Supplementary material of category-level 6D pose estimation using geometry-guided instance-aware prior and multi-stage reconstruction

Tong Nie, Jie Ma*, Yuehua Zhao, Ziming Fan, Junjie Wen, Mengxuan Sun

I. Detail of experiments on CAMERA and Real datasets

We trained our method on the CAMERA dataset and evaluated it on the CAMERA25 dataset. The evaluation results are compared with existing SOTA methods, and the comparison results are listed in Table I. The evaluation results demonstrate that our method is SOTA at 5°2cm, 10°2cm and 10°5cm, and achieves higher performance on all rotation translation metrics compared to the baseline method.

The mAP results of GIPN, MNRN, OURS and baseline (SGPA) trained on the CAMERA and REAL datasets are compared, as shown in Fig 1. The red, blue, orange and green lines plot OURS, SGPA, GIPN and MNRN mAP, respectively. The rotation mAP is more distinguishable than the translation and more influences the final pose mAP. The Sub1 (relatively strict metric part) shows that the green line (MNRN) is slightly higher than the blue line (SGPA) and the Sub2 (relatively relaxed metric part) shows that the green line is lower than the blue one. The Sub3 (relatively relaxed metric part) illustrates that the orange line (GIPN) is higher than the blue one and the Sub4 (relatively strict metric part) plots that the orange line is lower than the blue line. Rotation15 shows the red line (OURS) is close to the orange line in the relatively relaxed metric part and close to the green line in the relatively strict metric part. The experimental results demonstrate that the GIPN improves the performance on relaxed evaluation metrics, while MNRN improves the performance on the strict metrics. The GIPN+MNRN enhance the performance on the strict metrics after improvement on relaxed metrics, an overall improvement, which makes our method surpass the baseline in all metrics and beyond some SOTA methods in some metrics.

In addition, we compare the performance of a competitive method GPV, the baseline and our proposed method on the camera category. As shown in Table II, our method significantly improves the accuracy of baseline on all listed metrics, and outperforms GPV on evaluation metrics 25°5cm, 30°5cm, 35°5cm, 40°5cm and IoU50.

II. Detail of CAMERA-Light and REAL-Light datasets

To the best of our knowledge, until we completed the manuscript, there were few datasets for supervised learning of classification-level 6D pose estimation, only CAMERA and REAL were presented by NOCS, where CAMERA is a synthesized dataset and only REAL is a real-world dataset as shown in Fig 2 below. To more clearly investigate the effect of RGB influenced by varying light on 6D pose estimation, we present the CAMERA-Light and REAL-Light datasets including the CAMERA25-Light and REAL275-Light test datasets. The *-Light datasets synthesize RGB data with virtual illumination, as shown in Fig 3. The CAMERA-Light dataset contains 300K depth images and 300K synthesized virtual light RGB images, and REAL-Light contains 4300 and 2750 synthesized images for training and testing, respectively.

TABLE I
Comparison of our method with state-of-the-art methods on CAMERA25 benchmark.

Method	Data Setting	3D50	3D75	5°2cm	5°5cm	10°5cm
NOCS[17]	RGB-D	83.9	69.5	32.3	40.9	64.6
CASS[23]	RGB-D	-	-	-	-	-
SPD[18]	RGB-D	93.2	83.1	54.3	59.0	81.5
CR-Net[20]	RGB-D	**93.8**	88.0	72.0	**76.4**	87.7
SGPA[19]	RGB-D	93.2	**88.1**	70.7	74.5	88.4
DPN[25]	RGB-D	92.4	86.4	64.7	70.7	84.7
DO-Net[26]	D	-	-	-	-	-
FS-Net[36]	D	-	-	-	-	-
GPV[24]	D	-	-	-	-	-
OURS	D	92.1	87.7	**72.2**	76.3	**88.5**

III. Detail of experiments on *-Light datasets

We retrained the baseline method on the CAMERA and REAL-Light dataset in a 3:1 ratio as SGPA-Light and on the CAMERA-Light and REAL-Light dataset as SGPA-Light2. We also evaluated the original SGPA weights on the benchmark REAL275-Light as SGPA-275Light. The performance comparison of the *-Light and original datasets is listed in Table III. Compared to the baseline, SGPA-light, SGPA-light2 and SGPA-275light all degrade performance on all rotation translation metrics. In particular, SGPA-light2 decreased by 5.1%, 4.5% and 4.5% on 5°2cm, 5°5cm and 10°5cm. When the baseline method is affected by varying light, our methods GIPN and OURS, which do not rely on RGB data, outperform SGPA-light, SGPA-light2 and SGPA-275light in all the performance metrics listed. MNRN can refine the estimated 6D pose. Our proposed method combines the advantages of GIPN and MNRN, and improves on both relaxed and strict rotation translation performance and 3DIoU metrics. The experimental results demonstrate that the varying light will affect the results

This work is supported by Hebei Natural Science Foundation (Grant number [F2020202045]).

All author is with the School of Electronics and Information Engineering, Hebei University of Technology, Xiping Road, Tianjin, 300401, Tianjin, China.

*Corresponding author E-mail addresses: jma@hebut.edu.cn

The code is at https://github.com/nicu233/GIPMR

of classification-level 6D pose estimation, and our proposed method can avoid this impact and improve overall performance.

A comparison of the quality of SGPA before and after the effect of light is shown in Fig 4. The varying light leads to degrading category-level 6D pose prediction accuracy.

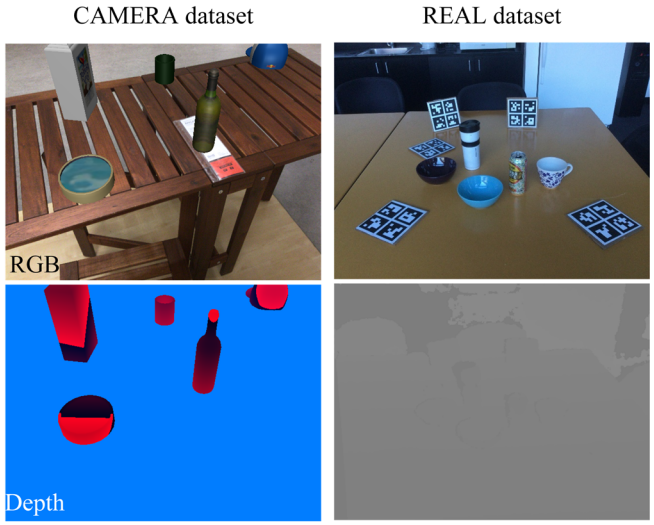


Figure 2. Visual comparison of CAMERA (left column) and REAL (right column) datasets

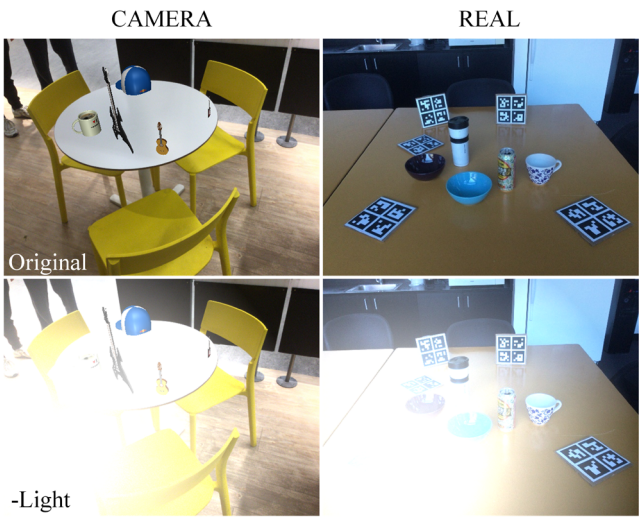


Figure 3. Visual comparison of RGB image of original and -Light datasets

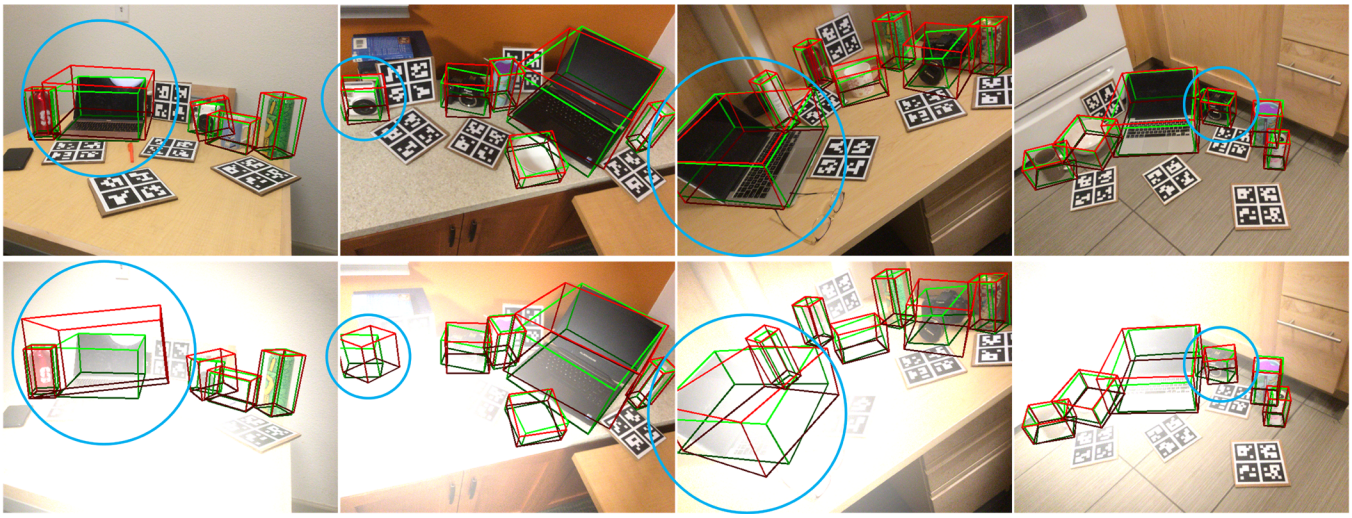


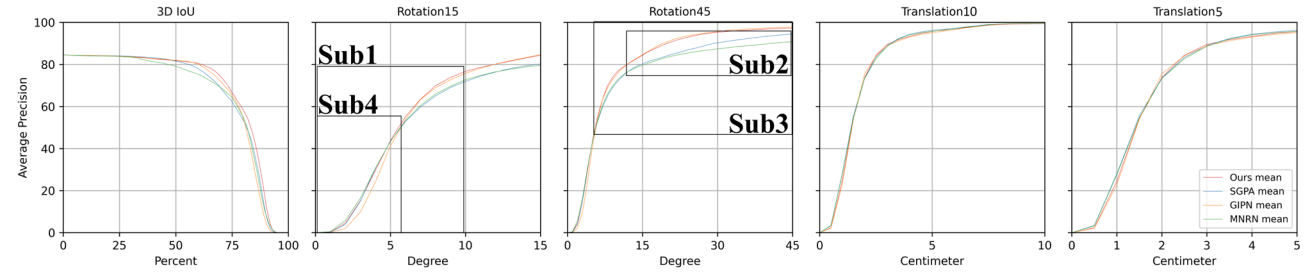
Figure 4. Visual qualitative comparison of estimated 6D object pose and size by SGPA on REAL275 (top row) and REAL275-Light dataset (bottom row). Green 3D bounding box is ground truth of 6D pose and size, and the red one is estimation result.

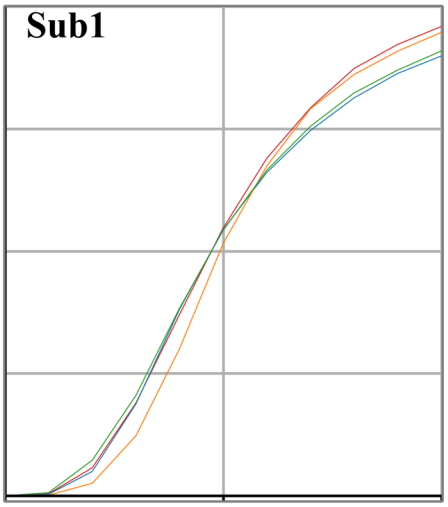
TABLE III
ABLATION STUDIES OF THE PROPOSED METHOD ON REAL275 BENCHMARK.

	REAL-Light	CAMERA-Light	REAL275-Light	3D50	3D75	5°2cm	5°5cm	10°5cm
SGPA	-	-	-	80.1	61.9	35.9	39.6	70.7
SGPA-Light	√	-	-	80.1	63.8	31.4	35.9	69.4
SGPA-Light2	√	√	-	80.8	65.1	30.8	35.1	66.2
SGPA-275Light	-	-	√	81.4	60.8	34.1	38.0	68.5
MNRN	-	-	-	79.1	63.9	**36.7**	**41.3**	69.7
GIPN	- /√	- /√	- /√	**82.1**	65.4	34.3	39.1	73.4
OURS	- /√	- /√	- /√	81.8	**66.8**	36.2	41.1	**74.2**

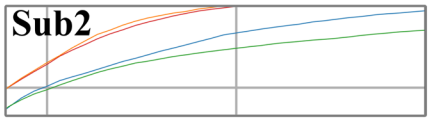
TABLE II
COMPARISON OF PERFORMANCE OF GPV, SGPA AND OURS ON CAMERA CATEGORY

Method	3D50	3D75	5°2cm	5°5cm	10°5cm	15°5cm	20°5cm	25°5cm	30°5cm	35°5cm	40°5cm
GPV	80.9	**53.1**	**0.4**	**0.5**	**11.0**	**13.5**	**39.**7	**57.3**	66.0	70.2	72.8
SGPA	81.7	13.9	0.0	0.0	0.4	0.6	5.1	17.4	32.5	46.1	54.4
OURS	**84.2**	37.6	0.2	0.2	3.9	5.2	26.2	54.3	**70.3**	**78.5**	**82.0**

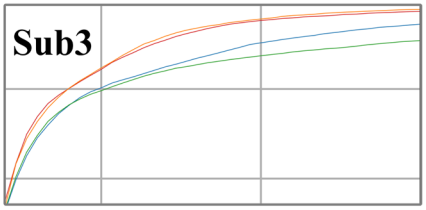




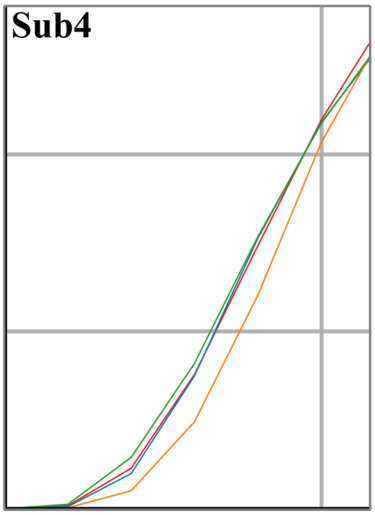
Green line is higher than the blue line shows that MNRN is better than SGPA on the relatively strict rotation metric.



Blue line is higher than the green line demonstrates that SGPA is better than MNRN on the relatively relaxed rotation metric.



Orange line is higher than the blue line shows that MNRN is More accurate than SGPA on the relatively relaxed rotation metric.



Orange line is lower than the blue line demonstrates that GIPN is worse than SGPA on the relatively strict rotation metric.

Figure 1. Visual comparison of mAP results of GIPN, MNRN, OURS and baseline (SGPA) trained on the REAL dataset