Supplementary material: DeepCO^3: Deep Instance Co-segmentation by Co-peak Search and Co-saliency Detection

Kuang-Jui Hsu[1,2]   Yen-Yu Lin[1]   Yung-Yu Chuang[1,2]
[1]Academia Sinica, Taiwan   [2]National Taiwan University, Taiwan

In the supplementary material, we first provide the statistics and image samples of the four datasets we collected. Then, the class-wise performance of our method for instance co-segmentation and object co-localization on the four datasets is reported. Finally, we display the results produced by our method.

1. Dataset Statistics and Image Samples

We provide the class-wise statistics, including the number of images, the average number of instances per image, and the average area ratio of an instance to the whole image, of four collected datasets. Figure 1 gives the statistics for the COCO-VOC, VOC12, and SOC datasets, and Figure 2 reports for the COCO-NONVOC dataset.

We randomly sample five images for each class from the four collected datasets and show them in Figure 3 (COCO-VOC), Figure 4 (COCO-NONVOC) Figure 5 (VOC12), and Figure 6 (SOC).

2. Class-wise Quantitative Performance

In Table 2 of the submitted paper, we report the average performance of our method for instance co-segmentation on the four datasets. Here we provide the detailed class-wise results. Table 1 reports the class-wise performance of our method on the COCO-VOC and SOC datasets. Table 2 and Table 3 show the class-wise performance of our method on the VOC12 and COCO-NONVOC datasets, respectively. Table 4 of the submitted paper provides the average performance of the variant of our method for object co-localization on the four datasets. Similarly, we report the class-wise performance of object co-localization on the four datasets in Table 4, Table 5, and Table 6.

From the results reported in the tables and the image samples of each class of the four collected datasets, we have some interesting observations as follows. First, our method carries out the co-peak detection, which seeks the discriminative and common parts of objects across image pairs. Objects with more distinctive parts often have better co-peak detection and hence reach better performance, such as categories *airplane* and *horse* on the COCO-VOC dataset and

bear and *parkingmeter* on the COCO-NONVOC dataset. In contrast, categories *cake, tie, sink,* and *refrigerator* on the COCO-NONVOC have less discriminative parts, and the detected co-peaks are not reliable enough, leading to the less satisfactory performance. Second, categories with instances of diverse appearance, such as *cake, suitcase, tie,* and *toothbrush* have the sub-optimal results since detecting reliable co-peaks in the images of these categories is difficult. Third, for object classes with complex shapes, such as *bicycle, chair, motorbike,* and *pottedplant* on the VOC12 dataset and *broccoli, carrot,* or *scissors* on the COCO-NONVOC datast, it is very difficult to retrieve the proposals that fit the object segments well. Therefore, the performance on these classes is inferior.

3. Qualitative Result

The results predicted by our method for instance co-segmentation on some images of the four collected datasets are shown in Figure 7, Figure 8, Figure 9, and Figure 10. It can be observed in these figures that the proposed method is robust to some major difficulties confronting instance co-segmentation, such as large intra-class variations, instance mutual occlusions, and cluttered scenes. For example, although the instances of categories *bird, cow, horse,* and *sheep* on the COCO-VOC dataset often have different colors and shapes, our method can still detect and segment these instances accurately. Our method is also able to handle the variations caused by different scales, poses, or views of instances, such as categories *bear, elephant,* and *zebra* on the COCO-NONVOC dataset. Instance mutual occlusions are often present in the aforementioned classes. Our method can deal with mutual occlusions and produce promising results.

The results by our method for object co-localization are shown in Figure 11, Figure 12, Figure 13, Figure 14, and Figure 15. Similar to instance co-segmentation, our results for object co-localization are also robust to the intra-class variations, instance mutual occlusions, and cluttered scenes. More results can be found at https://github.com/KuangJuiHsu/DeepCO3/.
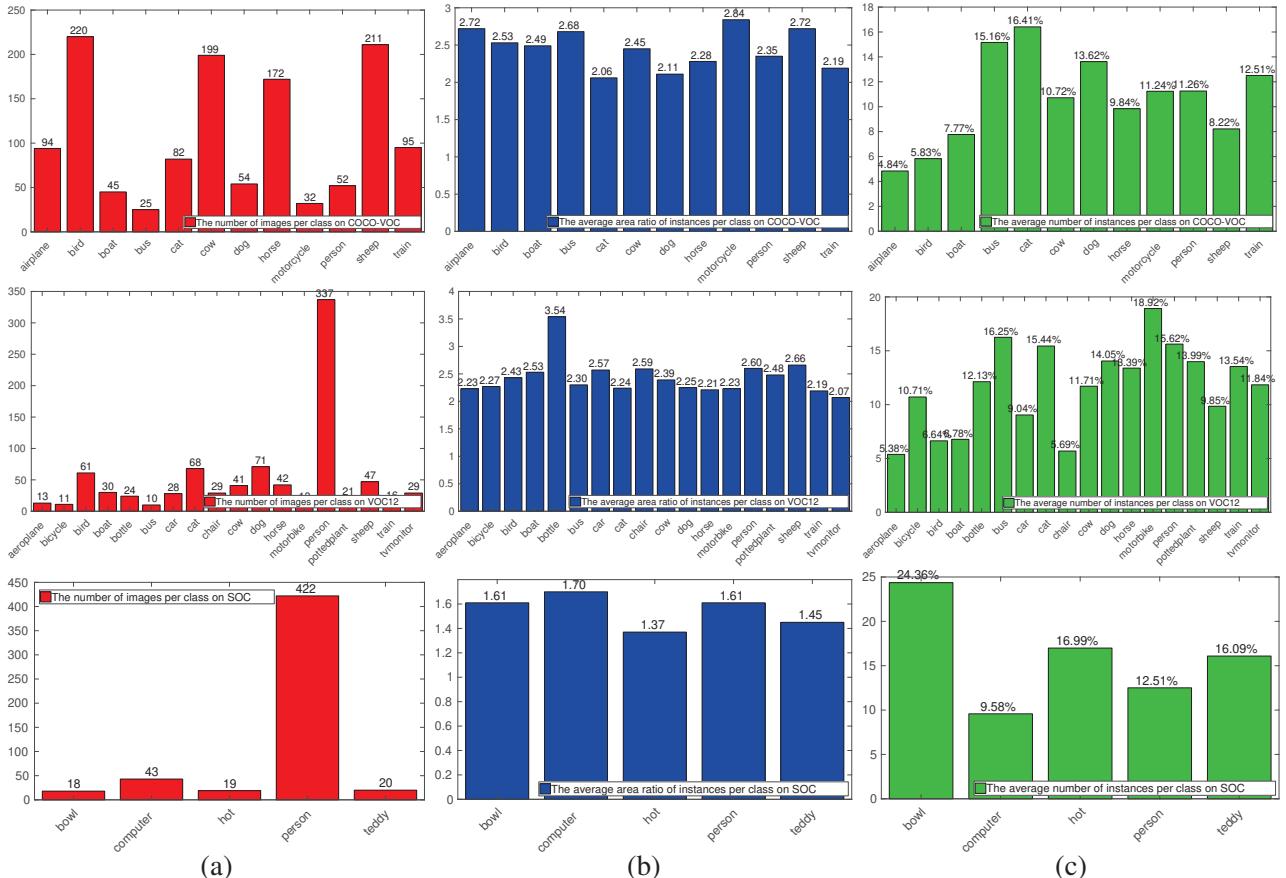
Figure 1. Class-wise statistics of the COCO-VOC (top row), VOC12 (middle row), and SOC (bottom row) datasets, including (a) the number of images, (b) the average number of instances per image, and (c) the average area ratio of an instance to the whole image.
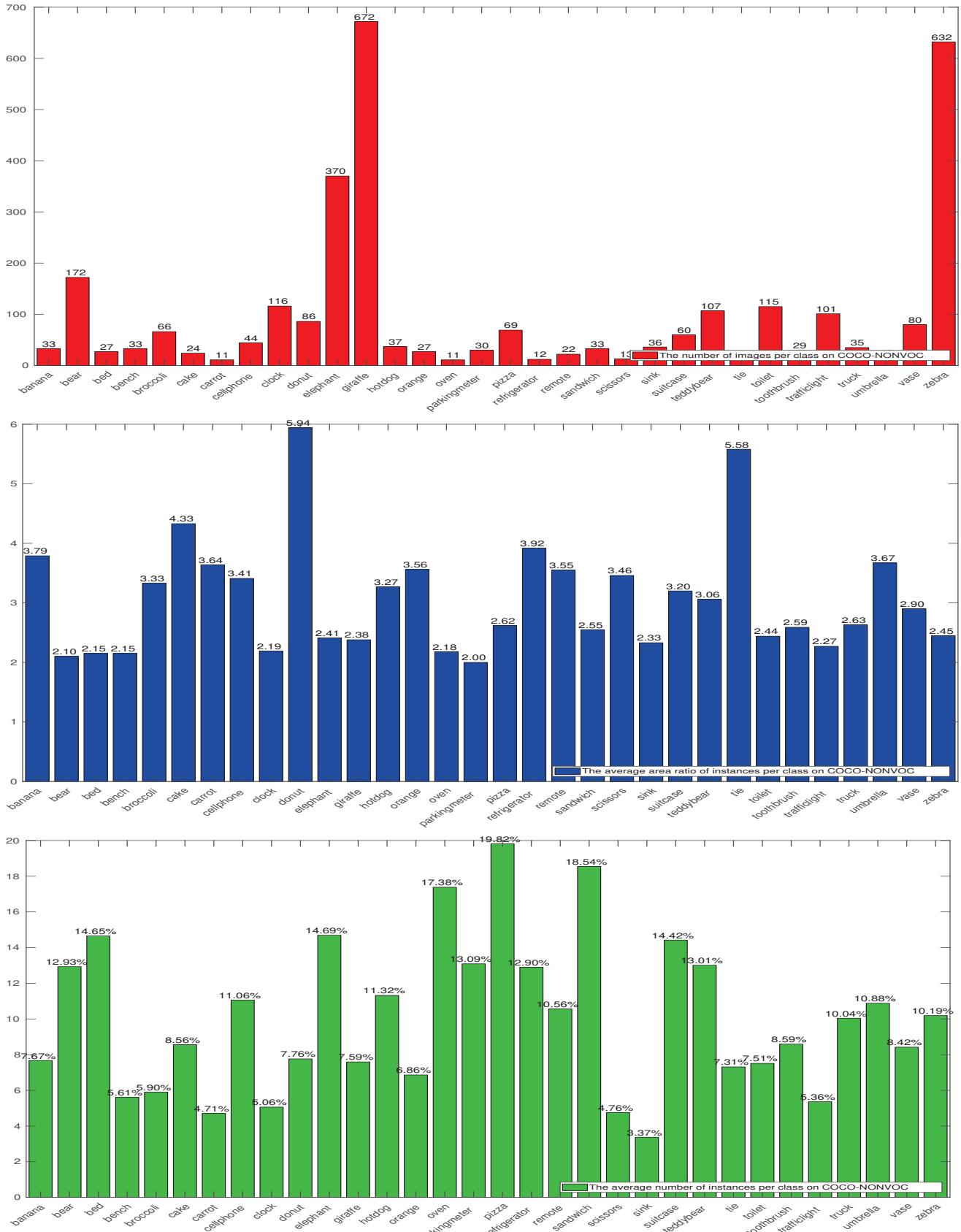
Figure 2. Class-wise statistics of the COCO-NONVOC dataset, including the number of images (top row), the average number of instances per image (middle row), and the average area ratio of an instance to the whole image (bottom row).
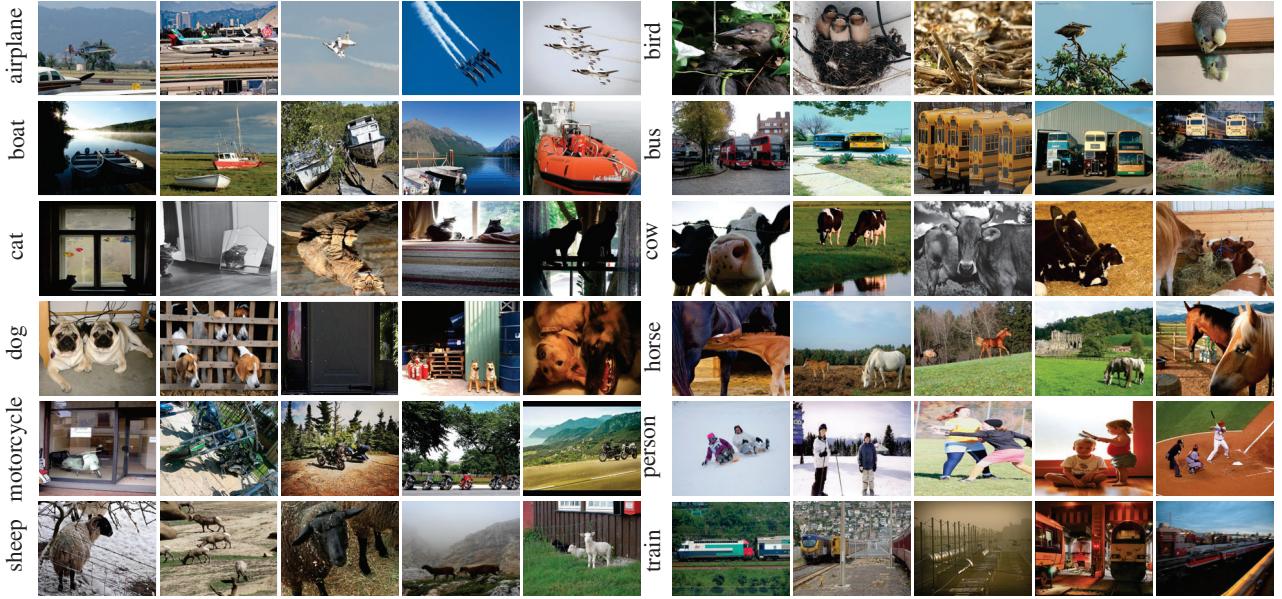
Figure 3. Five image samples of each class of the COCO-VOC dataset.

COCO-VOC	Avg.	airplane	bird	boat	bus	cat	cow	dog	horse	motorcycle	person	sheep	train
$mAP^r_{0.25}$	52.6	62.0	59.3	46.0	33.9	36.6	63.1	65.4	65.3	37.8	48.6	64.7	48.7
$mAP^r_{0.5}$	21.1	26.5	27.9	12.0	10.2	15.7	31.9	32.5	26.7	10.7	11.6	33.9	14.1

SOC	Avg.	bowl	computer	hot	person	teddy
$mAP^r_{0.25}$	54.2	50.8	39.7	79.7	31.4	69.5
$mAP^r_{0.5}$	26.0	31.8	16.0	47.1	9.0	25.3

Table 1. Class-wise performance in $mAP^r_{0.25}$ and $mAP^r_{0.5}$ of our method for instance co-segmentation on the COCO-VOC (left) and SOC (right) datasets.

	Avg.	aeroplane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dog	horse	motorbike	person	pottedplant	sheep	train	tvmonitor
$mAP^r_{0.25}$	45.6	46.7	28.5	52.1	35.3	25.1	43.1	41.3	55.4	22.0	60.9	72.9	67.7	43.5	44.2	30.0	53.5	39.9	57.8
$mAP^r_{0.5}$	16.7	18.0	3.3	29.1	6.0	6.6	4.8	12.1	28	2.2	22.9	38.7	35.9	2.8	8.7	8.9	29.2	20.2	23.2

Table 2. Class-wise performance in $mAP^r_{0.25}$ and $mAP^r_{0.5}$ of our method for instance co-segmentation on the VOC12 dataset.

	Avg.	banana	bear	bed	bench	broccoli	cake	carrot	cellphone	clock	donut	elephant	giraffe	hotdog	orange	oven
$mAP^r_{0.25}$	35.3	27.2	71.3	35.1	31.7	5.8	23.8	13.1	47.1	23.4	33.0	63.1	52.0	28.1	23.3	20.9
$mAP^r_{0.5}$	12.3	6.2	40.6	13.3	8.6	1.0	4.3	2.5	20.6	8.1	9.2	25.5	10.8	5.5	8.1	3.4
parkingmeter	pizza	refrigerator	remote	sandwich	scissors	sink	suitcase	teddybear	tie	toilet	toothbrush	trafficlight	truck	umbrella	vase	zebra
68.0	59.8	8.3	31.1	58.3	16.4	11.1	34.6	41.3	13.8	61.9	21.3	30.5	45.8	33.7	35.5	59.5
17.2	25.4	1.4	12.6	31.5	2.6	1.7	7.8	8.7	1.6	41.5	6.5	4.9	14.1	16.0	14.8	16.5

Table 3. Class-wise performance in $mAP^r_{0.25}$ and $mAP^r_{0.5}$ of our method for instance co-segmentation on the COCO-NONVOC dataset.

COCO-VOC	Avg.	airplane	bird	boat	bus	cat	cow	dog	horse	motorcycle	person	sheep	train
$CorLoc^b$	59.6	69.1	56.8	44.4	60.0	45.1	74.9	79.6	73.3	46.9	48.1	68.7	48.4
$CorLoc^r$	49.6	54.3	55.0	42.2	48.0	41.5	60.3	68.5	55.2	28.1	36.5	62.1	42.1

SOC	Avg.	bowl	computer	hot	person	teddy
$CorLoc^b$	46.6	44.4	37.2	73.7	32.7	45.0
$CorLoc^r$	43.1	38.9	37.2	68.4	25.8	45.0

Table 4. Class-wise performance in $CorLoc^b$ and $CorLoc^r$ of our method for object co-localization on the COCO-VOC (left) and SOC (right) datasets.

	Avg.	aeroplane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dog	horse	motorbike	person	pottedplant	sheep	train	tvmonitor
$CorLoc^b$	48.2	46.2	18.2	55.7	36.7	41.7	20.0	46.4	58.8	10.3	63.4	73.2	81.0	53.8	46.3	42.9	68.1	56.3	48.3
$CorLoc^r$	39.2	30.8	18.2	54.1	30.0	41.7	20.0	32.1	55.9	17.2	51.2	62.0	66.7	15.4	27.9	28.6	63.8	50.0	48.3

Table 5. Class-wise performance in $CorLoc^b$ and $CorLoc^r$ of our method for object co-localization on the VOC12 dataset.

	Avg.	banana	bear	bed	bench	broccoli	cake	carrot	cellphone	clock	donut	elephant	giraffe	hotdog	orange	oven
$CorLoc^b$	41.8	39.4	72.1	44.4	39.4	10.6	33.3	18.2	61.4	22.4	39.5	73.2	61.0	45.9	44.4	27.3
$CorLoc^r$	34.3	36.4	65.7	37.0	24.2	6.1	25.0	9.1	54.5	21.6	45.3	55.4	33.6	29.7	40.7	18.2
parkingmeter	pizza	refrigerator	remote	sandwich	scissors	sink	suitcase	teddybear	tie	toilet	toothbrush	trafficlight	truck	umbrella	vase	zebra
26.7	69.6	50.0	40.9	66.7	23.1	2.8	51.7	46.7	41.7	54.8	24.1	12.9	48.6	33.3	41.3	70.1
33.3	52.2	8.3	50.0	48.5	7.7	2.8	36.7	40.2	25.0	62.6	24.1	19.8	42.9	53.3	41.3	46.5

Table 6. Class-wise performance in $CorLoc^b$ and $CorLoc^r$ of our method for object co-localization on the COCO-NONVOC dataset.
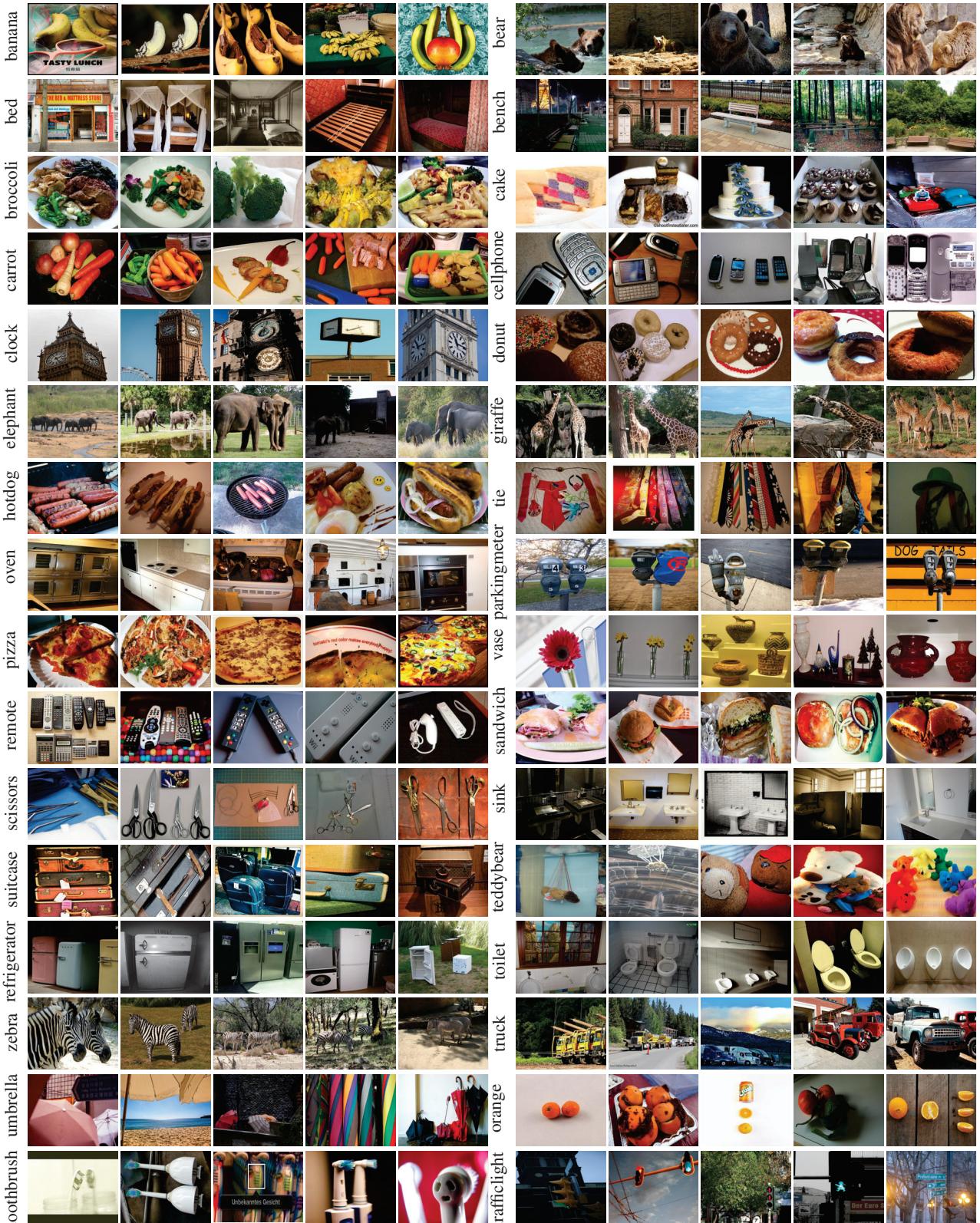
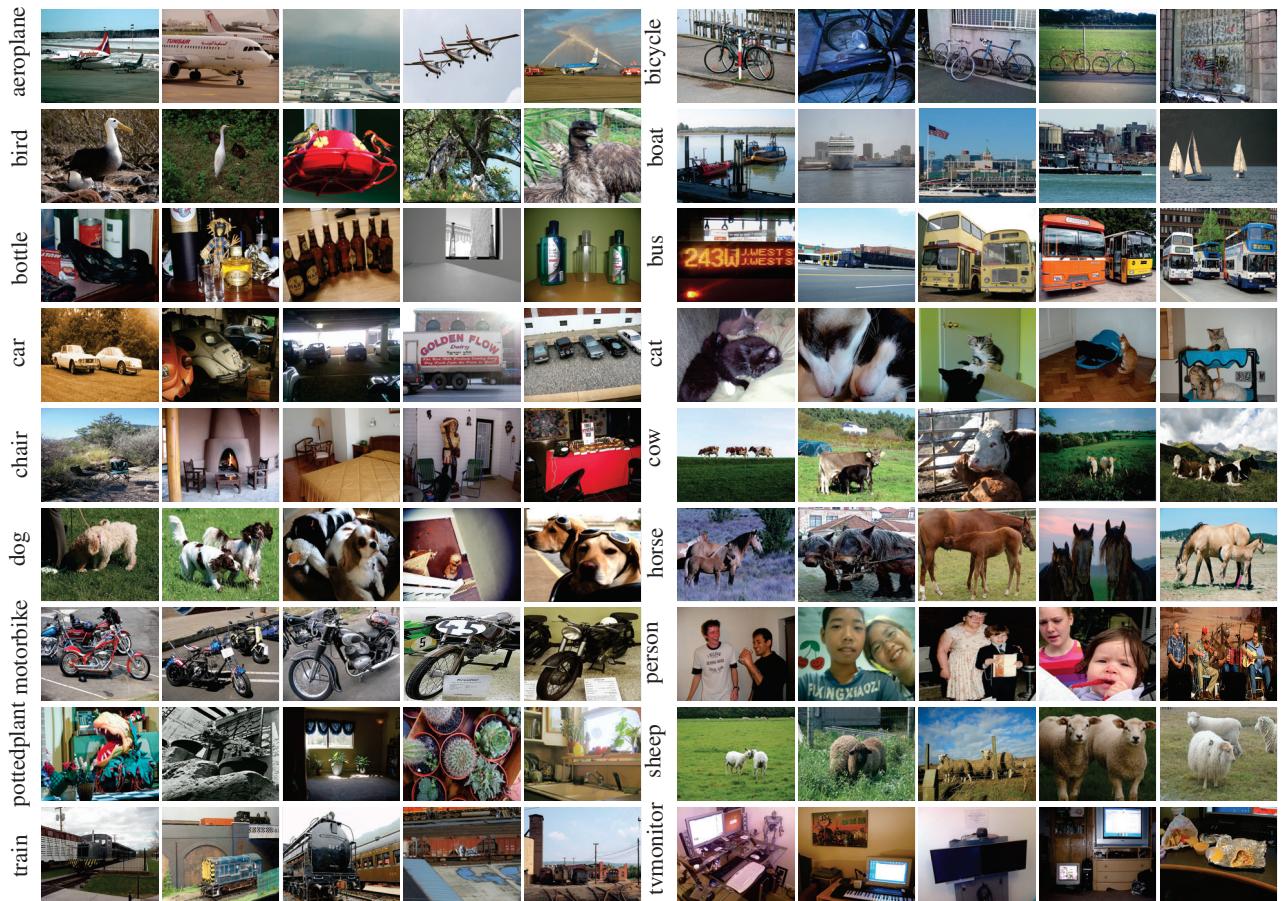Figure 4. Five image samples of each class of the COCO-NONVOC dataset.

Figure 5. Five image samples of each class of the VOC12 dataset.

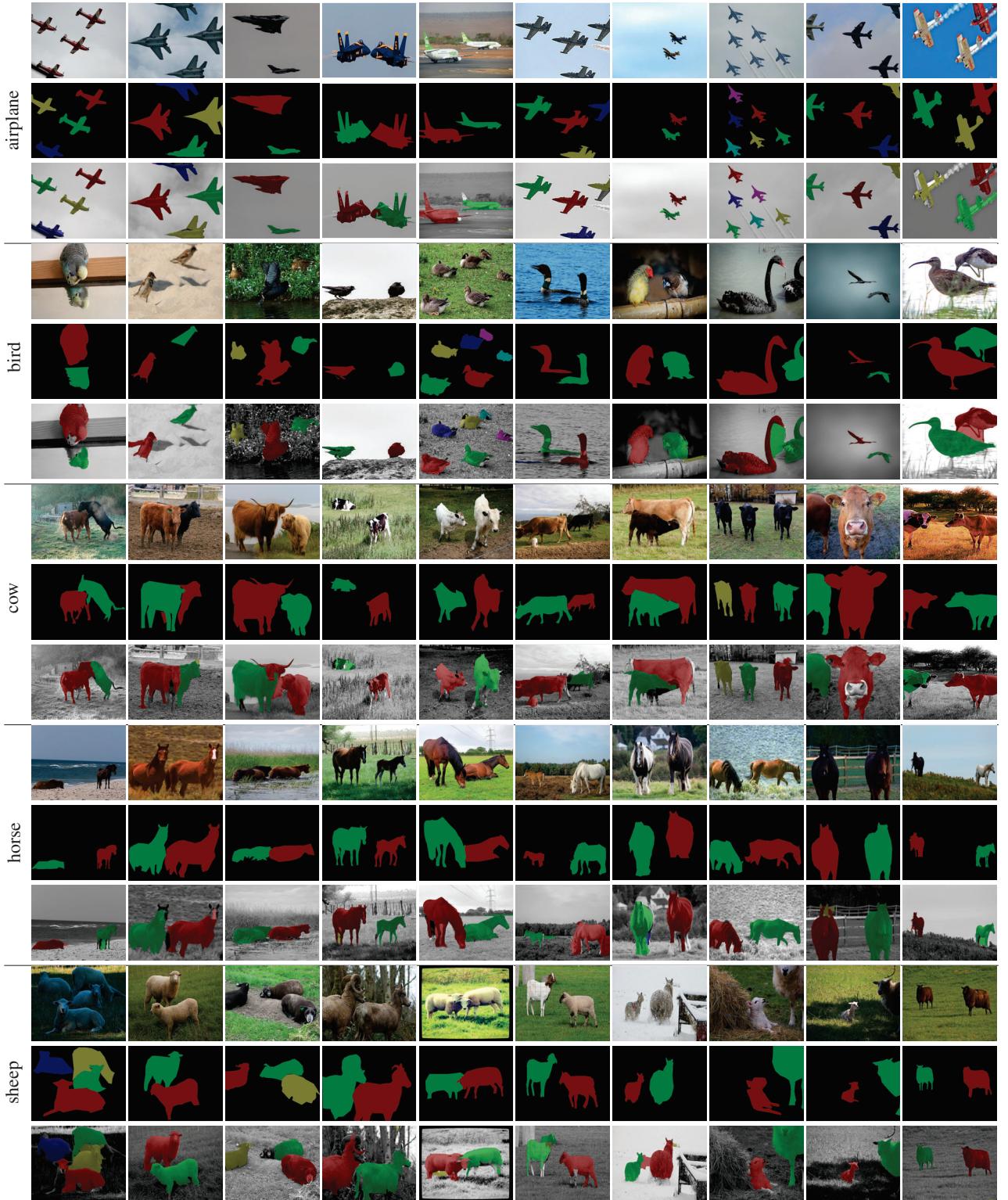Figure 6. Five image samples of each class of the SOC dataset.

Figure 7. Instance co-segmentation on some object categories of the COCO-VOC dataset. Each category is associated with three consecutive rows where from top to bottom, the input images, the ground truth, and the results predicted by our method, are shown respectively.
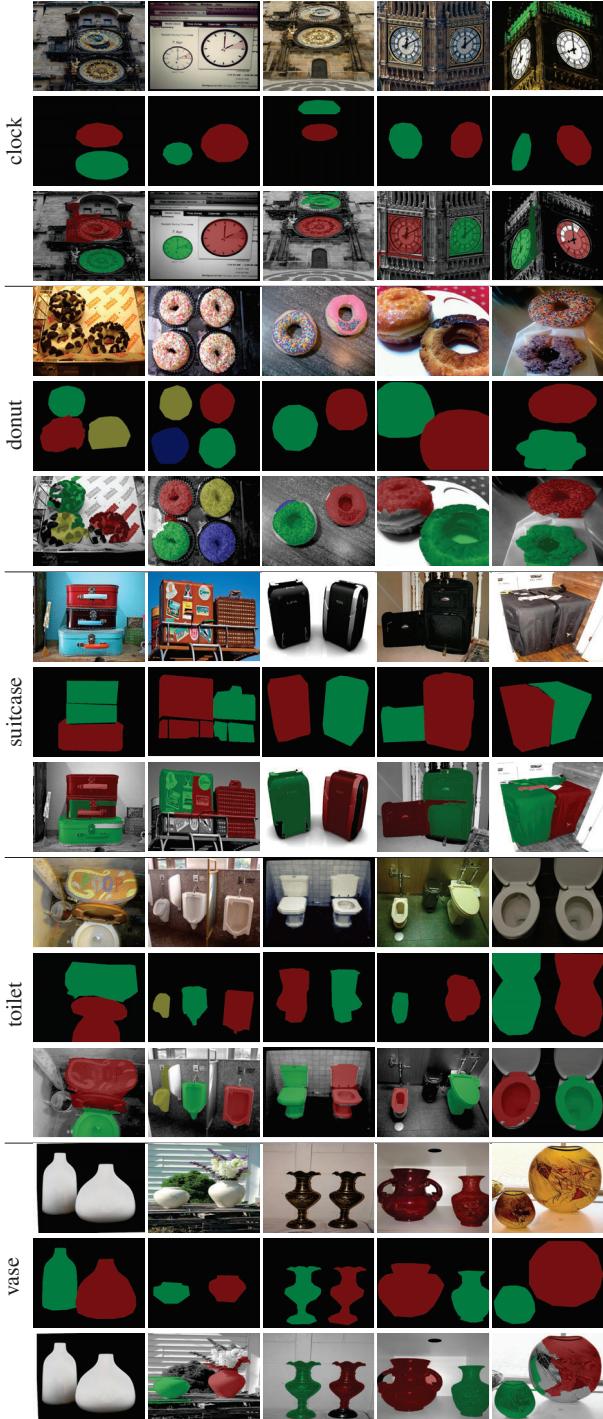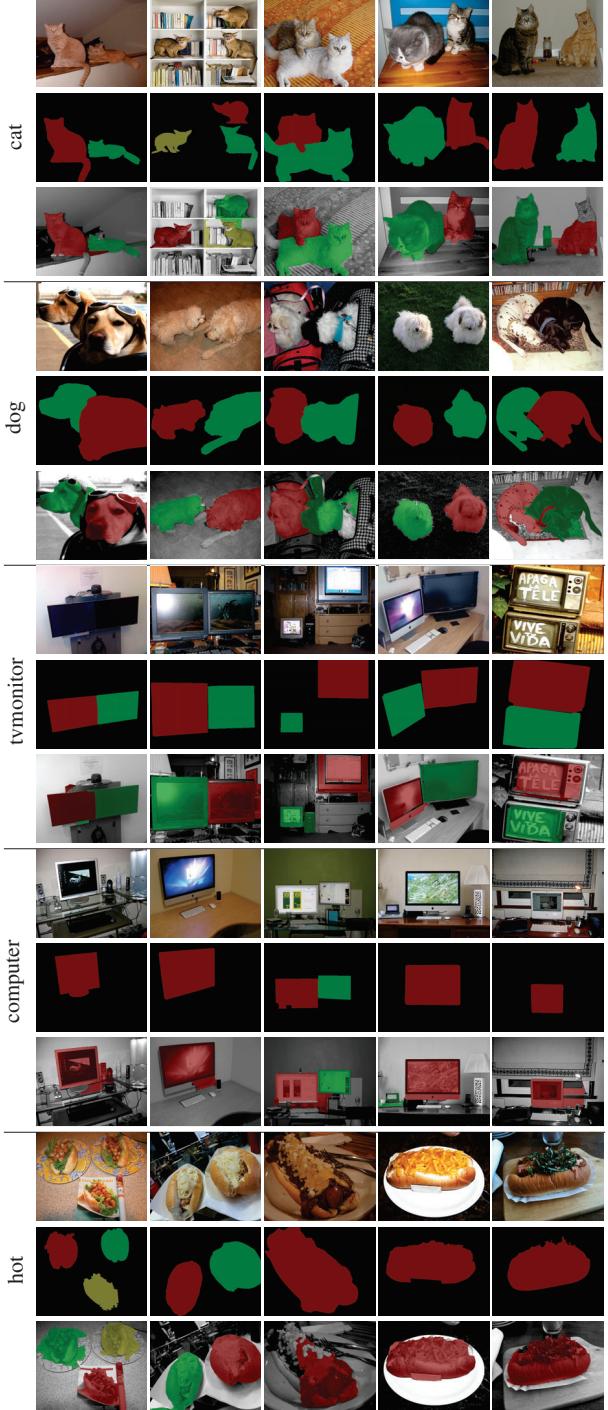
Figure 8. Instance co-segmentation on some object categories of the COCO-NONVOC dataset. Each category is associated with three consecutive rows where from top to bottom, the input images, the ground truth, and the results predicted by our method, are shown respectively.

Figure 9. Instance co-segmentation on some object categories of the COCO-NONVOC dataset. Each category is associated with three consecutive rows where from top to bottom, the input images, the ground truth, and the results predicted by our method, are shown respectively.

Figure 10. Instance co-segmentation on some object categories of the VOC12 and SOC datasets. Each category is associated with three consecutive rows where from top to bottom, the input images, the ground truth, and the results predicted by our method, are shown respectively.
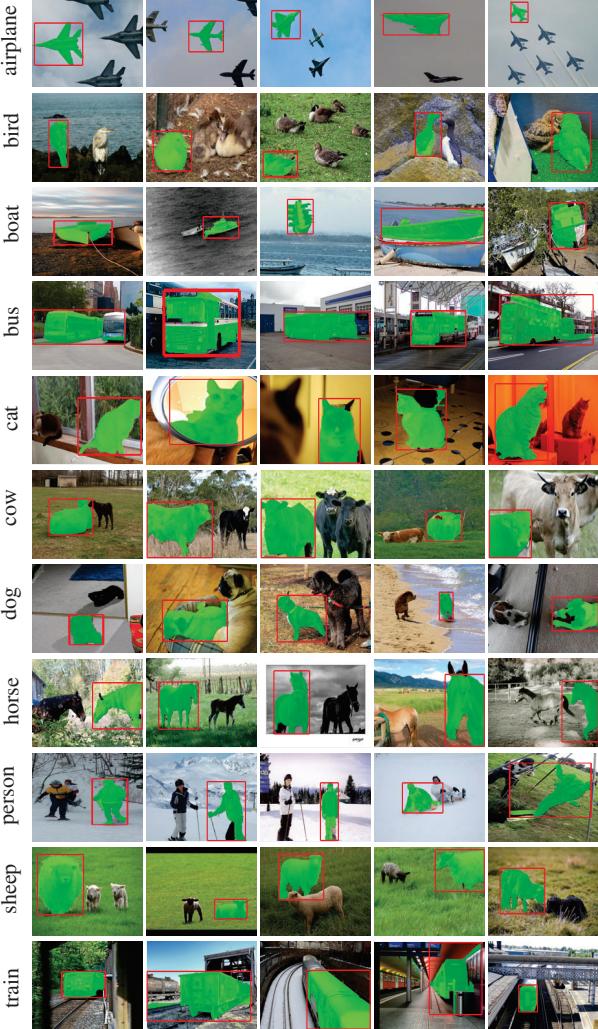
Figure 11. Some examples, one in each row, of the co-localization results by our method on the COCO-VOC dataset.
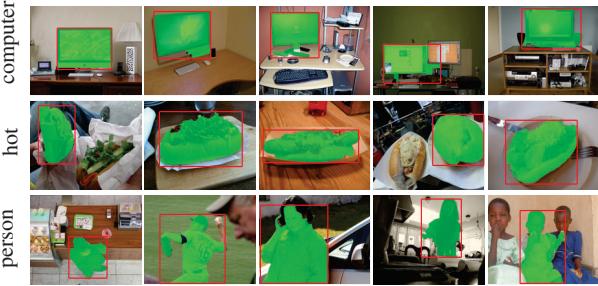

Figure 12. Some examples, one in each row, of the co-localization results by our method on the SOC dataset.
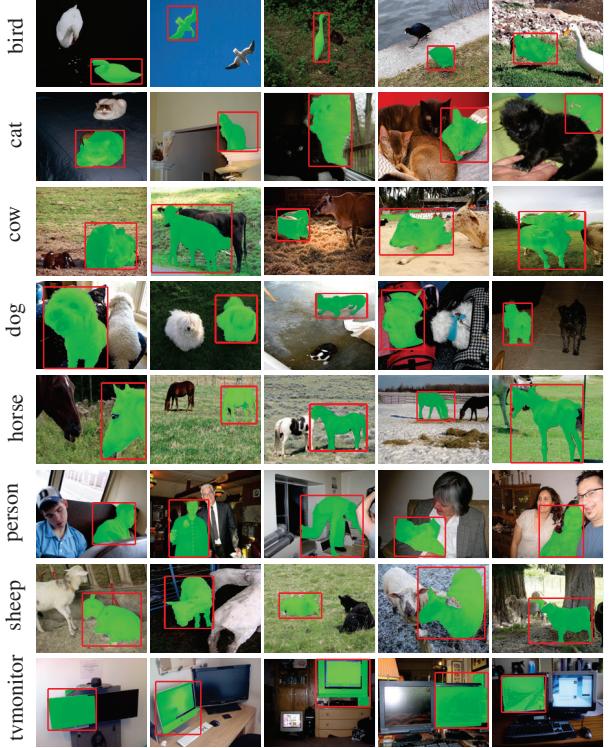

Figure 13. Some examples, one in each row, of the co-localization results by our method on the VOC12 dataset.
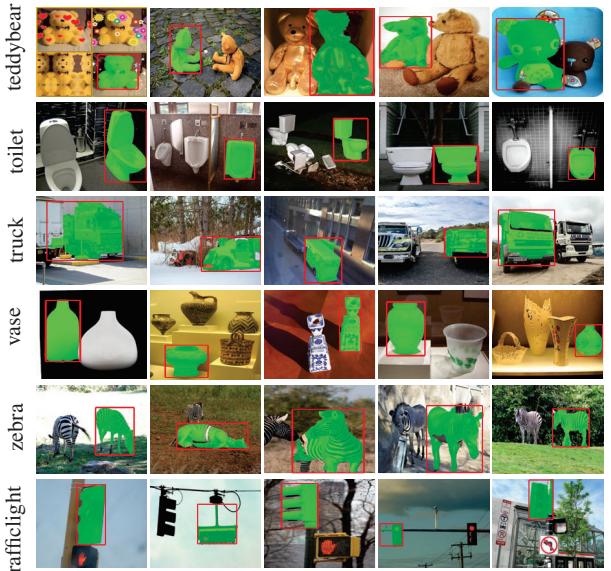

Figure 14. Some examples, one in each row, of the co-localization results by our method on the COCO-NONVOC dataset.
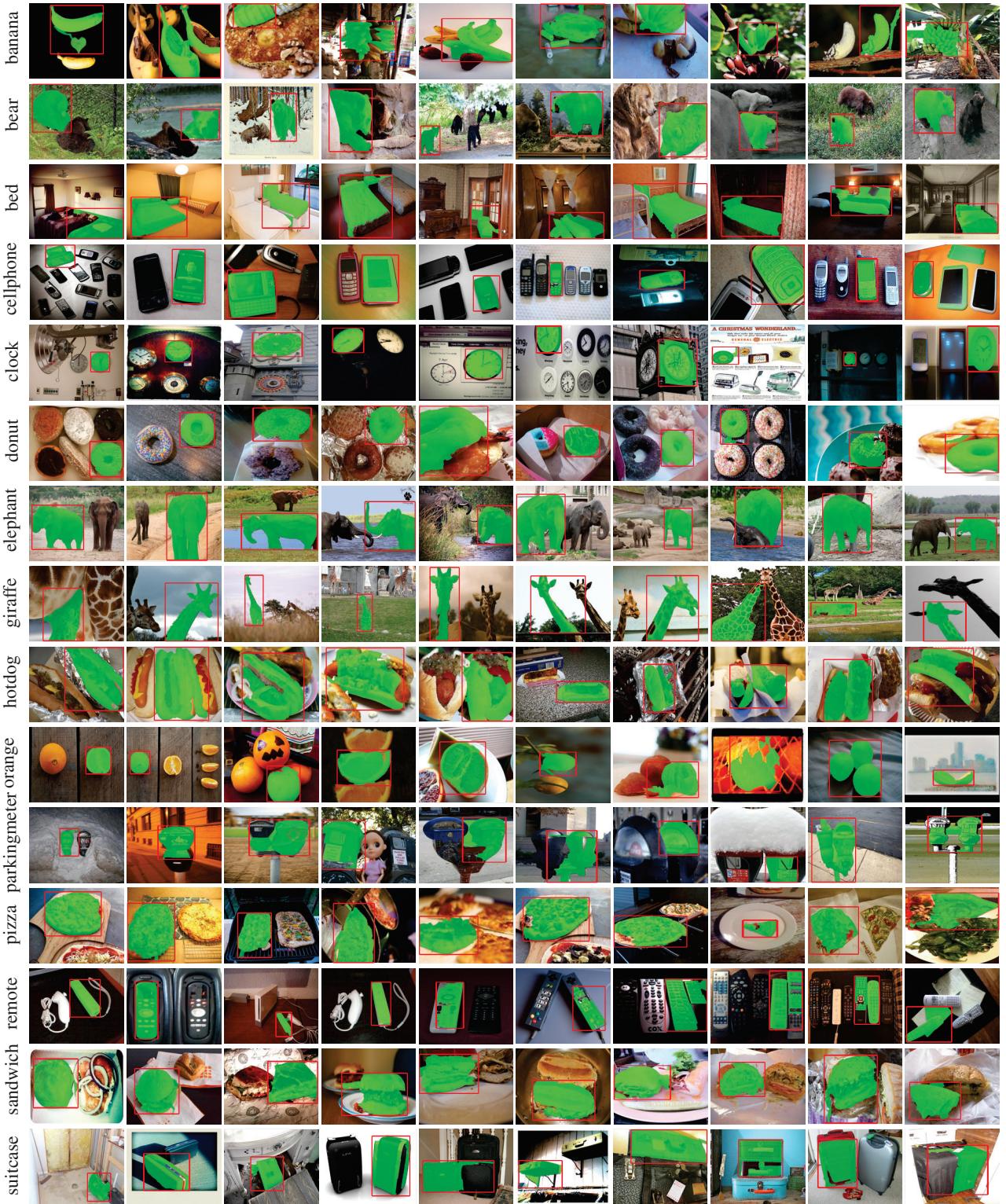
Figure 15. Some examples, one in each row, of the co-localization results by our method on the COCO-NONVOC dataset.