IDA-3D: Instance-Depth-Aware 3D Object Detection from Stereo Vision for Autonomous Driving (Supplemental Material)

Wanli Peng* Hao Pan* He Liu Yi Sun†
Dalian University of Technology, China
{1136558142,panhao15320,lhiceu}@mail.dlut.edu.cn, lslwf@dlut.edu.cn

1. Results on the KITTI test set

We report evaluation results on the KITTI [1] test set in Table 1. The detailed performance can be found at http://www.cvlibs.net/datasets/kitti/.

Methods	Sensor	AP_{bev}			AP_{3D}			Time	Environment
		Easy	Mode	Hard	Easy	Mode	Hard		
Ours	Stereo	61.87	42.47	34.59	45.09	29.32	23.13	83ms	RTX2080Ti

Table 1. Results of car category on the KITTI test set at IOU = 0.7.

2. Results on Pedestrian and Cyclist detection

We present our results on 3D pedestrian and cyclist detection, which are shown in Table 2. To the best of our knowledge, few prior works on image-based methods report this results. Both categories are much more challenging than car detection due to the small sizes of the objects. However, our method can still get promising performance on these two categories because our IDA module pays more attention to the global spatial information of the objects instead of predicting depth map and it is more effective to estimate the depth z of small objects compared with pseudo-LiDAR based methods.

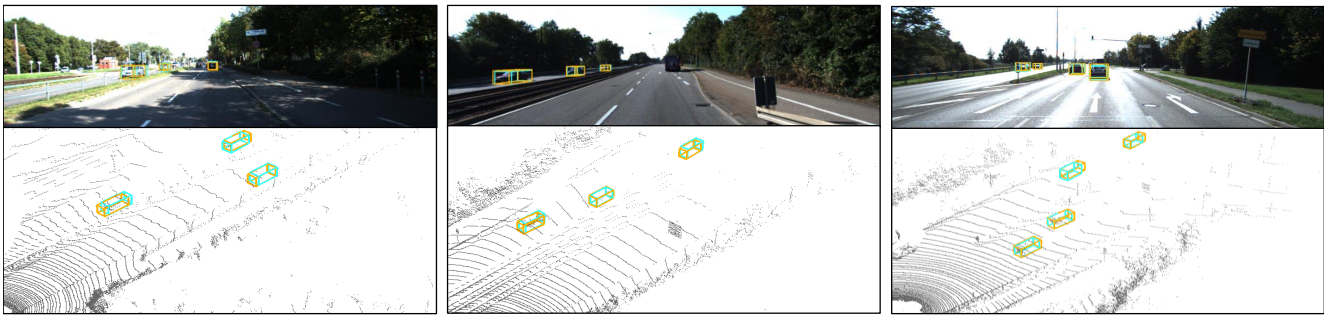
Methods	Sensor	AP_{bev}			AP_{3D}		
		Easy	Mode	Hard	Easy	Mode	Hard
Pedestrian							
Xinzhu et al. [2]	Mono+PL	14.30	11.26	9.23	11.29	9.01	7.04
PL+FP [3]	Stereo+PL	32.5	27.1	23.1	23.5	19.4	15.3
Ours	Stereo	**49.49**	**37.70**	**30.54**	**47.91**	**36.80**	**29.94**
Cyclist							
Xinzhu et al. [2]	Mono+PL	10.12	6.39	5.63	8.90	4.81	4.52
PL+FP [3]	Stereo+PL	35.4	23.7	22.0	28.5	19.3	18.2
Ours	Stereo	**42.84**	**24.23**	**23.87**	**41.25**	**23.17**	**22.96**

Table 2. Results on the KITTI validation set at IoU=0.5 (the standard metric), where PL donates pseudo-LiDAR based method.

3. More Qualitative Results

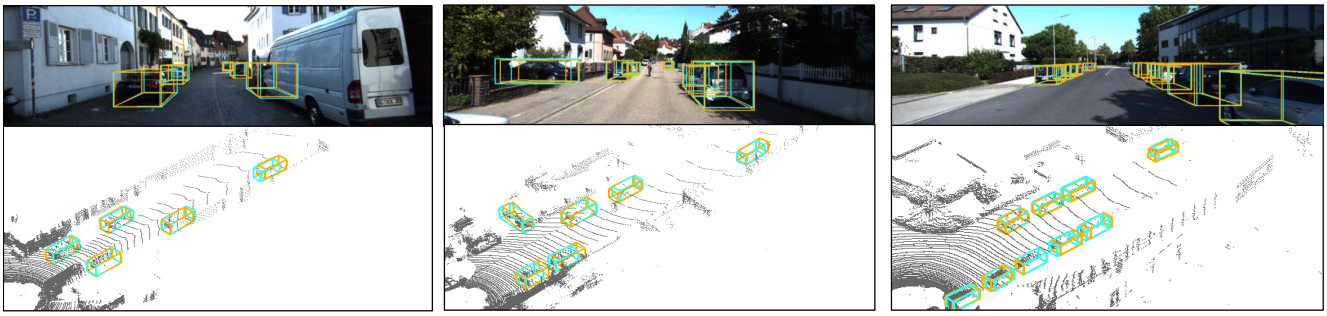
We visualize more detection results of hard examples on the KITTI dataset, where the predicted results are shown in yellow and the ground truth boxes are shown in blue. For the far-away objects as shown in Fig. 1, our method which benefits from our instance-depth-aware(IDA) module can accurately estimate their object locations. In the case there are too many vehicles in the scene as shown in Fig. 2, our method also has the potential to successfully detect these objects which is heavily occluded by others. Moreover, our method also has the ability to output accurate 3D bounding boxes even though the objects are seriously truncated by the image boundaries such as the object in lower-left corner of each image in Fig. 3, because it pays more attention to the global spatial information of the objects and doesn't predict the keypoints on the objects which may be truncated by image boundaries. As shown in Fig. 4, our method can also get accurate detection results on pedestrian and cyclist categories.

*The first two authors contributed equally to this work.
†Corresponding author.



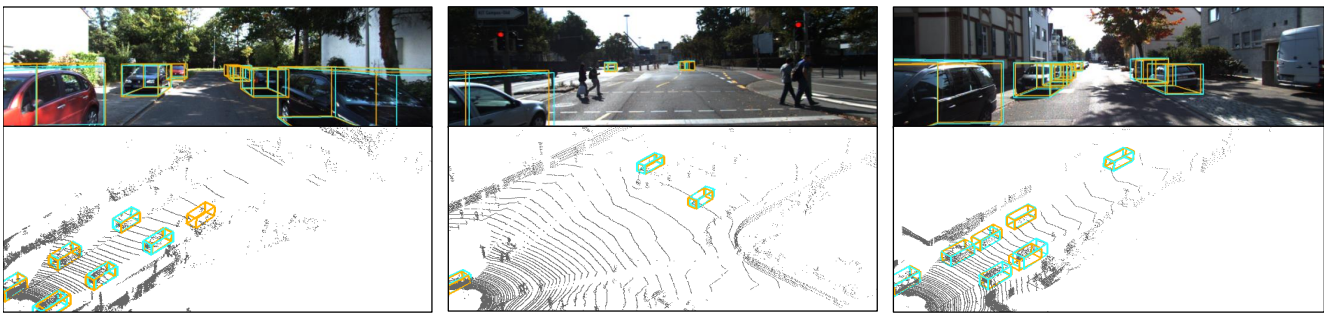
(a) (b) (c)

Figure 1. Detection results of far-away objects on the KITTI dataset.



(a) (b) (c)

Figure 2. Detection results of occluded objects on the KITTI dataset.



(a) (b) (c)

Figure 3. Detection results of truncated objects on the KITTI dataset.

(a) (b) (c)

Figure 4. Detection results of pedestrian and cyclist categories on the KITTI dataset.

References

[1] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[2] Xinzhu Ma, Zhihui Wang, Haojie Li, Pengbo Zhang, Wanli Ouyang, and Xin Fan. Accurate monocular 3d object detection via color-embedded 3d reconstruction for autonomous driving. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6851–6860, 2019.

[3] Yan Wang, Wei-Lun Chao, Divyansh Garg, Bharath Hariharan, Mark Campbell, and Kilian Q Weinberger. Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8445–8453, 2019.