SparseDet: Improving Sparsely Annotated Object Detection with Pseudo-positive Mining (Supplemental Material)

Contents

1. Recall analysis

To study the effect of missing annotations on the RPN, we compute the recall of the RPN trained using a sparsely annotated dataset. For a model trained using all the annotations, the recall of the RPN is 0.83. As we drop annotations progressively from 30% to 70% the recall drops to 0.79. This shows that there is no significant drop in recall due to missing annotations.

2. Additional Experiments

2.1. Experiments on publicly released splits

Authors of Co-mining [2] create a split similar to Split-4 but using COCO 2017 trainval set to perform ablation experiments. To the best of our knowledge, this is the only split released publicly for SAOD. We compare the performance of our approach with Co-Mining, which uses a RetineNet architecture, on this split and report results in Table 1. We observe a similar trend as reported in the main paper and outperform Co-mining by 1.38, 2.22 and 0.35 percentage points on the *Easy*, *Hard* and *Extreme* subsets respectively.

2.2. Effect of τ_m

We ablate over different values of τ_m and report the results in Table 5 . With a low threshold, the ground truth will be contaminated with noisy predictions and a high threshold will leave out a few positive missing annotations. We observed that a value of 0.9 is a good tradeoff between quality and recall. We use a threshold of 0.9 for all the datasets unless otherwise stated.

2.3. Warm-up

As we rely on the RPN objectness score, we employ a warmup strategy which switches on the pseudo-positive mining (PPM) during training. In Table 4 we start the PPM at different iterations and report results. This experiment is performed on the COCO-mini ablation dataset. Our experiments show that for models which are trained for fewer iterations, starting at 9000 iterations worked the best. For longer experiments we allow a warmup of 30000 iterations.

2.4. Additional ablations on C-RPN

We perform some additional analysis regarding C-RPN by using two sets of proposals, one from original and the other from augmented version of the image and combining them. This approach consists of $2\times$ more proposals than our approach and achieves an mAP of 45.25 (vs. **46.00** using ours). To make the comparison fairer, we generate half the number of proposals from each image and combine them (resulting in the same number of proposals as our approach) and this model achieves 45.46 (vs. **46.00** using ours) demonstrating the effectiveness of C-RPN.

2.5. Augmentation

The proposed approach process an input image and its augmented counterpart. For augmentation, a cascade of random contrast, brightness, saturation, lighting and bounding box erase are used. In Table 3, we analyze the effect of various augmentations on the performance of the model. This experiment was conducted on the same ablation set as the main paper. We observe that applying random saturation or random lighting alone are not as effective compared to other augmentations. Applying bounding box erase alone provides the most improvement. Finally, we achieve the best performance when we use all the augmentations. We randomly sample contrast, brightness, and saturation from

Table 1: Results on the splits released by authors of [2]. All methods use Res-50 FPN architecture.

Method	**Performance**		
	Easy	*Hard*	*Extreme*
Co-mining[2]	35.40	31.80	23.00
Ours	**36.78**	**34.02**	**23.35**

Table 2: Comparison with [1] on Split-2

Method	**Performance**		
	30%	*50%*	*70%*
Pseudo Label [1]	35.00	32.79	**29.03**
Soft Sampling [3]	33.98	31.39	27.30
Ours	**35.94**	**33.13**	28.88

Table 3: Analysis of different augmentations used in this work.

Contrast	**Brightness**	**Saturation**	**Lighting**	**Bbox-Erase**	**Performance**
✗	✗	✗	✗	✗	42.76
✓	✗	✗	✗	✗	43.99
✗	✓	✗	✗	✗	43.81
✗	✗	✓	✗	✗	42.70
✗	✗	✗	✓	✗	42.95
✗	✗	✗	✗	✓	45.11
✓	✓	✗	✗	✗	44.03
✓	✓	✓	✗	✗	43.90
✓	✓	✓	✓	✗	43.91
✓	✓	✓	✓	✓	46.00

Table 4: Analysis on warm-up iteration for performing PPM

Iteration	0	4000	9000	12000
AP	45.51	45.71	**46.00**	45.88

Table 5: Ablation for τ_m

Threshold	0.8	0.9	0.95
AP	45.04	**46.00**	45.38

$[0.5, 1.5]$. For lighting, a random scale of 1.2 was used. For bounding box erase, we randomly erase an area of of $[0.4, 0.7]$ at an aspect ratio of $[0.3, 3.3]$.

3. Additional details of splits and experiments

In Table 2, we compare our approach with Niitani *et al*. [1] and Wu *et al*. [3] using the same backbone for fair comparison. We outperform the best model [1] by 0.94, 0.34 points on the 30% and 50% splits respectively.

4. Results

In this section, we show some more results of the regions identified by the PPM in Figure 1 and a few failure cases in Figure 3. In Figures 1, 3 the red boxes correspond to the available annotations and the regions in white are the ones identified by PPM. Finally, we show detection results and failures of our approach in Figures 2 and 4 respectively.

References

[1] Yusuke Niitani, Takuya Akiba, Tommi Kerola, Toru Ogawa, Shotaro Sano, and Shuji Suzuki. Sampling techniques for large-scale object detection from sparsely annotated objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6510–6518, 2019.

[2] Tiancai Wang, Tong Yang, Jiale Cao, and X. Zhang. Co-mining: Self-supervised learning for sparsely annotated object detection. In *AAAI*, 2021.

[3] Zhe Wu, Navaneeth Bodla, Bharat Singh, Mahyar Najibi, Rama Chellappa, and Larry S. Davis. Soft sampling for robust object detection. In *BMVC*, 2019.

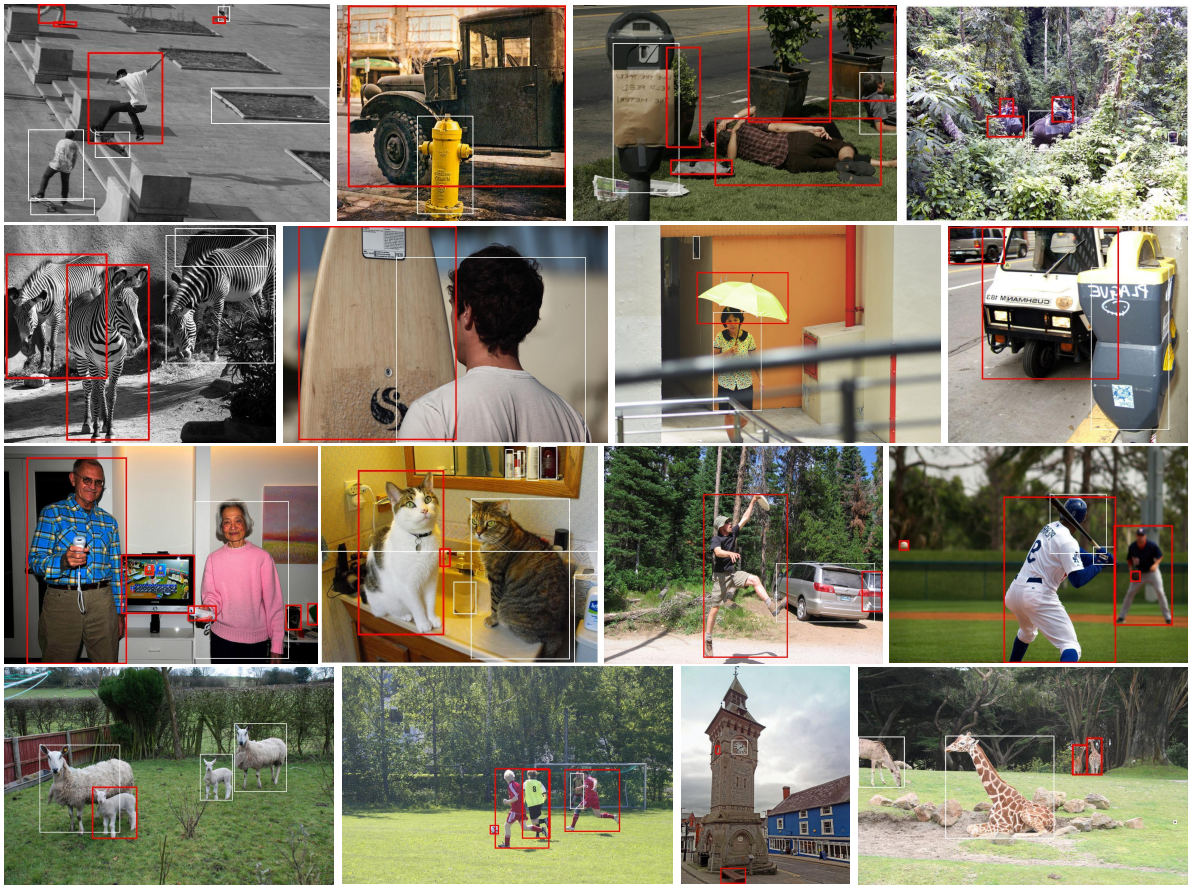

Figure 1: Figure showing regions (white) identified by the PPM step and available ground truth (red) during training.

Figure 2: Qualitative results comparing the output of a model trained using available ground truths (top) to a model trained using our approach (bottom). Predictions with a class confidence score greater than 0.9 are shown.

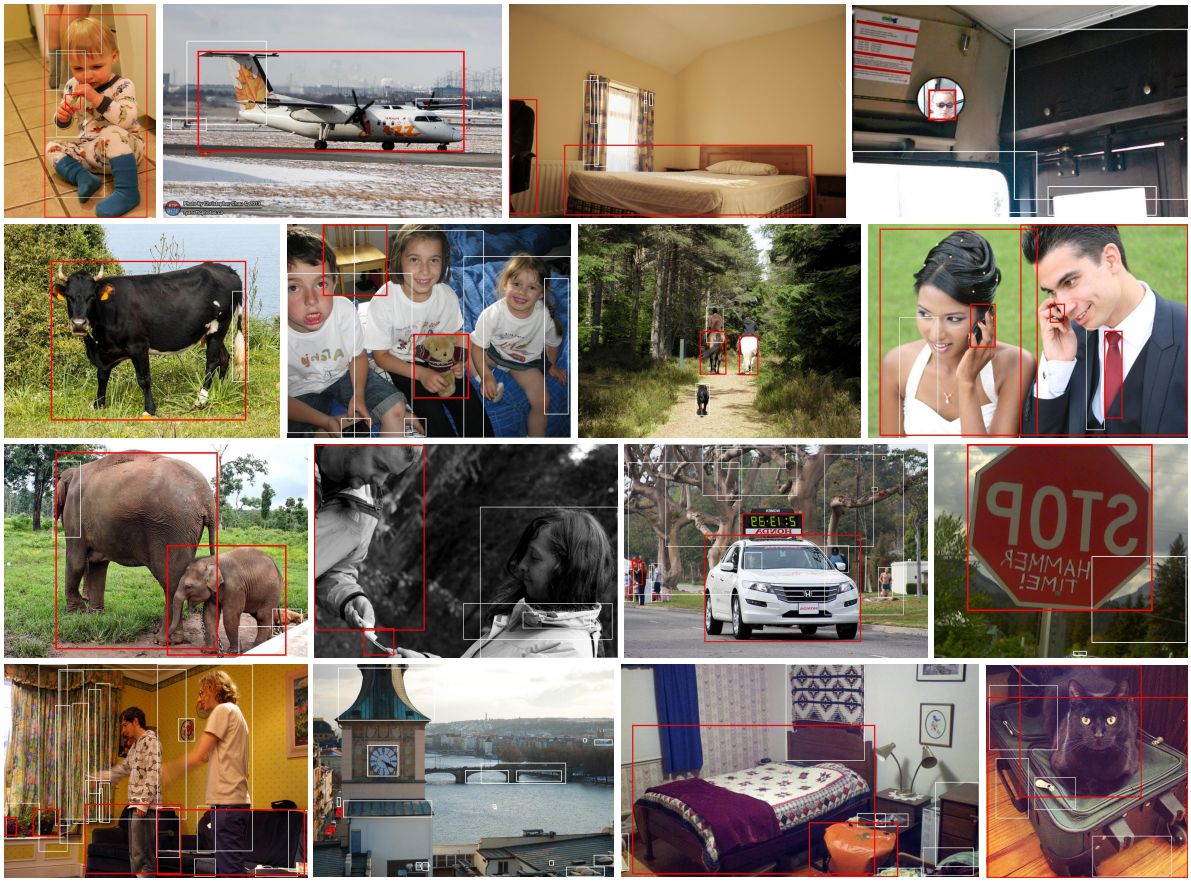

Figure 3: Failure cases of PPM.

Figure 4: Failure cases comparing the output of a model trained using available ground truths (top) to a model trained using our approach (bottom). Predictions with a class confidence score greater than 0.9 are shown.