CroCPS: Addressing Photometric Challenges in Self-Supervised Category-Level 6D Object Poses with Cross-Modal Learning

Intro

1. We propose **CroCPS**, **a multimodal self-supervision pipeline for category-level object pose estimation**, addressing the photometric challenges for metallic objects

2. We provide a synthetic dataset containing synthetic cutlery models and a real dataset with ground truth pose annotations, for exploring self-supervision of photometrically-challenging object poses from the synthetic to real images.

3. Our pipeline leveraging polarization images improved both 2D detections and 6D category-level object poses in the experiment.
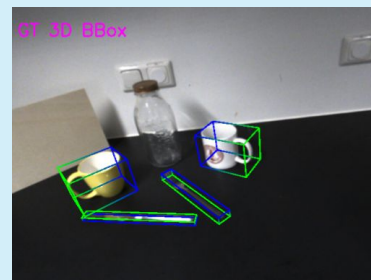
Datasets
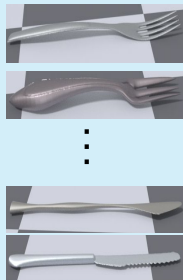
(a) Collected synthetic cutlery models
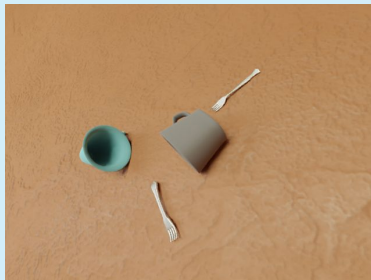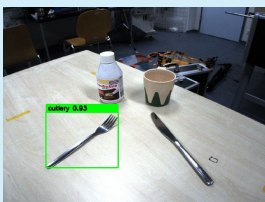
(b) Rendered synthetic images with Synthetic cutlery models

(c) Real dataset capture device

(d) One of the image in real dataset with accurate pose annotations

Self-Supervision of 2D Detections

(a) Detected 2D bounding box

(b) Degree of Linear Polarization

(c) Extracted mask from polarization

- The polarization image is used for verification of the detected 2D bounding box
- Verification results are leveraged as self-supervision signal for overcoming the domain gap

	AP$_{50}$	Recall$_{50}$	AP	Recall	F1
w/o self-supervision	60.47	63.75	36.28	38.75	37.47
with self-supervision	100	100	45.89	45.89	45.89

Table 1: Evaluation results of 2D detections

Self-Supervision of 6D Category-level Object Poses
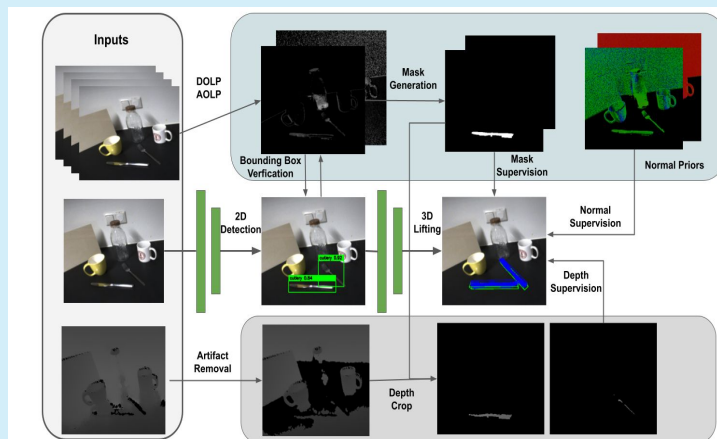
- Self-supervision of 6D object poses with normal priors, polarization masking and partial depth image

- Experiment Evaluations shows the effectiveness of the losses

3D$_{25}$ / 3D$_{50}$	3D$_{25}$	3D$_{50}$
mask+normal	0	0
mask+depth	81.25	25
mask+normal+depth	85	30