PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation

Yu Xiang[1], Tanner Schmidt[2], Venkatraman Narayanan[3] and Dieter Fox[1,2]

[1]NVIDIA Research, [2]University of Washington, [3]Carnegie Mellon University

6D Object Pose Estimation for Robotic Manipulation
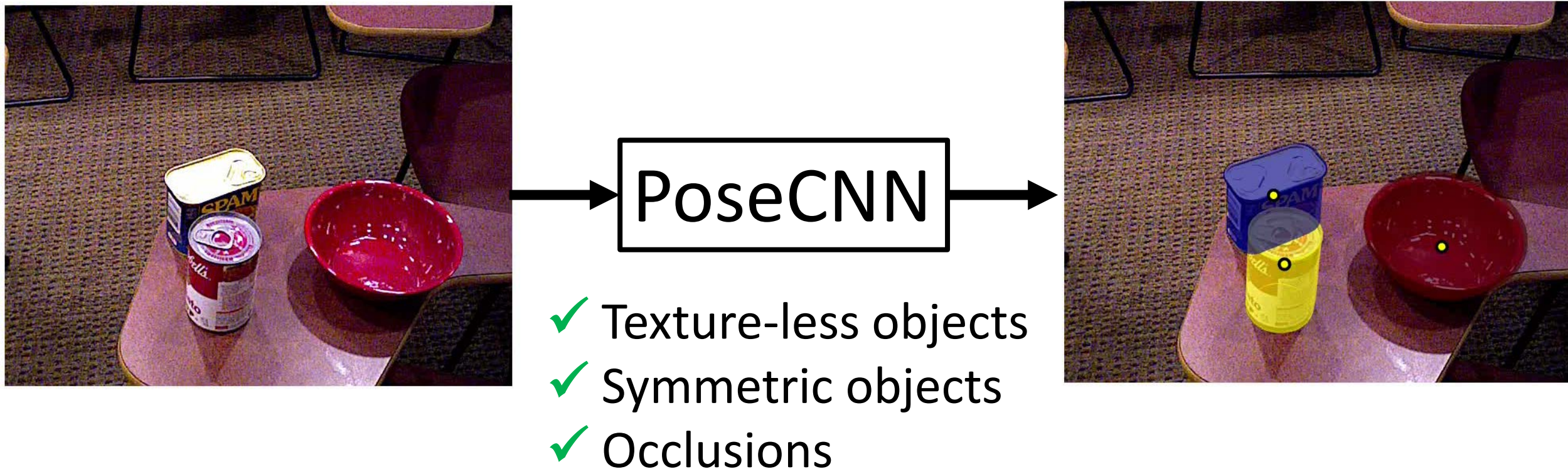
Camera

- **3D location**
- **3D orientation**

Input image

Challenge:
- **Texture-less objects**
- **Symmetric objects**
- **Occlusions**

Known 3D model

Our Contribution: A Generic Convolutional Neural Network for 6D Object Pose Estimation

PoseCNN

✓ Texture-less objects
✓ Symmetric objects
✓ Occlusions

PoseCNN: Decouple 3D Translation and 3D Rotation

- 3D Translation $\mathbf{T} = (T_x, T_y, T_z)^T$

2D center

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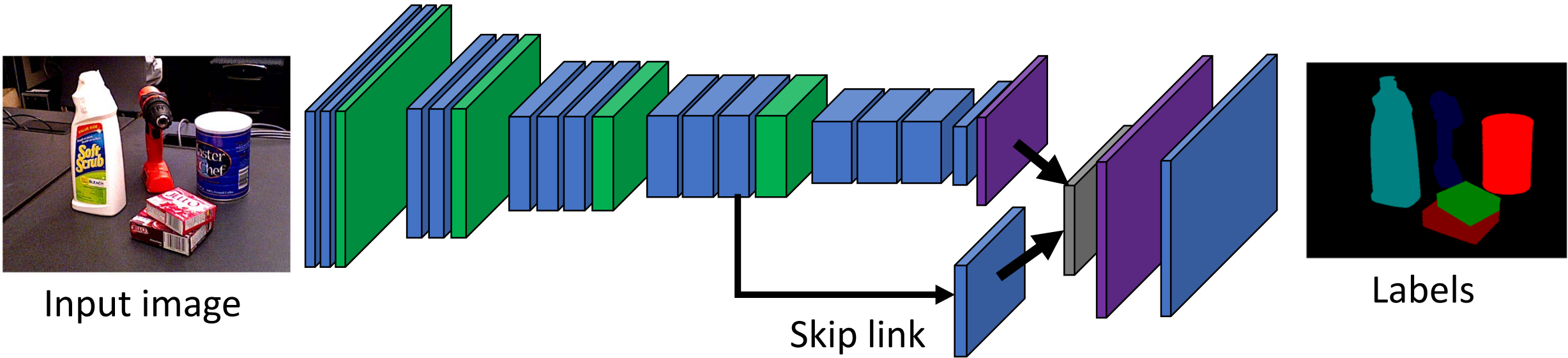
Distance T_z

2D Center Localization

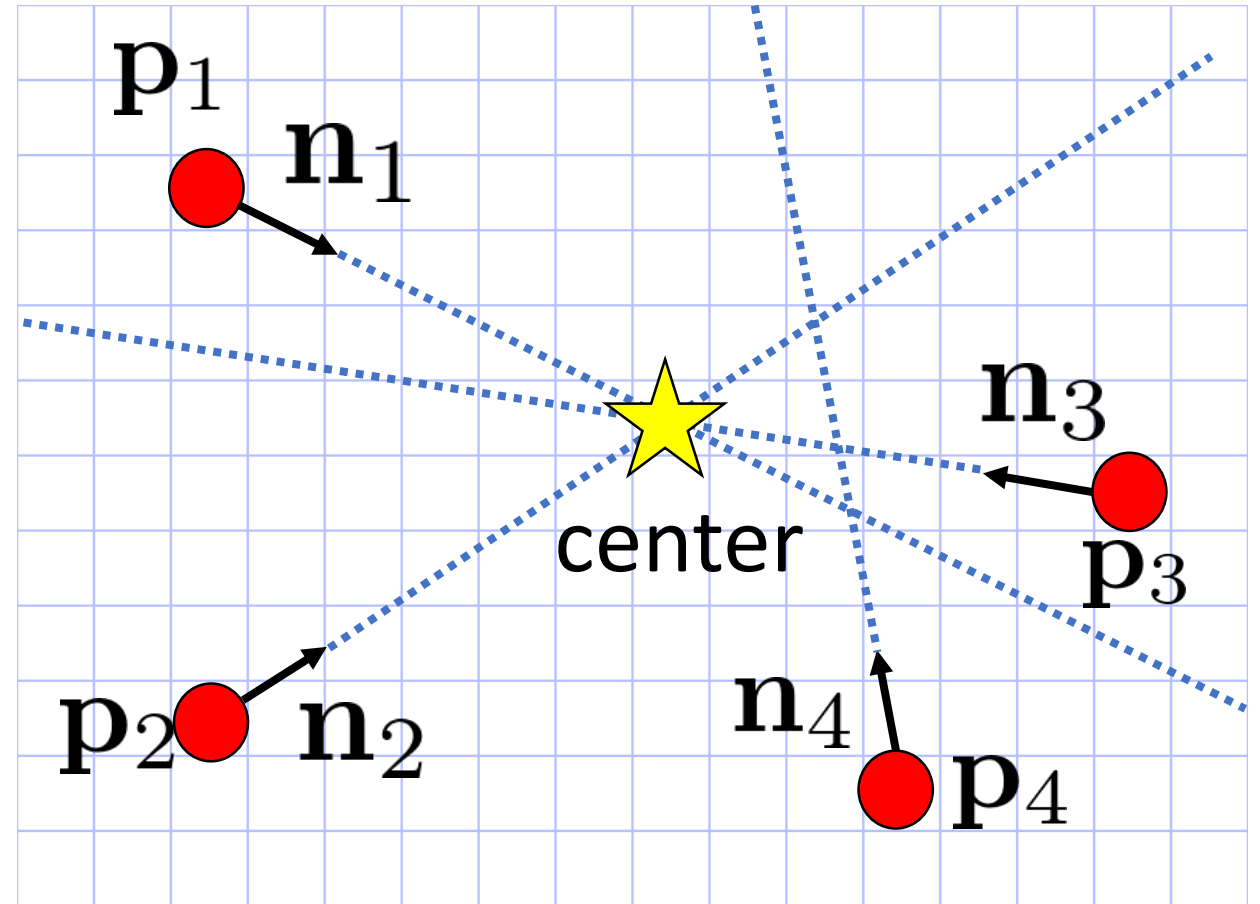- 3D Rotation $\mathbf{R}$

$$\mathbf{R}$$
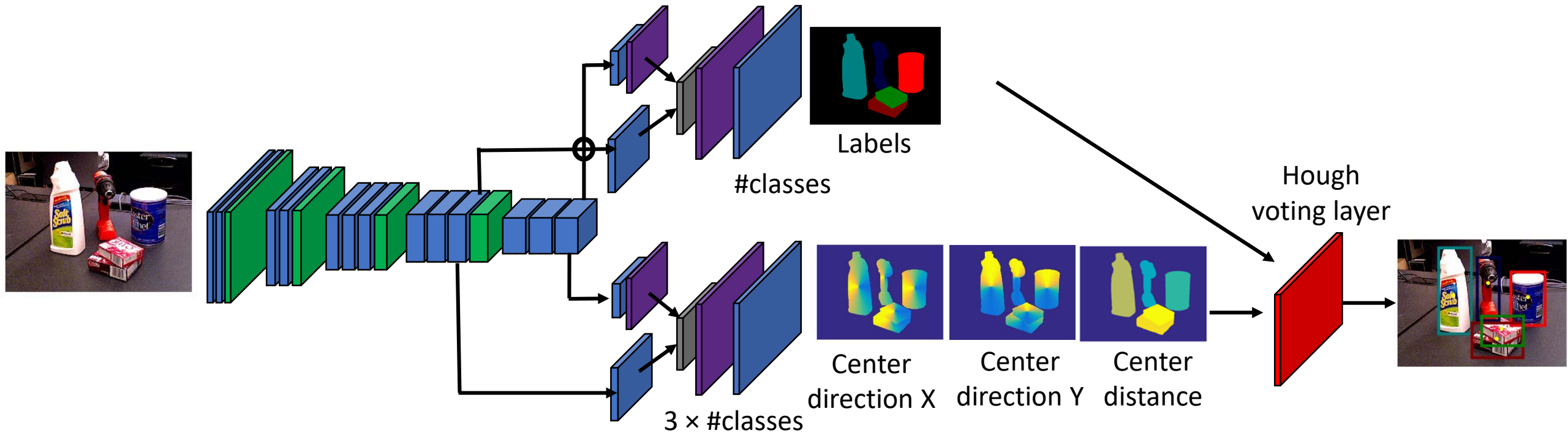
3D Rotation Regression

PoseCNN: Semantic Labeling

Input image

Skip link

Fully convolutional network

Labels

- Long et al., CVPR, 2015
- Xiang & Fox, RSS, 2017
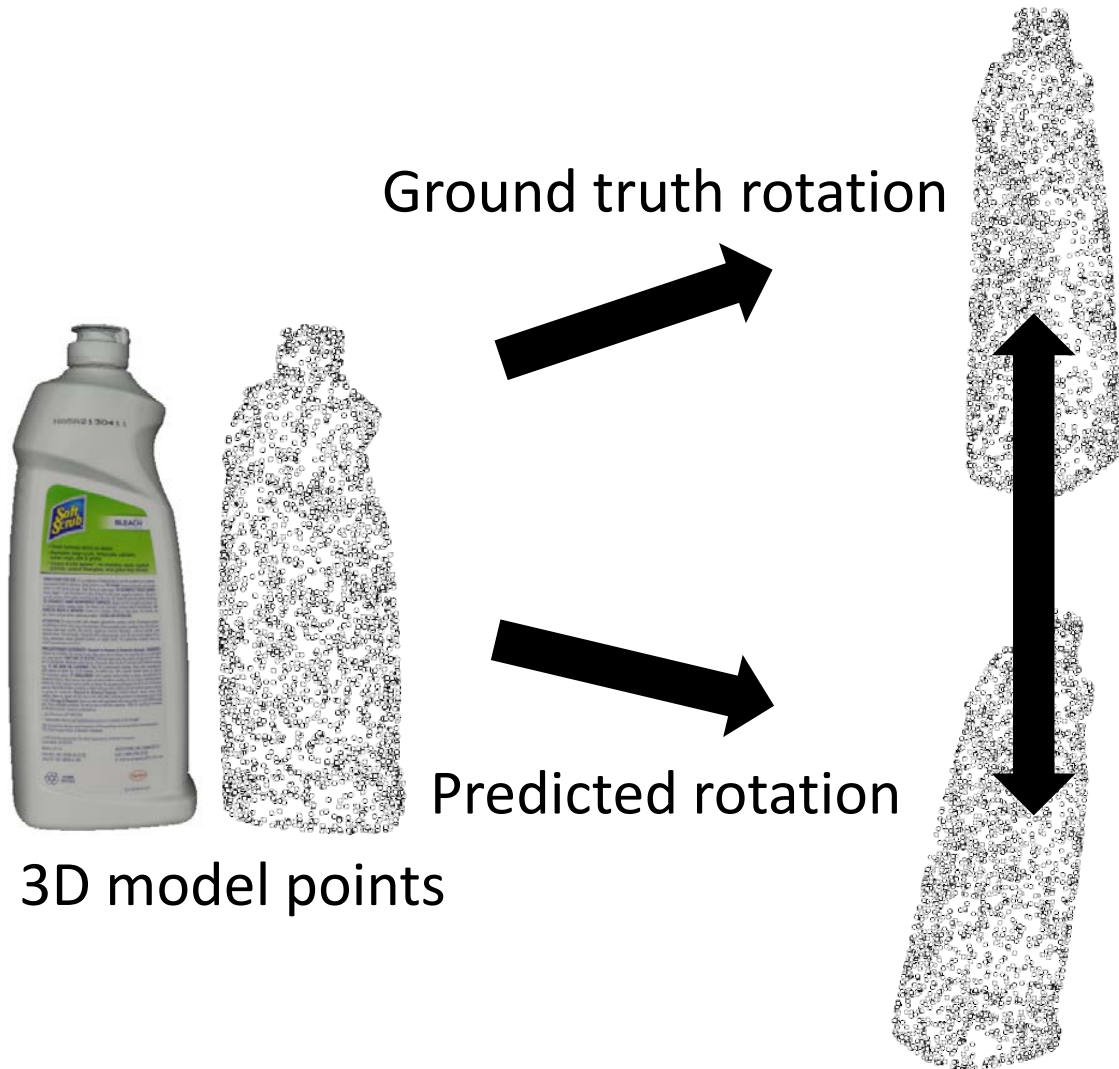
PoseCNN: 2D Center Voting for Handling Occlusions

PoseCNN: 3D Translation Estimation

Labels

#classes

3 × #classes

Center direction X

Center direction Y

Center distance

Hough voting layer

PoseCNN: 3D Rotation Regression

Labels

#classes

Hough voting layer

Center direction X

Center direction Y

Center distance

3 × #classes

RoIs

For each RoI

4 × #class

RoI pooling layers

6D Poses
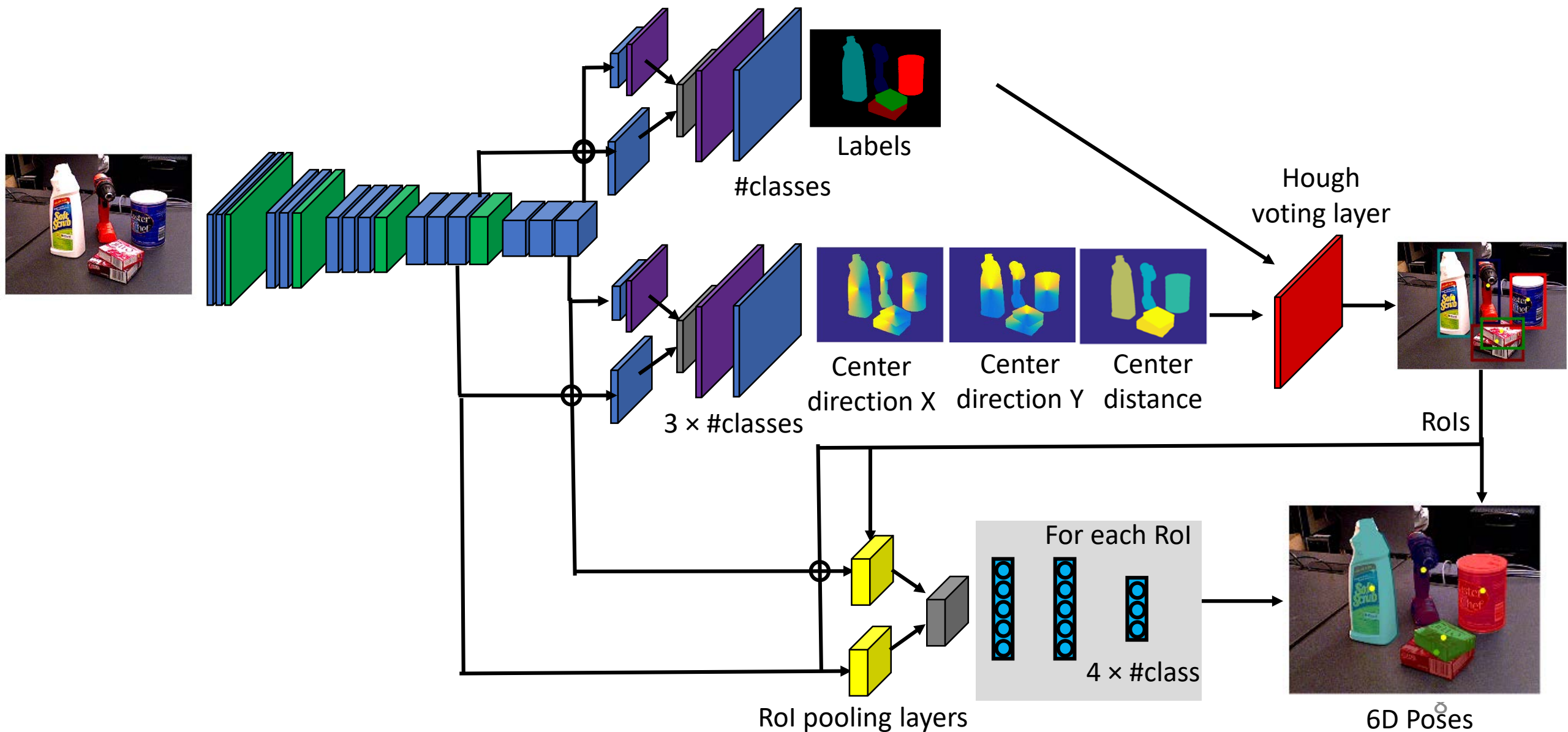
PoseCNN: 3D Rotation Regression Loss Functions

Ground truth rotation

Pose Loss (non-symmetric)

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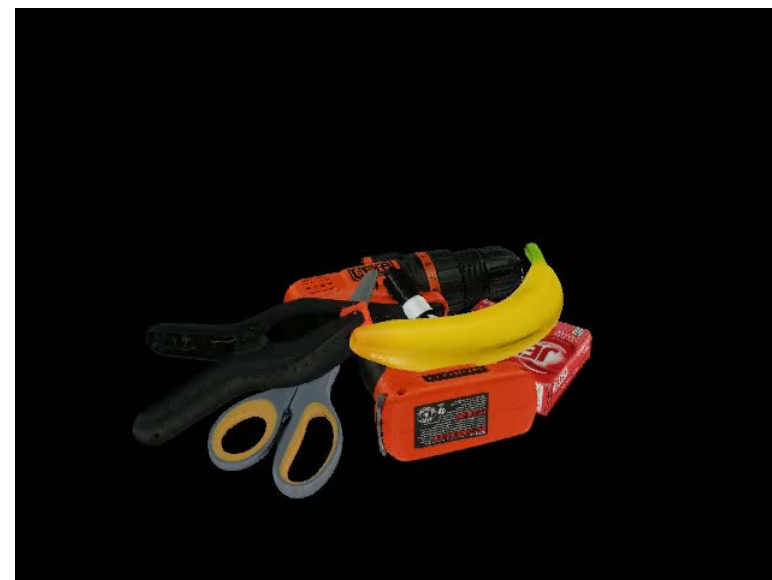
Predicted rotation

Shape-Match Loss for symmetric objects (symmetric)

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

3D model points
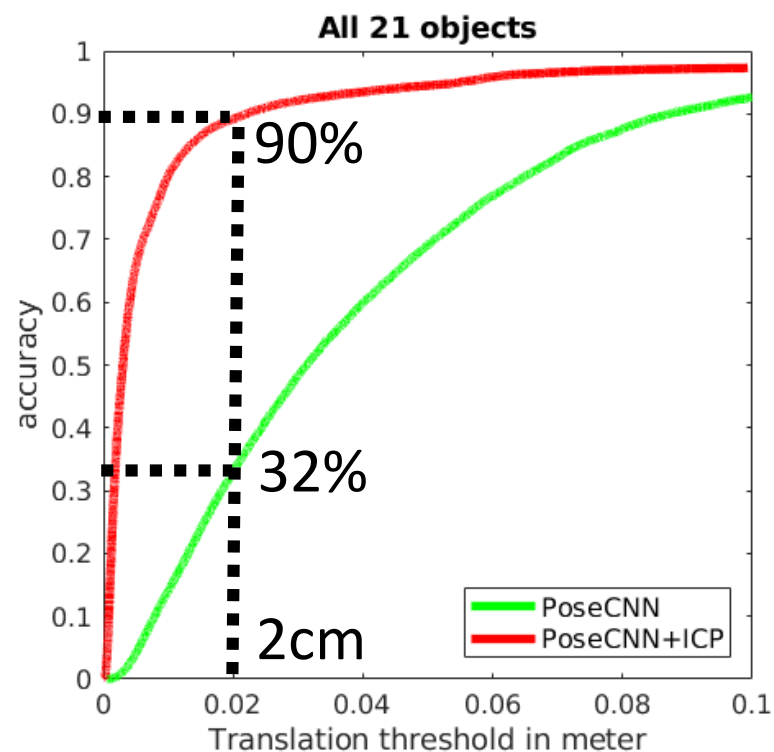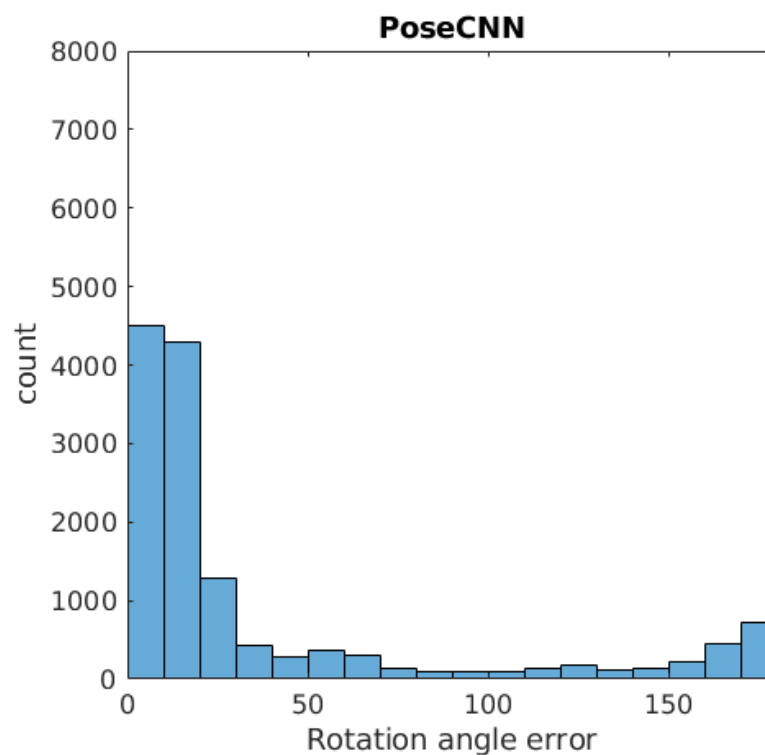
Our YCB-Video Dataset

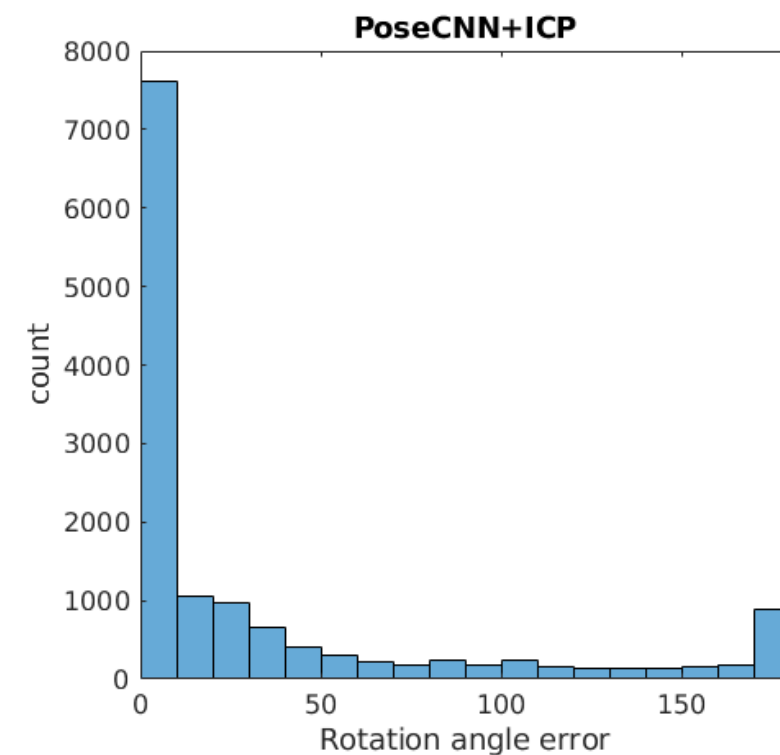
21 YCB Objects

92 Videos, 133,827 frames
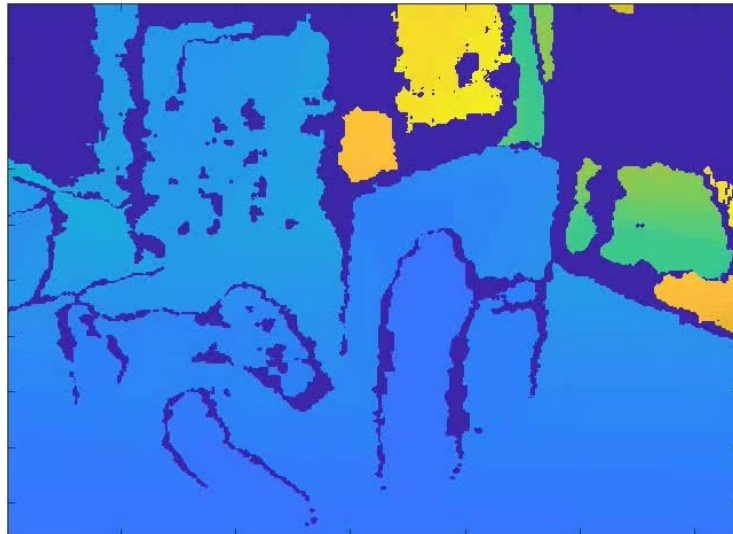
Results on the YCB-Video Dataset

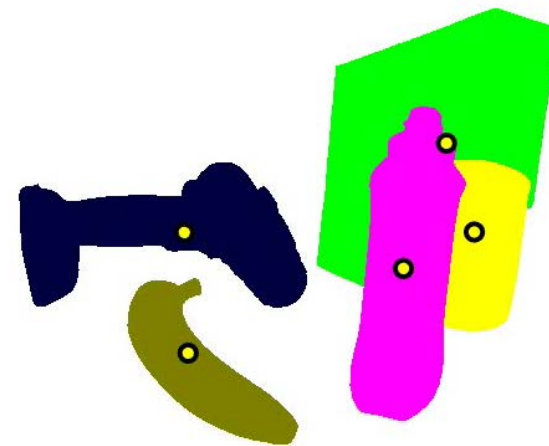3D Translation Error Analysis

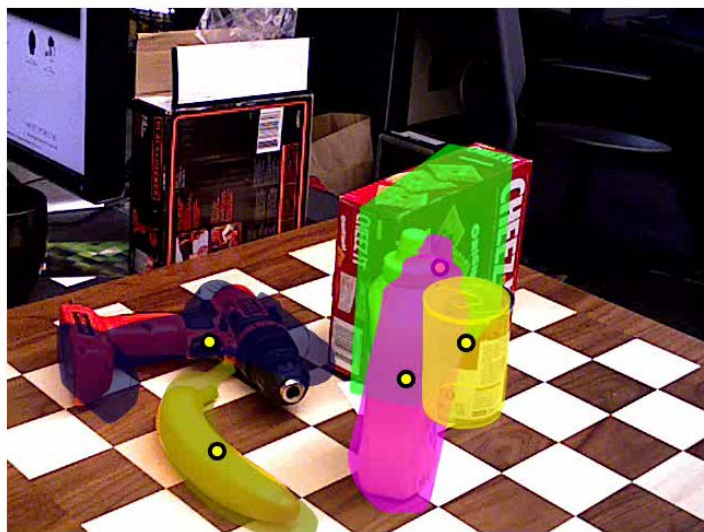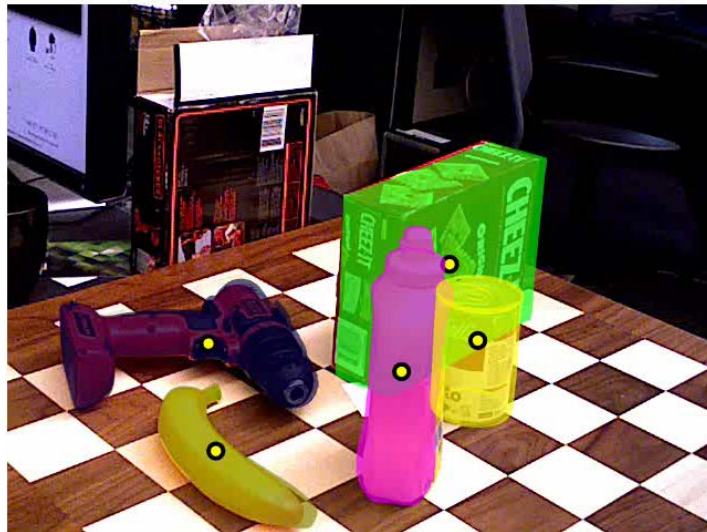3D Rotation Error Analysis
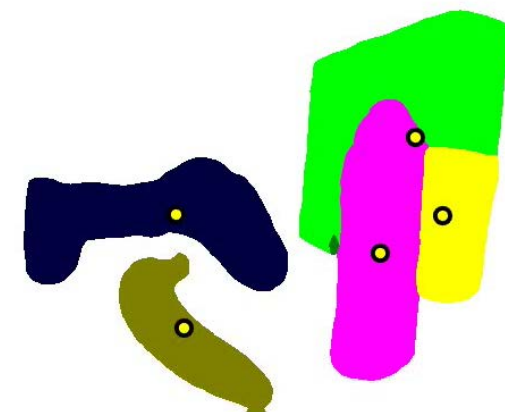
RGB

Depth

Groundtruth Labels

PoseCNN (RGB only)

PoseCNN + ICP

Predicted Labels

Conclusion

❑ PoseCNN

    o An end-to-end neural network for 6D pose estimation

    o Handle texture-less objects, symmetric objects and occlusions

    o Code and dataset are available online

Thank you!