

Automotive Radar Data Acquisition Using Object Detection

Madhumitha Sakthi¹, Ahmed Tewfik¹

Abstract—The growing urban complexity demands an efficient algorithm to acquire and process various sensor information from autonomous vehicles. In this paper, we introduce an algorithm to utilize object detection results from the image to adaptively sample and acquire radar data using Compressed Sensing (CS). This novel algorithm is motivated by the hypothesis that with a limited sampling budget, allocating more sampling budget to areas with the object as opposed to a uniform sampling ultimately improves relevant object detection performance. We improve detection performance by dynamically allocating a lower sampling rate to objects such as buses than pedestrians leading to better reconstruction than baseline across areas with objects of interest. We automate the sampling rate allocation using linear programming and show significant time savings while reducing the radar block size by a factor of 2. We also analyze a Binary Permuted Diagonal measurement matrix for radar acquisition which is hardware-efficient and show its performance is similar to Gaussian and Binary Permuted Block Diagonal matrix. Our experiments on the Oxford radar dataset show an effective reconstruction of objects of interest with 10% sampling rate. Finally, we develop a transformer-based 2D object detection network using the NuScenes radar and image data.

Index Terms—Compressed sensing, Object detection

I. INTRODUCTION

In an autonomous driving scenario, it is necessary to acquire various sensor data and process them efficiently to get a complete perspective about the surroundings. Radar [1] [2], LIDAR [3] data in addition to images has aided in improving the object detection performance. During onboard radar signal acquisition, CS is a technique used for recovering the original radar data with sampling rates much lower than the Nyquist rate [4] [5]. The authors in [6], showed successful radar reconstruction with 40% sampling rate. In our method, we show efficient reconstruction with 10% sampling rate. In adaptive CS, the CS technique is extended to focus on the important regions by allocating more sampling budget based on prior information [7] [8]. In [9] they used adaptive compressed sensing for a static tracker case and have shown improved target tracking performance. However, in our algorithm, we used adaptive compressed sensing for radar acquisition from an autonomous vehicle where, both the vehicle and the objects were moving.

In our work, we develop an adaptive block CS algorithm to acquire radar data using the object detection results from the image. The images were processed using Faster R-CNN [10] network to obtain the bounding box and the class of an object. The bounding boxes in the camera coordinates were

converted to bird-eye view coordinates and the azimuth of the object was identified. Therefore, the azimuth block with an object was sampled at a higher sampling rate compared to other regions which helped in the efficient reconstruction of objects such as pedestrians, bicycles, cars, trucks and buses while they were either missed or reconstructed poorly by the baseline, standard CS technique. In this work, we compare our method against the standard CS and another simpler algorithm called Algo-1. The standard CS acquisition has a uniform sampling rate of 10% across all the radar frames. In Algo-1, we use the prior image information to identify the important blocks with any object of interest but, the sampling rate is allocated manually and is fixed for a block size of 50x100. In Algo-2, proposed as the main contribution in this paper, compared to Algo-1, we 1) dynamically decrease the sampling rate for car or bus compared to pedestrian and use that budget to recover the trifling areas as well apart from having better reconstruction in the important regions, 2) we automated the dynamic sampling rate allocation using Linear Programming (LP), 3) we reduced the block size of radar reconstruction by a factor of 2 to 25x100 leading to a decrease in reconstruction time by 60% for 10% sampling rate. Therefore with the proposed method, we could reconstruct the trifling radar regions with the same or better reconstruction quality in comparison to the baseline & Algo-1 technique while preserving the reconstruction quality of the objects of interest. Therefore, our reconstruction technique *could* be reliably used for radar segmentation [11] apart from object detection task. In CS acquisition, it is important to design a hardware-efficient measurement matrix in order to operate the algorithm efficiently in real-time. Therefore, we analyze a Binary Permuted Diagonal (BPD) measurement matrix that simplifies hardware acquisition by directly measuring the radar data instead of acquiring a linear combination of the input radar and show that the reconstruction quality is similar to that of the Gaussian measurement matrix.

The binary measurement matrices in CS have been explored by several studies. In [12], the authors proposed a Binary Permuted Block Diagonal (BPBD) matrix with equal-sized diagonal blocks permuted along the columns to create randomness. Binary measurement matrices have been used for image [13], ECG [14] [15] and ground-penetrating radar data acquisition [16]. To the best of our knowledge, we are the first ones to implement the BPD matrix for automotive radar acquisition. Finally, we also develop an end-to-end transformer-based 2-D object detection (DETR) [17] network using the NuScenes [18] radar and image dataset. Numerous studies [19]

¹: Department of ECE, The University of Texas at Austin, United States, madhumithasakthi.iyer@utexas.edu, tewfik@austin.utexas.edu

[20] showed the advantage of using both radar and images in an object detection network for improved object detection performance. [1], [21], [22] showed that radar in addition to image improved distant vehicle object detection and occluded object detection due to adverse weather conditions. Again, to the best of our knowledge, we are the first ones to implement a DETR [17] based object detection network using both radar and image data and we show that the object detection performance of the model using both image and radar data performed better than the object detection model trained only on the image data.

In the next section, we explain the adaptive block CS algorithm followed by the sampling rate allocation using LP and the design of the measurement matrix along with the DETR-Radar architecture. In section 3, we present our results and finally, conclude the paper in section 4.

II. METHOD

A. Dataset

In the Oxford radar robotcar dataset [23], they provide radar, rear camera data, front camera data among other sensors, covering 280km in Oxford, UK. The front camera was captured at 16Hz Frames per second (FPS), the rear camera at 17Hz and radar at 4 Hz. To the best of our knowledge, the Oxford dataset was the only raw radar dataset. Since manual annotation was required to identify objects, we tested our algorithm on three random scenes with 11 radar frames each. To train our DETR-Radar object detection model, we used the NuScenes v0.1 dataset [18]. This is one of the publicly available datasets for autonomous driving with a range of sensors such as cameras, LiDAR and radar with 3D bounding boxes. Similar to [20], we converted all the 3D bounding box annotations to 2D bounding boxes and merged similar classes to 6 total classes, Car, Truck, Motorcycle, Person, Bus and Bicycle. This dataset consists of around 45k images split into 85% training and 15% validation.

B. Adaptive block compressed sensing

CS relies on the sparsity of the signal in some domain for robust reconstruction of data sampled at very low sampling rates. The signal $z \in R^n$ measured using the random measurement matrix $\phi \in R^{m \times n}$ gives $y \in R^m$ measurements. The original signal x is assumed to be sparse in Discrete Cosine Transform domain and it is recovered using Basis Pursuit (BP) algorithm as $\min_x \|\theta x\|_1$ s.t. $\phi x = y$ with θ as the domain transformation matrix [24]. In the adaptive block CS, the radar region is split into equal-sized blocks and varying measurements m are allocated based on prior knowledge.

The Oxford radar data was captured every 0.25s with a range resolution of 4.38cm and 0.9° azimuth resolution for 3768 range bins and 400 azimuth bins. Therefore, a total range of 163m and 360° horizontal field-of-view (HFOV) was captured. Also, the front camera has 66 degrees HFOV. The rear camera has 180 degrees HFOV. From this setup, there would be a blind spot of 57 degrees to the left and 57 degrees to the right. In both Algo-1 and Algo-2, at $t=0s$, the images

from the camera are captured and given to the Faster R-CNN object detection network. At $t=0.12s$, the object detection output is processed to obtain the important azimuth blocks. This data is processed to allocate more sampling budget to the important azimuth blocks and the sampling budget is used to capture the radar data at $t=0.25s$. This process is repeated for all the radar frames.

In Algo-1, the radar data was split into bins of size 50×100 creating 8 equal regions in azimuth and 37 regions in range. Therefore, from the camera image, the important azimuth sections to focus on would be derived based on the presence of objects in that section. Since the depth information is not available from the camera images, we can only choose azimuth sections and not the range. Also, since the radar data was acquired with a 163 m range, we allocated more sampling rate to the first 78.84m compared to the last 84.16m since there is not much useful information in the farther range values. The average driving speed in an urban environment is 40 miles per hour. Since radar is captured at 4 Hz, for every frame, the object could have moved 4.25m. Since the bin resolution is 4.38cm and for a particular block with 100 bins, the area spanned would be 4.38m. Since we are looking into the first 18 range blocks, that covers a total area of 78.84m. Hence, in an urban setting, the moving vehicle can be comfortably captured by focusing within the 78.84m range. In a freeway case, for an average speed of 65 miles per hour, the vehicle could have moved 7.2m per frame and again, this would be captured by focusing on the first 78.84m. Therefore, the radar regions are split into three ranges, R1, R2, R3 where, R1 is the chosen azimuth until 18th range block (78.84m), R2, the other azimuth regions until the 18th range block (78.84m) and finally, R3, all the azimuth blocks from 19th to 37th range block. Since across different scenes, there could have been a variable number of important azimuth blocks, if the chosen azimuth block was less than 50% i.e., less than 4 out of 8, we randomly sampled more azimuth blocks. Also, if it was more than 4, we avoided losing important information by reducing the sampling rate in the R1 region. Hence, when there were 4 important azimuth blocks, R1 was sampled at 30.8%, R2 at 5% and R3 at 2.5% sampling rate. In the case of 5 azimuth blocks, R1 was sampled at 25.5%, R2 and R3 at 5% and 2.5% respectively. In the case of 6 azimuths, R1 at 20.2%, R2 at 5% and R3 at 2.5%. Therefore, these sampling rates were set manually to determine the importance of image data as prior information for adaptive CS.

In Algo-2, the radar data was split into bins of size 25×100 , amounting to 22.5° azimuth and 4.38m range each. There were 16 blocks in azimuth and 37 blocks in range. The block size of 25×100 instead of 50×100 from Algo-1 reduced the processing time by 60% for a 10% sampling rate. This empirical change in processing time was measured using MATLAB on a 3.1 GHz Dual-Core Intel Core i7 processor. The dynamic sampling rate for Algo-2 was determined based on the importance of a region and was allocated using LP. Where, a_1 corresponds to blocks with pedestrians or bicycles, a_2 corresponds to blocks with the car and a_3 represents all the other regions. r_1 is the radar

region within the first 18 range blocks (78.84m) and r_2 is the next 19 range blocks (83.22m). x_1, x_2, x_3, x_4 is the sampling rate for each block, which is being optimized. S is the total radar data and $0.1S$ corresponds to the 10% sampling rate. The objective function is the total sampling budget allocated across various azimuth blocks for a given radar frame and it is optimized such that, for a given frame, the sampling rate does not exceed 10%. Since x_1 is the sampling rate for the region with small objects, the constraints are set such that, it is thrice as big as the least important region's sampling rate and x_2 is twice as big. Also, x_4 is the sampling rate for the region starting 78.84m from the autonomous vehicle and hence it is optimized to have the least sampling rate of less than 2.5%.

For any vector $x \in \mathbb{R}^4$, let

$$f(x) = a_1 r_1 x_1 + a_2 r_1 x_2 + a_3 r_1 x_3 + (a_1 + a_2 + a_3) r_2 x_4$$

Then, we have the following linear program

$$\begin{aligned} \max_{x \geq 0} f(x) \\ \text{s.t. } x_1 - 3x_3 = 0, \quad x_2 - 2x_3 = 0 \\ f(x) \leq 0.1S, \quad 0.02 \leq x_4 \leq 0.025, \\ 0.05 \leq x_i \leq 0.4 \quad \text{for } i = 1, 2, 3 \end{aligned}$$

Therefore, the bounding boxes and classes from the object detection network are used to determine the azimuth blocks a_1, a_2, a_3 , given as a constant to the LP. The LP algorithm in one case, gave 37.39%, 24.93% 12.46% and 2.5% as x_1, x_2, x_3, x_4 sampling rates.

C. Measurement matrix analysis

We compare the standard Gaussian measurement matrix and BPBD matrix with the BPD matrix. In the case of our BPD matrix, it is designed such that all elements of a row are 0 except for one randomly selected column. Therefore, when each row of the measurement matrix is used to measure the original signal, only one value of the original data is retained, even the addition of a few measurements in the case of BPBD is avoided and addition, multiplication of values in the case of the Gaussian measurement matrix is eliminated. Therefore, complex multipliers for measurement using the Gaussian matrix which consumes higher power are avoided and replaced with simpler elements like switches and selectors for binary matrix making them hardware efficient [13].

D. DETR-Radar

The Faster R-CNN [10] object detection network is one of the well-refined techniques for 2D object detection. However, it relies heavily on two components, non-maximum suppression and anchor generation. The end-to-end transformer-based 2D object detection introduced in [17], eliminates the need for these components. We included radar data in two ways. In the first case, we included the radar data as an additional channel to the image data [1]. In the second case, we rendered the radar data on the image. We used perspective transformation based on [20] to transform radar data points from the vehicle coordinates system to camera-view coordinates. In all the

above cases, the models were pre-trained on the COCO dataset and we fine-tuned them on the NuScenes data. To the best of our knowledge, we are the first ones to implement end-to-end transformer-based 2D object detection using both image and radar data.

III. EXPERIMENT

TABLE I

S-CS DENOTES STANDARD CS, A-1 IS ALGORITHM 1 AND A-2 IS ALGORITHM 2. THE PRESENCE OF AN OBJECT IS INDICATED AS 'YES' AND IF THE OBJECT IS FAINT OR ABSENT, IT IS INDICATED AS 'NO'.

Scene	Frame	Object	S-CS	A-1	A-2
Scene1	5,11	car (top)	no	yes	yes
	1-4	Person (left)	no	yes	yes
	5	Person(left)	no	yes	no
	2-5,8,10	Person (top-left)	no	yes	yes
	6,7,9	Person (top-left)	no	no	yes
	8-11	Bicycle (top)	no	yes	yes
Scene2	2,3,5	Pole (right)	yes	no	yes
	1-11	Cars (top-right)	no	yes	yes
	2,3,6	Car (rear)	no	yes	yes
	7-9	Bicycle (rear)	no	yes	yes
	10,11	Bicycle (rear)	no	no	yes
	1-3	Person (rear)	no	yes	yes
	5-7,9	Person (rear-left)	no	yes	yes
	7-11	Person(front)	no	yes	yes
	7,8	Traffic light (rear)	yes	no	yes
	1	Pole (top-right)	yes	no	yes
Scene3	3,5,6	car (rear)	no	yes	yes
	7,8	car (top-right)	no	yes	yes
	4	Bicycle (right)	yes	no	yes
	1,8	Person (left)	yes	no	yes
	10	Person (rear-left)	no	yes	no
	3,5,10	Pole (right)	yes	no	yes

The algorithms were tested on three random scenes with 11 radar frames each [23]. Since the important information could be captured within the first 78.84m in an urban setting, the results were analyzed in that region and are shown in the figure 1 as well. The standard CS refers to the baseline CS algorithm, with a uniform sampling rate of 10% on the entire radar frame (163m, 360° azimuth). The Algo-1 is a basic version with 50x100 block size and manual sampling rate allocation and Algo-2 is the promoted algorithm with 25x100 block size and LP-based dynamic sampling rate allocation. All the reconstruction algorithms have BPD as the measurement matrix. In table I, the presence of an object is indicated as 'yes' and if it is absent or faint, it is indicated as 'no'. In the first scene, the person in the top-left was missed by both standard CS and poorly reconstructed by Algo-1 in frame 6,7 and 9. However, it was recovered by Algo-2 because it had a significantly higher sampling rate allocated to regions with pedestrians. The person to the left of the vehicle appeared in the blind spot of the camera and it was missed by our Algo-2. But, it was captured by Algo-1 because the block adjacent to it was selected and it was twice as big as the second algorithm's block size. Therefore, by using additional camera information in Algo-2, the person could be captured. However, Algo-2 in general has better or similar reconstruction across all the regions of r_1 compared to baseline and reconstructed the pole to the right in frames 2,3 and 5. In scene 2, the bicycle to the rear was missed by the baseline and poorly reconstructed by

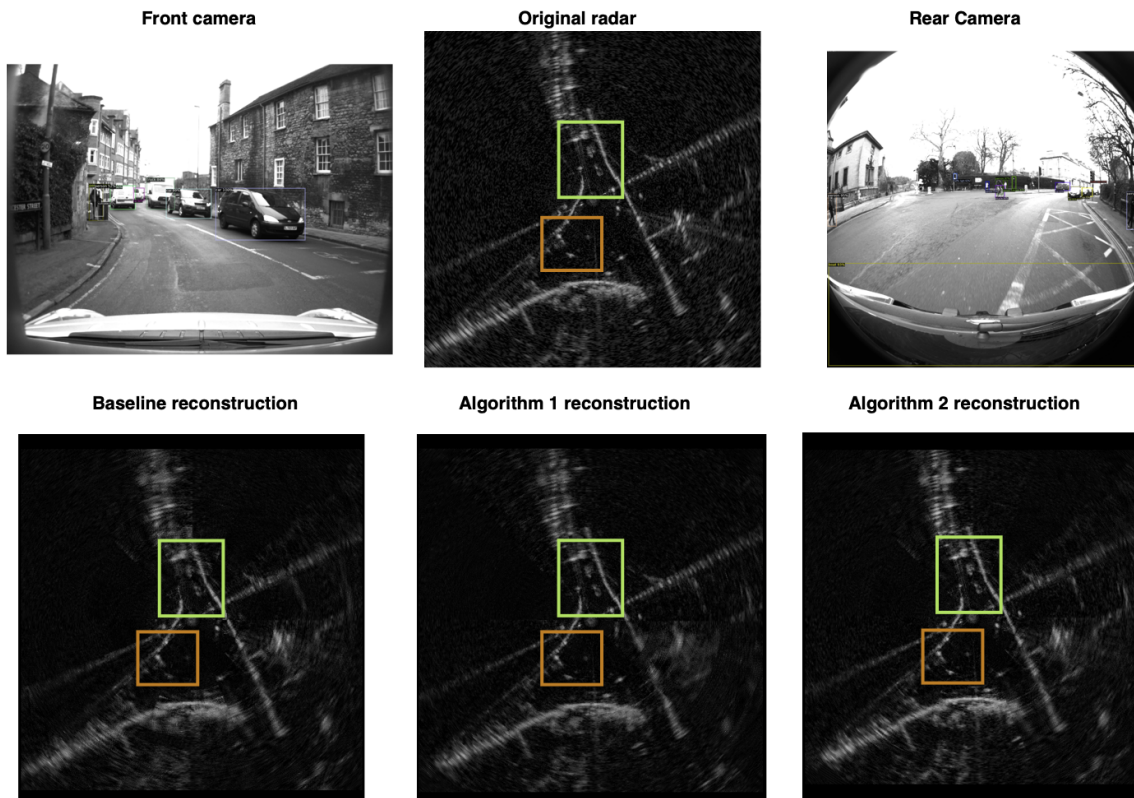


Fig. 1. In scene 2, frame 11, the front camera, rear camera, original radar and reconstructions are shown. The green box highlights the cars to the front and the pedestrian. The pedestrian was missed by the baseline reconstruction but, it is captured by Algorithm 1 and it is more clear in Algorithm 2. The orange box highlights the cars and bicycle to the rear. The bicycle was captured by Algorithm 2 but, it is hardly reconstructed by Algorithm 1 and baseline. *Best viewed as a digital copy by zooming in.*

Algo-1. However, it was reconstructed by Algo-2 in frames 10 and 11. The traffic light to the rear in frames 7, 8 and pole in frames 1 to the top-right was missed by Algo-1 but, it was captured by Algo-2. In scene 3 frame 4, the bicycle to the right was poorly reconstructed by Algo-1 but it was captured by Algo-2. The person to the left in frame 1 and 8 was missed by Algo-1 but it was captured by Algo-2. In Algo-1, apart from the important regions, the sampling rate was drastically reduced in other regions. The right and left of the vehicle is in the blind spot of the camera and would be classified as non-important regions. However, in Algo-2, the sampling rates saved from regions with bus or car were used to redistribute it across other regions and this helped in capturing the objects to the left and right. The person to the rear-left in frame 10 was missed by Algo-2 but it was captured by Algo-1. Although the person was captured by the object detection network, it was missed by our algorithm. Similar to the previous scenes, the pole to the right was captured by Algo-2 in frames 3,5 and 10 while it was missed by Algo-1. Also, from table I, considering yes as detected and no as not detected, Algo-1 detected 52 out of 60 objects across frames and scenes. The improved Algo-2, detected 58 out of 60 objects, improving from 86% detection using Algo-1 to 96% detection rate using our proposed Algo-2. Therefore, the Algo-2 has led to significant improvement in better reconstruction/detection.

In table II, the average peak signal-to-noise ratio (PSNR) is reported across three measurement matrices. In general, the BPD matrix has slightly lower PSNR compared to Gaussian measurement matrices. However, BPD is hardware efficient since it has binary elements as the measurement matrix.

TABLE II
THE THREE SCENES WERE RECONSTRUCTED USING THE STANDARD CS ALGORITHM AND AVERAGE PSNR IS REPORTED.

Sampling rate	Scene	Gaussian	BPBD	BPD
10%	scene1	29.88	29.91	29.87
	scene2	30.18	30.22	30.17
	scene3	29.89	29.93	29.88
20%	scene1	32.00	32.00	32.00
	scene2	32.26	32.26	32.26
	scene3	32.08	32.09	32.07
30%	scene1	34.30	34.30	34.27
	scene2	34.53	34.52	34.49
	scene3	34.43	34.42	34.38

Finally, we trained a separate object detection network using the NuScenes image and radar data. In this case, we limited our analysis to NuScenes image and original radar data because the Nuscenes radar that was available to us were processed pointclouds with annotations. Whereas, the Oxford data was the only available raw data on which we could apply CS but, without object annotations. Hence, we could not use Oxford data to train the object detection model. As shown in the table

TABLE III

I DENOTE MODEL TRAINED ON IMAGES, I+R INDICATED MODEL TRAINED WITH RADAR AS AN ADDITIONAL CHANNEL AND RONI IS FOR A MODEL TRAINED WITH THE RADAR RENDERED ON THE IMAGE. F-R IS THE FAST R-CNN NETWORK AND FA-R IS THE FASTER R-CNN NETWORK.

Network	AP	AP50	AP75	AR	ARs	ARm	ARI
F-R [20]	35.5	59.0	37.0	42.1	21.1	39.1	51.4
FA-R(I)	39.5	67.8	41.7	47.0	25.6	44.4	56.8
FA-R(I+R)	46.2	73.8	50.3	53.0	32.8	51.5	59.9
FA-R(RonI)	38.0	65.4	40.0	44.9	17.6	42.1	56.3
DETR(I)	47.1	80.2	50.4	61.6	38.4	57.2	72.5
DETR(RonI)	48.6	80.4	52.7	63.6	40.1	60.2	73.1
DETR(I+R)	44.8	76.3	46.8	58.2	29.7	54.9	68.8

III, our baseline comparison is with the [20] paper, where, they trained the model on Nuscenes v0.1 image dataset and used the radar data for anchor generation. The Faster R-CNN Img and Faster R-CNN RonImg models had ResNet-101 [25] as the backbone structure. The models with I+R were trained with radar as an additional channel. Therefore, the first layer of the backbone structure was changed to process the additional radar channel. The DETR network [17] had ResNet-50 [25] as the backbone structure. The I and RonI models were trained for the same number of epochs for a fair comparison. The I+R models were trained for additional epochs since the backbone structure's first layer was modified. In the Faster R-CNN case, I+R has better performance than I. While, in DETR, RonI has better performance. The Faster R-CNN I and RonI were trained for 25k iterations. The Faster R-CNN I+R was trained for 125k iterations. DETR I and DETR RonI models were trained for 160 epochs. While DETR I+R was trained for 166 epochs. The DETR RonI model performed better across various metrics compared to the baseline, Faster R-CNN and DETR I+R model. We believe that the attention heads in the transformer architecture helped in focusing object detection predictions around the radar points. However, the Faster R-CNN I+R was better than the Faster R-CNN RonI model. We used the standard evaluation metrics, mean average precision (AP), mean average recall (AR), average precision at 0.5, 0.75 IOU, small, medium and large AR [26].

IV. CONCLUSION

In this paper, we propose an adaptive block CS for radar data acquisition using an object detection result as prior information. Therefore, the main contribution of our work is to determine the areas with the object in the radar frame using prior image information while processing the data in a real-time scenario. The algorithm dynamically saves on sampling rate from big objects and redistributes to other regions to efficiently capture poles and other objects. This algorithm uses LP for dynamic sampling rate allocation and the reduction of block size helps in reducing the processing time significantly. Also, a hardware-efficient BPD measurement matrix is compared with a standard measurement matrix. Finally, our end-to-end transformer-based model trained on image and radar has better object detection performance than Faster R-CNN and transformer-based model trained on just images, validating the necessity for radar in addition to images.

REFERENCES

- [1] M. Meyer and G. Kusch, "Deep learning based 3d object detection for automotive radar and camera," in *2019 16th European Radar Conference (EuRAD)*, 2019, pp. 133–136.
- [2] Shuo Chang et al., "Spatial attention fusion for obstacle detection using mmwave radar and vision sensor," *Sensors (Basel, Switzerland)*, vol. 20, no. 4, February 2020.
- [3] Gledson Melotti et al., "Multimodal deep-learning for object recognition combining camera and lidar data," in *IEEE International Conference on Autonomous Robot Systems and Competitions*, 2020.
- [4] Z. Slavik et al., "Compressive sensing-based noise radar for automotive applications," in *2016 12th IEEE International Symposium on Electronics and Telecommunications (ISETC)*, 2016, pp. 17–20.
- [5] A. Correas-Serrano, "Experimental evaluation of compressive sensing for doa estimation in automotive radar," in *2018 19th International Radar Symposium (IRS)*, 2018, pp. 1–10.
- [6] F. Roos et al., "Data rate reduction for chirp-sequence based automotive radars using compressed sensing," in *2018 11th German Microwave Conference (GeMiC)*, 2018, pp. 347–350.
- [7] A. M. Assem, "Adaptive sub-nyquist sampling based on haar wavelet and compressive sensing in pulsed radar," in *2016 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, 2016, pp. 173–177.
- [8] J. Zhang, "Adaptive compressed sensing radar oriented toward cognitive detection in dynamic sparse target scene," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, 2012.
- [9] I. Kyriakides, "Adaptive compressive sensing and processing for radar tracking," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 3888–3891.
- [10] Kaiming He et al., "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv*, 2015.
- [11] Prannay Kaul et al., "Rss-net: Weakly-supervised multi-class semantic segmentation with fmcw radar," *arXiv*, 2020.
- [12] Z. He, "The simplest measurement matrix for compressed sensing of natural images," in *2010 IEEE International Conference on Image Processing*, 2010, pp. 4301–4304.
- [13] A. Akbari, "Robust image reconstruction for block-based compressed sensing using a binary measurement matrix," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018.
- [14] Xianxiang et al. Chen, "Compressed sensing based method for electrocardiogram monitoring on wireless body sensor using binary matrix," *International Journal of Wireless and Mobile Computing*, vol. 8, no. 2, pp. 114–121, 2015.
- [15] A. Ravelomanantsoa, "Compressed sensing: A simple deterministic measurement matrix and a fast recovery algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 12, 2015.
- [16] L. Miccinesi et al., "Compressive sensing for no-contact 3d ground penetrating radar," in *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, 2018, pp. 1–5.
- [17] Nicolas Carion et al., "End-to-end object detection with transformers," *arXiv*, 2020.
- [18] Holger Caesar et al., "nuscenes: A multimodal dataset for autonomous driving," *arXiv*, 2020.
- [19] Shuo Chang et al., "Spatial attention fusion for obstacle detection using mmwave radar and vision sensor," *Sensors (Basel, Switzerland)*, vol. 20, no. 4, February 2020.
- [20] R. Nabati and H. Qi, "Rrpn: Radar region proposal network for object detection in autonomous vehicles," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 3093–3097.
- [21] S. Chadwick, W. Maddern, and P. Newman, "Distant vehicle detection using radar and vision," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 8311–8317.
- [22] Felix Nobis et al., "A deep learning-based radar and camera sensor fusion architecture for object detection," in *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2019, pp. 1–7.
- [23] Dan Barnes et al., "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," *arXiv*, 2019.
- [24] Irfan Mehmood et al., "Adaptive compressive sensing of images using spatial entropy," *Computational Intelligence and Neuroscience*, 2017.
- [25] Kaiming He et al., "Deep residual learning for image recognition," *arXiv*, 2015.
- [26] Tsung-Yi Lin et al., "Microsoft coco: Common objects in context," *arXiv*, 2015.