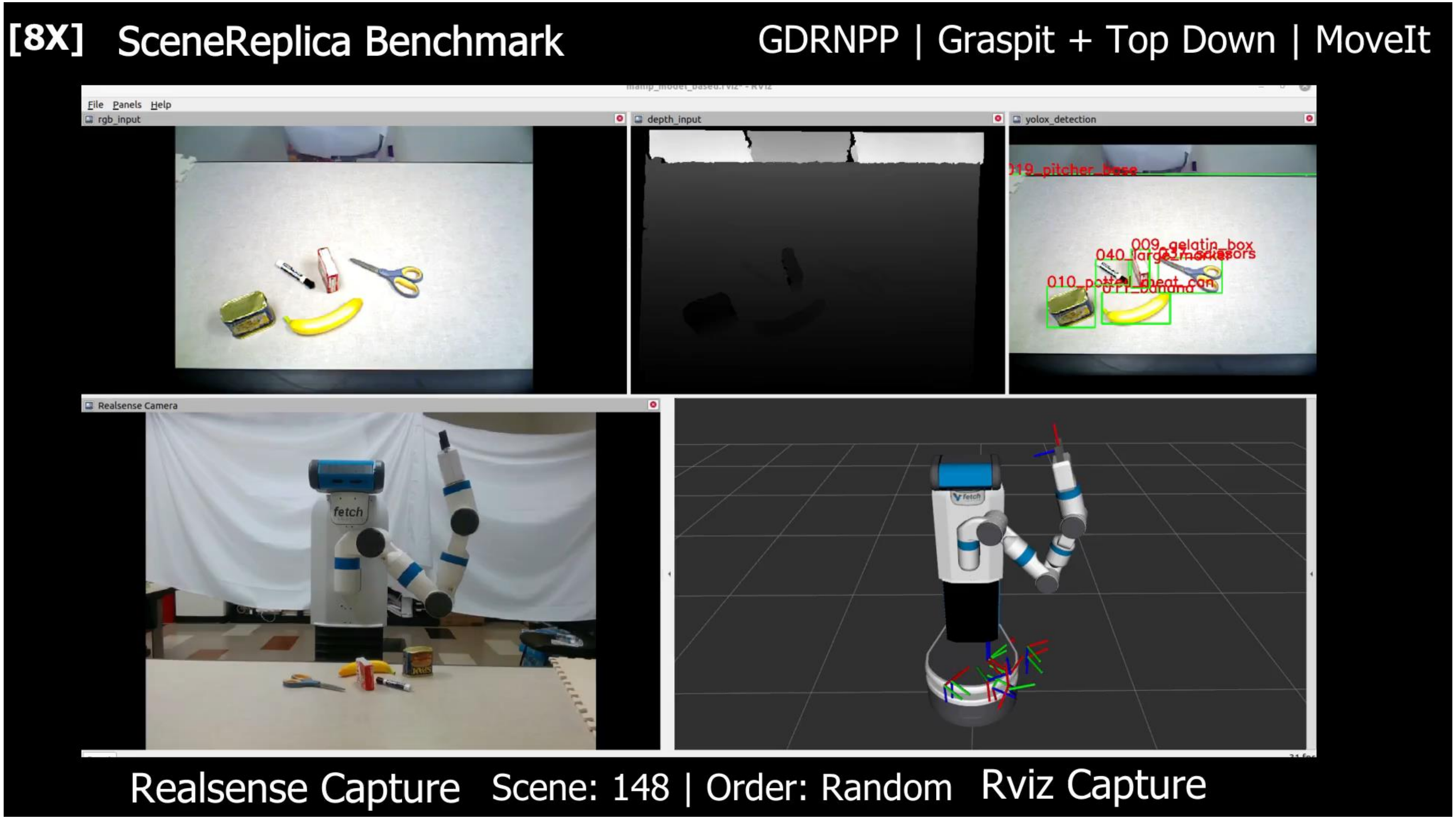
Motion Planning Setup

Grasping & Lifting

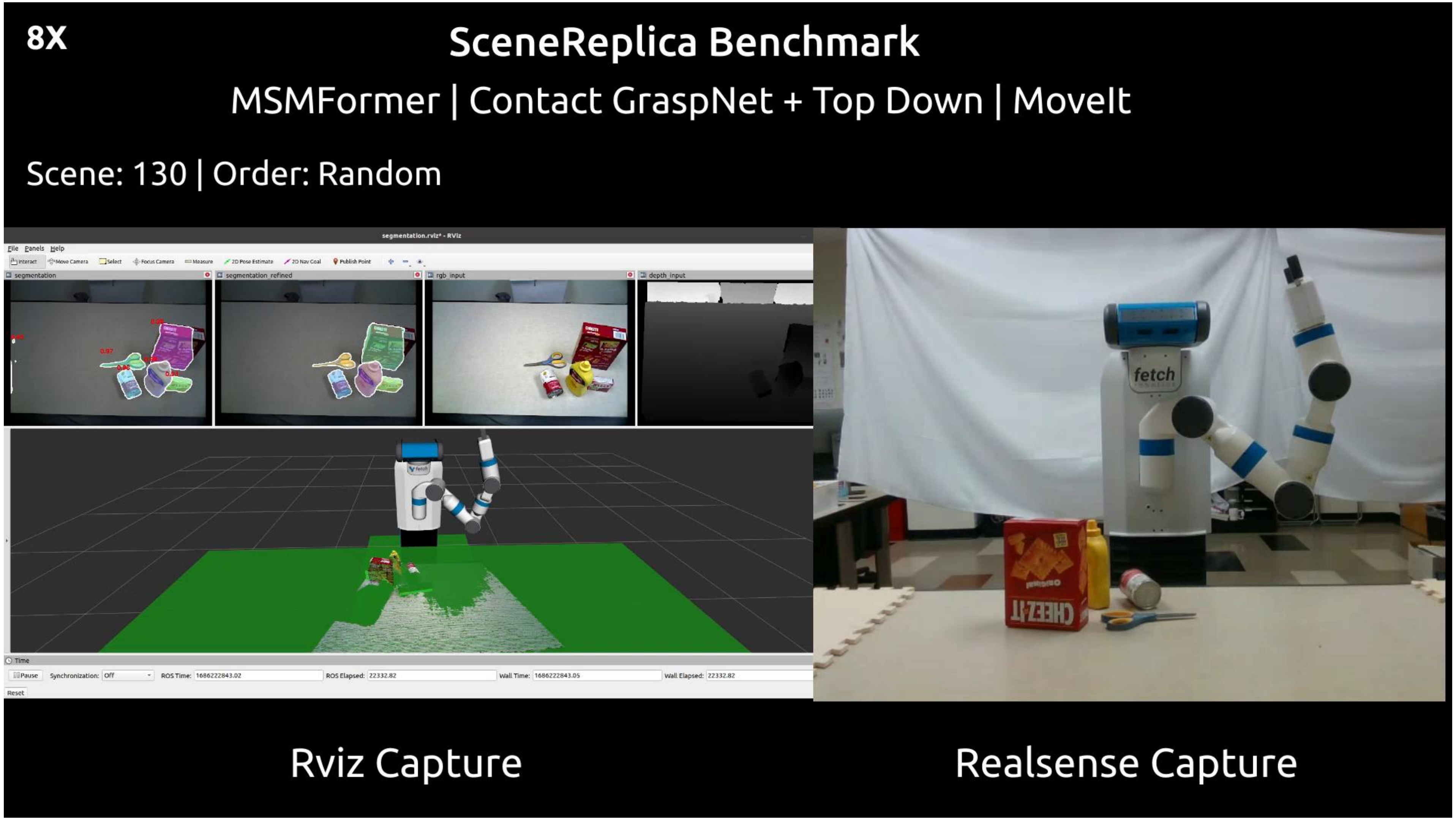


Moving arm for Dropoff

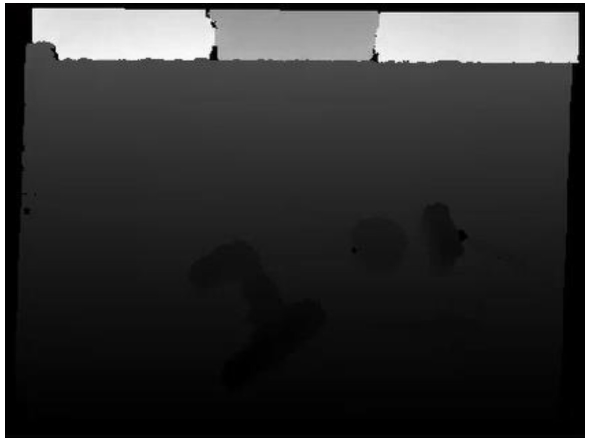
Model-based Grasping Example



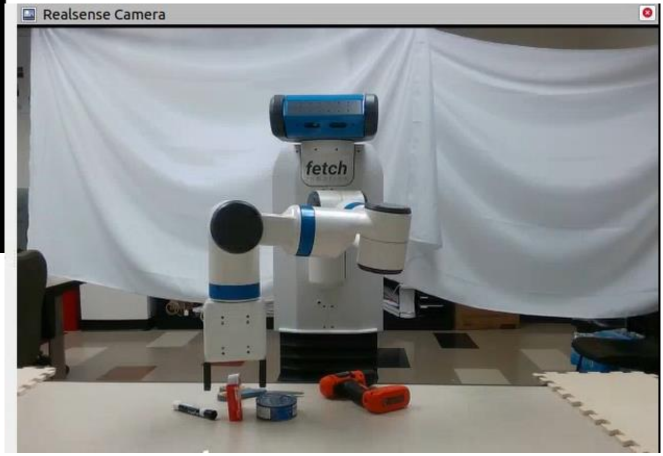
Model-free Grasping Example



End-to-end Learning-based Grasping

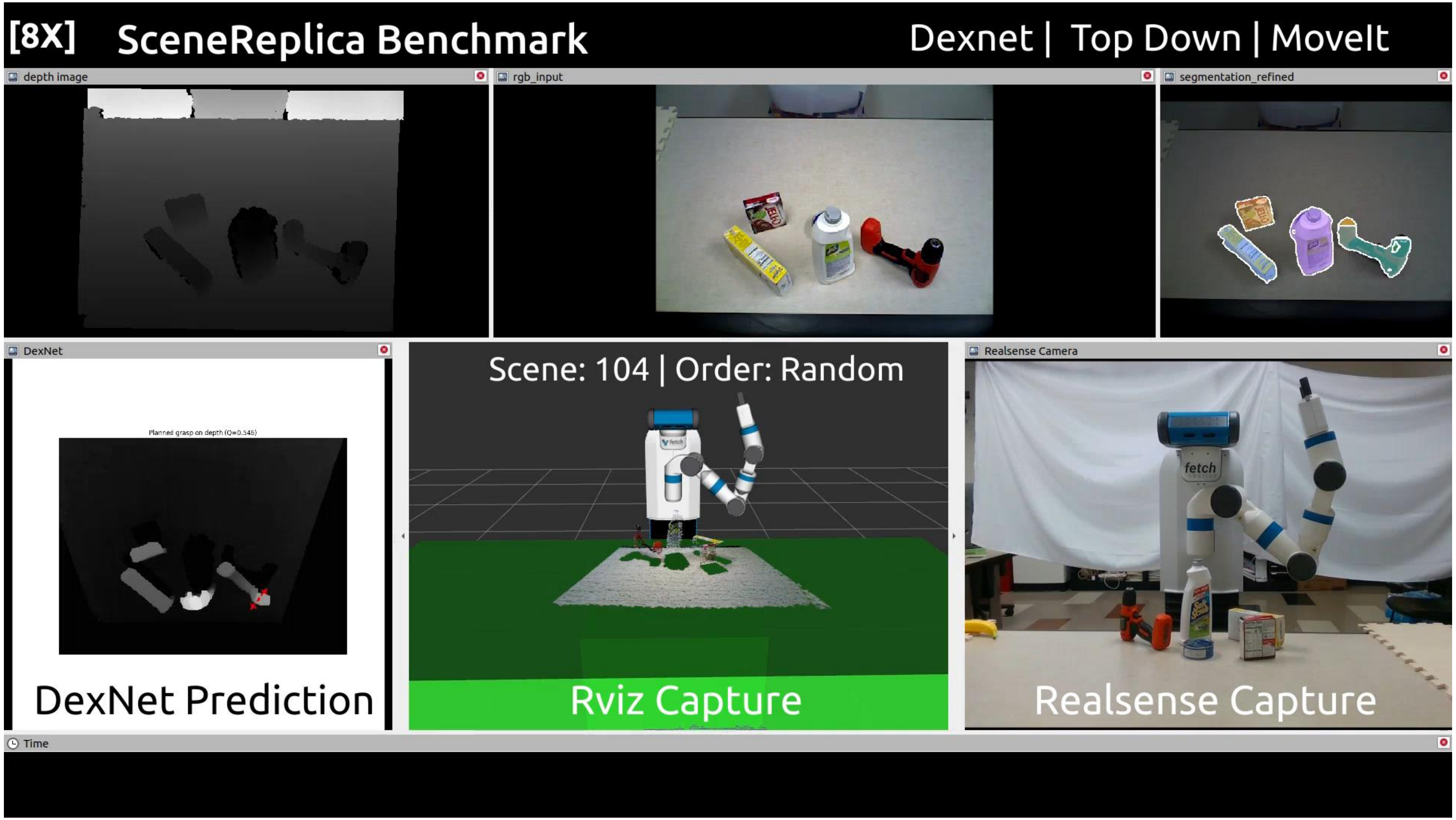


Dex-Net 2.0



J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, “Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics”, arXiv preprint arXiv:1703.09312, 2017.

Dex-Net 2.0 Grasping Example



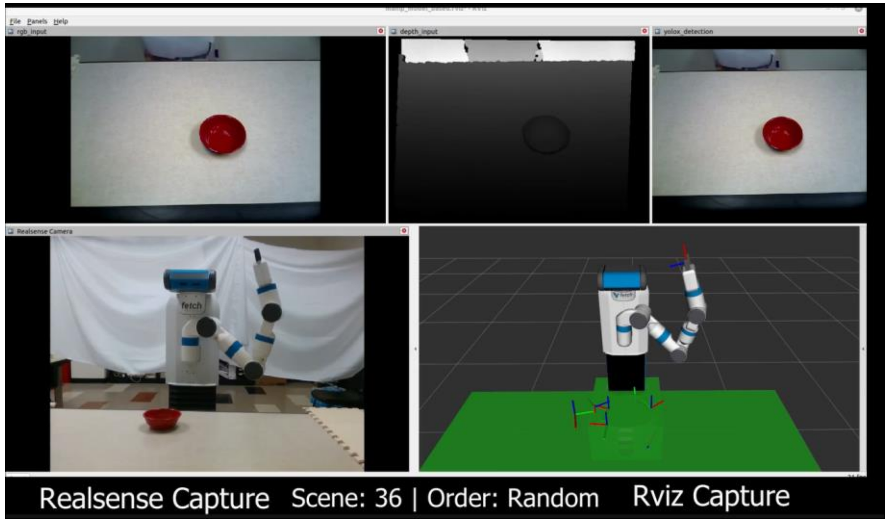
Current Leaderboard

Overall best method in the BOP Challenge 2022

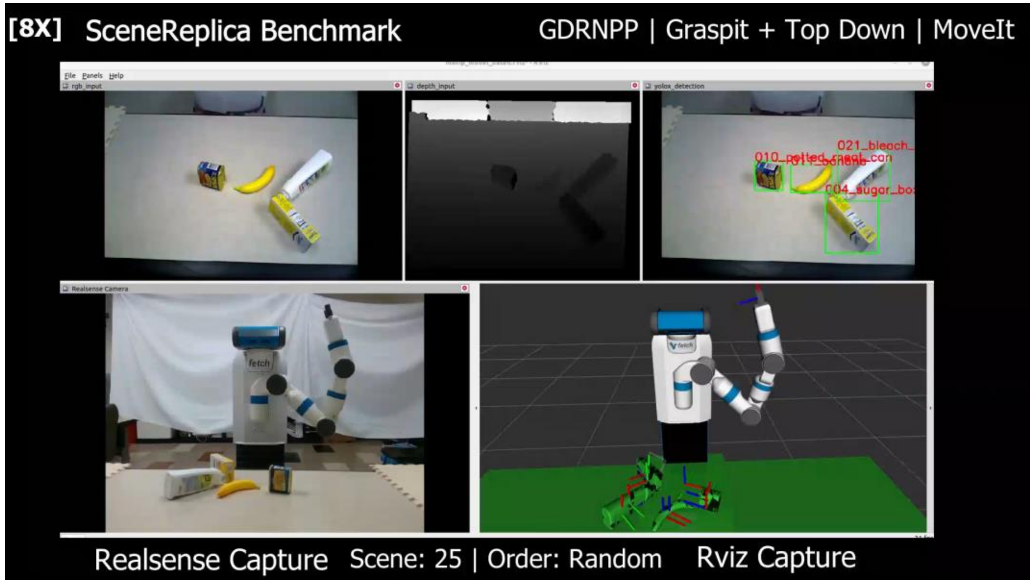
#	Perception	Grasp Planning	Motion Planning	Control	Ordering	Grasping Type	Pick & Place Success 🥇	Grasping Success	Videos
3	GDRNPP [9]	GraspIt! [2] + Top-Down	OMPL [3]	MoveIt	Near-to-Far	Model-Based	66/100	69/100	🔗
3	GDRNPP [9]	GraspIt! [2] + Top-Down	OMPL [3]	MoveIt	Fixed Random	Model-Based	62/100	64/100	🔗
7	MSMFormer [8]	Contact-graspnet [7] + Top-Down	OMPL [3]	MoveIt	Fixed Random	Model-Free	61/100	70/100	🔗
5	UCN [5]	Contact-graspnet [7] + Top-Down	OMPL [3]	MoveIt	Near-to-Far	Model-Free	60/100	63/100	🔗
5	UCN [5]	Contact-graspnet [7] + Top-Down	OMPL [3]	MoveIt	Fixed Random	Model-Free	60/100	64/100	🔗
1	PoseRBPF [1]	GraspIt! [2] + Top-Down	OMPL [3]	MoveIt	Fixed Random	Model-Based	59/100	59/100	🔗
1	PoseRBPF [1]	GraspIt! [2] + Top-Down	OMPL [3]	MoveIt	Near-to-Far	Model-Based	58/100	64/100	🔗

https://irvlutd.github.io/SceneReplica/

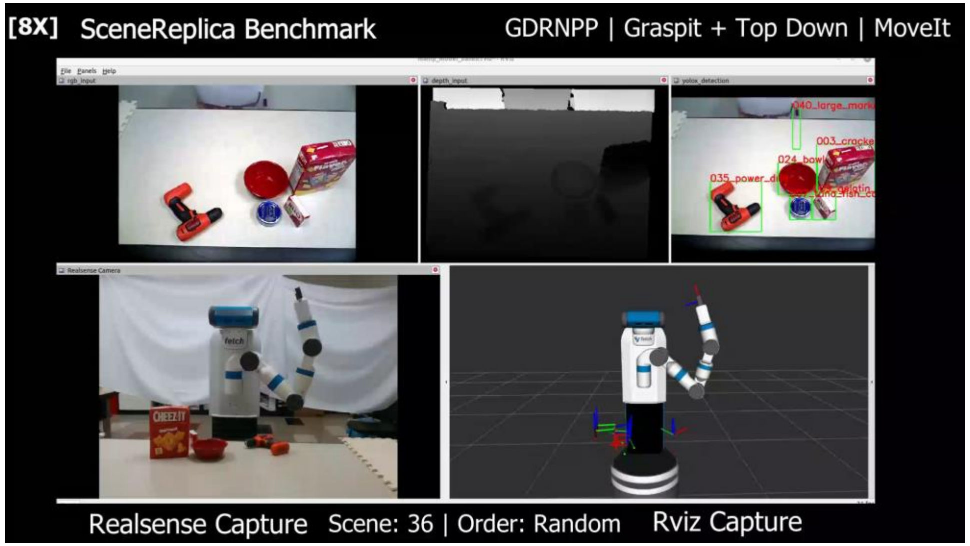
Failure Analysis (GDRNPP + GraspIt! + Top-Down)



Object Detection Error



Grasp Planning Error



Pose Estimation Error

Object	Method 3			
	S	P_EF	P_LF	EF
003 cracker box	3	2	1	-
004 sugar box	5	-	-	-
005 tomato soup can	5	1	-	1
006 mustard bottle	7	-	-	-
007 tuna fish can	1	5	-	-
008 pudding box	5	-	-	-
009 gelatin box	6	-	1	-
010 potted meat can	7	-	-	-
011 banana	6	-	1	-
021 bleach cleanser	3	1	-	1
024 bowl	2	4	1	-
025 mug	4	-	1	-
037 scissors	4	3	-	-
035 power drill	1	3	2	1
040 large marker	2	4	-	-
052 extra large clamp	5	1	-	-
ALL	66	24	7	3

P_EF: #perception failure

P_LF: #planning failure

EF: #execution failure

Conclusion

- 6D object pose estimation can facilitate robot manipulation
- The performance of 6D object pose estimation is not saturated yet
- Connecting BOP and SceneReplica to evaluate object pose estimation and robot grasping in the real world

Thank you!