

Weakly Supervised Object Boundaries

Supplementary material

Anna Khoreva¹ Rodrigo Benenson¹ Mohamed Omran¹ Matthias Hein² Bernt Schiele¹

¹Max Planck Institute for Informatics, Saarbrücken, Germany

²Saarland University, Saarbrücken, Germany

1. Content

This supplementary material provides both additional quantitative and qualitative results:

- Section 2 (Figure 5) shows visualization of the different variants of the proposed weakly supervised boundary annotations.
- Section 3 provides additional results for VOC (Figures 1 - 3, Table 1). Qualitative results of boundary detection for VOC can be found in Section 4 (Figure 6).
- Detailed results for COCO are shown in Section 5 (Figure 4). Visualization of boundary detection for COCO can be found in Section 5 (Figure 7).
- Section 7 reports SBD results per class (Figure 8). Boundary detection examples for SBD are shown in Section 8 (Figure 9).

2. Weakly supervised boundary annotations

In this work we propose to train boundary detectors using weakly supervised annotations. We propose and evaluate multiple strategies to generate annotations fusing different sources, such as unsupervised image segmentation [2], object proposal methods [10, 5], and object detectors [3, 6] (trained on bounding boxes). Figure 5 illustrates the examples of the proposed weakly supervised boundary annotations, these extend the example in Figure 4 of the main paper. See Section 5 of the main paper for more details.

3. Detailed results for VOC

We provide here some quantitative results mentioned in Section 5, and in Table 2 of the main paper.

Detection BBs versus GT BBs To generate weakly supervised boundary annotations we explore class-specific object detectors, such as Fast-RCNN [3] with MCG [5] or Selective Search [10] object proposals and Faster-RCNN

[6]. We also experiment using directly the ground truth bounding box annotation variants: bounding boxes tight around segmentation annotations and bounding boxes from detection annotations. This results in a minor drop in the performance. Using object detector allows to filter out hard cases by discarding images which have zero bounding boxes with confidence scores above 0.8. from the training set, thus reducing training data noise and resulting in the performance improvement. The results are shown in Figure 1 and in Table 1.

BBs source	F	AP
Fast-RCNN	42.6	34.3
Faster-RCNN	42.4	34.2
GT segm. masks	42.0	33.6
GT bound. boxes	41.6	32.8

Table 1: VOC results for weakly supervised SE ($MCG \cap BBs$) models with different bounding boxes sources.

GrabCut, DenseCut and CNN+GraphCut For generating weakly supervised annotations in addition to GrabCut[7] we also experimented with DenseCut [1] and CNN+GraphCut [8]. Employing DenseCut or CNN+GraphCut does not bring any gain opposed to GrabCut. The results are presented in Figure 2.

Using VOC₊ Since we generate boundary annotations in a weakly supervised fashion, we are able to generate boundaries over arbitrary image sets. In our experiments we consider VOC (Pascal VOC12 segmentation task) and VOC₊ (VOC plus images from Pascal VOC12 detection task). Figure 3 presents the results using VOC and VOC₊. Methods using VOC₊ are denoted by \cdot_+ (e.g. SE (SeSe₊ \cap BBs)). Using a larger set of images for training allows to further improve the performance of SE trained with the generated boundary annotations.

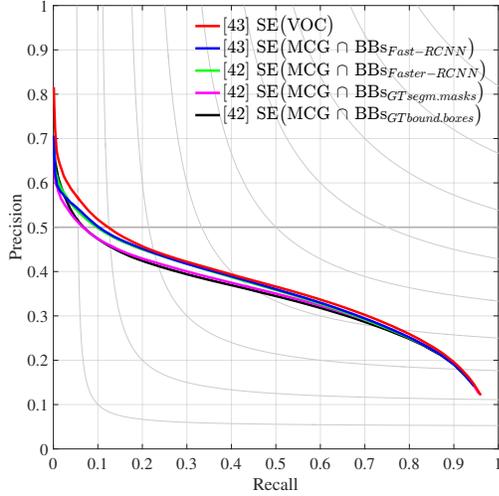


Figure 1: VOC results: Detection BBs versus GT BBs. (·) denotes the data used for training. Legend indicates AP numbers. $BBS_{GT \text{ bound. boxes}}$ denotes GT bounding boxes, $BBS_{GT \text{ segm. masks}}$ boxes obtained from GT segmentations.

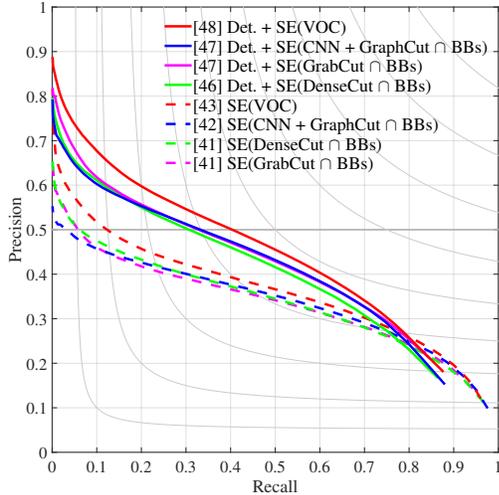


Figure 2: VOC results: GrabCut versus DenseCut and CNN+GraphCut. (·) denotes the data used for training. Continuous/dashed line indicates models using/not using a detector at test time. Legend indicates AP numbers. DenseCut and CNN+GraphCut perform on par with GrabCut.

4. VOC boundary detection examples

Figure 6 presents qualitative results of boundary detection on VOC. The presented boundary estimate examples show that high quality object boundaries can be achieved using only detection bounding box annotations. This figure extends Figure 7 of the main paper.

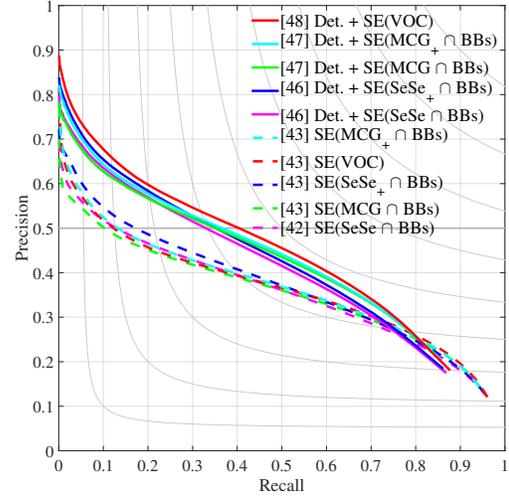


Figure 3: VOC results using additional VOC₊ images. (·) denotes the data used for training. Continuous/dashed line indicates models using/not using a detector at test time. Legend indicates AP numbers.

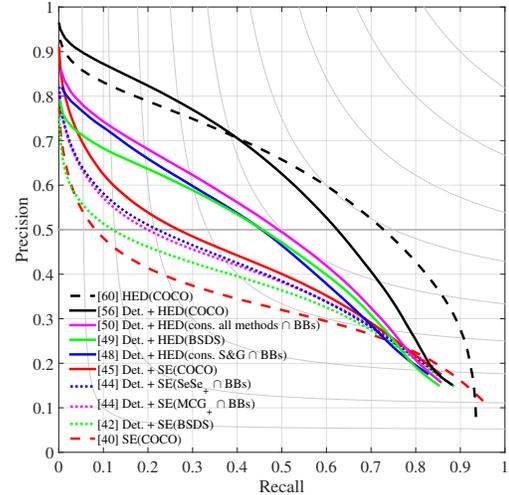


Figure 4: COCO results. (·) denotes the data used for training. Continuous/dashed line indicates models using/not using a detector at test time. Legend indicates AP numbers. For weakly supervised cases the results are shown with the models trained on VOC, without re-training on COCO.

5. Detailed results for COCO

Figure 4 shows the generalization of the proposed weakly supervised variants for object boundary detection on the COCO dataset. For weakly supervised cases the results are shown with the models trained on VOC, without re-training on COCO. These curves complement Table 3 from the main paper.

For both SE and HED the models trained on the proposed

weak annotations perform as well as the fully supervised SE models. Similar to the VOC benchmark the HED model trained on ground truth shows superior performance.

6. COCO boundary detection examples

Figure 7 shows examples of boundary detection on COCO. This figure complements Table 3 from the main paper. Our proposed weak-supervision techniques achieve competitive performance with fully supervised results for object-specific boundaries.

7. Detailed results for SBD

Per-class curves Figure 8 shows the per class performance of the proposed weakly supervised boundary variants trained with SE and HED on the SBD dataset [4] (this figure is a breakdown of Figure 9 in the main paper). In contrast to VOC and COCO we move from object boundaries to class specific object boundaries. We are interested in external boundaries of all annotated objects of the specific semantic class and all internal boundaries are ignored during evaluation following the benchmark [4]. As Figure 8 shows our weakly supervised approach considerably outperforms [9, 4] on all 20 classes.

Compared to VOC, SE and HED results are most similar between each other because the evaluation protocol focuses on the external object boundaries (ignoring internal object boundaries), where both methods equally well. Compared to VOC, we also notice that Det.+HED (cons. S&G \cap BBs) performs better than Det.+HED (SBD), we attribute this to the “consensus” aspect of our generated annotations (see BSDS results, Section 4 of main paper).

8. SBD boundary detection examples

Figure 9 shows examples of boundary detection on the SBD dataset. As the quantitative results indicate, qualitatively our weakly supervised results are on par to the fully supervised ones.

References

- [1] M.M. Cheng, V. Prisacariu, S. Zheng, P. Torr, and C. Rother. Denscut: Densely connected crfs for real-time grabcut. *Computer Graphics Forum*, 2015. 1
- [2] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 2004. 1
- [3] R. Girshick. Fast R-CNN. In *ICCV*, 2015. 1
- [4] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, and J. Malik. Semantic contours from inverse detectors. In *ICCV*, 2011. 3
- [5] J. Pont-Tuset, P. Arbeláez, J. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping for image segmentation and object proposal generation. In *arXiv:1503.00848*, March 2015. 1
- [6] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, 2015. 1
- [7] C. Rother, V. Kolmogorov, and A. Blake. Grabcut - interactive foreground extraction using iterated graph cuts. *SIGGRAPH*, 2004. 1
- [8] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *ICLR Workshop*, 2014. 1
- [9] J.R.R. Uijlings and V. Ferrari. Situational object boundary detection. In *CVPR*, 2015. 3
- [10] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, and A.W.M. Smeulders. Selective search for object recognition. *IJCV*, 2013. 1

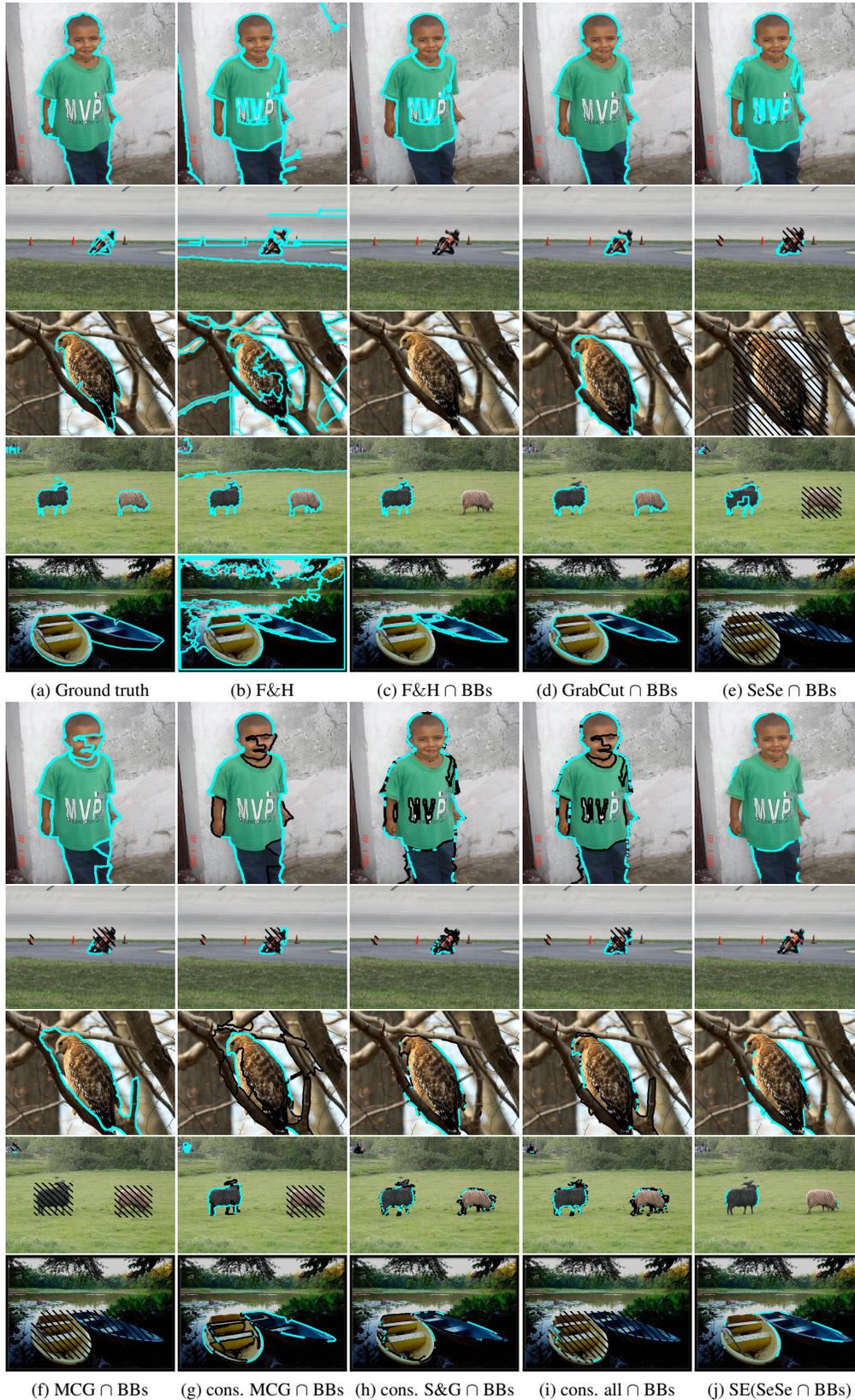


Figure 5: Different generated boundary annotations. Cyan/black indicates positive/ignored boundaries.

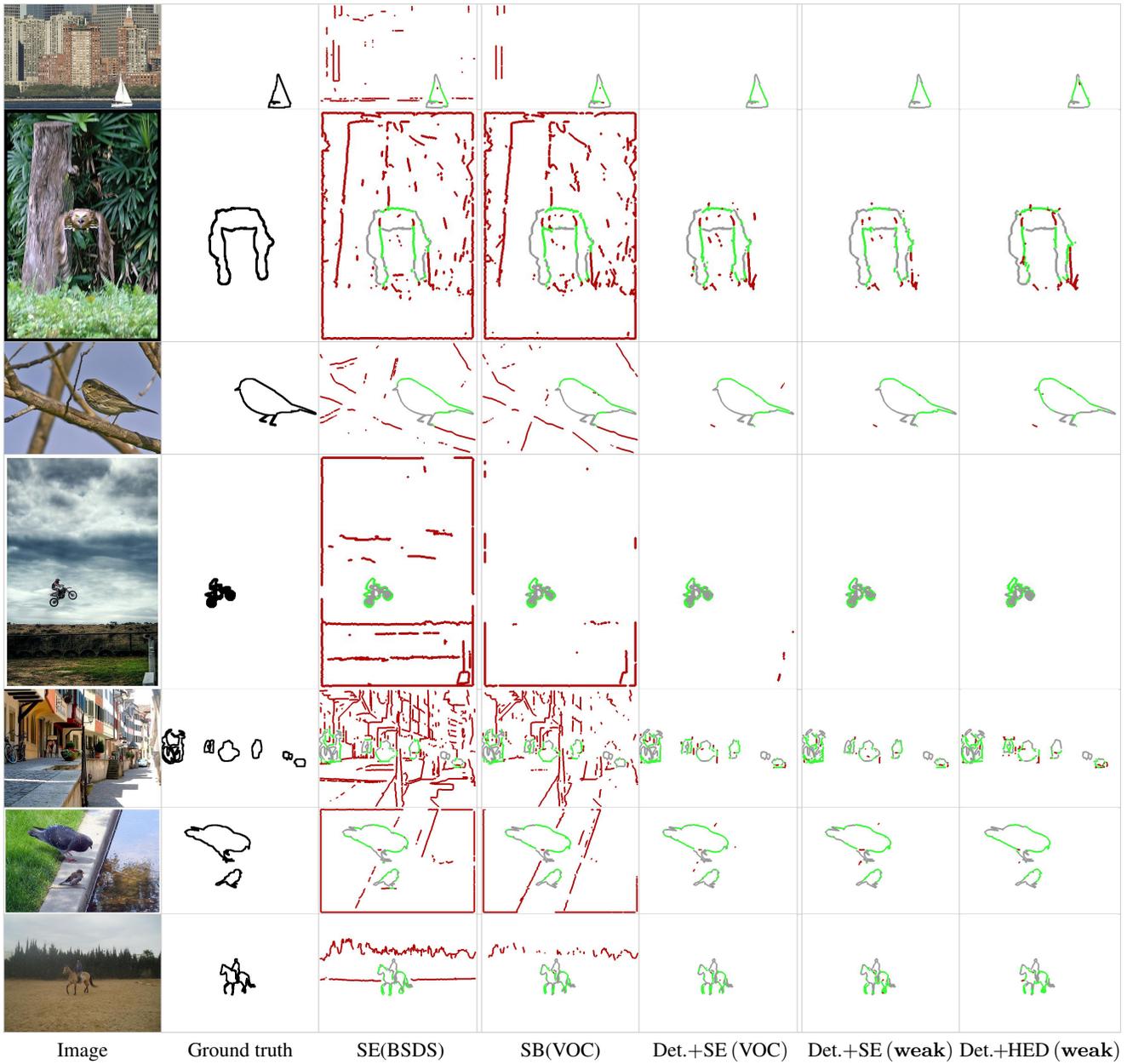


Figure 6: Qualitative results on VOC. (·) denotes the data used for training. Red/green indicate false/true positive pixels, grey is missing recall. All methods are shown at 50% recall. Det.+SE (weak) denotes the model Det.+SE (SeSe₊ ∩ BBs) and Det.+HED (weak) denotes Det.+HED (cons. S&G ∩ BBs). Object-specific boundaries differ from generic boundaries (such as the ones detected by SE(BSDS)). By using an object detector we can suppress non-object boundaries and focus boundary detection on the classes of interest. The proposed weakly supervised techniques allow to achieve high quality boundary estimates that are similar to the ones obtained by fully supervised methods.

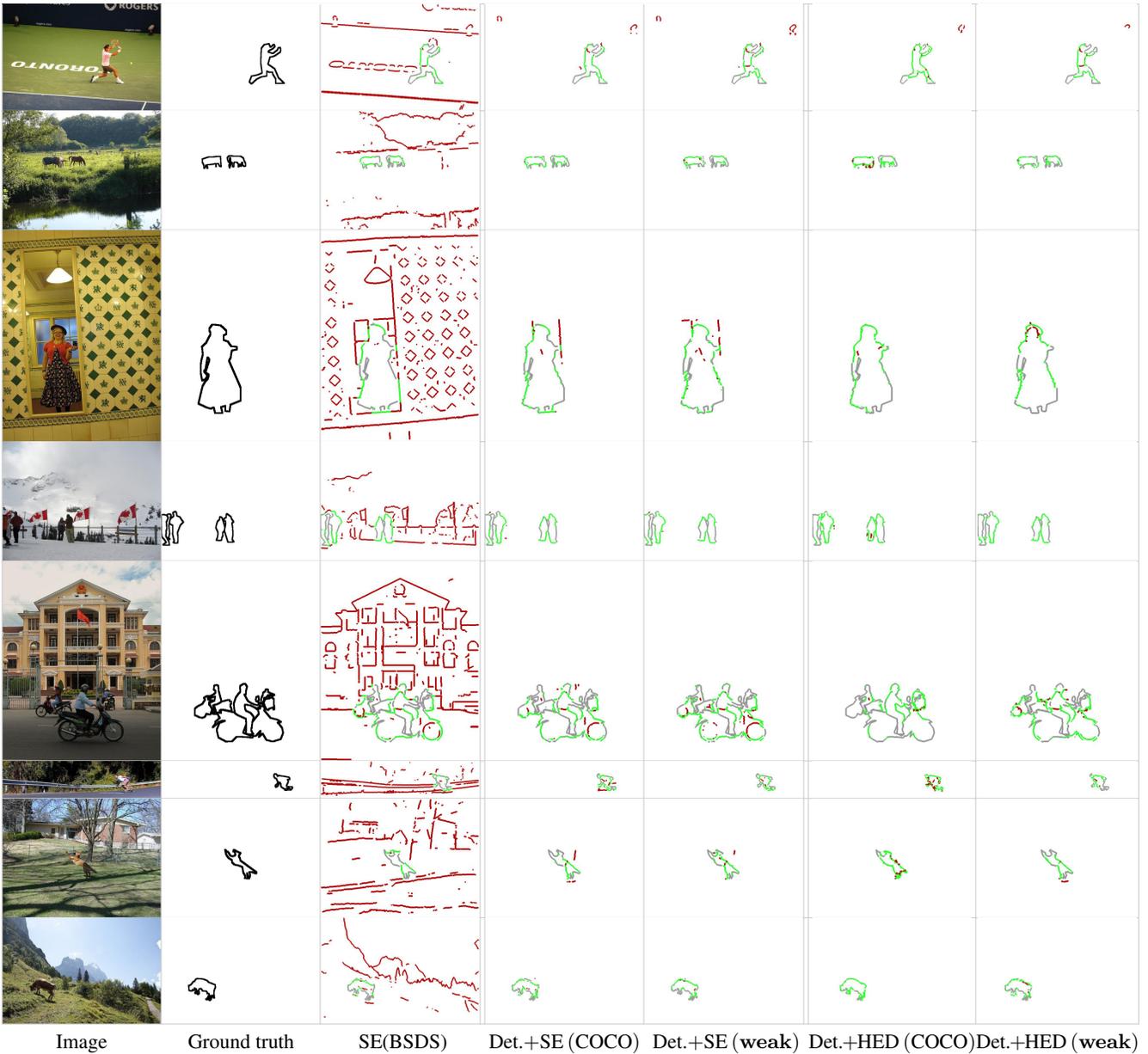


Figure 7: Qualitative results on COCO. (·) denotes the data used for training. Red/green indicate false/true positive pixels, grey is missing recall. All methods are shown at 50% recall. Det.+SE (weak) denotes the model Det.+SE ($SeSe_+ \cap BBs$) and Det.+HED (weak) denotes Det.+HED (cons. S&G \cap BBs). Object-specific boundaries differ from generic boundaries (such as the ones detected by SE(BSDS)). By using an object detector we can suppress non-object boundaries and focus boundary detection on the classes of interest. The proposed weakly supervised techniques allow to achieve high quality boundary estimates that are similar to the ones obtained by fully supervised methods.

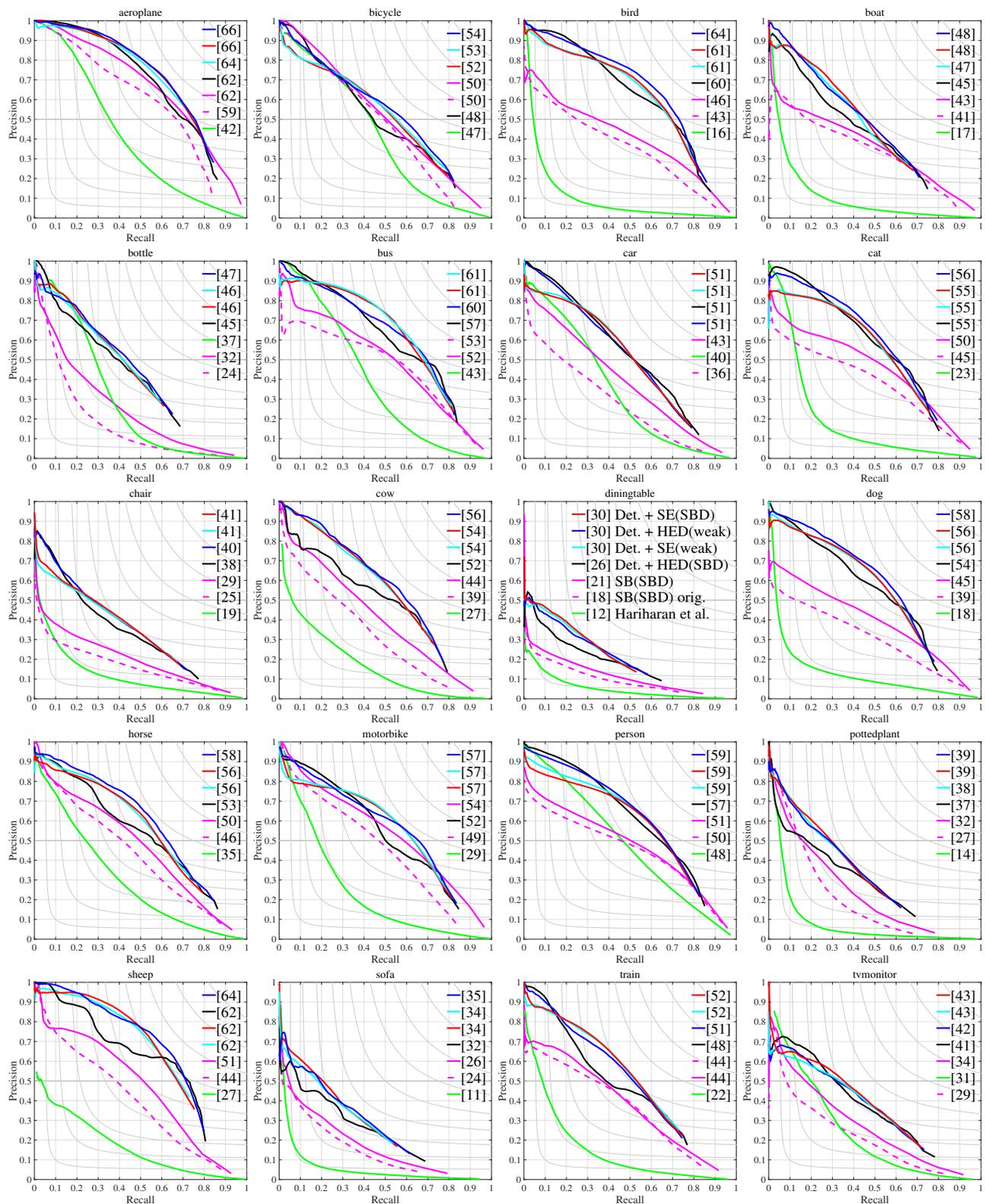


Figure 8: SBD boundary PR curves per class.(·) denotes the data used for training. Legend indicates AP numbers. Det.+SE(weak) denotes the model Det.+SE(MCG₊ ∩ BBs) and Det.+HED(weak) denotes Det.+HED(cons. S&G ∩ BBs). For all classes our weakly supervised results are on par to the fully supervised ones.

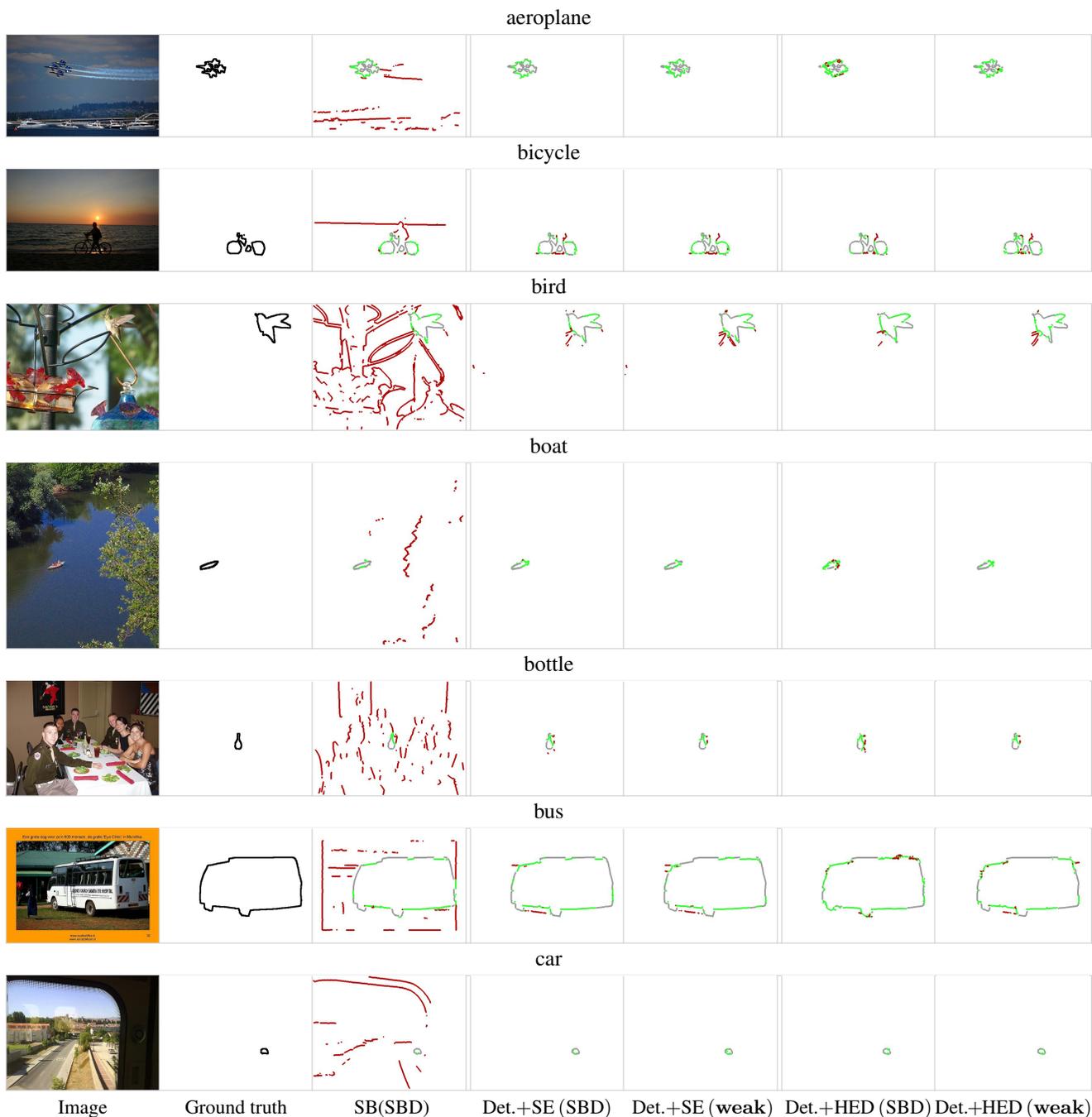


Figure 9: Qualitative results on SBD. Red/green indicates false/true positives pixels, grey is missing recall. All methods shown at 50% recall. Det.+SE (weak) denotes the model Det.+SE ($MCG_+ \cap BBs$) and Det.+HED (weak) denotes Det.+HED (cons. $S \& G \cap BBs$). As the quantitative results indicate, qualitatively our weakly supervised results are on par to the fully supervised ones.

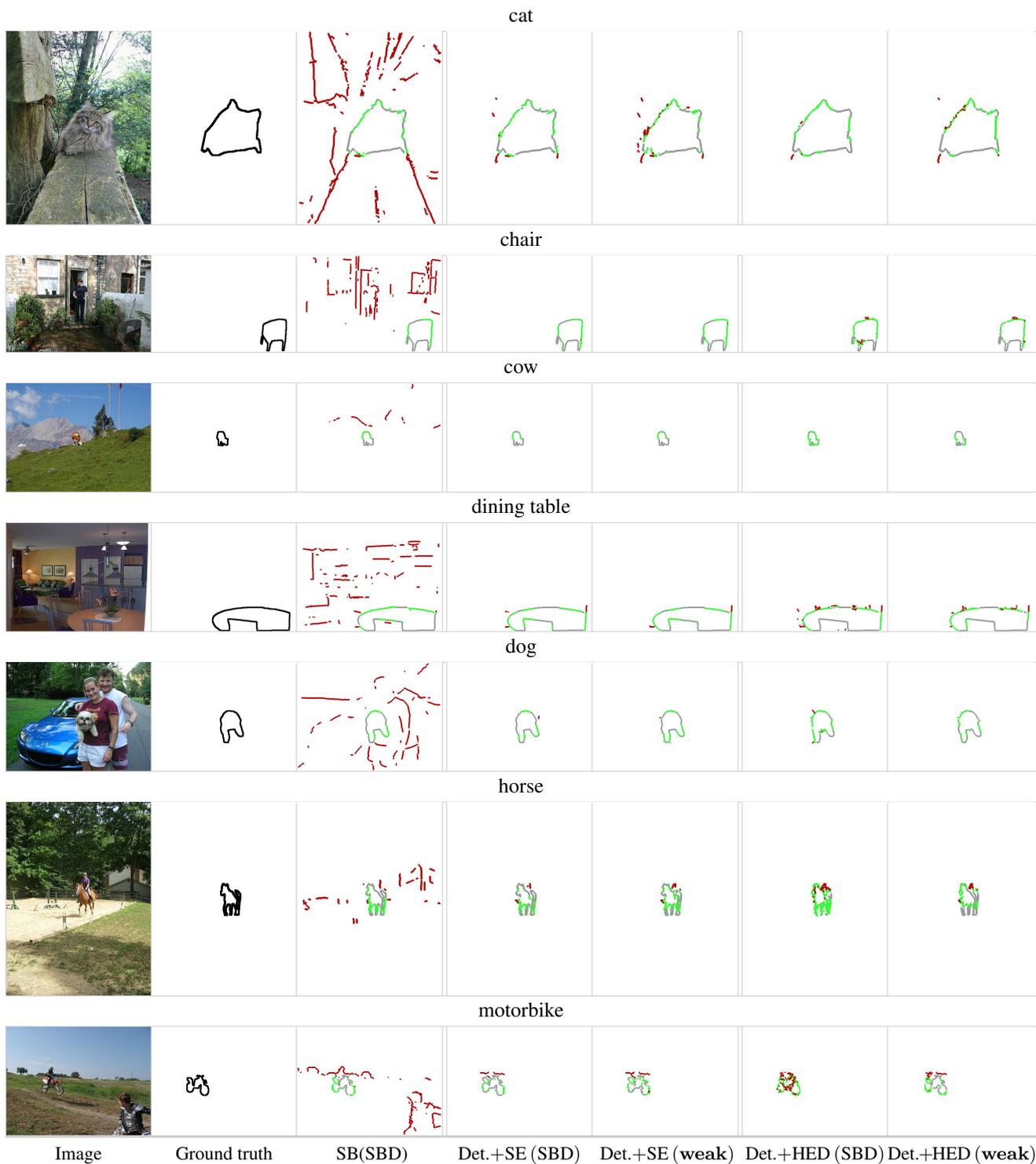


Figure 9: Qualitative results on SBD. Red/green indicates false/true positives pixels, grey is missing recall. All methods shown at 50% recall. Det.+SE (weak) denotes the model Det.+SE ($MCG_+ \cap BBs$) and Det.+HED (weak) denotes Det.+HED (cons. S&G $\cap BBs$). As the quantitative results indicate, qualitatively our weakly supervised results are on par to the fully supervised ones.



Figure 9: Qualitative results on SBD. Red/green indicates false/true positives pixels, grey is missing recall. All methods shown at 50% recall. Det.+SE (weak) denotes the model Det.+SE ($MCG_+ \cap BBs$) and Det.+HED (weak) denotes Det.+HED (cons. $S \& G \cap BBs$). As the quantitative results indicate, qualitatively our weakly supervised results are on par to the fully supervised ones.