Supplementary material

A. Evaluation of single object discovery methods on the multiple objects discovery task

In our study, we have compared our approach mainly to the multi-object discovery (MOD) methods. The comparison with single object discovery (SOD) approaches has only been performed after training an object detector with pseudo labels from each method. Here we raise the question of what happens if SOD models are applied to MOD tasks. To answer this question in a quantitative aspect, we provide in table 6 a comparison of our method with LOST[24] and the more recent baseline TokenCut[31]. Since these methods output a single object, odAP is no longer adapted and we use the AP50 metric for comparison. Note that AP50 is more relevant when a fixed number of objects are returned. Considering these results (table 6) and those of table 1, we conclude that our approach provides a better precision-recall trade-off than existing object discovery methods.

Method	VOC07	VOC12	COCO20k
LOST [24]	15.7	17.7	4.1
TokenCut [31]	18.6	22.6	5.8
Ours	**22.5**	**23.1**	**6.3**

Table 6. Multiple objects discovery results in AP@50%

In the following, we provide additional qualitative results on VOC12 and COCOC20k datasets, for the tasks addressed in this work. In particular, in figures 4, 5, we compare our results with LOST[24], a single object discovery method that achieves the second best mAP after training the class-agnostic object detector (see table 2). We can see from the visual results that LOST usually groups several instances, even of different semantics, in complex scenes, with spatially close objects. This explains the superiority of our approach in the initialization of the class-agnostic object detector. We also show in figures 6, 7 that after training an object detector with our pseudo-labels, we improve the separation of instances, even of the same semantic category.

B. More qualitative results for multiple objects discovery on VOC12 dataset

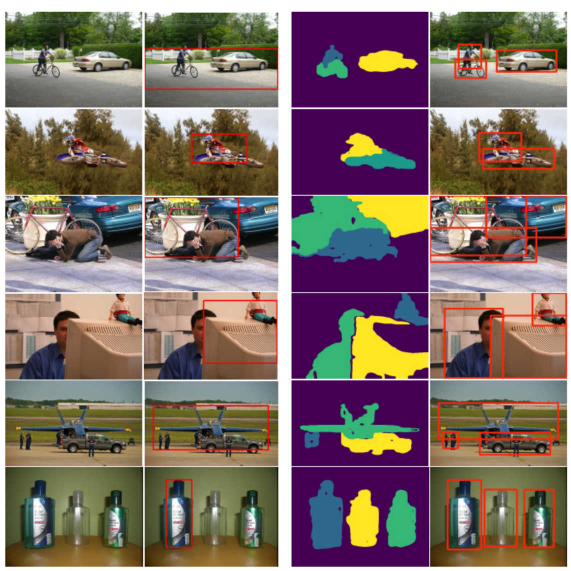

Figure 4. **Qualitative results on VOC12 of the multiple objects discovery.** By column: original image, the predicted bounding-box from LOST[24], our segmentation result, our pseudo-boxes.

C. More qualitative results for multiple objects discovery on COCO20k dataset

Figure 5. **Qualitative results on COCO20K of the multiple objects discovery.** By column: original image, the predicted bounding-box from LOST[24], our segmentation result, our pseudo-boxes.

D. Qualitative results for class-agnostic object detection on VOC12 dataset

Figure 6. Qualitative results for class-agnostic object detection on VOC12 dataset.

E. Qualitative results for class-agnostic object detection on COCO20k dataset

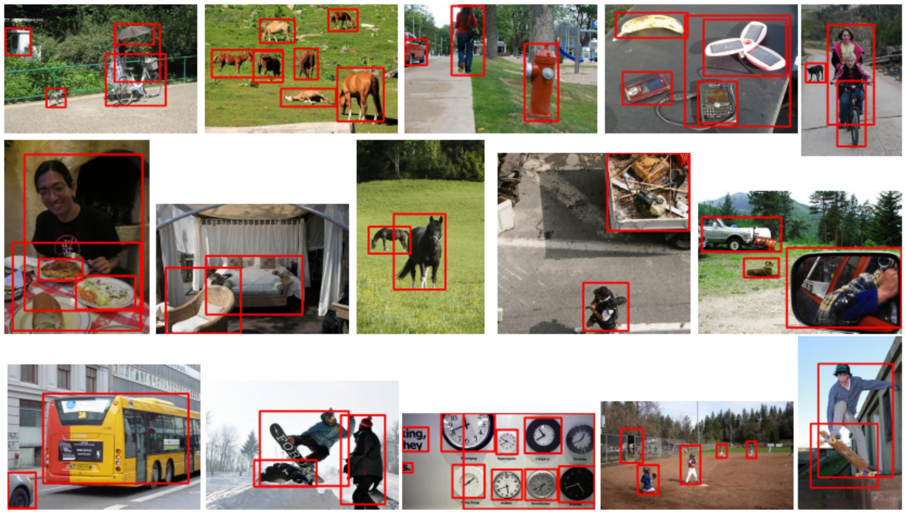

Figure 7. Qualitative results for class-agnostic object detection on COCO20k dataset

F. Failure cases

Our approach, although effective, shows some limitations that can be investigated in a future work. In particular, in the merging process, since nearby segments belonging to the same class are merged, this leads to the merging of nearby instances of the same category, see figure 8. This problem occurs in all existing object discovery methods [29, 24, 31], and comes from the semantic information contained in both supervised and self-supervised features. Solving this problem requires, for example, the learning of an instance-variant representation, which is a challenging task. We hope that this work will stimulate interest in this research direction.

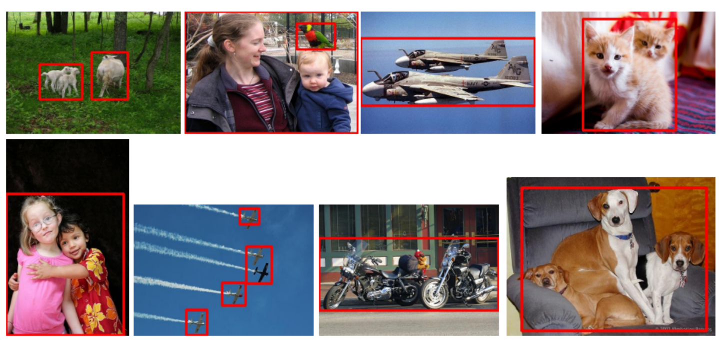

Figure 8. Examples from VOC07 dataset where our approch fails to discover objects, by merging nearby instances of the same category.