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

Ning Gao[1,2] Ngo Anh Vien[1] Hanna Ziesche[1] Gerhard Neumann[2]

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

Abstract

Goal:

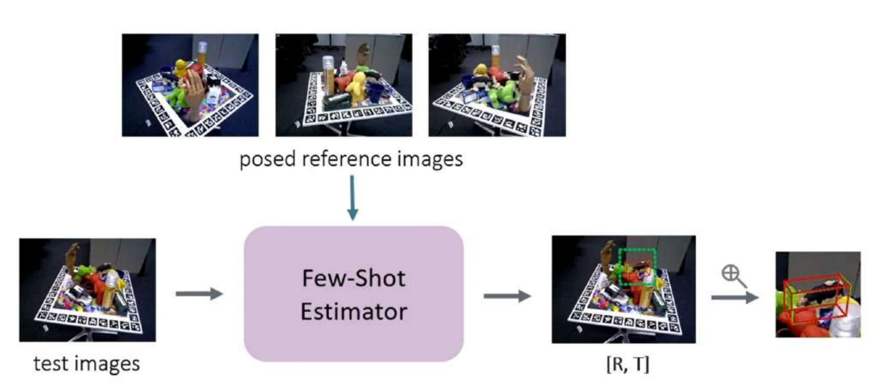
Exploit generic few-shot 6D pose estimation of novel objects from unseen categories, especially under severe clutter.

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.

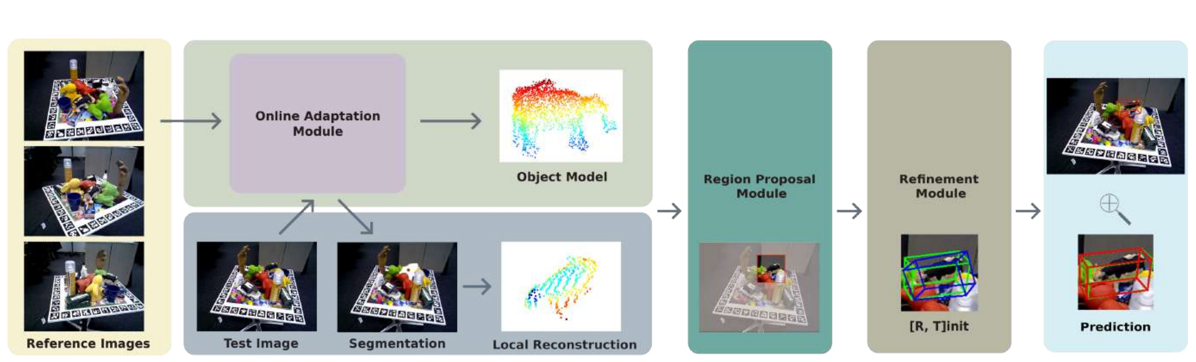
Task Design

Few-Shot 6D Novel Object Pose Estimation:



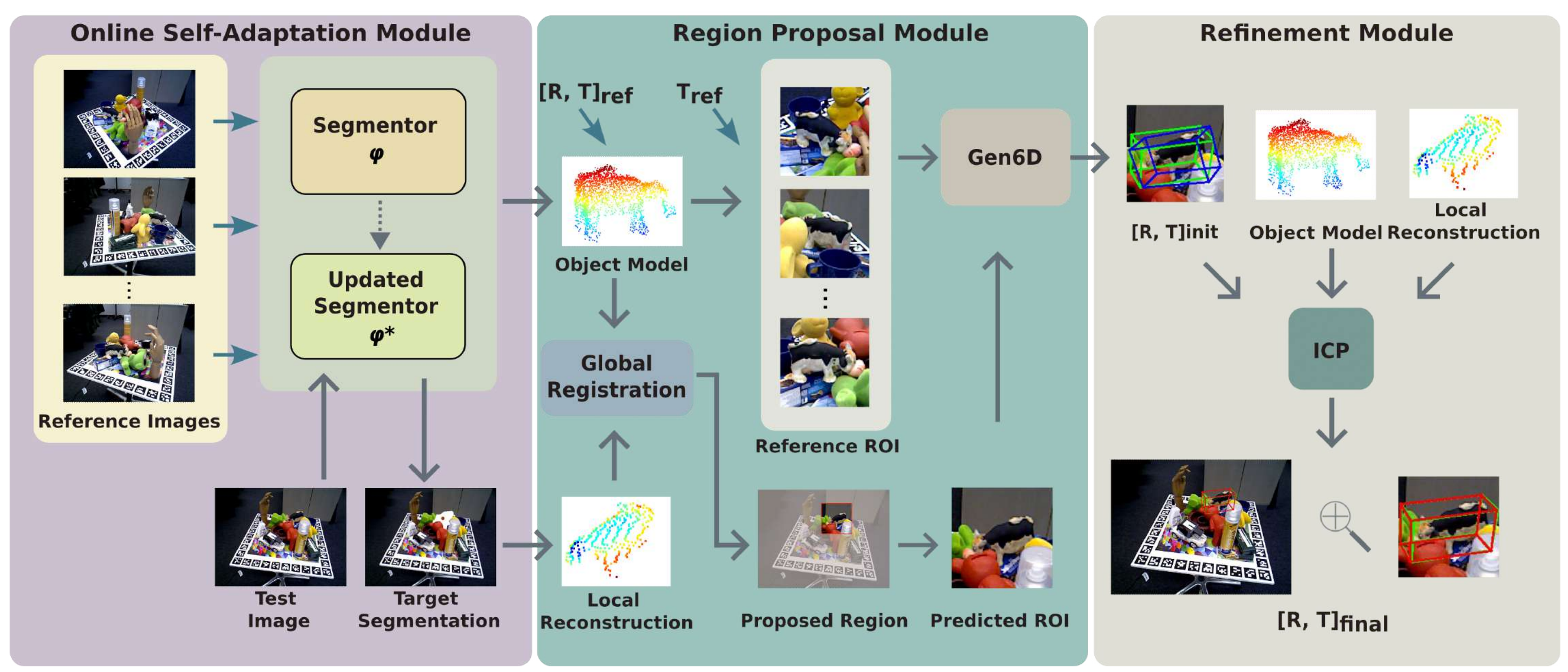
Predict 6D pose of a novel object in a new image (test image) from a few reference images with the known pose of the object (reference images).

Overview



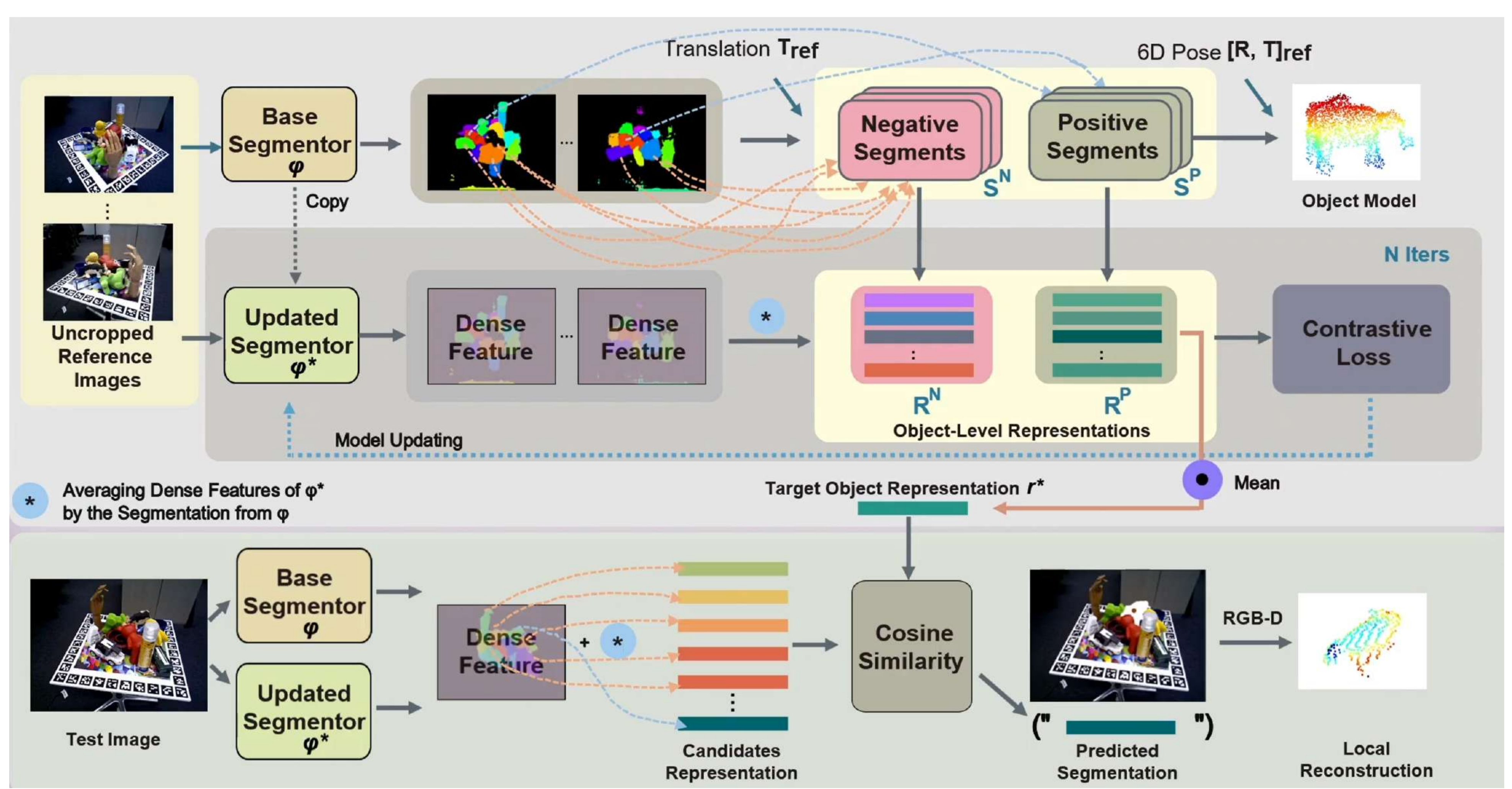
We present a generalizable and category-agnostic few-shot 6D object pose estimator using a small number of posed RGB-D images as references. Compared to existing methods, our approach provides robust and accurate predictions of novel objects against occlusions without requiring retraining or any object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

Onlie Adaptation Segmentation:



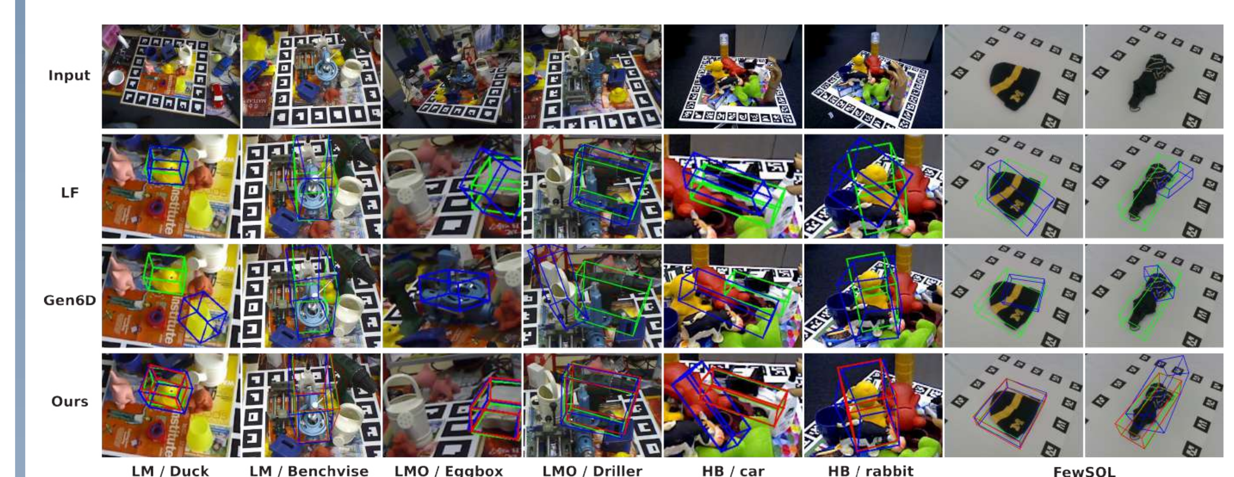
Online self-adaptation module. A pretrained segmentor φ is applied on reference images to predict segmentations. With the ground-truth translation of the target object in the reference images T_{ref}, the object center can be reprojected to the image. For each reference image, one segment is chosen as 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 $\varphi*$. A contrastive loss is then applied over the positive and negative object representations and updates $\varphi*$ iteratively. After adaptation, $\varphi*$ 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

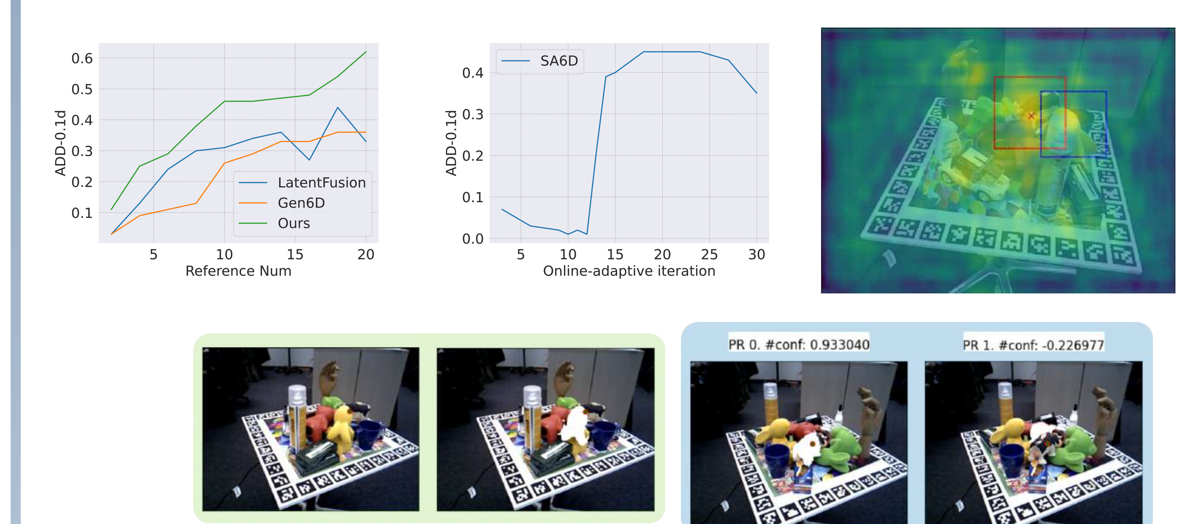
Contrastive loss:

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

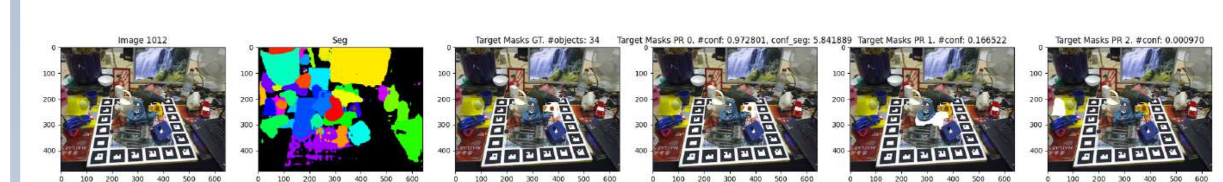
Qualitative results:



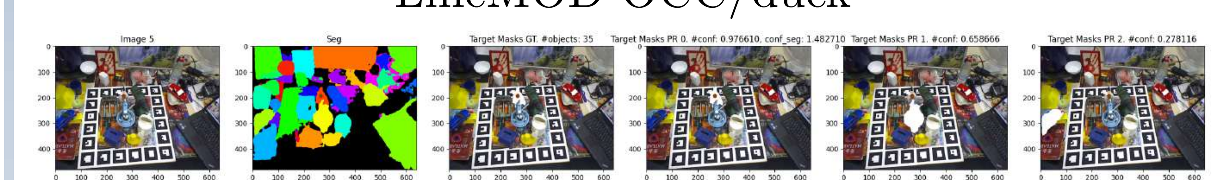
More analysis:



Robust prediction of target segmentation:



LineMOD-OCC/duck



LineMOD-OCC/glue

Quantitative results:

Method	GT-Mask	Ref. Num	LineMODlinemod-dataset							LineMOD-OCClinemod-occ							HomeBrewedDBHomeBrewedDB				
			eggbox	duck	benchvise	cam	cat	glue	avg.	driller	eggbox	duck	glue	ape	can	avg.	cow	flange	car	rabbit	avg.
Gen6DGen6D		20	0.63	0.30	0.45	0.29	0.25	0.26	0.36	0.09	0.02	0.07	0.03	0.12	0.21	0.09	0.36	0.15	0.15	0.52	0.30
SA6D (ICP only)		20	0.53	0.31	0.37	0.25	0.21	0.17	0.31	0.17	0.16	0.10	0.08	0.14	0.22	0.14	0.23	0.17	0.20	0.44	0.26
SA6D (wo/ RPM)		20	0.63	0.47	0.50	0.37	0.36	0.38	0.45	0.19	0.15	0.13	0.10	0.17	0.28	0.17	0.37	0.19	0.21	0.61	0.35
SA6D (wo/ RFM)		20	0.57	0.36	0.45	0.34	0.29	0.26	0.38	0.15	0.08	0.09	0.04	0.10	0.28	0.12	0.39	0.12	0.20	0.55	0.32
SA6D		20	**0.73**	**0.73**	**0.55**	**0.50**	**0.47**	**0.72**	**0.62**	**0.45**	**0.26**	**0.30**	**0.21**	**0.32**	**0.53**	**0.35**	**0.62**	**0.35**	**0.33**	**0.78**	**0.52**
Gen6D		64	0.74	0.40	**0.73**	0.65	0.65	0.53	0.62	0.27	0.09	0.23	0.03	0.11	0.50	0.21	0.38	0.06	0.49	**0.78**	0.43
SA6D		64	**0.80**	**0.84**	**0.73**	**0.80**	**0.84**	**0.75**	**0.79**	**0.44**	**0.41**	**0.38**	**0.31**	**0.33**	**0.66**	**0.42**	**0.72**	**0.49**	**0.72**	0.69	**0.66**
LFLatentFusion		20	0.61	**0.61**	0.68	0.65	**0.72**	**0.78**	0.67	0.28	0.01	0.00	0.18	**0.45**	0.17	0.18	0.33	0.00	0.00	0.16	0.12
SA6D (wo/ RFM)		20	0.56	0.32	0.54	0.30	0.26	0.29	0.38	0.10	0.06	0.08	0.04	0.14	0.24	0.11	0.41	0.13	0.15	0.54	0.31
SA6D		20	**0.68**	0.58	**0.80**	**0.73**	**0.72**	**0.78**	**0.72**	**0.33**	**0.26**	**0.29**	**0.30**	0.19	**0.45**	**0.30**	**0.58**	**0.17**	**0.44**	**0.76**	**0.49**

Evaluation of ADD-0.1d

Method	ADD-0.1d	ADD-0.3d	ADDs-0.1d	ADDs-0.3d
LFLatentFusion	0.1162	0.1738	0.1299	0.1907
Gen6DGen6D	0.3571	0.6399	0.6399	0.7530
SA6D (wo/ RFM)	0.4018	0.7292	0.6964	**0.8780**
SA6D	**0.5595**	**0.7887**	**0.8393**	**0.8780**

Evaluation on FewSOL

Method	$IOU_{0.5}$	5°2cm	5°5cm	10°5cm
CASScass	0.01	0.0	0.0	0.0
Shape-Priorshapeprior	0.33	0.03	0.04	0.14
DualPoseNetdualposenet	0.70	0.18	0.23	0.37
RePoNetWild6D	**0.71**	0.30	0.34	**0.43**
SA6D	0.65	**0.37**	**0.40**	0.42

Evaluation on Wild6D