

## A. More Visualization of COCO-O

We present the class distribution of COCO-O in Figure 10. The number of instances per image in COCO-O are shown in Figure 8. COCO-O have around 5% more iconic images compared with COCO validation set. To study the impact, we decrease the proportion by randomly deleting iconic images in COCO-O, and re-run the evaluation. The ablation in Table 5 suggests more iconic images in COCO-O will not bring great change in result.

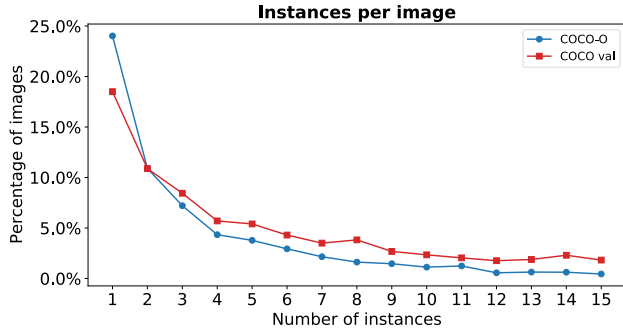


Figure 8: Number of annotated instances per image for COCO and COCO-O.

Detectors	Percent of iconic images			
	24%	22%	20%	18%
Faster R-CNN	16.4	16.3	16.5	16.4
RetinaNet	16.6	16.5	16.4	16.5

Table 5: Impact of the more iconic images in COCO-O.

## B. Detailed Results of Compared Detectors

Table 6 presents the detailed results of compared detectors in Figure 1. For compared SOTA detectors, we introduce their details in Table 7.

## C. Comparison of Results on COCO-C and COCO-O

Figure 9 suggests COCO-C has a positive growth trend with clean performance, while such trend in COCO-O is more weak. It demonstrates COCO-O measures the robustness aspect out of standard COCO test set.

## D. Visualization of Detection Results

We adopt five detectors in Table 6: Faster R-CNN, DETR, YOLOX-S, PVTv2-B2, GLIP-T, and visualize their results on our COCO-O in Figure 12. The score threshold is set as 0.5, which means that only bounding box with classification confidence over 0.5 will be visualized. Among

the compared detectors, Faster R-CNN has failed to localize most objects and leads to a low recall rate. Detectors such as DETR, YOLOX-S recall some wrong bounding box with misclassified category or meaningless positions. GLIP-T outputs more accurate detection results. It can even recognize some abstract objects, such as the person with umbrella (the third line in figure).

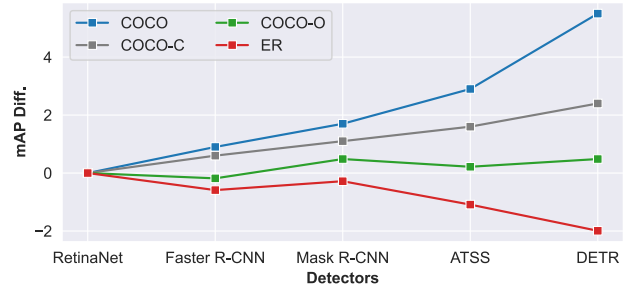


Figure 9: The results of COCO-C and COCO-O on different detectors.

## E. Visualization of Bad Cases

Among COCO-O images, we have found some bad cases which make most detectors fail. They have been shown Figure 11. Although the presented images contain common objects that can be easily recognized by humans, machines are hard to understand them. A vivid example is the anime characters in “The Powerpuff Girls” or “Dora the Explorer”, which should be detected as person while machine looks unaware of such art form of expression. In addition, some traditional factors, *e.g.*, perspective change, severe haze or blur, very poor lighting conditions, will also increase the difficulty of detection task.

## F. Usage of COCO-O

- Example code for testing with MMDetection:

```
$ git clone https://github.com/open-mmlab/mmdetection.git
$ cd mmdetection
$ mkdir data
$ ln -s /path/to/coco-o/sketch data/coco
$ python tools/test.py ${CONFIG_FILE} \
${CHECKPOINT_FILE} --eval bbox
```

- Example code for testing with Detectron2:

```
$ git clone https://github.com/facebookresearch/detectron2.git
$ cd detectron2
$ mkdir datasets
$ ln -s /path/to/coco-o/sketch datasets/coco
$ ./tools/lazyconfig_train_net.py \
--config-file ${CONFIG_FILE} \
train.init_checkpoint=${CHECKPOINT_FILE} \
--eval-only
```

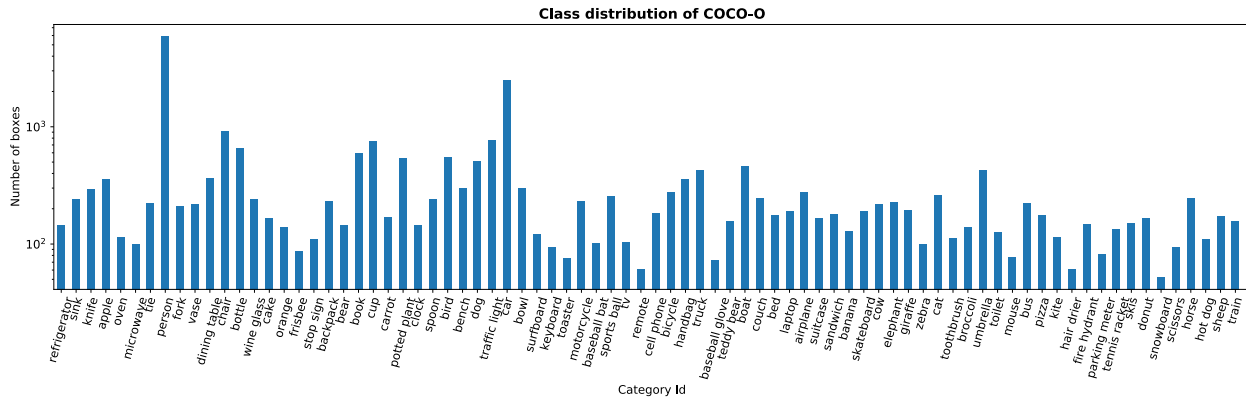


Figure 10: Class distribution of our COCO-O.

Detection Methods	BackBone	COCO mAP	COCO rank	COCO-O mAP	COCO-O rank	Effective Robustness	Rank Diff.	Use Mask	Additional Data	FPS
Mask R-CNN	ViT-B	51.6	1	24.3	8	+1.06	-7	✓	✗	12.7
YOLOv7	-	51.2	2	29.0	3	+5.99	-1	✗	✗	161.0
FIBER-B	Swin-B	49.5	3	33.7	1	+11.43	+2	✗	✓	3.1
Mask R-CNN	Swin-S	48.5	4	25.5	6	+3.71	-2	✓	✗	10.1
Mask R-CNN	MViTV2-T	48.3	5	25.7	5	+3.98	0	✓	✗	14.6
GLIP-T	Swin-T	46.6	6	29.1	2	+8.11	+4	✗	✓	5.5
Mask R-CNN	Uniformer-S	45.6	7	26.2	4	+5.65	+3	✓	✗	15.6
Mask R-CNN	PVTv2-B2	45.3	8	24.9	7	+4.52	+1	✓	✗	15.8
Deformable-DETR	ResNet-50	44.5	9	18.5	13	-1.49	-4	✗	✗	15.0
DyHead	ResNet-50	42.5	10	19.3	10	+0.16	0	✗	✗	13.2
HTC	ResNet-50	42.3	11	19.1	11	+0.08	0	✓	✗	5.8
DETR	ResNet-50	42.0	12	17.1	17	-1.82	-5	✗	✗	27.8
VarifocalNet	ResNet-50	41.6	13	18.8	12	+0.05	+1	✗	✗	19.4
YOLOX-S	-	40.5	14	20.6	9	+2.48	+5	✗	✗	102.0
Cascade R-CNN	ResNet-50	40.3	14	18.2	14	+0.02	0	✗	✗	16.1
ATSS	ResNet-50	39.4	16	16.8	19	-0.91	-3	✗	✗	19.7
Mask R-CNN	ResNet-50	38.2	17	17.1	17	-0.11	0	✓	✗	16.1
Sparse R-CNN	ResNet-50	37.9	18	17.5	16	+0.41	+2	✗	✗	18.3
YOLOv5s	-	37.4	19	17.8	15	+1.00	+4	✗	✗	156.3
Faster R-CNN	ResNet-50	37.4	19	16.4	22	-0.41	-3	✗	✗	21.4
FCOS	ResNet-50	36.6	21	16.7	20	+0.25	+1	✗	✗	22.7
RetinaNet	ResNet-50	36.5	22	16.6	21	+0.18	+1	✗	✗	19.0
YOLOS-S	-	36.1	23	15.5	23	-0.80	0	✗	✗	5.7
YOLOv3	DarkNet-53	33.7	24	14.8	24	-0.37	0	✗	✗	48.1
SSD	VGG16	29.5	25	13.6	25	+0.36	0	✗	✗	30.7

Table 6: Detailed results of the compared detectors in Figure 1.

Detection Methods	BackBone	COCO mAP	COCO rank	COCO-O mAP	COCO-O rank	Effective Robustness	Rank Diff.	Use Mask	Additional Data	FPS
EVA	-	64.2	1	57.8	1	+28.86	0	✓	✓	0.8
DETA	Swin-L	62.9	2	48.5	2	+20.15	0	✗	✓	1.2
GRiT	ViT-H	60.4	3	42.9	4	+15.72	-1	✓	✓	0.7
CBNetV2	Swin-L	59.1	4	39	7	+12.36	-3	✓	✓	2
ViT-Adapter	BEiTv2-L	58.8	5	34.25	13	+7.79	-8	✓	✓	5.6
ViTDet	ViT-H	58.7	6	34.3	12	+7.89	-6	✓	✗	4.8
DINO	Swin-L	58.5	7	42.1	5	+15.76	+2	✗	✓	1.8
H-Deformable-DETR	Swin-L	57.6	8	41.4	6	+15.45	+2	✗	✓	1.9
YOLOv6-L6	-	57.2	9	32.5	18	+6.73	-9	✗	✗	26
YOLOR-D6	-	57	10	33	17	+7.3	-7	✗	✗	34
YOLOv7-E6E	-	56.8	11	32	19	+6.42	-8	✗	✗	36
SOLQ	Swin-L	56.5	12	38.8	8	+13.41	+4	✓	✓	2.3
DyHead	Swin-L	56.2	13	35.3	11	+10	+2	✗	✓	6.3
Cascade Mask R-CNN	InternImage-L	56.1	14	37	10	+11.72	+4	✓	✓	9.1
Cascade Mask R-CNN	MViTV2-H	56.1	14	30.9	21	+5.62	-7	✓	✓	3.3
CenterNet2	R2-101-DCN	56.1	14	29.5	25	+4.29	-11	✗	✗	11.8
QueryInst	Swin-L	55.5	17	33.2	16	+8.26	+1	✓	✓	3.3
Cascade Mask R-CNN	ConvNeXt-XL	55.2	18	37.5	9	+12.68	+9	✓	✓	8.6
YOLOv4-P6	-	54.4	19	30.4	23	+5.89	-4	✗	✗	30
YOLOv5l6	-	53.7	20	29.3	26	+5.17	-6	✗	✗	63
Cascade Mask R-CNN	UniFormer-B	53.4	21	31.8	20	+7.77	+1	✓	✗	12.4
DiffusionDet	Swin-B	52.2	22	34.2	14	+10.69	+8	✗	✓	10.8
Det-AdvProp	EfficientNet-B5	52.2	22	30.8	22	+7.34	0	✗	✗	13.8
Cascade Mask R-CNN	Swin-B	51.9	24	25.5	35	+2.18	-11	✓	✗	11.6
GLIP-L	Swin-L	51.4	25	48	3	+24.89	+22	✗	✓	0.5
EfficientDet-D5	EfficientNet-B5	51.2	26	28.5	29	+5.44	-3	✗	✗	13.8
YOLOX-X	-	51.1	27	30.3	24	+7.26	+3	✗	✗	57.8
Cascade Mask R-CNN	ResNeSt-200-DCN	50.9	28	26.1	33	+3.18	-5	✓	✗	12.4
UniverseNet	R2-101-DCN	50.9	28	24.8	38	+1.86	-10	✗	✗	11.7
HTC	RX-101-64x4d-DCN	50.4	30	28.4	30	+5.67	0	✓	✗	2.1
VFNet	RX-101-64x4d	50.4	30	28	32	+5.27	-2	✗	✗	6.7
GFLv2	R2-101-DCN	49.9	32	25.1	36	+2.6	-4	✗	✗	10.9
FIBER-B	Swin-B	49.5	33	33.7	15	+11.43	+18	✗	✓	3.1
Mask R-CNN	MogaNet-L	49.4	34	28.8	28	+6.57	+6	✓	✗	11.2
RepPointsV2	RX-101-64x4d-DCN	49.3	35	24.9	37	+2.7	-2	✗	✗	5.8
GCNet	RX-101-32x4d-DCN	48	36	26	34	+4.38	+2	✓	✗	7.5
TOOD	RX-101-64x4d	47.6	37	23.1	40	+1.7	-3	✗	✗	6.1
Mask R-CNN	PVTv2-B5	47.4	38	28.2	31	+6.85	+7	✓	✗	12.8
Deformable DETR	ResNet-50	46.9	39	20.1	45	-1.01	-6	✗	✗	14.5
GLIP-T	Swin-T	46.6	40	29	27	+8.01	+13	✗	✓	5.5
Cascade Mask R-CNN	RX-101-64x4d	46.6	40	22.2	41	+1.23	-1	✓	✗	6.7
HTC	HRNetV2-W40	46.4	42	24	39	+3.07	+3	✓	✗	4.3
Mask R-CNN	RX-101-64x4d	44.5	43	21.5	42	+1.48	+1	✓	✗	8
Faster R-CNN	RX-101-64x4d	43.1	44	20.2	44	+0.76	0	✗	✗	9.4
FCOS	RX-101-64x4d	42.6	45	21	43	+1.83	+2	✗	✗	9.7
YOLOS-B	ViT-B	42	46	20	46	+1.05	0	✗	✗	2.7
DETR	ResNet-50	42	46	17.1	49	-1.82	-3	✗	✗	27
RetinaNet	RX-101-64x4d	41.6	48	19.9	47	+1.16	+1	✗	✗	8.7
RepPointsV2	MobileNetV2	36.8	49	16.4	51	-0.18	-2	✗	✗	32.1
NanoDet	ShuffleNetV2	34.1	50	17.3	48	+1.92	+2	✗	✗	256
YOLOv3	DarkNet-53	33.7	51	14.8	52	-0.37	-1	✗	✗	48.1
YOLACT	ResNet-101	30.4	52	16.8	50	+3.15	+2	✓	✗	33.5
SSD	VGG16	29.5	53	13.6	53	+0.36	0	✗	✗	30.7

Table 7: Detailed results of the compared SOTA detectors in Figure 6.

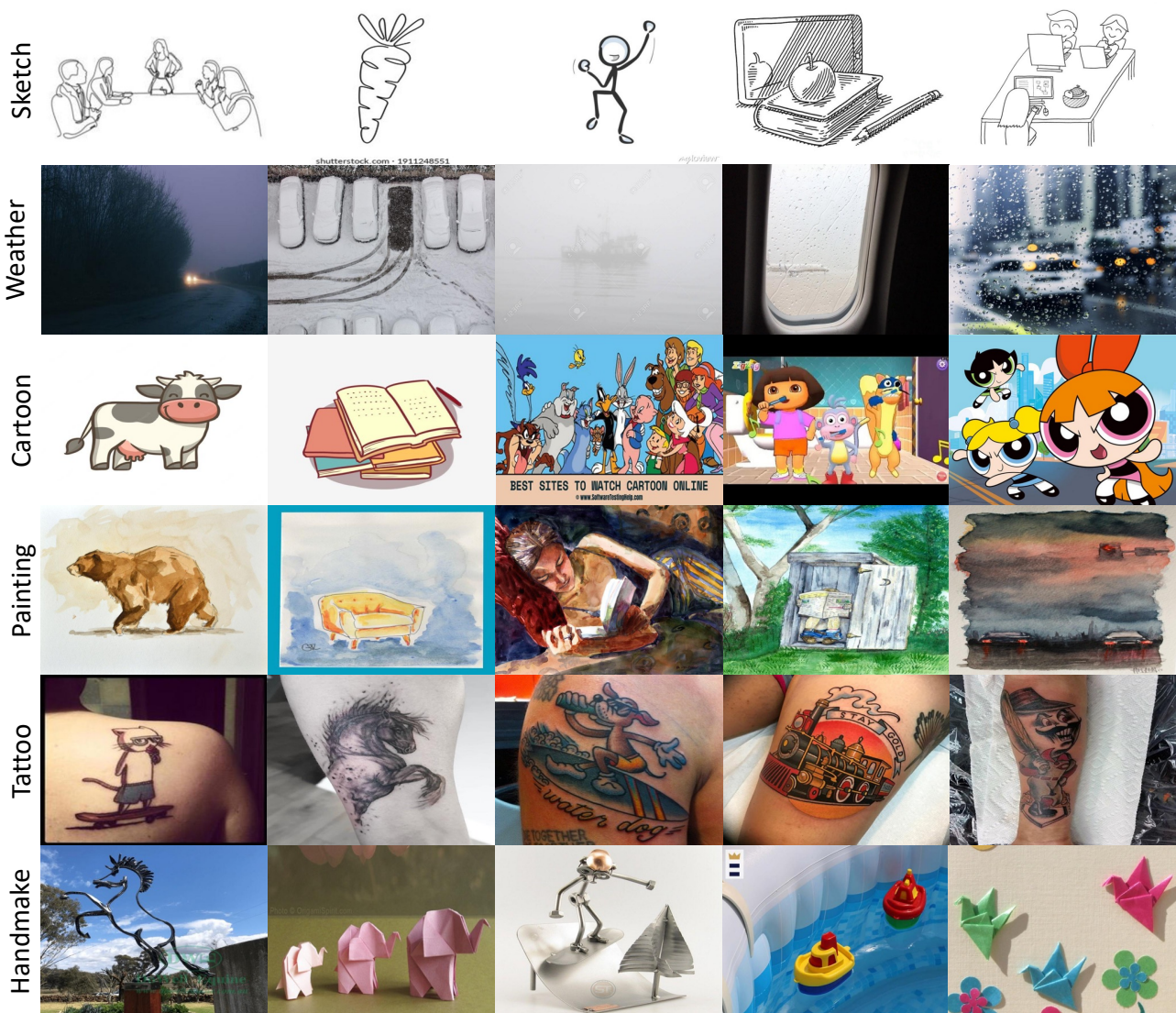


Figure 11: Hard examples in COCO-O which make detection failed.

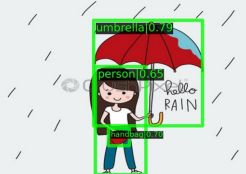
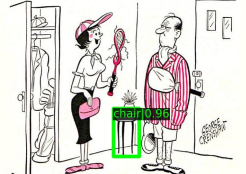
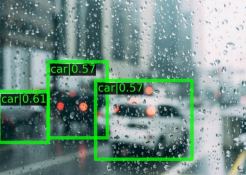
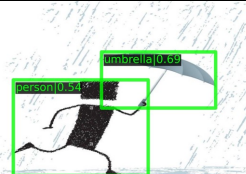
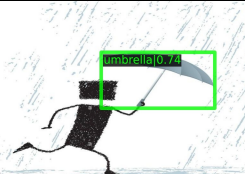
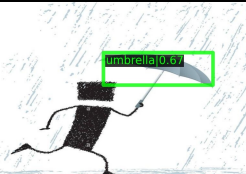
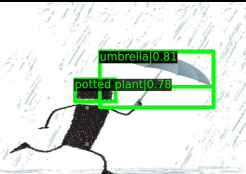
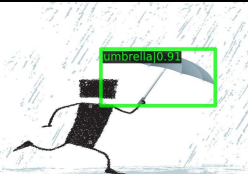
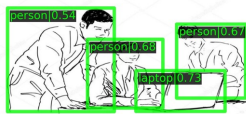
Faster R-CNN

DETR

YOLOX-S

PVTv2-B2

GLIP-T





## G. Queries for Collecting COCO-O

We show the query keywords used for collect sketch images in COCO-O. Similarly, for other test domains, we create queries by just replacing with the corresponding domain keywords.

```
sketch person
sketch bicycle
sketch car
sketch motorcycle
sketch airplane
sketch bus
sketch train
sketch truck
sketch boat
sketch traffic light
sketch fire hydrant
sketch stop sign
sketch parking meter
sketch bench
sketch bird
sketch cat
sketch dog
sketch horse
sketch sheep
sketch cow
sketch elephant
sketch bear
sketch zebra
sketch giraffe
sketch backpack
sketch umbrella
sketch handbag
sketch tie
sketch suitcase
sketch frisbee
sketch skis
sketch snowboard
sketch sports ball
sketch kite
sketch baseball bat
sketch baseball glove
sketch skateboard
sketch surfboard
sketch tennis racket
sketch bottle
sketch wine glass
sketch cup
sketch fork
sketch knife
sketch spoon
sketch bowl
sketch banana
sketch apple
sketch sandwich
sketch orange
sketch broccoli
sketch carrot
sketch hot dog
sketch pizza
sketch donut
sketch cake
sketch chair
sketch couch
sketch potted plant
sketch bed
```

```
sketch dining table
sketch toilet
sketch tv
sketch laptop
sketch mouse
sketch remote
sketch keyboard
sketch cell phone
sketch microwave
sketch oven
sketch toaster
sketch sink
sketch refrigerator
sketch book
sketch clock
sketch vase
sketch scissors
sketch teddy bear
sketch hair drier
sketch toothbrush
sketch person chair
sketch person car
sketch person handbag
sketch person dining table
sketch chair dining table
sketch person cup
sketch person backpack
sketch person bottle
sketch cup dining table
sketch person cell phone
sketch person tennis racket
sketch person bench
sketch person truck
sketch person sports ball
sketch car truck
sketch bowl dining table
sketch person umbrella
sketch person bus
sketch person surfboard
sketch person tie
sketch person bowl
sketch cup chair
sketch bottle cup
sketch person motorcycle
sketch person skateboard
sketch person bicycle
sketch bottle dining table
sketch person skis
sketch knife dining table
sketch cup bowl
sketch car traffic light
sketch person traffic light
sketch fork dining table
sketch car bus
sketch person knife
sketch bottle chair
sketch person book
sketch person baseball glove
sketch bottle bowl
sketch person remote
sketch person baseball bat
sketch person pizza
sketch person couch
sketch chair couch
sketch chair book
sketch person horse
sketch spoon bowl
```

sketch spoon dining table  
sketch cup knife  
sketch person tv  
sketch chair tv  
sketch backpack handbag  
sketch fork knife  
sketch person kite  
sketch cup spoon  
sketch sports ball tennis racket  
sketch person wine glass  
sketch person laptop  
sketch bowl chair  
sketch person clock  
sketch chair potted plant  
sketch person train  
sketch pizza dining table  
sketch person dog  
sketch person suitcase  
sketch person boat  
sketch person potted plant  
sketch person spoon  
sketch bottle sink  
sketch person cake  
sketch person frisbee  
sketch person fork  
sketch cup fork  
sketch cake dining table  
sketch mouse keyboard  
sketch toilet sink  
sketch tv book  
sketch tv keyboard  
sketch chair laptop  
sketch wine glass dining table  
sketch laptop keyboard  
sketch couch remote  
sketch couch book  
sketch laptop mouse  
sketch car motorcycle  
sketch car handbag  
sketch baseball bat baseball glove  
sketch chair vase  
sketch bottle knife  
sketch potted plant dining table  
sketch couch tv  
sketch knife chair  
sketch knife bowl  
sketch person bed  
sketch cup pizza  
sketch wine glass cup  
sketch tv mouse  
sketch fork bowl  
sketch bottle wine glass  
sketch couch dining table  
sketch tv laptop  
sketch bicycle car  
sketch potted plant vase  
sketch dining table vase  
sketch sandwich dining table  
sketch umbrella handbag  
sketch laptop book  
sketch person chair dining table  
sketch person cup dining table  
sketch person car truck  
sketch cup chair dining table  
sketch person cup chair  
sketch person bottle dining table  
sketch person car bus

sketch person backpack handbag  
sketch person sports ball tennis racket  
sketch cup bowl dining table  
sketch person car traffic light  
sketch person bottle cup  
sketch person bottle chair  
sketch person knife dining table  
sketch fork knife dining table  
sketch person bowl dining table  
sketch bottle chair dining table  
sketch bottle cup dining table  
sketch bowl chair dining table  
sketch cup knife dining table  
sketch cup fork dining table  
sketch person cup bowl  
sketch person car handbag  
sketch person baseball bat baseball glove  
sketch cup spoon dining table  
sketch person bottle bowl  
sketch person fork dining table  
sketch spoon bowl dining table  
sketch person pizza dining table  
sketch tv mouse keyboard  
sketch person car motorcycle  
sketch person umbrella handbag  
sketch person cup knife  
sketch knife chair dining table  
sketch bottle cup bowl  
sketch bottle bowl dining table  
sketch cup fork knife  
sketch person spoon dining table  
sketch person cup spoon  
sketch cup spoon bowl  
sketch person fork knife  
sketch chair potted plant dining table  
sketch person spoon bowl  
sketch person bicycle car  
sketch laptop mouse keyboard  
sketch person cup fork  
sketch person car backpack  
sketch person wine glass dining table  
sketch person cake dining table  
sketch person wine glass cup  
sketch bottle cup chair  
sketch fork chair dining table  
sketch chair couch dining table  
sketch person couch remote  
sketch person knife chair  
sketch person bowl chair  
sketch chair couch tv  
sketch fork bowl dining table  
sketch person tennis racket chair  
sketch person handbag chair  
sketch person bottle wine glass  
sketch chair dining table vase  
sketch person chair book  
sketch tv laptop keyboard  
sketch cup pizza dining table  
sketch person bottle knife  
sketch bottle knife dining table  
sketch car truck traffic light  
sketch person sports ball baseball glove  
sketch pizza chair dining table  
sketch person cup pizza  
sketch cup bowl chair  
sketch chair dining table book  
sketch knife bowl dining table



sketch person bicycle backpack  
sketch chair tv book  
sketch person wine glass chair  
sketch person fork chair  
sketch tv laptop mouse  
sketch person handbag cell phone  
sketch person bus truck  
sketch person umbrella chair  
sketch person pizza chair  
sketch bottle fork dining table  
sketch wine glass cup dining table  
sketch person handbag suitcase  
sketch chair couch book  
sketch knife spoon dining table  
sketch person chair laptop  
sketch fork pizza dining table  
sketch bottle cup knife  
sketch person bicycle handbag  
sketch car bus truck  
sketch fork spoon dining table  
sketch person bench handbag  
sketch bottle spoon bowl  
sketch person chair potted plant  
sketch wine glass chair dining table  
sketch cup fork chair  
sketch person fork pizza  
sketch person chair cell phone  
sketch person bottle spoon  
sketch cup knife chair  
sketch chair potted plant vase  
sketch person cup cake  
sketch chair couch potted plant  
sketch person bus handbag  
sketch person car bench  
sketch person sports ball baseball bat  
sketch person sports ball chair