OBJECT PERCEPTION FOR ROBOT MANIPULATION

Yu Xiang, 7/12/2019

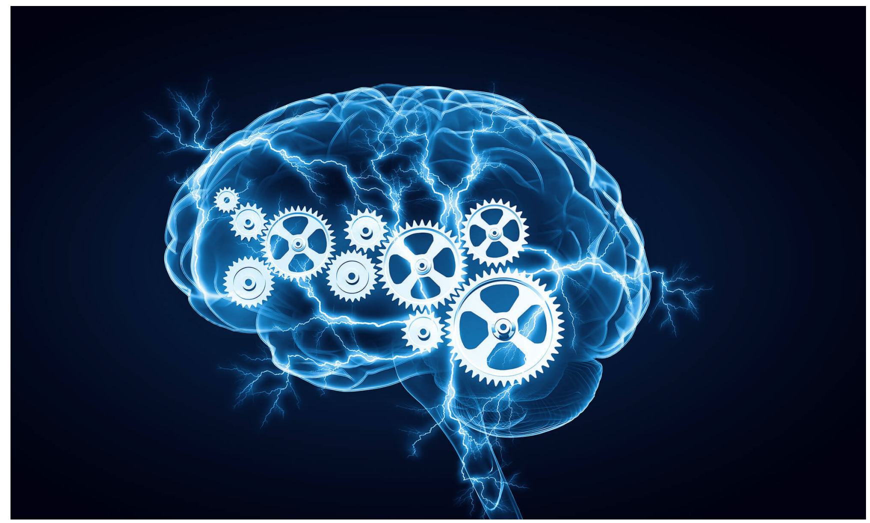
MANIPULATION

- The way of making physical changes to the world around us

Vs. question answering or
autonomous driving

MANIPULATION REQUIRES INTELLIGENCE

- Understanding the 3D environment from sensing
  - E.g., Vision, Tactile
- Grasp and motion planning / decision making
  - E.g., Obstacle avoidance
- Dynamics / Control
- Learning from experience



ROBOT MANIPULATION

Sensing | Perception | Planning | Control

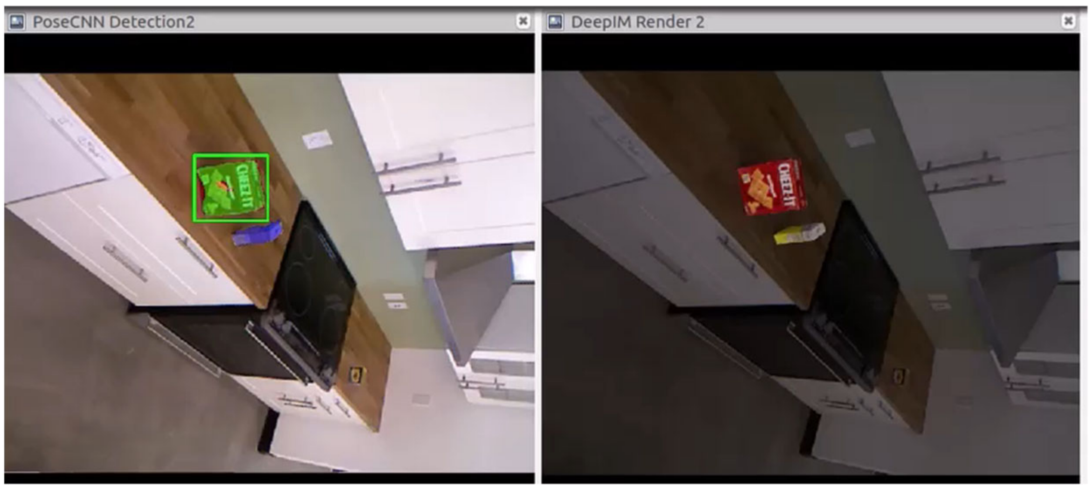
Learning

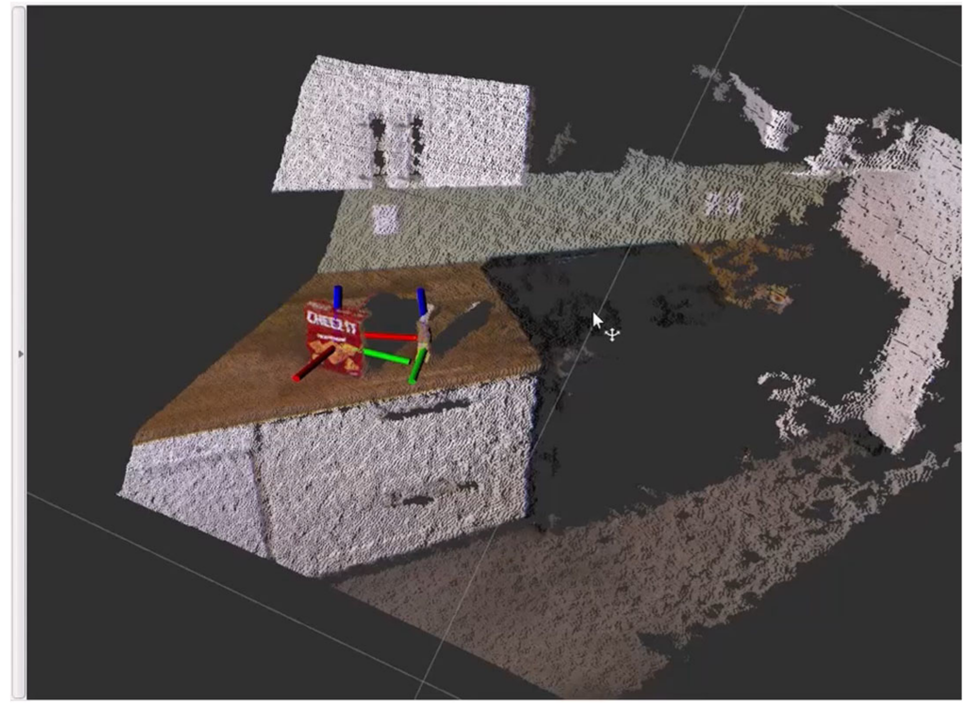
Intelligent System

3D Environment


CHEEZ-IT
Original

6D OBJECT POSE ESTIMATION



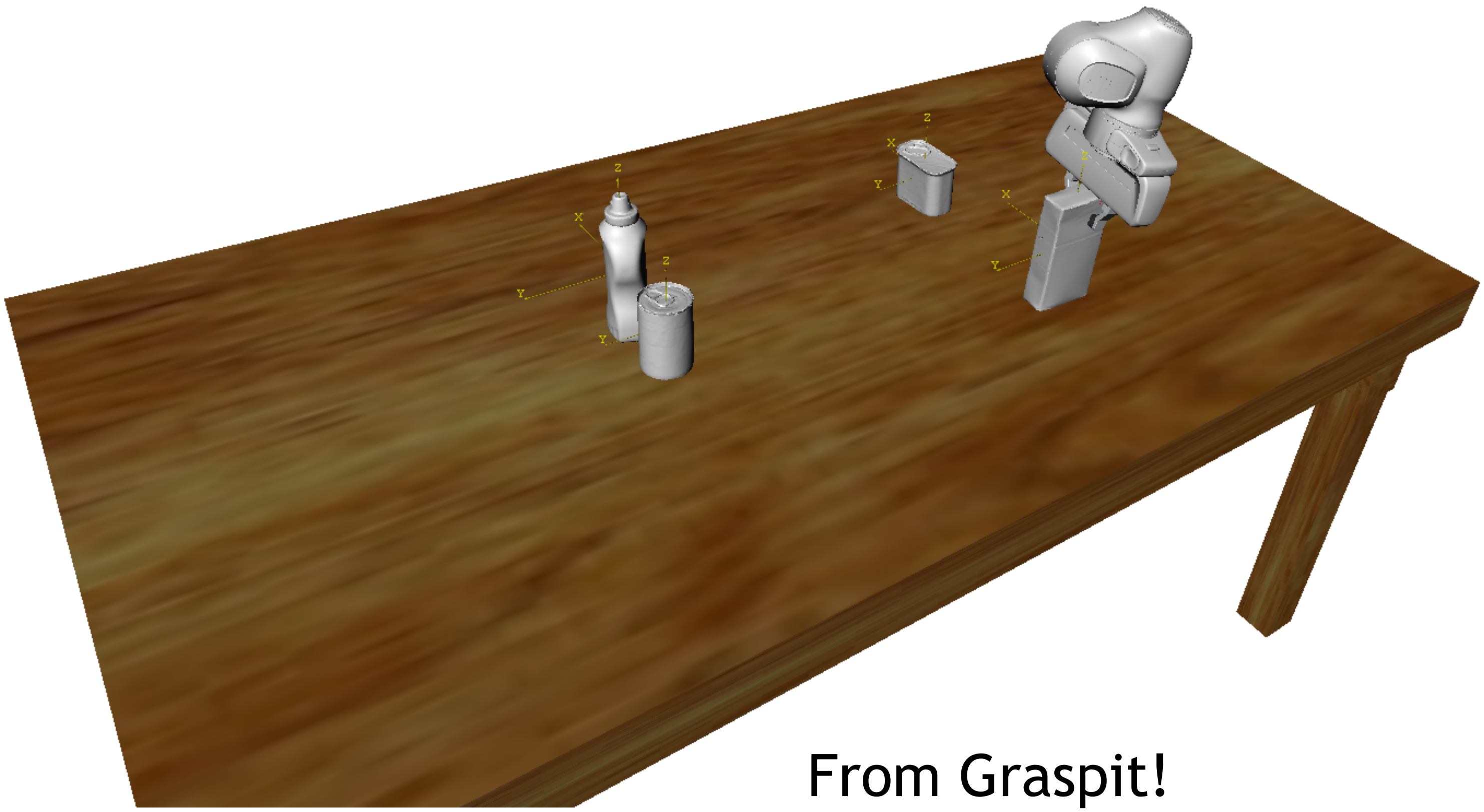


USING 3D MODELS OF OBJECTS

- The YCB Object and Model Set

B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel and A. M. Dollar, "The YCB object and Model set: Towards common benchmarks for manipulation research," International Conference on Advanced Robotics (ICAR), 2015.

6D POSE ESTIMATION FOR GRASP PLANNING



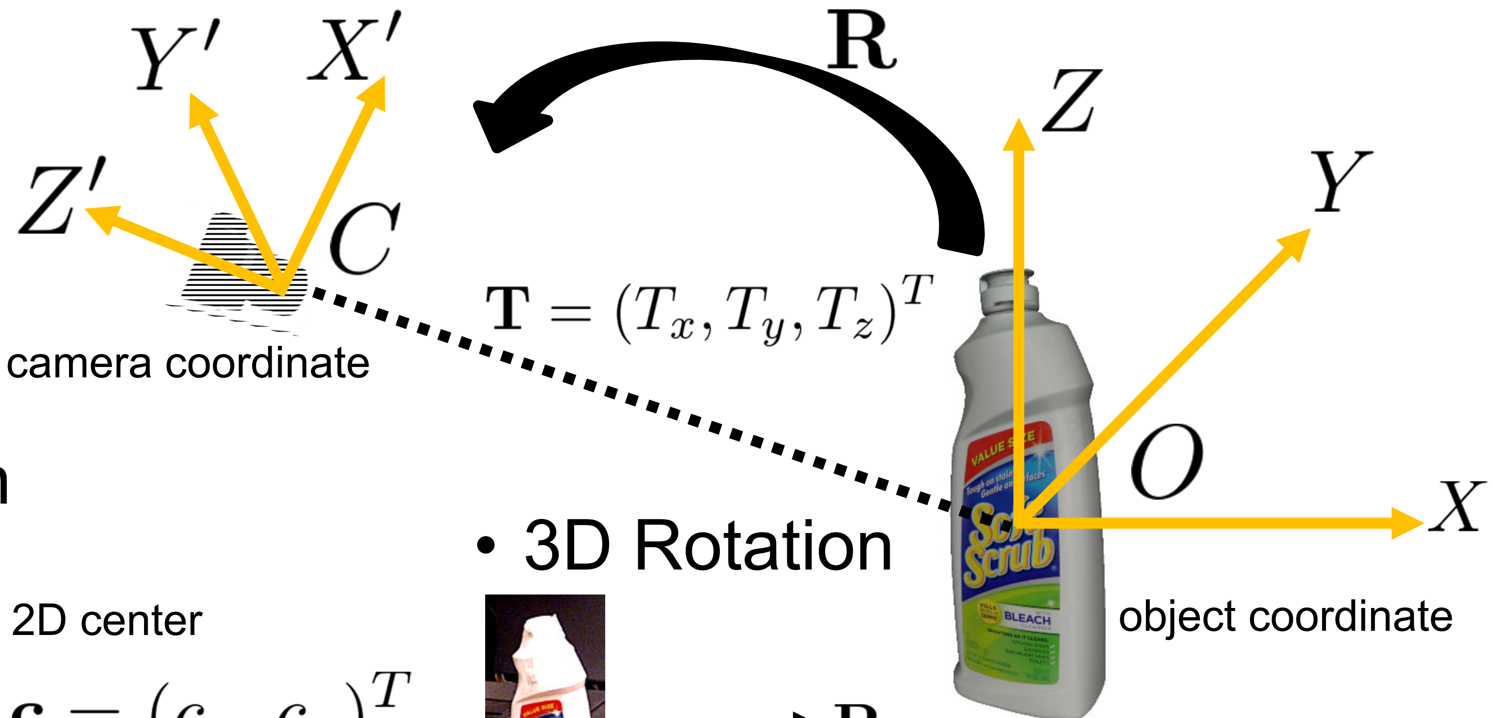
POSECNN

PoseCNN

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

Yu Xiang, Tanner Schmidt, Venkatraman Narayanan and Dieter Fox. PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes. In RSS, 2018.

POSECNN: DECOUPLE 3D TRANSLATION AND 3D ROTATION



- 3D Translation

2D center

$$\mathbf{c} = (c_x, c_y)^T$$

$$T_z$$

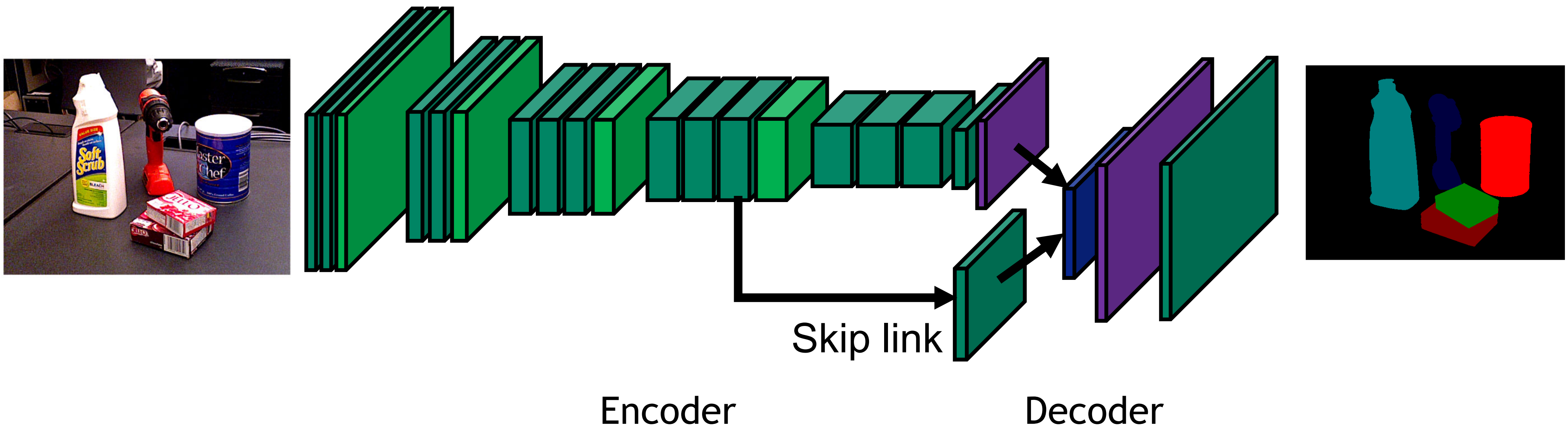
2D Center Localization

- 3D Rotation

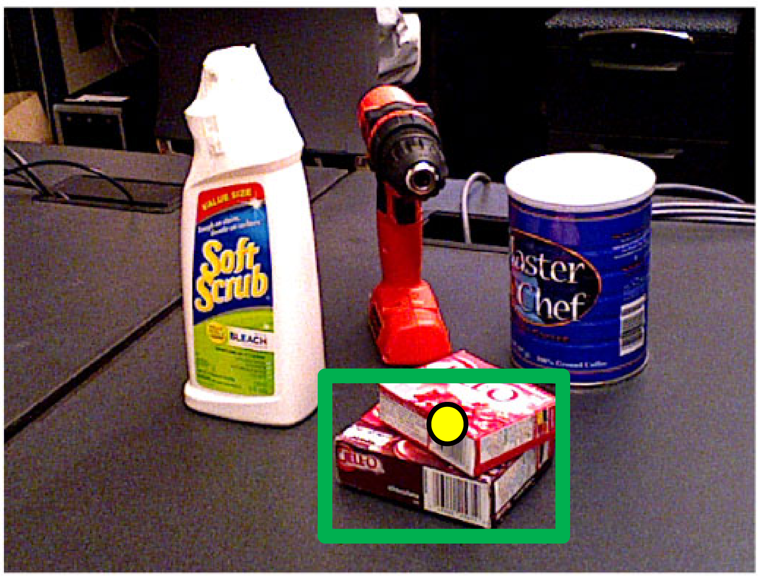
$\longrightarrow \mathbf{R}$

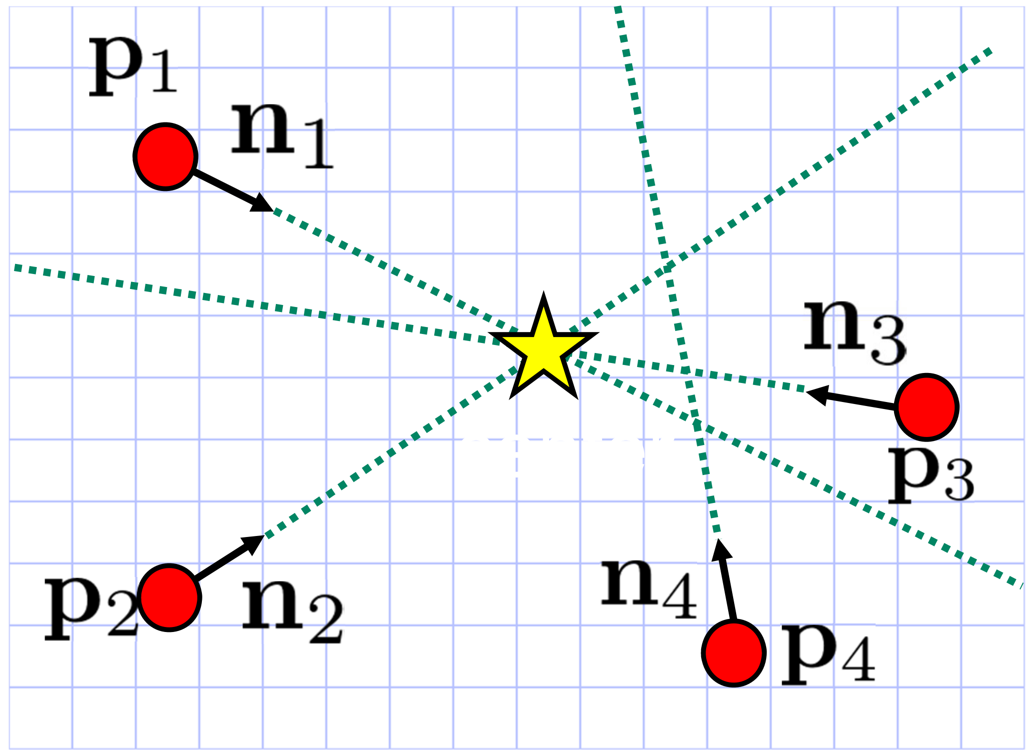
3D Rotation Regression

POSECNN: SEMANTIC LABELING

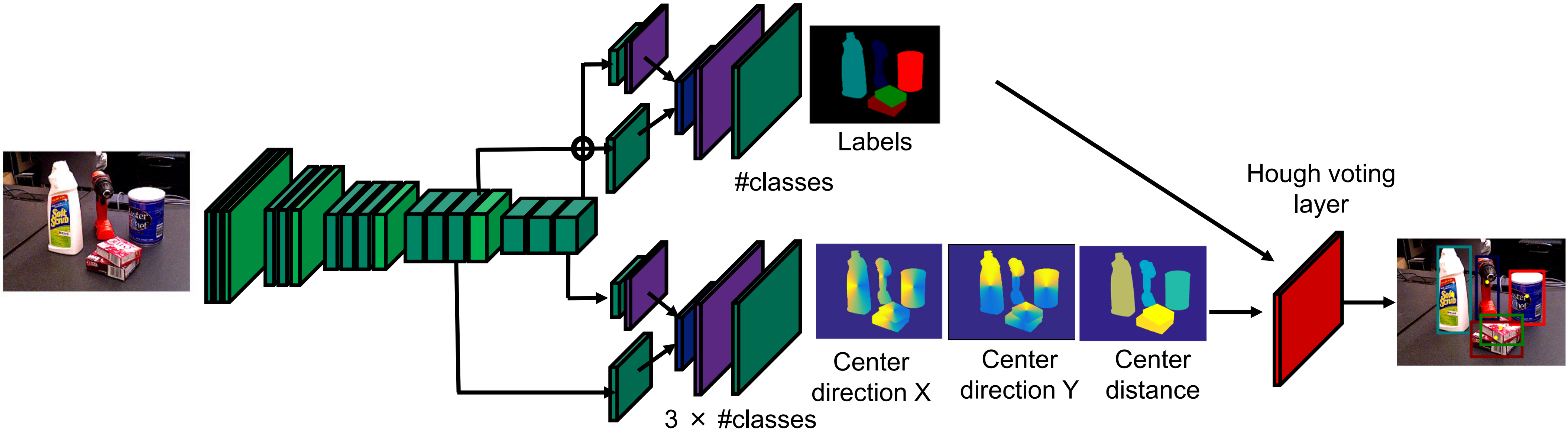


POSECNN: 2D CENTER VOTING FOR HANDLING OCCLUSIONS

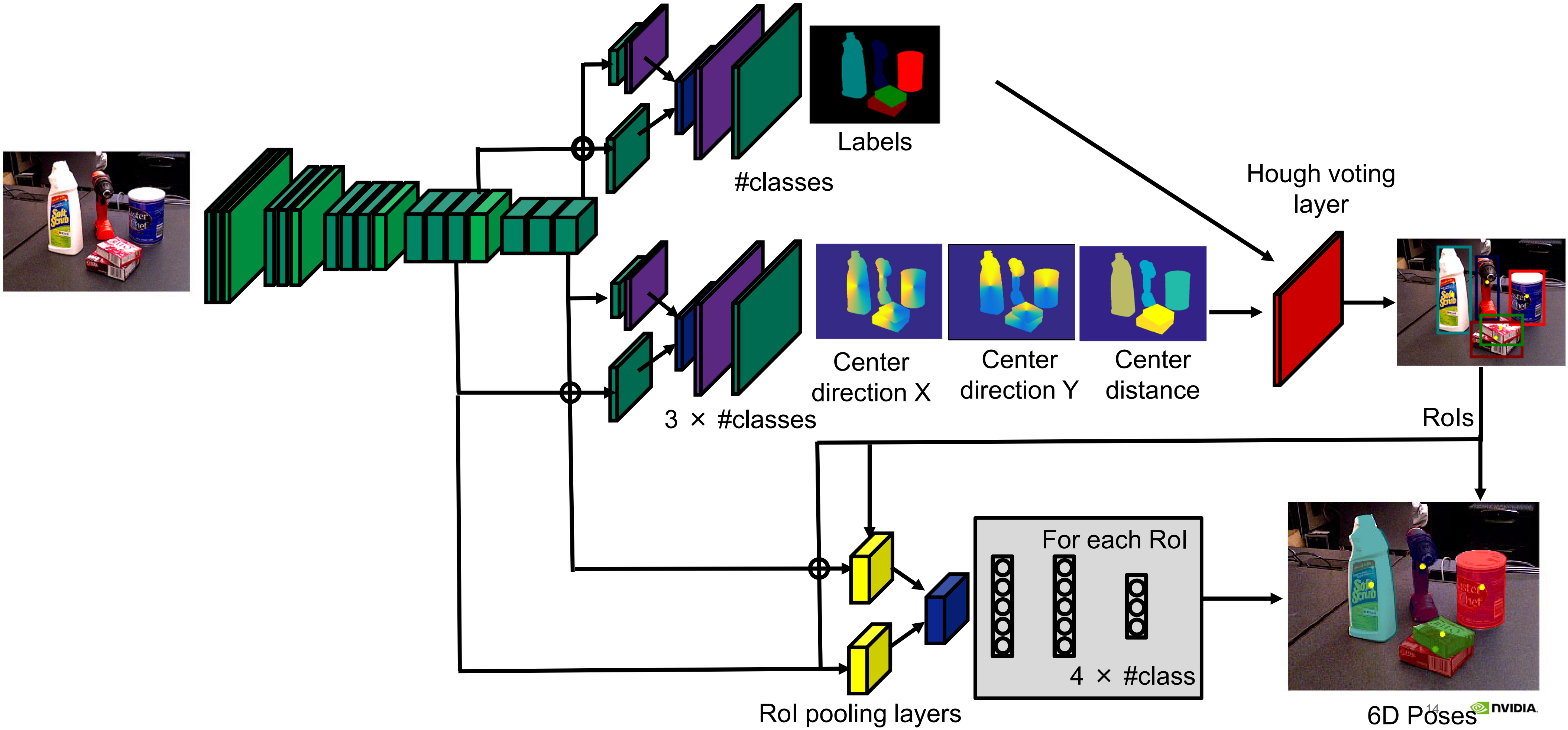




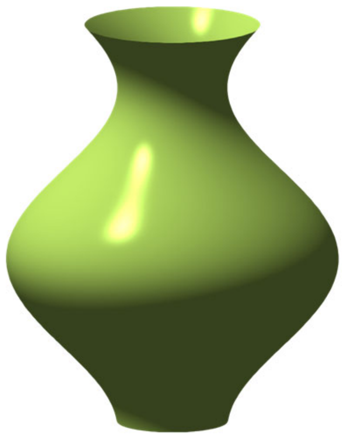
POSECNN: 3D TRANSLATION ESTIMATION



POSECNN: 3D ROTATION REGRESSION

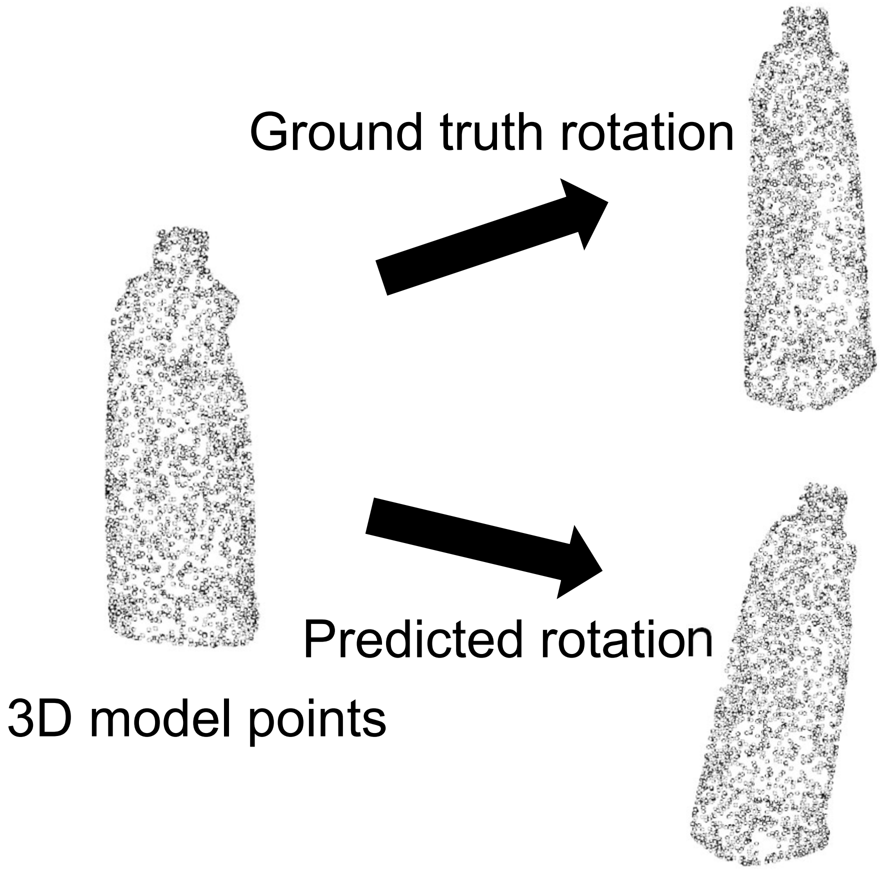


POSECNN: HANDLE SYMMETRIC OBJECTS






POSECNN: 3D ROTATION REGRESSION LOSS FUNCTIONS



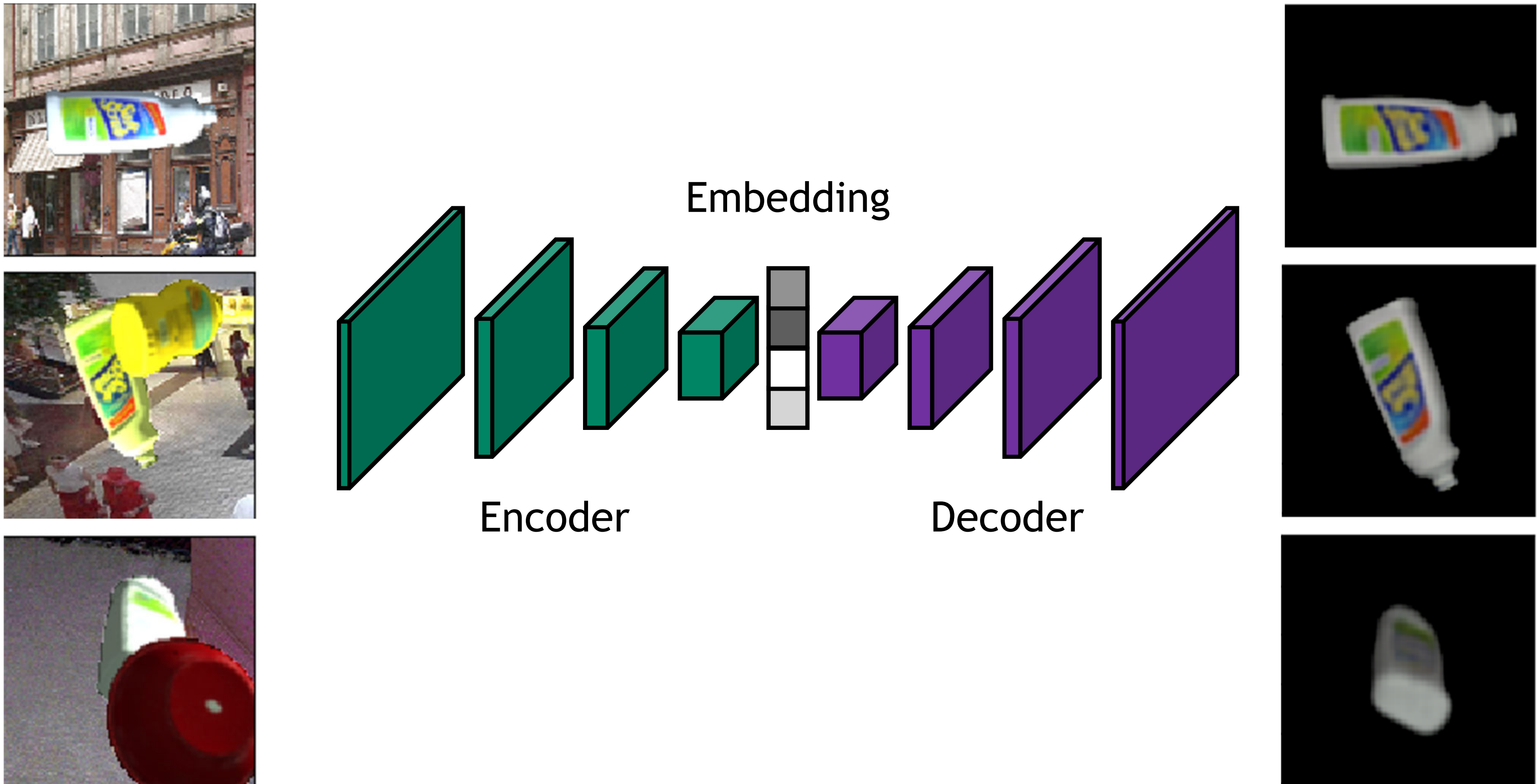
Pose Loss (non-symmetric)

$$\mathrm{PLoss}(\tilde{\mathbf{q}}, \mathbf{q}) = \frac{1}{2m} \sum_{\mathbf{x} \in \mathcal{M}} \|R(\tilde{\mathbf{q}})\mathbf{x} - R(\mathbf{q})\mathbf{x}\|^2$$

Shape-Match Loss for symmetric objects (symmetric)

$$\mathrm{SLoss}(\tilde{\mathbf{q}}, \mathbf{q}) = \frac{1}{2m} \sum_{\mathbf{x}_1 \in \mathcal{M}} \min_{\mathbf{x}_2 \in \mathcal{M}} \|R(\tilde{\mathbf{q}})\mathbf{x}_1 - R(\mathbf{q})\mathbf{x}_2\|^2$$

IMPLICIT ROTATION LEARNING

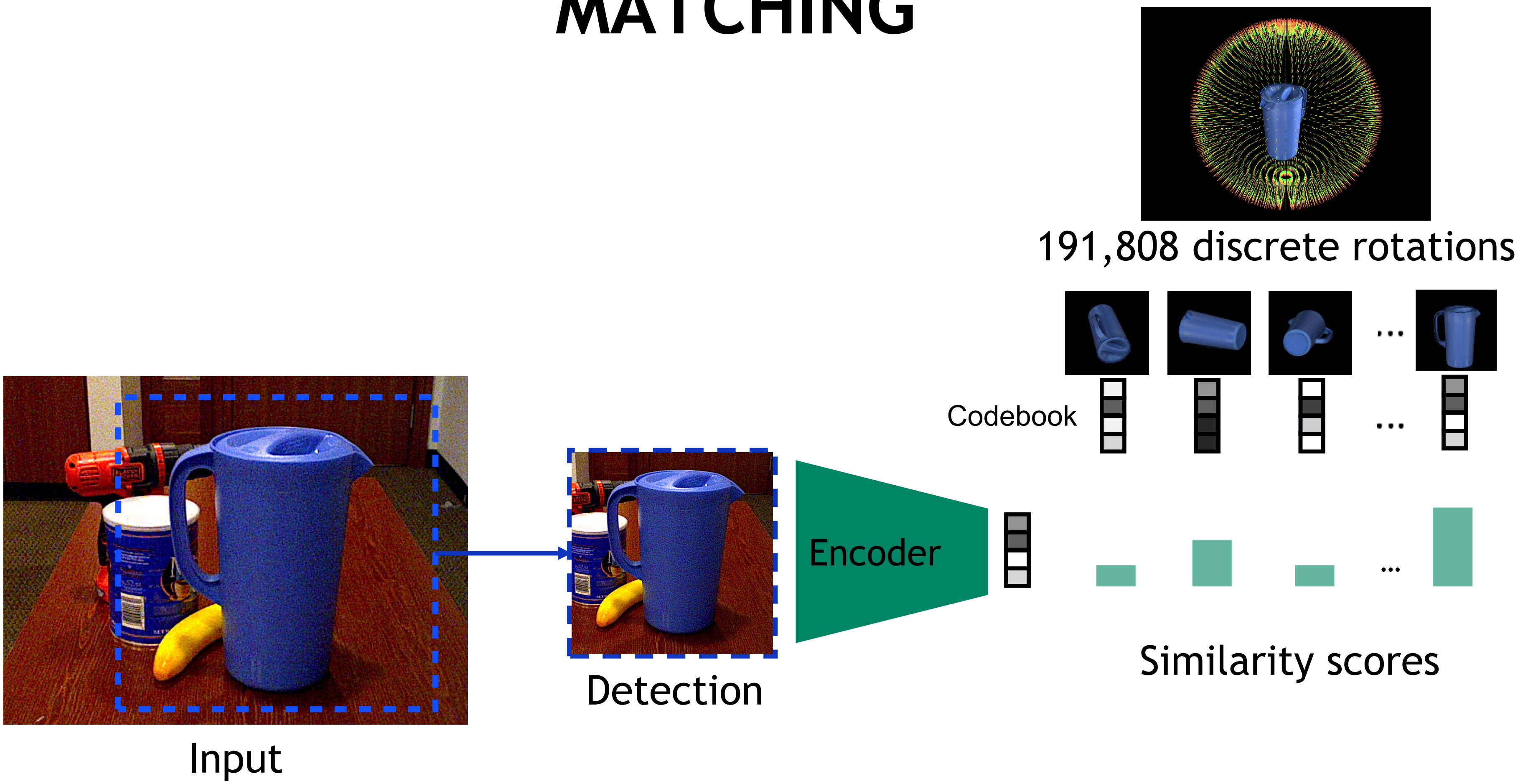


Input

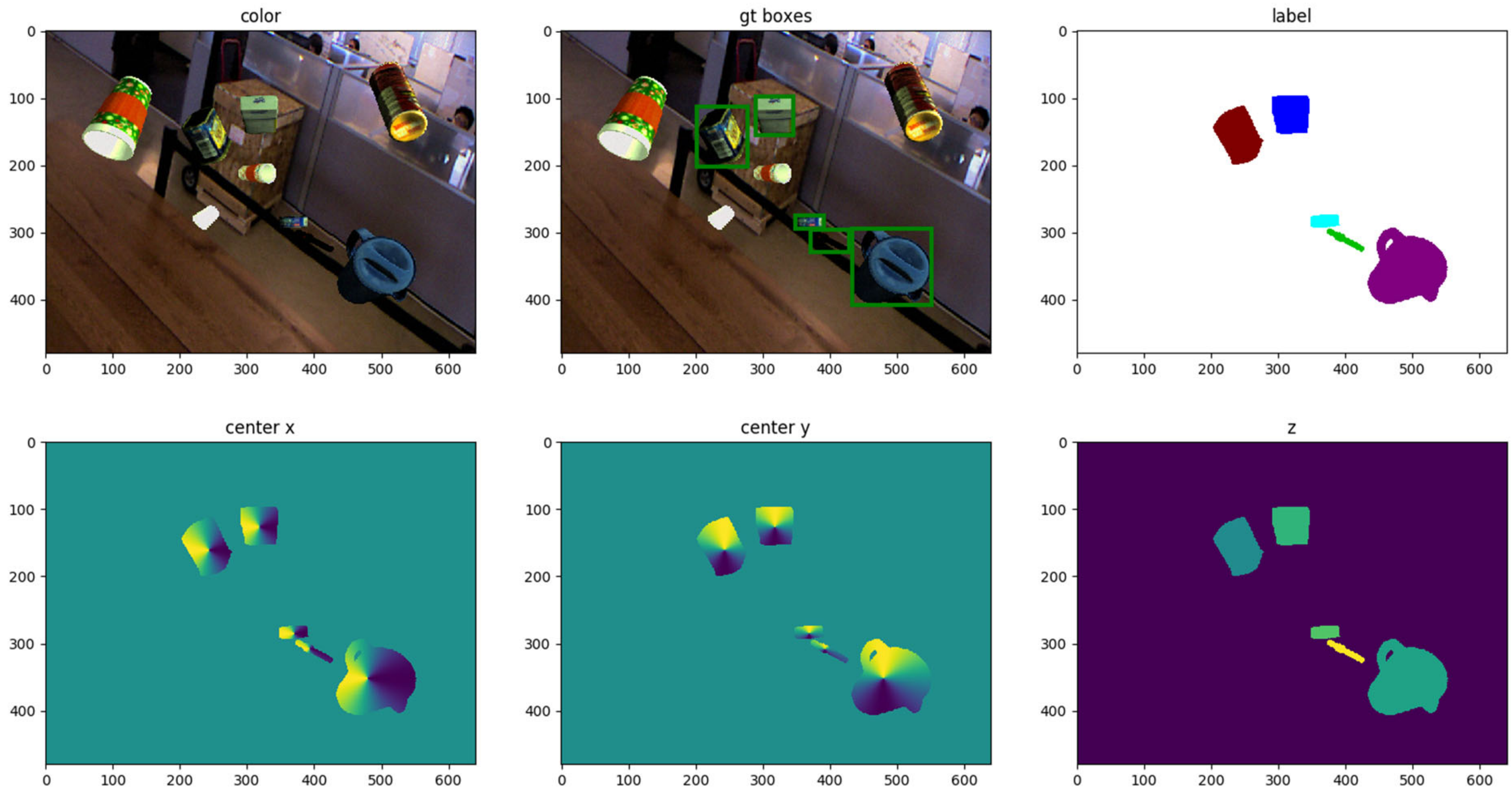
Reconstruction

Sundermeyer et al. Implicit 3D orientation learning for 6D object detection from RGB images. In ECCV, 2018.

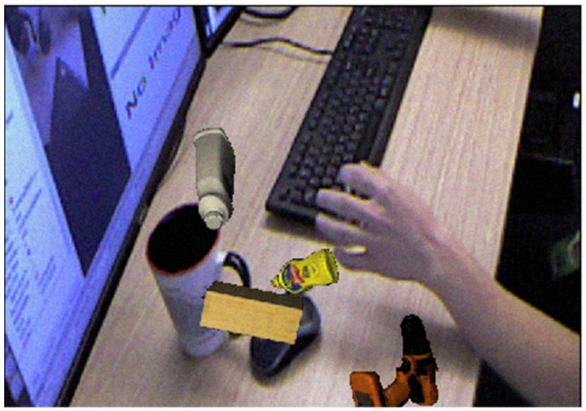
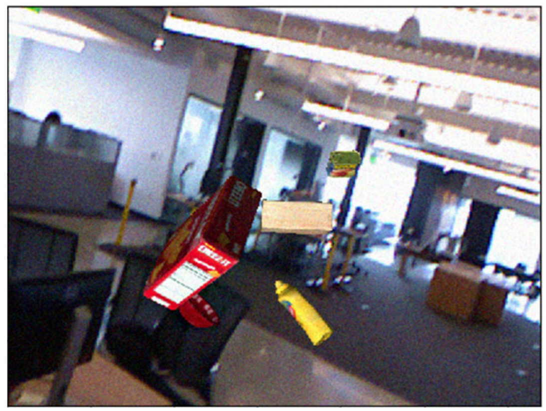
ROTATION ESTIMATION WITH CODEBOOK MATCHING

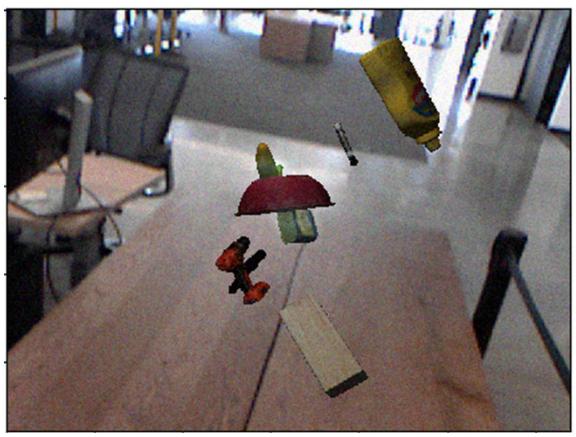


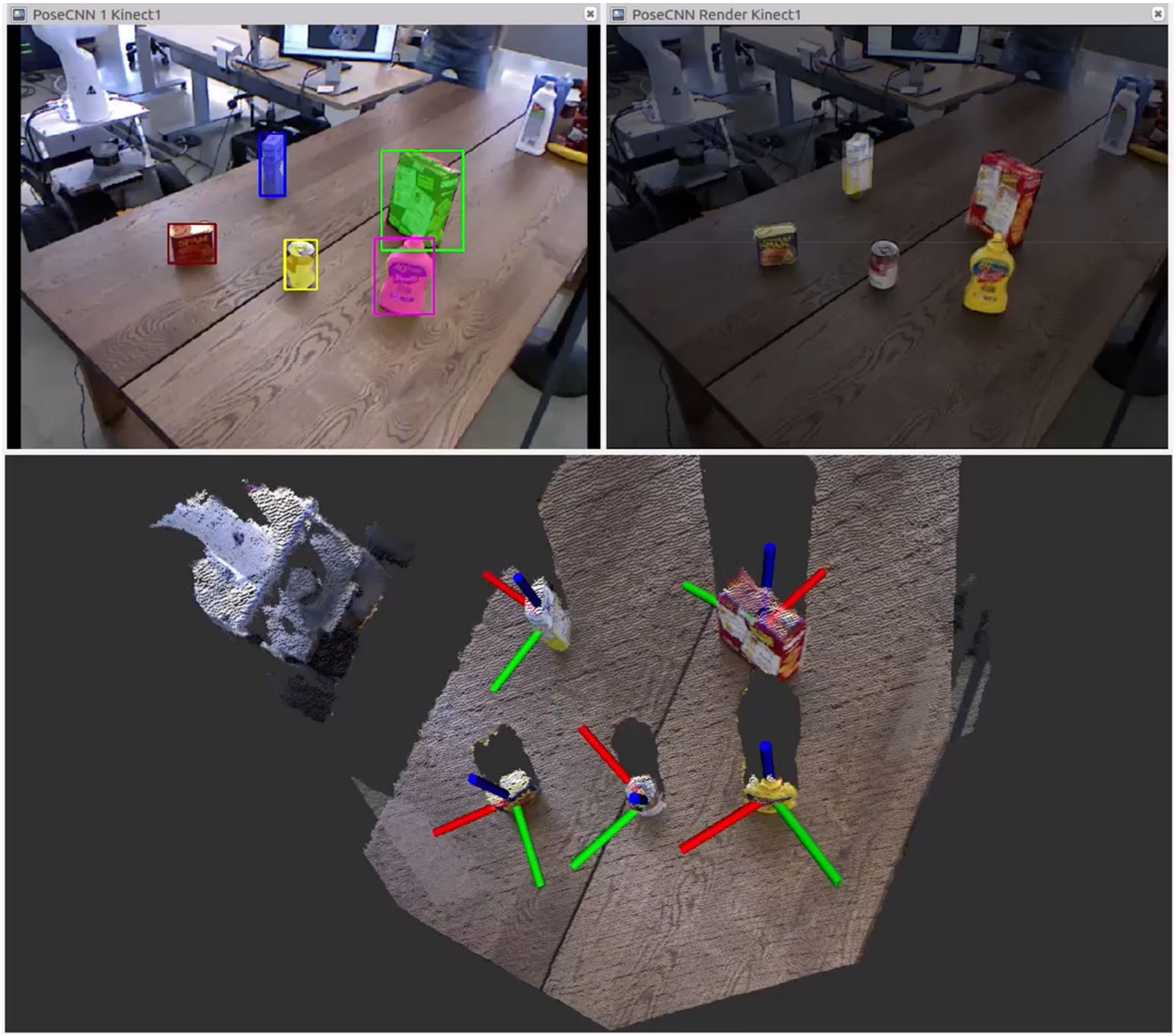
TRAINING DATA: DOMAIN RANDOMIZATION



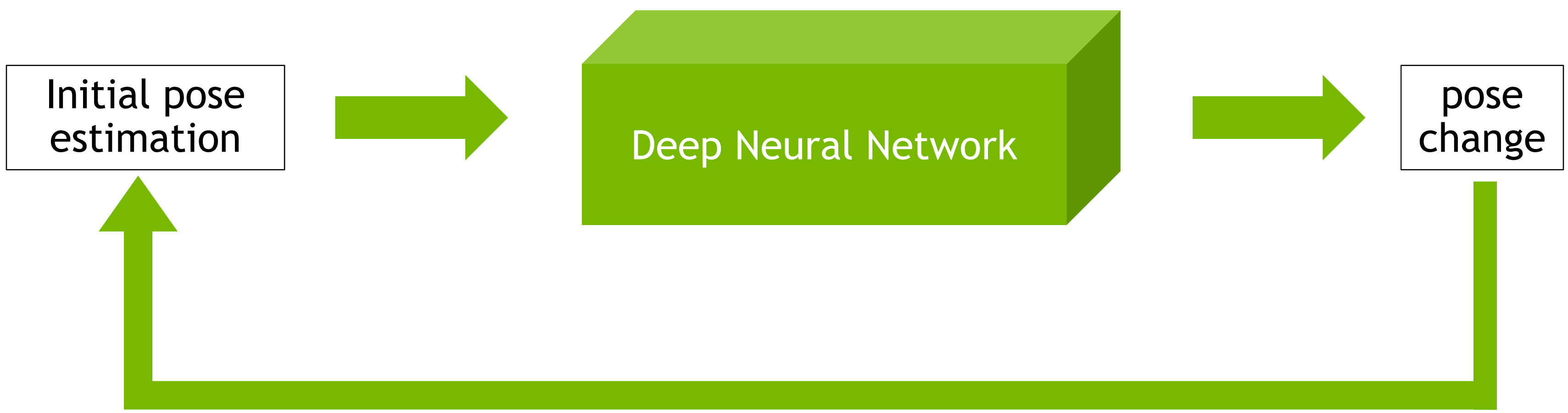
TRAINING DATA: DOMAIN RANDOMIZATION



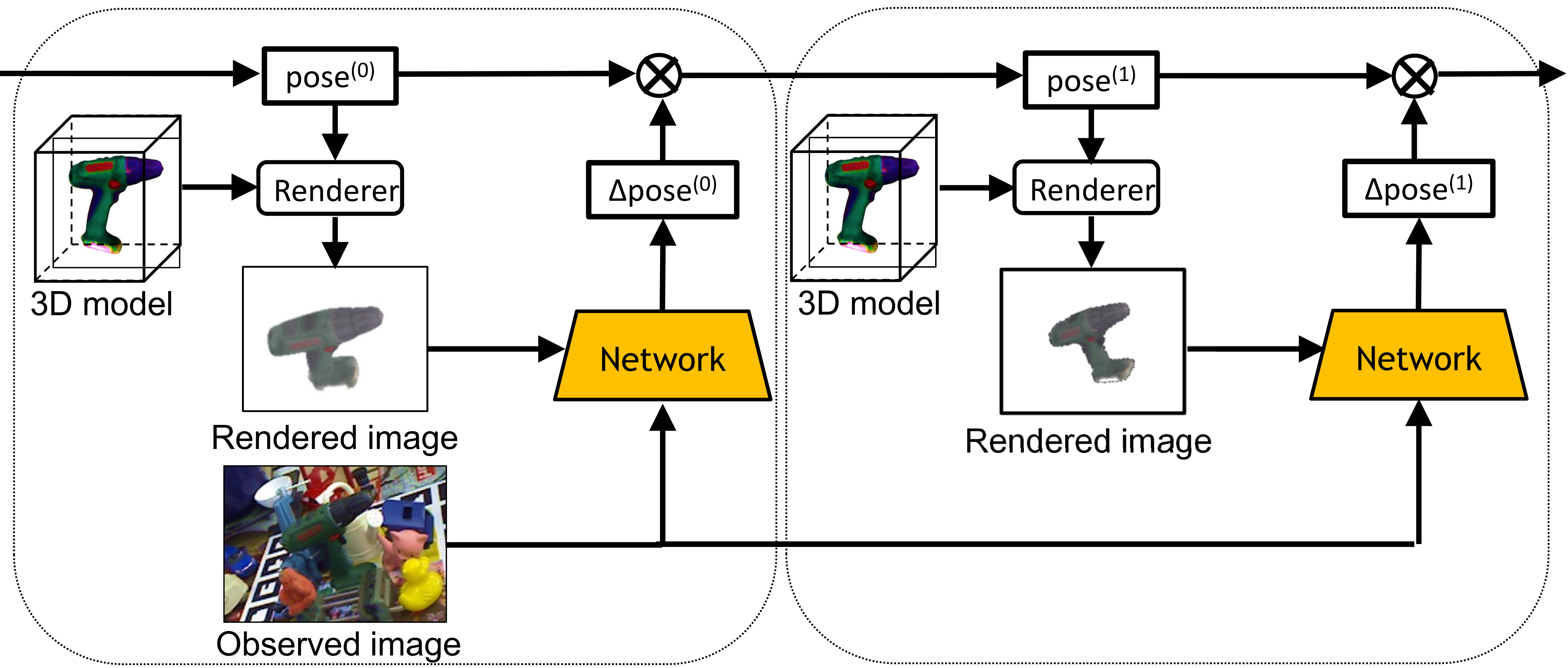

PoseCNN 1 Kinect1
PoseCNN Render Kinect1

DEEP ITERATIVE MATCHING FOR 6D OBJECT POSE ESTIMATION

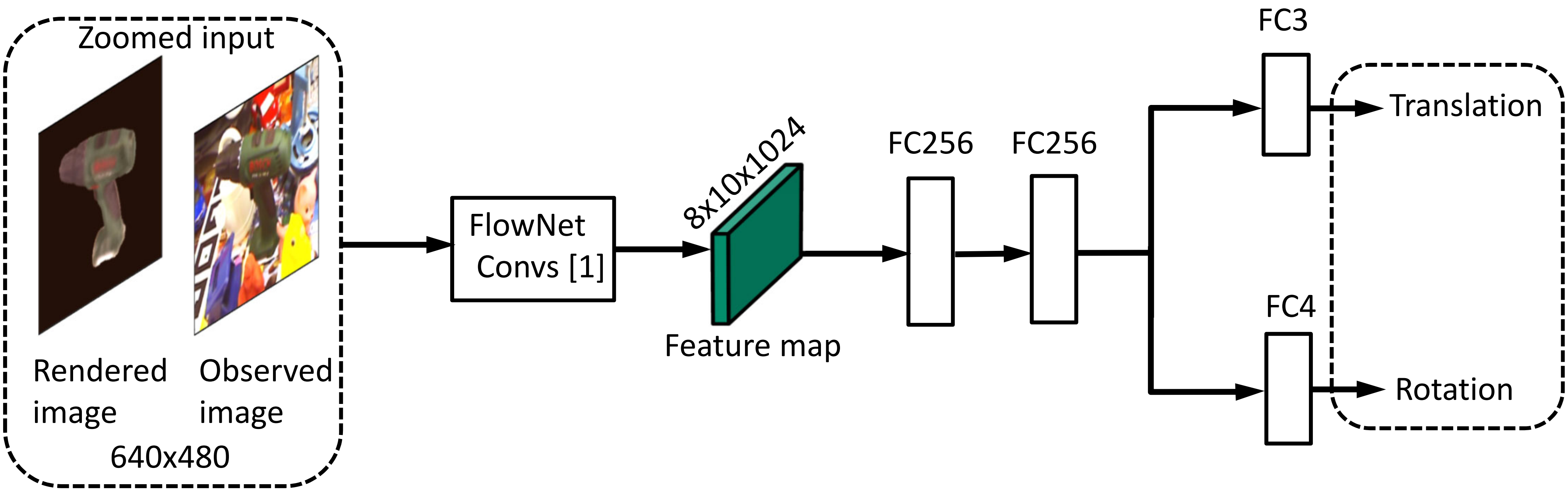


Yi Li*, Gu Wang, Xiangyang Ji, **Yu Xiang** and Dieter Fox. DeepIM: Deep Iterative Matching for 6D Pose Estimation. In ECCV, 2018 (Oral) (*PhD student at UW).

DEEPIM PIPELINE



NETWORK STRUCTURE



[1] Dosovitskiy, Alexey and Fischer, Philipp and Ilg, Eddy and Hausser, Philip and Hazirbas, Caner and Golkov, Vladimir and Van Der Smagt, Patrick and Cremers, Daniel and Brox, Thomas. Flownet: Learning optical flow with convolutional networks. In ICCV, 2015.

TRAINING DATA: YCB OBJECTS

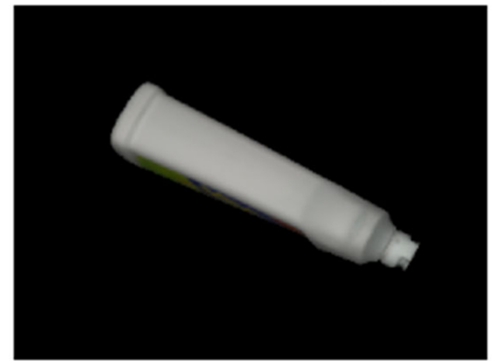
zoomed source image

zoomed target image

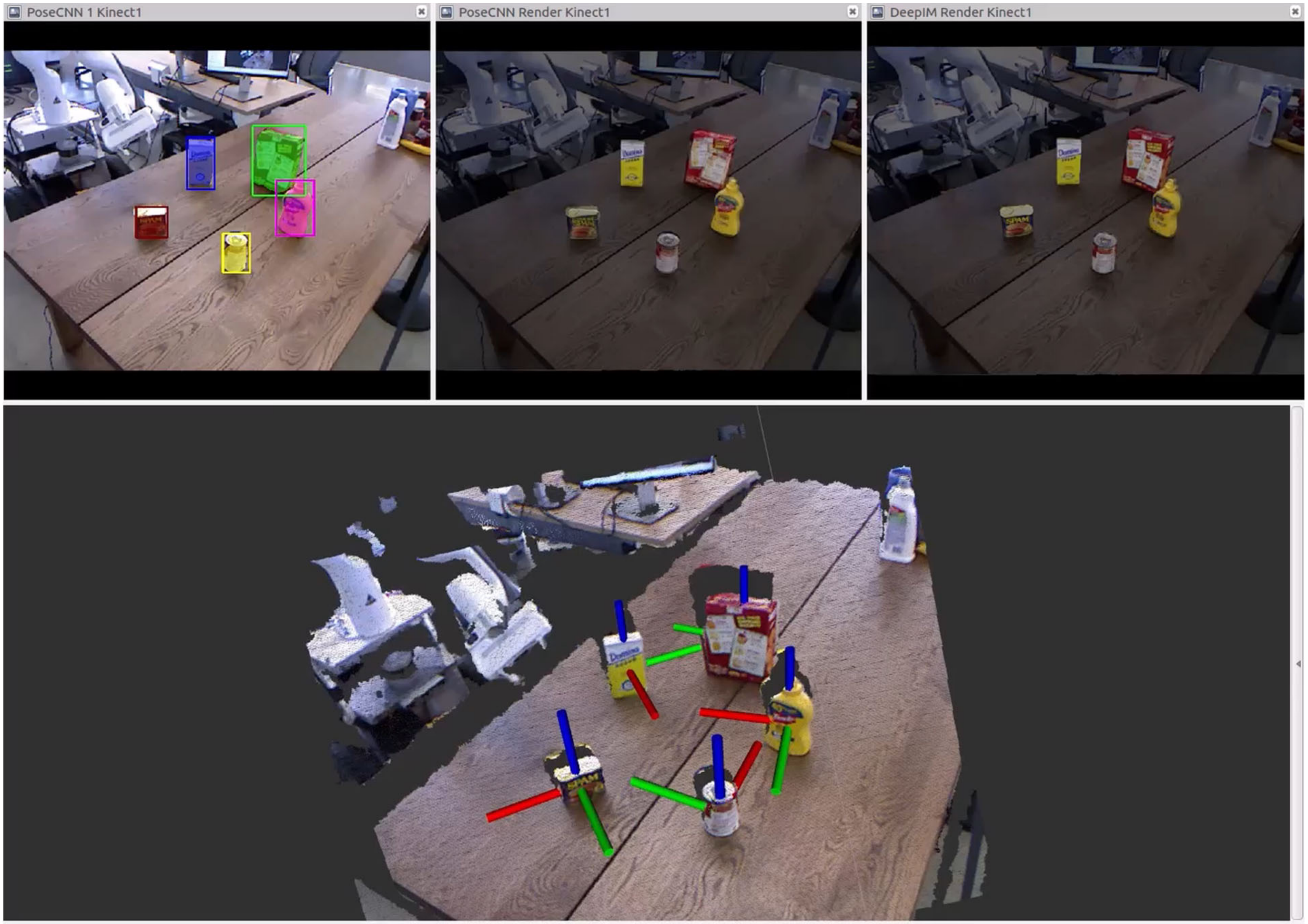
zoomed source image

zoomed target image

zoomed source image



zoomed target image


PoseCNN 1 Kinect1
PoseCNN Render Kinect1
DeepIM Render Kinect1

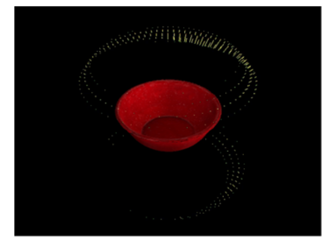
6D OBJECT POSE TRACKING

Input images

PoseRBPF

Translation

Orientation **distribution**

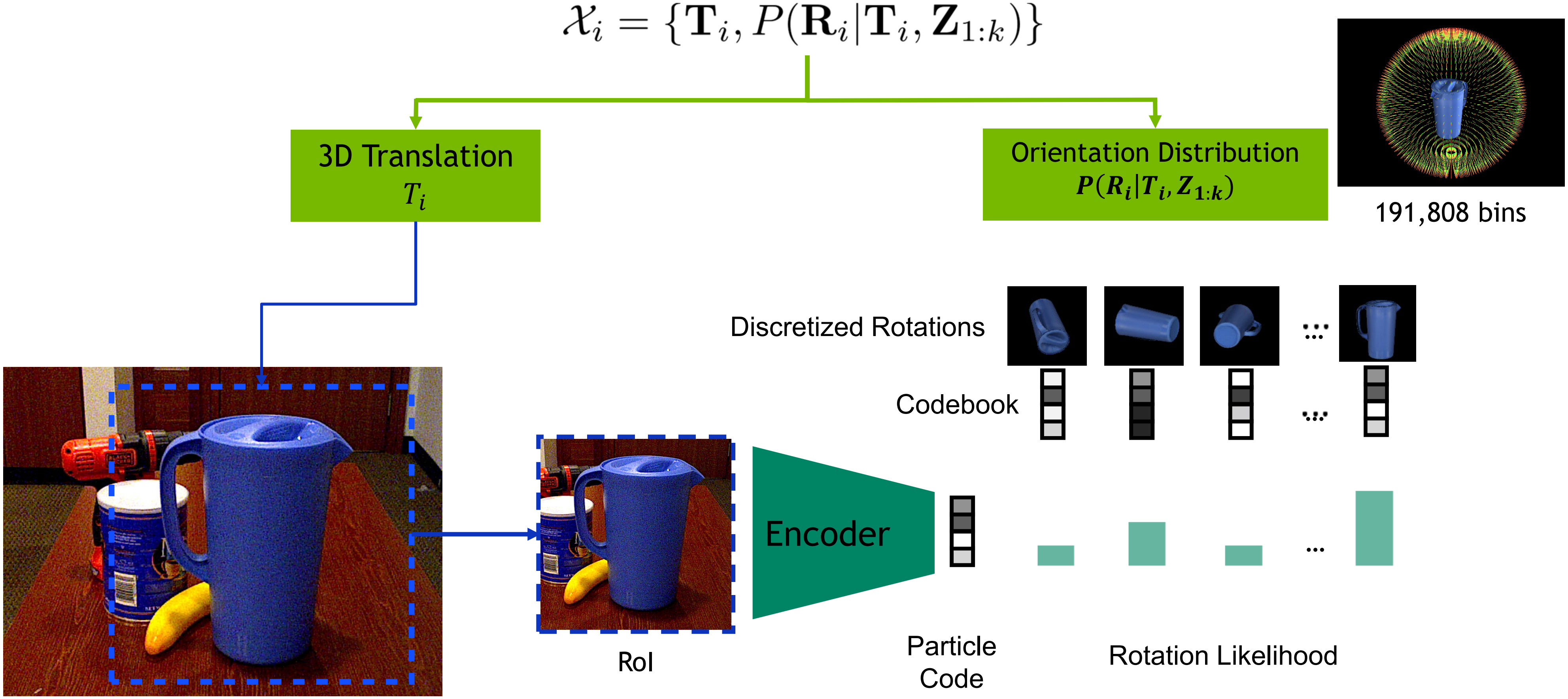


✓ Uncertainty in Pose Estimation

Xinke Deng*, Arsalan Mousavian, **Yu Xiang**, Fei Xia*, Timothy Bretl and Dieter Fox. PoseRBPF: A Rao-Blackwellized Particle Filter for 6D Object Pose Tracking. In RSS, 2019 (*intern at NVIDIA).

PoseRBPF: Particle Representation

$$\mathcal{X}_i = \{\mathbf{T}_i, P(\mathbf{R}_i | \mathbf{T}_i, \mathbf{Z}_{1:k})\}$$

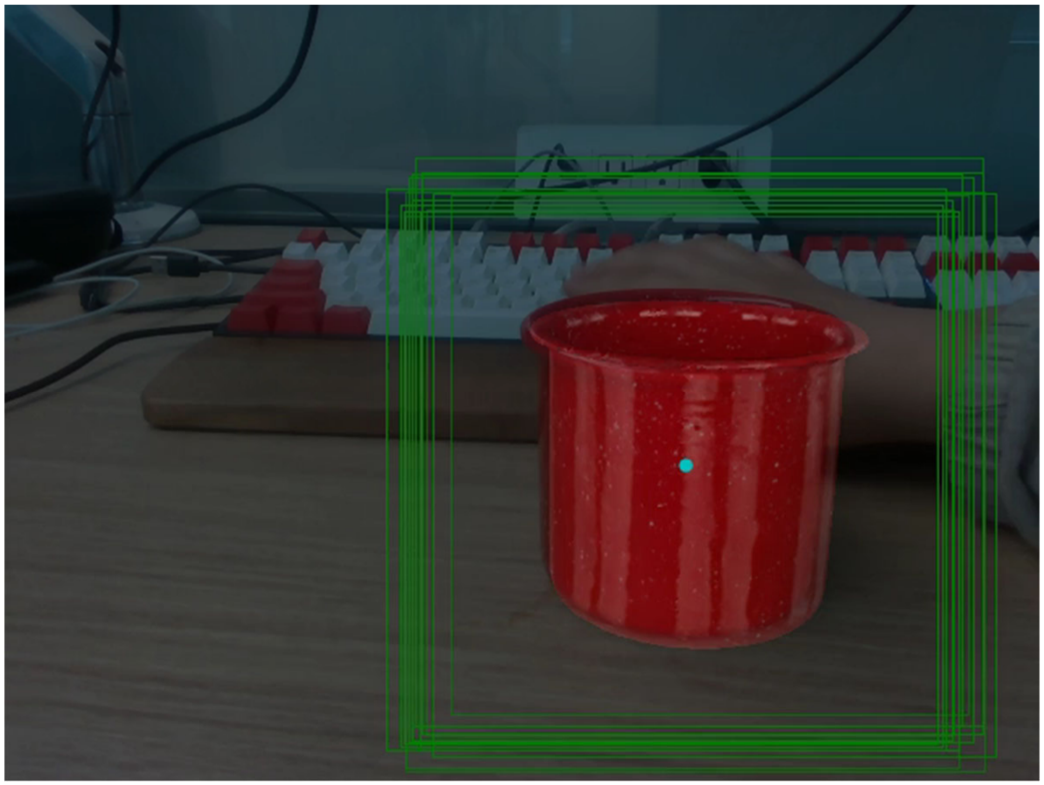


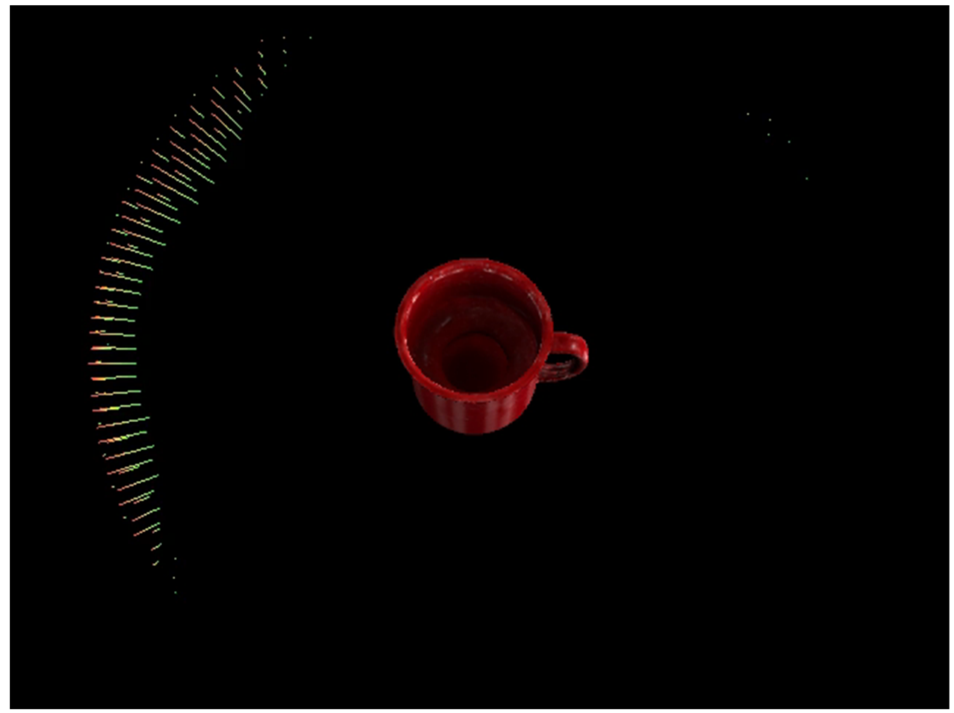
Results: YCB Objects

Example: YCB mug (50 particles, ~20fps)

YCB-Video RGB

- PoseRBPF:
  ADD: 62.1, ADD-S: 78.4
- PoseCNN:
  ADD: 53.7, ADD-S: 75.9





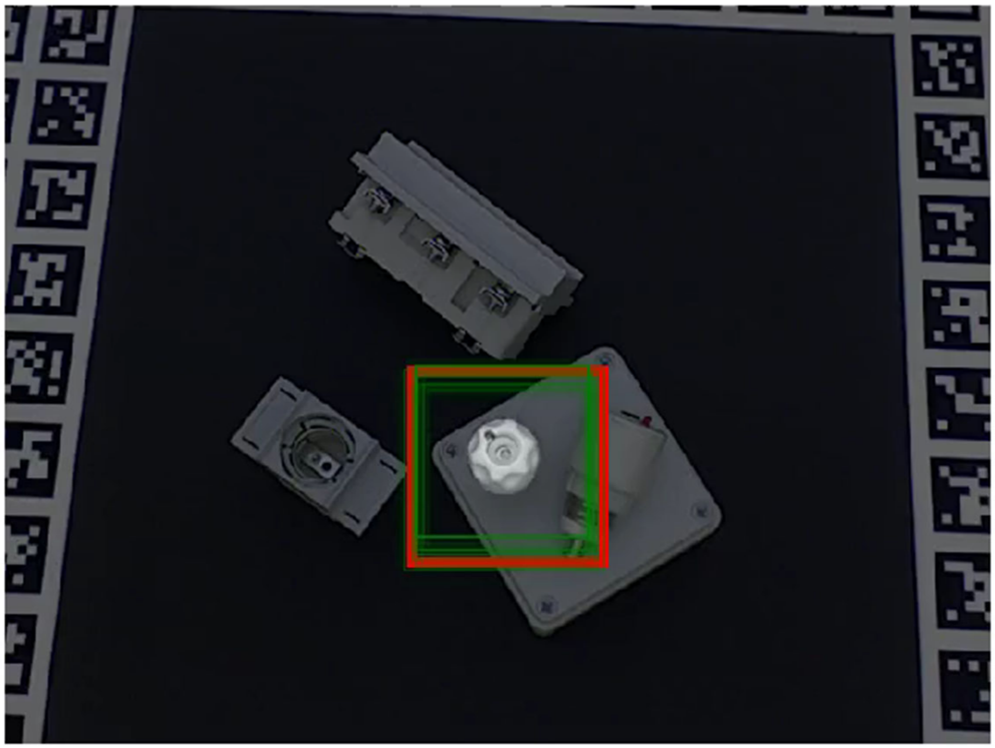
Results: TLess Objects

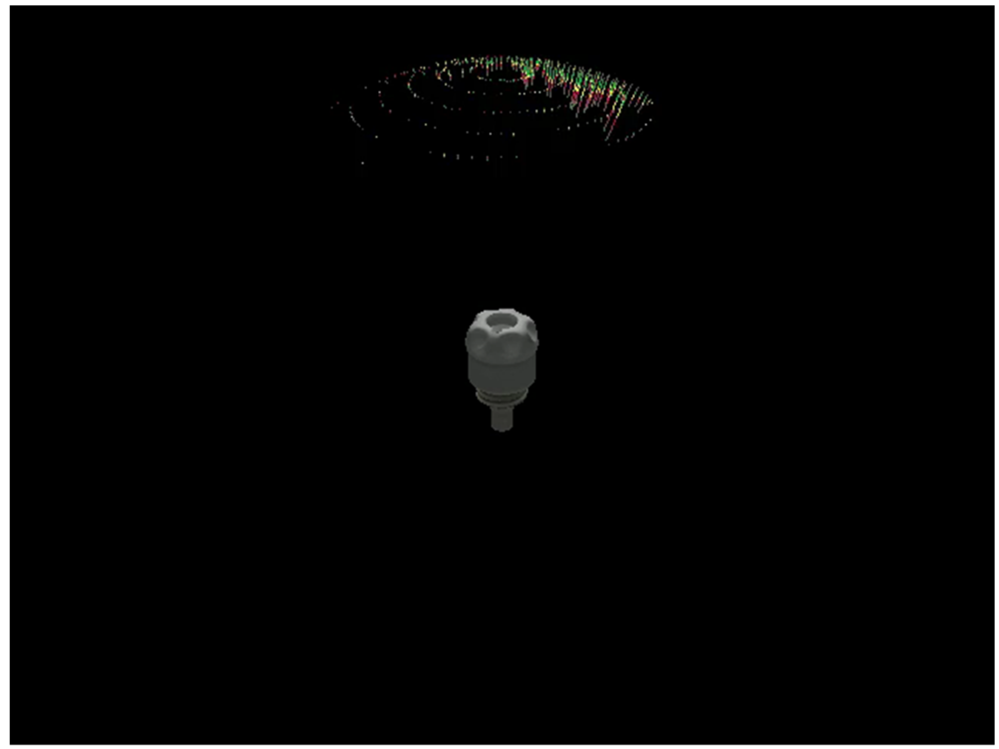
Example: TLess 01 (100 particles, ~11fps)

TLess RGB

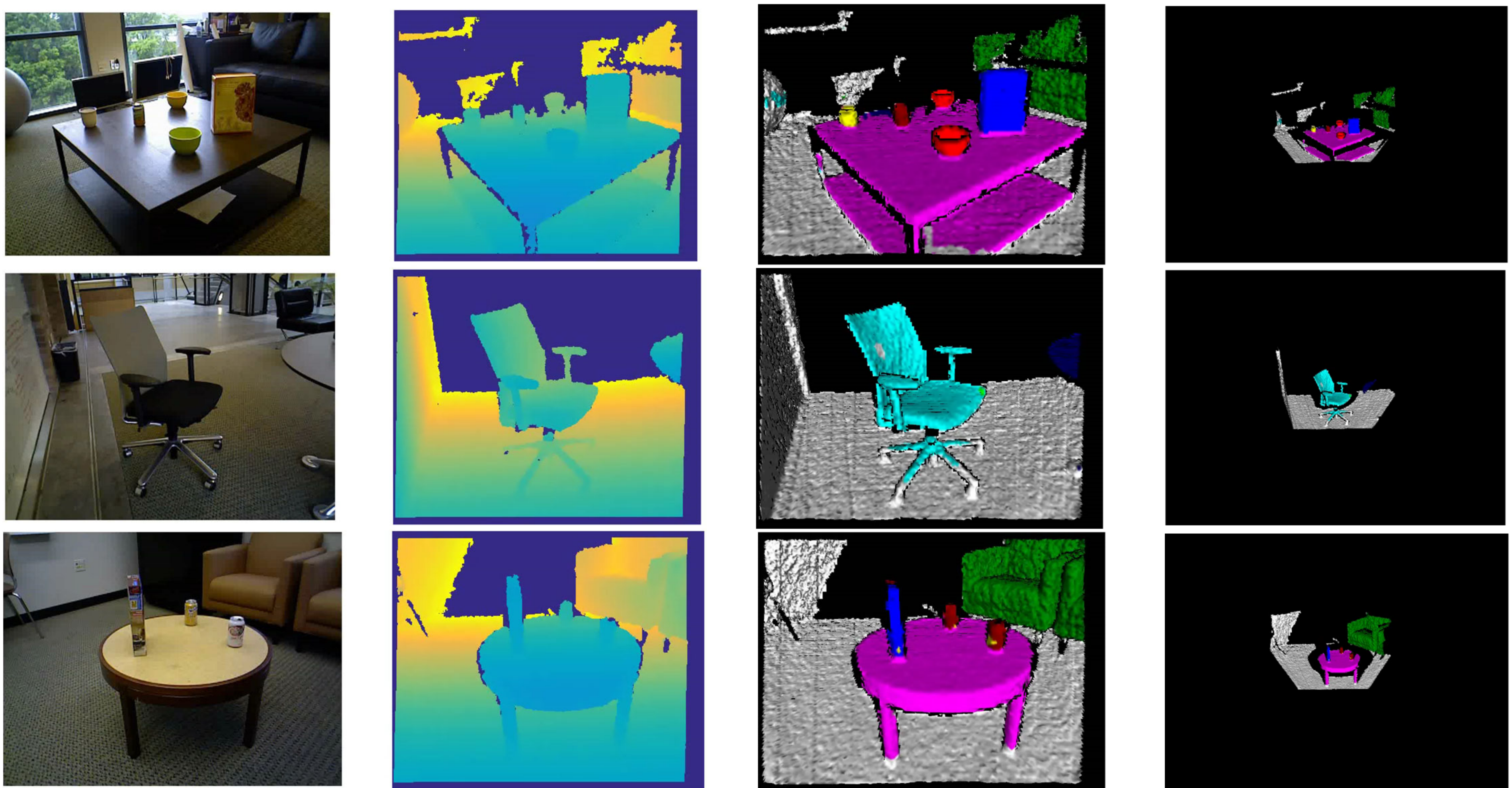
Object recall for Err_vsd < 0.3:

- PoseRBPF: 41.47%
- Sundermeyer et al: 18.35%





SEMANTIC MAPPING



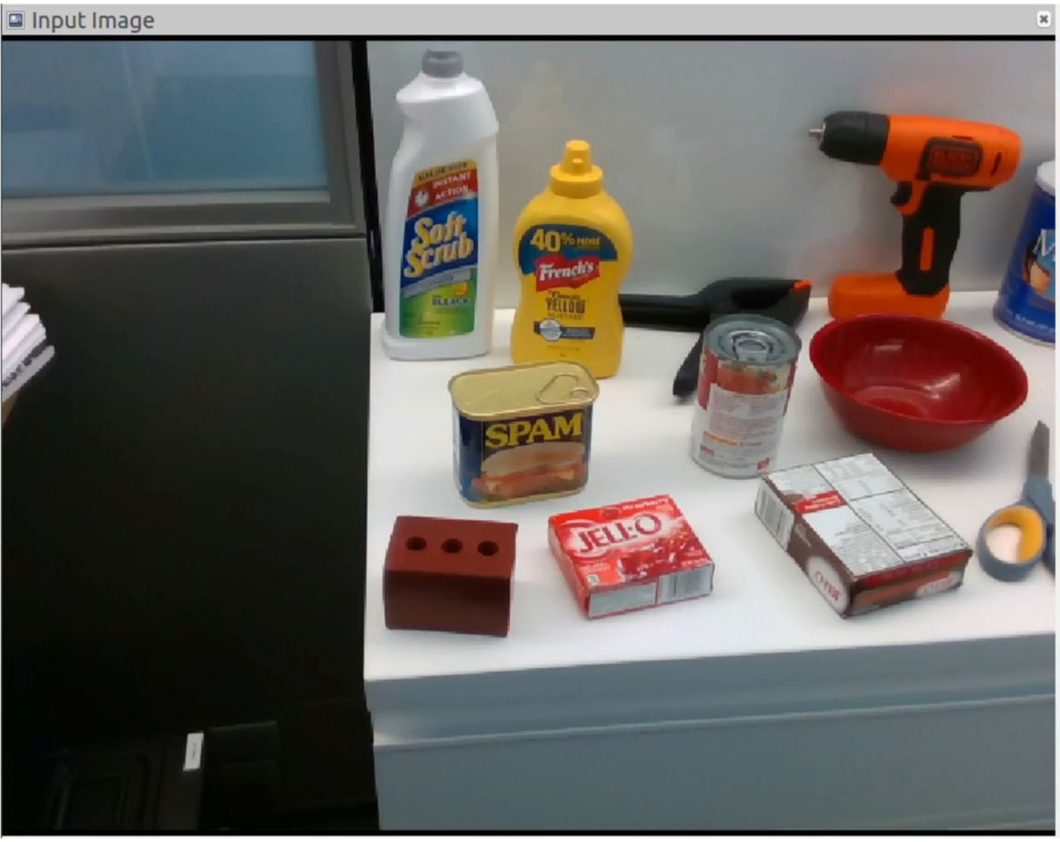
Yu Xiang and Dieter Fox. DA-RNN Semantic Mapping with Data Associated Recurrent Neural Network, RSS, 2017.

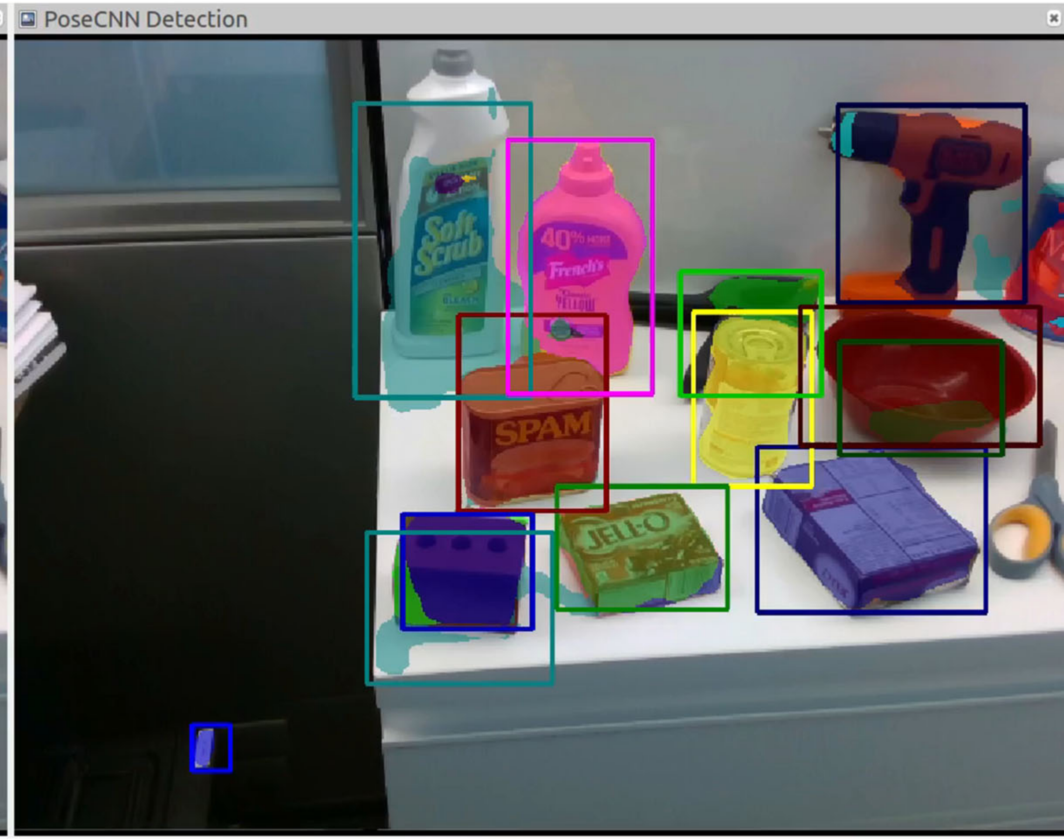
UNSEEN OBJECT INSTANCE SEGMENTATION

Failures

Christopher Xie*, **Yu Xiang**, Arsalan Mousavian and Dieter Fox. The Best of Both Modes: Separately Leveraging RGB and Depth for Unseen Object Instance Segmentation. Under Review, 2019 (*PhD student at UW).

POSECNN FOR 20 YCB OBJECTS





FUTURE WORK: SELF-SUPERVISED LEARNING

Simulation

Synthetic data

Leaning

model

Apply

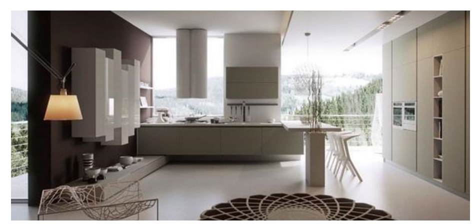
Interacting

Training data in the real world

Updating

- Reason about uncertainty/failure in the real world to obtain annotations
- Interact with the real world to collect more data

Real environment



New environment? Learning in that environment to adapt the model!

Questions?