Deep learning for Object Detection using RADAR Data

Ahmed M. Reda[1,2]*, Naser El-Sheimy[1], Adel Moussa[1,3]

[1] Dept. of Geomatics Engineering, University of Calgary, Calgary, AB, Canada – (ahmed.hegazy@ucalgary.ca, elsheimy@ucalgary.ca, amelsaye@ucalgary.ca)
[2] Dept. of Geomatics Engineering, Benha University, Benha, Egypt – (ahmed.hegazy@ucalgary.ca)
[3] Dept. of Electrical and Computer Engineering, Port-Said University, Port-Said, Egypt – (amelsaye@ucalgary.ca)

KEY WORDS: Autonomous Driving, Deep learning, LiDAR, RADAR, Faster-RCNN, Resnet-50.

ABSTRACT:

Recently, Deep learning algorithms are becoming increasingly instrumental in autonomous driving by identifying and acknowledging road entities to ensure secure navigation and decision-making. Autonomous car datasets play a vital role in developing and evaluating perception systems. Nevertheless, the majority of current datasets are acquired using Light Detection and Ranging (LiDAR) and camera sensors. Utilizing deep neural networks yields remarkable outcomes in object recognition, especially when applied to analyze data from cameras and LiDAR sensors which perform poorly under adverse weather conditions such as rain, fog, and snow due to the sensor wavelengths. This paper aims to evaluate the ability to use RADAR dataset for detecting objects in adverse weather conditions, when LiDAR and Cameras may fail to be effective. This paper presents two experiments for object detection using Faster-RCNN architecture with Resnet-50 backbone and COCO evaluation metrics. Experiment 1 is object detection over only one class, while Experiment 2 is object detection over eight classes. The results show that as expected the average precision (AP) of detecting one class is (47.2) which is better than the results from detecting eight classes (27.4). Comparing my results from experiment 1 to the literature results which achieved an overall AP (45.77), my result was slightly better in accuracy than the literature mainly due to hyper-parameters optimization. The outcomes of object detection and recognition based on RADAR indicate the potential effectiveness of RADAR data in automotive applications particularly in adverse weather conditions, where vision and LiDAR may encounter limitations.

1. INTRODUCTION

Recently, the technology of autonomous driving has received much attention. The system of autonomous driving mainly consists of three sequential modules of perception, planning, and control. As the modules of planning and control rely on the output of the perception module, The perception module needs to be robust even under all driving weather conditions. Identifying objects is a fundamental task for autonomous vehicles. In the process of generating a virtual map of the environment, it is essential to recognize pivotal elements like vehicles, pedestrians, street fixtures, walls, traffic signs, junctions, and more. Therefore, focusing on these objects and forecasting their movements becomes imperative for developing a secure perception system for autonomous cars (Sheeny, 2020).

The pivotal question at hand is whether we are prepared to introduce fully autonomous vehicles. It is well-known that the majority of self-driving cars rely on cameras and Light Detection and Ranging (LiDAR) systems, which are not resilient in adverse weather conditions. A telling instance is the Competition of Automobile Technology in South Korea in 2014 (KAIST, 2014), where four teams, each with twelve autonomous cars, were tasked with completing various challenges in an urban setting. Despite the success of all four teams on the first day, which featured favorable weather conditions, the second day's rain, slippery roads, and wet conditions resulted in crashes for two of the initially successful teams. This competition underscored the ongoing lack of readiness to deploy fully autonomous cars on public roads during bad weather. Addressing the challenges posed by adverse weather conditions is essential in the development of autonomous vehicles (Sheeny, 2020).

Radio Detection and Ranging (RADAR) sensors have the ability to penetrate rain, snow, and fog (Skolnik, 1980). Establishing robust approaches based on RADAR will result in the creation of a more secure perception system, facilitating the realization of complete autonomy under diverse weather conditions. RADAR, while producing images with lower resolution in contrast to video and LiDAR, poses a challenging task in designing an object recognition system tailored for adverse weather (Sheeny, 2020).

The computer vision community and RADAR community have different meaning of detection term. In RADAR communities, detection is confined to identifying regions without assigning classes (ex. Constant False Alarm Rate (CFAR) (Skolnik, 1980) algorithm). Conversely, in computer vision communities, detection encompasses both the localization of regions (usually rectangular boxes) and their classification (ex. Faster R-CNN algorithm). There are three important terms which are:

- Classification: In the context of classification, the term involves categorizing the entire image without specifying the object's location (ex. AlexNet (Krizhevsky et al., 2017)). The main metric employed for classification is accuracy.
- Detection: Concerning detection, it encompasses the localization of potential regions followed by their classification (ex. Faster R-CNN (Ren et al., 2016)). The key metric for classification in this scenario is Average Precision (AP) (Everingham et al., 2010).
- Recognition: as a term, is applied in a broader context, encompassing instances where classification occurs within a detection framework or independently.

1.1 Problem Statement

The systems of autonomous cars usually used LiDAR and video sensors. Is it feasible to employ LiDAR and video in bad weather conditions? Figure 1 illustrates the application of the state-of-the-art object detection algorithm (Faster R-CNN (Ren et al., 2016)) trained on MS-COCO (Fleet et al., 2014) in various foggy road scenarios. The results depicted in Figure 1 indicate that the network successfully detects vehicles in only two images, while it struggles to identify pedestrians or vehicles in the remaining

images. In adverse weather conditions, a perception system relying on a video sensor might fail to detect crucial objects due to signal attenuation and the need for alternative sensing methods, on the other hand, the RADAR sensor is a key aspect to treat the bad weather problem. The problem addressed in this paper is object recognition for all weather scenarios (night, sunny, rain, fog, and snow) based on RADAR sensing. Despite the capability of RADAR sensors to penetrate fog, rain, and snow, they offer relatively limited spatial resolution (especially in cross-range) (Sheeny, 2020).
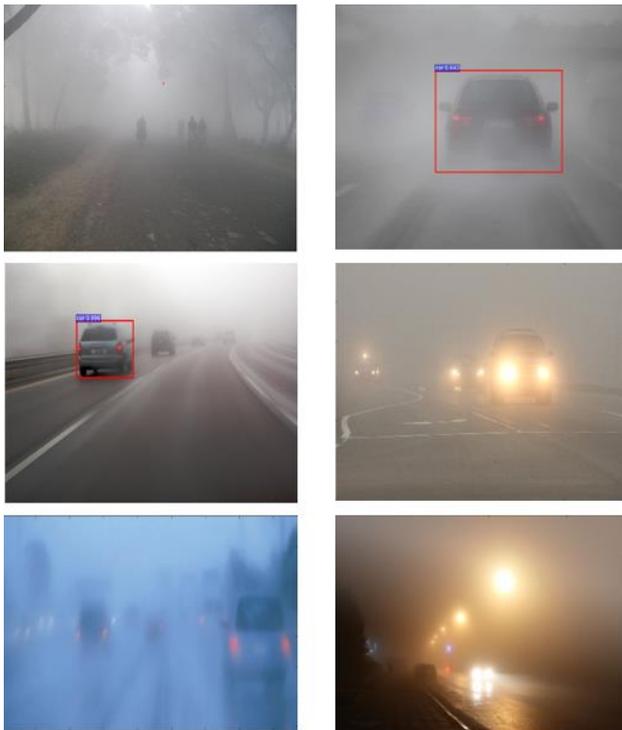

Figure 1. Faster R-CNN algorithm was utilized on numerous images during inclement weather. The training for this network was conducted using MS-COCO (Fleet et al., 2014). This figure shows that camera sensor fails to recognise objects under adverse weather conditions.

1.2 Objective

The primary goal of this paper is to detect and recognize objects for automotive applications by using resilient RADAR images across diverse weather conditions.

1.3 OBJECT DETECTION CHALLENGES

Deep learning and Computer vision methods can be used for solving the challenges of object detection. These challenges are:

1.3.1 Object classification in digital image: Object classification involves determining the category of an object within an image. Traditional machine learning algorithms are outperformed by conventional neural networks (CNNs) in achieving more accurate image classification, in achieving more accurate image classification. For multi-class object classification, the CNN-RNN (Recurrent Neural Network) model, as proposed by (Wang et al., 2013), yields improved results.

1.3.2 Object localization in digital image: Object localization issue entails determining the precise location of an object within a digital image. It involves identifying the existence of an object by precisely determining its position in the digital image, typically represented using bounding boxes. These bounding boxes are defined by the coordinates of the object within the image. Object localization combines both classification and pinpointing the object's location. According to (Bazzani et al., 2016), an effective object localization model should have the capability of predicting the object class along with its associated bounding box.

1.3.3 Object detection in image: Within the field of Computer vision, the object detection challenge focuses on the identification and recognition of multiple objects within a single image. Objects in digital images may be associated with specific classes. Various machine learning algorithms, such as support vector machines (SVMs), CNN-based deep learning models, and naïve Bayes, are employed for object detection and identification (classification). Object detection involves outlining the detected object with a bounding box, while object recognition involves labeling the object with a tag. The accuracy of the bounding boxes predicted by deep learning models can be assessed using the Intersection over Union (IoU) technique. This evaluation method measures how precisely the model delineates a box around the object, with IoU scores ranging from 0 to 1. The score is computed by dividing the common area of the two boxes by the area of their union (Bazzani et al., 2016). A higher IoU score indicates a more accurate bounding box, typically a score exceeding 0.5 is considered indicative of a more precise prediction.

2. LITERATURE REVIEW

Presently, various methods have been employed to facilitate the successful detection of vehicles by Autonomous Vehicles (AV) under challenging lighting conditions (night scene) and under adverse weather conditions (sunny, rainy, and snowy) (Kelly et al., 2006). LiDAR (Ramasamy et al., 2016) detects vehicles during bad weather however, it encounters difficulties in accurately interpreting road lane markers crucial for vehicle detection. Additionally, LiDAR proves ineffective during heavy rain or when low-hanging clouds affect it due to refraction effects (Ramasamy et al., 2016) (Wang et al., 2013). Moreover, LiDAR requires supplementary cameras to identify obstacles. Consequently, the deployment of the LiDAR and camera sensor combination may be challenging in various situations, and it is highly susceptible to sensor failures. Moreover, rain and fog can disrupt the laser light emitted by LiDAR sensors (Sucgang et al., 2017) (Wang et al., 2013). Another approach is using RADAR for vehicle detections, where radio waves are employed to identify the presence of objects in the atmosphere (Reina et al., 2015). In the AV sector, RADAR is employed to determine the distance, angle, and speed of vehicles. Additionally, it is utilized for detecting precipitation as well as other meteorological events, issuing warnings to vehicles about potential impacts and enabling drivers to apply brakes. Nevertheless, RADAR usually has constrained coverage, extending up to a distance of 200 meters (Appiah and Bandaru, 2011).

Over the past two decades, research in the field of autonomous cars has experienced rapid growth. The DARPA Grand Challenge, established with the aim of supporting significant technological advancements with military applications, has been instrumental in this growth. The challenges conducted in 2004, 2005, and 2007 specifically concentrated on the development of autonomous cars capable of covering extensive distances in both

urban and off-road environments. Numerous universities participated in these competitions, contributing to the progress of cutting-edge tasks like localization, mapping, and obstacle detection. (Thrun et al., 2006).

(Pfennigbauer et al., 2014) outlines the capabilities of LiDAR in scenarios with fog. The study involved an experiment conducted in a chamber filled with fog of varying visibilities, while an object was positioned at a distance of 100 meters. The results revealed that the full waveform of the LiDAR generated a pronounced unwanted return signal, representing the presence of fog. Despite this, it also successfully captured signals from objects in conditions with a measured visibility of 40 meters. LiDAR demonstrates potential for sensing in moderately foggy weather. Through appropriate full waveform processing, it is possible to mitigate the effects of fog on the image. However, in scenarios with dense fog, the detection of objects at long distances becomes challenging.

(Premebida et al., 2007) noted that video and LiDAR serve as the primary sensors in the ongoing development of autonomous vehicles. Video offers high-color resolution, while LiDAR provides accurate 3D point cloud estimation. When two cameras are employed, video can also supply 3D information. Both sensors find application in tasks such as vehicle detection, pedestrian recognition (Ren et al., 2016), (Bartsch et al., 2012), and 3D mapping (Krishnan and Kollipara, 2014), (Cadena et al., 2016). Leveraging its ability to capture color information, video is additionally employed for tasks like traffic sign recognition (Wu et al., 2013) and lane detection (Ghafoorian et al., 2019). Often, to achieve a more comprehensive scene representation, a fused representation of both sensors is employed. (Premebida et al., 2007).

(Radecki et al., 2016) conveyed that both video and LiDAR sensors exhibit suboptimal performance in bad weather conditions due to their reliance on the optical electromagnetic spectrum, which is impeded by rain, fog, and snow. In contrast, RADAR sensors utilize radio waves that can penetrate through such weather conditions. However, the trade-off is that RADAR provides lower resolution compared to LiDAR and video. To harness the strengths of both sensor types, sensor fusion approaches have incorporated RADAR sensors.

(Grimes and Jones, 1974) provide a comprehensive overview of autonomous RADAR, examining the challenges and potential advancements in RADAR technology. The paper specifically delves into applications related to speed sensing, predictive crash sensing, and obstacle detection. It also explores the impact of various weather conditions on RADAR systems, demonstrating how attenuation varies with different types of weather. Ultimately, the authors highlight the promising prospect of achieving target recognition through RADAR technology.

(Rasshofer, 2007) examines the functional requirements for RADAR in automotive contexts, addressing aspects such as high performance, cost, and system architecture. The paper initially explores the feasibility of RADAR imaging using current technology, highlighting the challenge of poor angle resolution. The authors emphasize that enhancing this aspect could greatly improve the potential of high-resolution RADARs. Additionally, the paper explores the use of low-THz bands in automotive scenarios, deeming it promising. However, the potential impact of rain, fog, and snow must be meticulously considered during the development of such sensors.

(Bartsch et al., 2012) employed a 24 GHz RADAR to classify pedestrians based on the object area's shape and Doppler spectrum features. The classification process involved analyzing the probability of each feature, and a straightforward decision model based on these features was utilized. The outcomes revealed that under optimal scenarios, they attained a classification rate of 95%, However, in situations where pedestrians appeared in gaps between cars due to the low resolution of the RADAR sensors, the classification rates dropped significantly to 29.4%.

(Nordenmark and Forsgren, 2015) employed four short-range mono-pulse Doppler RADARs operating at 77 GHz to categorize four classes (car, truck, pedestrian, and bicycle). Each image in their dataset contained one target. Detection was performed using the Density-based spatial clustering of applications with noise algorithm (DBSCAN). Feature extraction relied on parameters such as minimum object length, object area, number of detections, density, mean Doppler velocity, variance Doppler velocity, amplitude per distance, and variance of amplitude. The classification experiment employed Support Vector Machines (SVM). The outcomes demonstrated an impressive average accuracy of 95%.

The current limitations in the range and azimuth capabilities of RADAR technology present a significant challenge in achieving reliable target recognition rates for autonomous vehicles across various scenarios. Currently, sensor fusion with LiDAR and video is employed to attain dependable recognition. However, in adverse weather conditions, the performance of LiDAR and video degrades. The advancement of high-resolution RADAR sensors is expected to enhance the development of a more reliable recognition system suitable for use in all weather conditions. In the paper by (Roos et al., 2019), the authors outline the current challenges in developing RADAR systems for autonomous vehicles. They illustrate the evolution of RADAR from a detection-only system to a high-resolution perception sensor, highlighting the ongoing progress in achieving high resolution in current RADAR systems.

Cutting-edge algorithms for object recognition and detection in LiDAR-based automotive scenarios rely on deep neural networks. Research papers such as (Engelcke et al., 2017)(Li et al., 2016) have introduced convolutional neural networks specifically designed for detecting and recognizing cars. Another notable method is SA-SSD3D developed by He et al (He et al., 2020), which adapts the Single Shot Multibox Detector (SSD) to a 3D point cloud. This approach incorporates an auxiliary network to convert the voxel representation into a 3D point cloud, enhancing the accuracy of predicted bounding box locations.

3. DATA SET (RADIATE)

In many existing datasets offering RADAR data for automotive applications, the RADAR is typically employed solely as a basic detector., e.g., the NuScenes dataset (Caesar et al., 2020) gives sparse 2D point clouds (does not contain odometry data, and data during Fog and Snow). Recently, both Oxford Robotcar (does not contain data during Fog and Snow and object detection, object tracking) and MulRan datasets (excluding data from nighttime, fog, rain, and snow, as well as object detection and tracking) furnish information obtained through a scanning Navtech RADAR across diverse weather conditions. Nonetheless, these datasets do not include object annotations, as their primary purpose is geared towards Simultaneous Localization and Mapping (SLAM) and place recognition for long-term autonomy. The Astyx dataset (acks data during nighttime, fog,

rain, and snow, and does not include object tracking or odometry). It offers denser data but has annotations for only 500 frames and is limited in terms of weather variability. In contrast, the RADIATE dataset is extensive (encompasses radar, lidar, camera, nighttime, fog, rain, snow, as well as object detection, object tracking, and odometry) (Sheeny, 2020), (Sheeny et al., 2021).

The RADIATE dataset obtained through the Navtech CTS350-X (Sheeny et al., 2021) RADAR between February 2019 and February 2020. This scanning RADAR generates 360° high-resolution range-azimuth images. The RADAR boasts a maximum range of 100 m, along with 17.5 cm range resolution, 1.8° azimuth resolution, in addition to 1.8° elevation resolution. The RADIATE data was collected in seven different scenarios (overcast (motorway), rain (suburban), sunny (parked), sunny/overcast (urban), fog (suburban), night (motorway), and snow (suburban)) with Eight different labeled objects (vans, truck, buses, cars, motorbikes, bicycle, pedestrian, and group of pedestrians). Examples of RADAR images in polar (raw polar format with resolution 400 x 576 (azimuth x range)). Subsequent to projecting each point, bilinear interpolation is employed to fill voids without values, yielding images with a resolution of 1152 x 1152 (Sheeny, 2020).

4. METHODOLOGY AND EXPERIMENTS

4.1 Select Object detection architecture & Backbone networks.

The methods of object detection can be classified into two groups based on deep neural networks: one-stage detector and two-stage detector (Sheeny, 2020).

- Two-stage detectors employ a dual-step process for object detection. In the initial stage, they identify potential regions, and in the subsequent stage, they classify each identified region. Examples of two-stage detectors include Overfeat (Sermanet et al., 2014), R-CNN, Fast R-CNN, and Faster R-CNN (Ren et al., 2016).
- One-stage detectors employ an end-to-end network for simultaneous detection and classification in a single pass. The primary advantage of these methods lies in their speed, as they operate in a single pass without encountering bottlenecks associated with a region proposal algorithm. Examples of one-stage detectors include SSD (Fleet et al., 2014), RetinaNet (Lin et al., 2017), and YOLO (Redmon et al., 2016).

The assessment of various deep learning-based object detectors with different Convolutional Neural Network (CNN) architectures, focusing on the Average Precision (AP) as a primary evaluation factor for detectors and feature extractors, reveals an enhancement in AP values with increased depth of the backbone network. As a result, there is an enhancement in AP when transitioning from AlexNet to the ResNet-50 architecture. However, further increasing depth leads to a decline in performance, as observed in the AP results for Resnet-101, Inception-v3, and EfficientNet-B0. This decline is attributed to the reliance of object detectors on features at the end of the backbone architecture for object detection. Excessive depth can lead to feature maps with very low resolution, contributing to a deterioration in object detection performance. Faster RCNN surpassed other detectors, achieving the highest AP of 0.97 when utilizing ResNet-50 as the backbone feature extraction network (Azam et al., 2022).

The object detection processing time is a crucial consideration in evaluating various detectors and feature extractors The comparison involves assessing the average time of detection for each deep learning-based object detector In addition to each backbone feature extractor calculated on a single GPU shows that both YOLOv2 and v3 take similar time to detect objects. Furthermore, these object detectors outperform region proposal-based object detectors by at least a factor of five. On the other hand, Faster-RCNN with Resnet 50 got 1.7 seconds per frame which is slow for RADAR detection but as RADAR data is very challenging, so we need high performance, which is provided on faster RCNN with Resnet 50, in addition to the comparison done using only single GPU but in the case of RADAR detection it will be available hardware with high specifications which will increase the detection speed(Azam et al., 2022).

Prior to the introduction of faster R-CNN, advanced models for vehicle detection relied on selective search to approximate target locations. The networks like SPPnet(He et al., 2015) and fast R-CNN succeeded in reducing the operational time of detection networks, but the computations remained time-intensive. In comparison to R-CNN and fast R-CNN, faster R-CNN consolidates the four fundamental steps of object detection— candidate region generation, feature extraction, classification, and location refinement—into an integrated deep network framework. This integration is particularly beneficial because the candidate region selection algorithm (Selective Search) consumed a significant amount of time during detection. Faster R-CNN introduces the use of CNN RPNs (Region Proposal Networks) to handle candidate region selection. The RPN network initially generates region proposals and subsequently classifies them, employing a two-stage processing approach. The feature map of the convolutional layer of the region detector in the fast R-CNN serves to generate candidate regions for the RPNs. Building upon feature mapping, several backward convolutional layers are added to create a regional recommendation network, representing a fully convolutional network (FCN) (Long et al., 2015).

The core concept of the RPN network is as follows: within the extracted feature map, a feature vector is derived through a sliding window and then sent to two layers, the bounding box regression layer and the bounding box classification layer. A sliding window is applied to traverse every point on the feature map, establishing k anchor boxes at each point. Although these k anchor boxes are used to extract features from the feature map, their effectiveness is limited. To enhance performance, a classifier and a box regression are employed. There are two parallel loss functions, softmax, and smoothL1, that classify and regress each region of interest (RoI) respectively. In this way, the model can get a real category and more precise coordinates, length, and width of each RoI (Long et al., 2015).

The RPN is equipped with a loss function that considers both the object's class and its location. The formulated loss functions for RPN are represented by Eq. (1):

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{loc}} \sum_i p_i^* L_{loc}(t_i, t_i^*) \quad (1)$$

In this context, "i" represents the index of an anchor within a mini-batch, and p_i denotes the predicted probability of anchor "i" being an object. The label of ground-truth p_i^* is assigned 1 if the anchor is positive and is assigned 0 if the anchor is negative. The vector t_i signifies the 4 parameterized coordinates of the predicted bounding box, while t_i^* represents the corresponding parameters of the ground-truth box associated with a positive anchor. The classification loss denoted as L_{cls} is log loss

computed over two classes (object vs. not object). Regarding the regression loss, denoted as L_{loc} (t_i, t_i^*) = R (t_i − t_i^*) where R denotes the robust loss function (smooth L1). The term $p_i^* L_{loc}$ indicates that the regression loss is activate only for positive anchors ($p_i^* = 1$) and is inactive otherwise ($p_i^* = 0$). The cls and loc layers produce outputs represented by {p_i} and {t_i} correspondingly.

ResNet-50 (He et al., 2016) employs residual layers, which are convolutional neural networks (CNNs) featuring "Shortcut connections". These connections bypass the current layer, and the excluded output is incorporated into the output after the convolution is executed. ResNet involves a trade-off between accuracy and network depth: a smaller network leads to faster performance. 50-layer ResNet formed by replacing each 2-layer block in the 34-layer net with this 3-layer bottleneck block, resulting in a 50-layer ResNet. The 50-layer ResNets is more accurate than the 34-layer ones by considerable margins (Sheeny, 2020) .

So, I adopted the Faster R-CNN (Ren et al., 2016) architecture with Resnet-50 as a backbone to illustrate the application of RADIATE for RADAR-based object detection. Two adjustments were incorporated into the original architecture to enhance its compatibility with RADAR detection:

- Pre-defined sizes were used for anchor generation [8, 16, 32, 64, 128] because vehicle volumes are typically well-known and RADAR images provide metric scales, different from camera images.
- I modified the Region Proposal Network (RPN) from Faster R-CNN to output the bounding box and a rotation angle in which the bounding boxes are represented by x, y, width, height, and angle.

4.2 Experiments

This section discusses the experimental setup, dataset, hyperparameters, and evaluation metrics.

4.2.1 Platform specifications and requirements: The experiments are performed on Google Colab and the evaluation is done by using COCO evaluation Metrics. The COCO evaluation library needs two requirements first, the annotation file should be in coco format but the annotation file which provided with the RADAR data was in another format, so I design a code to convert the data annotation file format to COCO format. Second, the data images should contain 3 channels, but the RADAR data is gray images that have 1 channel, so I design a small code to build 3 channels for each image.

4.2.2 Dataset specifications: The applied systems of Vehicle detection undergo training and testing on RADAR data (RADIATE Dataset) which contain 8 classes (Car, Van, Truck, Bus, Motorbike, Bicycle, Pedestrian, Group of Pedestrian) with labels. It consists of 5459 RADAR images with a resolution of 1152 × 1152 pixels, employed for training the deep learning-based object detectors (Incorporates data captured under both favorable and adverse weather conditions (night, rain, fog, and snow)), A set of 5,000 images, each with a resolution of 1152 × 1152 pixels, was utilized for performance assessment and testing (data collected in both favorable and adverse weather conditions serving as a means for evaluation and benchmarking).

4.2.3 Hyperparameters: The hyperparameters of the Convolutional Neural Network (CNN) can be fine-tuned to achieve optimal training by selecting the best parameters, including the optimizer function, number of epochs, and learning rate. An epoch represents a complete pass of the data through the architecture. Typically, optimizer values are chosen to be large enough to minimize the loss for the system but not excessively large to prevent overfitting. In my experiments, the Stochastic Gradient Descent Method (SGDM) proved to be a more effective optimizer than ADAM, particularly for transfer learning. However, ADAM exhibited superior performance when training the model from the beginning. An appropriate initial learning rate of 0.001 is selected, and adjustments are made based on the observed higher false alarm rate for larger values, then divided by 2 after 30% of the iterations and then divided again by 2 after 70% of the total iterations. The momentum value is 0.9 and the value of weight decay is 0.0001. Anchor boxes play a crucial role in tuning Faster RCNN object detectors, influencing their efficiency and precision. But in this paper, I predefined the size anchor boxes [8, 16, 32, 64, 128], since the sizes of vehicles are generally established, and RADAR images offer metric scales.

4.2.4 Evaluation metrics: In this paper, the COCO evaluation Metrics (average precision (AP) metric with Intersection over Union (IoU) equal to 0.5) to assess and compare the accuracy of various object detection models. which is the same evaluation metrics used in the PASCAL VOC and DOTA. Therefore, Initially, the IoU is calculated for each bounding box with a confidence score exceeding a threshold γ, in addition to computing the ground truth. Let A denote the detected bounding box and B represent the ground truth bounding box. The IoU is calculated as the ratio of the overlap area between the ground truth and predicted bounding boxes to their union, as described in Eq. (2). If the computed IoU exceeds the threshold γ, the object is labeled as a true positive (TP) detection; otherwise, it is categorized as a false positive (FP). Following this, the precision value is calculated using Eq. (3). Throughout the two experiments, a threshold value of $\gamma = 0.5$ is employed to derive the results.

The evaluation was applied after parts of iterations over 7 different scenarios (Urban, Fog, Static, Motorway, Night, Rain, Snow. The accuracy assessment was done over 7 scenarios to identify the improvement opportunities.

$$IoU = \frac{A \cap B}{A \cup B} \qquad (2)$$

$$Precision = \frac{TP}{TP + FP} \qquad (3)$$

4.2.5 Experiment 1: First, this experiment started with training the RADAR data at which all the 8 classes were combined and named as vehicle which means detecting only one class (vehicle). This experiment is the same as the literature but with different hyper-parameters to check the efficiency of the model in detection.

During the training with 130000 iterations, changing the hyper-parameters and then evaluating the model after 40000, 80000, and 130000 iterations over 7 different scenarios (Urban, Fog, Static, Motorway, Night, Rain, Snow) to see if the hyper-parameters accepted and give good results and there is no overfitting or saturation happen or need to change them. This modification continues until we get to the best hyper-parameters which achieved the best performance and accuracy.

4.2.6 Experiment 2: First, by checking the data and annotation file, it's found that the annotation file contains 8 different classes so, why we have 8 classes and detect only 1 class especially that the autonomous driving system needs to identify all the different objects to predict their movement and their volume, then beginning for training the RADAR data for all the 8 classes to detect 8 classes (Car, Van, Truck, Bus, Motorbike, Bicycle, Pedestrian, Group of Pedestrian). During my research, I did not find any experiments on RADAR data detecting more than 1 class. During the training with 230000 iterations, changing the hyper-parameters and evaluating the model after 25000, 60000, 90000,130000, 170000, and 230000 iterations over 7 different scenarios (Urban, Fog, Static, Motorway, Night, Rain, Snow) to see if the hyper-parameters accepted and give good results and there is no overfitting of saturation happen or need to change them. This modification continues until we get to the best hyper-parameters which achieved the best performance and accuracy.

5. RESULTS AND DISCUSSION

5.1 Experiment 1

no. of iteration	Urban	Fog	static	Motor-way	night	rain	snow
40000	**41.2**	84.8	94.7	22.4	46.1	28.3	3.7
80000	37.5	**92.8**	**95.8**	**23.2**	44.4	23.7	1.5
130000	36.3	90.7	95.3	21.3	**49.5**	**32.2**	**5.6**

Table 1. Average precision over different scenarios

The outcomes, as presented in TABLE 1, indicate that the bias is predominantly associated with the data type rather than the weather conditions. The static scenario when the vehicle is parked, stands out as the most straightforward, attaining nearly 95% Average Precision (AP), likely due to the consistency in RADAR returns from the surrounding environment. Conversely, performance in snow and motorway data scenarios was suboptimal. The smaller size of the snow dataset, coupled with its absence from any training sets, likely impacted the results. Notably, the foggy scenario achieved an impressive 90% AP, showcasing RADAR's efficacy in challenging scenarios for optical sensors, such as dense fog. In the case of night data, a commendable result of close to 50% AP was achieved, highlighting RADAR's resilience to illumination challenges as an active sensor. While the results showed that the rain and urban scenarios achieved considerably better results which are 32.2%, and 36.3% respectively. The results show that, by increasing the training iterations, the average precision of some scenarios decreases, and it is an indication of overfitting and some increases. Finally, comparing my results from experiment 1 achieved an overall AP of (47.2) to the literature results (Sheeny et al., 2021) which achieved an overall AP of (45.77) TABLE 2, my result (2) was slightly better in accuracy over the literature (1) mainly due to hyperparameters optimization.

	overall	Urban	Fog	static	motor way	night	rain	snow
1	45.7	36.1	48.2	78.9	**42.1**	**54.7**	**33.5**	**12.8**
2	**47.2**	**36.3**	**90.7**	**95.3**	21.3	49.5	32.2	5.6

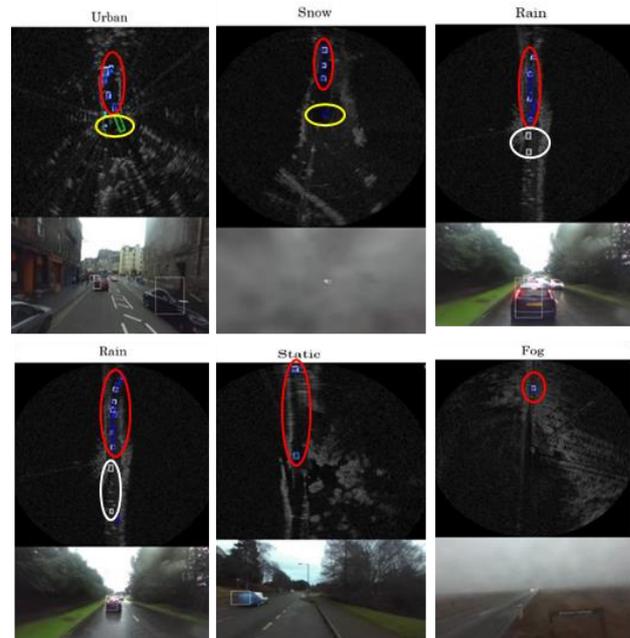Table 2. comparison between this experimental results and literature results.

Figure 3. Visualization of Experiment 1 Results (Red means correct detection, White means false detection, Yellow means misdetection).

5.2 Experiment 2

no. of iteration / scenario	Urban	Fog	static	Motor-way	night	rain	snow
25000	0.4	36.5	3	2.7	2.3	0.5	0.96
60000	0.45	37.8	1.8	2.6	2.2	0.22	0.65
90000	0.45	38.6	2.1	3	2.6	0.33	0.7
130000	8.9	34.2	54.2	5.1	15.2	11.6	21.7
170000	8.6	34.4	57.7	6.7	16.9	10.9	24.5
230000	**8.9**	**41.4**	**59.7**	**14.7**	**25.3**	**11.5**	**30.4**

Table 3. Average precision over different scenarios.

Concerning the results across 8 classes in each scenario, the bias is primarily associated with the data type rather than the weather conditions, as illustrated in TABLE 3. The static scenario (parked) is demonstrated to be the least challenging, achieving an approximate 60% Average Precision (AP). Results in rain, urban and motorway data performed worse. In the foggy scenario, I achieved 41% AP which is good results. Given the considerable challenge that fog poses for optical sensors, RADAR proves to be an effective solution for robust perception in dense fog conditions. While using the night data I achieved considerable results close to 25% AP which is good results. Finally, comparing my results from experiment 1 achieved an overall AP of (47.2) to the results from experiment 2 which achieved an overall AP of (27.4). As expected, the results from experiment 1 over 1 class were better than the results from experiment 2 over 8 classes in, as the model used to detect more than one object, if it detects the object but can't recognize its class it will cause false negative will make the detection confusion, so it decreases the average precision and also, the representations of the eight classes in the data are not equivalent some classes have thousands of objects, and some have just tens of objects which make the average precision not very high. the result showed that there is still potential for improvement.
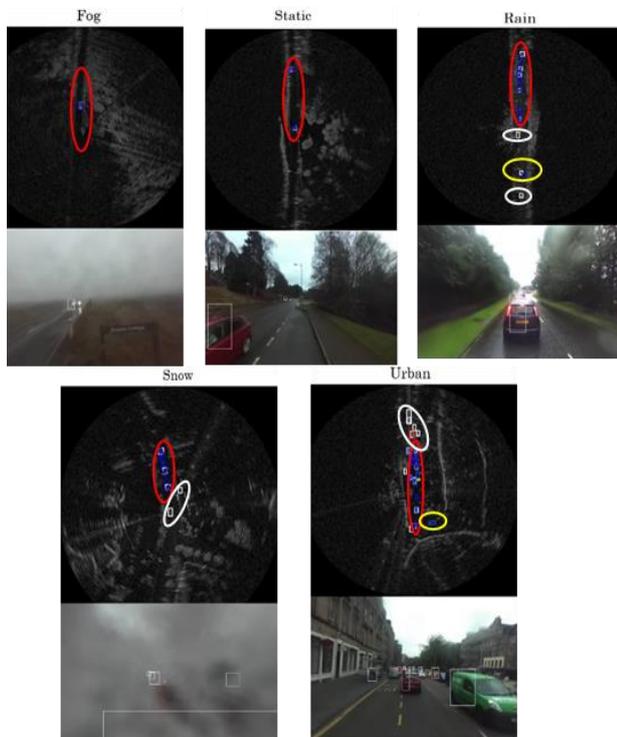
Figure 4. Visualization of Experiment 2 Results (Red means correct detection, White means false detection, Yellow means misdetection).

The results from the two experiments showed that the high AP achieved during the static scenario (parked), which means that collecting data while moving affected the quality of the images which affects the accuracy of detection objects.

6. CONCLUSION

The findings indicated that RADAR has the potential to serve as a perception sensor for autonomous vehicles, proving its functionality in different weather conditions. The results indicate that object detection utilizing RADAR is minimally impacted by weather conditions, particularly in foggy scenarios. When training the model under varying weather conditions, I obtained a 47.2% Average Precision (AP) for experiment 1. On the other hand, I achieved 27.4% AP for experiment 2, as expected the results of detecting one class are better than detecting 8 classes for two reasons. Firstly, when we are detecting one class, any object detected is related to this class but when we are detecting eight classes if the model detects an object without detecting its class, it will make the model confused which makes it misclassify.

Secondly, as the representations of the eight classes in the data are not equivalent some classes have thousands of objects, and some have just tens of objects which makes the average precision not very high (decreasing the accuracy). Comparing my results from experiment 1 achieved an overall AP of (47.2) to the literature results which achieved an overall AP of (45.77), My result was slightly better in accuracy than the literature mainly due to hyper-parameters optimization. I trained the model on all objects (8 classes) in experiment 2, according to the result, there is still potential for improvement.

I did an accuracy assessment over 7 scenarios to identify the improvement opportunities, I discovered that collecting data while moving affected the quality of the images which affects the accuracy of detection objects.

REFERENCES

Appiah, N., Bandaru, N., 2011. Obstacle detection using stereo vision for self-driving cars 926--932.

Azam, B., Khan, M.J., Bhatti, F.A., Maud, A.R.M., Hussain, S.F., Hashmi, A.J., Khurshid, K., 2022. Aircraft detection in satellite imagery using deep learning-based object detectors. Microprocessors and Microsystems 94, 104630. https://doi.org/10.1016/j.micpro.2022.104630.

Bartsch, A., Fitzek, F., Rasshofer, R.H., 2012. Pedestrian recognition using automotive radar sensors. Advances in Radio Science 10, 45–55.

Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O., 2020. nuScenes: A multimodal dataset for autonomous driving.

Engelcke, M., Rao, D., Wang, D.Z., Tong, C.H., Posner, I., 2017. Vote3Deep: Fast object detection in 3D point clouds using efficient convolutional neural networks, in: 2017 IEEE International Conference on Robotics and Automation (ICRA). Presented at the 2017 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Singapore, Singapore, pp. 1355–1361. https://doi.org/10.1109/ICRA.2017.7989161.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The Pascal Visual Object Classes (VOC) Challenge. Int J Comput Vis 88, 303–338. https://doi.org/10.1007/s11263-009-0275-4.

Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), 2014. Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V, Lecture Notes in Computer Science. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-10602-1.

Li, B., Zhang, T., Xia, T., 2016. Vehicle Detection from 3D Lidar Using Fully Convolutional Network.
Ghafoorian, M., Nugteren, C., Baka, N., Booij, O., Hofmann, M., 2019. EL-GAN: Embedding Loss Driven Generative Adversarial Networks for Lane Detection, in: Leal-Taixé, L., Roth, S. (Eds.), Computer Vision – ECCV 2018 Workshops, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 256–272. https://doi.org/10.1007/978-3-030-11009-3_15.

Kelly, A., Amidi, O., Bode, M., Happold, M., Herman, H., Pilarski, T., Rander, P., Stentz, A., Vallidis, N., Warner, R., 2006. Toward Reliable Off Road Autonomous Vehicles Operating in Challenging Environments, in: Ang, M.H., Khatib, O. (Eds.), Experimental Robotics IX, Springer Tracts in Advanced Robotics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 599–608. https://doi.org/10.1007/11552246_57.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90. https://doi.org/10.1145/3065386.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation.

Premebida, C., Monteiro, G., Nunes, U., Peixoto, P., 2007. A Lidar and Vision-based Approach for Pedestrian and Vehicle Detection and Tracking, in: 2007 IEEE Intelligent Transportation Systems Conference. Presented at the 2007 IEEE Intelligent Transportation Systems Conference, IEEE, Bellevue, WA, USA, pp. 1044–1049. https://doi.org/10.1109/ITSC.2007.4357637.

Radecki, P., Campbell, M., Matzen, K., 2016. All Weather Perception: Joint Data Association, Tracking, and Classification for Autonomous Ground Vehicles.

Ramasamy, S., Sabatini, R., Gardi, A., Liu, J., 2016. LIDAR obstacle warning and avoidance system for unmanned aerial vehicle sense-and-avoid. Aerosp Sci Technol 55, 344–358. https://doi.org/10.1016/j.ast.2016.05.020.

Reina, G., Johnson, D., Underwood, J., 2015. Radar Sensing for Intelligent Vehicles in Urban Environments. Sensors 15, 14661–14678. https://doi.org/10.3390/s150614661.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.

Roos, F., Bechter, J., Knill, C., Schweizer, B., Waldschmidt, C., 2019. Radar Sensors for Autonomous Driving: Modulation Schemes and Interference Mitigation. IEEE Microwave 20, 58–72. https://doi.org/10.1109/MMM.2019.2922120.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2014. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks.

Skolnik, M.I., 1980. Introduction to radar systems, 2d ed. ed. McGraw-Hill, New York.

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., Lau, K., Oakley, C., Palatucci, M., Pratt, V., Stang, P., Strohband, S., Dupont, C., Jendrossek, L.-E., Koelen, C., Markey, C., Rummel, C., Van Niekerk, J., Jensen, E., Alessandrini, P., Bradski, G., Davies, B., Ettinger, S., Kaehler, A., Nefian, A., Mahoney, P., 2006. Stanley: The robot that won the DARPA Grand Challenge. J. Field Robotics 23, 661–692. https://doi.org/10.1002/rob.20147.

Wang, J., Sun, X., Guo, J., 2013. A Region Tracking-Based Vehicle Detection Algorithm in Nighttime TrafficScenes. Sensors 13, 16474–16493. https://doi.org/10.3390/s131216474.

KAIST, "Korean competition shows weather still a challenge for autonomous cars." http://spectrum:ieee:org/cars-that-think/transportation/advanced-cars/japan-competition shows weather-still-a-challenge-for-autonomous cars, 2014:

Sucgang, N.J., Jr, M.R., Arriola, N.A., 2017. Road Surface Obstacle Detection using Vision and LIDAR for Autonomous Vehicle. Hong Kong.

Pfennigbauer, M., Wolf, C., Weinkopf, J., Ullrich, A., 2014. Online waveform processing for demanding target situations, in:

Laser Radar Technology and Applications XIX; and Atmospheric Propagation XI. SPIE, pp. 142–151.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J., 2016. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. IEEE Transactions on robotics 32, 1309–1332.

Wu, Y., Liu, Y., Li, J., Liu, H., Hu, X., 2013. Traffic sign detection based on convolutional neural networks, in: The 2013 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–7.

Grimes, D.M., Jones, T.O., 1974. Automotive radar: A brief review. Proceedings of the IEEE 62, 804–822.

Rasshofer, R.H., 2007. Functional requirements of future automotive radar systems, in: 2007 European Radar Conference. IEEE, pp. 259–262.

He, C., Zeng, H., Huang, J., Hua, X.-S., Zhang, L., 2020. Structure aware single-stage 3d object detection from point cloud, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11873–11882.

Sheeny, M., De Pellegrin, E., Mukherjee, S., Ahrabian, A., Wang, S., Wallace, A., 2021. RADIATE: A Radar Dataset for Automotive Perception in Bad Weather.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37, 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Krishnan, A.B., Kollipara, J., 2014. Intelligent Indoor Mobile Robot Navigation Using Stereo Vision. SIPIJ 5, 45–53. https://doi.org/10.5121/sipij.2014.5405.

Nordenmark, V., Forsgren, A., 2015. Radar-detection based classification of moving objects using machine learning methods.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

Bazzani, L., Bergamo, A., Anguelov, D., Torresani, L., 2016. Self-taught object localization with deep networks, in: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). Presented at the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, Lake Placid, NY, USA, pp. 1–9. https://doi.org/10.1109/WACV.2016.7477688.

Sheeny, M., 2020. All-Weather Object Recognition Using Radar and Infrared Sensing, PhD Thesis, Heriot-Watt University, Edinburgh EH14 4AS, UK.