

MonoCInIS: Camera Independent Monocular 3D Object Detection using Instance Segmentation : Supplementary

Jonas Heylen¹ Mark De Wolf¹ Bruno Dawagne¹ Marc Proesmans^{1,2} Luc Van Gool^{2,3}
Wim Abbeloos⁴ Hazem Abdelkawy⁴ Daniel Olmeda Reino⁴

¹TRACE vzw ²KU Leuven/ESAT-PSI ³ETHZ/CVL ⁴Toyota Motor Europe

In this supplementary, we provide some more qualitative results of the approach on the KITTI3D dataset. Figure 1 shows the results for the ResNet101 architecture. Only KITTI3D data is used for these trainings. All figures show that even small cars are well detected. Even when instances are not perfect, our method can still predict valid 3D bounding boxes. Also, occlusions and truncations are handled well.



Figure 1: Qualitative results on the KITTI3D test set on *Car* for the ResNet101 architecture: instance segmentation (top), 2D reference points (mid) and 3D bounding boxes (bottom).

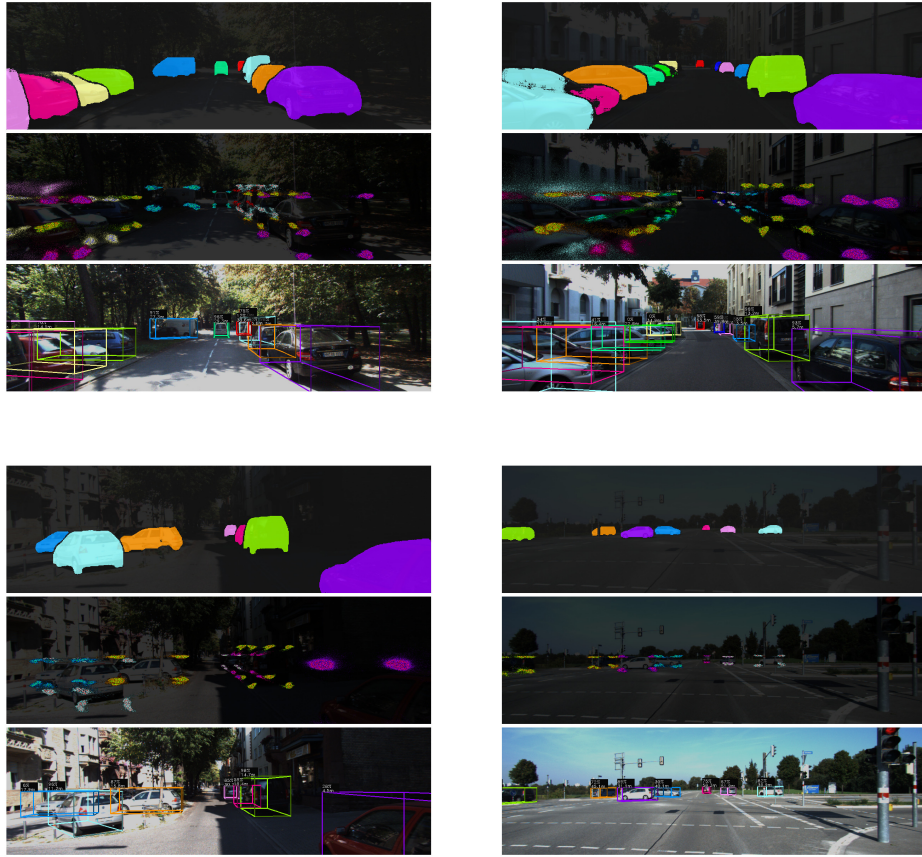


Figure 2: Qualitative results on the KITTI3D test set on *Car* for the ERFNet architecture: instance segmentation (top), 2D reference points (mid) and 3D bounding boxes (bottom).

Figure 2 shows the results for the ERFNet architecture. As expected and indicated in the paper, the performance for smaller architectures can drop substantially in the very strict KITTI3D benchmark. In contrast to these relatively low numbers, ERFNet performs reasonably well qualitatively for cars closer by.

The supplementary video, some snapshots below in fig. 3, shows the performance of the approach on video data. The images are processed frame-by-frame, and no temporal constraints have been used. Nevertheless, the predictions remain quite consistent over time.

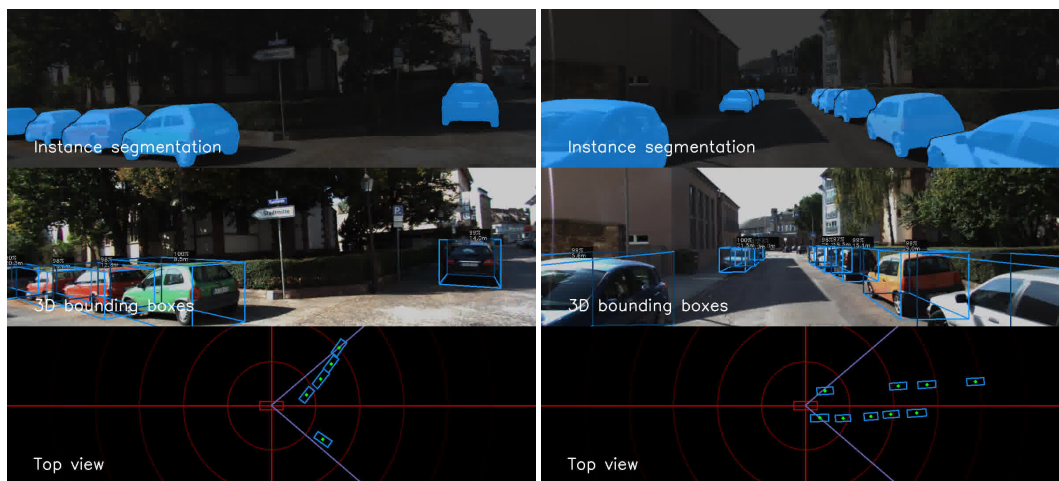


Figure 3: snapshots from the supplementary video