Far3Det: Towards Far-Field 3D Detection
(Supplementary Material)

Shubham Gupta[1*] Jeet Kanjani[1*] Mengtian Li[1] Francesco Ferroni[2]
James Hays[2,3] Deva Ramanan[1,2 †] Shu Kong[4 †]
[1]CMU [2]Argo AI [3]Gatech [4]Texas A&M University

Overview

In this document, we first include detailed studies w.r.t different evaluation metrics to supplement Table 5 of the main paper. Then, we include open-source code for far-field 3D detection (Far3Det) benchmarking, including Far nuScenes frame ids, metrics, and different fusion baselines (NMS, AdaNMS, and CLOCs3D). Lastly, we attach a video demo to demonstrate the results of our lidar-based detector (CenterPoint [5]), image-based detector (FCOS3D [4]), and our NMS-based late-fused detector.

1. Far Field Annotations

We analyze the total number of far-field annotations per class as shown in Table 1.

2. Evaluation Metric

We evaluate all the baseline models and our proposed NMS and AdaNMS fusion methods for various thresholding schemes. For nuScenes default metric (average of AP at 0.5, 1, 2 and 4 meters), we can observe that the lidar-based method (CenterPoint-VoxelNet) has higher AP compared to the image-based method (FCOS3D), (16.5 vs 11.5) for distant objects (50-80m) in contrast to what we observed at our proposed linear adaptive threshold-based metric. This occurs due the higher noise in the models prediction as the distance from the ego-vehicle increases. Table 2 shows the performance of the various models at the evaluated on the 4m threshold. We can observe that this thresholding scheme follows the similar trend as our proposed linear and qudratic adaptive thresholding scheme.

3. Zero lidar Point Objects

As discussed in Section 3.2 of the main paper, we include the unoccluded zero-lidar point objects in our validation set. Table 3 summarizes our findings for this data. Fig. 1 and 2 visualize some examples that have zero lidar points on far-field objects. Table 4 supplements Table 7 of main paper, comparing different methods on far-field objects that have zero lidar points.

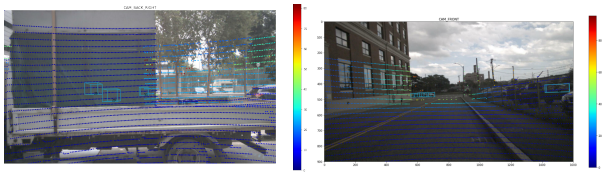

Figure 1: Visualization of objects (annotated with bounding boxes) that have zero lidar points. Left: objects do not have lidar returns because they are occluded by a vehicle in front; they are not included in Far nuScenes validation set. Right: objects in the far side of road are unoccluded and are included in our Far nuScenes validation set.

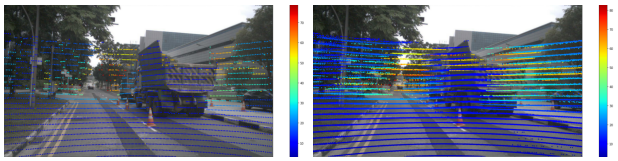

Figure 2: Comparison between single lidar sweep and multiple lidar sweeps. The left image has objects with zero lidar points on the far side of the road whereas when we consider multiple sweeps (right image), the objects has non-zero number of points on it.

4. Visualization, Demo, and Code

Fig. 3 supplements the Fig. 4 in the main paper, demonstrating a "failure mode" of image-based detectors as below. The detections are quite good in the projected image space but might have notable errors in the BEV. Existing evaluation protocols penalize such errors unreasonably too heavily, hence motivating our metrics that use distance adaptive thresholds.

We also attach a video demo (`demo.mp4`) in this supplemental material, visualizing how our method improve 3D detections in the far-field. We further attach our code. We will publish them along our paper.

*Co-first authors. †Co-last authors.

Table 1: Class-specific far-field annotations (50-80m) in the nuScenes [1] dataset. Here "Construct" and "Cone" are short for "Construction Vehicle" and "Traffic Cone" respectively. We perform this analysis for nuScenes Training set, nuScenes Validation set, and Far nuScenes. In general, there are fewer annotations for small objects ("Motorcycle", "Bicycle", "Traffic Cone") than big ones ("Car", "Truck", "Bus").

	50-80m									
Dataset	Car	Truck	Bus	Trailer	Construct	Pedestrian	Motorcycle	Bicycle	Cone	Barrier
Training	49758	15443	3751	5072	2345	16106	508	334	1424	5382
Validation	10435	3541	673	1084	740	3339	123	85	169	1469
Far nuScenes (Validation)	5043	1791	184	743	401	1427	9	35	81	765

Table 2: AP computed with 4m threshold. All numbers improve compared to the default nuScenes metric that averages AP over thresholds *up* to 4m. Here, image-based far-field 3D detections are *far* more accurate than lidar detections (37 vs 28.1 for distant Cars from 50-80m). Fusion once again further improves results. Similar trends hold for Truck and Pedestrian. Overall, results and trends are qualitatively similar to those for a linearly-growing threshold (the default used in our paper).

	Modality		Car		Truck		Pedestrian	
Model	lidar	Camera	0-50m	50-80m	0-50m	50-80m	0-50m	50-80m
CP-VoxelNet* [5]	✓		**95.3**	28.1	**71.7**	12.3	94.1	17.3
CP-PointPillars [5]	✓		93.7	14.3	63.4	2.7	87.4	6.7
PointPillars-FPN [3]	✓		91.2	7.7	54.3	0.7	85.3	1.9
PointPillars-SECFPN [3]	✓		89.7	5.8	62.4	1.4	81.3	1.9
SSN-SECFPN [7]	✓		91.1	8.3	61.4	1.2	78.1	1.0
FCOS3D [4]		✓	87.2	37.0	48.5	8.1	67.9	12.6
Bayesian Fusion [2]	✓	✓	94.8	50.8	68.7	18.2	94.3	22.5
CLOCs3D**	✓	✓	94.8	50.8	68.7	18.2	**94.3**	22.5
NMS Fusion (Ours)	✓	✓	95.3	50.5	71.7	20.3	94.1	**23.8**
AdaNMS Fusion (Ours)	✓	✓	95.3	53.1	71.7	21.0	94.1	19.7
MVP [6]	✓	✓	**96.47**	69.33	**76.12**	43.21	**96.08**	**58.89**
AdaNMS (MVP, FCOS3D)	✓	✓	**96.47**	**69.89**	**76.12**	**43.44**	**96.08**	56.61

References

[1] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[2] Yi-Ting Chen, Jinghao Shi, Christoph Mertz, Deva Ramanan, and Shu Kong. Multimodal object detection via probabilistic ensembling. In *European Conference on Computer Vision (ECCV)*, 2022.

[3] Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollár. Designing network design spaces. In *CVPR*, 2020.

[4] Tai Wang, Xinge Zhu, Jiangmiao Pang, and Dahua Lin. Fcos3d: Fully convolutional one-stage monocular 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 913–922, 2021.

[5] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Center-based 3d object detection and tracking. In *CVPR*, 2021.

[6] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Multimodal virtual point 3d detection. *NeurIPS*, 2021.

[7] Xinge Zhu, Yuexin Ma, Tai Wang, Yan Xu, Jianping Shi, and Dahua Lin. Ssn: Shape signature networks for multi-class object detection from point clouds. In *ECCV*, 2020.

Table 3: Annotations count for Zero-lidar point boxes in the nuScenes [1] dataset. We can observe that the unoccluded lidar point objects account for $\geq$10% of the annotations in the far-field (50-80m)

	50-80m		
Dataset	Non-zero lidar Point Boxes	Zero-lidar Point Boxes	Unoccluded Zero-lidar Point Boxes
Training	100123	45050	10430
Validation	21658	10069	2655
Far nuScenes (Validation)	10470	5126	1623

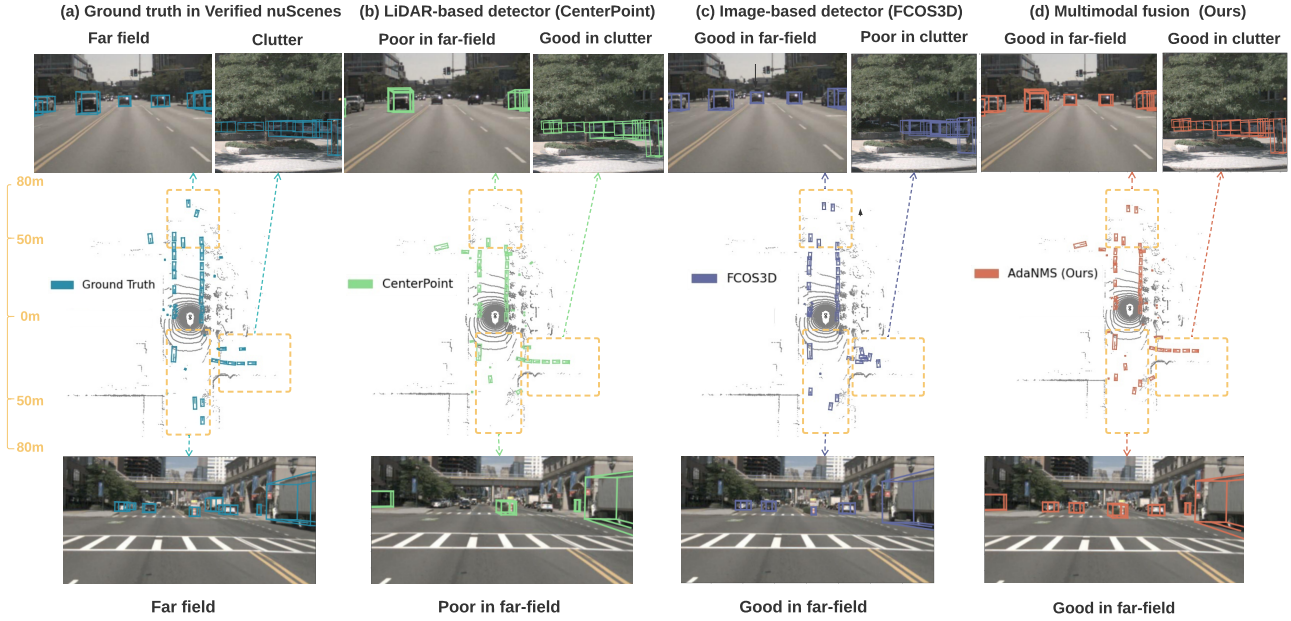

Figure 3: Qualitative results on Far nuScenes. First, we note that our Far nuScenes contain clean annotations at a distance and for scenes with cluttered objects (a). By comparing the predictions between CenterPoint (b) and FCOS3D (c), we observe the image-based FCOS3D contains higher quality predictions for Far3Det, which is not reflected in the standard evaluation (Table 3 in the main paper). Our proposed fusion method AdaNMS (d) leverages their respective advantages and greatly improves detection of far-field objects.

Table 4: (This table supplements Table 7 of main paper.) We include the unoccluded objects with zero lidar points using strategy described in Section 3.2 and calculate the mAP for 50-80m distance range. Clearly, our AdaNMS (MVP+FCOS3D) outperforms others on all categories except the Pedestrian and Trailer category. We see a slight decline in performance on this class as a distance adaptive IOU hurts the recall in a cluttered scene. Note, we don't include the classes Motorcycle, Bicycle and Traffic Cone as the number of annotations are quite low Table 1.

Method	50-80m						
	Car	Truck	Bus	Trailer	Construction Vehicle	Pedestrian	Barrier
CP [5]	20.5	8.6	0.1	1.1	0.0	11.2	18.0
FCOS3D [4]	35.2	9.7	6.9	11.2	0.0	11.7	38.6
AdaNMS (CP, FCOS3D)	45.2	18.5	7.1	5.4	0.0	14.4	41.4
MVP [6]	55.2	35.5	26.0	**17.5**	**1.6**	**44.2**	37.0
AdaNMS (MVP, FCOS3D)	**60.3**	**36.0**	**27.0**	16.1	**1.6**	42.6	**48.8**