Weakly Supervised Object Boundaries

Anna Khoreva[1], Rodrigo Benenson[1], Mohamed Omran[1], Matthias Hein[3], Bernt Schiele[1]
[1]Max Planck Institute for Informatics, Germany. [2]Saarland University, Germany.

State-of-the-art learning based boundary detection methods require extensive training data. Since labelling object boundaries is one of the most expensive types of annotations, there is a need to relax the requirement to carefully annotate images to make both the training more affordable and to extend the amount of training data.

In this paper we focus on learning object boundaries in a weakly supervised fashion and show that high quality object boundary detection can be obtained without using any object-specific boundary annotations. We propose several ways of generating object boundary annotations with different levels of supervision, from just using a bounding box oriented object detector to using the boundary detector trained on generic boundaries. For generating weak object boundary annotations we consider different sources, fusing unsupervised image segmentation [5] and object proposal methods [8, 12] with object detectors [6, 9]. We show that bounding box annotations alone suffice to achieve objects boundary estimates with high quality.

We present results using a decision forest (SE) [3] and a convnet edge detector (HED) [13]. We report top performance on Pascal object boundary detection [4, 7] with our weak-supervision approaches already surpassing previously reported fully supervised results.

Our main contributions are summarized below:

- We introduce the problem of weakly supervised object-specific boundary detection.

- We show that good performance can be obtained on BSDS, Pascal VOC12, and SBD boundary estimation using only weak-supervision (leveraging bounding box detection annotations without the need of instance-wise object boundary annotations).

- We report best known results on PascalVOC12, and SBD datasets. Our weakly supervised results alone improve over the previous fully supervised state-of-the-art.

Further information is available at https://goo.gl/kDVZwS.

Robustness to annotation noise

We start by exploring weakly supervised training for generic boundary detection, as considered in BSDS.

Model based approaches such as Canny [2] and F&H [5] are able to provide low quality boundary detections. We notice that correct boundaries tend to have consistent appearance, while erroneous detections are mostly inconsistent. Robust training methods should be able to pick-up the signal in such noisy detections.

In Figure 2 we report our results when training a structured decision forest (SE) and a convnet (HED) with noisy boundary annotations. When training SE using either Canny ("SE(Canny)") or F&H ("SE(F&H)") we observe a notable jump in boundary detection quality. HED(SE(F&H)) provides better boundaries than SE(F&H) alone, and reaches quality comparable to the classic gPb method [1].

Conclusion SE is surprisingly robust to annotation noise during training. HED is also robust but to a lesser degree. By using noisy boundaries generated from unsupervised methods, we can reach a performance comparable to the bulk of current methods.

Weakly supervised boundary annotations

We propose to train boundary detectors using data generated from weak annotations. Our weakly supervised models are trained in a regular fashion, but use generated (noisy) training data as input instead of human annotations. We consider boundary annotations generated with three different levels of supervision: fully unsupervised, using only detection annotations, and using both detection annotations and BSDS boundary annotations (e.g. using generic boundary annotation, but zero object-specific boundaries). Different variants of weakly supervised boundary annotations are illustrated in Figure 3.

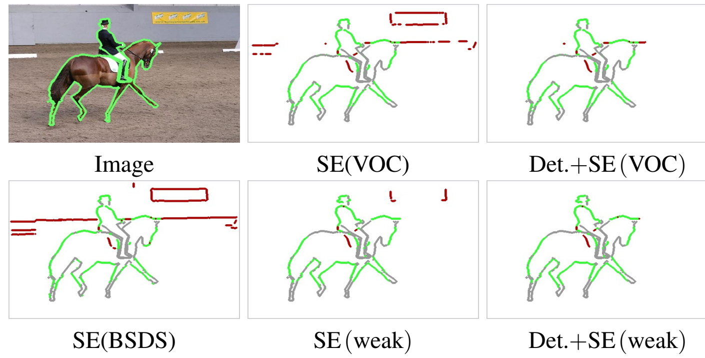

Figure 1: Object boundaries differ from generic boundaries. The proposed weakly supervised approach drives boundary detection towards the objects of interest. Red/green indicate false/true positive pixels, grey is missing recall. All methods are shown at 50% recall.

BBs We use the bounding box annotations to train a class-specific object detector [6, 9]. We then apply this detector over the training set (and possibly a larger set of images), and retain boxes with confidence scores above 0.8.

F&H As a source of unsupervised boundaries we consider the image segmentation technique proposed by [5] (F&H). We intersect these boundaries with detection bounding boxes from [9] (F&H$\cap$BBs). Only the segment boundaries contained inside a bounding box are retained.

GrabCut A way to exclude internal object boundaries, is to extract object contours via figure-ground segmentation of the detection bounding box [9]. We use GrabCut [10] for this purpose. For GrabCut$\cap$BBs a segment is only accepted if a detection has the IoU ≥ 0.7. Otherwise, the whole region is marked as ignore.

Object proposals Another way to bias generation of boundary annotations towards object contours is to consider object proposals: SeSe [12] and MCG [8]. SeSe$\cap$BBs and MCG$\cap$BBs are generated by matching proposals to bounding boxes [6] (if IoU ≥ 0.9).

Consensus boundaries We consider using the consensus between object proposal boundaries. The boundary is considered to be present if the agreement is higher than 70%, otherwise the boundary is ignored. We denote such annotations as "cons.", e.g. cons. MCG$\cap$BBs. Another way to generate sparse (consensus-like) boundaries, is to threshold the boundary probability map out of SE$(\cdot)$ model. SE(SeSe$\cap$BBs) uses the top 15% quantile per image as weakly supervised annotations. We can also do consensus between methods. cons. S&G$\cap$BBs is the intersection between

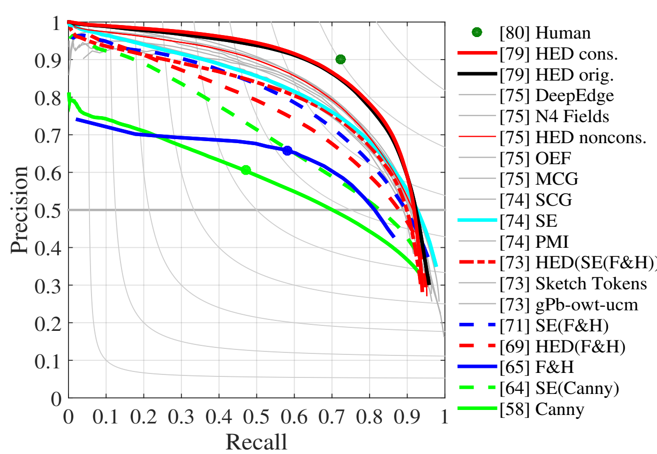

Figure 2: BSDS results. Canny and F&H points indicate the boundaries used as noisy annotations.

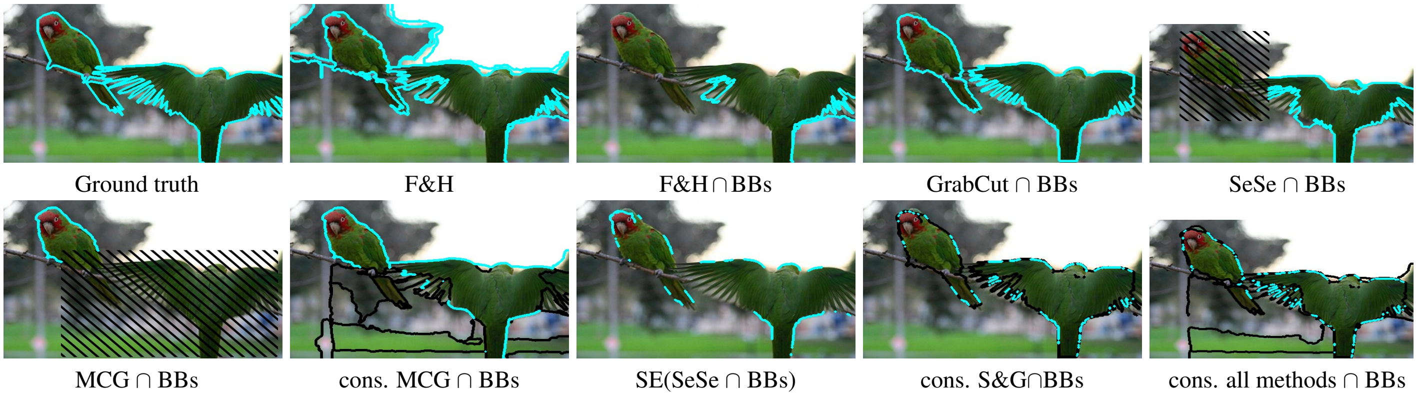

Figure 3: Different generated boundary annotations. Cyan/black indicates positive/ignored boundaries.

	Family	Method	mF	mAP
Other	GT	Hariharan et al. [7]	28	21
SE	GT	SB(SBD) orig. [11]	39	32
		SB(SBD)	43	37
		Det.+SE (SBD)	51	45
	Other GT	Det.+SE (BSDS)	51	44
		Det.+MCG (BSDS)	50	42
	Weakly supervised	SB(SeSe ∩ BBs)	40	34
		SB (MCG ∩ BBs)	42	35
		Det.+SE (SeSe ∩ BBs)	48	42
		Det.+SE (MCG ∩ BBs)	**51**	**45**
HED	GT	HED (SBD)	44	41
		Det.+HED (SBD)	49	45
	Other GT	HED(BSDS)	38	32
		Det.+HED (BSDS)	49	44
	Weakly supervised	HED(cons. MCG ∩ BBs)	41	37
		HED (cons. S&G ∩ BBs)	44	39
		Det.+HED (cons. MCG ∩ BBs)	48	44
		Det.+HED (cons. S&G ∩ BBs)	**52**	**47**

Table 1: SBD results. Results are mean F(ODS)/AP across all 20 categories. $(\cdot)$ denotes the data used for training. See also Figure 4. Bold indicates our best weakly supervised results.

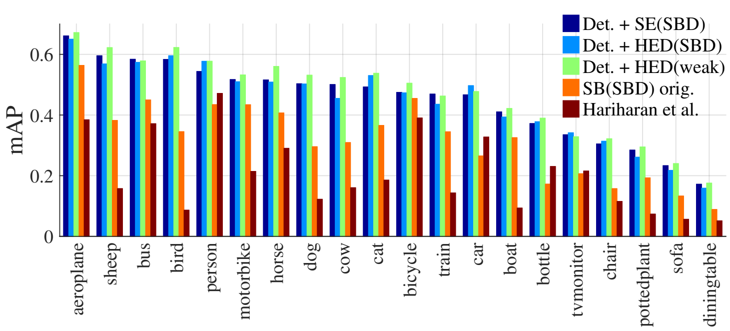

Figure 4: SBD results per class. $(\cdot)$ denotes the data used for training. Det.+ HED (weak) refers to the model Det.+HED (cons. S&G ∩ BBs).

SE (SeSe ∩ BBs), SeSe and GrabCut boundaries (fully unsupervised); while cons. all methods ∩ BBs is the intersection between MCG, SeSe and GrabCut (uses BSDS data).

SBD boundary detection results

In this section we analyse the performance of our weakly supervised variants trained with SE and HED on SBD [7]. We are interested in external object boundaries of the specific semantic class. Internal boundaries are ignored during evaluation [7]. The results are presented in Figure 4 and in Table 1.

Fully supervised Rather than training/testing with 20 SE models plus an image classifier [11], we propose to use a single SE model with a detector [6]. By computing a per-pixel maximum among all detection boxes and their score, we construct an "objectness map" that we multiply with the boundary probability map. False positive boundaries are thus down-scored, and boundaries in high confidence regions get boosted.

Applying SE model plus object detection at test time outperforms the situational boundary detector [11] as well as the Inverse Detectors [7]. The model trained with SE on ground truth performs as well as the HED detector. Both of the models are good at detecting external object boundaries; however SE, being a more local, triggers more on internal boundaries than HED. Even so, in the SBD evaluation these are ignored. This explains the small gap in the performance between SE and HED on this benchmark.

Weakly supervised The models trained with the proposed weakly supervised boundary variants perform on par with the fully supervised detectors, while only using bounding boxes or generic boundary annotations. We show in Table 1 the top result with the Det. + HED(cons. S&G∩BBs) model, achieving the state-of-the-art performance on the SBD benchmark. Figure 4 shows our weakly supervised approach considerably outperforms [7, 11] on all 20 classes.

Conclusion

The presented experiments show that when using the bounding box annotations for training an object detector, one can also train a high quality object boundary detector without additional annotation effort. Using boxes alone, our proposed weak-supervision techniques improve over previously reported fully supervised results for object-specific boundaries. When using generic boundary or ground truth annotations, we also achieve the top performance on the object boundary detection task, outperforming previously reported results by a large margin.

[1] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *PAMI*, 2011.

[2] J. Canny. A computational approach to edge detection. *PAMI*, 1986.

[3] P. Dollár and C. L. Zitnick. Fast edge detection using structured forests. *PAMI*, 2015.

[4] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *IJCV*, 2015.

[5] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 2004.

[6] R. Girshick. Fast R-CNN. In *ICCV*, 2015.

[7] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, and J. Malik. Semantic contours from inverse detectors. In *ICCV*, 2011.

[8] J. Pont-Tuset, P. Arbeláez, J. Barron, F. Marques, and J. Malik. Multi-scale combinatorial grouping for image segmentation and object proposal generation. In *arXiv:1503.00848*, March 2015.

[9] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, 2015.

[10] C. Rother, V. Kolmogorov, and A. Blake. Grabcut -interactive foreground extraction using iterated graph cuts. *SIGGRAPH*, 2004.

[11] J.R.R. Uijlings and V. Ferrari. Situational object boundary detection. In *CVPR*, 2015. URL `http://arxiv.org/abs/1504.06434`.

[12] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, and A.W.M. Smeulders. Selective search for object recognition. *IJCV*, 2013.

[13] S. Xie and Z. Tu. Holistically-nested edge detection. In *ICCV*, 2015.