Supplementary of MonoPair: Monocular 3D Object Detection Using Pairwise Spatial Relationships

Yongjian Chen Lei Tai Kai Sun Mingyang Li
Alibaba Group
{yongjian.cyj, tailei.tl, sk157164, mingyangli}@alibaba-inc.com

	AP_{bv} IoU≥0.5			AP_{3D} IoU≥0.5			AP_{bv} IoU≥0.7			AP_{3D} IoU≥0.7		
Methods	E	M	H	E	M	H	E	M	H	E	M	H
Mono3D [2]	30.50	22.34	19.16	25.19	18.20	15.52	5.22	5.19	4.13	2.53	2.31	2.31
OFTNet [6]	-	-	-	-	-	-	11.06	8.79	8.91	4.07	3.27	3.29
MF3D [8]	55.02	36.73	31.27	47.88	29.48	26.44	22.03	13.63	11.60	10.53	5.69	5.39
MonoPSR [3]	56.97	43.39	36.00	49.65	41.71	29.95	20.63	18.67	14.45	12.75	11.48	8.59
TLNet(mono) [5]	52.72	37.22	32.16	48.34	33.98	28.67	21.91	15.72	14.32	13.77	9.72	9.29
MonoGRNet [4]	54.21	39.69	33.06	50.51	36.97	30.82	24.97	19.44	16.30	13.88	10.19	7.62
MonoDIS [7]	-	-	-	-	-	-	24.26	18.43	16.95	18.05	14.98	13.42
M3D-RPN [1]	55.37	42.49	35.29	48.96	39.57	33.01	25.94	21.18	17.90	20.27	17.06	15.21
Baseline	54.50	40.87	34.45	48.22	36.80	31.97	25.69	18.97	16.48	14.69	10.27	9.06
$+\sigma^z + \sigma^{uv}$	58.49	48.57	43.20	54.68	42.43	40.17	26.22	21.09	19.78	19.27	16.57	14.50
MonoPair	**59.66**	**49.52**	**43.76**	**55.88**	**43.32**	**40.94**	**28.97**	**22.65**	**21.10**	**22.26**	**18.42**	**16.49**

Table 1: AP_{11} scores on KITTI3D validation set for car. E, M and H represent *Easy*, *Moderate* and *Hard* samples.

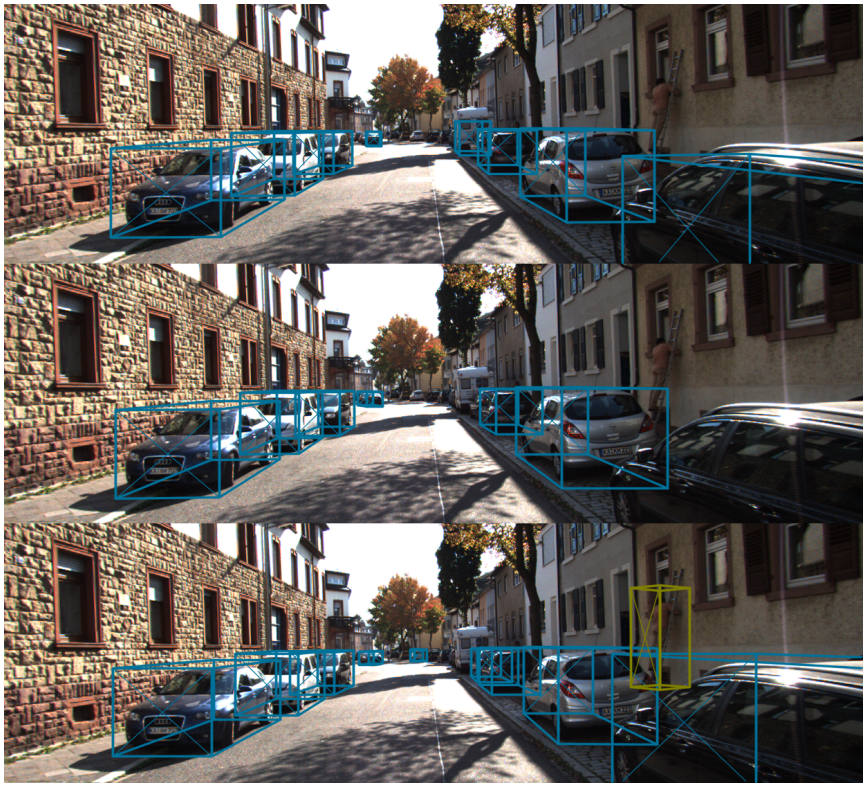
1. Additional Validation Results through AP_{11}

As mentioned in the main paper, previous methods mainly conduct evaluation experiments through the old metric AP_{11} on KITTI3D benchmark. Thus, To compare our method with more monocular 3D object detectors, we also show results on the KITTI3D validation set through AP_{11} as shown in Table 1. Our **Baseline**, as mentioned in the main paper, is mainly derived from CenterNet [9], which is designed specifically for 2D object detection. It can not catch a similar performance as state-of-the-art monocular 3D object detectors. However, with the proposed uncertainty-aware spatial constraint optimization, our **MonoPair** finally outperforms all of the other methods with a large margin.

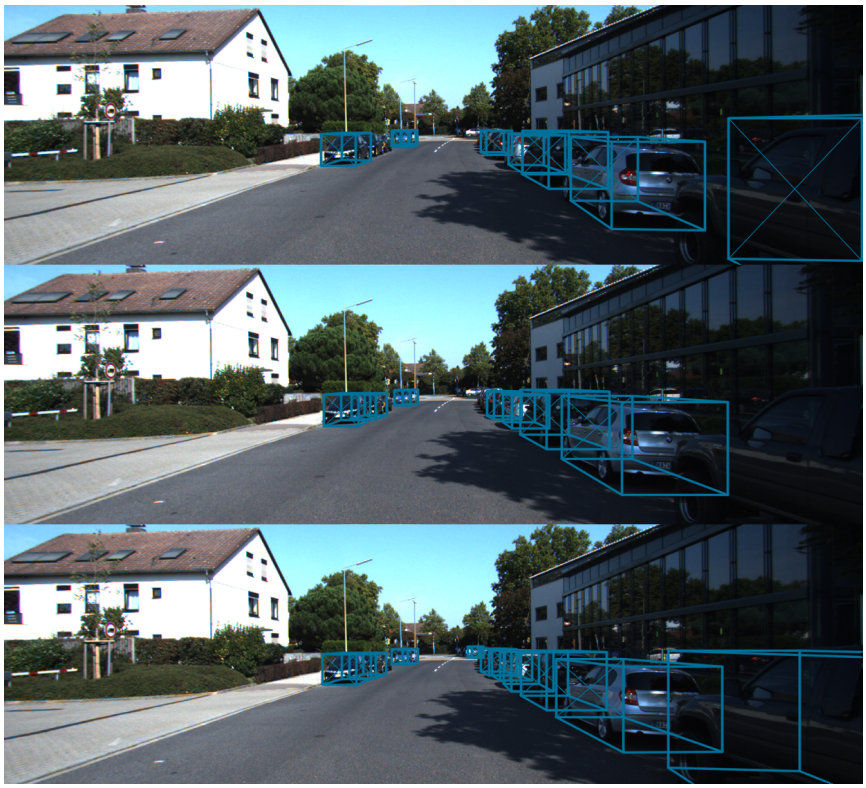
2. Additional Qualitative Results

We also present additional qualitative results on KITTI validation set as shown in Figure 1-4. We choose four different scenarios (block, town road, highway and city center), where two samples are selected from each of the scenarios. Predictions from MonoGRNet [4] and M3D-RPN [1], and our MonoPair are presented in each figure from top to down. Blue boxes mean predictions from cars. Yellow and gray boxes are predictions of pedestrians and cyclists respectively. The cross shows the predicted orientation of the 3D object. More results are also uploaded in the supplementary directory.

Compared with results from other detectors as shown in these figures, MonoPair shows a great ability to detect seriously occluded samples. It also provides a considerable bounding box for samples far away from the camera. However, MonoGRBet [4] and M3D-RPN [1] always neglect those occluded or further samples. Besides, MonoPair also provides much better orientation predictions as shown in all the figures. Figure 4 is mainly to show the detection ability for pedestrians and cyclists, which are trained from much fewer samples compared with cars. The proposed spatial constraint from the same category provides more information for training and shows much accurate predictions.



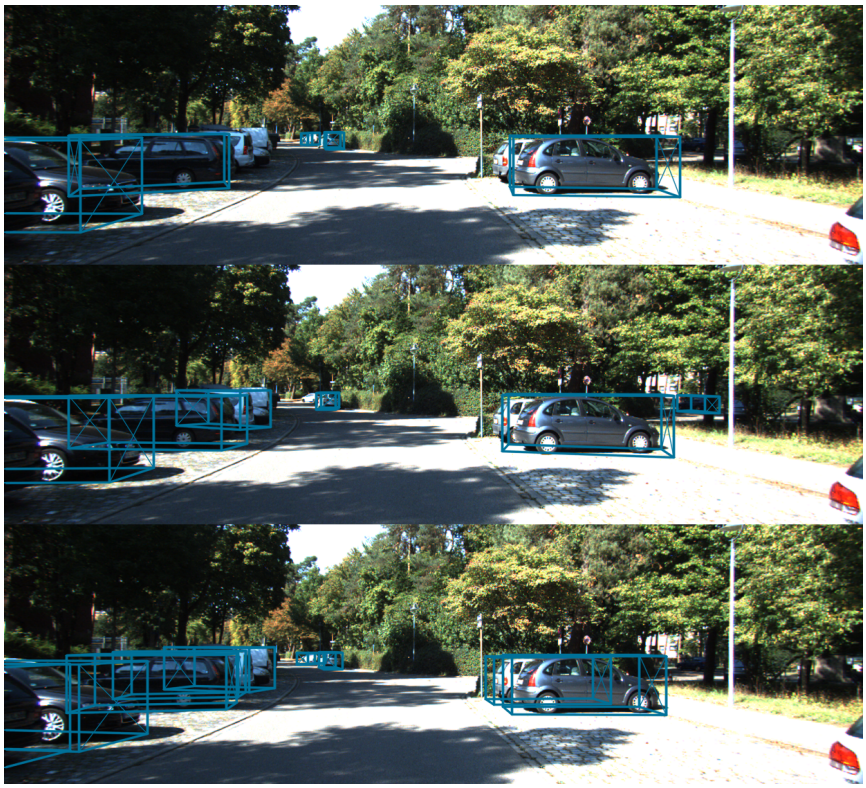
(a)



(b)

Figure 1: Predictions on two block scenarios from the KITTI validation set. Results are from MonoGRNet [4], M3D-RPN [1], and our MonoPair from top to down in both (a) and (b).

(a)



(b)

Figure 2: Predictions on two town road scenarios from the KITTI validation set. Results are from MonoGRNet [4], M3D-RPN [1], and our MonoPair from top to down in both (a) and (b).

(a)

(b)

Figure 3: Predictions on two highway scenarios from the KITTI validation set. Results are from MonoGRNet [4], M3D-RPN [1], and our MonoPair from top to down in both (a) and (b).

(a)

(b)

Figure 4: Predictions on two city center scenarios especially for pedestrians and cyclists from the KITTI validation set. Results are from MonoGRNet [4], M3D-RPN [1], and our MonoPair from top to down in both (a) and (b).

References

[1] Garrick Brazil and Xiaoming Liu. M3d-rpn: Monocular 3d region proposal network for object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019. 1, 2, 3, 4, 5

[2] Xiaozhi Chen, Kaustav Kundu, Ziyu Zhang, Huimin Ma, Sanja Fidler, and Raquel Urtasun. Monocular 3d Object Detection for Autonomous Driving. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2147–2156, Las Vegas, NV, USA, June 2016. IEEE. 1

[3] Jason Ku, Alex D. Pon, and Steven L. Waslander. Monocular 3d object detection leveraging accurate proposals and shape reconstruction. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1

[4] Zengyi Qin, Jinglu Wang, and Yan Lu. Monogrnet: A geometric reasoning network for monocular 3d object localization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8851–8858, 2019. 1, 2, 3, 4, 5

[5] Zengyi Qin, Jinglu Wang, and Yan Lu. Triangulation learning network: From monocular to stereo 3d object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1

[6] Thomas Roddick, Alex Kendall, and Roberto Cipolla. Orthographic feature transform for monocular 3d object detection. *arXiv preprint arXiv:1811.08188*, 2018. 1

[7] Andrea Simonelli, Samuel Rota Bulo, Lorenzo Porzi, Manuel Lopez-Antequera, and Peter Kontschieder. Disentangling monocular 3d object detection. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019. 1

[8] Bin Xu and Zhenzhong Chen. Multi-level Fusion Based 3d Object Detection from Monocular Images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2345–2353, Salt Lake City, UT, USA, June 2018. IEEE. 1

[9] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as Points. *arXiv:1904.07850 [cs]*, Apr. 2019. arXiv: 1904.07850. 1