Supplementary Material: Annotating Object Instances with a Polygon-RNN

Anonymous CVPR submission

Paper ID 2187

1. Additional examples of full image annotation

We show additional qualitative results from our approach in Fig. 1, 2. In the first column, we show the GT provided by the Cityscapes dataset [1], and in the second column, we show results from our approach, obtained without any corrections (i.e., 0 clicks).

2. Additional crop examples

In Fig. 3, 4, we show visualizations of the instances inside the crop of the GT boxes. In the first column, we show the GT annotation, while in the second column, we show the output from SharpMask [2]. In the third column we report the our predictions without any human intervention. Finally, in the fourth column, we show a corrected prediction. Since the output of SharpMask is a dense pixel labeling, we draw the boundary based on the connectivity in the 8-neighborhood.

3. Performance analysis

GT | Ours (0 clicks)

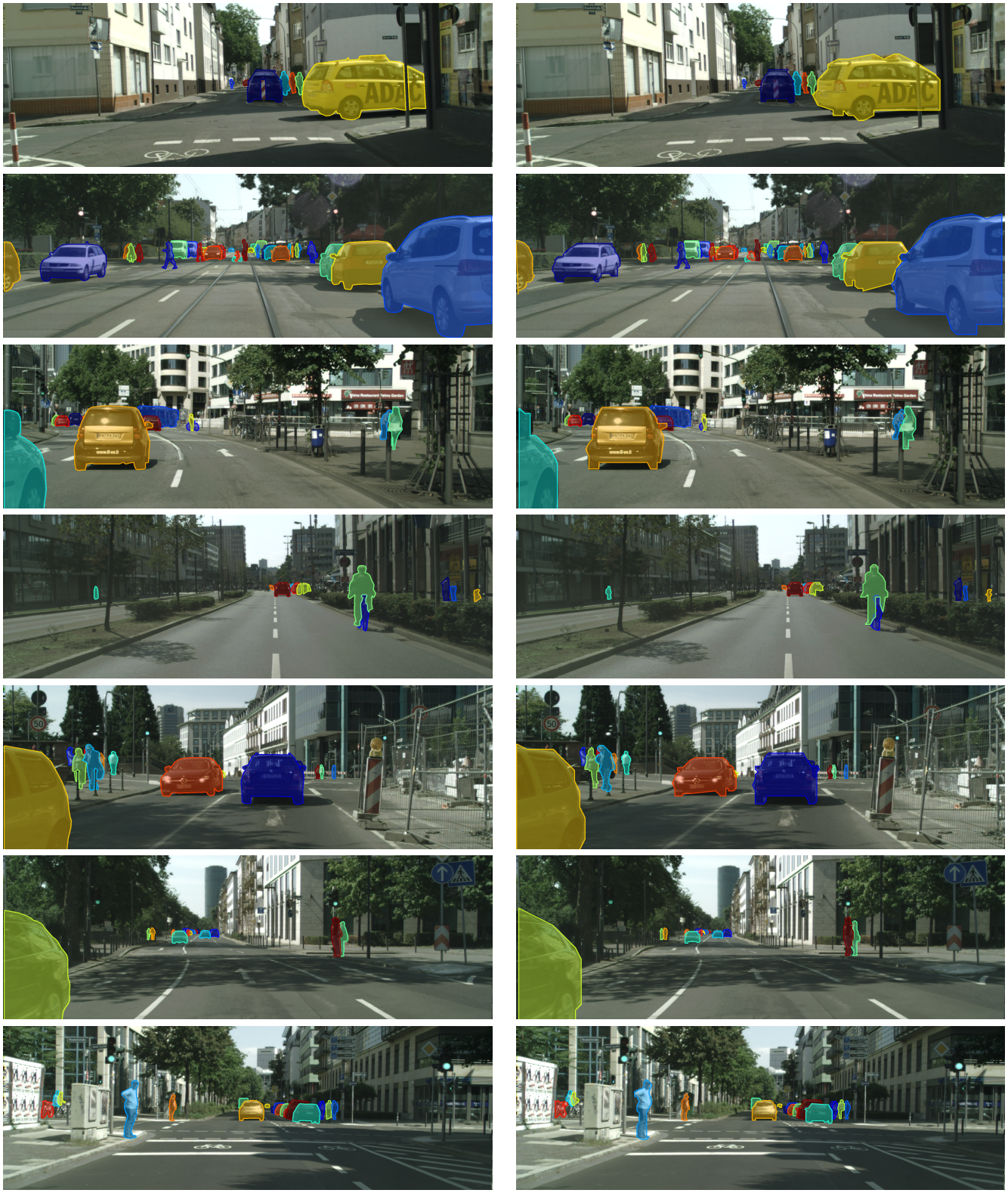

Figure 1. Here we show all our polygons (for all classes) in the original image. We remind the reader that that our approach exploited (ground-truth) boxes to be provided as input. On the **left** we show the ground-truth labeling of the image, while on the **right** we show our polygons in the 0-click regime (running in automatic prediction mode). The GT images contain 38, 12 and 28 instances and required 985, 308 and 580 clicks respectively from their Cityscapes annotators.

GT

Ours (0 clicks)

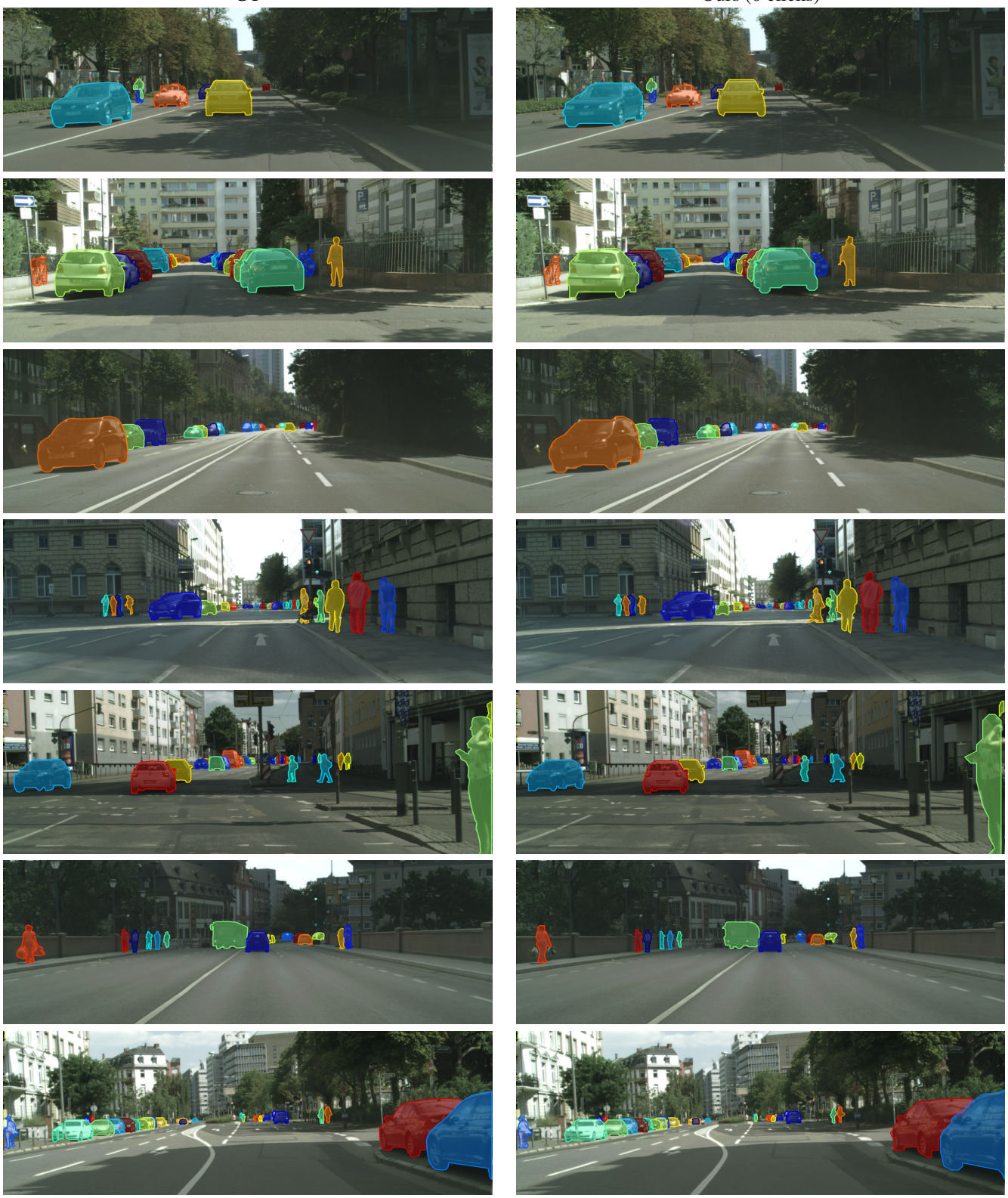

Figure 2. Here we show all our polygons (for all classes) in the original image. We remind the reader that that our approach exploited (ground-truth) boxes to be provided as input. On the **left** we show the ground-truth labeling of the image, while on the **right** we show our polygons in the 0-click regime (running in automatic prediction mode). The GT images contain 38, 12 and 28 instances and required 985, 308 and 580 clicks respectively from their Cityscapes annotators.

Figure 3. Here we look at a few instances in more detail. On the **first column** we show the GT annotation, while on the **second column**, we show the output from SharpMask. On the **third column** we report the PolygonRNN prediction without human intervention. Finally, on the **fourth column** we show a corrected prediction, in which we can observe how the segmentation is refined to better surround the car mirrors or their wheels.

Figure 4. Here we look at a few instances in more detail. On the **first column** we show the GT annotation, while on the **second column**, we show the output from SharpMask. On the **third column** we report the PolygonRNN prediction without human intervention. Finally, on the **fourth column** we show a corrected prediction, in which we can observe how the segmentation is refined to better surround the car mirrors or their wheels.

References

[1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, 2016. 1

[2] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár. Learning to refine object segments. *ECCV 2016*, 2016. 1