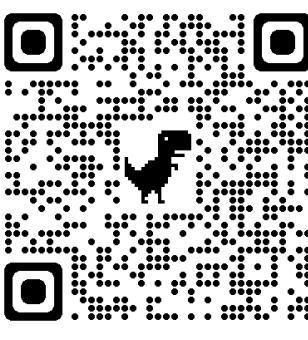


SA6D: Self-Adaptive Few-Shot 6D Pose Estimator for Novel and Occluded Objects

Ning Gao^{1,2} Ngo Anh Vien¹ Hanna Ziesche¹ Gerhard Neumann²

¹Bosch Center for Artificial Intelligence ²Autonomous Learning Robots, KIT



ABSTRACT

Goal:

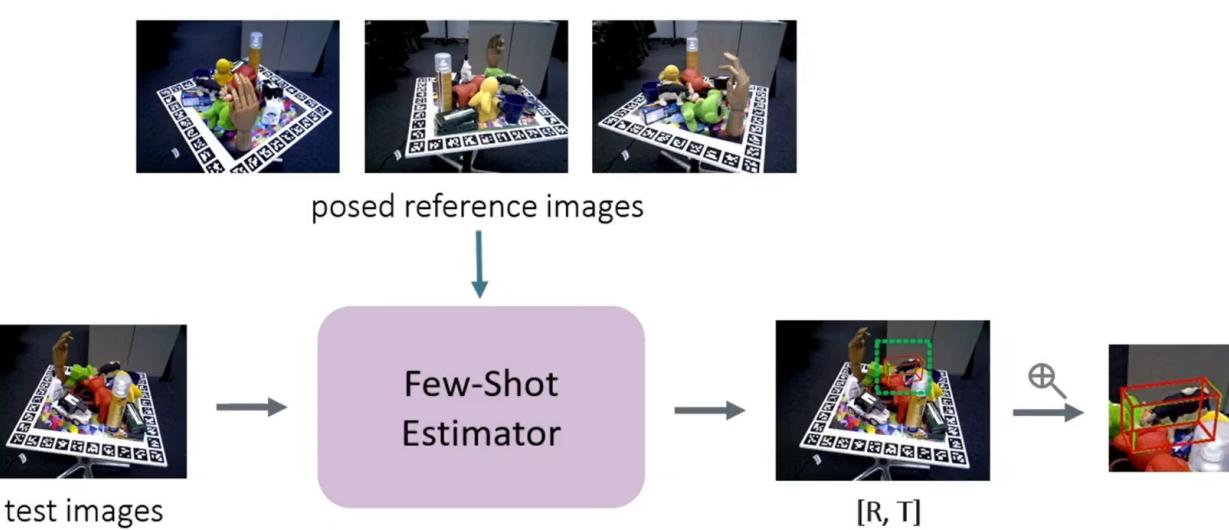
Exploit generic few-shot 6D pose estimation of novel objects from unseen categories, especially under severe clutter.

Contributions:

- SA6D increases the performance and robustness against heavy occlusions while *not* requiring any object information (3D model, object diameter or ground-truth mask) or object-centric images, while only requiring a small set of posed RGB-D reference images with known poses of the novel object.
- SA6D employs an online-adaptive segmentation module to identify the target object during inference.
- SA6D utilizes pretrained models from prior work *without* any retraining processes.
- SA6D significantly outperforms current state-of-the-art methods against occlusion in real-world scenarios while trained entirely on synthetic data.

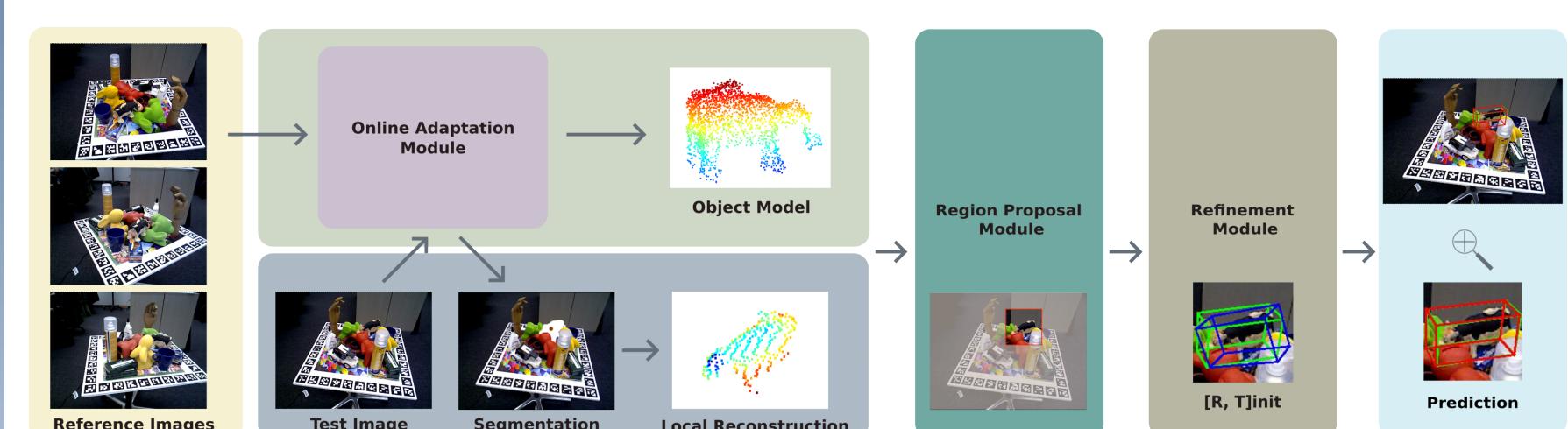
TASK DESIGN

Few-Shot 6D Novel Object Pose Estimation:



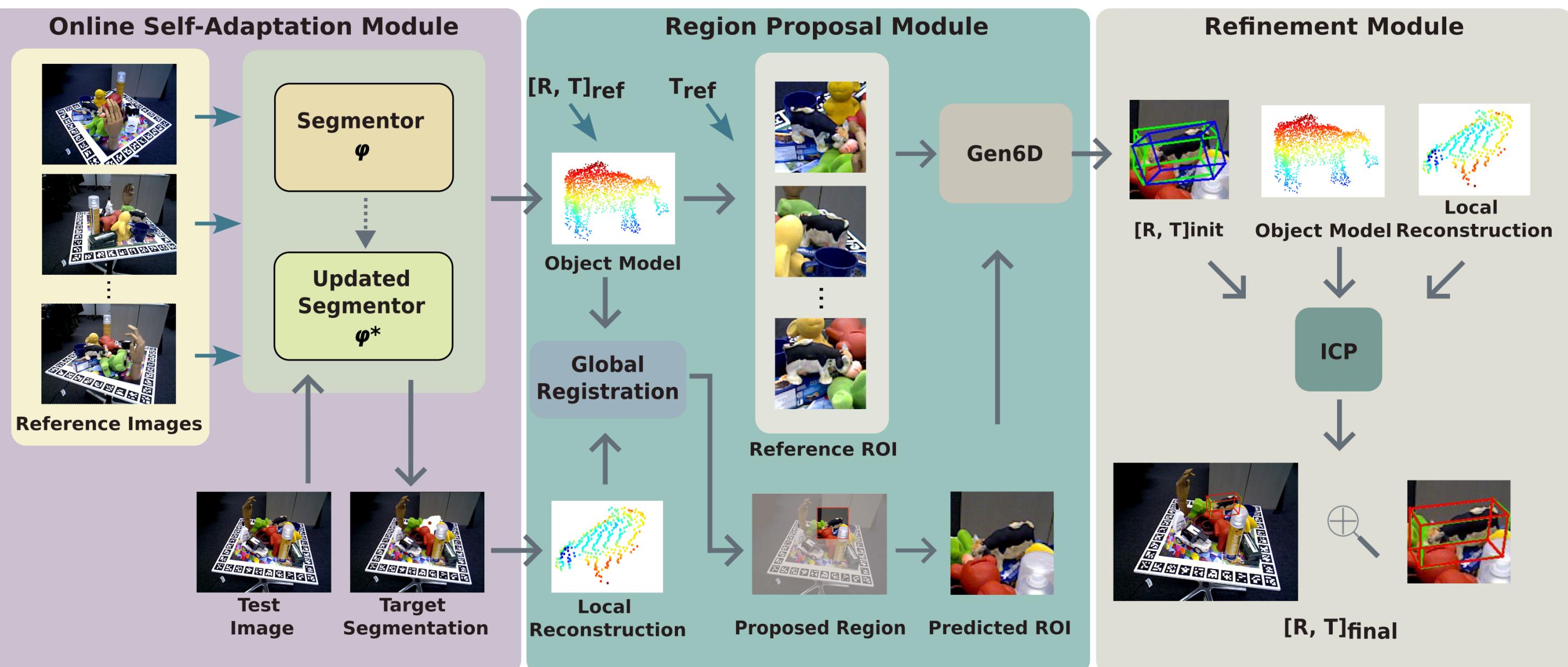
Predict 6D pose of a novel object in a new image (**test image**) from a few reference images with the known pose of the object (**reference images**).

OVERVIEW



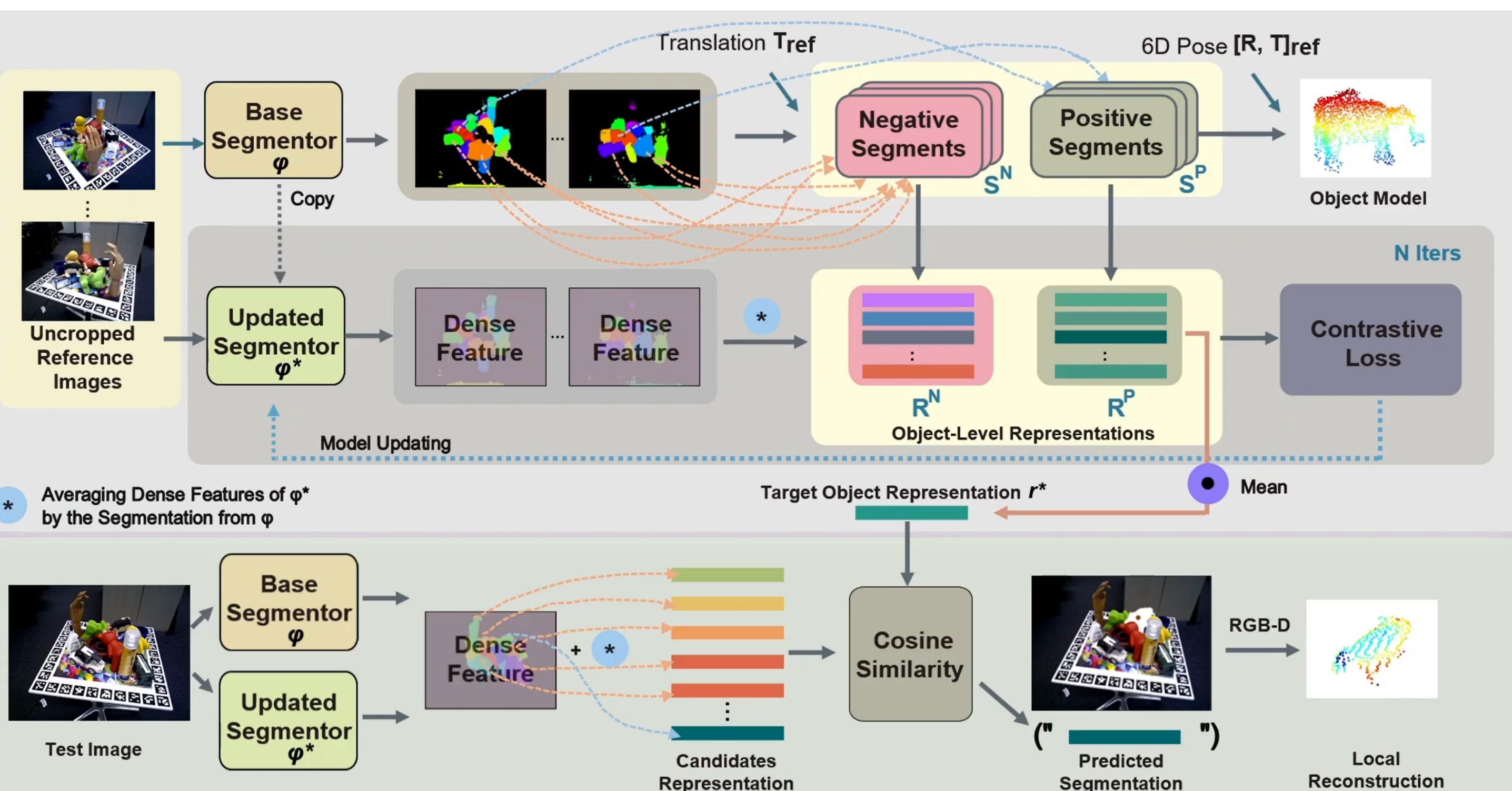
We present a generalizable and category-agnostic few-shot 6D object pose estimator using a small number of posed RGB-D images as references. Compared to existing methods, our approach provides robust and accurate predictions of novel objects against occlusions without requiring retraining or any object information.

PIPELINE



Overview. SA6D includes three modules: i) The *online self-adaptation module* discovers and segments the target object (*milk cow*) from a cluttered scene giving a few posed RGB-D images as reference. ii) The *region proposal module* outputs a robust region of interest (ROI) of the target object against occlusion by incorporating visual and geometric features. A coarse 6D pose is then estimated by Gen6D and iii) further fine-tuned using ICP.

Online Adaptation Segmentation:



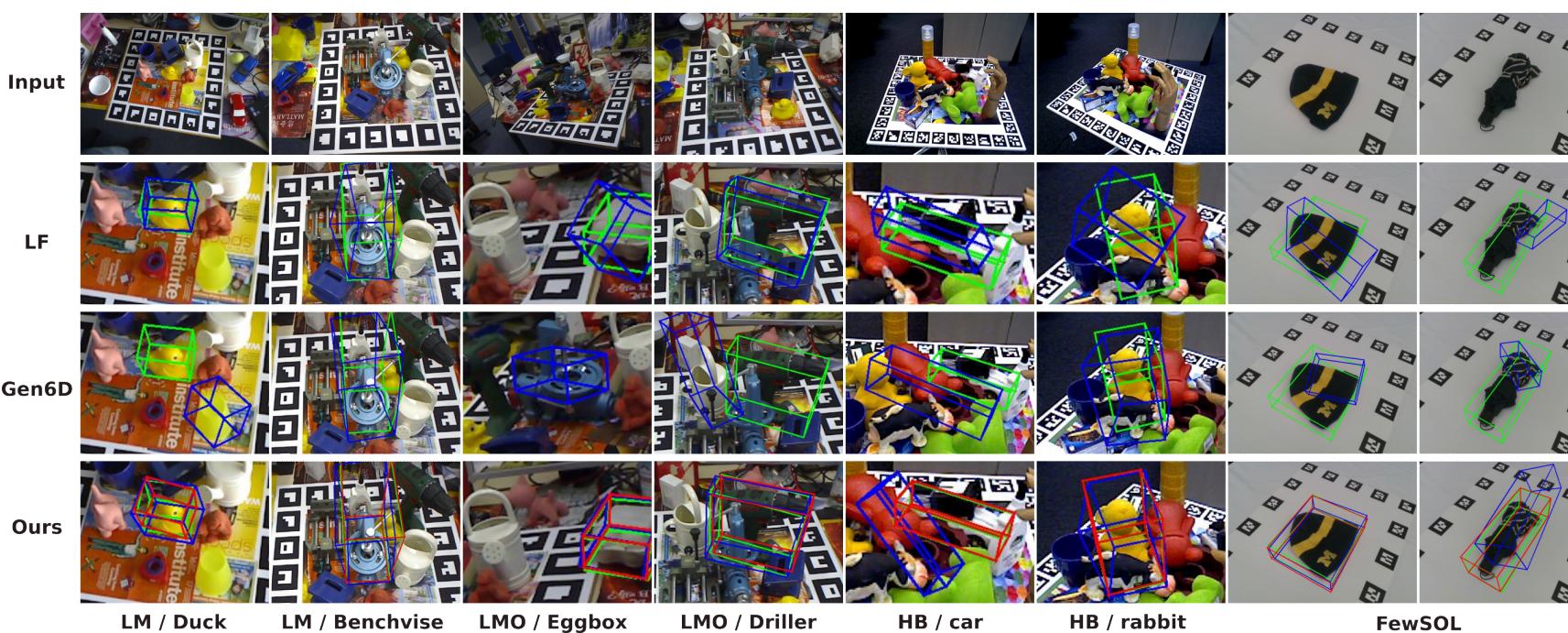
Online self-adaptation module. A pretrained segmentor φ is applied on reference images to predict segmentations. With the ground-truth translation of the target object in the reference images T_{ref} , the object center can be reprojected to the image. For each reference image, one segment is chosen as a positive sample if it includes the reprojected object center while the remaining segments are considered as negative samples. Subsequently, an object-level representation of each segment is computed by averaging the pixel-wise dense features from φ^* . A contrastive loss is then applied over the positive and negative object representations and updates φ^* iteratively. After adaptation, φ^* generates the target object representation r^* by averaging over all positive representations from reference images. Given a test image, we compute the cosine similarity between each candidate and r^* and the most similar candidate is chosen as the segment of the target object.

EXPERIMENTS

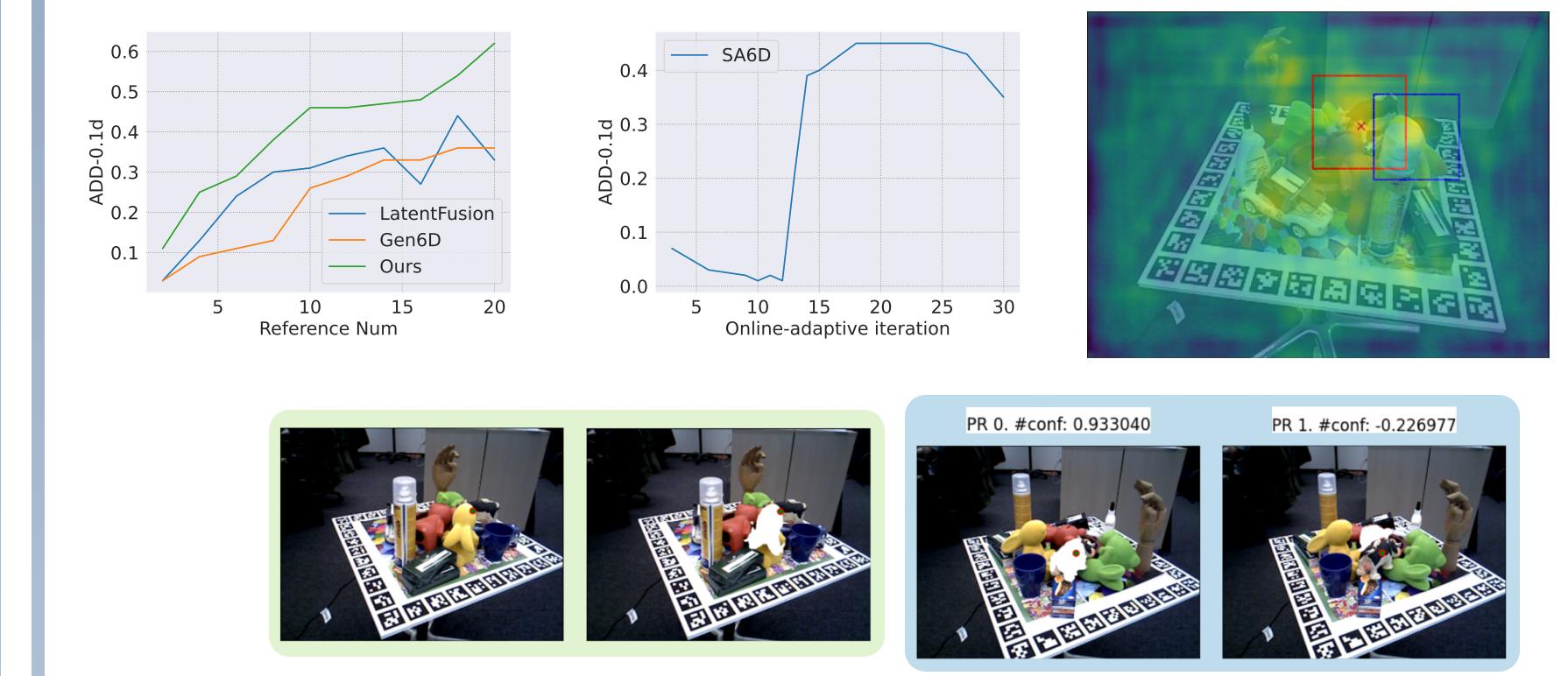
Contrastive loss:

$$l_{ij} = -\log \frac{\exp(\text{sim}(r_i^P, r_j^P)/\tau)}{\sum_{r' \in R^N \cup \{r_j^P\}} \exp(\text{sim}(r_i^P, r')/\tau)}, \quad (1)$$

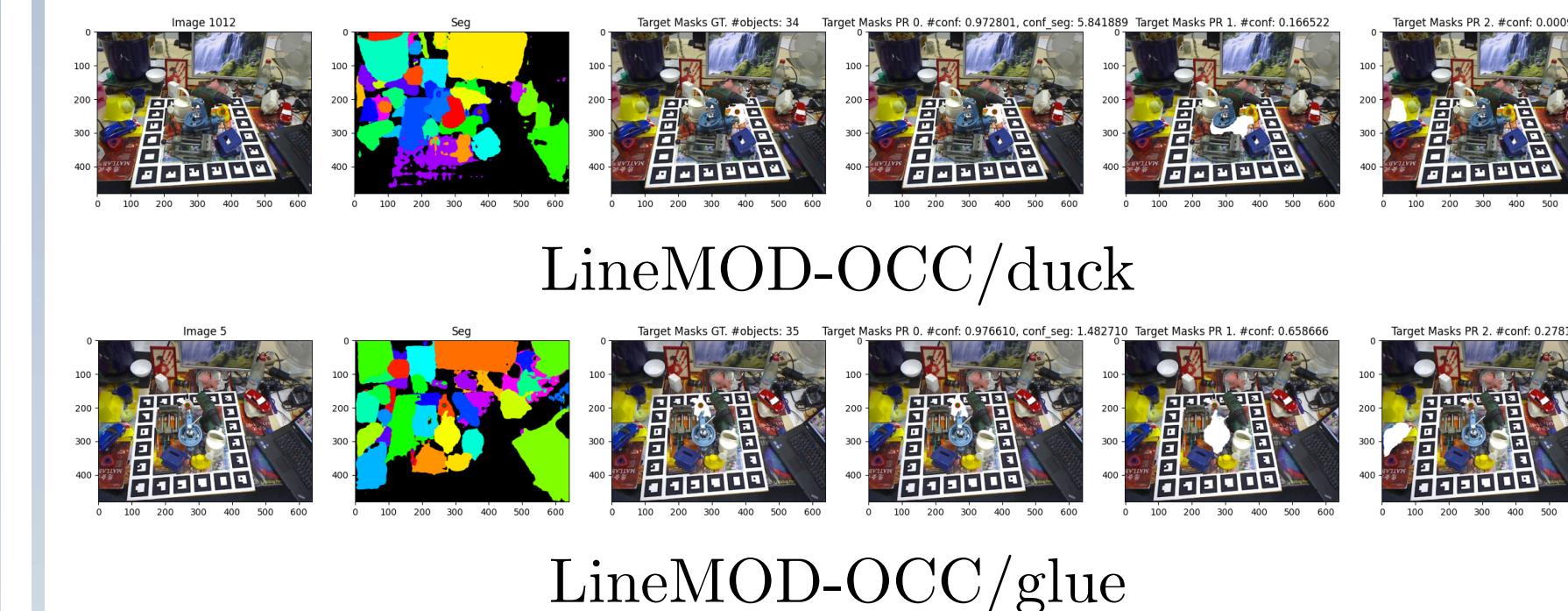
Qualitative results:



More analysis:



Robust prediction of target segmentation:



Quantitative results:

Method	GT-Mask	Ref. Num	LineMOD-OCC/line-mod-dataset						Home-RevivedDB/line-mod-revivedDB						
			eggbox	duck	benchvis	can	cat	glue	avg.	eggbox	duck	glue	ape	avg.	
Gen6DGen6D	20	0.63	0.45	0.29	0.25	0.26	0.36	0.07	0.09	0.12	0.21	0.26	0.15	0.20	0.30
SAGD (ICP)	20	0.53	0.31	0.25	0.21	0.17	0.31	0.17	0.18	0.14	0.23	0.17	0.20	0.44	0.26
SAGD (wo/ RPM)	20	0.53	0.37	0.27	0.21	0.17	0.31	0.17	0.18	0.14	0.23	0.17	0.20	0.45	0.35
SAGD (wo/ RFM)	20	0.57	0.36	0.24	0.20	0.18	0.38	0.15	0.18	0.14	0.28	0.12	0.20	0.55	0.32
SAGD	20	0.73	0.55	0.50	0.47	0.72	0.62	0.45	0.26	0.30	0.21	0.53	0.42	0.32	0.33
Gen6D	64	0.80	0.74	0.71	0.68	0.75	0.79	0.44	0.41	0.38	0.31	0.36	0.42	0.72	0.69
LatentFusion	20	0.61	0.62	0.68	0.75	0.78	0.67	0.28	0.09	0.18	0.45	0.17	0.33	0.09	0.16
SAGD	64	0.80	0.84	0.73	0.80	0.84	0.75	0.79	0.44	0.41	0.38	0.31	0.42	0.72	0.69
LFLatentFusion	20	0.61	0.56	0.32	0.54	0.30	0.26	0.29	0.38	0.10	0.06	0.08	0.04	0.14	0.11
SAGD (wo/ RFM)	20	0.56	0.55	0.32	0.54	0.30	0.26	0.29	0.38	0.10	0.06	0.08	0.04	0.14	0.11
SAGD	64	0.68	0.58	0.60	0.73	0.72	0.78	0.72	0.33	0.26	0.29	0.30	0.34	0.58	0.42

Evaluation of ADD-0.1d

Method	IOU _{0.5}				5'2cm				5'5cm				10'5cm			
	ADD-0.1d	ADD-0.3d	ADDs-0.1d	ADDs-0.3d	eggbox	duck	glue	ape	eggbox	duck	glue	ape	eggbox	duck	glue	ape
CASScass	0.01	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Shape-Prismapeprior	0.33	0.03	0.04	0.14	0.14	0.23	0.17	0.20	0.44	0.26	0.23	0.17	0.20	0.44	0.26	0.20
DualPoseNetDualposenet	0.70	0.18	0.23	0.37	0.15	0.10	0.15	0.15	0.35	0.15	0.10	0.15	0.15	0.35	0.15	0.15
ReFoWWild6D	0.71	0.30	0.42	0.43	0.42	0.47	0.42	0.49	0.52	0.42	0.47	0.49	0.42	0.49	0.52	0.42
SA6D	0.65	0.37	0.40	0.42	0.68	0.78	0.83	0.87	0.90	0.68	0.78	0.83	0.68	0.78	0.83	0.78

Evaluation on FewSOL

Evaluation on Wild6D