6D Object Pose Estimation in Cluttered Scenes from RGB Images

Xiaolong Yang[1,2] **Xiaohong Jia**[1,2]

Yuan Liang[3] **Lubin Fan**[3]

[1]Academy of Mathematics and Systems Science,
Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]Alibaba Group

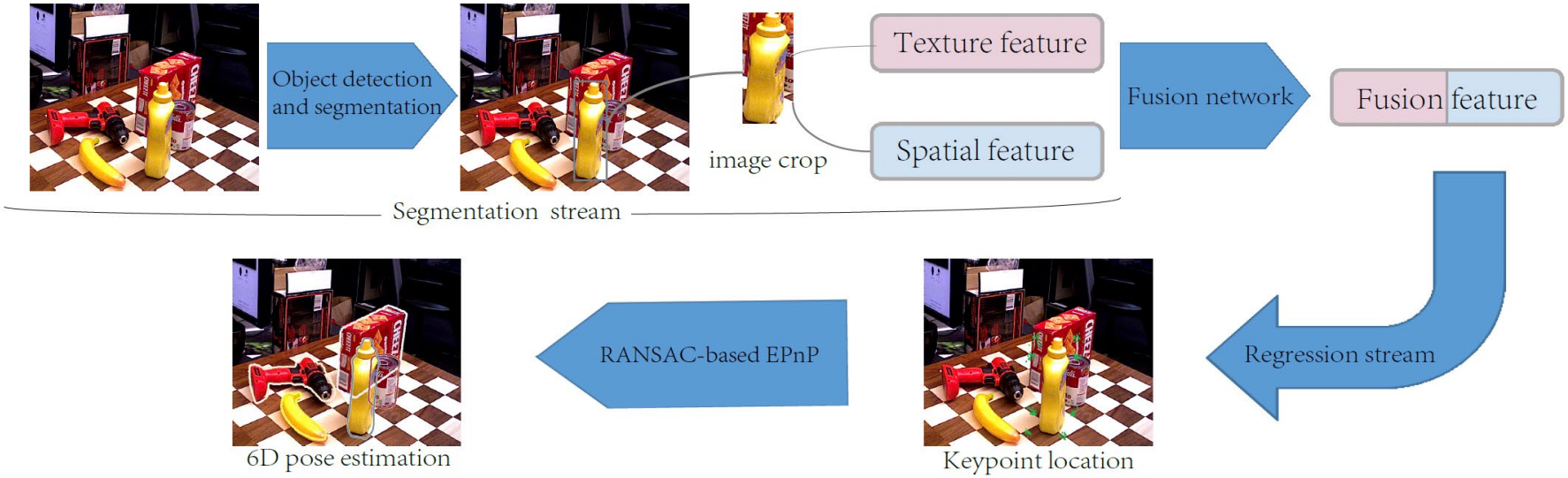
Research Problem

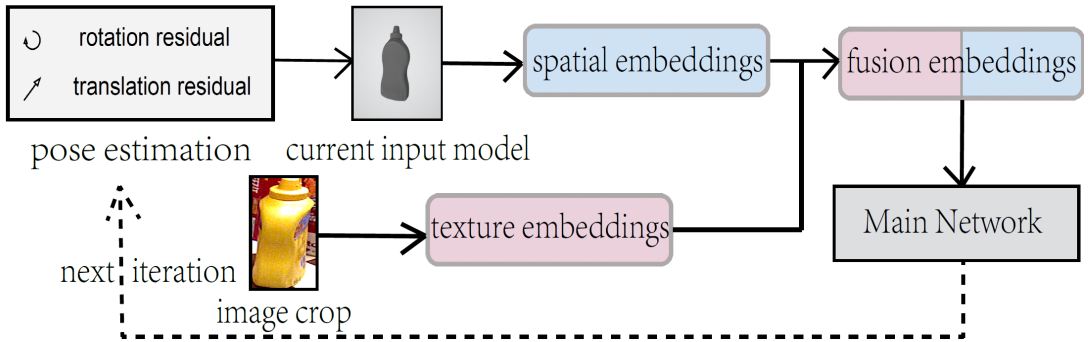
Higher requirements are put forward for the various needs of 6D object pose estimation:

- properly handle objects with irregular shapes , low-resolution textures and different materials;
- be robust to heavy occlusion, changing lighting in various environments, and potential technical noise;
- achieve real-time speed as possible.

An example of pose estimation with occluded known objects.

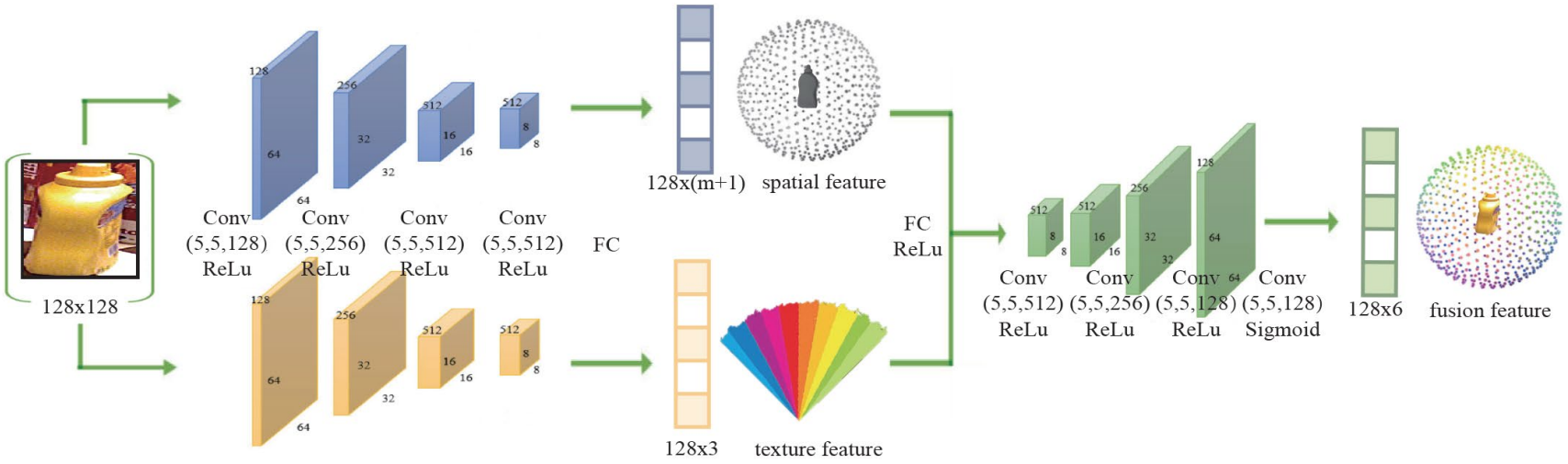
Proposed Method





The workflow of our pipeline

Proposed Method



Feature extraction in fusion network

Feature regression on grid mask

Results

Visualization results on the Occluded-LineMOD dataset

Results

Visualization results on the YCB-Video dataset

Results

Table 1. Ablation study of 6D Pose (ADD-0.1d) on Occluded-LINEMOD. Bold-name objects are symmetric. SF: spatial feature, TF:texture feature, IR:iterative refinement.

Methods	SF	TF	IR	Ape	Can	Cat	Driller	Duck	**Eggbox**	**Glue**	Holepun	Average
XYZ only	✓			14.9	60.1	16.5	48.2	25.1	35.6	44.4	35.9	35.1
XYZ+Refine	✓		✓	17.5	63.1	16.6	48.3	25.4	**35.8**	44.6	38.6	36.2
Fusion	✓	✓		37.4	68.6	26.0	48.5	25.0	30.3	45.2	45.1	40.8
Fusion+Refine	✓	✓	✓	**39.1**	**69.8**	**26.9**	**49.0**	**25.6**	31.9	**46.5**	**53.1**	**42.7**

Table 2. Quantitative Evaluation on Occluded-LINEMOD dataset. Symmetric objects are bold-name. The best and second best results are shown in red and blue, respectively.

Object	PoseCNN	Tekin	BB8	Pix2Pose	Heatmap	Seg-Driven	PVnet	CDPN	Ours
Ape	9.6	7.0	28.5	8.3	16.5	12.1	15.81	28.92	39.1
Can	45.2	1.2	11.2	12.1	42.5	39.9	63.31	55.98	69.8
Cat	0.9	3.6	9.6	9.3	2.8	8.2	16.68	13.24	26.9
Driller	41.4	1.4	0.2	10.9	47.1	45.2	65.65	51.37	49.0
Duck	19.6	5.1	6.8	6.3	11.0	17.2	25.24	22.97	25.6
Eggbox	22.0	9.6	4.0	13.8	24.7	22.1	50.17	35.98	31.9
Glue	38.5	6.5	4.7	11.3	39.5	35.8	49.62	39.68	46.5
Holepun	22.1	8.3	8.1	10.7	21.9	36	39.67	51.06	53.1
Average	24.9	5.3	9.1	10.3	25.8	27.0	40.77	37.4	42.7

Table 3. Efficiency Comparisons on Occluded-LINEMOD dataset. All methods run on the same environment.

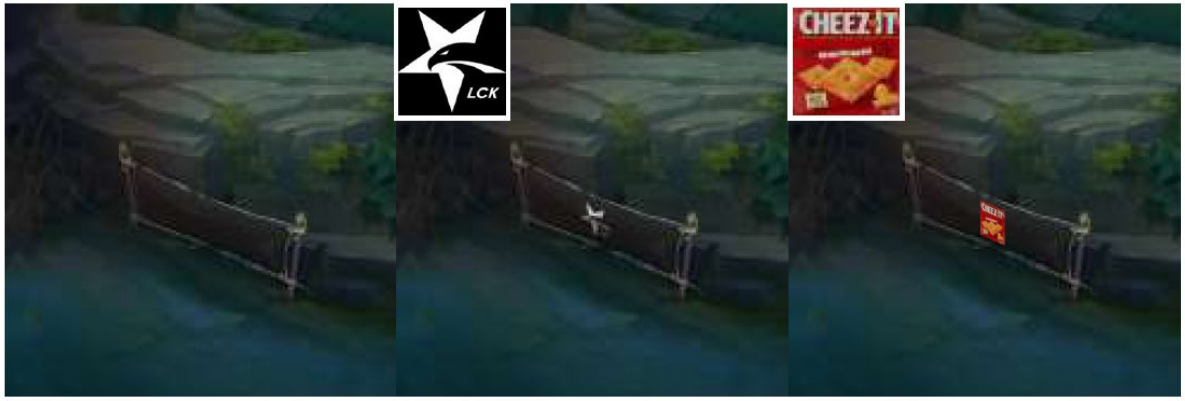
Method	PoseCNN	Tekin	BB8	Pix2Pose	Heatmap	Seg-Driven	PVnet	Ours
FPS	4	40	3	-	4	22	8	30

Table 4. Quantitative Evaluation of 6D Pose (ADD-0.1d) on YCB. Objects with bold name are symmetric. The best and second best results are shown in red and blue, respectively.

Object	PoseCNN	SilhoNet	Heatmap	Seg-Driven	CDPN	Ours
master_chef_can	3.6	23.8	32.9	33.0	35.5	37.1
cracker_box	25.1	20.1	62.6	44.6	45.6	45.8
sugar_box	40.3	48.5	44.5	75.6	71.5	69.4
tomato_soup_can	25.5	25.1	31.1	40.8	49.6	52.3
mustard_bottle	61.9	60.8	42.0	70.6	74.8	78.2
tuna_fish_can	11.4	25.3	6.8	18.1	25.2	24.1
pudding_box	14.5	17.0	58.4	12.2	48.4	32.6
gelatin_box	12.1	26.2	42.5	59.4	57.8	46.9
potted_meat_can	18.9	22.2	37.7	33.3	37.6	40.1
banana	30.3	32.8	16.8	16.6	25.1	27.5
pitcher_base	15.6	25.9	57.2	90.0	80.8	82.0
bleach_cleanser	21.2	20.8	65.3	70.9	81.9	82.1
bowl	12.1	22.5	25.6	30.5	22.6	23.0
mug	5.2	12.3	11.6	40.7	51.2	55.6
power_drill	29.9	26.0	46.1	63.5	57.4	59.8
wood_block	10.7	18.7	34.3	27.7	25.1	27.5
scissors	2.2	3.4	0.1	17.1	11.6	12.4
large_marker	3.4	3.0	3.2	4.8	4.5	6.1
large_clamp	28.5	29.7	10.8	25.6	25.9	27.7
extra_large_clamp	19.6	20.4	29.6	8.8	13.4	16.4
foam_brick	54.5	42.0	51.7	34.7	40.7	44.8
Average	21.3	25.1	33.6	39.0	42.2	42.4

Quantitative results exhibition

Applications



(a) advertising space (b) officical logo implant (c) our algorithm

Advertisement replacement

(d) advertising space (e) recommended decoration (f) our algorithm

Decoration recommendation

Conclusions

- An efficient and high-quality two-stream architecture for cluttered or texture-less object pose estimation.
- A feature embedding method to connect spatial information and appearance texture.
- A strategy based on RANSAC E-PnP and an end-to-end refinement to enhance performance.