

# AUTOMOTIVE RADAR BASED LEAN DETECTION OF VEHICLES

A. Moussa<sup>1,2,\*</sup>, N. El-Sheimy<sup>1</sup>

<sup>1</sup> Dept. of Geomatics Engineering, University of Calgary, Calgary, T2N 1N4 Canada - [amelsaye@ucalgary.ca](mailto:amelsaye@ucalgary.ca), [elsheimy@ucalgary.ca](mailto:elsheimy@ucalgary.ca)

<sup>2</sup> Department of Electrical Engineering, Port-Said University, Port Said, Egypt - [amelsaye@ucalgary.ca](mailto:amelsaye@ucalgary.ca)

Commission I, WG I/7

**KEY WORDS:** Vehicle, Detection, Automotive Radar.

## ABSTRACT:

One of the most critical features of autonomous vehicles is the detection of road active objects such as vehicles and pedestrians. The autonomous vehicles' navigation planning and manoeuvre decision-making are aided by the detection of such active objects, resulting in safe and efficient navigation. Deep Convolutional Neural Networks (CNNs) have recently advanced to become one of the state-of-the-art ways to solving detection challenges, particularly in the autonomous vehicle area. Deep CNNs typically use a large number of processing layers with a high number of kernels per layer to enable detection of the target classes which also demands the use of powerful hardware units. In this research, we present a tailored lean detection strategy for vehicle detection using radar observations. The proposed method employs a compact set of convolutions, as well as pixel classification and a customized selection of kernels and kernel sizes, to provide an efficient technique that greatly decreases detection burden and enables real-time processing on average processing units. A training dataset is used to train the convolution window sizes and the pixel classifiers. Finally, the pixel classified grids are processed to identify the vehicles' bounding boxes. Experimental data sets have been collected using medium-range radar sensors mounted on top of a vehicle to evaluate the suggested approach, the Intersection over Union (IoU) values of the test scenes' detections range from 0.51 to 0.78.

## 1. INTRODUCTION

Detection of the road active objects such as vehicles and pedestrians is one of the crucial requirements of autonomous vehicles. The detection of such active objects assists the autonomous vehicles in navigation planning and manoeuvre decision making towards safe and efficient navigation. Moreover, road mapping using the numerous available modalities requires reliable detection of non-stationary road objects to be excluded from the reconstructed map.

While cameras can be considered the main perception sensors in autonomous vehicle systems, radar sensors are increasingly adopted by many manufacturers to aid the process of perception of the vehicle surroundings. Typically, Frequency Modulated Continuous Wave (FMCW) radar sensors are mounted on such autonomous vehicles to offer depth and Doppler measurements of the hit surrounding objects. Unlike other ultrasonic or laser sensors, radar sensors' measurements are not degraded in harsh conditions such as fog, rain, and snow. These unique characteristics enable radar sensors to play a significant role in environment awareness and detection process either solely or collaboratively with other sensors such as cameras to compensate the camera's lack of depth and Doppler measurements.

The recent advancements of deep Convolutional Neural Networks (CNNs) place it among the state of art approaches to address detection problems especially in the autonomous vehicles' domain. Numerous deep learning approaches have been proposed for detection of road actors from images using different network architectures such Regions with CNN features (R-CNN), faster R-CNN and You Only Look Once (YOLO).

Radar sensors typically offer various measurements about the hit target which include range, azimuth angle, elevation angle, power of reflected signal, cross section of target, and Doppler measurements. These measurements offer 3D perception of the surrounding targets but with much sparse fashion compared to

LiDAR measurements. These 3D radar measurements can be stacked into 2D layers representing the collected measurements where the popular CNNs models such R-CNN, faster R-CNN and YOLO can be applied towards the road actors detection objective.

Generally, deep CNNs employ a large number of processing layers with larger number of kernels per layer in order to capture the features of the target classes, to abstract these features and finally to offer scores for potential locations of these classes in the final layers to be processed to detect the class instances in the scene. This requires a massive number of convolution calculations throughout the network architecture to reach the final result. While this deep architecture offers a generic framework for feature extraction and instance localization under different input transformations, it also requires powerful processing units to match its needed computation load that exceeds billions of floating points operations per image.

In this paper, we propose a tailored lean detection approach for vehicle detection from radar measurements. The proposed approach adopts a concise set of convolutions with tuned selection of kernels, kernel sizes along with pixel classification to offer an efficient approach that significantly reduces the detection computation load and enables real-time vehicle detection on average processing units.

## 2. RELATED WORK

The recent developments of radar sensors and its increasing adoption by autonomous vehicle systems motivate increasing research effort towards employing radar sensors in environment awareness and detection tasks.

Object detection is defined as the process of identifying where objects exist in an input dataset and to which class each object belongs. Convolutional neural networks are the most widely

\* Corresponding author

used tools for addressing detection problems using different modalities in the different domains (Zhao et al. 2019).

Radar sensors have been employed solely by many researchers for vehicle detection to help the perception of the vehicle surroundings. (Ouaknine et al. 2021) introduced multiple deep neural network architectures that employ range-angle-Doppler (RAD) tensor for vehicle detection. (Cennamo et al. 2021) proposed a neural network architecture named RadarPCNN based on PointNet++ (Qi et al. 2017) as a building-block to perform semantic segmentation of radar point clouds.

(Schumann et al. 2018) have also introduced a neural network architecture named SegNet-2 for segmentation of multiple road classes including vehicles. (Major et al. 2019) used image-like range-azimuth-Doppler tensors instead of the radar point cloud and proposed a Recurrent Neural Network (RNN) that uses Long Short-Term Memory (LSTM) modules for vehicle detection. (Fang et al. 2007) utilized Doppler signature of low-cost K-band unmodulated CW radar to detect moving vehicles using signal processing.

Furthermore, radars have been widely fused with cameras to compensate for the absence of depth information in cameras. (Chadwick et al. 2019) fused cameras of different focal lengths with radars and proposed a neural network detection framework based on SSD architecture. (Lim et al. 2019) proposed Fusion net inspired by SSD architecture to fuse camera with radar for vehicle detection. (Cui et al. 2021) proposed a CNN based vehicle detection using both cameras and radars. (Bombini et al. 2006) employed radar data to identify potential vehicles' locations in camera images to support a more robust visual vehicle detection. (Liu et al. 2011) proposed a cross-verification vehicle detection that fuses cameras and radar where Support Vector Machine (SVM) based vehicle detection using images is verified using radar data.

There has been a plenty of successful research developments in the object detection tasks in the past few years. Numerous neural network architectures have been proposed to offer increasingly accurate and reliable object detection such as Region-based Fully Convolutional Network (R-FCN) (Dai et al. 2016), Feature Pyramid Network (FPN) (Lin et al. 2017), Auto-FPN (Xu et al. 2019), RCNN (Girshick et al. 2014), Fast-RCNN (Girshick 2015), Faster-RCNN (Ren et al. 2015), Mask-RCNN (He et al. 2017), Path Aggregation Network (PANet) (Liu et al. 2018), Spatial Pyramid Pooling network (SPP-net) (He et al. 2015). These network architectures identify the potential regions of the target classes in the first phase then these regions are processed afterwards to determine the classes and their bounding boxes coordinates.

Another set of neural network architectures emphasizes the fast processing to meet the real-time requirements, These networks such as YOLO (Redmon et al. 2016), YOLO-V3 (Redmon and Farhadi 2018), Single Shot multiBox Detector (SSD) (Liu et al. 2016) alternatively adopt a faster single processing phase to identify and locate the classes. Despite the architectural reduction adopted by these faster networks, these networks still have large number of consecutive convolution layers. As an example, a minimalistic implementation such as the Tiny-YOLO-v2 network consists of 9 convolution layers with 16 to 1024 filters/kernels per each layer. This tiny network takes approximately 7.3 billion operations to process a 416x416 pixels image (Wai et al. 2019) which still constitutes a substantial processing load given that it covers a single task among many other tasks needed by the overall system. The convolution layers accounts for approximately 90% of the feed-forward computation (Cong and Xiao 2014). Therefore, further

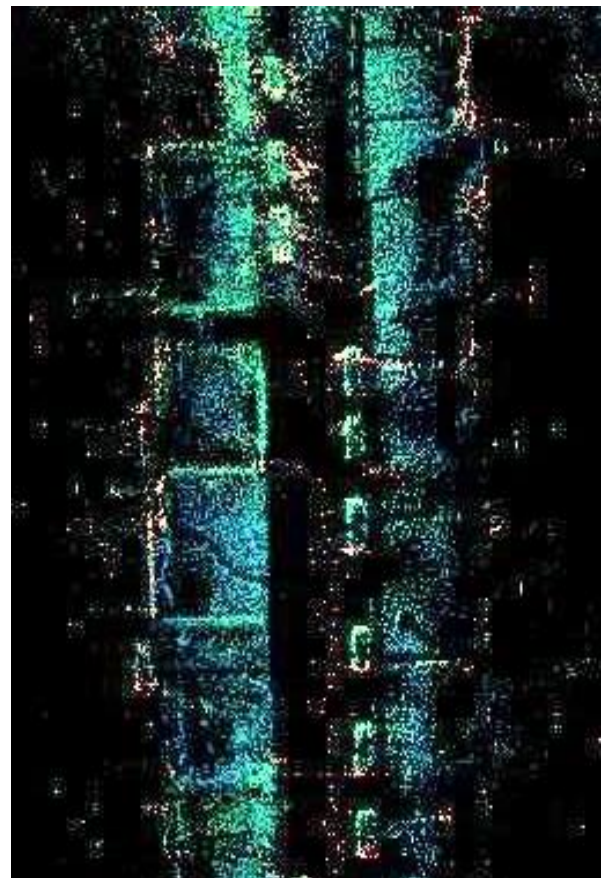
reduction of the convolution networks could significantly decrease the processing burden of the detection task.

In this research, a tailored lean detection approach for vehicle detection from radar measurements is proposed. This approach employs a concise set of convolutions with simple kernels to considerably decrease the detection processing load and to enable operation on average processing units.

### 3. METHODOLOGY

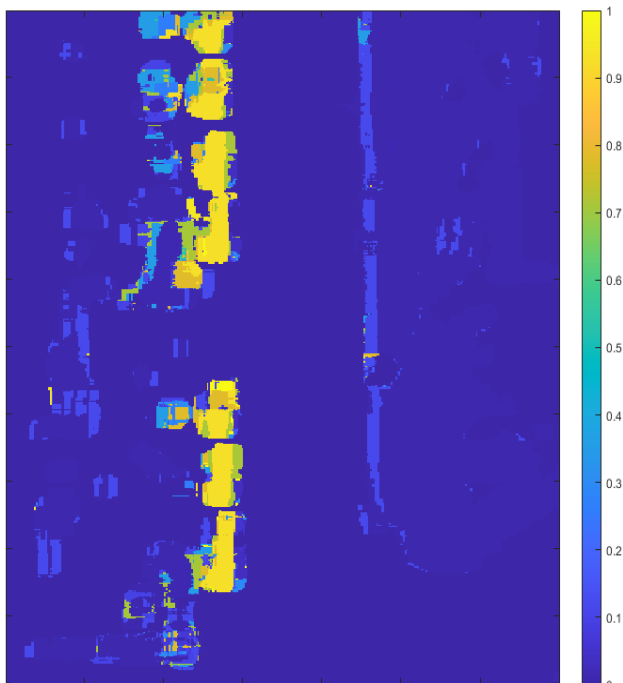
Radar sensors can offer range, azimuth and elevation angles, power of reflected signal, target cross section, and Doppler measurements, among other information about the impacted target. Such measurements offer a sparse three-dimensional view of the targets. The proposed approach stacks these measurements into 2D grids representing the radar reflections' characteristics such as occupancy, height, cross-section, power, and noise level. Figure 1 depicts a sample false-color image of radar points representing the cross section, power, and the noise level of the radar reflections.

A considerable number of noisy measurements are scattered all over the scene as shown in Figure 1. The characteristics of the parked vehicles in the scene (as depicted by the false-colors) are not easily distinguishable as pixels from the other objects in the scene. Primary shape features such as gradients computed in the neighbourhood of each pixel can be of more importance than individual pixel characteristics. Convolution layers of proper window sizes and with suitable kernels can provide these features for the classification purpose.



**Figure 1.** Sample false-color image of radar points representing the cross section, power, and the noise level of the radar reflections.

Instead of the typical numerous layers of the deep learning approaches, the proposed approach includes a single layer of compact number of convolution operations with a limited set of simple kernels of different window sizes. Since the radar points of the vehicles exhibit simple noisy rectangle shapes, simple kernels could be sufficient for representing the pixels neighbourhood. Haar-like kernels that represent average, first-order gradients in x and y directions are adopted since its computation could be significantly accelerated using integral images approach (Tapia 2011). The outputs of the kernel convolutions are fed into a pixel classifier that provides a vehicle classification probability/score to each pixel as shown in Figure 2.



**Figure 2.** Pixel based classification probability/score.

Preliminary pixel-based potential detections are obtained using the pixels above a threshold probability. The preliminary potential detection regions are morphologically processed to discard tiny regions and to find bounding boxes around potential regions to form the final detections. Figure 3 illustrates the detection results of the scene in Figure 2, cyan represents ground truth labels and yellow represents detection results.

The proposed detection pipeline steps are summarized as follows:

---

#### Detection pipeline steps

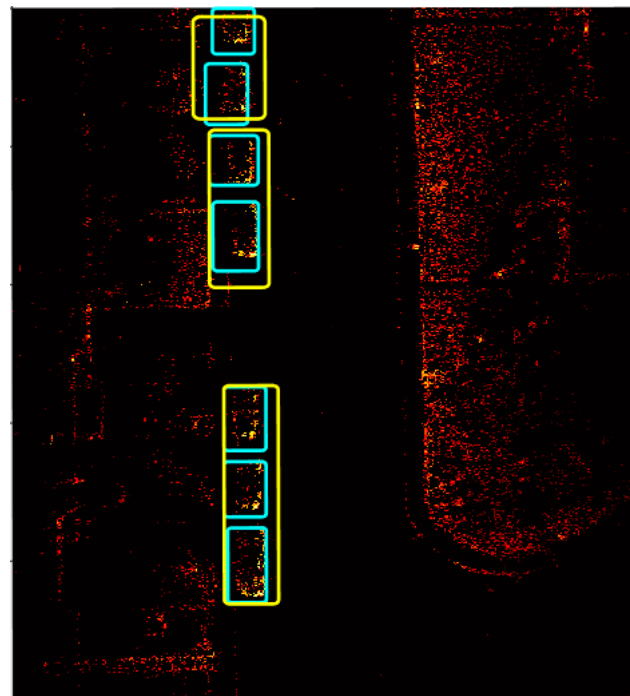
---

- For each selected input radar characteristic 2D grid (height, power, cross section, noise level, occupancy):
    - For each selected window size:
      - For each used kernel:
        - Compute corresponding convolution.
  - Use the convolution outputs as input to the pixel classifier to obtain pixel classification probability.
  - Apply threshold to obtain the potential detection regions.
  - Morphologically process the potential detection regions
- 

---

to discard tiny regions and to delineate detection bounding boxes.

---



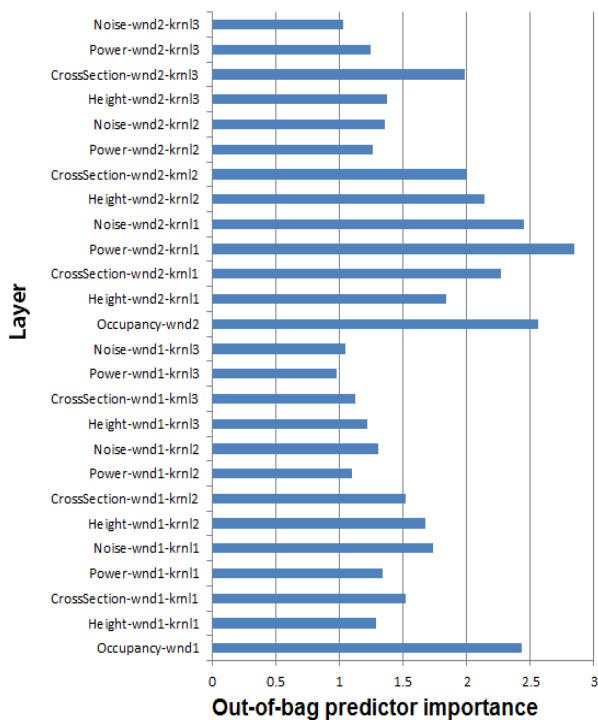
**Figure 3.** Sample detection results, cyan represents ground truth labels and yellow represents detection results.

The performance of the vehicle detection depends significantly on the choice of the parameters such as the convolution window sizes, included radar characteristics, and pixel classification probability threshold.

In order to tune such parameters, some sort of training takes place using labelled dataset with both training and validation sets. The training dataset are employed through the detection pipeline to compute the convolutions, and to train the pixel classifier and processing the potential classification into final detection. The Intersection over Union (IoU) metric is used throughout the tuning process to evaluate the detection performance.

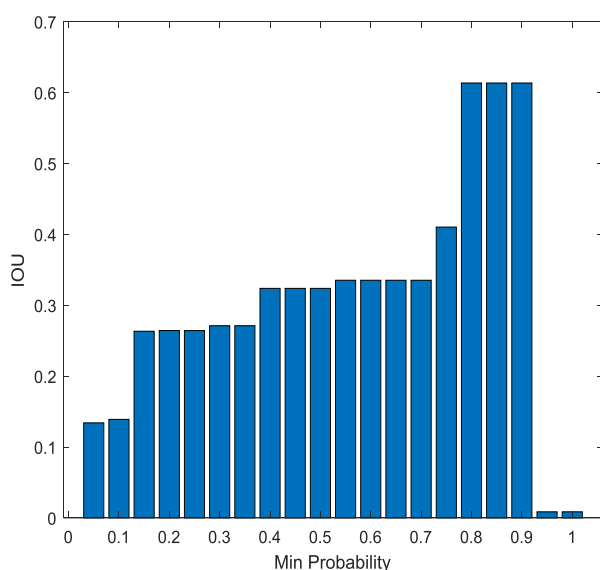
This process iterates through different combinations of window sizes. Moreover, the computed convolutions are also assessed to rank their importance to the pixel classification step using out-of-bag predictor importance estimation. This technique permutes the computed convolutions/predictors during ensemble bag classification to quantify how effective each predictor is for classification. Permuting the values of a predictor should influence the classification error if it is important for classification while permuting the values of a predictor that isn't influential should have little to no effect on the classification error.

The mean difference between classification accuracies in both normal and permuted scenarios divided by its standard deviation indicates the importance of each predictor as shown in Figure 4. This step helps to further prune the employed kernel convolutions by excluding the least contributing.



**Figure 4.** Out-of-bag predictor importance of convolution outputs of different radar characteristics, window sizes and kernels.

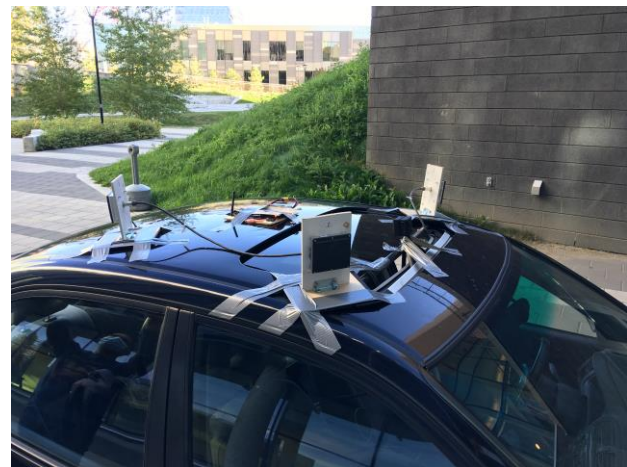
Finally the detection performance is also investigated for different values of probability thresholds to obtain the optimal value. Figure 5 depicts the detection performance in terms of IoU using different values of pixel classification probability threshold values. Using low threshold values include non-correctly classified regions while very high threshold values exclude most of the correct potential detection areas. Threshold values between 0.8 and 0.9 offers the highest average performance.



**Figure 5.** The detection performance in terms of IoU using different values of pixel classification probability values.

#### 4. RESULTS AND DISCUSSION

To assess the proposed approach, experimental data sets have been collected in Calgary city using three smartmicro UMRR-11 type-132 radar sensors mounted on top of a vehicle as shown in Figure 6. The radar sensors have maximum range of 64 meters for the medium resolution with range accuracy of less than 0.25 m and with 100° and 15° horizontal and vertical Field of View (FOV) respectively. The radar points have been georeferenced with the help of low-cost GNSS/IMU/Odometer on-board sensors.

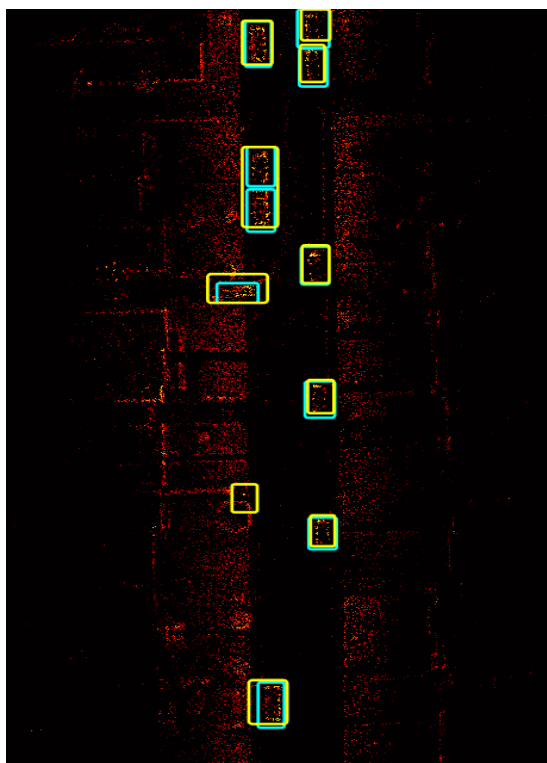


**Figure 6.** Radar sensors mounted on test vehicle.

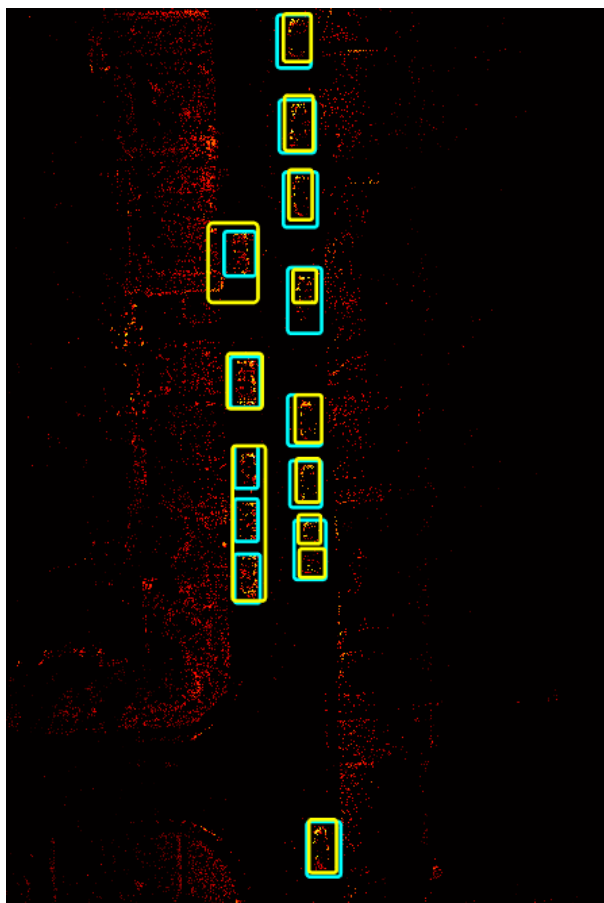
Figures 7, 8 show sample detection results where cyan rectangles represent ground truth labels and yellow rectangles represent the achieved detections. As shown in these samples, all vehicles in both figures have been successfully detected. Some vehicles have been precisely localized with very small boxes deviations. Some other vehicles have been detected with slightly larger boxes. Many vehicles have been grouped with other nearby vehicles into larger detection boxes.

The noises of radar measurements could be the most probable reason of such behaviour of oversizing of detection boxes. The radar measurements noises between nearby vehicles reduce the ability to isolate such vehicles. Additionally, the post-processing of the pixel classification includes dilation step that groups nearby pixels into more solid regions in preparation for determination of boxes' coordinates. Such step could also glue nearby regions into larger detected boxes. Such grouping behaviour is not of great concern to many applications that focuses on the vehicles locations more than its count. A single false positive case exists near the left middle of Figure 7 where the nearby fence and road furniture has been confused by the detection scheme as a vehicle.

Intersection over Union (IoU) metric is used to evaluate the accuracy of the detection results. IoU is simply computed by dividing the area of overlap between the predicted bounding box and the ground-truth bounding box by the area encompassed by both the predicted bounding box and the ground-truth bounding box. The more accurate the matching between the predicted detections boxes and the ground-truth, the higher the value of IoU metric. IoU values for the detections of the test scenes vary between 0.51 and 0.78.



**Figure 7.** Sample detection results, cyan represents ground truth labels and yellow represents detection results.



**Figure 8.** Sample detection results, cyan represents ground truth labels and yellow represents detection results.

## 5. CONCLUSIONS

In this paper, we present a customised lean detection approach for vehicle detection using radar observations. The suggested method uses a 2D grid of radar measurements as an input and employs a compact set of convolutions to capture the primary features needed for pixel classification. Due to the simple rectangular shapes of the vehicles, simple Haar-like kernels are used as their usage for convolution could be accelerated using the integral images technique. The convolution window sizes and the pixel classifiers are trained using a training dataset. The pixel classified grids are finally processed to detect the locations of the vehicles' bounding boxes. The proposed approach has been implemented and tested on real data set acquired using three medium resolution radar sensors mounted on top of a vehicle. The mean IoU of the detection of the test data set is 0.62 which proves the potential of the proposed approach.

## ACKNOWLEDGEMENTS

This work was supported by Dr. Naser El-Sheimy research funds from NSERC CREATE and Canada Research Chairs programs.

## REFERENCES

- Bombini, L., P. Cerri, P. Medici and G. Alessandretti, 2006. Radar-vision fusion for vehicle detection, Proceedings of the Proceedings of International Workshop on Intelligent Transportation, 2006, pp. 70.
- Cennamo, A., F. Kaestner and A. Kummert, 2021. A Neural Network Based System for Efficient Semantic Segmentation of Radar Point Clouds. *Neural Processing Letters*, 53(5):3217-3235.
- Chadwick, S., W. Maddern and P. Newman, 2019. Distant vehicle detection using radar and vision, Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 8311-8317.
- Cong, J. and B. Xiao, 2014. Minimizing computation in convolutional neural networks, Proceedings of the International conference on artificial neural networks, 2014, pp. 281-290.
- Cui, H., J. Wu, J. Zhang, G. Chowdhary and W.R. Norris, 2021. 3D Detection and Tracking for On-road Vehicles with a Monovision Camera and Dual Low-cost 4D mmWave Radars, Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 2021, pp. 2931-2937.
- Dai, J., Y. Li, K. He and J. Sun, 2016. R-fcn: Object detection via region-based fully convolutional networks. *Advances in neural information processing systems*, 29.
- Fang, J., H. Meng, H. Zhang and X. Wang, 2007. A low-cost vehicle detection and classification system based on unmodulated continuous-wave radar, Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, 2007, pp. 715-720.
- Girshick, R., J. Donahue, T. Darrell and J. Malik, 2014. Rich feature hierarchies for accurate object detection and semantic segmentation, Proceedings of the Proceedings of the IEEE

- conference on computer vision and pattern recognition, 2014, pp. 580-587.
- Girshick, R., 2015. Fast r-cnn, Proceedings of the Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448.
- He, K., X. Zhang, S. Ren and J. Sun, 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9):1904-1916.
- He, K., G. Gkioxari, P. Dollár and R. Girshick, 2017. Mask r-cnn, Proceedings of the Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961-2969.
- Lim, T.-Y., A. Ansari, B. Major, D. Fontijne, M. Hamilton, R. Gowaikar and S. Subramanian, 2019. Radar and camera early fusion for vehicle detection in advanced driver assistance systems, Proceedings of the Machine Learning for Autonomous Driving Workshop at the 33rd Conference on Neural Information Processing Systems, 2019.
- Lin, T.-Y., P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, 2017. Feature pyramid networks for object detection, Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117-2125.
- Liu, S., L. Qi, H. Qin, J. Shi and J. Jia, 2018. Path aggregation network for instance segmentation, Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8759-8768.
- Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A.C. Berg, 2016. Ssd: Single shot multibox detector, Proceedings of the European conference on computer vision, 2016r, pp. 21-37.
- Liu, X., Z. Sun and H. He, 2011. On-road vehicle detection fusing radar and vision, Proceedings of the Proceedings of 2011 IEEE International Conference on Vehicular Electronics and Safety, 2011, pp. 150-154.
- Major, B., D. Fontijne, A. Ansari, R. Teja Sukhvasi, R. Gowaikar, M. Hamilton, S. Lee, S. Grzechnik and S. Subramanian, 2019. Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors, Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.
- Ouaknine, A., A. Newson, P. Pérez, F. Tupin and J. Rebut, 2021. Multi-View Radar Semantic Segmentation, Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 15671-15680.
- Qi, C.R., L. Yi, H. Su and L.J. Guibas, 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems, 30.
- Redmon, J., S. Divvala, R. Girshick and A. Farhadi, 2016. You only look once: Unified, real-time object detection, Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779-788.
- Redmon, J. and A. Farhadi, 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
- Ren, S., K. He, R. Girshick and J. Sun, 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.
- Schumann, O., M. Hahn, J. Dickmann and C. Wöhler, 2018. Semantic segmentation on radar point clouds, Proceedings of the 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 2179-2186.
- Tapia, E., 2011. A note on the computation of high-dimensional integral images. Pattern Recognition Letters, 32(2):197-201.
- Wai, Y.J., Z.M. Yussof and S.I.M. Salim, 2019. A scalable FPGA based accelerator for Tiny-YOLO-v2 using OpenCL. Int. J. Embed. Syst, 8:206.
- Xu, H., L. Yao, W. Zhang, X. Liang and Z. Li, 2019. Auto-fpn: Automatic network architecture adaptation for object detection beyond classification, Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6649-6658.
- Zhao, Z.-Q., P. Zheng, S.-t. Xu and X. Wu, 2019. Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems, 30(11):3212-3232.