BoostRad: Enhancing Object Detection by Boosting Radar Reflections Supplementary Material

Yuval Haitman* and Oded Bialer*
General Motors, Technical Center Israel
haitman@post.bgu.ac.il, oded.bialer8@gmail.com

1. Supplementary Performance Evaluation

In this section we provide a performance analysis of *BoostRad*, expanding upon the insights provided in Section 4 of the main paper.

1.1. Reflection Boosting Performance Evaluation

In this section, we assess the boosting DNN's accuracy in detecting reflection points. Notably, objects consist of multiple reflection points. Our evaluation involves comparing the boosting DNN's output to ground truth reflection points using a specific metric. The ground truth reflection points were determined as the central positions of pixels in the 'super-radar' reflection image with intensities surpassing a predefined noise threshold. On the other hand, detection points in the boosting DNN's output were considered as the centers of pixels surpassing a detection threshold. A precision-recall curve was generated for varying detection thresholds. Detection points within a 25cm radius of a ground truth point were considered true positive detections. The average precision (AP) was computed by the area under the precision-recall curve.

Fig. 1 presents the precision-recall curve of the boosting DNN output. The boosting DNN was trained as outlined in Section 4.1 of the main paper. The precision-recall were calculated with a set of 5,000 synthetic examples. These examples were generated through the simulation detailed upon in Section 3.3 of the main paper and were distinct from the training dataset. The figure offers a comparative perspective by presenting the precision-recall curve of the original radar reflection intensity images from the same test set, prior to undergoing the boosting DNN's influence.

The results presented in Fig. 1 clearly indicate a substantial enhancement in the accuracy of radar reflection detection achieved by the boosting DNN when compared to the original radar reflection intensity image. Notably, the average precision (AP) attained by the boosting DNN exceeds that of the original image by more than twofold. This improvement can be attributed to the boosting DNN's ability to generate a radar image that is both sharper and cleaner than the original image, as evidenced by the qualitative examples presented in Section 4.2 of the main paper and in Section 1.2 below.

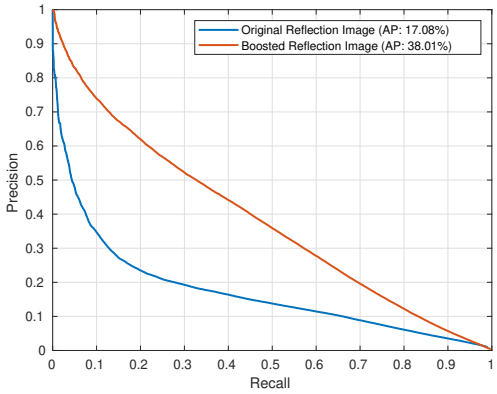

Figure 1. Reflection detection accuracy of the boosting DNN output image compared to the original intensity image.

1.2. Qualitative Evaluation With High-End Radar

In Figure 6 of the main paper, we showcase qualitative examples of *BoostRad* using the RADDet dataset. This section presents additional qualitative examples, this time employing a higher-end radar compared to the one in RADDet. The specifications of this higher-end radar system can be found in Table 1. Notably, it has three times higher range and four times higher azimuth angle resolution than the RADDet radar. We trained the boosting DNN using synthetic radar images produced via simulation of the higher-end radar. Subsequently, we tested both the boosting DNN and the object detection DNN using real measurements collected from this higher-end radar.

Fig. 2 presents qualitative examples of real measure-

*Both authors contributed equally to this work.
Both authors are with General Motors, Yuval Haitman is also with the School of Electrical and Computer Engineering in Ben Gurion University of the Negev.

ments from the higher-end radar. The original radar reflection image with the 'U-Net' object detection reference (see section 4.1 in the main paper) is presented in column (a). The output image of the boosting DNN and the BoostRad detected objects are presented in column (b), while column (c) showcases a scene photograph and an enlarged view of the scene's vehicles. Each row is a different scenario. In (a) and (b), the detected bounding boxes are denoted in purple, and the ground truth bounding boxes are denoted in white. In (c), the ground truth boxes are marked in red. Notably, the wide point spread function introduces clutter that obscures vehicles with lower reflectivity in (a). Consequently, this leads to the miss-detection of 'car 4' in Scenario 1, 'car 1' in Scenario 2, and 'car 3' in Scenario 3. In contrast, images in (b) exhibit considerably reduced clutter due to the narrower spreading function, facilitating the detection of all vehicles. Another observation of the narrower spreading function in (b) compared to (a) is apparent when examining the lighting poles that are located along a line on the right side of the image in scenario 1 and on the left side of the images in scenario 2 and 3 (as indicated in the figure). These lighting poles possess strong reflectivity, causing their spreading function main-lobe and side-lobes to stand out in (a). In contrast, (b) presents sharper lighting poles with significantly narrower spreading functions, resulting in reduced masking of other objects within the scene.

Radar parameter	
Maximal range	$150m$
Range resolution	$0.28cm$
Azimuth field of view	$(-60° : +60°)$
Azimuth resolution	$3.9°$
Elevation field of view	$(-20° : +20°)$
Elevation resolution	none
Number of Tx antennas	4
Number of Rx antennas	8
Carrier frequency	$77GHz$
Sampling rate	$25Mhz$
Waveform	Fast chirps FMCW

Table 1. Specifications of radar used in Sections 1.2.

1.3. Object Detection AP at Relatively High IOU

Table 1 of the main paper (in Section 4.1) presented Average Precision (AP) comparison between *BoostRad* and the reference methods for IOU thresholds 0.1 and 0.3. Additional AP results are presented in Table 2 of the supplementary material, specifically at an IOU of 0.5 for the 'car' class within the RADDet dataset. As expected, the AP scores for all methods exhibit a reduction at an IOU of 0.5 in comparison to IOU values of 0.1 and 0.3. Nevertheless, the performance advantage of *BoostRad* compared to the reference methods is still apparent.

Table 2. Object Detection Average Precision on RADDet Dataset for class 'car' in all ranges ($R > 0$) at IOU 0.5.

Method	Boosting	@0.5
RADDet [4]	✗	47.98
	✓	50.70 (+2.72)
Probalistic [2]	✗	40.68
	✓	45.16 (+4.48)
U-Net	✗	51.50
BoostRad	✓	**52.75** (**+1.25**)

2. Radar Simulation Details

Section 3.3 of the primary paper offers a description of the simulation processing procedures. In this accompanying section, we expound upon the mathematical expressions employed in the simulation's implementation. The simulation involves a Multiple-Input Multiple-Output (MIMO) radar that emits a rapid chirp Frequency-Modulated Continuous-Wave (FMCW) waveform [1, 3]. The transmission signal as a function of time, t, is mathematically represented as follows:

$$x(t) = \sum_{m=0}^{M-1} s(t - mT_c), \tag{1}$$

where T_c is the chirp duration, M is the number of chirps per frame, and

$$s(t) = \begin{cases} e^{-j2\pi(f_c t + \frac{1}{2}\alpha t^2)}, & \text{if } 0 \le t \le T_c \\ 0, & \text{otherwise ,} \end{cases} \tag{2}$$

is a single chirp, where $j = \sqrt{-1}$ is the imaginary unit, f_c is the carrier frequency and α is the chirp slope. The received signal at the k^{th} receive antenna is a superposition of all the reflected signals from all transmit antennas, which is expressed by

$$y_k(t) = \sum_{i=0}^{I-1} \sum_{q=0}^{Q-1} c^q x(t - \tau_{k,i}^q), \tag{3}$$

where Q are the number of reflection points, c_q is the q^{th} reflection point's intensity, I is the number of transmit antennas, i is the transmit antenna index, and $\tau_{k,i}^q$ is the round

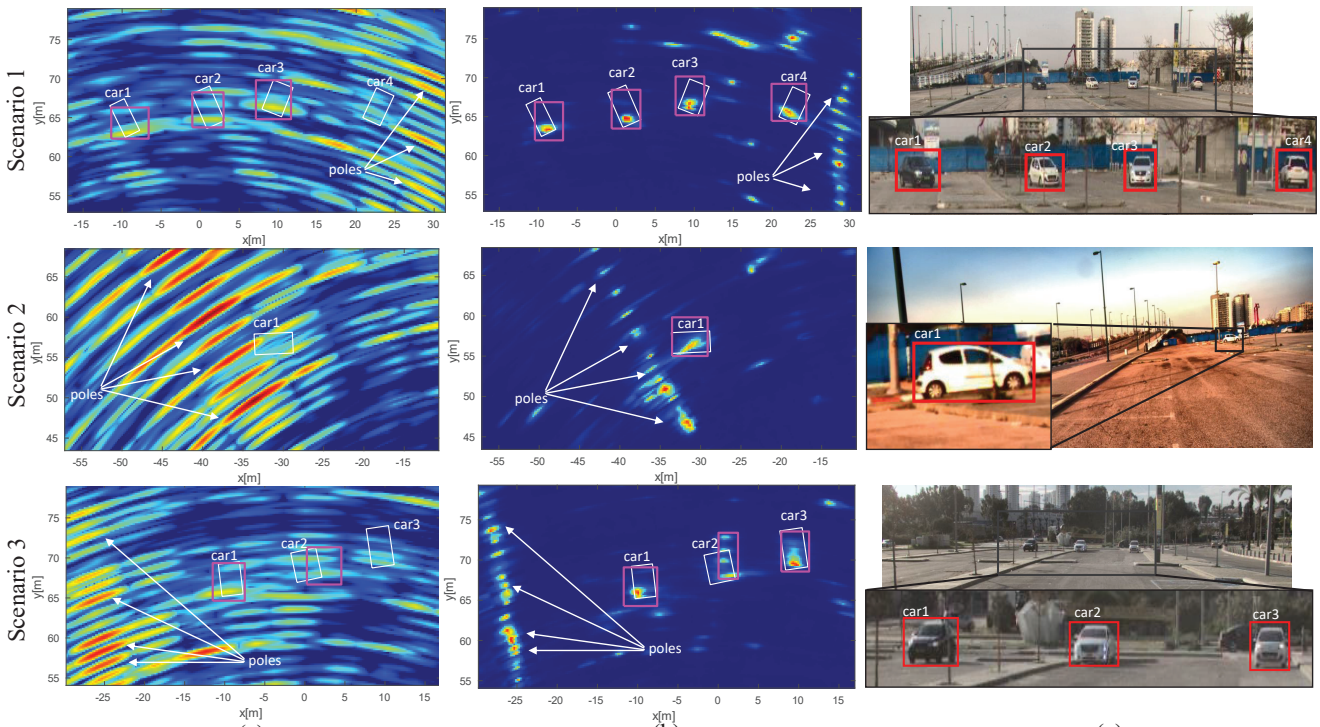

Figure 2. Qualitative examples with a higher-end radar than the radar used in RADDet. Each row is a different scenario. (a) original radar reflection image with the 'U-Net' reference detections, (b) boosting DNN output image with BoostRad detections, (c) camera image of the scene with ground truth bounding boxes. In (a) and (b), detection bounding boxes are purple, and ground truth bounding boxes are white. The vehicles have an identification number per each scenario that is written next to the ground truth bounding box.

trip delay between the i^{th} transmit antenna to the q^{th} reflection point and back to the k^{th} received antenna.

To derive the radar reflection intensity image, the received signal undergoes conventional FMCW radar processing [1]. This process encompasses down-conversion of the received signal through multiplication with the transmit reference signal (1), a range Fast Fourier Transform (FFT) for each chirp, a Doppler FFT across range bins over chirps, and beamforming for each range-Doppler bin across antennas.

3. Additional Implementation Details

Additional implementation details to those given in Section 4 of the main paper are specified below. The probability mapping of the ground truth reference given in Eq. 1 of the main paper was calculated with a fixed noise variance $\sigma_n^2 = 8 \times 10^{-5}$ and signal variance $\sigma_s^2 = 100 R_{max}^2 \sigma_n^2 / r^2$, where $R_{max} = 50[m]$ is the maximal range of the TI prototype radar (used in RADDet and CARRADA) and r is the range of the ground truth reference pixel. The hyper-parameters for the Boosting DNN loss in Eq. 2 of the main paper are: $(\rho_r, \rho_n, \rho_s) = (0.1, 5, 1)$. For the object detection loss, we weighted the L2 regression loss with factor 10^{-3} compared to the classification loss.

4. RADDet Train-Test Set Split

As outlined in Section 4 of the main paper, the performance evaluation conducted in the main paper involved a train-test set split of the RADDet dataset that is different than the split proposed in [4]. Table 3 presents the results of *BoostRad* and the reference methods on the original RADDet train-test split. The results show that all methods achieved higher results compared to Table 1 in the main paper, and *BoostRad* does not attain a performance gain compared to the reference methods. These results can be attributed to the fact that the test and training images in the original split [4] were derived from the same scenarios, often with small temporal gaps. As a result, a strong correlation is established between the test and training samples, leading to overfitting of all methods on the test set.

To address the issue of overfitting, we adopted a train-test set partitioning approach that ensures that the training and testing sets encompass disparate scenarios. The RADDet dataset comprises 15 distinct scenes, each detailed in Table 4. In our partitioning scheme, we designated scenes 9 and 11 for the test set, while the remaining scenes were uti-

Table 3. Object Detection Average Precision on RADDet With Original Train-Test split in [4]

Method	Boosting	Car					Person	
		$R \geq 0$			$R \geq 40$		$R \geq 0$	
		@ 0.1	@ 0.3	@ 0.5	@ 0.1	@ 0.3	@ 0.1	@ 0.3
RADDet [4]	✗	93.82	88.03	68.71	85.76	62.73	69.40	43.48
Probalistic [2]	✗	92.87	86.36	66.30	81.55	58.58	64.29	48.10
U-Net	✗	94.68	89.59	71.00	88.35	68.09	67.00	50.21
BoostRad	✓	93.58	88.59	67.29	85.40	64.96	66.66	51.39

Scene ID	Frame Numbers
0	$0-439, 559-724, 1549-1971$
1	$440-555, 731-1548, 1972-2571$
2	$2572-3038$
3	$3039-3437$
4	$3438-3653$
5	$3654-4073$
6	$4074-4331$
7	$4332-5018, 5623-6243$
8	$5019-5622, 6244-6608$
9	$6609-8046$
10	$8047-8634$
11	$8635-9158$
12	$9159-9437$
13	$9438-9745, 10175-10292$
14	$9746-10174$

Table 4. RADDet partition into distinct scenes

lized for the training set. Our modified training set encompasses a total of 17,021 cars (compared to the original partition's count of 16,755) and 5,240 pedestrians (compared to the original 5,210). For the test set, we have 4,094 cars (as opposed to 4,135 in the original) and 1,011 pedestrians (as opposed to 1,280).

5. Probability Mapping Derivation

In this section we provide a detailed derivation of the mapping of ground truth 'super-radar' intensity to probability given in Eq. 1 of the main paper. Denote by z_i the complex value of the i^{th} pixel of the ground truth reference image. We assume that z_i is a complex Gaussian random variable with zero mean and variance that could be either of noise or a signal, denoted by σ_n^2 and σ_s^2, respectively. The intensity (energy) of a complex Gaussian random variable has Chi-square distribution with 2 degrees of freedom. Let H_0 and H_1 denote the hypotheses that the i^{th} pixel is a noise or signal pixel, respectively. Then the probability density function of the i^{th} pixels' intensity given each hypothesis can be expressed as

$$p(|z_i|^2|H_0) = \frac{1}{2\sigma_n^2} e^{-|z_i|^2/(2\sigma_n^2)}, \tag{4}$$

$$p(|z_i|^2|H_1) = \frac{1}{2\sigma_s^2} e^{-|z_i|^2/(2\sigma_s^2)}. \tag{5}$$

The probability mapping of the ground truth reference intensity is the posteriori probability density function $p(H_1||z_i|^2)$. From Bayes rule and the law of total probability we have that

$$p(H_1||z_i|^2) = \frac{p(|z_i|^2|H_1)p(H_1)}{p(|z_i|^2)} = \frac{p(|z_i|^2|H_1)p(H_1)}{p(|z_i|^2|H_1)p(H_1) + p(|z_i|^2|H_0)p(H_0)}. \tag{6}$$

By substituting (4) and (5) into (6), and by assuming that $p(H_0) = p(H_1) = 0.5$ we arrive at

$$p(H_1||z_i|^2) = \frac{e^{-|z_i|^2/(2\sigma_{s_i}^2)}}{e^{-|z_i|^2/(2\sigma_s^2)} + \frac{\sigma_{s_i}^2}{\sigma_n^2} e^{-|z_i|^2/(2\sigma_n^2)}}, \tag{7}$$

which is the probability mapping given in Eq. 1 of the main paper.

References

[1] Oded Bialer, Amnon Jonas, and Tom Tirer. Code optimization for fast chirp FMCW automotive MIMO radar. *IEEE Transactions on Vehicular Technology*, 70(8):7582–7593, 2021. 2, 3

[2] Xu Dong, Pengluo Wang, Pengyue Zhang, and Langechuan Liu. Probabilistic oriented object detection in automotive radar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 102–103, 2020. 2, 4

[3] Ziqiang Tong, Ralf Renter, and Masahiko Fujimoto. Fast chirp FMCW radar in automotive applications. In *IET International Radar Conference*, pages 1–4. IET, 2015. 2

[4] Ao Zhang, Farzan Erlik Nowruzi, and Robert Laganiere. Raddet: Range-azimuth-doppler based radar object detection for dynamic road users. In *2021 18th Conference on Robots and Vision (CRV)*, pages 95–102. IEEE, 2021. 2, 3, 4