Multi-View RGB-D Fusion for 6D Pose Estimation

BOSCH

Introduction
Current Challanges with 6D Pose Estimation

6D(oF) Pose Estimation → Estimate Rotation (SO3) and Translation (R3) of objects in the scene

Heavy Occlusions

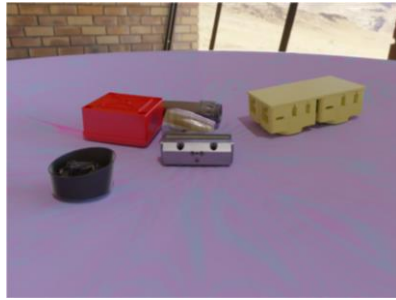▶ Some objects are not visible enough

Symmetric Objects

▶ Loss function favors keypoints on symmetry axis

BOSCH

Dataset
Overview

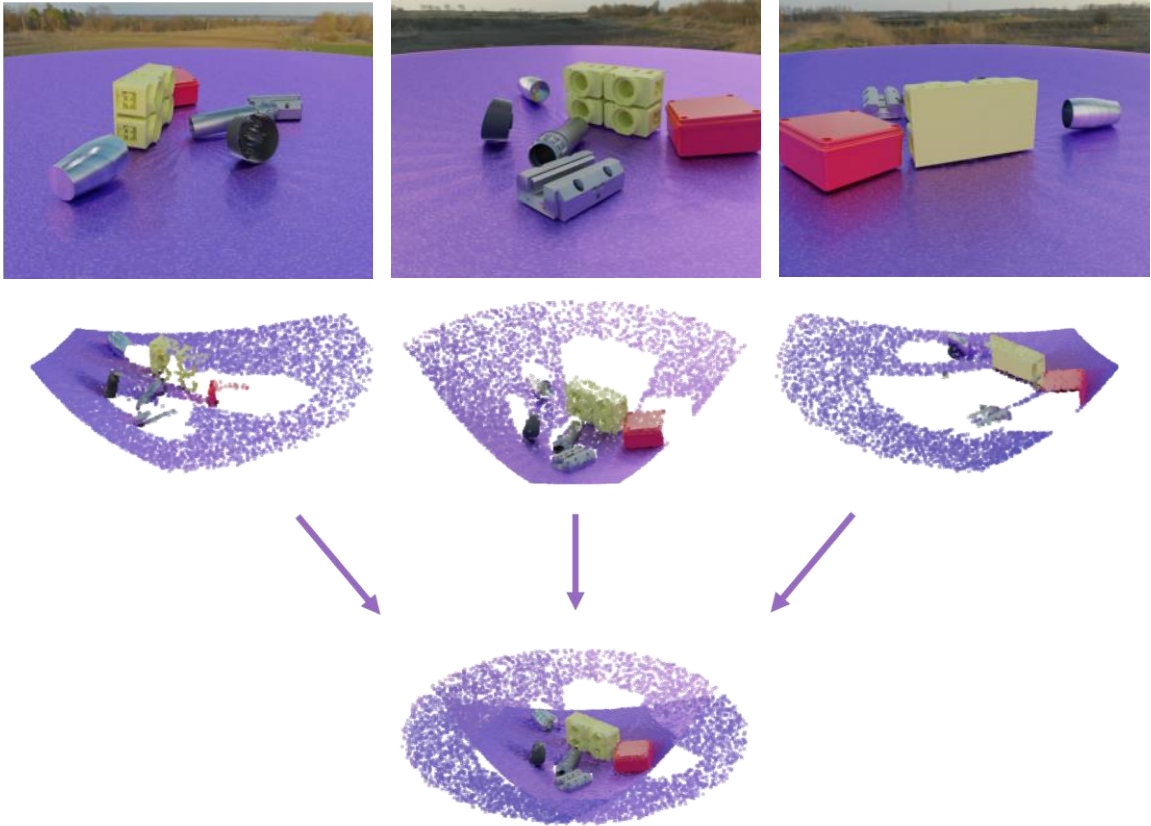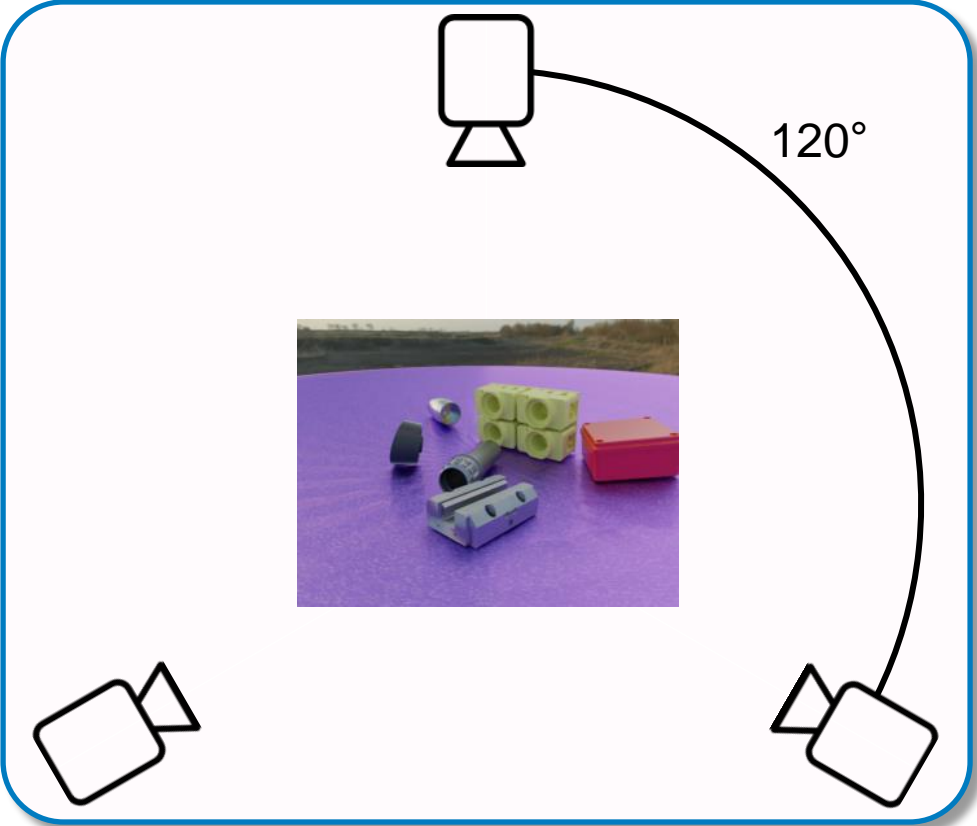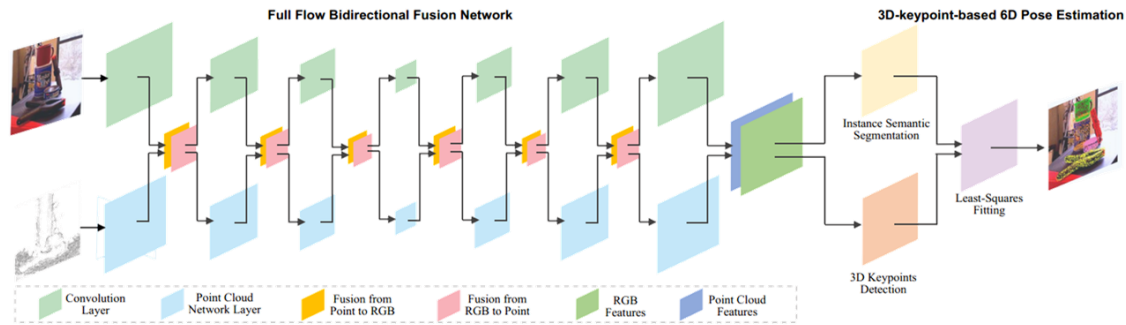|	YCB-Video	SCAPE YCB	SCAPE 2	SCAPE YCB2
Real/Synthetic	Real	Synthetic	Synthetic	Synthetic
Views	1	3	3	5
Occlusions	Few	Many	Few	Many
Symmetries	Few	None	Many	None
Sample				

BOSCH

Dataset
SCAPE 2 Visualization

▶ Using known camera positions for optimal depth fusion

BOSCH

Method
FFB6D: Multi-View Extension

Full Flow Bidirectional Fusion Network

3D-keypoint-based 6D Pose Estimation

Instance Semantic Segmentation

Least-Squares Fitting

3D Keypoints Detection

Convolution Layer   Point Cloud Network Layer   Fusion from Point to RGB   Fusion from RGB to Point   RGB Features   Point Cloud Features

[He et al.: FFB6D: A Full Flow Bidirectional Fusion Network for 6D Pose Estimation, CVPR21]

3. Novel Fusion Module

Fusion module

CNN Encoder   CNN Decoder

Instance Segmentation

Least-Squares Fitting

Testing

Training

Loss

3D Keypoint Estimation

1. Fused Point Cloud

Concatenate

2. Batched Img Branch

BOSCH
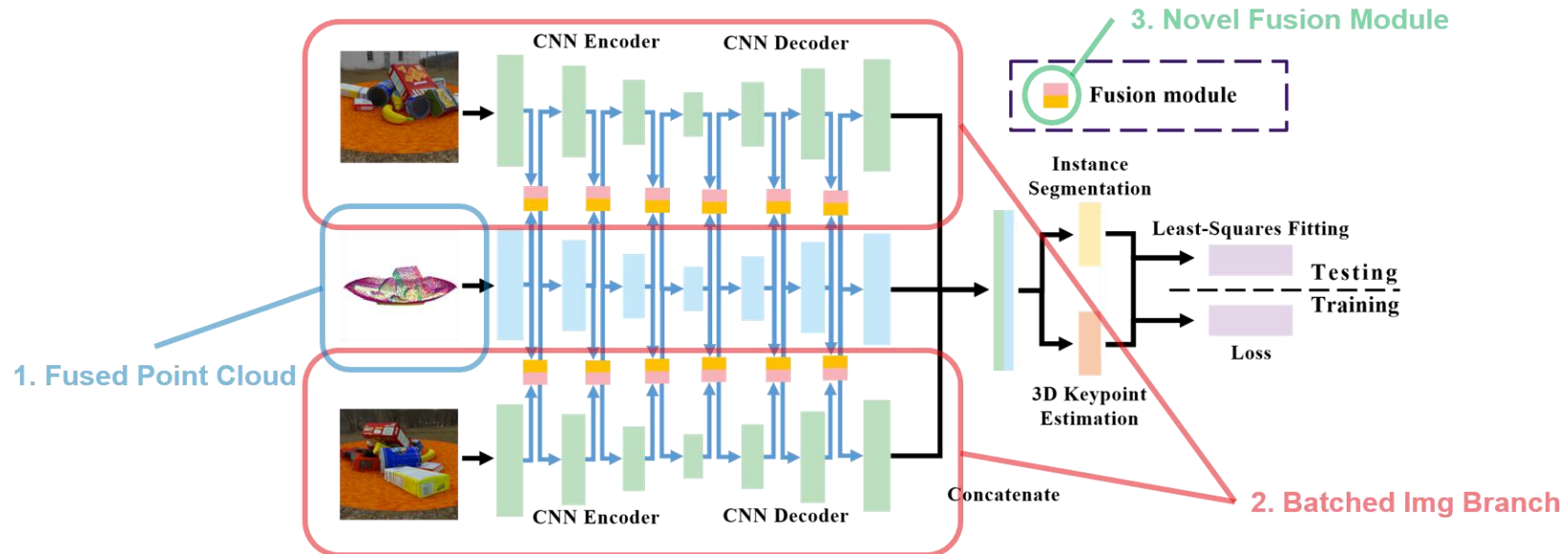
Method
FFB6D Symmetry Extension
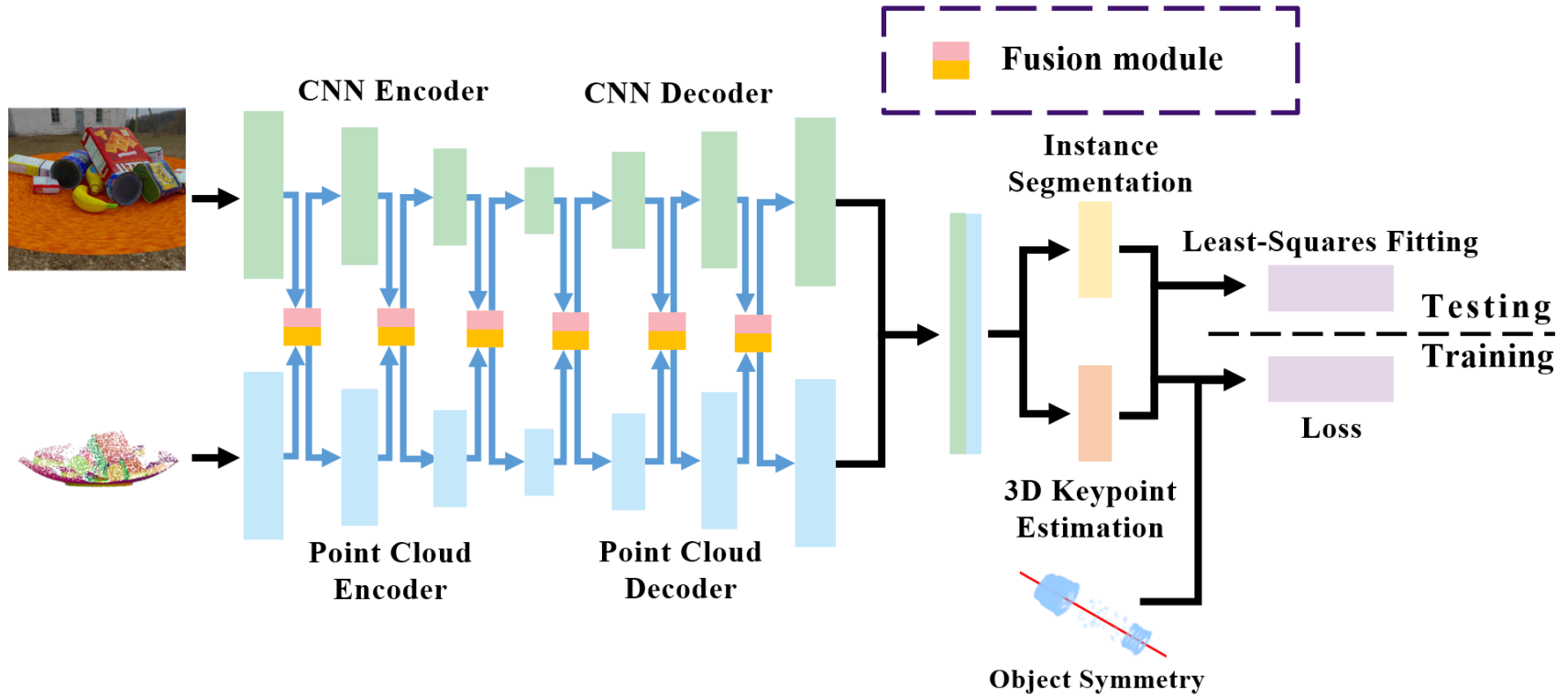
$$L_{keypoints} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} \|of_i^j - of_i^{j*}\| \mathbb{I}(p_i \in I) \quad \rightarrow \quad L_{keypoints} = \frac{1}{N} \min_{S \in S_O} \sum_{i=1}^{N} \sum_{j=1}^{M} \|of_i^j - Sof_i^{j*}\| \mathbb{I}(p_i \in I)$$
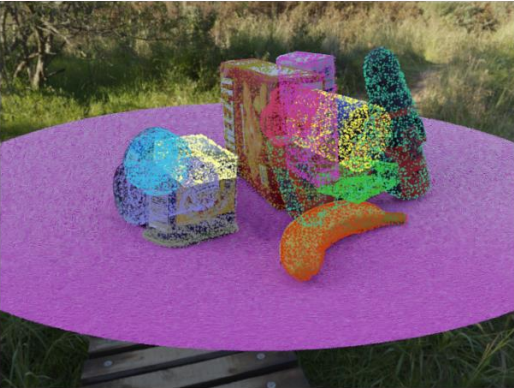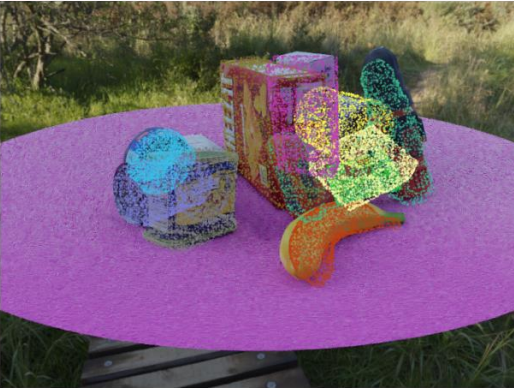
BOSCH

Multi-View − Experiments
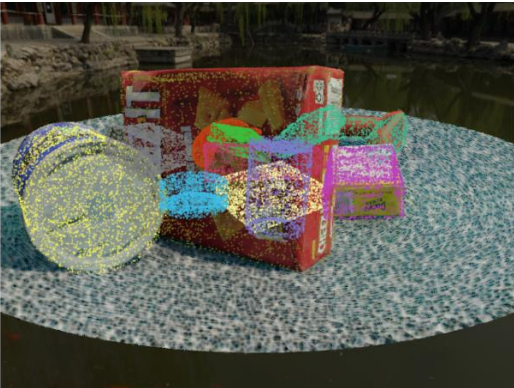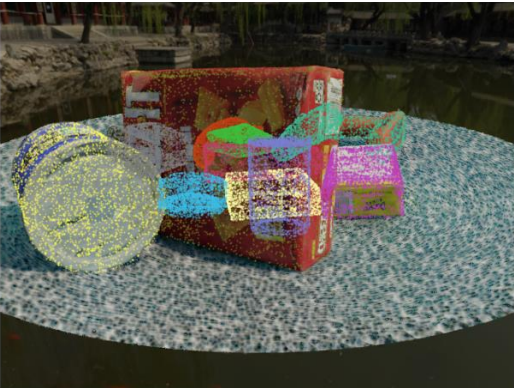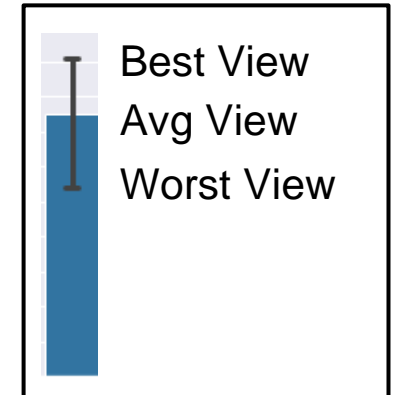## SCAPE YCB: Qualitative Results MV-FFB6D

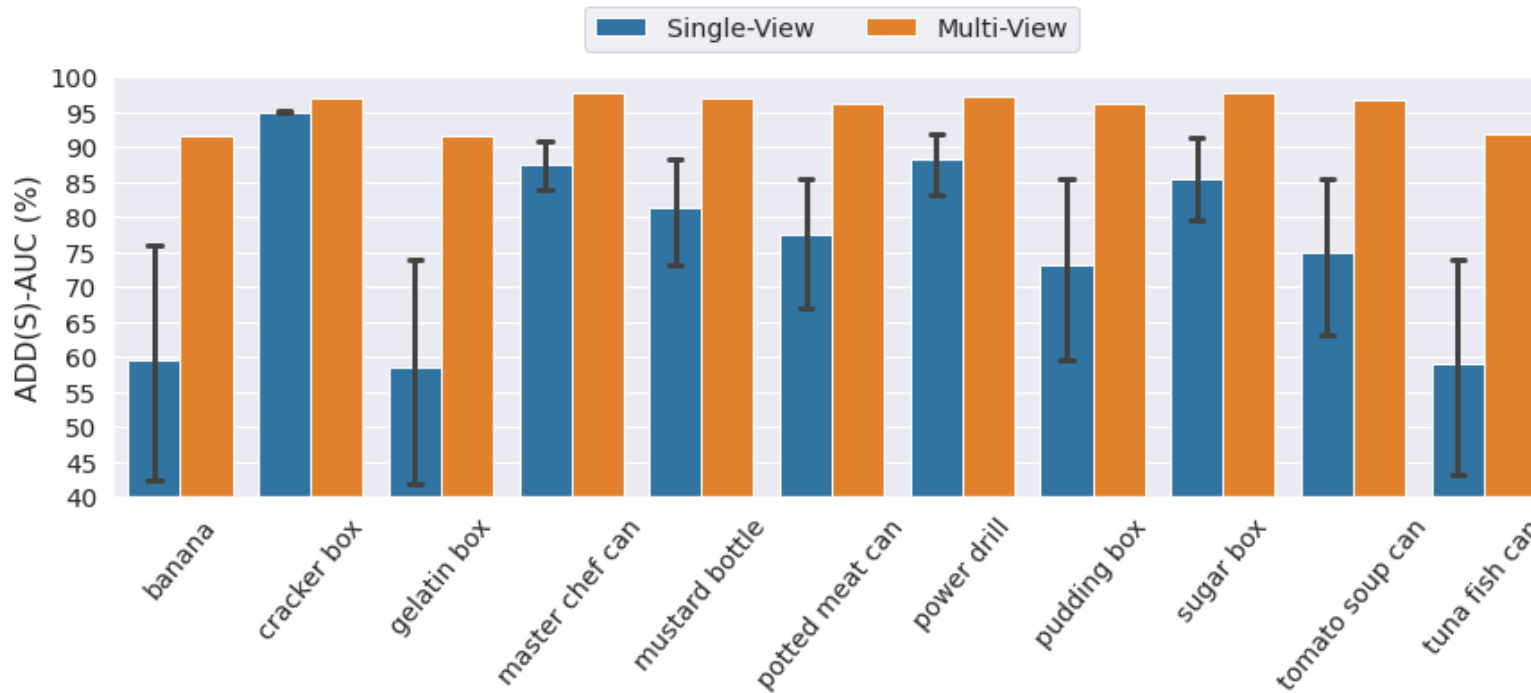Source | Ground Truth | Single-View Prediction | Multi-View Prediction

BOSCH

Multi-View − Experiments
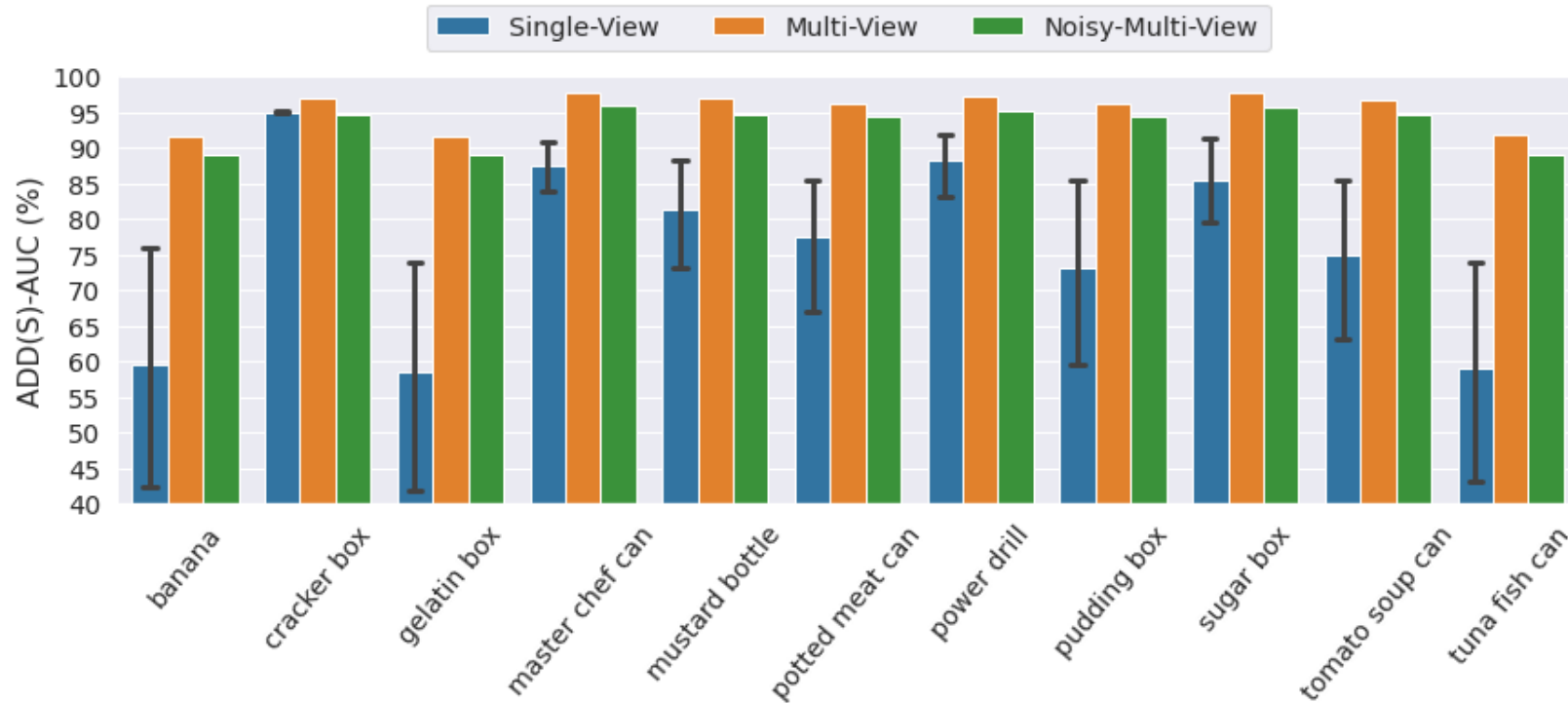## SCAPE YCB: Single-View vs Multi-View

▶ The single-view model is evaluated on all multi-view frames individually and ranked by accuracy

BOSCH

Multi-View − Experiments
SCAPE YCB: Single-View vs Multi-View vs Multi-View Noisy

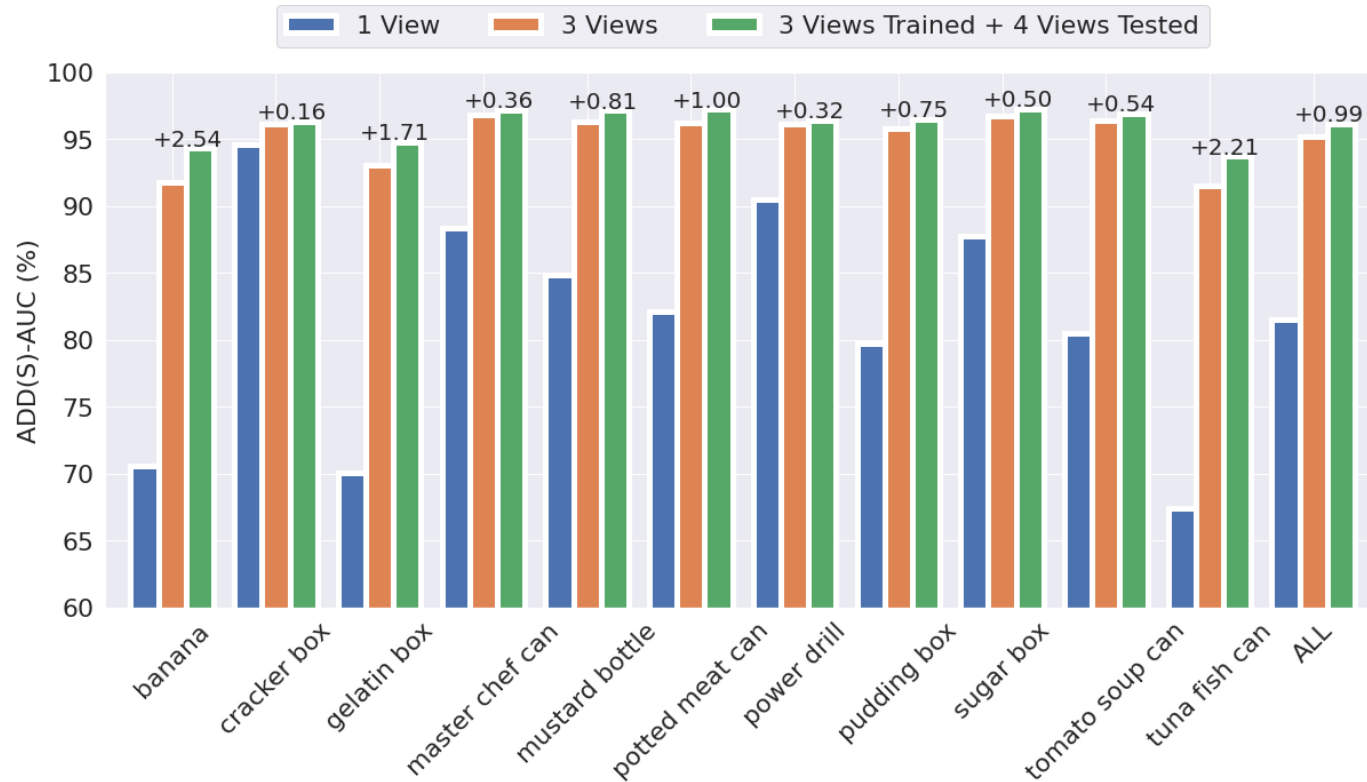▶ Multi-View Fusion relies heavily on known camera positions
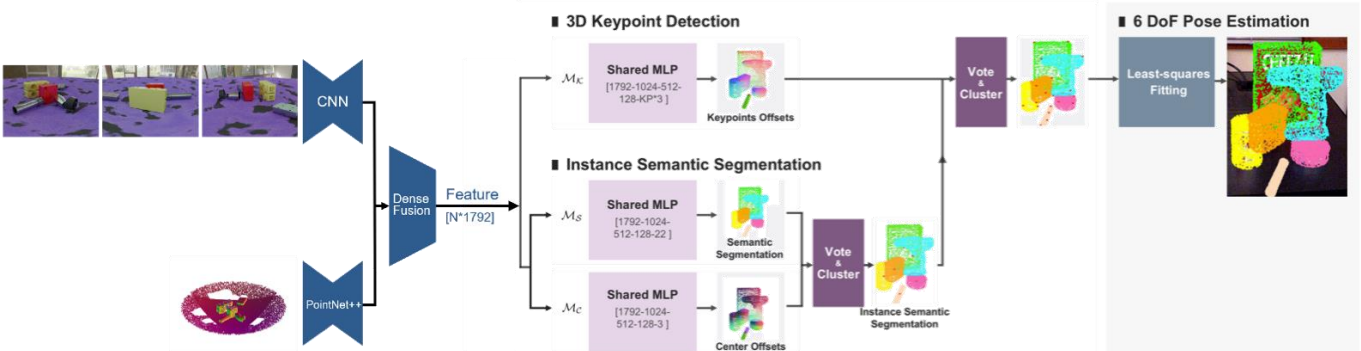

Accurate Poses

Noisy Poses

BOSCH

Experiments
SCAPE YCB2: Dynamic Setup

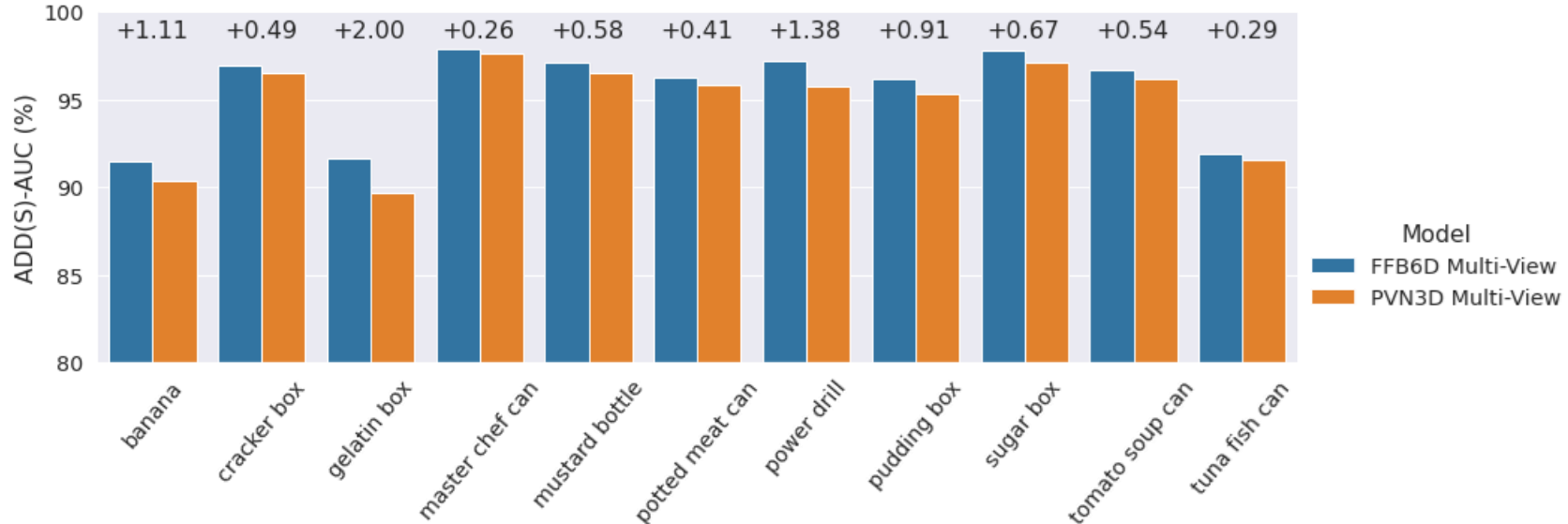▶ In SCAPE YCB2 each scene has a different camera setup!

BOSCH

Multi-View − Experiments
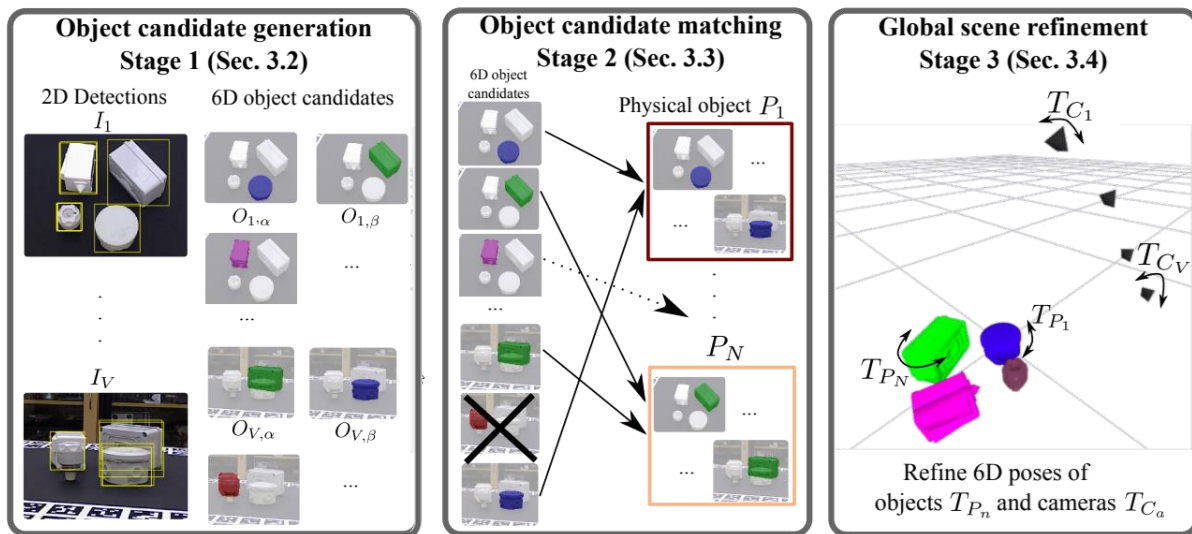## SCAPE YCB: MV-FFB6D vs MV-PVN3D

[Demmler: Multi-view 6D Pose Estimation on RGB-D Frames using a Deep Point-wise Voting Network, 2021]

BOSCH

Multi-View − Experiments
YCB-Video: MV-FFB6D vs CosyPose

| | Object candidate generation Stage 1 (Sec. 3.2) | Object candidate matching Stage 2 (Sec. 3.3) | Global scene refinement Stage 3 (Sec. 3.4) |

[Labbe et al.: CosyPose: Consistent multi-view multi-object 6D pose estimation, ECCV20]

YCB-Video-Dataset		3 views	5 views
CosyPose	ADD-S↑	92.29	93.40
	ADD(S)↑	87.66*	88.80*
MV-FFB6D	ADD-S↑	**95.16**	**95.29**
	ADD(S)↑	**91.37**	**91.58**

*Reimplemented results

ADD-S – Average Closest Point Distance

ADD(S) – Average (Closest Point) Distance

▶ Comparison not totally fair:

  ▶ RGB vs RGB-D
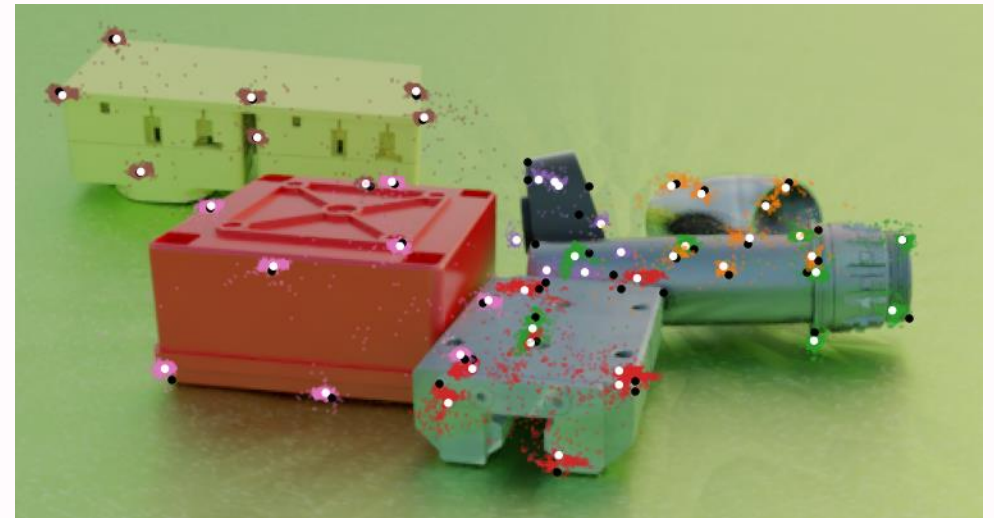
  ▶ Unknown camera poses vs known camera poses

BOSCH

Symmetry – Experiments
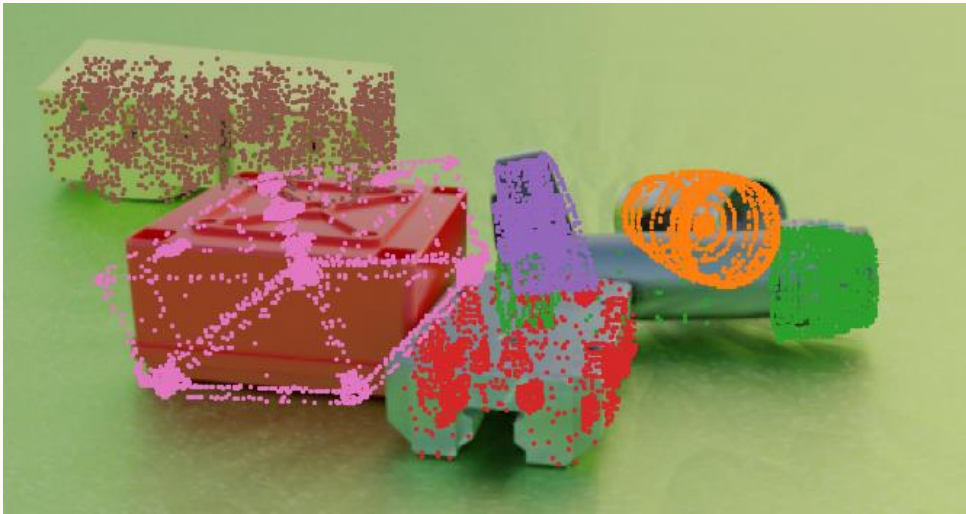## SCAPE 2: Symmetry Keypoint Proposals
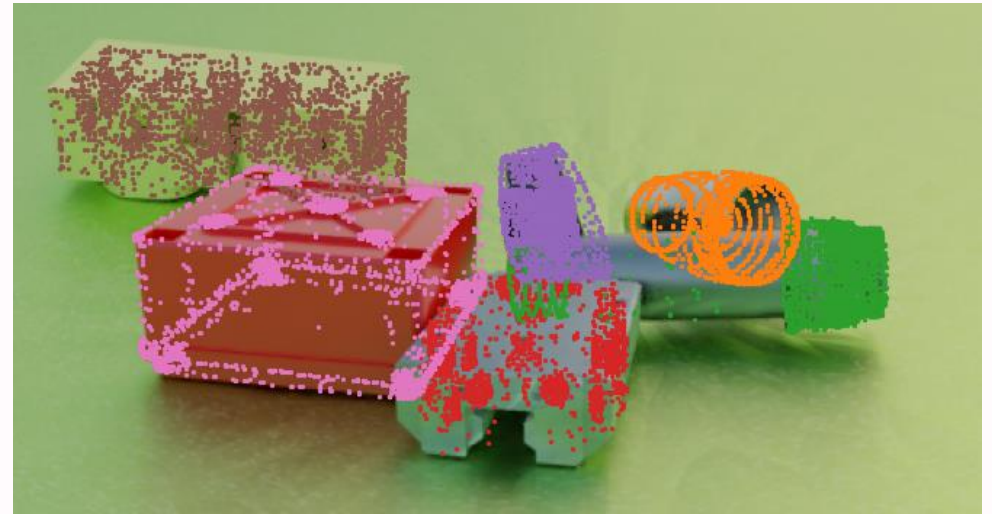

Standard Training

Symmetry-aware training

BOSCH

Symmetry – Experiments
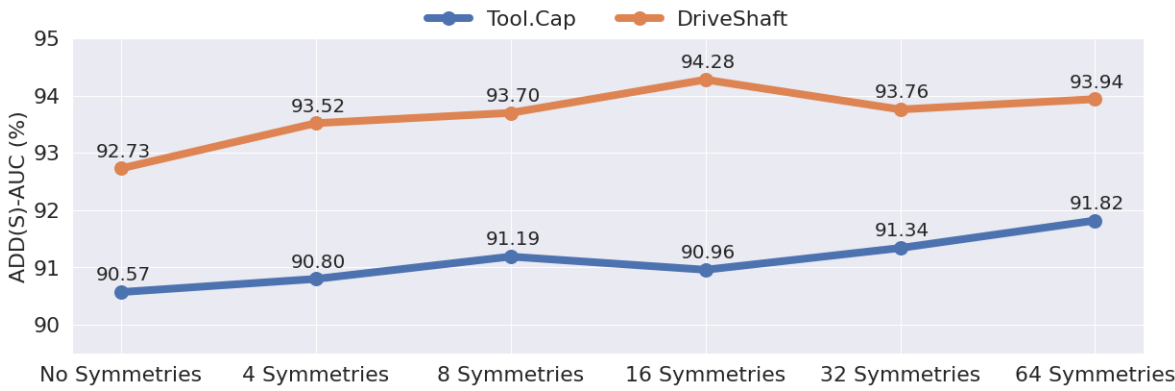## SCAPE 2: Poses with Symmetry
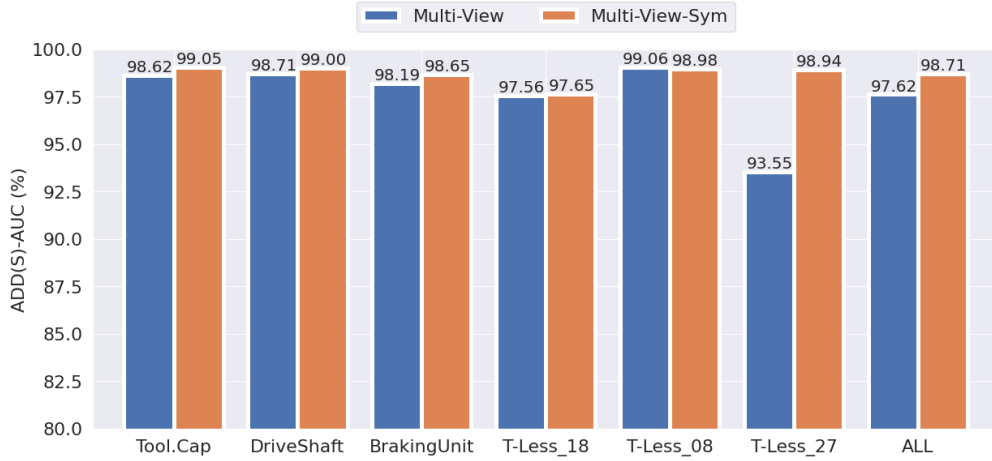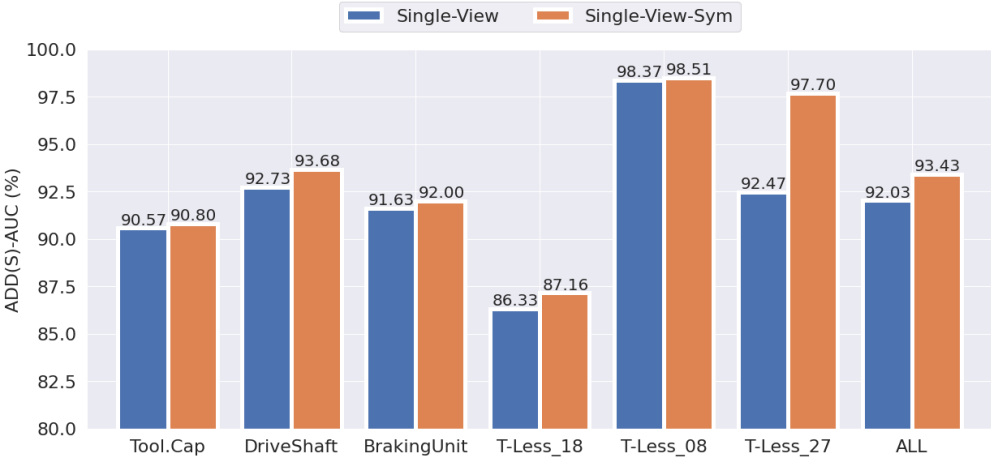

Standard Training

Symmetry-aware training

BOSCH

Symmetry – Experiments
SCAPE 2: Symmetry Results

BOSCH

Conclusion

▶ **State-of-the-art multi-view 6D Pose Estimation architecture**

   ▶ Robustness towards noisy camera positions

   ▶ Robustness towards changing camera poses → SCAPE YCB2

   ▶ Robustness towards changing amount of multi-view cameras → SCAPE YCB2

▶ **Novel Symmetry-Aware training**

▶ **Open research directions**

   1. Multi-View/Symmetry 6D Pose Estimation on real dataset

   2. Fairer comparison with CosyPose (RGB-D + known camera positions) → Using FFB6D as first stage

BOSCH