Supplementary Materials

Figure 1. Visualization of given bounding-boxes (red box) and generated pseudo labels (green region) on **LINEMOD**.

1. Bounding-box Visualization

As shown in Fig. 1, we do not assume the given 2D bounding-boxes tightly cover object regions, which further relaxes the labeling requirements. For generating loose bounding-boxes, we first extract tight bounding-boxes from the ground-truth object masks. Then we randomly expand the size of tight bounding-boxes by $0\% \sim 15\%$ of the width and height.

2. Iterative Weakly-supervised Segmentation

In the weakly-supervised segmentation step, we iterate the pseudo-label generation and fine-tuning process by $T = 5$ steps. As seen in Table 1, the quality of pseudo segmentation labels progressively improves as the iteration step increases. The IoU score always converges after $T = 3$.

3. More Results

Qualitative Results on LINEMOD. Fig. 2 demonstrates our qualitative results on LINEMOD. The ensembled predictions are always better than the results predicted by a single scale (*i.e.*, either un-normalized scale or normalized scale).

Fine-tuning on HomebrewedDB. Following Self6D, we provide a self-supervised result by using 15% of real data from HomebrewedDB [2] without pose labels. As shown in Table 2, the self-supervised fine-tuning significantly improves the performance of our RGB based DSC-PoseNet on HomebrewedDB, marked as Ours^{FT}. Moreover, Ours^{FT} also outperforms RGBD based Self6D. Fig. 3 further demonstrates our qualitative results on HomebrewedDB.

Evaluation on BOP [1] protocols. We also use the Average Recall (AR) used by BOP to evaluate our performance. The AR scores of our DSC-PoseNet on LINEMOD, OCC-LINEMOD and HomebrewedDB are 71.0%, 45.0% and 67.4%, respectively. For comparison, DPOD (synthetic) [5] achieves an inferior result 16.9% on OCC-LINEMOD without using depth and pose annotations.

T	Ape	Bvise	Cam	Can	Cat	Driller	Duck	Eggbox	Glue	Holep	Iron	Lamp	Phone	Mean
1	95.2	92.0	86.2	93.3	92.1	92.9	94.7	95.6	86.9	89.9	**90.5**	72.8	83.0	89.6
2	95.2	92.1	87.3	93.7	92.3	93.5	**95.0**	95.8	87.2	90.4	**90.5**	74.4	83.9	90.1
3	95.4	92.4	88.3	93.8	**92.7**	93.7	94.8	**95.9**	87.4	90.5	**90.5**	75.2	84.2	90.4
4	95.6	**92.5**	88.8	93.8	**92.7**	**93.9**	94.5	**95.9**	**87.5**	**90.6**	90.4	75.8	84.9	90.5
5 (Ours)	**95.7**	**92.5**	**89.4**	**93.9**	92.6	**93.9**	94.2	**95.9**	**87.5**	90.5	90.4	**76.2**	**85.2**	**90.6**

Table 1. The impact of different iteration step numbers (T) on our weakly-supervised segmentation on **LINEMOD**. We evaluate the Intersection over Union (IoU) scores between pseudo-labeled masks and ground-truth masks on the training split.

Figure 2. Qualitative results on **LINEMOD** dataset. *Green*: the ground truth pose. *Red*: un-normalized scale prediction. *Yellow*: normalized scale prediction. *Blue*: ensembled prediction by averaging the keypoints predicted at both the scales.

Method	RGB-based				RGBD
	DPOD [5]	SSD+Ref. [3]	**Ours**	**Ours**^{FT}	Self6D [4]
Bvise	52.9	**82.0**	72.9	78.2	72.1
Drill	37.8	22.9	40.6	**76.2**	65.1
Phone	7.3	24.9	18.5	**42.9**	41.8
Mean	32.7	43.3	44.0	**65.8**	59.7

Table 2. Comparisons with state-of-the-art on **HomebrewedDB** dataset. Ours^{FT}: self-supervised fine-tuning by using 15% of real data from HomebrewedDB [2]

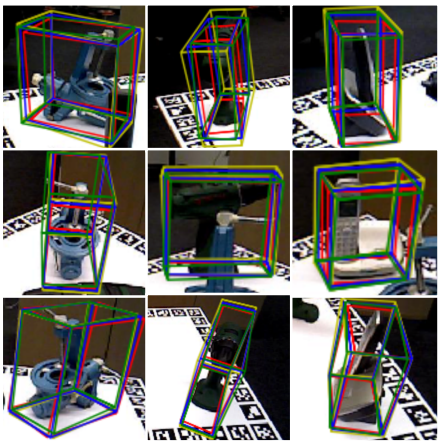

Figure 3. Qualitative results on **HomebrewedDB** dataset.

References

[1] Tomas Hodan, Frank Michel, Eric Brachmann, Wadim Kehl, Anders GlentBuch, Dirk Kraft, Bertram Drost, Joel Vidal, Stephan Ihrke, Xenophon Zabulis, et al. Bop: Benchmark for 6d object pose estimation. In *ECCV*, pages 19–34, 2018. 1

[2] Roman Kaskman, Sergey Zakharov, Ivan Shugurov, and Slo-

bodan Ilic. Homebreweddb: Rgb-d dataset for 6d pose estimation of 3d objects. In *ICCV Workshops*, 2019. 1, 2

[3] Wadim Kehl, Fabian Manhardt, Federico Tombari, Slobodan Ilic, and Nassir Navab. Ssd-6d: Making rgb-based 3d detection and 6d pose estimation great again. In *ICCV*, pages 1521–1529, 2017. 2

[4] Gu Wang, Fabian Manhardt, Jianzhun Shao, Xiangyang Ji, Nassir Navab, and Federico Tombari. Self6d: Self-supervised monocular 6d object pose estimation. In *ECCV*, 2020. 2

[5] Sergey Zakharov, Ivan Shugurov, and Slobodan Ilic. Dpod: 6d pose object detector and refiner. In *ICCV*, 2019. 1, 2