BoostRad: Enhancing Object Detection by Boosting Radar Reflections

Yuval Haitman* and Oded Bialer*

General Motors, Technical Center Israel

haitman@post.bgu.ac.il, oded.bialer8@gmail.com

Abstract

Automotive radars have an important role in autonomous driving systems. The main challenge in automotive radar detection is the radar's wide point spread function (PSF) in the angular domain that causes blurriness and clutter in the radar image. Numerous studies suggest employing an 'end-to-end' learning strategy using a Deep Neural Network (DNN) to directly detect objects from radar images. This approach implicitly addresses the PSF's impact on objects of interest. In this paper, we propose an alternative approach, which we term "Boosting Radar Reflections" (BoostRad). In BoostRad, a first DNN is trained to narrow the PSF for all the reflection points in the scene. The output of the first DNN is a boosted reflection image with higher resolution and reduced clutter, resulting in a sharper and cleaner image. Subsequently, a second DNN is employed to detect objects within the boosted reflection image. We develop a novel method for training the boosting DNN that incorporates domain knowledge of radar's PSF characteristics. BoostRad's performance is evaluated using the RADDet and CARRADA datasets, revealing its superiority over reference methods.

1. Introduction

Automotive radars play an important role in autonomous driving systems, offering extended detection range and resilience to adverse weather and lighting conditions. These radars emit RF signals, which interact with objects and bounce back to the receiving antennas as echoes. By applying signal processing algorithms [6], the received signals are coherently combined over a brief duration (e.g., 20ms) and across multiple antennas to generate a radar image. This image, characterized by dimensions of range and angle, captures the intensity of the radar's received energy at range and angle coordinates of each pixel. Serving as a visual representation of the scene, the radar image is further

---
*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.
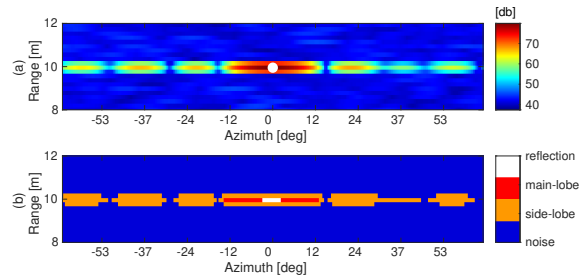
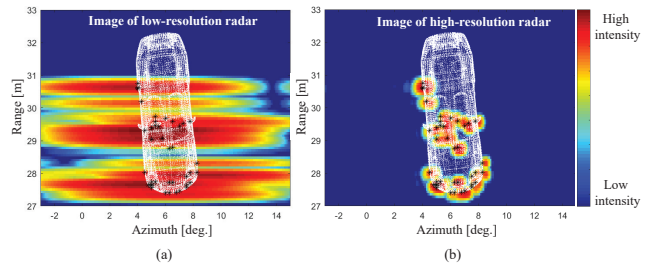Figure 1. Radar's point spread function (PSF).

Figure 2. Comparison of low-resolution realistic automotive radar reflection image (a) and high-resolution 'super-radar' image (b). White points depict vehicle, black asterisk shows radar reflection point from vehicle. Reflection points in (a) have wide PSF, while in (b) the PSF is narrow.

processed using computer vision algorithms to extract valuable scene information, thereby enhancing the perception of the environment.

One of the key obstacles in object detection from radar images is the presence of blurriness and excessive clutter noise. These issues arise due to the wide point spread function (PSF) of the radar in the angular domain, which is a consequence of its coarse angular resolution. Fig.1(a) illustrates the spreading function of a single reflection point, represented by a white point. The PSF consists of a main-lobe centered on the reflection point, along with multiple angular side-lobes that are attenuated replicas of the main-lobe at angle offsets from the main-lobe. These main-lobe and side-lobe components are depicted with different colors in Fig.1(b). In radar images, reflections from objects appear as multiple points, each exhibiting a wide PSF in the angular domain. Fig. 2(a) shows a radar image of a simu-

lated car (white dots) with black asterisks marking the reflection points. The PSF centered at each black asterisk has an angular resolution of $8°$, which is commonly found in production-grade automotive radars. The wide main-lobe of the PSF causes image blurriness, making it difficult to accurately determine object shape, orientation and class, as well as discriminating between closely located objects. Additionally, the PSF's side lobes contribute to image clutter in scenarios involving multiple objects, resulting in both miss-detection of objects with low reflection intensity and false detection of clutter as actual objects.

Fig.2(b) illustrates the radar image for the same reflection points as in Fig.2(a). However, in this case, the radar used has an angular resolution of $0.1°$, resulting in a PSF with significantly narrower main-lobe and lower side-lobes. It is clear that this narrower PSF allows for better estimation of object characteristics and eliminates clutter. Nevertheless, achieving such high angular resolution through hardware in automotive radars presents practical challenges. It would necessitate a larger physical size and a greater number of antennas, making it difficult to mount on vehicles while also increasing system complexity and cost [4]. Therefore, it is crucial to explore algorithmic alternatives that can reduce the PSF of practical radars, such as demonstrated in Fig. 2(a).

Several notable works [7, 10, 18, 25, 31, 39, 40, 42, 43] have proposed an *'end-to-end'* learning approach for object detection, which involves training a Deep Neural Network (DNN) to estimate objects directly from the radar image using ground truth reference bounding boxes. In this approach, the adverse effects of the radar's PSF on the objects of interest are implicitly mitigated. In contrast, our paper presents a distinct approach called Boosting Radar Reflections (*BoostRad*). In *BoostRad*, a boosting DNN is trained to narrow the main-lobe and reduce the side-lobes of the PSF for all reflection points in the scene. The output, referred to as the boosted image, is a higher-resolution and cleaner radar image with reduced clutter. Subsequently, a second DNN is employed to perform object detection from the boosted image. The boosting DNN aims to transform a low-resolution image, illustrated in Fig. 2(a), into a representation that closely resembles the high-resolution radar image depicted in Fig.2(b). This transformation enables enhanced object detection capabilities.

We present a novel approach for training the boosting DNN, distinguished by two key aspects. Firstly, it utilizes a deep learning technique that incorporates domain knowledge of the physical PSF characteristics of the radar. This is achieved through the integration of a unique ground truth reference of a high-resolution radar image with narrow PSF that is converted to a reflection probability map, and a customized loss function, enabling effective narrowing of the PSF. Secondly, to overcome the practical issue of acquiring real reference images with narrow PSF, we have developed a radar simulation that generates synthetic data. This allows us to train the boosting DNN exclusively on synthetic samples. Notably, the trained model demonstrates promising performance when applied to real radar images.

The performance of *BoostRad* was assessed using the RADDet [42] and CARRADA [28] datasets, which comprise automotive radar images taken from various scenarios. The evaluation clearly demonstrates the performance advantage of *BoostRad* over multiple reference *'end-to-end'* object detection methods.

The main contributions of this paper are:

1. A novel technique that narrows the PSF in radar images, leading to improved object detection. This technique stands out due to its deep learning approach that incorporates domain knowledge of the physical sensor's PSF characteristics. The method can also be extended to tackle the wide PSF issue in other image-producing sensors, such as ultrasound, MRI, CT, telescopes, and low-end cameras.

2. Insight into the debate between *'end-to-end'* and multi-stage object detection approaches. The paper challenges the prevailing trend of *'end-to-end'* object detection approaches by demonstrating the superior performance of multi-stage methods in radar images. By enhancing the image prior to object detection, the proposed technique surpasses *'end-to-end'* approaches, prompting further investigation of multi-stage approaches in the field of computer vision.

3. Highlighting the value of synthetic simulation data for radar images. This paper successfully trained a DNN to narrow the PSF using synthetic data alone. The demonstrated effectiveness of the trained DNN on real radar data serves as a motivation for further exploration and utilization of radar simulation in computer vision. Additionally, the details of the developed radar simulation are disclosed to facilitate its utilization, thereby promoting further progress in radar-based computer vision research.

2. Related Work

2.1. Radar Object Detection

One approach to integrate DNNs into radar object detection involves using radar detection points, known as the radar point cloud, as input for the DNN [9,12,15,19,34,36]. These points are identified using the Constant False Alarm Rate (CFAR) algorithm [32], which detects peaks in the radar image surpassing a local noise threshold. However, CFAR introduces information loss during subsequent DNN processing.

To address the information loss caused by CFAR, an alternative approach directly applies DNNs to the radar image for object detection, bypassing CFAR. Kim et al. [18] utilized YOLO [31] on radar images, demonstrating better performance than the conventional CFAR-based method. Xu et al. [10] adapted a ResNet-18 encoder with a CNN decoder to output 3D bounding box center, size, and orientation. RADDet [42] employed residual blocks and YOLO detection heads [7]. Zhang et al. [43] used a CNN version of U-Net for radar object detection. Meyer et al. [25] proposed graph convolution networks. Additional enhancements have been made by integrating temporal processing [39, 40] and by fusing radar and camera images [5, 14, 16, 17, 20, 24, 26, 27].

2.2. Radar Datasets

Numerous publicly available automotive radar datasets with object detection annotations exist, which can be categorized to two groups based on the type of radar data they offer. The first group includes datasets with radar point cloud data (the CFAR detections), such as nuScenes [8], Radar Scenes [37], Pointillism [3], Cooperative Radars [35], aiMotive [22], and Astyx [23].

The second group consists of datasets providing radar reflection intensity images before CFAR processing. The RADIATE dataset [38] captures unique $360°$ scanning radar reflection images, distinguishing it from conventional automotive radars with antenna arrays. The CRUW dataset [41] offers radar images with a limited range of up to $25m$ and the annotations include object center points without bounding box information. RADIal [30] and K-Radar [29] contain higher-resolution radar reflection intensity images, but low level radar hardware details are undisclosed. As a result, simulating the radar and generating synthetic data for these datasets is unattainable.

The CARRADA dataset [28] provides radar reflection images from 30 controlled scenarios in an open environments with few objects per scene using a Texas Instruments (TI) automotive radar prototype [1, 2]. RADDet [42] offers TI radar-derived reflection intensity images across 15 diverse automotive scenarios. Both datasets have up to $50m$ range and 2D bounding box annotations. An important distinction lies in the availability of hardware specifications for the TI prototype radar used in these datasets. This availability enables radar simulation and the generation of synthetic radar data, offering a notable advantage.

3. Object Detection With Reflection Boosting

A block diagram of the proposed object detection system is depicted in Fig. 3. The input radar image with range and angle dimensions undergoes processing by the reflection boosting DNN, which enhances resolution and reduces clutter by narrowing the main-lobe and damping side-lobes
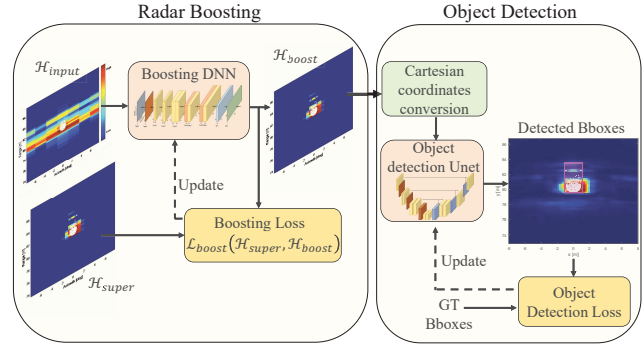

Figure 3. BoostRad Overview: The boosting DNN enhances input reflection image $\mathcal{H}_{input}$ to a radar image $\mathcal{H}_{boost}$ with a narrower PSF, resembling the 'super-radar' image $\mathcal{H}_{super}$. Subsequently, object detection DNN identifies objects in the boosted image.

of the PSF across all reflection points in the scene. Subsequently, another DNN detects objects within the boosting DNN output.

The boosting DNN is trained with a ground truth reference image from a high-resolution radar referred to as the 'super-radar'. This radar captures images from the same scene as the automotive radar but with a PSF featuring a narrower main-lobe and lower side-lobes. However, a notable challenge lies in the unavailability of a 'super-radar' hardware. Moreover, the selection of 'super-radar' resolution enhancement factor and a suitable loss function play a crucial role in successfully training the boosting DNN to closely align automotive radar images with the 'super-radar' images. These challenges are addressed comprehensively in Sections 3.1 and 3.3 that follow.

3.1. Reflection Boosting Network

Architecture. The reflection boosting network is depicted in Fig. 4. The input to the network is the reflection intensity image with a uniform grid in the dimensions range and sinus of the azimuth angle. In this coordinate system, the radar's PSF is spatially invariant and separable in range and angle dimensions. These properties are essential for narrowing the PSF. While our method primarily addresses radars without elevation data, its applicability can extend to radars featuring azimuth and elevation by increasing the input tensor's dimensions. The input image comprises three channels: the real and imaginary parts of the highest intensity along the Doppler domain for each range-angle pixel, and the Doppler frequency corresponding to the most intense Doppler bin. In this paper, we refer to the latter channel as the Doppler map.

The boosting network is a 2D convolutional neural network tailored to preserve the radar image's range dimension while expanding the azimuth angular dimension. This approach retains the fine radar range resolution (usually below

$25cm$) while increasing the initially lower angular resolution. This enhancement in angular resolution is achieved through successive layers of transpose convolutions in the angular domain. The network's output head consists of two layers of 1×1 convolution filters followed by the Sigmoid function.

Ground Truth Reference and Loss Function. The boosting network's training involves generating the probability of reflection for each output range-azimuth pixel, with an azimuth resolution κ times higher than the input. The ground truth reference is a reflection image constructed from a 'super-radar' that corresponds to the input image scenario. The 'super-radar' image angular resolution is κ times greater than the input radar image. As a result, the PSF in the 'super-radar' image is narrower by factor κ and has lower side-lobes than the input radar image. However, obtaining a practical 'super-radar' is unfeasible due to the complex and costly requirements, including a larger aperture with numerous antennas. To circumvent this limitation, we've developed a radar simulation that generates synthetic pairs of automotive and 'super-radar' images originating from the same scene. These synthetic pairs are subsequently employed for training the boosting DNN. A comprehensive explanation of this simulation process is provided in Section 3.3.

The proposed ground truth reference holds two significant advantages. Firstly, it shares the same sensor physical modality as the radar that produced the input image. In both images, the positions and intensity of reflection points remain consistent; the sole distinction lies in the spreading functions. Notably, this differs from using a LIDAR as a reference, where the reflection points of the LIDAR deviate considerably from the radar's. Secondly, this reference allows for the careful selection of achievable angular resolution, serving as a realistic goal for the network. While the ideal scenario would entail an extremely high κ, akin to LIDAR output resolution, physical constraints limit achievable output resolution, and complete elimination of the radar spreading function remains unattainable. As shown in the ablation study in Section 4.3, selecting an excessively high κ results in a decline in performance.

For calculating the loss function, we transform the 'super-radar' reflection intensity for each range and angle pixel into reflection probability. Let the intensity at the nth pixel of the 'super-radar' image be denoted as z_n. We make the assumption that z_n follows a chi-square distribution [13] with variance σ_s^2 when the pixel contains a reflection point (where σ_s^2 varies with range), and has a noise variance of σ_n^2 when the pixel lacks a reflection point. Under this assumption, we deduce that the ground truth reference probability is formulated as

$$p_n = \frac{e^{-|z_n|^2/(2\sigma_s^2)}}{e^{-|z_n|^2/(2\sigma_s^2)} + \frac{\sigma_s^2}{\sigma_n^2}e^{-|z_n|^2/(2\sigma_n^2)}}. \quad (1)$$

For a more comprehensive explanation of the derivation of (1), please refer to the supplementary material.

The reflection boosting DNN is trained with a weighted cross-entropy loss between the pixel-wise probabilities of the boosting DNN output image and the ground truth reference image. The loss function is given by

$$\mathcal{L}_{boost} = \rho_n \sum_{n \in \Omega_n} \mathcal{L}_{bce}(p_n, \hat{p}_n) +$$
$$\rho_r \sum_{n \in \Omega_r} \mathcal{L}_{bce}(p_n, \hat{p}_n) + \rho_s \sum_{n \in \Omega_s} \mathcal{L}_{bce}(p_n, \hat{p}_n), \quad (2)$$

where $\mathcal{L}_{bce}(p_n, \hat{p}_n)$ is the cross entropy loss between p_n and $\hat{p}_n$. The notations in (2) are as follows. The symbol n is the boosting DNN output pixel index, $\hat{p}_n$ is the DNN estimated probability of a reflection in the nth pixel, and p_n is n^{th} pixel reflection probability from the ground truth reference image given in (1). The symbols Ω_n, Ω_r, and Ω_s denote three different sets of pixels, which will be explained below, and ρ_n, ρ_r, ρ_s correspond to weight factors associated with each respective set.

The loss function in (2) comprises three loss function terms corresponding to three sets of pixels from the input radar image denoted by Ω_n, Ω_r and Ω_s. The set Ω_r is the image's pixels containing reflection points. The set Ω_s are pixels belonging to a reflection point's spreading function but not the reflection point's pixel itself. These pixels can be distinguished as pixels not included in Ω_r that have intensity above a noise level threshold. For example, we set the threshold 8 dB above the noise standard deviation (calculated empirically). The set Ω_n comprises pixels containing noise; these are the remaining pixels that do not belong to the sets Ω_r and Ω_s. We refer the reader to Fig. 1(b) for an illustration of the pixel sets. The set Ω_r is colored white in the figure, the set Ω_s includes the red (main-lobe) and orange (side-lobes) pixels, and the set Ω_n is colored blue.

The partitioning into these three sets facilitates the assignment of distinct weight factors (ρ_r, ρ_s, ρ_n) to the corresponding loss terms for each set. Notably, this partitioning empowers the enhancement of the weight factor associated with the pixels within the spreading function (ρ_s), thereby compelling the boosting DNN to effectively narrow the spreading function and reduce its side-lobes. The effectiveness of this enhancement will be illustrated in the ablation study detailed in Section 4.3.

3.2. Object Detection Network

The object detection DNN features four input channels. The first channel comprises the output of the reflection boosting DNN, converted to Cartesian coordinates. This transformation is important for object detection since object shape and size remain consistent in these coordinates. The second channel contains the Cartesian coordinates transformation of the Doppler map, which also serves
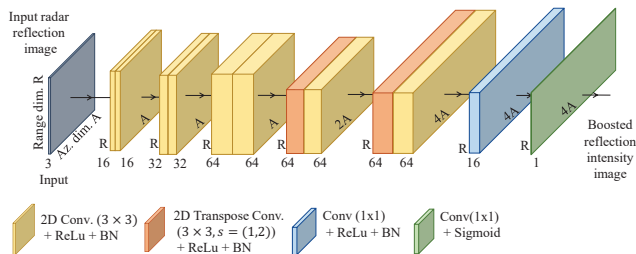
Figure 4. Diagram depicting reflection boosting DNN architecture with 4x expansion in output angular dimension compared to input.

as an input channel for the boosting DNN. Each pixel in the Doppler map represents the Doppler frequency of the highest-intensity Doppler bin within the Doppler domain of each range-angle pixel in the original reflection intensity image. The remaining two channels consist of the Cartesian coordinates (x, y) for each cell. While object shapes remain invariant in Cartesian coordinates, their surface reflection intensity varies based on their position relative to the radar. Radar reflection intensity from surfaces depends on their angle and distance from the radar. This information relies on the object's position in relation to the radar. By providing Cartesian input coordinate values for each pixel, the network learns these relationships and applies them to object detection.

The object detection DNN is a standard U-Net architecture [33] that has an encoder-decoder network with skip connections. The encoder has ten convolution layers that gradually down-sample the spatial features by a factor of 16. The decoder has eight convolution layers that gradually up-sample the spatial features back to the original input dimension. Batch normalization and ReLU activation functions are applied after each layer. The final layer is a 1×1 convolution layer, which outputs the probability of an object class (class score) and the 2D bounding box parameters of the vehicle in a bird's-eye view for each spatial cell. The bounding box specifications encompass the offset of the object center point from the cell center, the object's width and height, and the orientation represented by $cos(\theta), sin(\theta)$, with θ being the object's orientation angle. When the dataset lacks the bounding box's orientation angle, θ remains fixed at zero. During inference, non-maximal suppression is applied to filter out overlapping bounding boxes. The object detection network is trained using standard cross-entropy loss for classification and L2 regression loss for bounding box parameters.

3.3. Synthetic Data Generation

Training the boosting DNN in Section 3.1 requires a unique ground truth reference reflection image of the 'super-radar'. However, the 'super-radar' hardware is not a commodity available and may not be practical for imple-

mentation due to the large aperture, high number of antennas, and system complexity. To overcome this limitation, we developed a high-fidelity radar simulation that generates synthetic radar reflection intensity images of realistic scenes for any desired radar specification.

Fig. 5 presents the simulation procedure for generating radar images, which is elaborated below. In the first part, the CARLA simulation [11] generates a realistic driving scenario with roads, bridges, pavements, buildings, signs, poles, persons, vehicles, etc. From this CARLA simulation, a dense 3D point cloud is generated, capturing positions from all visible surfaces in the scene that can reflect radar signals. In the subsequent phase, each of these points is assigned a reflection intensity value using standard radar signal propagation and reflectivity formulas [21]. The reflection point intensity is a function of three factors: (a) the surface material, (b) the angle between the normal vector of the point's surface and the direction from the point to the radar, and (c) the distance between the point and the radar.

In the subsequent phase, the radar-received signal is synthesized using the reflection points and specific radar parameters, such as the antenna array layout and the transmitted waveform. The received signal is a sum of the received signals from all the reflection points in the scene with additional Gaussian noise. The received signal of each reflection point is generated by the point's reflection intensity, the distance between the point to the transmit and the receive antennas, and the radar's transmit waveform [6, 13]. In the final step of the simulation, standard radar signal processing techniques are applied to the received signals collected from all antennas, resulting in the production of the radar reflection intensity image. For more details on the mathematical formulas utilized in the simulation's implementation, please refer to the supplementary material.

To enable boosting DNN inference on an actual physical radar, it's essential to train the DNN using synthetic simulation data that closely aligns with the physical radar characteristics. This simulation requires knowledge of the antenna array layout and the transmit waveform employed by the physical radar. For the ground truth reference, we simulate a higher resolution radar with narrow PSF, featuring an increased number of transmit and receive antennas spanning a broader aperture than the physical radar being tested.

4. Experiments and Results

To assess the performance of BoostRad, we utilize the RADDet [42] and CARRADA [28] datasets. Both of these datasets comprise radar reflection images alongside 2D bounding box annotations for objects. The radar images were captured using a TI prototype automotive radar [1, 2], that has $50m$ range capability. The hardware specifications of the TI radar are publicly accessible, enabling us
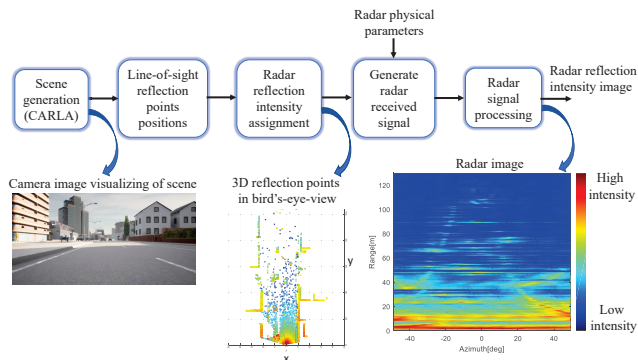
Figure 5. Radar simulation pipeline with illustrated simulated scenario: camera image, reflection points, and radar reflection image.

to simulate it and generate synthetic training data for the reflection boosting DNN (as detailed in Section 3.3). These datasets provide comprehensive testing coverage for automotive radar scenarios. RADDet encompasses images from 15 different densely populated automotive scenarios under favorable weather conditions. On the other hand, CARRADA contains 30 staged scenarios with fewer objects per scenario, yet these are captured in varying conditions, including fair and more challenging weather conditions such as snow.

In the RADDet dataset utilized in [42], the test and training images were sourced from the same scenarios, often with small time gaps between them. This split creates a strong correlation between test and training examples, leading to overfitting on the test set. To address this, we adjusted the RADDet data split for our performance evaluations, ensuring different scenarios in the train and test sets. Unlike RADDet, the CARRADA dataset didn't display this issue, so we retained the original data partitioning as proposed in [28]. For dataset partitioning details, including ablation tests with the original RADDet split, see the supplementary material.

Implementation details. The boosting and object detection models were trained on single NVIDIA Tesla V100 GPU, for 100 epochs with batch size 32, and ADAM optimizer with learning rate of 10^{-4}. More implementation details can be found in the supplementary material.

4.1. Object Detection Performance Evaluation

The boosting DNN's training set was generated by simulating the TI radar prototype used in RADDet and CARRADA datasets (see Section 3.3). It included 15,000 synthetic radar images and their corresponding 'super-radar' reference images, with a 12 times higher angular resolution ($\kappa = 12$) than the TI radar prototype. After training the boosting DNN, its weights were frozen for subsequent object detection training. Radar images from RADDet or CARRADA were passed through the boosting DNN,

and its output replaced the original radar image for the object detection DNN. It's important to highlight that the same boosting DNN was utilized for both datasets. Conversely, the object detection DNN was optimized independently for RADDet and CARRADA, and their respective performances were evaluated separately.

We assess BoostRad's object detection performance against three reference methods: 'U-Net', 'RADDet', and 'Probabilistic'. The first method employs the same U-Net as BoostRad's, but with the original radar reflection image as input instead of the boosting DNN output image. 'RADDet' refers to the object detection network introduced in the RADDet paper [42], while 'Probabilistic' refers to the network from [10]. To examine the boosting DNN's generalization across different architectures, we also evaluated 'RADDet' and 'Probabilistic' using the boosting DNN output as input instead of the original radar reflection image

Table 1 summarizes the average precision (AP) results for the tested methods on RADDet and CARRADA at IOU thresholds 0.1 and 0.3. Supplementary material contains further results for larger IOU values. AP represents the area under the precision-recall curve and is provided for the 'car' and 'person' classes. In the car class, we also include trucks, which are larger vehicles such as vans. Cars are easier to detect than persons since their reflection intensity is significantly higher. To assess the performance on more challenging cases within the 'car' class, we also present the AP for cars at range (denoted by R) exceeding $40m$. The 'Boosting' in the second column specifies whether the input to the object detection network was the boosting DNN output (✓) or the original reflection image (✗). Green values indicate the AP difference between using boosting DNN output and original reflection image for the same detection method.

We first examine in Table 1 the results of BoostRad compared to the reference methods 'U-Net', 'Probalistic', and 'RADDet' when the reference methods use the original reflection intensity image and not the boosting DNN output. This comparison reveals that BoostRad consistently achieves higher AP values than the reference methods, across all IOU values, object classes, detection distances, and datasets. For class car, the performance advantage of BoostRad over the best reference method, 'U-Net', is relatively small when calculating the AP for all ranges. The reason is that most cars are at relatively close distances and could be relatively easily detected by all methods. However, in more challenging scenarios such as 'person' detection at all distances and 'car' detection beyond $40m$ with an IOU of 0.3, BoostRad demonstrates a notable performance gain over the reference methods. 'U-Net' stands out as the best-performing reference method, potentially due to its supplementary channels containing (x, y) Cartesian coordinates, which contributes to object detection. Notably, the 'RADDet' method exhibits lower AP than the values re-

ported in [42] due to differences in the train-test set split, preventing overfitting (details in the supplementary material).

Next, we evaluate the performance gain using the boosting DNN output instead of the original reflection intensity image as input for the 'RADDet' and 'Probalistic' reference methods. The results in Table 1 show that the AP of both reference methods improves when using the boosting DNN output instead of the original reflection intensify image. The performance gain increases as the detection cases are harder, i.e., persons at all distances and cars at a distance greater than $40m$.

4.2. Qualitative performance evaluation

Fig. 6 provides a qualitative comparison between Boost-Rad and the 'U-Net' reference method, using two examples from the RADDet test set. Additional qualitative examples can be found in the supplementary material. Each row within the figure corresponds to a distinct scenario. For each scenario, we present the original radar reflection image (a), the image after passing through the boosting DNN (b), and the camera image of the scene (c). Detected bounding boxes from the 'U-Net' reference method are plotted in purple in (a), while those from BoostRad are plotted in purple in (b). Ground truth bounding boxes are indicated in white within (a) and (b), and in red in (c). Car and person bounding boxes are labeled with text and identification numbers. Other objects in the environment, such as poles and structures, are marked on the figure to aid orientation, even though they aren't part of the dataset and thus aren't classified.

From Fig. 6, it's evident that the original radar reflection intensity image (a) appears blurred and cluttered, attributed to the radar's wide PSF. In contrast, the image at the boosting DNN's output (b) is sharper, cleaner, and displays a considerably narrower PSF. In the first scenario, Boost-Rad detects all objects, whereas 'U-Net' misses 'car 4' and the 'person'. This discrepancy might be due to differences between the boosting DNN image (b) and the reflection intensity image (a). Specifically, in scenario 1 (a), 'Car 4' is masked by the side-lobe of a pole at a similar range, and the person's reflection is relatively weak in comparison to nearby clutter arising from side-lobes of adjacent reflections. Moving to scenario 2, BoostRad successfully detects all objects, while 'U-Net' miss-detects 'car 3' and falsely detects a 'car' on the right side of 'car 2'. Such errors could stem from the higher side-lobes of the spreading function in (a) as opposed to (b).

4.3. Ablation Study

The effective training of the boosting DNN can be attributed to its distinct loss function and ground truth reference, as examined in the ablation study outlined below. The boosting DNN loss function expressed in (2) comprises two key components. The first component involves mapping the ground truth 'super-radar' reference intensity to probabilities using (1), facilitating pixel-wise cross-entropy computation between it and the boosting DNN's output. The second component involves segmenting the pixel-wise cross-entropy into three sets of pixels $(\Omega_r, \Omega_s, \Omega_n)$, each assigned distinct weights (ρ_r, ρ_s, ρ_n). This segmentation mechanism empowers the enforcement of PSF narrowing by amplifying the loss weight associated with the PSF's pixels.

Table 2 presents a comparison of object detection average precision for IOU 0.3 on the RADDet dataset, evaluating the impact of training the boosting DNN with alternative loss functions from those employed in BoostRad. The table examines L1 loss instead of cross entropy, as indicated in the first column. Additionally, it investigates the use of cross entropy with binary mapping of the ground truth 'super-radar' reference intensity image via an optimized noise level threshold, instead of the probability mapping in (1), as indicated in the second column. The table also compares results when partitioning cross-entropy pixels into sets $\Omega_r, \Omega_s, \Omega_n$ as in (2), separating side-lobe pixels from noise pixels, versus when partitioning only to signal and noise pixels, including side-lobe pixels in the noise category. The former case is indicated by ✓ and the later by ✗ in the third column of the table, and a hyphen indicates the case of no segmentation to different pixel sets. The final row showcases performance with all proposed loss components of BoostRad. This configuration significantly outperforms alternative loss functions, underlining the significance of each component within the boosting DNN's loss function for achieving maximal object detection accuracy.

Next, we investigate the importance of selecting an appropriate resolution enhancement factor κ for the 'super-radar' ground truth reference. Fig. 7 depicts the AP results for class 'car' at distance greater than $40m$ as a function of κ. The analysis reveals that $\kappa = 12$ achieves optimal performance. A lower κ fails to fully exploit the network's resolution enhancement capabilities. Conversely, selecting an excessively high κ value leads to performance degradation, possibly due to attempting to surpass the physical limitations of resolution. Thus, the choice of κ significantly affects performance, and its optimization was made possible through radar simulation, enabling the generation of ground truth references with various angular resolutions.

Additional analysis of the boosting DNN image enhancement can be found in the supplementary material.

5. Conclusion

We present a novel method to enhance radar image quality using a DNN that narrows the radar PSF, resulting in improved object detection. Our approach integrates domain knowledge, utilizing a high-resolution radar image with op-

Table 1. Object Detection Average Precision on RADDet and CARRADA Datasets

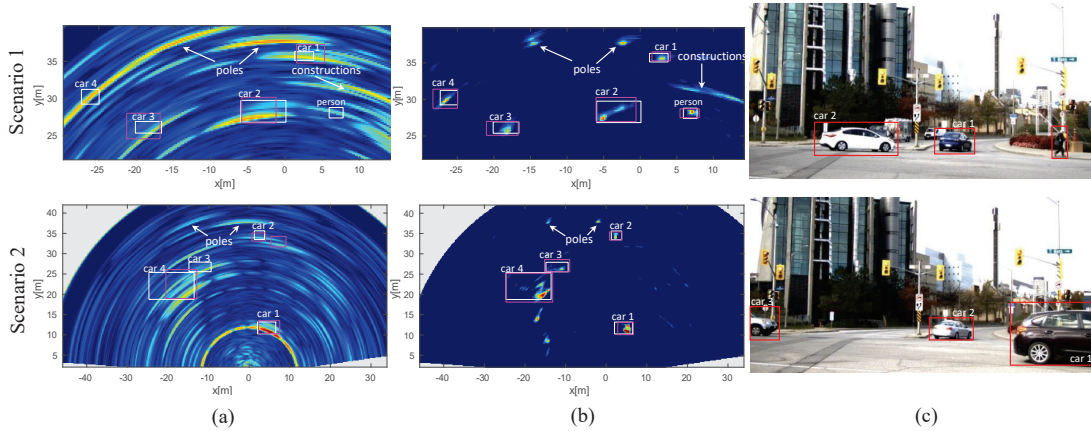|		RADDet Dataset						CARRADA Dataset					
		Car				Person		Car				Person	
		$R \geq 0$		$R \geq 40$		$R \geq 0$		$R \geq 0$		$R \geq 40$		$R \geq 0$	
Method	Boost.	@ 0.1	@ 0.3	@ 0.1	@ 0.3	@ 0.1	@ 0.3	@ 0.1	@ 0.3	@ 0.1	@ 0.3	@ 0.1	@ 0.3
RADDet	✗	83.69	72.96	42.45	21.27	29.76	15.38	87.59	70.90	76.46	68.74	29.10	10.24
	✓	85.91	78.19	51.65	28.25	31.50	18.90	88.48	73.79	80.10	78.02	35.59	22.39
		+2.22	+5.23	+9.20	+6.98	+1.74	+3.52	+0.89	+2.89	+3.64	+9.28	+6.49	+12.15
Probalistic	✗	83.32	74.80	43.28	28.50	21.30	17.10	84.72	67.56	69.40	63.92	22.23	10.99
	✓	89.17	79.74	68.19	45.85	41.07	31.23	86.12	78.19	81.55	73.29	32.29	22.30
		+5.85	+4.94	+24.91	+17.35	+19.77	+14.13	+1.40	+10.63	+12.15	+9.37	+10.06	+11.31
U-Net	✗	88.90	81.36	71.90	47.90	25.50	19.60	87.86	81.05	75.56	64.35	25.50	23.74
BoostRad	✓	**90.00**	**83.55**	**79.50**	**62.10**	**41.20**	**33.00**	**88.86**	**81.43**	**82.23**	**75.87**	**40.45**	**35.46**
		+1.10	+2.19	+7.60	+14.20	+15.70	+13.40	+1.00	+0.38	+6.67	+11.52	+14.95	+11.72

Figure 6. Qualitative examples from RADDet dataset. 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. Detections in (a) and (b) are purple, while ground truth bounding boxes are white.

Table 2. Boosting Ablation Study (RADDet, AP @0.3)

	GT		Car		Person
Loss	Map.	Ω_s	$R \geq 0$	$R \geq 40$	$R \geq 0$
L1	✓	-	79.46	43.26	23.43
	✓	✓	78.20	43.98	26.55
CE	✓	✗	81.54	50.64	25.98
	✗	✓	68.14	19.43	5.18
	✓	✓	**83.55**	**62.10**	**33.00**

Figure 7. Effect of κ on object detection. Raddet cars at $R > 40$

timized resolution enhancement (κ), a unique intensity-to-probability mapping, and a tailored cross entropy loss that enforces the attenuation of the PSF side-lobes. An ablation study confirms the importance of these components. Addressing the lack of narrow PSF radar hardware, we develop a simulation to generate synthetic data for the boost-
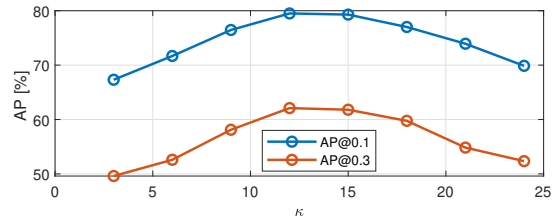
ing DNN training. In testing on real data from RADDet and CARRADA datasets, *BoostRad* outperforms reference methods in object detection.

Our work challenges prevailing *'end-to-end'* object detection trends, encouraging exploration of multistage approaches. The PSF narrowing technique holds potential beyond radar, benefiting sensors with wide PSFs. Successful synthetic radar image utilization prompts further investigation of synthetic data for similar challenges in computer vision tasks involving radar and other sensors.

References

[1] AWR1843 data sheet, product information and support, https://www.ti.com/product/awr1843. 3, 5

[2] TI mmwave-sdk software development kit (SDK), https://www.ti.com/tool/mmwave-sdk. 3, 5

[3] Kshitiz Bansal, Keshav Rungta, Siyuan Zhu, and Dinesh Bharadia. Pointillism: Accurate 3d bounding box estimation with multi-radars. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 340–353, 2020. 3

[4] Oded Bialer, Amnon Jonas, and Tom Tirer. Super resolution wide aperture automotive radar. *IEEE Sensors Journal*, 21(16):17846–17858, 2021. 2

[5] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11682–11692, 2020. 3

[6] Igal Bilik, Oded Bialer, Shahar Villeval, Hasan Sharifi, Keerti Kona, Marcus Pan, Dave Persechini, Marcel Musni, and Kevin Geary. Automotive mimo radar for urban environments. In *2016 IEEE Radar Conference (RadarConf)*, pages 1–6. IEEE, 2016. 1, 5

[7] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020. 2, 3

[8] 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*, pages 11621–11631, 2020. 3

[9] Andreas Danzer, Thomas Griebel, Martin Bach, and Klaus Dietmayer. 2d car detection in radar data with pointnets. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 61–66. IEEE, 2019. 2

[10] 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, 3, 6

[11] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017. 5

[12] Maria Dreher, Emeç Erçelik, Timo Bänziger, and Alois Knol. Radar-based 2d car detection using deep neural networks. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–8. IEEE, 2020. 2

[13] Jerry Eaves and Edward Reedy. *Principles of modern radar*. Springer Science & Business Media, 2012. 4, 5

[14] Di Feng, Christian Haase-Schütz, Lars Rosenbaum, Heinz Hertlein, Claudius Glaeser, Fabian Timm, Werner Wiesbeck, and Klaus Dietmayer. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1341–1360, 2020. 3

[15] Zhaofei Feng, Shuo Zhang, Martin Kunert, and Werner Wiesbeck. Point cloud segmentation with a high-resolution automotive radar. In *AmE 2019-Automotive meets Electronics; 10th GMM-Symposium*, pages 1–5. VDE, 2019. 2

[16] Jyh-Jing Hwang, Henrik Kretzschmar, Joshua Manela, Sean Rafferty, Nicholas Armstrong-Crews, Tiffany Chen, and Dragomir Anguelov. Cramnet: Camera-radar fusion with ray-constrained cross-attention for robust 3d object detection. In *European Conference on Computer Vision*, pages 388–405. Springer, 2022. 3

[17] Jinhyeong Kim, Youngseok Kim, and Dongsuk Kum. Low-level sensor fusion network for 3d vehicle detection using radar range-azimuth heatmap and monocular image. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 3

[18] Woosuk Kim, Hyunwoong Cho, Jongseok Kim, Byungkwan Kim, and Seongwook Lee. Yolo-based simultaneous target detection and classification in automotive fmcw radar systems. *Sensors*, 20(10):2897, 2020. 2, 3

[19] Florian Kraus, Nicolas Scheiner, Werner Ritter, and Klaus Dietmayer. Using machine learning to detect ghost images in automotive radar. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–7. IEEE, 2020. 2

[20] Teck-Yian Lim, Amin Ansari, Bence Major, Daniel Fontijne, Michael Hamilton, Radhika Gowaikar, and Sundar Subramanian. Radar and camera early fusion for vehicle detection in advanced driver assistance systems. In *Machine Learning for Autonomous Driving Workshop at the 33rd Conference on Neural Information Processing Systems*, volume 2, page 7, 2019. 3

[21] Bassem R Mahafza and Atef Elsherbeni. *MATLAB simulations for radar systems design*. Chapman and Hall/CRC, 2003. 5

[22] Tamás Matuszka, Iván Barton, Ádám Butykai, Péter Hajas, Dávid Kiss, Domonkos Kovács, Sándor Kunsági-Máté, Péter Lengyel, Gábor Németh, Levente Pető, Dezső Ribli, Dávid Szeghy, Szabolcs Vajna, and Bálint Varga. aimotive dataset: A multimodal dataset for robust autonomous driving with long-range perception. 2022. 3

[23] Michael Meyer and Georg Kuschk. Automotive radar dataset for deep learning based 3d object detection. In *2019 16th european radar conference (EuRAD)*, pages 129–132. IEEE, 2019. 3

[24] Michael Meyer and Georg Kuschk. Deep learning based 3d object detection for automotive radar and camera. In *2019 16th European Radar Conference (EuRAD)*, pages 133–136. IEEE, 2019. 3

[25] Michael Meyer, Georg Kuschk, and Sven Tomforde. Graph convolutional networks for 3d object detection on radar data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3060–3069, 2021. 2, 3

[26] Ramin Nabati and Hairong Qi. Radar-camera sensor fusion for joint object detection and distance estimation in au-

tonomous vehicles. *arXiv preprint arXiv:2009.08428*, 2020. 3

[27] Ramin Nabati and Hairong Qi. Centerfusion: Center-based radar and camera fusion for 3d object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1527–1536, 2021. 3

[28] Arthur Ouaknine, Alasdair Newson, Julien Rebut, Florence Tupin, and Patrick Perez. Carrada dataset: Camera and automotive radar with range-angle-doppler annotations. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 5068–5075. IEEE, 2021. 2, 3, 5, 6

[29] Dong-Hee Paek, Seung-Hyun Kong, and Kevin Tirta Wijaya. K-radar: 4d radar object detection dataset and benchmark for autonomous driving in various weather conditions. *arXiv preprint arXiv:2206.08171*, 2022. 3

[30] Julien Rebut, Arthur Ouaknine, Waqas Malik, and Patrick Pérez. Raw high-definition radar for multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17021–17030, 2022. 3

[31] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 2, 3

[32] Hermann Rohling. Radar cfar thresholding in clutter and multiple target situations. *IEEE transactions on aerospace and electronic systems*, (4):608–621, 1983. 2

[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 5

[34] Nicolas Scheiner, Nils Appenrodt, Jürgen Dickmann, and Bernhard Sick. Radar-based road user classification and novelty detection with recurrent neural network ensembles. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 722–729. IEEE, 2019. 2

[35] Johannes Schlichenmaier, Maximilian Steiner, Timo Grebner, and Christian Waldschmidt. Radar measurements of two vehicles with three cooperative imaging sensors, 2022. 3

[36] Ole Schumann, Markus Hahn, Jürgen Dickmann, and Christian Wöhler. Semantic segmentation on radar point clouds. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 2179–2186. IEEE, 2018. 2

[37] Ole Schumann, Markus Hahn, Nicolas Scheiner, Fabio Weishaupt, Julius F Tilly, Jürgen Dickmann, and Christian Wöhler. Radarscenes: A real-world radar point cloud data set for automotive applications. In *2021 IEEE 24th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2021. 3

[38] Marcel Sheeny, Emanuele De Pellegrin, Saptarshi Mukherjee, Alireza Ahrabian, Sen Wang, and Andrew Wallace. Radiate: A radar dataset for automotive perception in bad weather. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–7. IEEE, 2021. 3

[39] Ana Stroescu, Liam Daniel, and Marina Gashinova. Combined object detection and tracking on high resolution radar imagery for autonomous driving using deep neural networks and particle filters. In *2020 IEEE Radar Conference (Radar-Conf20)*, pages 1–6. IEEE, 2020. 2, 3

[40] Yizhou Wang, Zhongyu Jiang, Xiangyu Gao, Jenq-Neng Hwang, Guanbin Xing, and Hui Liu. Rodnet: Radar object detection using cross-modal supervision. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 504–513, 2021. 2, 3

[41] Yizhou Wang, Zhongyu Jiang, Yudong Li, Jenq-Neng Hwang, Guanbin Xing, and Hui Liu. Rodnet: A real-time radar object detection network cross-supervised by camera-radar fused object 3d localization. *IEEE Journal of Selected Topics in Signal Processing*, 15(4):954–967, 2021. 3

[42] 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, 5, 6, 7

[43] Guoqiang Zhang, Haopeng Li, and Fabian Wenger. Object detection and 3d estimation via an fmcw radar using a fully convolutional network. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4487–4491. IEEE, 2020. 2, 3