

Beyond mAP: Towards better evaluation of instance segmentation

Supplementary Material

A. Behavior of PR curves in SOTA

We make an unusual observation that state-of-the-art instance segmentation frameworks improve mAP, but also worsen the amount of *hedged predictions*. Fig.7 shows a side-by-side comparison between SOLOv2 with Matrix NMS and Mask NMS on individual images from the COCO validation dataset. We observe that at an individual image level, AP is unable to detect or penalize hedging by construction. This is detected using DC, NE, and LRP, and can be mitigated by our proposed Semantic Sorting and Semantic NMS modules.

B. More Qualitative results in COCO validation dataset

In this section, we compare segmentation quality in SOLOv2 and our method where we augment SOLOv2 with a shallow semantic segmentation module, which allows us to perform Semantic Sorting and Semantic NMS. Results are shown in Fig.8, 9, 10. We prevent both spatial and category hedging on all images, leading to better qualitative and interpretable segmentation outputs. In particular, we observe that our method outputs accurate counts of objects in the scene, which may be useful for applications like tracking and crowd counting.

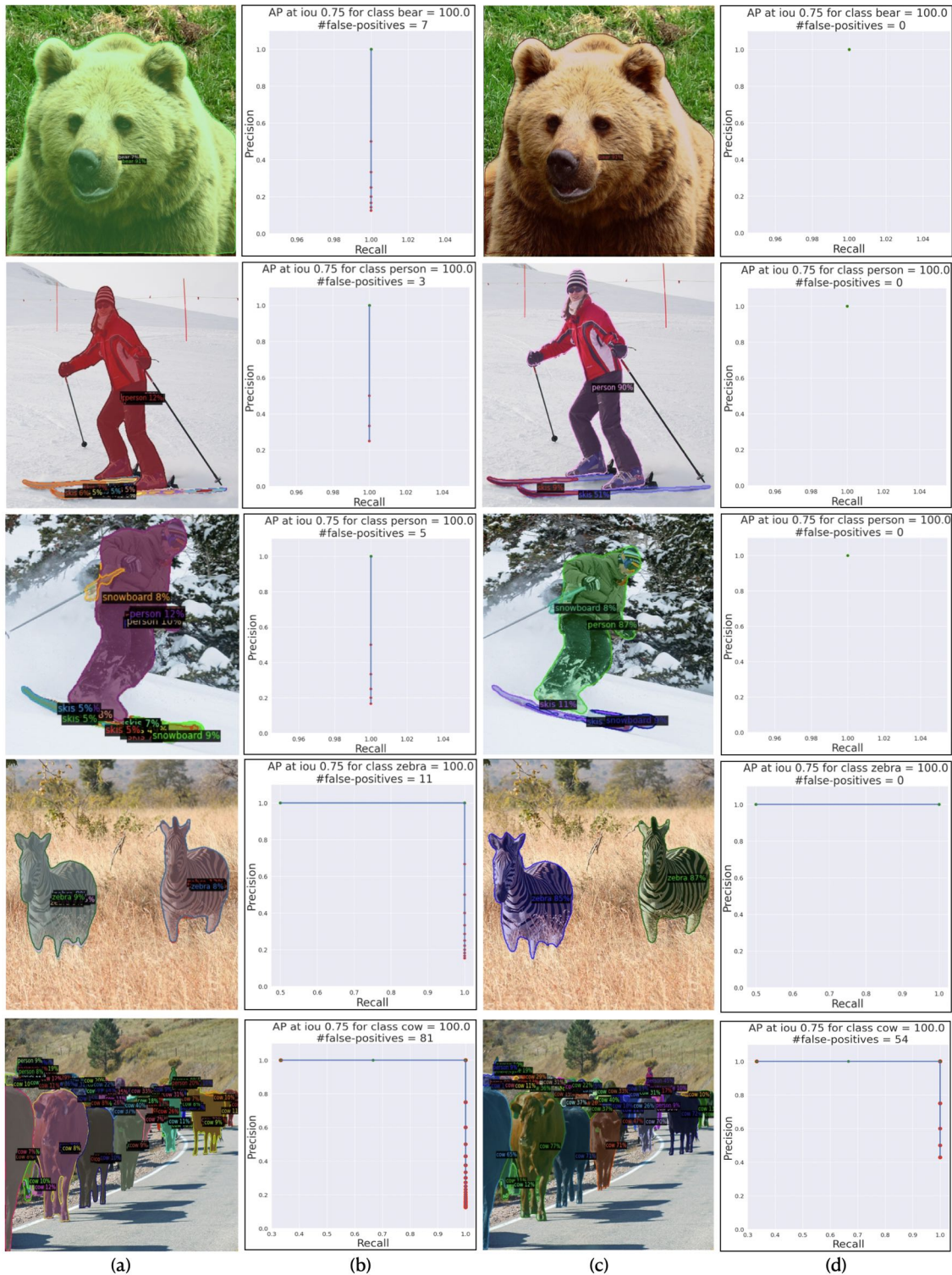


Figure 7. **AP and hedging:** (a) shows the prediction of SOLOv2 model with Matrix NMS, (b) shows the corresponding P/R curve. (c) shows the prediction with the same network but with Mask NMS, (d) shows the corresponding P/R curve. Note that despite having hedging (overcounting) in the first case, the AP scores are the same for both cases. However, they exhibit drastically different qualitative behavior, showing that AP is not an adequate metric for evaluating the hedging problem.



Figure 8. **Qualitative comparison on COCO-val-2017 dataset:** Images on left are predictions made by SOLOv2, images on right are predictions by our model with Semantic Sorting and Semantic NMS.



Figure 9. **Qualitative comparison on COCO-val-2017 dataset:** Images on left are predictions made by SOLOv2, images on right are predictions by our model with Semantic Sorting and Semantic NMS.

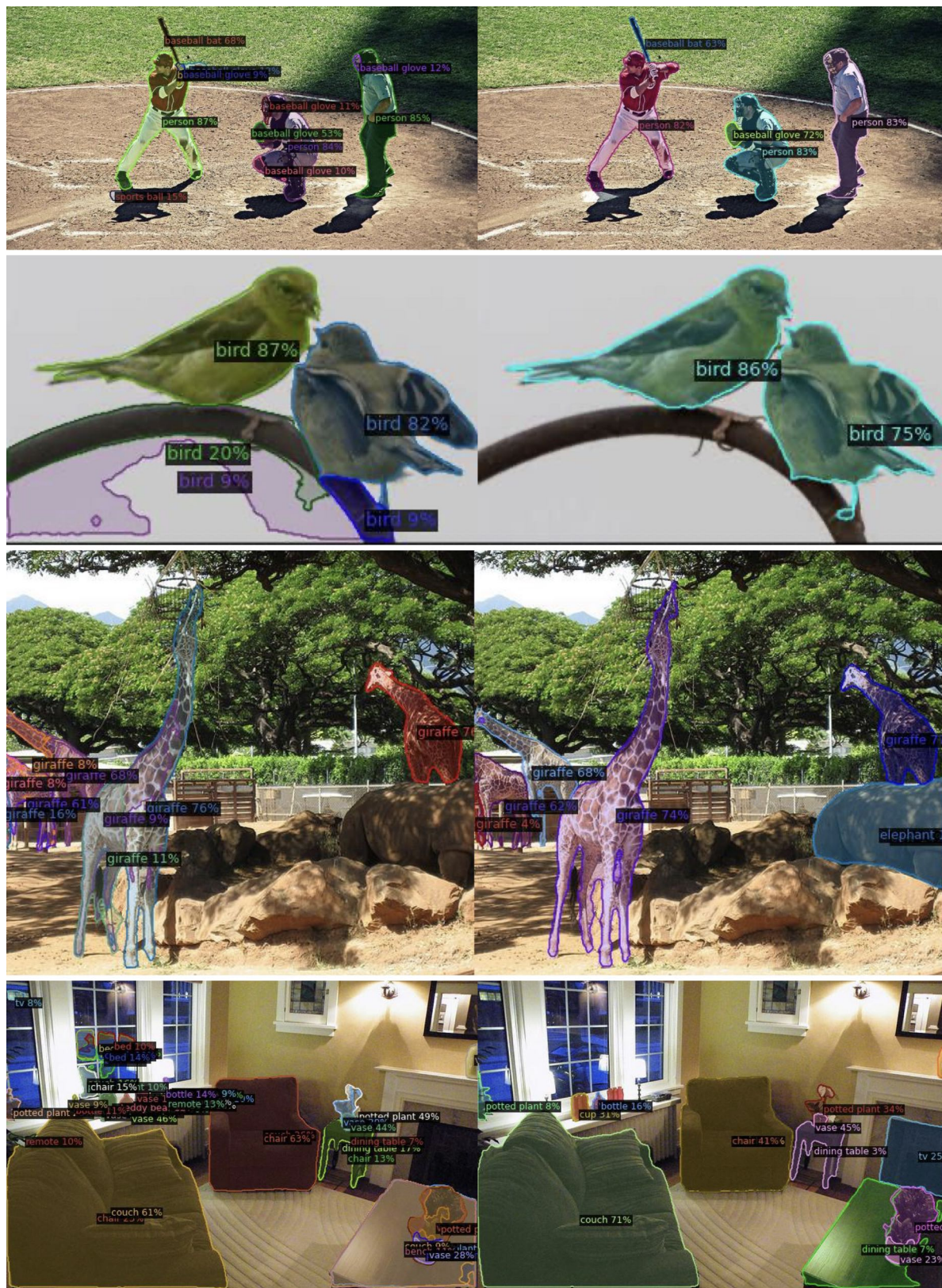


Figure 10. **Qualitative comparison on COCO-val-2017 dataset:** Images on left are predictions made by SOLOv2, images on right are predictions by our model with Semantic Sorting and Semantic NMS.