Supplementary Material: Object Detection and Parsing Results

Fig. 1 and Fig. 2 show results for object detection and parsing with the suggested randomized max-margin compositions (see also Fig. 1 and Fig. 6 in the main paper submission). The results in Fig. 2 show the ability of our approach to handle intra-class variations and viewpoint changes.

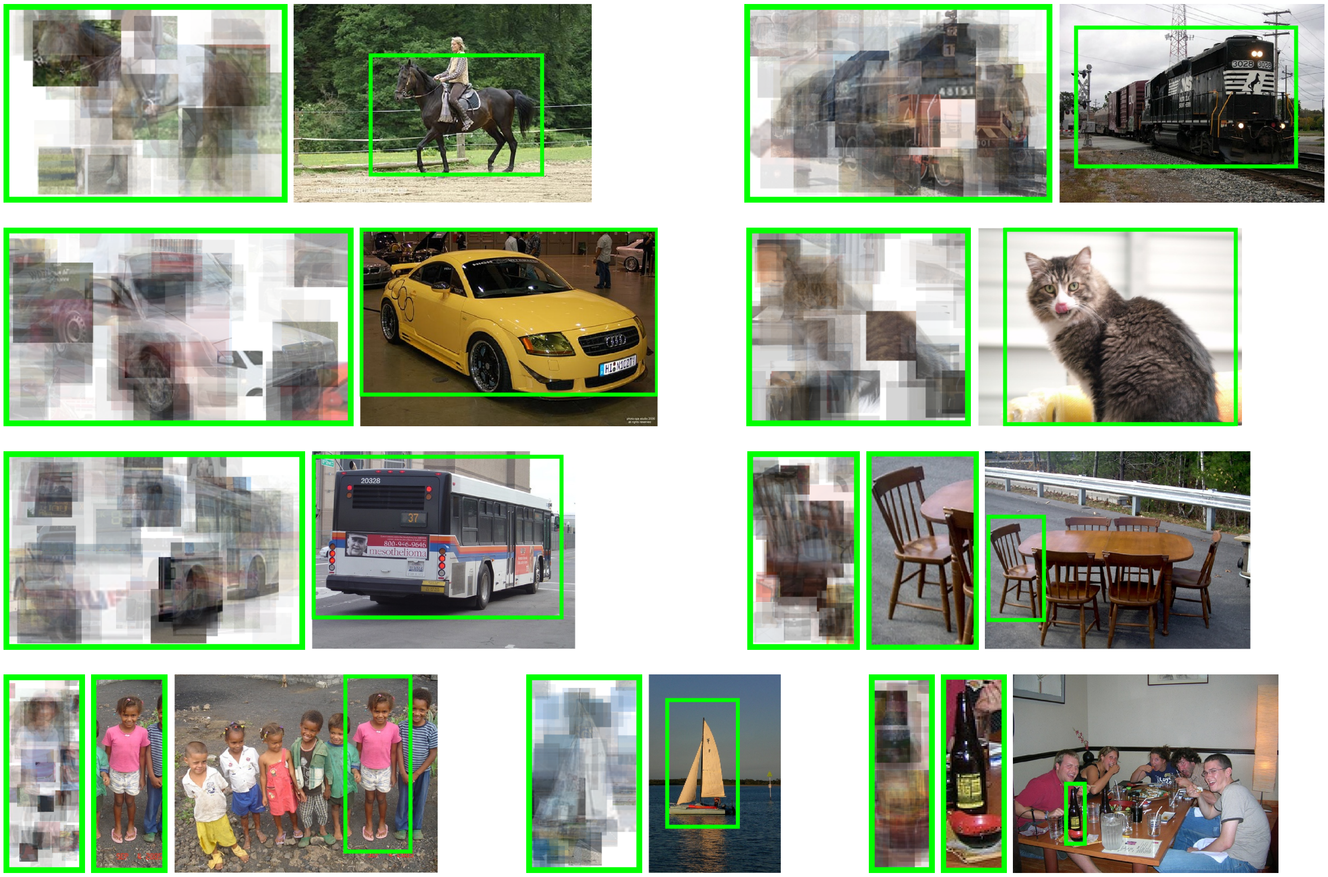

Figure 1. Object detection and parsing with randomized max-margin compositions as shown in Fig. 1 and Fig. 6 in the main paper submission. For a true positive detection in a test image we show (from left to right) the parsing result provided by our algorithm, for small detections the corresponding cropped out image region and the detection (green box) in the full-sized test image.

Fig. 1 & 2. Reconstruction Process. The figures show the result of applying the recognition process detailed in Fig. 2 and Sec. 3 in the main paper: First we extract HOG descriptors x_j and run part classifiers $h_i(x_j)$ from Eq. 4. Then we pool part responses into $\pi_i(\nu)$ using Eq. 3 before running the composition classifiers $f_k(\cdot)$ (Eq. 1). Finally we evaluate $g(F(\cdot))$ (Eq. 2) to combine all compositions using the non-linear classifier.
The parsing results at the left then show the responses of compositions and their constituent parts. The non-linear classifier $g(F(\cdot))$ weights the compositional classifiers $f_k(\cdot)$ which in turn activate their constituent part classifiers $\pi_i(\nu)$ with weights w_k^T. The weights indicate the importance for separating positive and negative training samples. For each part i and site ν, we go to the location of the part x_j in the test image that wins the pooling of Eq. 3. At this location in the test bounding box $\mathcal{I}$ we then place the single positive patch x_p from the training data that defines the i-th part classifier. The transparency of this training patch is the importance that the compositional model assigns to it, i.e., it is proportional to $g(F(\mathcal{I}))\, f_k(\gamma_k(\mathcal{I}))\, w_{k,i}\, \pi_i(\nu)$.

Observations Fig. 1: Parsing tries to explain test data using training samples. Due to large intra-class variation in PASCAL VOC, test samples are quite diverse from training data and we observe generalization artifacts: e.g. the radiator grill of the

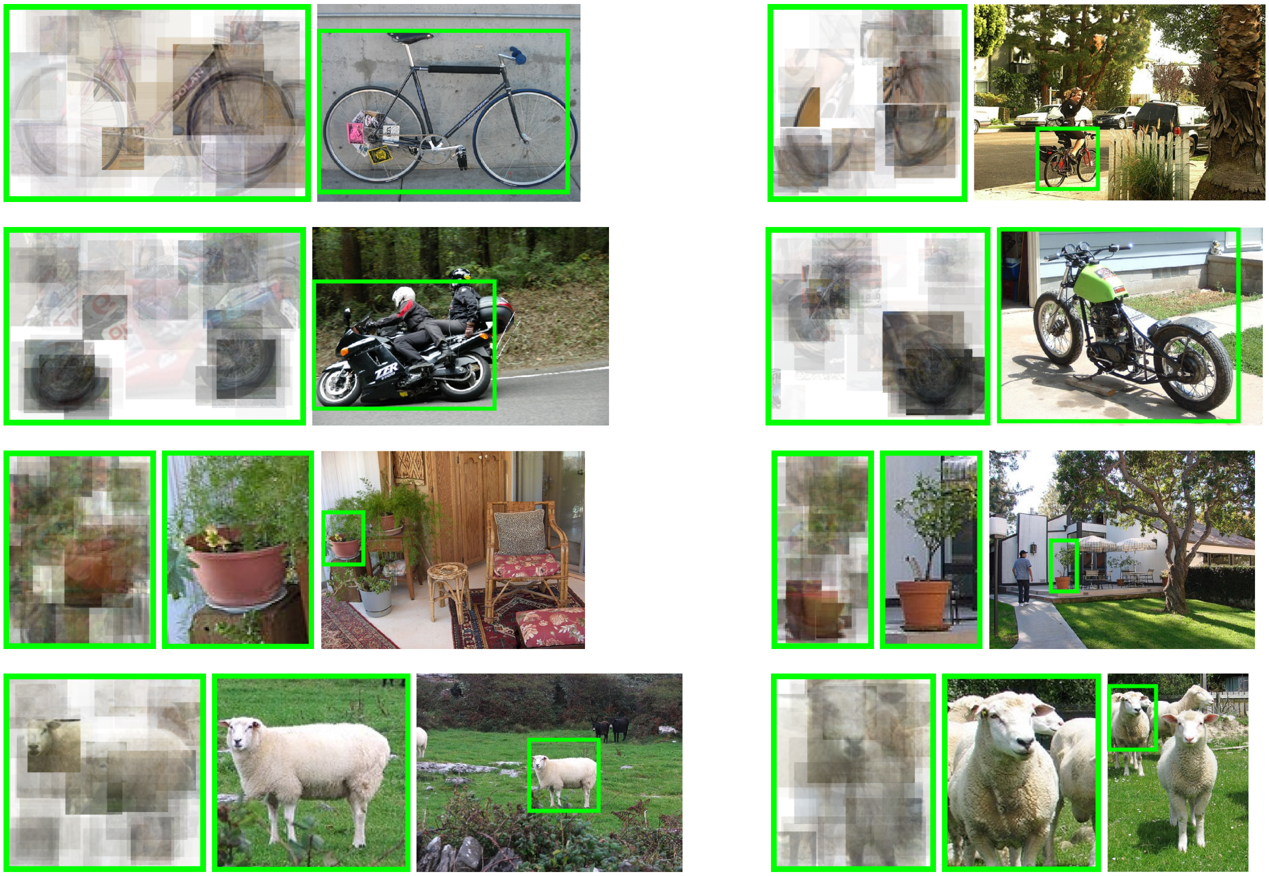

Figure 2. Object detection and parsing with randomized max-margin compositions as shown in Fig. 1 and Fig. 6 in the main paper submission as in Fig. 1. Each row shows two example detections for the same category thus illustrating how the model deals with large intra-class variabilities and viewpoint changes.

car is changing the brand from “Audi” to “Mercedes”; the t-shirt of the little girl (last row, left) is changing its color from pink to blue; the train in the first row slightly changes in style from a modern locomotive to a steam train.

Fig. 2 shows the importance the model assigns to different object regions: For the two motorbikes (second row) the middle part of the motorbikes, which is often covered by a sitting person is assigned lower weight. This region is less reliable than the rest of the bike, as it is covered by bike riders in some samples, thus leading to larger variability. Similar behavior can be seen for the bicycles and horses. Moreover, we see that if a query region does not fit to what is expected, there is not arbitrary hallucination, but rather the region is down weighted: Whereas the right tire of the first bicycle is completely reconstructed, the middle of the other tire is down weighted. This is due to the large discrepancy to what the model has learned from training data for this region of a bike.