

Adaptive Region Pooling for Object Detection

Supplementary Material

Yi-Hsuan Tsai
UC Merced

ytsai2@ucmerced.edu

Onur C. Hamsici
Qualcomm Research, San Diego

ohamsici@qti.qualcomm.com

Ming-Hsuan Yang
UC Merced

mhyang@ucmerced.edu

1. Discovering Parts

We present additional examples of the object parts obtained by our algorithm in Figure 1 and Figure 2. For those categories with high detection rates, e.g., aeroplane, dog and dining table, the discovered parts of segmentation are discriminative.

2. Object Localization

Table 1 shows the results of coarse region matching on the test set for top $N = \{10, 20, \dots, 50\}$ selected regions. The recall rate is consistently increased as using more proposals are used. There is a significant improvement from top 10 to top 20 for most of the categories (we use top 20 proposals in the experiments).

3. Object Detection

Figure 3 illustrates detection results. Our algorithm is able to detect objects under different scales, lighting conditions, or partial occlusions. Some failure cases are also shown in Figure 4, where most of them are caused by heavy occlusions or objects with similar appearance. Moreover, we present the region proposal results that generate the bounding boxes. It shows that our SVR models are able to select high-quality region proposals.

4. Object Transfer via Exemplars

Figure 5 presents more visualization results for keypoint transfer. The recall-error curves for some categories are shown in Figure 6. Overall, our algorithm performs well against the ESVM [1] method for transferring keypoints on deformable objects such as cow, sheep, and dog. In addition, we present more results for transferring parts and masks in Figure 7 and Figure 8.

Table 1. Recall rate for coarse region matching with top N proposals on the PASCAL VOC 2007 test set.

	Top 10	Top 20	Top 30	Top 40	Top 50
aeroplane	90.2	90.9	91.2	91.2	91.2
bicycle	95.6	97.9	99.1	99.1	99.4
bird	85.0	90.2	92.4	93.3	93.9
boat	85.6	89.0	91.6	92.0	92.8
bottle	61.2	72.3	78.7	80.6	82.3
bus	85.0	90.6	91.6	92.0	92.5
car	83.5	88.7	91.2	92.2	92.8
cat	93.6	95.8	96.4	96.7	97.2
chair	83.6	89.0	92.9	95.2	96.0
cow	87.3	90.6	91.8	94.7	95.1
dtable	93.2	95.6	97.1	98.1	98.5
dog	92.4	96.1	97.1	97.6	97.8
horse	92.5	94.0	94.5	95.7	95.7
motorbike	85.2	90.2	92.3	92.9	93.2
person	73.8	81.1	84.6	86.6	87.7
plant	71.5	80.0	84.8	87.1	88.8
sheep	92.6	96.7	97.1	97.5	98.4
sofa	97.1	98.7	98.7	98.7	99.2
train	91.8	94.0	95.0	95.4	96.1
tv	89.9	93.8	94.5	94.8	95.5
mean	86.5	90.8	92.6	93.6	94.2

References

- [1] T. Malisiewicz, A. Gupta, and A. Efros. Ensemble of exemplar-svms for object detection and beyond. In *ICCV*, 2011. 1



Figure 1. Object parts generated by our algorithm. The most left column shows the union region from all the parts, and the rest of regions are parts.

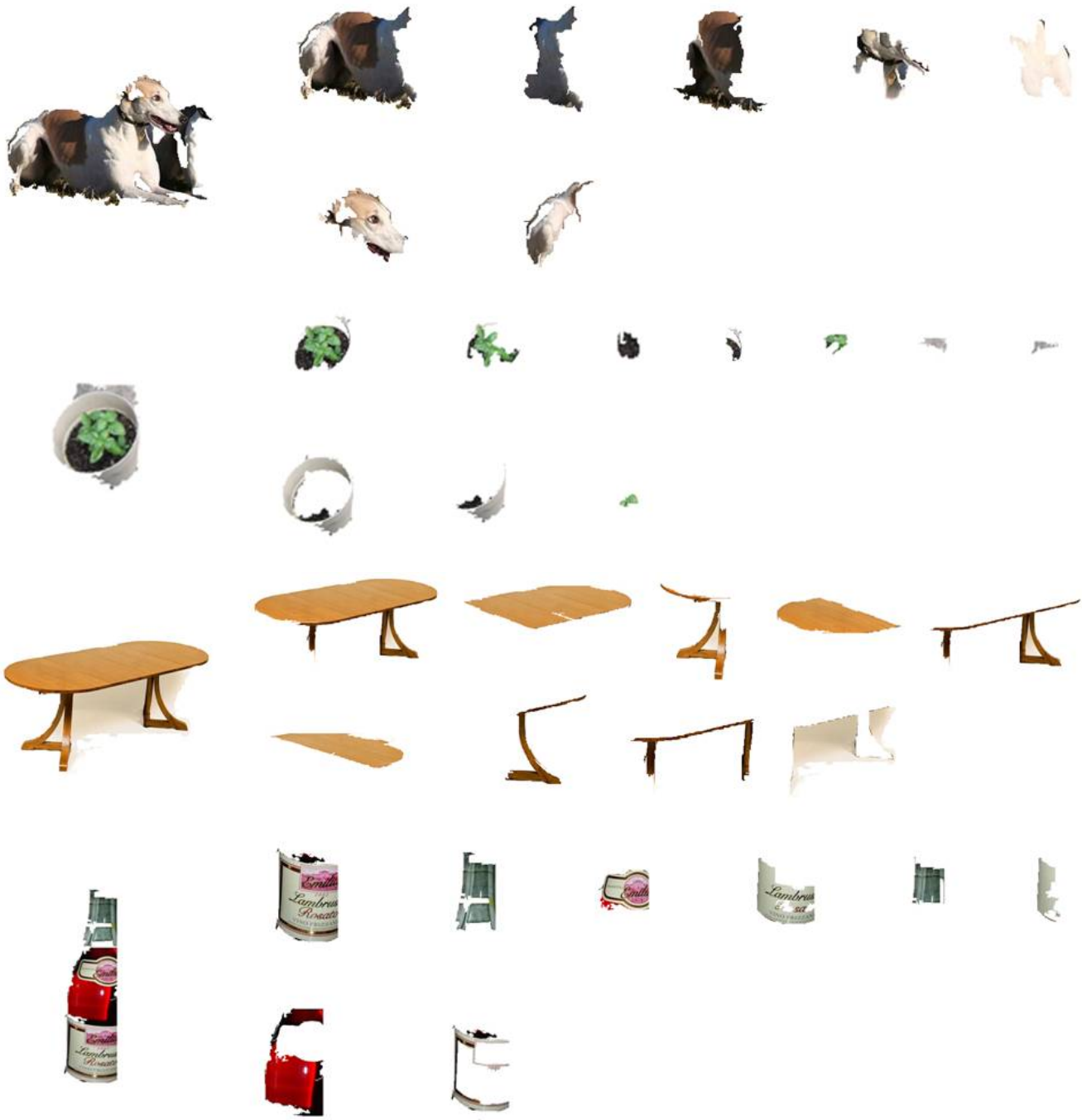


Figure 2. Object parts generated by our algorithm. The most left column shows the union region from all the parts, and the rest of regions are parts.

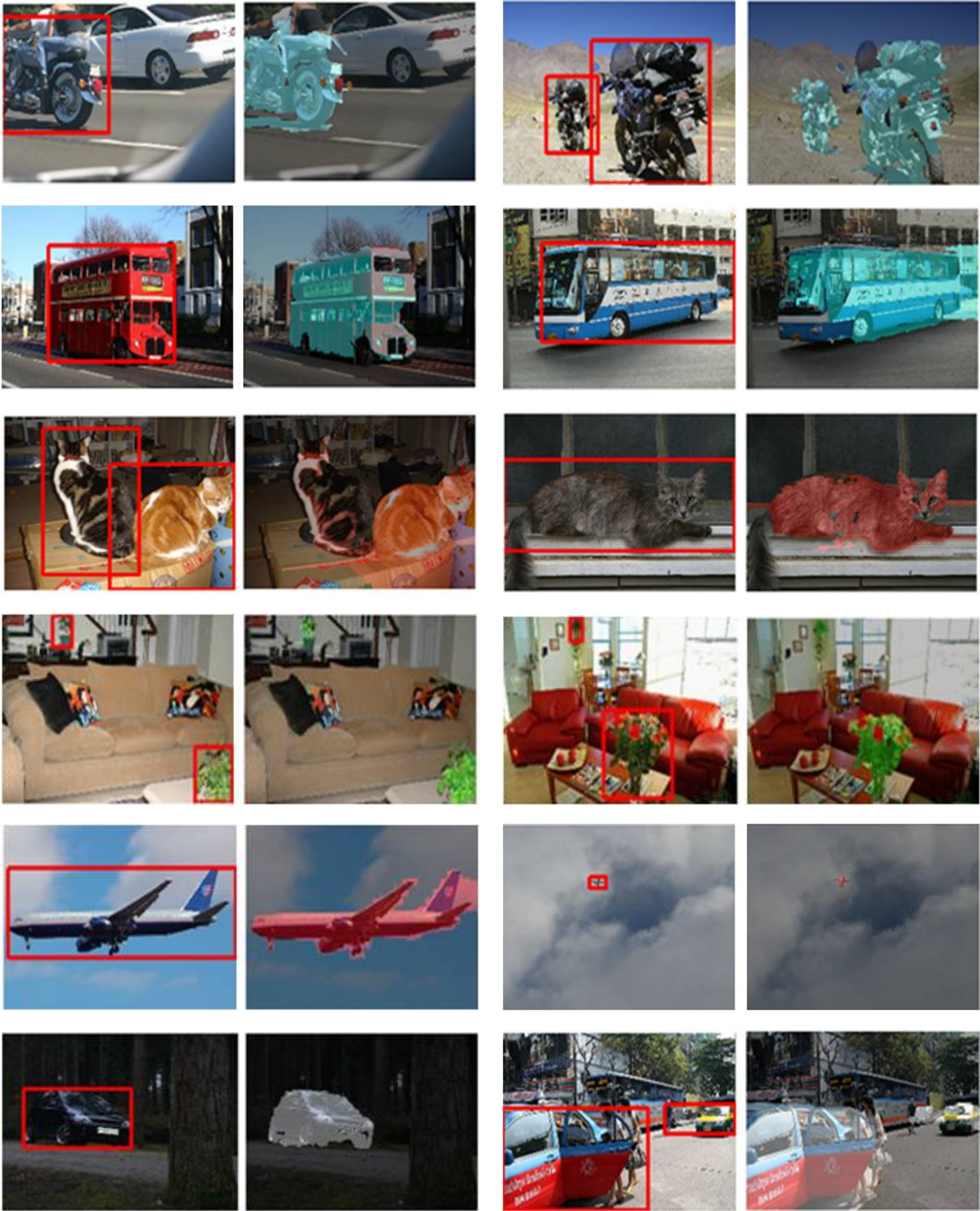


Figure 3. Visualization for detection results on the PASCAL VOC 2007 dataset. Each row presents the result for one category. For each pair of images, the left side is the detection result generated from the region proposal (with color mask) on the right side. The masks for different categories follow the VOC color codes. Best viewed in color.

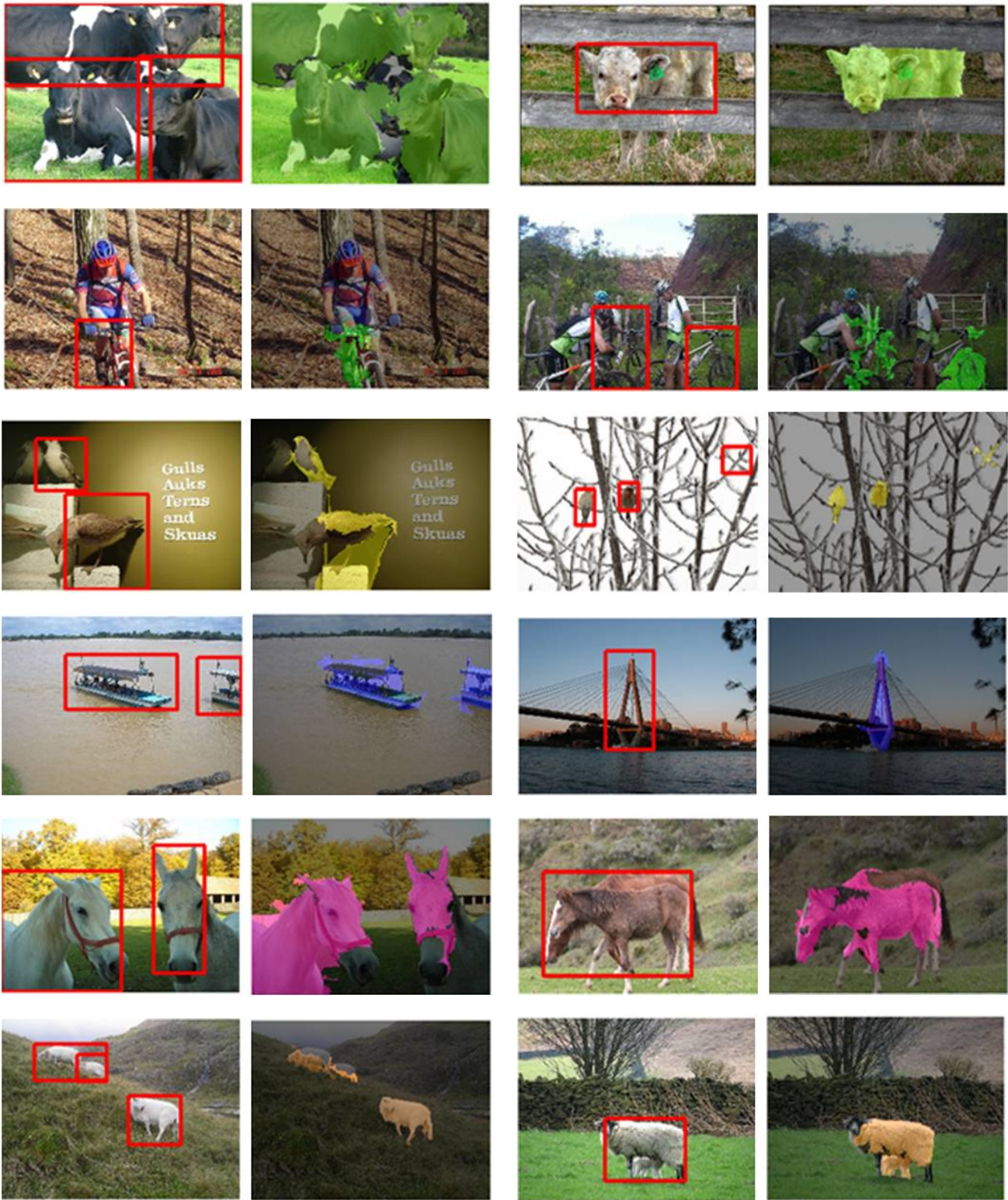


Figure 4. Additional visualization of detection results on the PASCAL VOC 2007 dataset. Image pairs in the right column are some failure cases caused by heavy occlusions (cow and bike: incomplete bounding boxes; horse and sheep: missing objects) or similar appearance to objects (bird and boat). The masks for different categories follow the VOC color codes. Best viewed in color.



Figure 5. Keypoints transfer results on the PASCAL VOC 2007 dataset. For each pair of the result, the right figure is the exemplar (keypoints marked in blue) that transfers keypoints to the detected object (keypoints marked in pink) in the left figure. Best viewed in color with enlarged images.

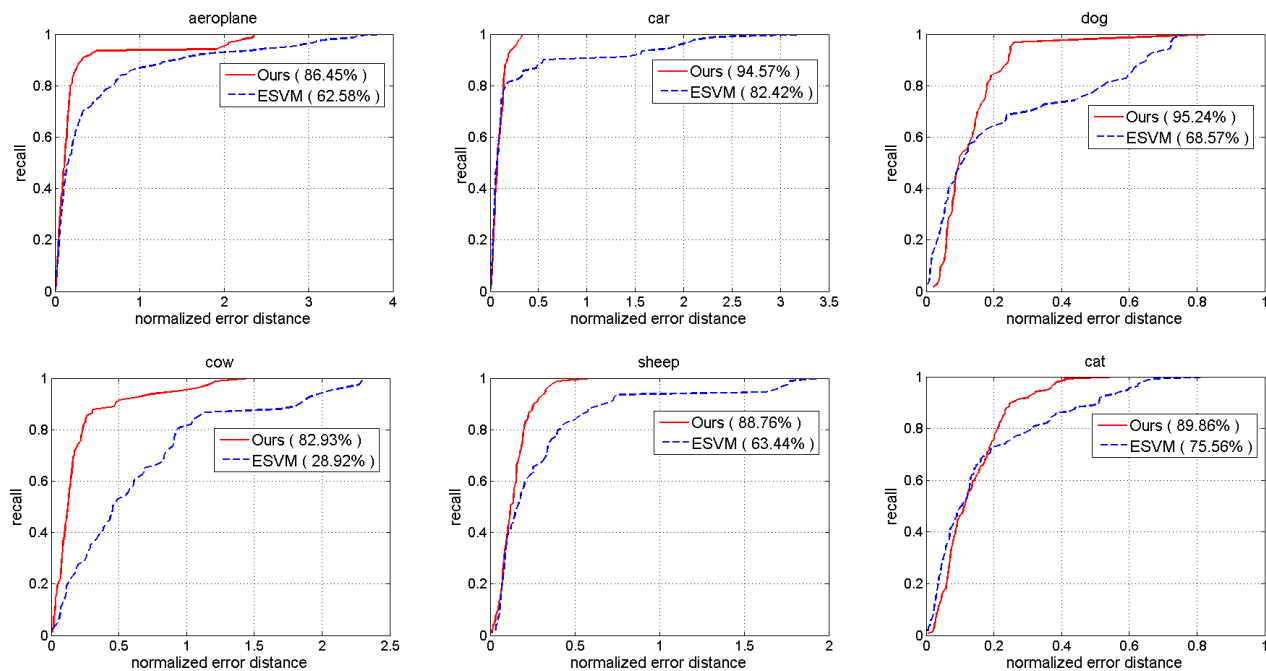


Figure 6. The recall-error curve comparing to the ESVM method for different categories. The number in the legend indicates the recall rate when the error distance is 0.25.

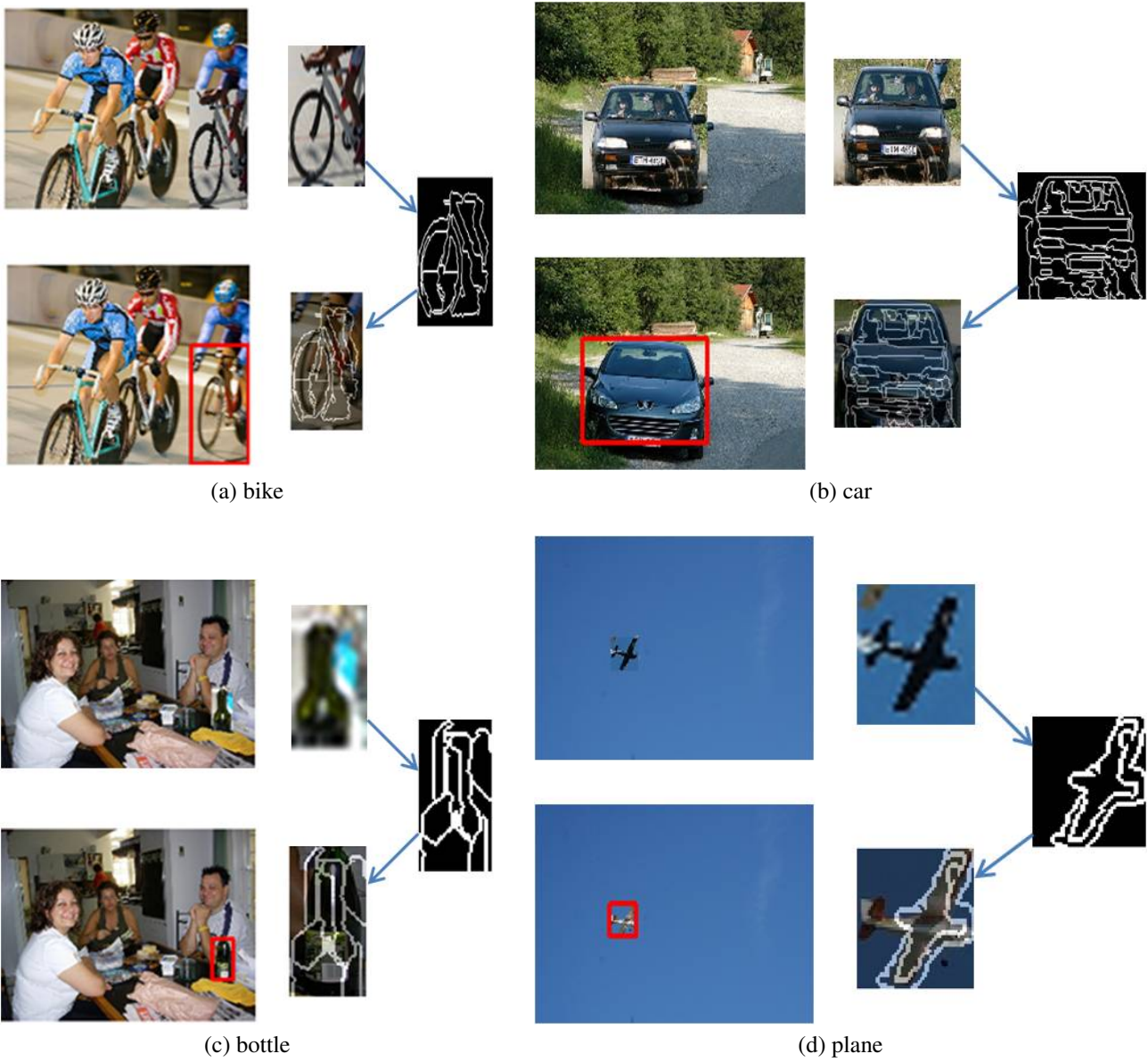


Figure 7. For each category, we show the detected object and the transferred mask. From (a) to (d), the red bounding box is the detected object and the top-left figure shows the transferred mask. The part information of the exemplar is transferred to the detected object, which has a similar region structure with the exemplar. Best viewed in color with enlarged images.



Figure 8. For each category, we show the detected object and the transferred mask. From (a) to (d), the red bounding box is the detected object and the top-left figure shows the transferred mask. The part information of the exemplar is transferred to the detected object, which has a similar region structure with the exemplar. Best viewed in color with enlarged images.