

# The Role of Context for Object Detection and Semantic Segmentation in the Wild

Roozbeh Mottaghi<sup>1</sup>   Xianjie Chen<sup>2</sup>   Xiaobai Liu<sup>2</sup>   Nam-Gyu Cho<sup>3</sup>   Seong-Wan Lee<sup>3</sup>  
 Sanja Fidler<sup>4</sup>   Raquel Urtasun<sup>4</sup>   Alan Yuille<sup>2</sup>  
 Stanford University<sup>1</sup>   UCLA<sup>2</sup>   Korea University<sup>3</sup>   University of Toronto<sup>4</sup>

Due to a bug in the evaluation code and also inaccuracies in the original annotations, the results that we published in Tables 1 and 2 are not accurate. The correct accuracies can be found below:

	Recall	IOU		Recall	IOU
<b>bag</b>	2.1	1.2	<b>food</b>	16.4	10.7
<b>bed</b>	2.8	0.7	<b>mouse</b>	1.0	0.9
<b>bedcloth</b>	0.0	0.0	<b>plate</b>	10.2	5.6
<b>bench</b>	0.2	0.1	<b>platform</b>	9.9	7.5
<b>book</b>	13.5	5.0	<b>rock</b>	8.0	6.7
<b>cabinet</b>	6.7	4.4	<b>shelves</b>	15.1	3.7
<b>clothes</b>	3.3	1.8	<b>sidewalk</b>	0.6	0.5
<b>computer</b>	0.0	0.0	<b>sign</b>	11.2	7.0
<b>cup</b>	1.9	1.4	<b>snow</b>	20.8	16.4
<b>curtain</b>	22.1	11.6	<b>truck</b>	0.6	0.2
<b>door</b>	3.6	2.3	<b>window</b>	31.7	14.6
<b>fence</b>	10.9	6.6	<b>wood</b>	1.2	0.8
<b>flower</b>	14.6	6.8	<b>light</b>	14.3	8.5
			<b>Avg.</b>	8.6	4.8

Table 1. The subset of 59 most frequent classes that have low segmentation accuracy according to O<sub>2</sub>P [1] results.

## References

- [1] J. Carreira, R. Caseiroa, J. Batista, and C. Sminchisescu. Semantic segmentation with second-order pooling. In *ECCV*, 2012. 1
- [2] J. Tighe and S. Lazebnik. Superparsing: Scalable nonparametric image parsing with superpixels. In *ECCV*, 2010. 1

	Recall		IOU	
	SuperParsing [2]	O <sub>2</sub> P [1]	SuperParsing [2]	O <sub>2</sub> P [1]
<b>sky</b>	88.8	93.9	65.6	75.6
<b>grass</b>	68.0	77.7	45.3	56.0
<b>water</b>	44.8	72.0	34.5	54.8
<b>person</b>	72.8	57.6	30.1	44.5
<b>tree</b>	66.2	66.7	37.8	44.3
<b>bus</b>	22.8	70.1	14.0	43.2
<b>wall</b>	66.6	68.1	30.8	40.5
<b>cat</b>	36.5	66.4	20.1	36.7
<b>aeroplane</b>	29.3	67.2	19.5	36.4
<b>car</b>	31.2	55.5	15.0	33.5
<b>motorbike</b>	25.7	66.1	14.3	32.8
<b>road</b>	22.8	50.0	15.8	31.2
<b>track</b>	22.9	44.3	17.5	29.5
<b>ground</b>	48.9	41.8	24.0	27.6
<b>dog</b>	18.6	46.3	11.5	26.9
<b>train</b>	16.6	47.9	10.4	26.7
<b>horse</b>	2.2	44.8	2.0	26.4
<b>floor</b>	25.6	46.1	14.4	25.7
<b>bird</b>	4.9	42.7	4.1	24.6
<b>building</b>	45.7	31.4	19.8	24.3
<b>tvmonitor</b>	10.5	48.9	9.0	24.3
<b>sheep</b>	5.0	38.0	4.2	23.7
<b>bicycle</b>	16.6	52.5	11.3	23.5
<b>boat</b>	0.1	37.8	0.0	22.3
<b>mountain</b>	10.3	30.4	8.8	19.2
<b>keyboard</b>	0.1	34.6	0.1	18.2
<b>cow</b>	0.1	24.6	0.1	16.2
<b>sofa</b>	4.4	29.2	3.6	16.1
<b>pottedplant</b>	1.2	40.7	1.1	15.9
<b>bottle</b>	1.3	35.8	1.2	15.0
<b>ceiling</b>	9.6	20.1	6.4	12.7
<b>table</b>	9.6	11.3	6.4	7.0
<b>chair</b>	3.5	10.1	2.9	6.8
<b>Avg.</b>	25.3	47.6	15.2	29.1

Table 2. **Segmentation:** Nearest-neighbor methods such as [2] do not work well on PASCAL due to the high variability of images. In contrast the O<sub>2</sub>P classifier [1] on superpixels performs well.