Since there are no real-world semantic part datasets with different occlusion levels, we randomly choose several images with partial occlusion and/or truncation from MS COCO [1] 2014 val. We run DeepVoting on these images and plot the detection heatmap (yellow for high scores and blue for low scores). Note that no ground-truth annotations are available.

Figure 1: Detecting wheels of cars. The left rear wheel of the left car is partially occluded by a chair, and the right wheels of the right car are occluded by the same chair and the left car.

Figure 2: Detecting wheels of a bike. The rear wheel is occluded by a blue bag.

*The first three authors contributed equally to this work.
Figure 3: Detecting license plates of a car. The license plate is occluded by a woman.

Figure 4: Detecting side windows of a car. The rear side window is truncated.

Figure 5: Detecting headlights of a car. The headlight is both occluded and truncated.
Figure 6: Detecting wheels of a motorbike. Most of the wheels are occluded by the legs of two men.

Figure 7: Detecting headlights of a car. The headlight is occluded by the head of a man.

In these heatmaps, we can see strong responses in the area of occluded semantic parts. These heatmaps are then used to generate bounding boxes for the target part, followed by bounding box regression and non-maximum suppression.

References