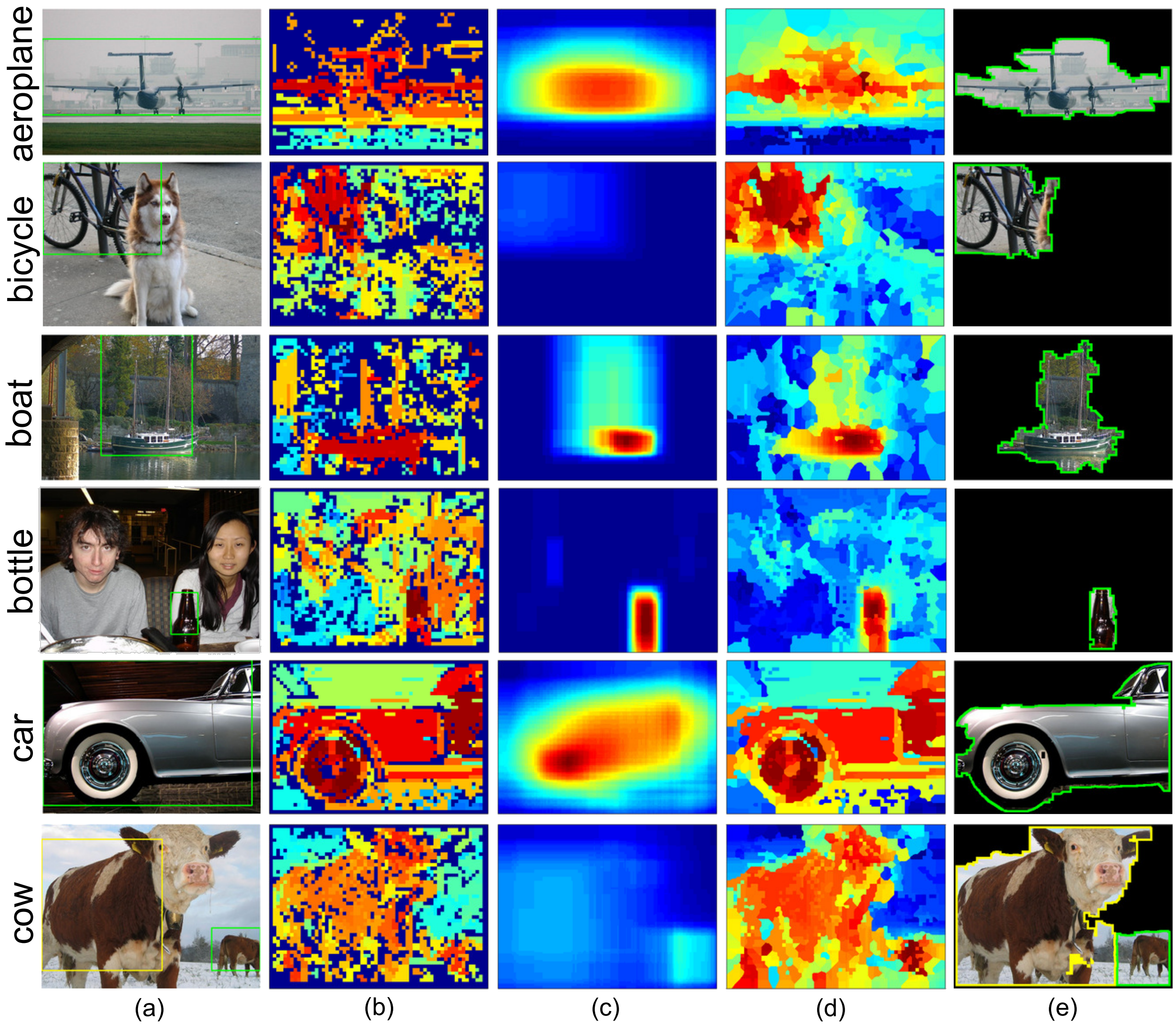
Augmenting Deformable Part Models with Irregular-shaped Object Patches
Supplementary Material

Roozbeh Mottaghi
University of California, Los Angeles
roozbehm@cs.ucla.edu

1. Additional Detection Results



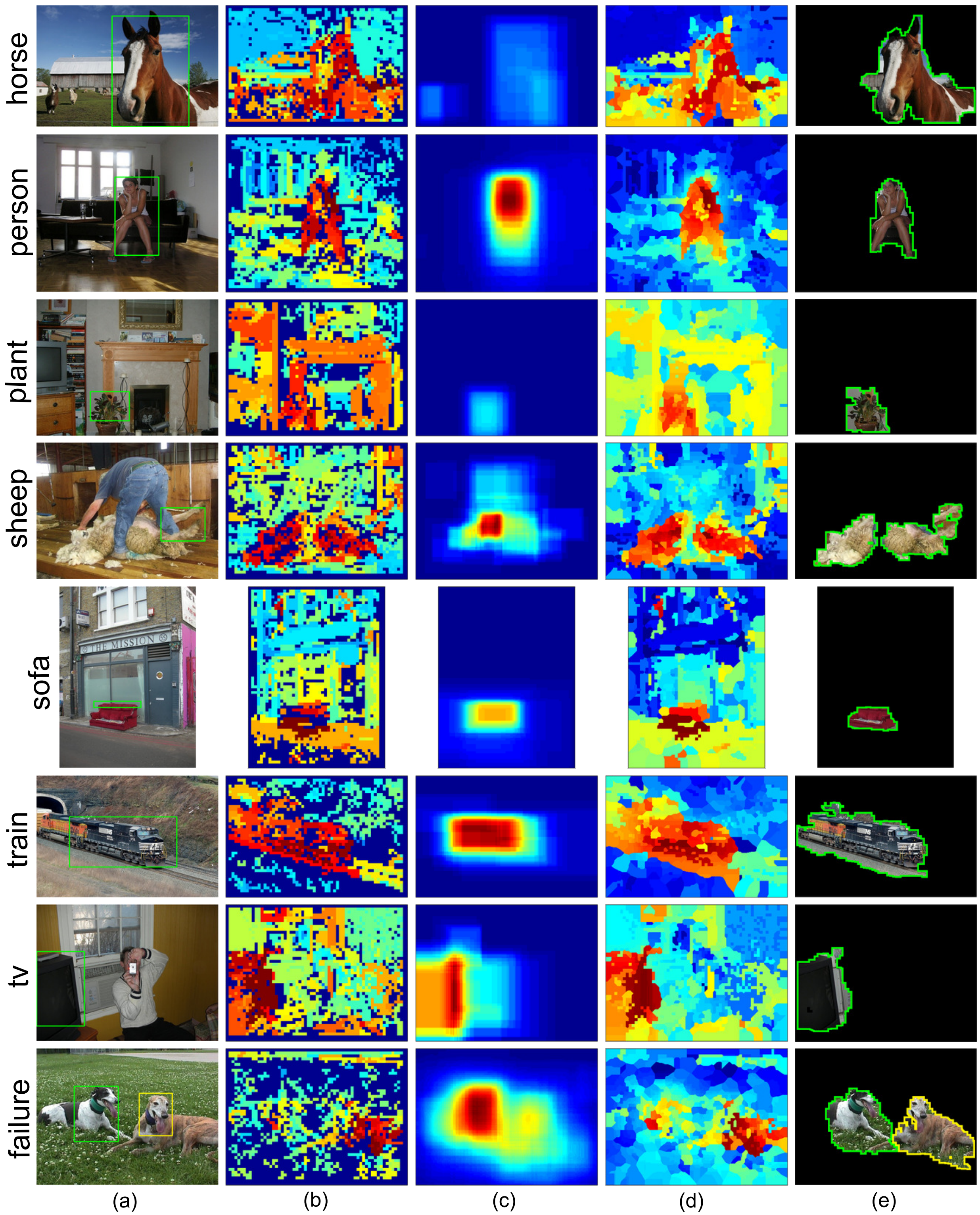


Figure 1. (a) Original detections of the base detector, (b) Heat map of the HOG-bundle term (P_h), (c) Heat map of the Superposition term (P_s), (d) Smoothed heat map of the likelihood map (P_{lm}), (e) Final output of our method. The boundary for each object instance is illustrated by a different color (corresponding to the color of the detections in column (a)).

2. HOG-bundles

Some parts have been copied from reference [18].

HOG-bundles are computed by grouping neighboring HOG cells which share similar properties using the following grouping rules: Two HOG cells are grouped if they are neighbors in the grid image and satisfy the following criteria:

1. The difference between the feature vectors of two cells are small, as computed by the χ^2 distance function over the feature vectors.

2. The orientation with the maximum magnitude should be similar for two HOG cells. Usually, the cells that belong to a part have a similar orientation.

3. HOG cells with gradients in many directions will not be grouped, since they usually correspond to randomly textured areas such as grass. This is quantified by the squared difference $|\Omega - \omega|^2$, where Ω is the maximum magnitude of the orientation part of the feature vector, and ω is the mean of the magnitudes. However, the cells that correspond to uniform intensity regions are grouped (low-magnitude gradient cells).

The HOG-bundles are built by choosing an arbitrary HOG cell in the image and checking if its neighbors satisfy the grouping criteria. If they do so, they are grouped and the neighbors of neighbors are checked until all of the cells in the image are processed. Each HOG-bundle has the following attributes: position of the center ($\mathbf{r}$), width (w), height (h), and orientation (θ). The orientation θ is defined as the mean of the orientations with the maximum magnitude in the histograms of the constituent cells. Height h is the length of the longer side of the bundle bounding box oriented along the dominant orientation. Similarly, width w is the length of the shorter side of that box.

The HOG feature has 31 dimensions if the orientations are quantized to 18 bins. The first 18 components correspond to contrast sensitive gradients i.e. the gradient orientation is specified according to the contrast direction. The next 9 components correspond to contrast insensitive gradients i.e. the gradient orientation does not depend on the contrast. The last four components are related to the gradient magnitudes. Dimensions 1 to 9 correspond to half of a circle, and dimensions 10 to 18 correspond to the other half.

To construct the bundles, the HOG feature vector is divided into two parts based on the orientations corresponding to different contrasts. The grouping rules are applied to these parts separately. Hence, two sets of overlapping HOG-bundles are generated that we refer to as *positive* and *negative* bundles in our paper. For the first grouping criteria, the χ^2 distance is applied to the first nine components

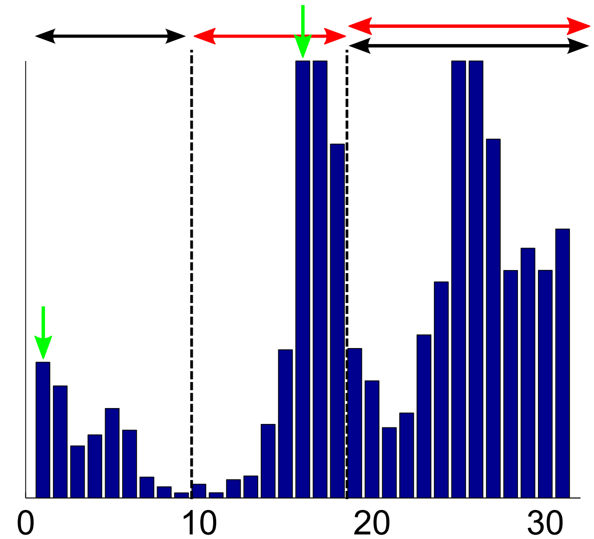


Figure 2. Different parts of a HOG feature vector used for constructing HOG bundles are shown. The black and red arrows represent different parts of the feature vector that the first grouping rule is based on. The dominant orientations of the HOG cell are represented by the green arrows.

plus the last thirteen components (forming a vector with 22 components). Then, it is applied to the second nine components plus the last thirteen components. The black and red arrows in Fig. 2 represent these different divisions. Regarding the second rule, each HOG cell has two dominant orientations, one is computed according to the first nine components and the other one is computed from the second nine components. The orientations with maximum magnitude are shown by the green arrows in Fig. 2. Also, the third grouping rule is applied to dimensions 1 to 9 and 10 to 18 separately.

To group the HOG cells and construct the HOG-bundles, an image is processed twice independently i.e. to compute the positive bundles (corresponding to dimensions 1 to 9 of the feature vector), no information from the construction of negative bundles (corresponding to dimensions 10 to 18 of the feature vector) is used.

3. Instance-based Pixel-wise Scoring

PASCAL evaluation method is not suitable for our method because it is based on 50% intersection over union overlap between the bounding boxes, while our method is able to generate object masks. The segmentation evaluation methods are not suitable for our method either because they evaluate only the correctness of the object category and ignore object instance information. Therefore, we design a new evaluation method such that pixel-wise mismatch between the generated object masks and the groundtruth is penalized.

We sort all of the generated detections for all of the images by their score. Note that these detections correspond to a particular object category. We start from the top-scoring detection and find the groundtruth object segment that has the highest overlap with it. Intersection over union between the segments is used as the measure of overlap. The matched pixels are true positives, and unmatched pixels of the detection contribute to the false positives. A groundtruth segment can only match to one of the detections so we remove the groundtruth object from the available segments set when we find a match for it. For some of the detections, we can not find any match. All of the pixels of these detections are considered as false positives.

To compute the precision-recall curve, we vary a threshold over the detection scores, where only the detections whose score is above the threshold contribute to the true and false positive computations. Note that all of the pixels of a detection have the same score.