Supplementary Material for "ObjectNet3D: A Large Scale Database for 3D Object Recognition"

Yu Xiang, Wonhui Kim, Wei Chen, Jingwei Ji, Christopher Choy, Hao Su, Roozbeh Mottaghi, Leonidas Guibas, and Silvio Savarese

Stanford University

1 Statistics

We present detailed statistics of our dataset in Table 1. For each category, we list the number of objects in images and the number of 3D shapes in the dataset. For 42 categories among the 100 categories, we have additional 3D shapes from the ShapeNet repository [1].

2 Viewpoint Distributions

In Fig. 1 - Fig. 4, we visualize the viewpoint distributions of all 100 categories in our database. From these distributions, we can see which viewpoints are more likely for different objects in the real world.

3 Detection and Viewpoint Estimation Examples

In Fig. 5 - Fig. 14, we show detection and viewpoint estimation results from our baseline method. Please refer to the main paper for detailed description of the method, where we add a viewpoint regression branch to the Fast R-CNN framework [2].

4 3D Shape Retrieval Examples

In Fig. 15 - Fig. 17, we show 3D shape retrieval results using our learned feature embedding from [3].

5 Annotation Demo

Please see the attached video for a demo of the 2D-3D alignment in creating annotations in our database.

References

1. : Shapenet. `http://shapenet.cs.stanford.edu/`
2. Girshick, R.: Fast r-cnn. In: ICCV. (2015) 1440–1448
3. Song, H.O., Xiang, Y., Jegelka, S., Savarese, S.: Deep metric learning via lifted structured feature embedding. In: CVPR. (2016)

Category	# objects	# 3D shapes	Category	# objects	# 3D shapes
aeroplane	2062	8 + 4046	ashtray	1112	10
backpack	1152	16	basket	1887	15
bed	2023	10 + 254	bench	1814	7 + 1816
bicycle	1802	7 + 59	blackboard	1055	11
boat	2944	6 + 3076	bookshelf	1226	8 + 466
bottle	5684	8 + 498	bucket	1702	4
bus	1304	6 + 939	cabinet	4988	15 + 1572
calculator	995	5	camera	1352	11 + 113
can	2149	6 + 108	cap	3036	15 + 56
car	12886	10 + 6591	cellphone	1631	11 + 527
chair	14042	10 + 6778	clock	1146	9 + 655
coffee maker	1200	7	comb	1189	9
computer	1272	11	cup	6022	10
desk lamp	2221	8	dining table	2818	6
dishwasher	1067	4 + 96	door	2845	14
eraser	1763	15	eyeglasses	2504	11
fan	1336	13	faucet	1696	11 + 744
filing cabinet	982	8 + 298	fire extinguisher	811	9
fish tank	902	6	flashlight	1150	6
fork	1912	9	guitar	1027	5 + 797
hair dryer	872	4	hammer	809	6
headphone	1163	5 + 73	helmet	2445	8 + 162
iron	559	5	jar	2129	7 + 597
kettle	2279	7	key	1064	13
keyboard	3106	12	knife	2151	8 + 424
laptop	1770	5 + 460	lighter	1177	6
mailbox	1384	8 + 94	microphone	963	7 + 67
microwave	1150	6 + 152	motorbike	1486	5 + 337
mouse	2079	5	paintbrush	1399	6
pan	791	5	pen	1738	4
pencil	1683	4	piano	907	5 + 293
pillow	5593	6 + 96	plate	4766	6
pot	1573	7	printer	1150	6 + 167
racket	1032	3	refrigerator	1247	12
remote control	1298	7 + 67	rifle	1146	8 + 2373
road pole	2832	16	satellite dish	351	5
scissors	1370	6	screwdriver	1628	7
shoe	8323	10	shovel	1063	5
sign	1991	11	skate	701	2
skateboard	610	2 + 152	slipper	1188	6
sofa	2412	6 + 3173	speaker	3226	9 + 1618
spoon	2708	7	stapler	999	5
stove	2038	6 + 218	suitcase	1139	9
teapot	1117	7	telephone	1067	9 + 1052
toaster	859	10	toilet	882	7
toothbrush	1126	5	train	1392	4 + 389
trash bin	1811	10 + 343	trophy	719	11
tub	835	9	tvmonitor	3063	4 + 1095
vending machine	1281	11	washing machine	1022	5
watch	1429	9	wheelchair	1054	5

Table 1. The number of objects in images and the number of 3D shapes for each of the 100 object categories in our database. For the 3D shapes, 42 categories have 3D shapes from the ShapeNet repository [1].

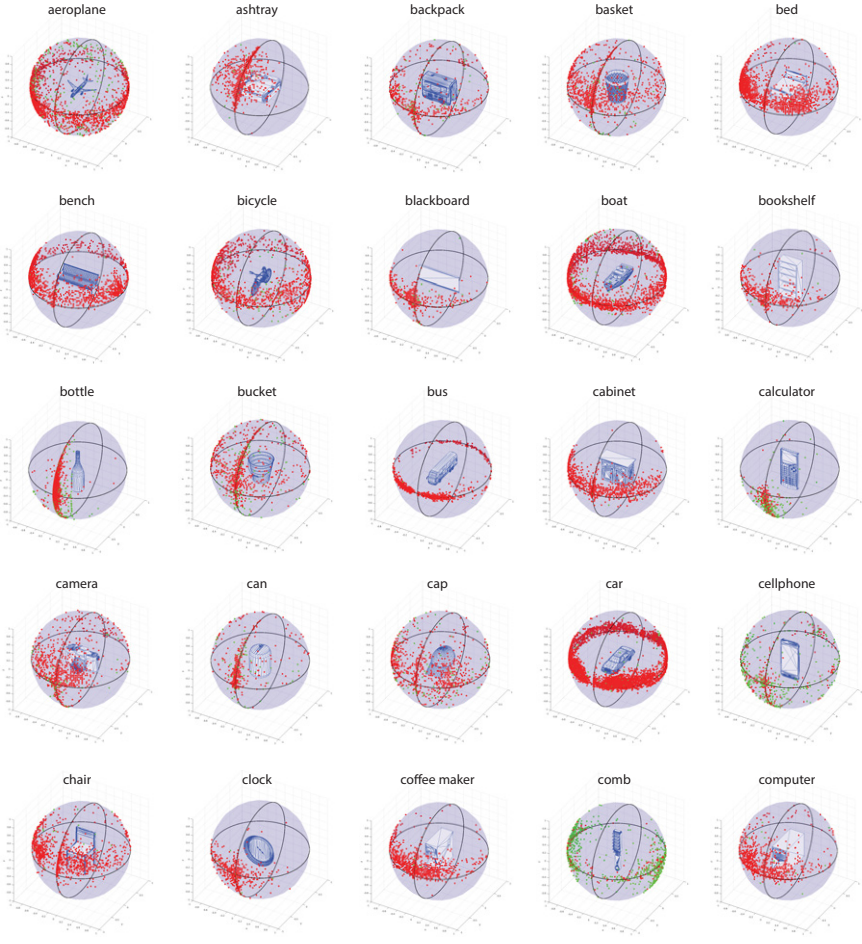

Fig. 1. Viewpoint distributions of different categories in our database. We visualize the camera position as a point on the unit sphere (red points: in-plane rotation $< 15°$; green points: in-plane rotation $> 15°$).

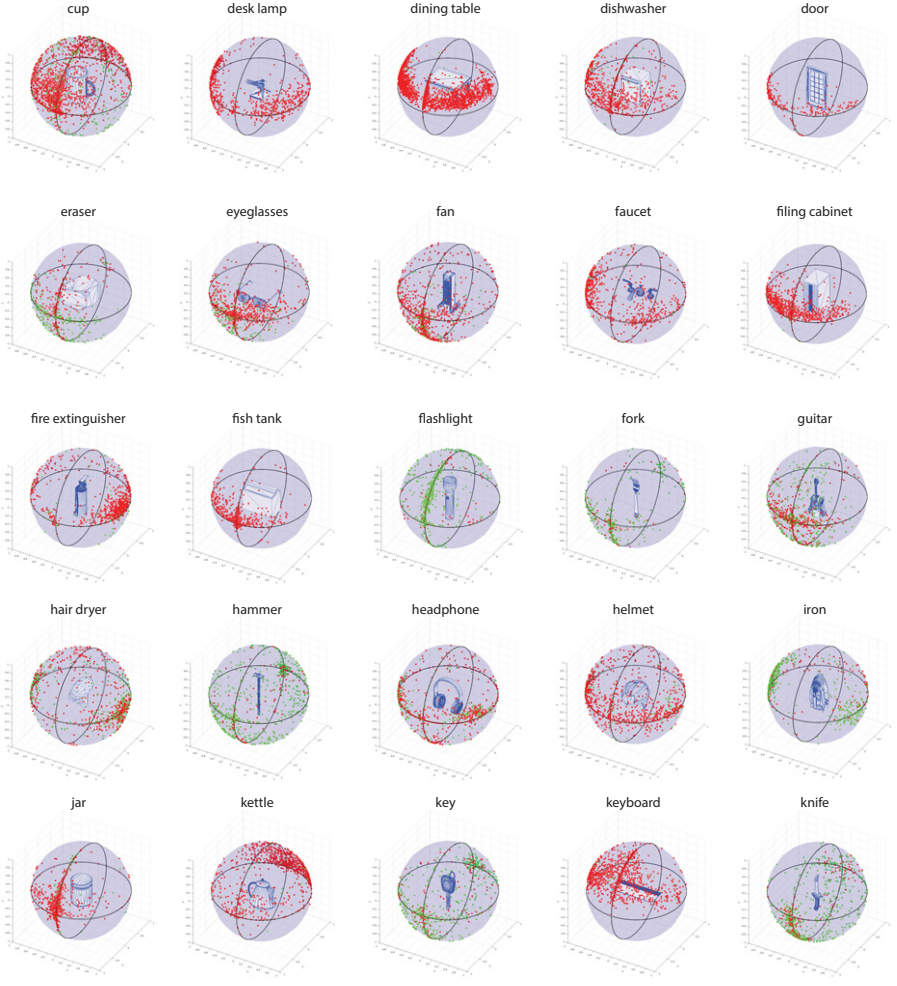

Fig. 2. Viewpoint distributions of different categories in our database. We visualize the camera position as a point on the unit sphere (red points: in-plane rotation $< 15°$; green points: in-plane rotation $> 15°$).

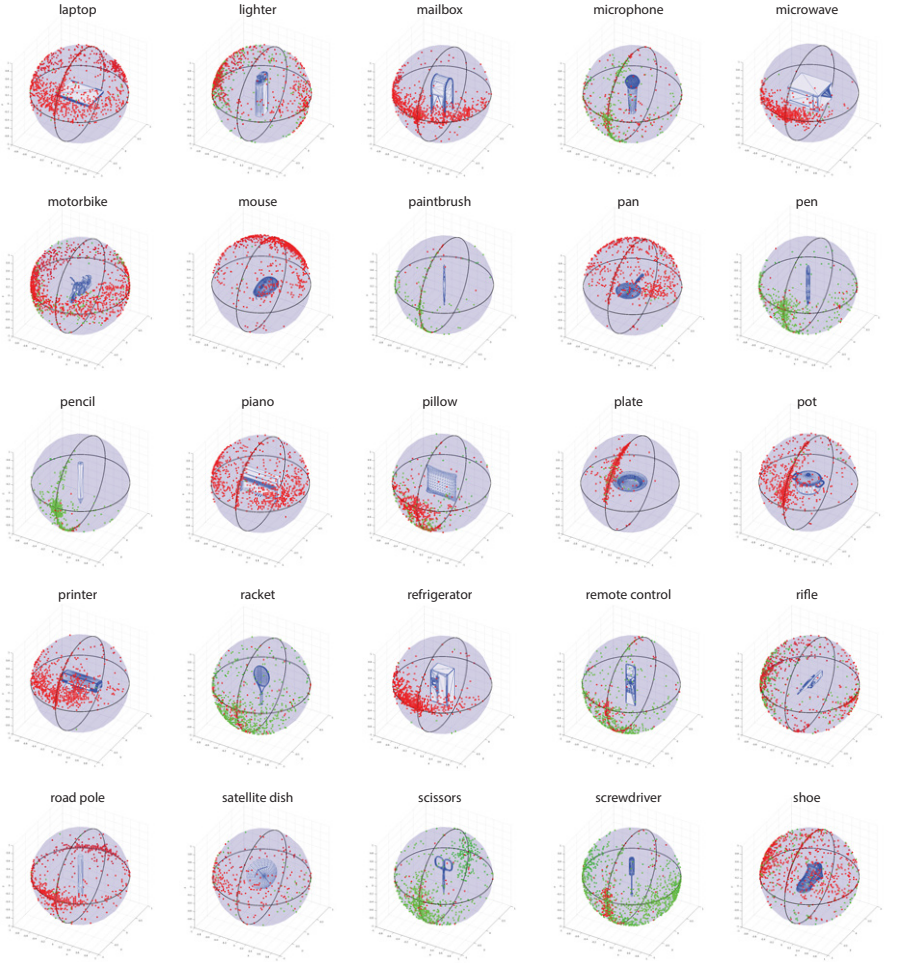

Fig. 3. Viewpoint distributions of different categories in our database. We visualize the camera position as a point on the unit sphere (red points: in-plane rotation $< 15°$; green points: in-plane rotation $> 15°$).

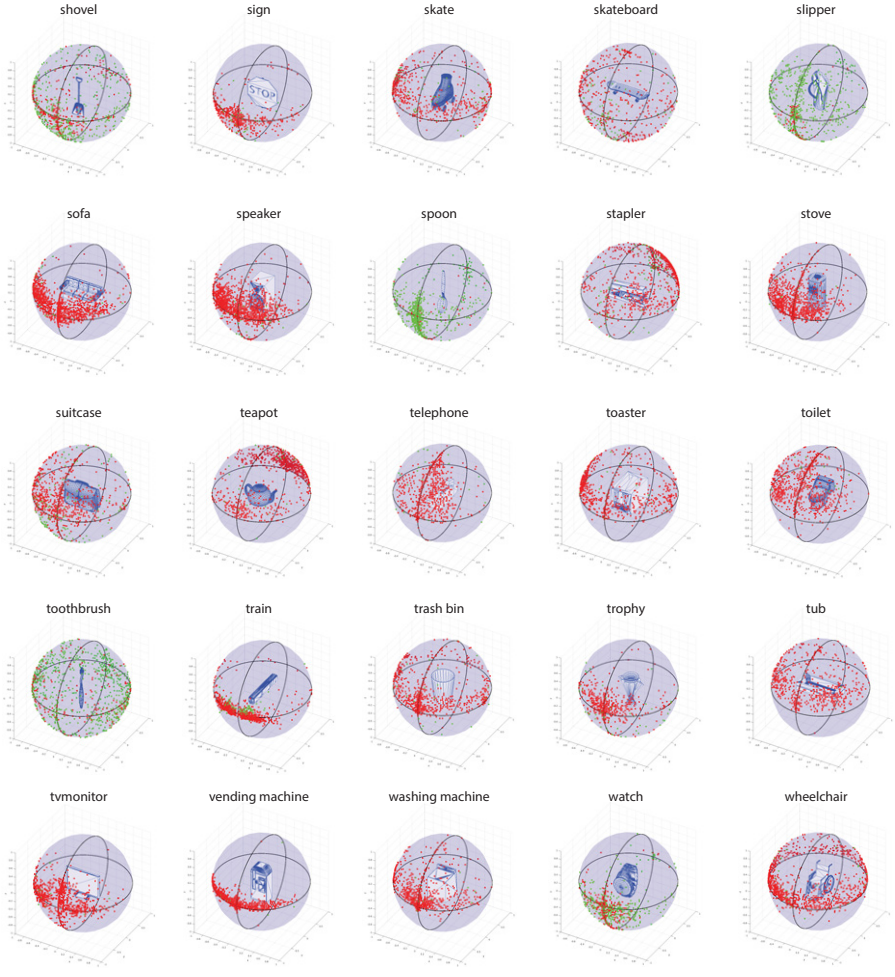

Fig. 4. Viewpoint distributions of different categories in our database. We visualize the camera position as a point on the unit sphere (red points: in-plane rotation $< 15°$; green points: in-plane rotation $> 15°$).

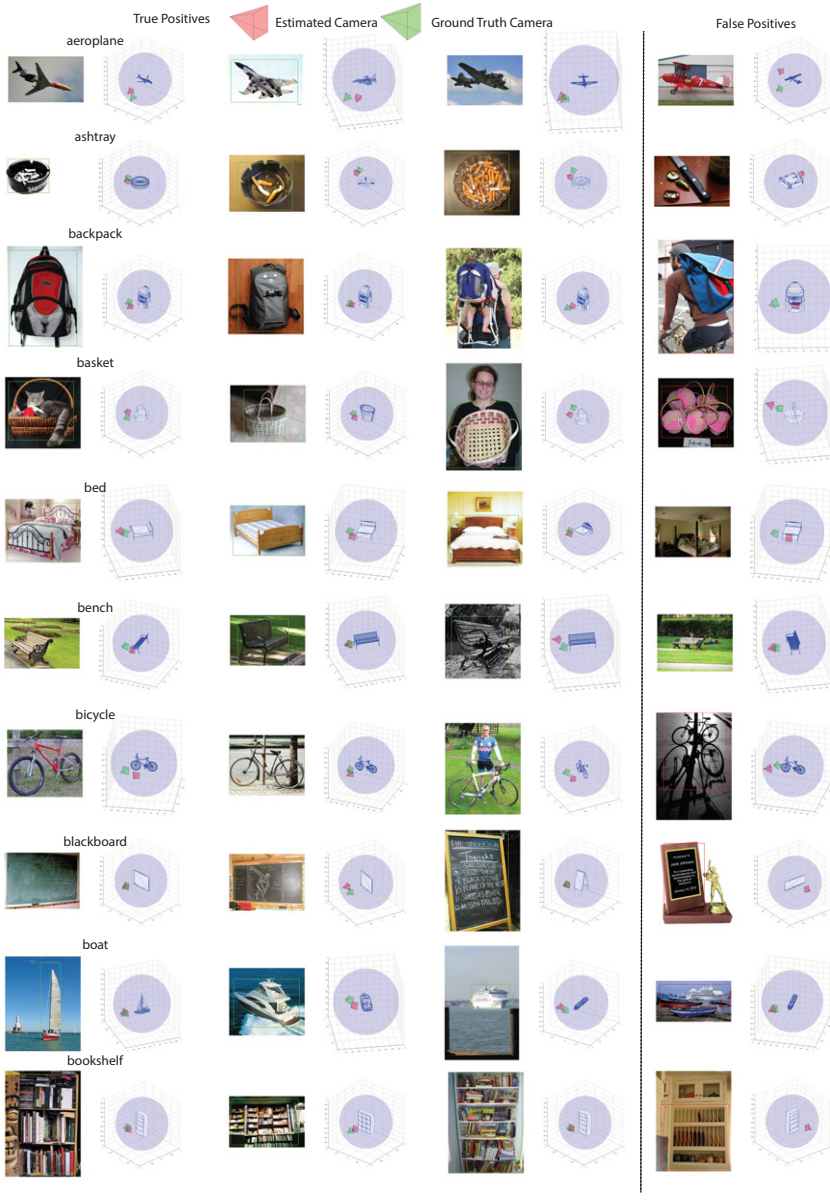

Fig. 5. Object detection and pose estimation examples using our baseline method.

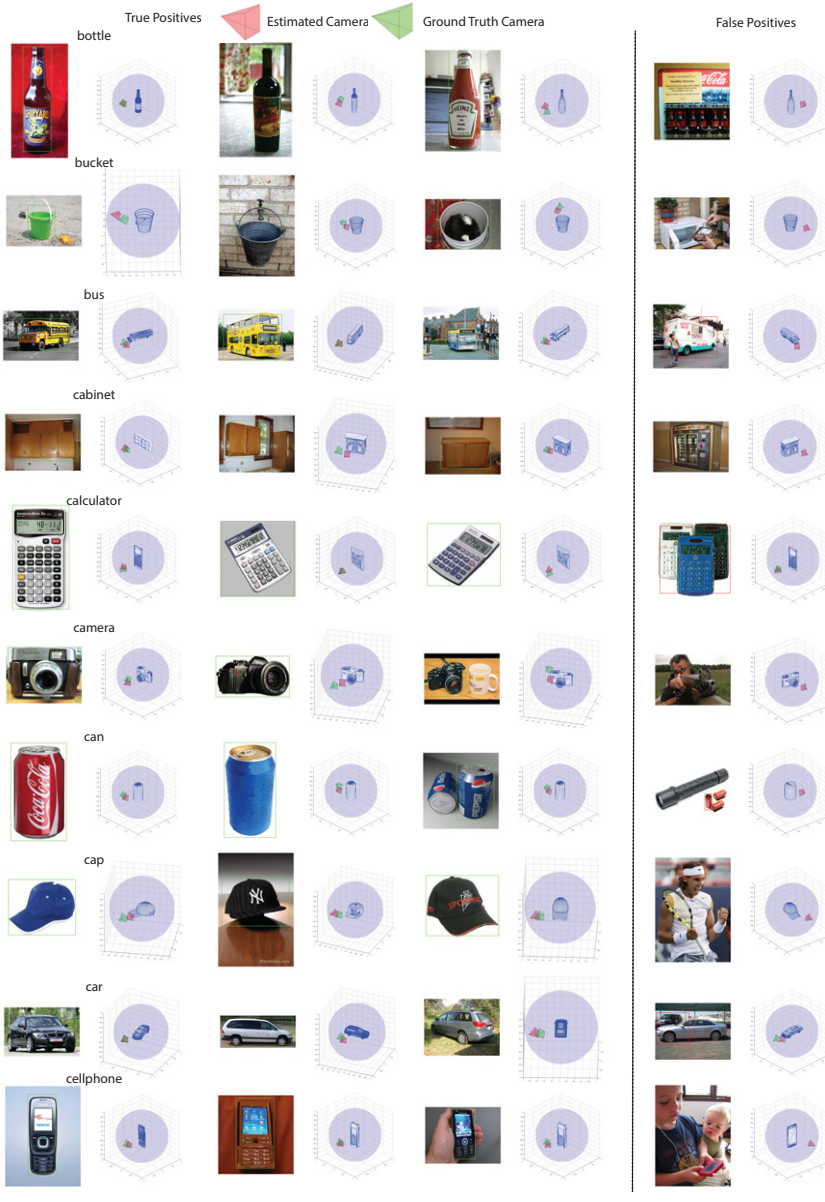

Fig. 6. Object detection and pose estimation examples using our baseline method.

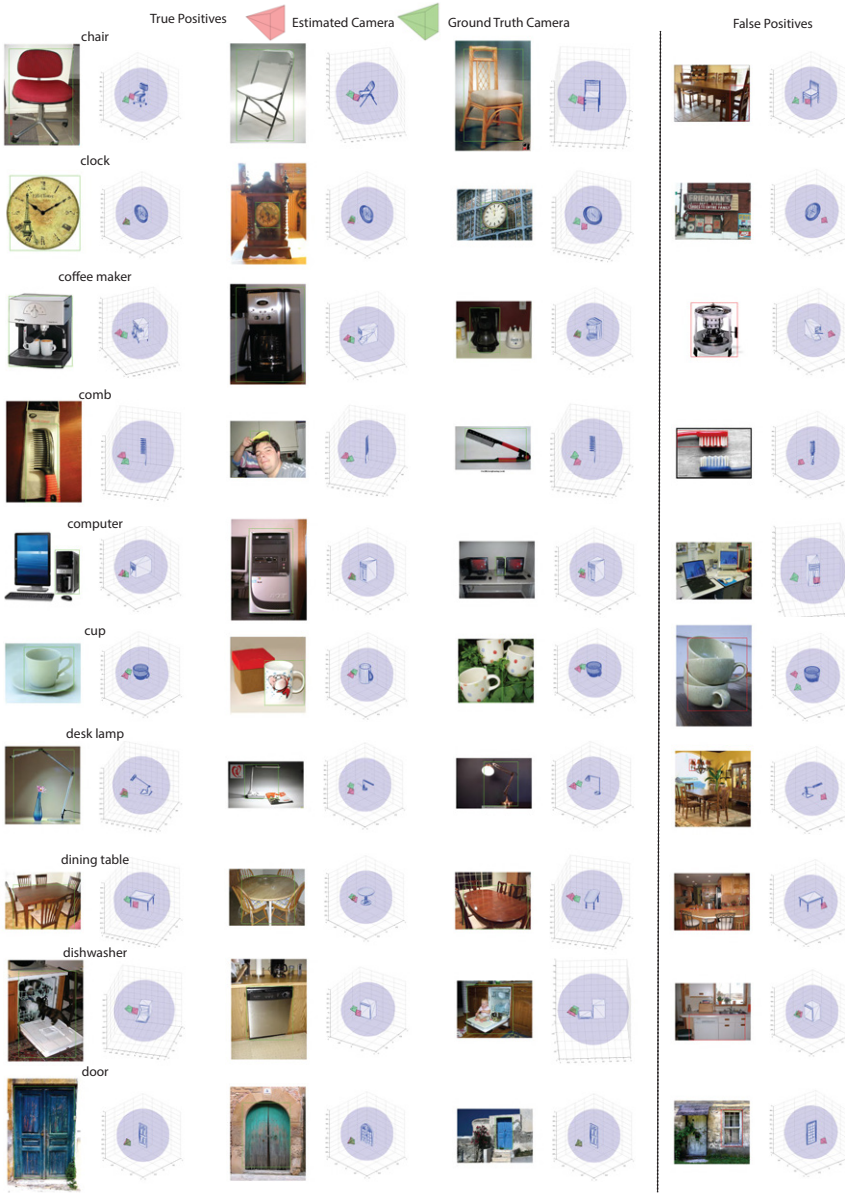

Fig. 7. Object detection and pose estimation examples using our baseline method.

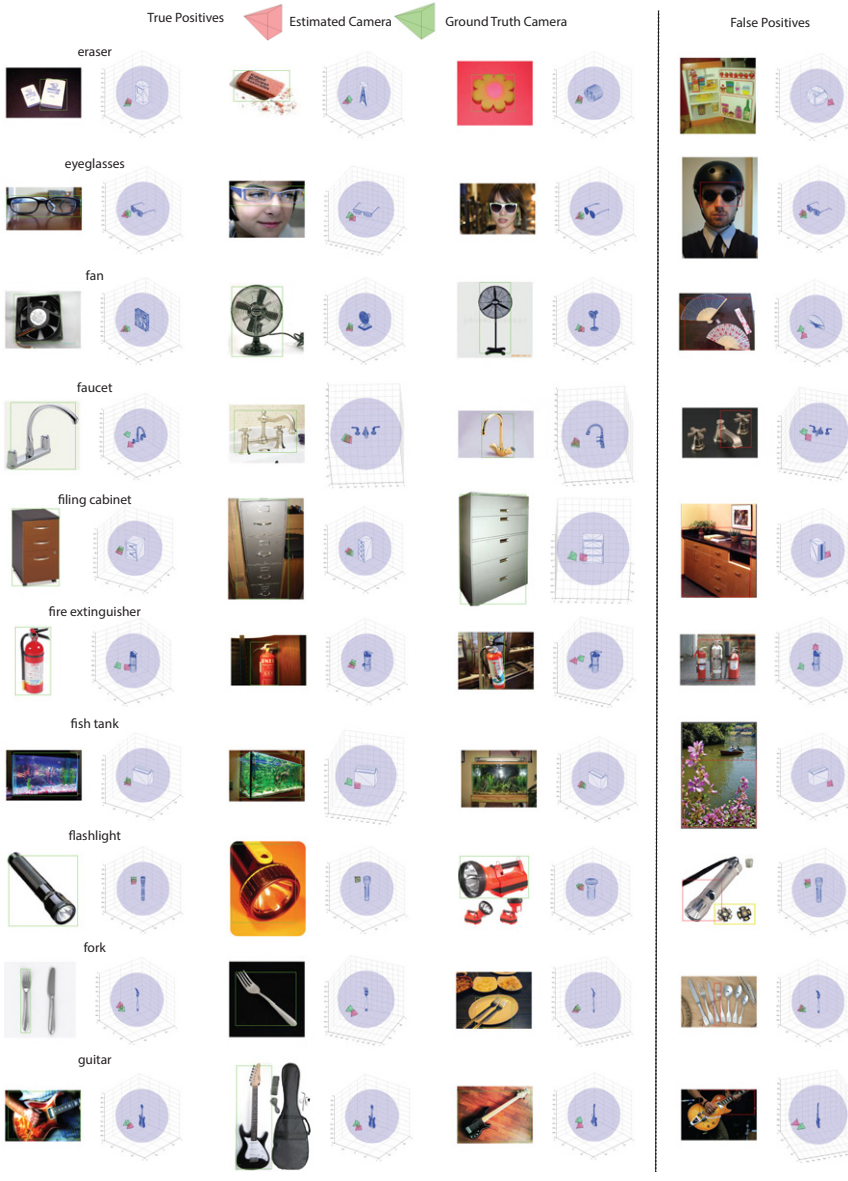

Fig. 8. Object detection and pose estimation examples using our baseline method.

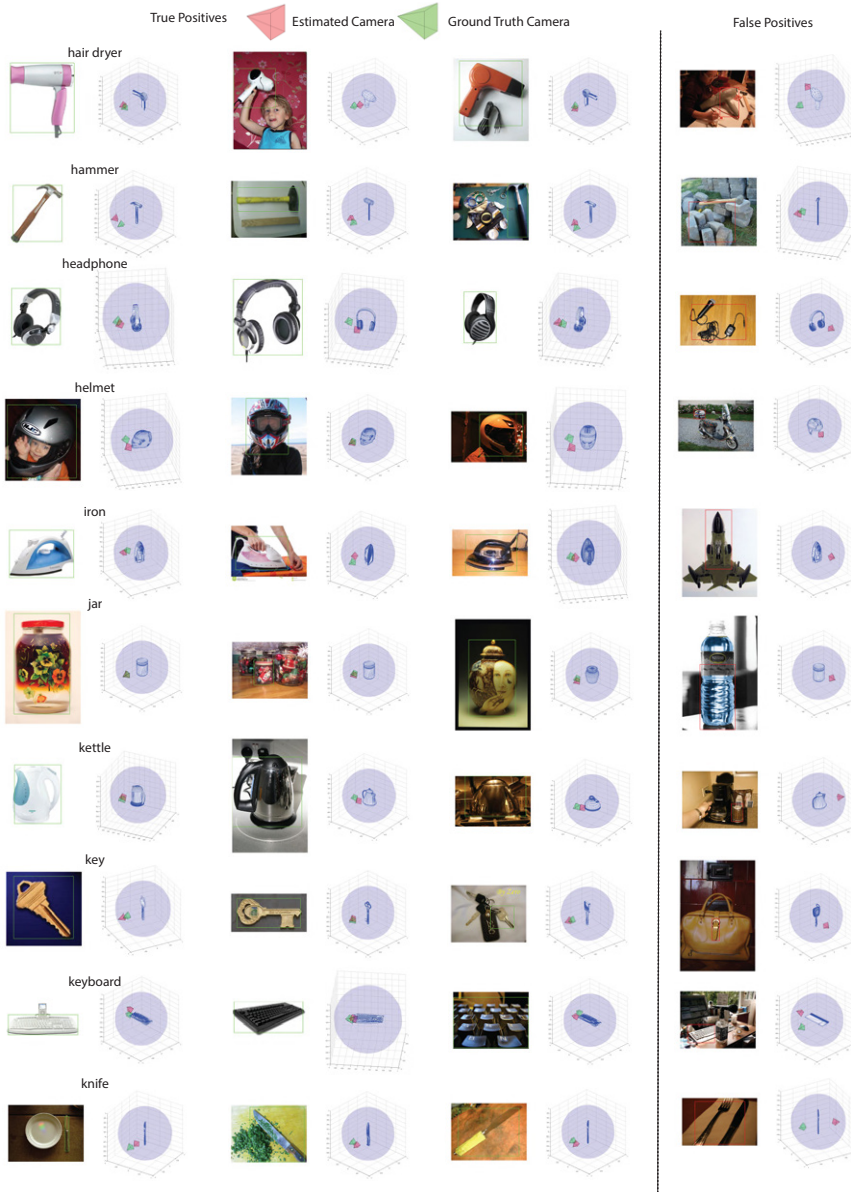

Fig. 9. Object detection and pose estimation examples using our baseline method.

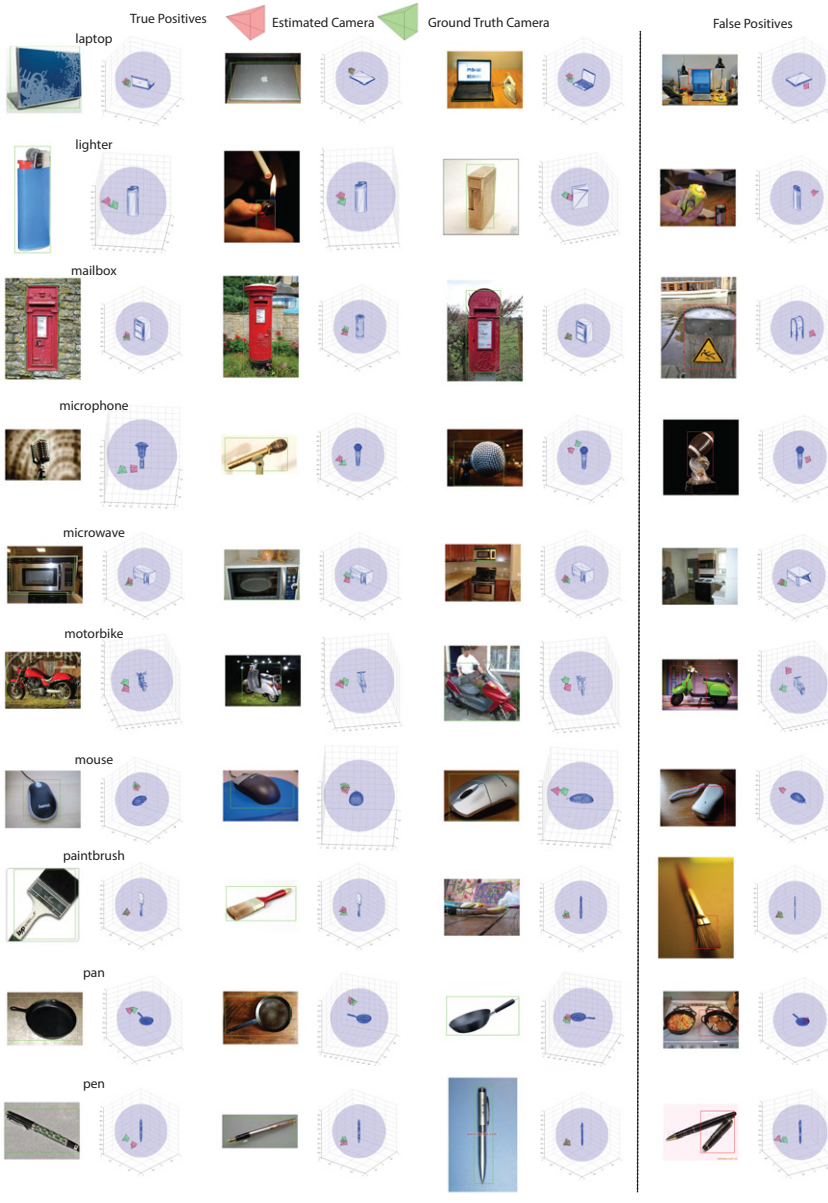

Fig. 10. Object detection and pose estimation examples using our baseline method.

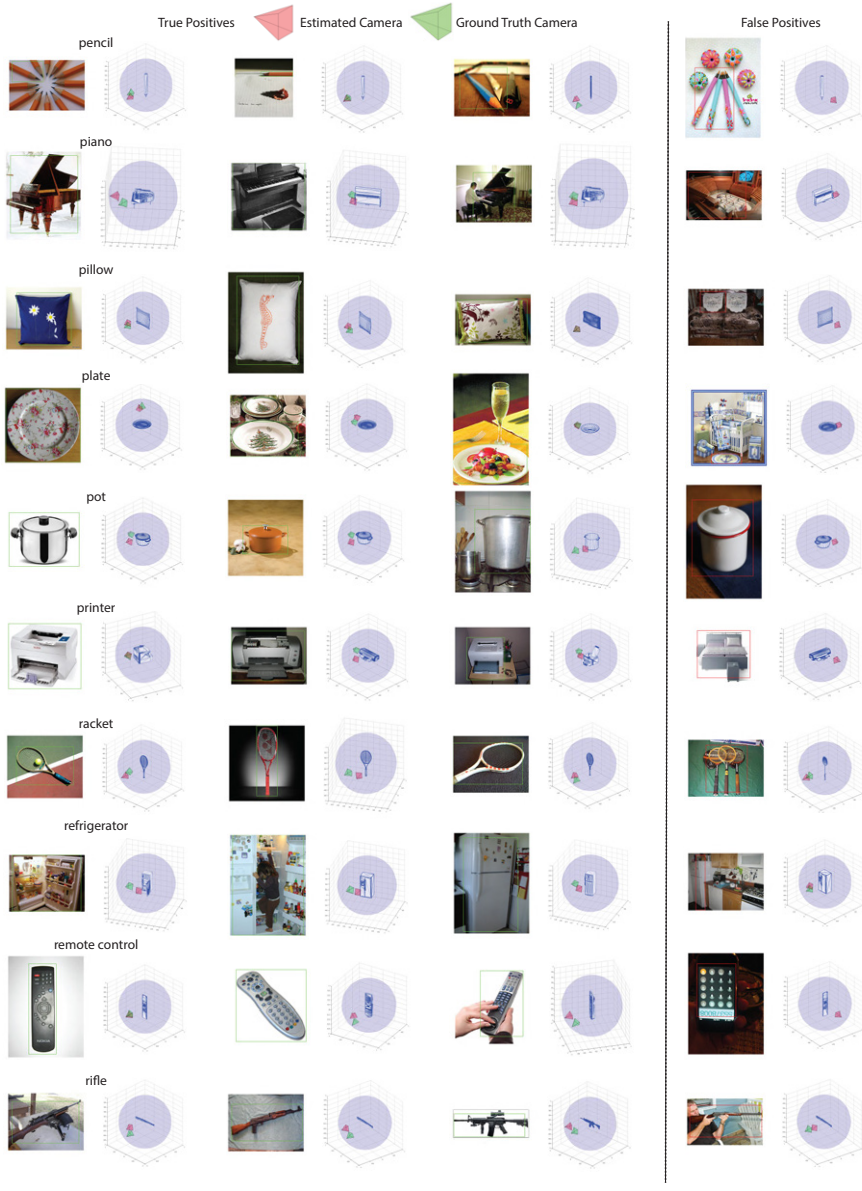

Fig. 11. Object detection and pose estimation examples using our baseline method.

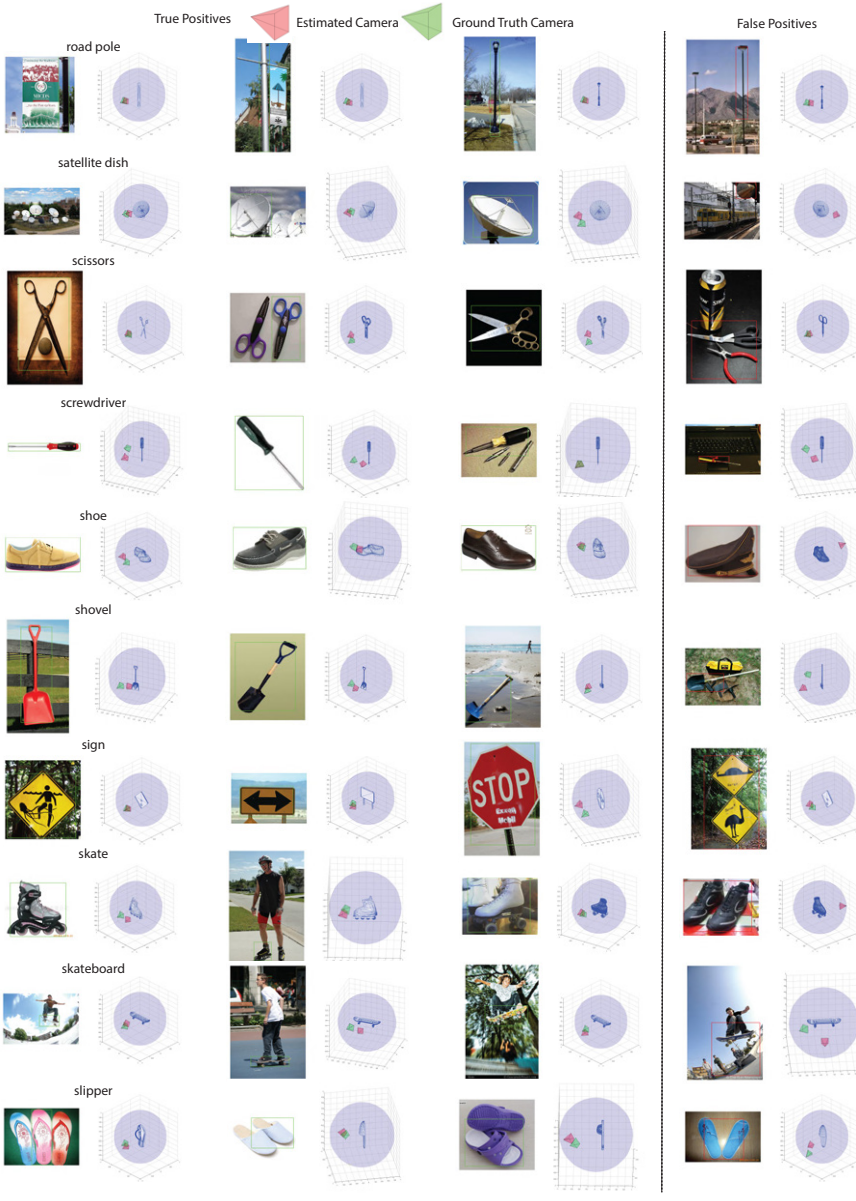

Fig. 12. Object detection and pose estimation examples using our baseline method.

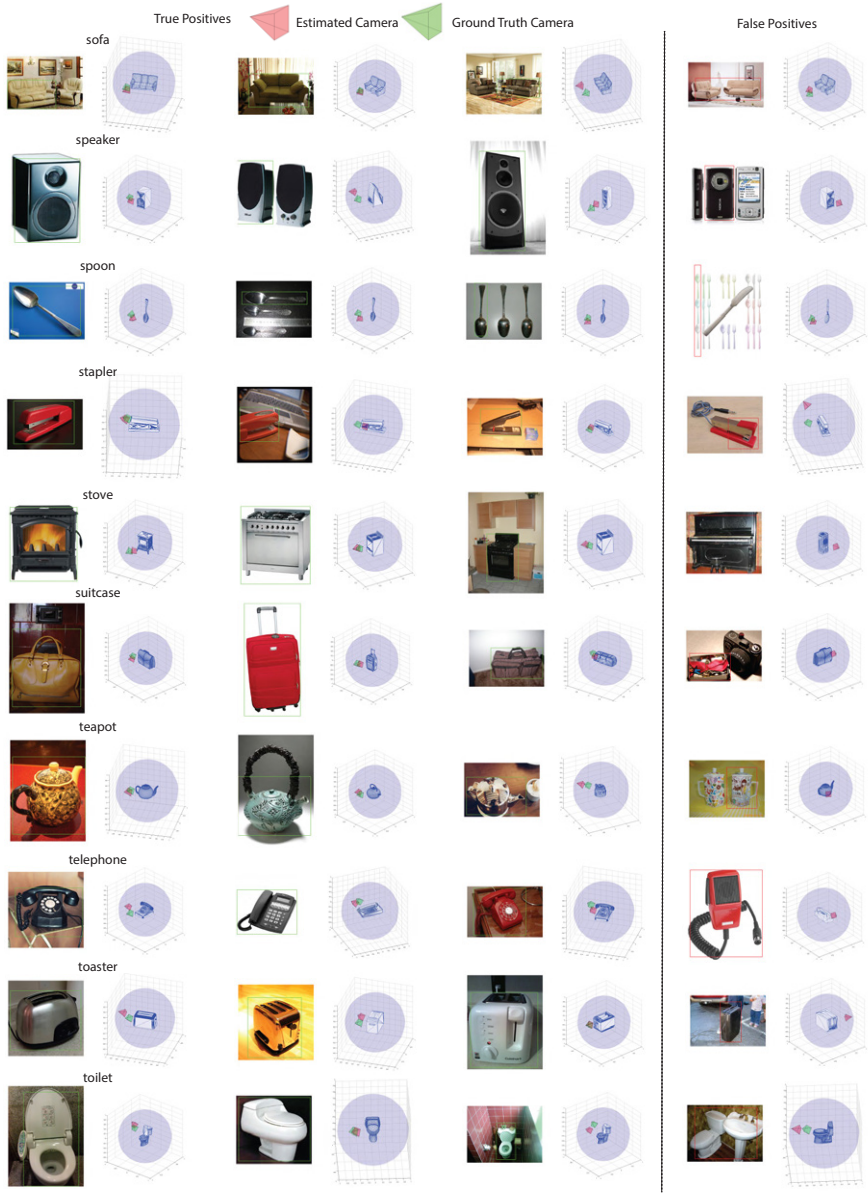

Fig. 13. Object detection and pose estimation examples using our baseline method.

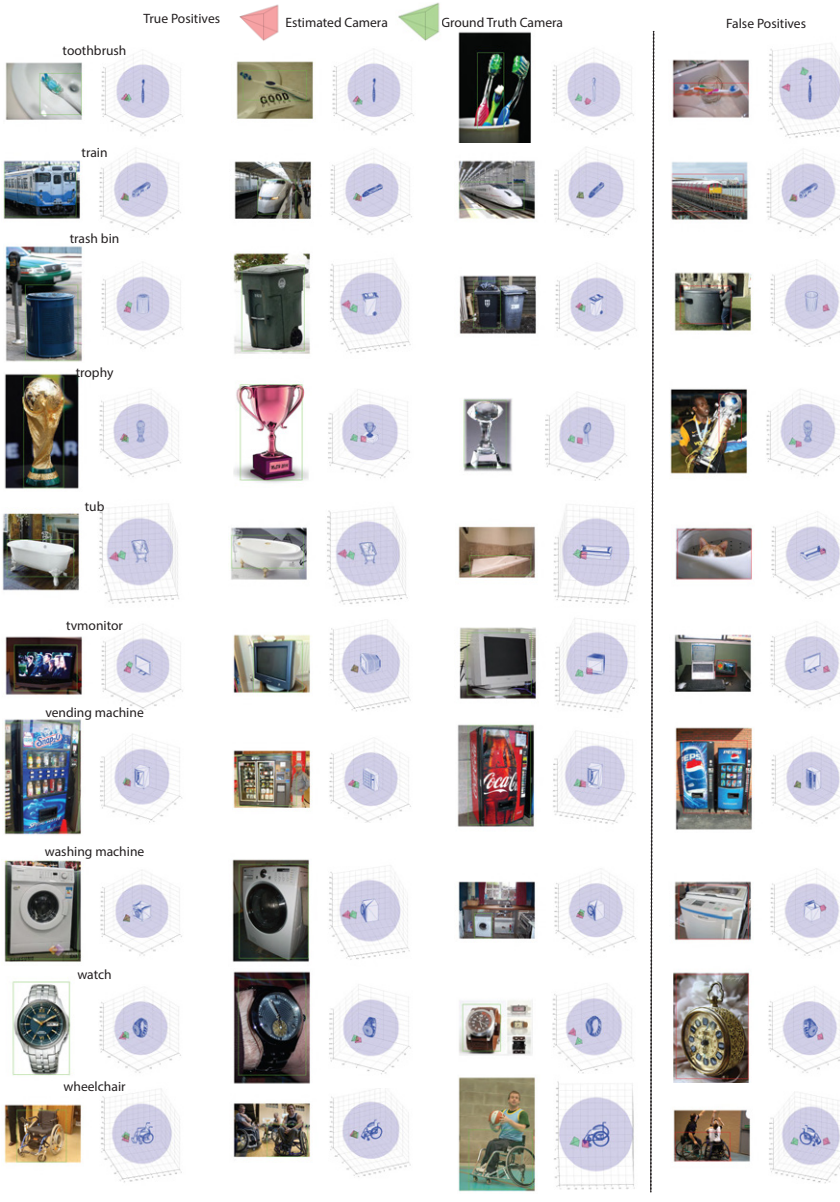

Fig. 14. Object detection and pose estimation examples using our baseline method.

Fig. 15. 3D shape retrieval examples. Green boxes are the selected 3D shapes.

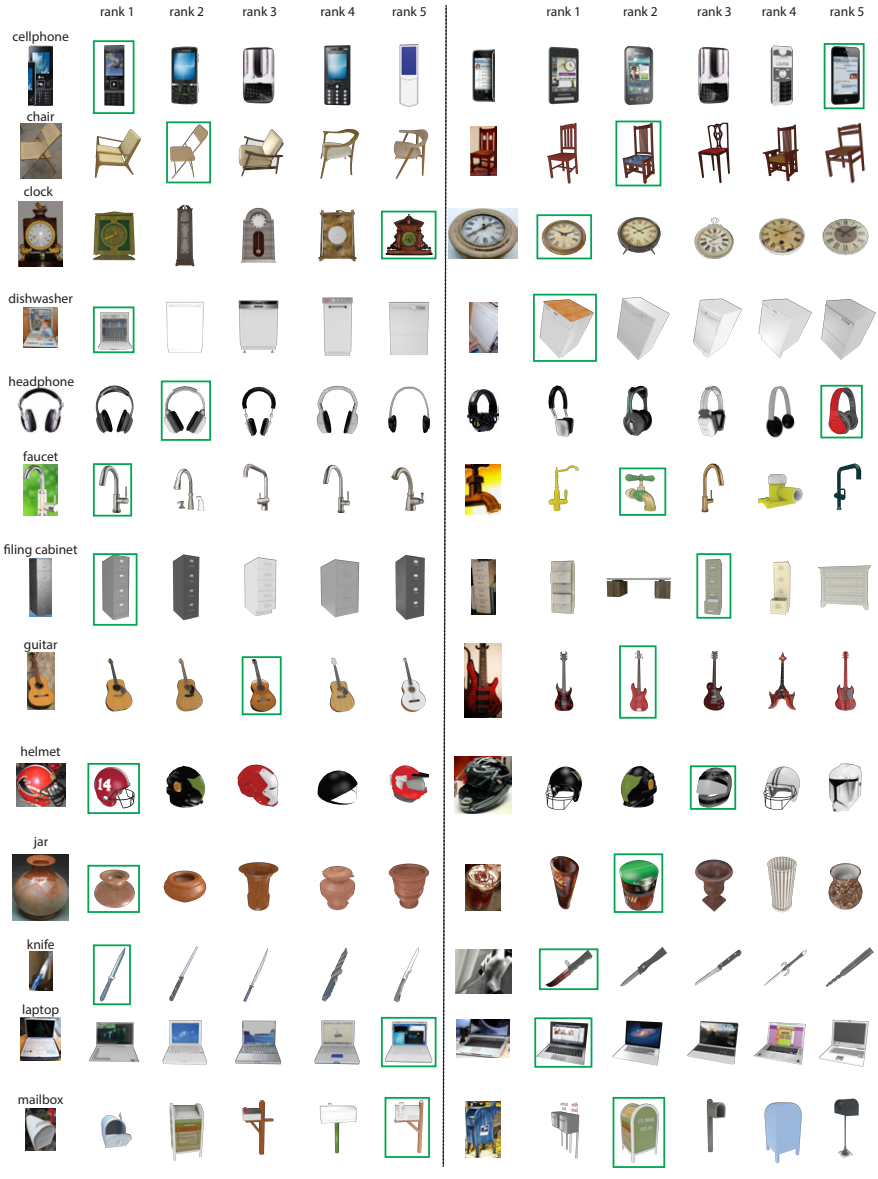

Fig. 16. 3D shape retrieval examples. Green boxes are the selected 3D shapes.

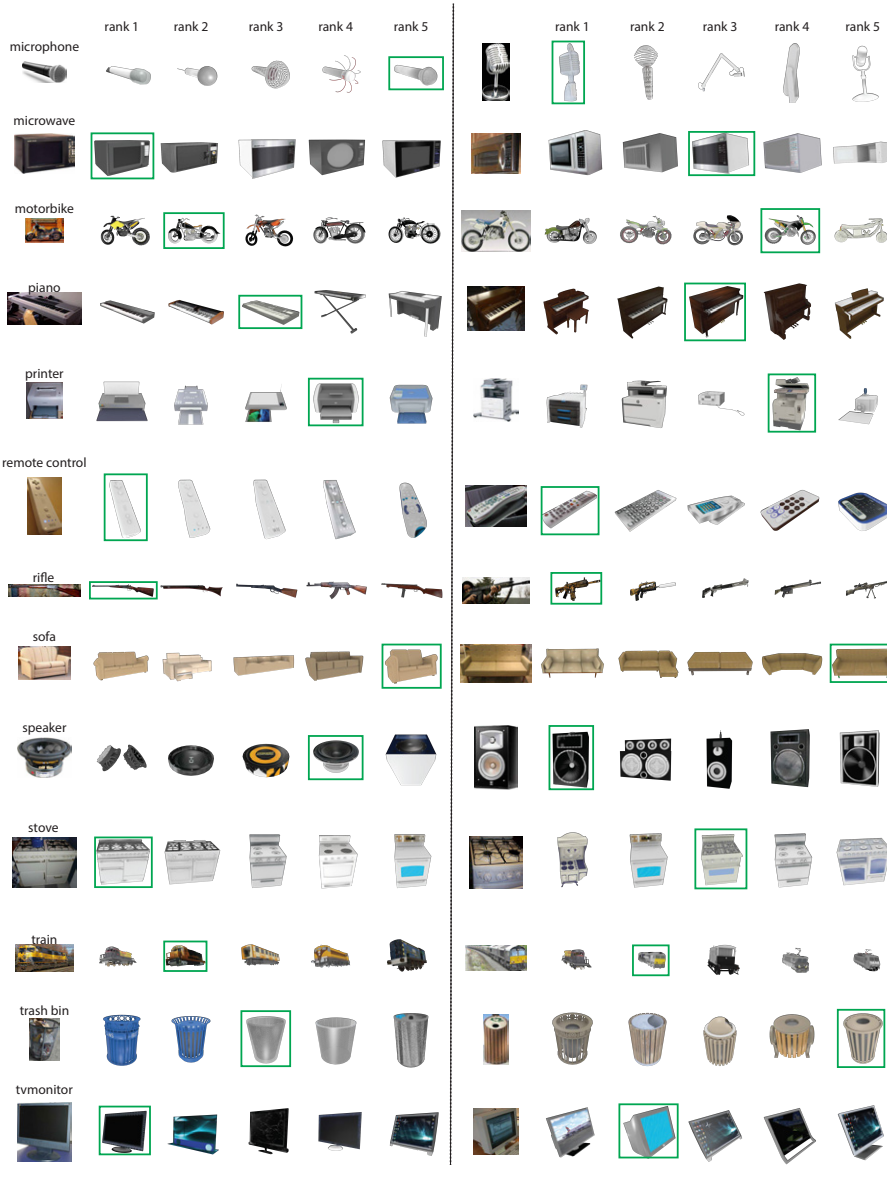

Fig. 17. 3D shape retrieval examples. Green boxes are the selected 3D shapes.