End-to-end system for object detection from sub-sampled radar data

Madhumitha Sakthi, Ahmed Tewfik, Marius Arvinte, Haris Vikalo

***Abstract*—Robust and accurate sensing is of critical importance for advancing autonomous automotive systems. The need to acquire situational awareness in complex urban conditions using sensors such as radar has motivated research on power and latency-efficient signal acquisition methods. In this paper, we present an end-to-end signal processing pipeline, capable of operating in extreme weather conditions, that relies on sub-sampled radar data to perform object detection in vehicular settings. The results of the object detection are further utilized to sub-sample forthcoming radar data, which stands in contrast to prior work where the sub-sampling relies on image information. We show robust detection based on radar data reconstructed using 20% of samples under extreme weather conditions such as snow or fog, and on low-illuminated nights. Additionally, we generate 20% sampled radar data in a fine-tuning set and show 1.1% gain in AP50 across scenes and 3% AP50 gain in motorway condition.**

***Index Terms*—Deep learning, compressed sensing, object detection, radar.**

I. Introduction

A thorough understanding of surroundings is vital for the safety of autonomous driving systems. Similar to humans who while driving rely on multiple sensor information such as sound and vision, these systems acquire data from a variety of sensors including e.g. image, radar and LIDAR. Among those, radar was demonstrated to enable accurate object detection whether used in conjunction with other sensing modalities (e.g., images) [1]–[4] or alone [5]–[11]. The ability to achieve accurate object detection without relying upon data other than radar is critical in extreme weather conditions such as snow, fog or rain where the image sensors may struggle to provide the information needed to develop situational awareness [12].

When a sensor rapidly collects information about a vehicle's environment, it is necessary to reduce the data rate while maintaining the quality of acquired information. To this end, signal acquisition often relies on compressed sensing (CS) to collect data at a sub-Nyquist rate without compromising the quality of information [13]. The CS algorithms typically exhibit a trade-off between signal reconstruction quality and sampling rate. In [14], the CS framework is utilized to acquire Synthetic Aperture Radar (SAR) data and achieve robust reconstruction with 70% of the samples. In [15], authors demonstrate efficient reconstruction of a frequency-modulated continuous wave (FMCW) radar using 40% of the samples. A CS-based signal acquisition methodology for noise radar with a 30% sampling rate was presented in [16]. The authors of [17] compare the performance of Orthogonal Matching Pursuit (OMP) and Basis Pursuit De-noising (BPDN) in applications to the direction-of-arrival estimation problem; note that while OMP achieves more accurate reconstruction, it generally requires more measurements than basis pursuit [18]. This motivates the use of the basis pursuit (BP) algorithm in our current work.

Adaptive CS is a technique used to increase the sampling rate for important signal regions [19], [20], [21]. In [19], the authors present a pulsed radar undersampled acquisition method that utilizes the previously received pulse interval and applies a constant false alarm rate (CFAR) detection technique to determine the importance coefficients for the present interval. In [20], adaptive CS is utilized in static settings to improve target tracking performance. In contrast, in the current paper for an autonomous vehicle, while both the vehicle and objects are potentially moving, we use adaptive CS algorithm for radar acquisition. Finally, [21] presents an adaptive CS algorithm which aims to optimize the measurement matrix in a setting where only the targets are moving, leading to increased performance but, this also increased the computational complexity of the algorithm. In the proposed approach, the measurement matrix size is increased for certain regions of a radar frame using linear programming (LP) problem formulation; this allocates a larger sampling budget to important regions while keeping the overall sampling budget and reconstruction complexity under control.

In this paper, we develop an efficient signal acquisition/processing pipeline for radar-based object detection that uses the detection result from the current frame to sub-sample the subsequent radar frame via adaptive compressed sensing. In particular, we build upon [22] where the results of image-based object detection were used to identify the important regions in radar. To accommodate all weather conditions, the approach in the current paper removes the need for image data and instead relies on radar-based object detection results for the subsequent radar frame. We test our sub-sampled radar on an object detection task and show performance comparable to the fully-sampled radar. In particular, we test the method on the RADIATE dataset [23] collected in extreme weather conditions and demonstrate robust detection based on radar data acquired using only 20% of the samples. Finally, since task-based fine-tuning is known to improve the network performance, we generated 20% sampled radar data and used the sub-sampled radar frames to fine-tune the object detection

The authors are with the Department of Electrical and Computer Engineering, The University of Texas at Austin. E-mails: madhumithasakthi.iyer@utexas.edu, tewfik@austin.utexas.edu, arvinte@utexas.edu, hvikalo@ece.utexas.edu. code:https://github.com/Madhusakth/RADIATE-Adaptive-CS

Fig. 1. The overall sampling pipeline for every 20 frames across scenes.

network. This resulted in 1.1% gain in AP50 across all the scenes and 3% gain in AP50 in the motorway condition.

II. Method

A. Data

The RADIATE dataset [23] was collected in multiple extreme weather conditions, and consists of radar, lidar, camera, GPS sensor data. The radar data was collected by the Navtech CTS350-X with 360°Horizontal Field of View (HFoV) and 100-meter range at 4Hz, resulting in a range-azimuth image of size 400x576 where the rows represent the angle and the column represent the range. The set contains 300 hours of annotated radar data. In addition to the sensor data, the authors released object annotation on the Cartesian radar images. Although the released annotations are fine-grained, similar to the [23], we classify objects as vehicle or background by defining the vehicle class as either car, bus, bicycle, truck, van or motorbike. We selected about 20 frames for each night, snow, foggy, motorway and city scene conditions, and used them to test our acquisition algorithm. In the fine-tuning case, we generated 200 20% sampled radar data based on the previous frame's object detection result.

B. Adaptive radar sampling

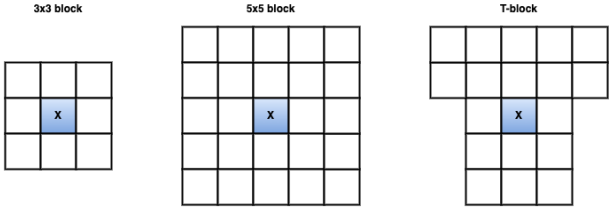

Fig. 2. The 3x3 sampling pattern for small objects, 5x5 sampling pattern for big object and T-shaped pattern for objects beyond 50m from the AV.

In adaptive CS, the measurement region is split into blocks and the number of measurements allocated to each block varies according to a parameter. In each of the blocks, standard CS is performed by collecting m measurements. To ensure robust reconstruction using the compressed measurements, the following assumptions are made. First, the measured signal is sparse in some domains. Specifically, we assume that the signal is sparse in the Discrete Cosine Transform (DCT) domain. Second, the measurement matrix exhibits the restricted isometry property [24]. In each block, given the original signal $x \in R^n$ acquired using the random measurement matrix $\phi \in R^{mxn}$, we obtain measurements $y \in R^m$. The original signal x is recovered using the basis pursuit (BP) algorithm as $\min_x \|\theta x\|_1 \; s.t. \; \phi x = y$ with θ as the DCT transformation matrix [25].

In the RADIATE dataset, the radar data is captured at 4Hz, i.e., every 0.25 seconds. The radar frame in the polar domain of size 400×576 is split into 20×48 sized blocks; this amounts to 18°and 8.4 meters in range. In the baseline acquisition algorithm (Standard-CS) that deploys a uniform sampling rate, all the blocks were acquired using the same sampling rate of 10%, 20% or 30%. That is, for each 20×48 sized block, in the case of 10% sampling rate, 96 samples were acquired using the measurement matrix and this information was used to reconstruct the original block using the BP algorithm. Similarly, for a radar frame of size 400×576, all 240 blocks were sampled and reconstructed to form the 10% uniform sampled radar data. In the proposed algorithm, as shown in Figure 1 the first frame is fully sampled and processed by the object detection network. However, the acquisition of radar data takes place in the polar domain. Specifically, in case of Navtech hardware, 400 measurements are acquired in each cycle, resulting in a 360°HFoV and 100 meters in range data. The polar data is converted to the Cartesian domain which leads to bird-eye view radar images. In [23], object annotations are provided in the Cartesian domain and thus the network is trained to perform object detection by processing this Cartesian radar data. Once the bounding boxes are predicted for a given frame, the object's center is converted to polar coordinates which are then used to identify the block in which the object is present. Since the next frame is acquired in 0.25 seconds, the region surrounding the current radar block is also marked important in order to account for the motion of the detected object as well as the motion of the Autonomous Vehicle (AV).

Each polar block covers about 8.4 meters in length; this is sufficient to capture the moving vehicle within the adjacent block while the object moves at most 80 miles per hour. Hence, in Rad-info-1, if the detected object is small – for example, a car or motorbike – then the 3×3 block sampling pattern shown in Figure 2 is chosen to identify important blocks. However, if the object is large – such as a truck or bus – the 5×5 block sampling pattern is selected instead. Since a larger object may span across two radar blocks (the length of a truck or a bus could be around 14 meters), it is necessary to cover two adjacent blocks around the vehicle. Also, while the radar data is converted from polar to Cartesian domain, the polar blocks closer to the autonomous vehicle will translate to a smaller area in the Cartesian domain while the more distant polar blocks will occupy a larger area in the Cartesian domain. In the case of Rad-Info-2, motivated by the above reasoning, a T-block sampling pattern is used, where only adjacent polar

blocks that are close to the autonomous vehicle are chosen while considering additional polar blocks which are far away from the vehicle. This ensures that a considerably larger area around the important object is sampled while preserving sampling rates from closer polar blocks that translate to a smaller area in the Cartesian domain. The T-block sampling pattern is used instead of the 5×5 block and we use the 3×3 block in case of small objects that appear within 50 meters range. In order to give importance to objects that appear very close to the AV, we also sample 3 adjacent blocks on either side of the predicted object while the object is within 16 meters of the AV in a given frame.

Once the number of important and other polar blocks are identified, we use LP with the constraints specified below to determine the sampling rate for the important and other blocks. For any vector $x \in \mathbb{R}^2$, let

$$f(x) = I.w.h.x_1 + O.w.h.x_2.$$

Then, we have the following linear program

$$\begin{aligned} &\max_{x \geq 0} f(x) \\ &s.t.\ x_1 >= 1.1x_2 \\ &\quad f(x) \leq S,\ x_1 l \leq x_1 \leq x_1 u, \\ &\quad x_2 l \leq x_2 \leq x_2 u. \end{aligned}$$

In the above LP, w is the width of 48 (range), h is the height of 20 (azimuth) of the block, I denotes the total number of important blocks, and O is the total number of other blocks. In total, since the 400×576 frame is split into 20×48, there are 240 blocks. x_1 is the sampling rate for important blocks and x_2 is the sampling rate for the other blocks. The $f(x) < S$ condition is used to limit the number of samples – e.g., to 10% or 20% or 30% of the total samples (400x576). The condition $x_1 >= 1.1x_2$ aids in ensuring that the sampling rate for the important regions is higher than for the other regions. Finally, the lower bound for x_1 is 0.1, 0.2, and 0.3 for the three sampling rates 10%, 20%, and 30%, respectively, while the upper bound is 0.55. The lower bounds were chosen to ensure that the sampling rate is at least as in the standard-CS case, while the upper bound was determined such that the reconstruction matches that of the original. In case of x_2, the lower bound was set to 0.07 to ensure there are enough samples to support reconstruction, and the upper bounds were set to 0.1, 0.2 and 0.3 for 10%, 20% and 30% sampling rates, respectively, since the number of measurements could be limited due to lack of object of interest in the other regions. Once the sampling rates are determined by solving the LP, they can be used for the subsequent radar frame and the reconstructed radar is used anew for object detection; this, in turn, is further used for important region determination, and the loop continues for 20 frames. Therefore, in case of 10% sampling rate, S is set to 23040 (10% of 400×576) and these measurements are adaptively allocated across the important and other regions of the radar frame based on the LP results while maintaining the overall sampling budget to be within 10% or 23040 measurements for 10% sampled radar data.

C. RAD-Net

The network, proposed in [23], takes a single Cartesian radar frame as input and predicts bounding boxes and classes. The RAD-Net is a modification of the Faster-RCNN [26] network with pre-defined anchor sizes and a single class prediction head. The network is trained on both good and bad weather condition data using ResNet-50 [27] as a backbone. We generate the fine-tuning dataset using a pipeline similar to the one proposed in Figure 1. First, the fully sampled radar data is used as the first frame to predict important radar regions for the second frame; that information is then utilized to sub-sample the second radar frame. For the third radar frame, we rely on the original second radar frame to predict bounding boxes and important blocks, and use this information to reconstruct the frame. Proceeding in this way, 200 images were generated as the fine-tuning set.

III. Experimental Results

We tested the above algorithm on 5 scenes: city, motorway, night and snow with 20 frames each, and foggy with 18 frames. We report the standard AP50 metric [23], the average precision at IoU of 0.5 and AP, the average precision calculated and averaged across IoU ranging from 0.50 to 0.95. The standard-CS resulted in a poor reconstruction quality and hence low AP50 of 6.3 while using 10% of the samples, 34.6 for 20% of the samples, and 47.6 for 30% of the samples. Additional reconstruction experiments that relied on 40% of the samples in the standard-CS setting led to 54.5 AP50 and 21.5 AP. Moreover, using 50% of the samples yielded 56.3 AP50 and 23.1 AP, while using 55% of the samples yielded 56.7 AP50 and 23.3 AP. Given these performances, we chose 55% as the upper bound sampling rate for our algorithm. Across all the sampling rates shown in Table I, our algorithm consistently outperforms the standard-CS algorithm with at least 10% AP50 difference in case of 20% and 30% data sampling rates. For Radar-info-2, the T-block sampling pattern and assigning importance to the region closer to the AV yielded 54.2 AP50 and 22.5 AP in case of 20% sampling rate, outperforming Radar-info-1 which achieved 47.9 AP50 and 20.9 AP. We conjecture that by saving on sampling rate in the regions adjacent to the object and thus reducing the total number of important blocks helps maintain a higher sampling rate at the remaining important blocks, leading to an improved reconstruction quality and object detection. This emphasises the need to precisely identify the region wherein an object of interest is present, and allocate most of the sampling resources to that region to ensure accurate reconstruction. In the case of the 30% sampling rate, the reconstruction using Rad-info-2 achieves AP almost as same as that of the scheme using all radar data. In Figure 3, we show detection results across the baseline and our proposed algorithm on a frame from fog and city conditions.

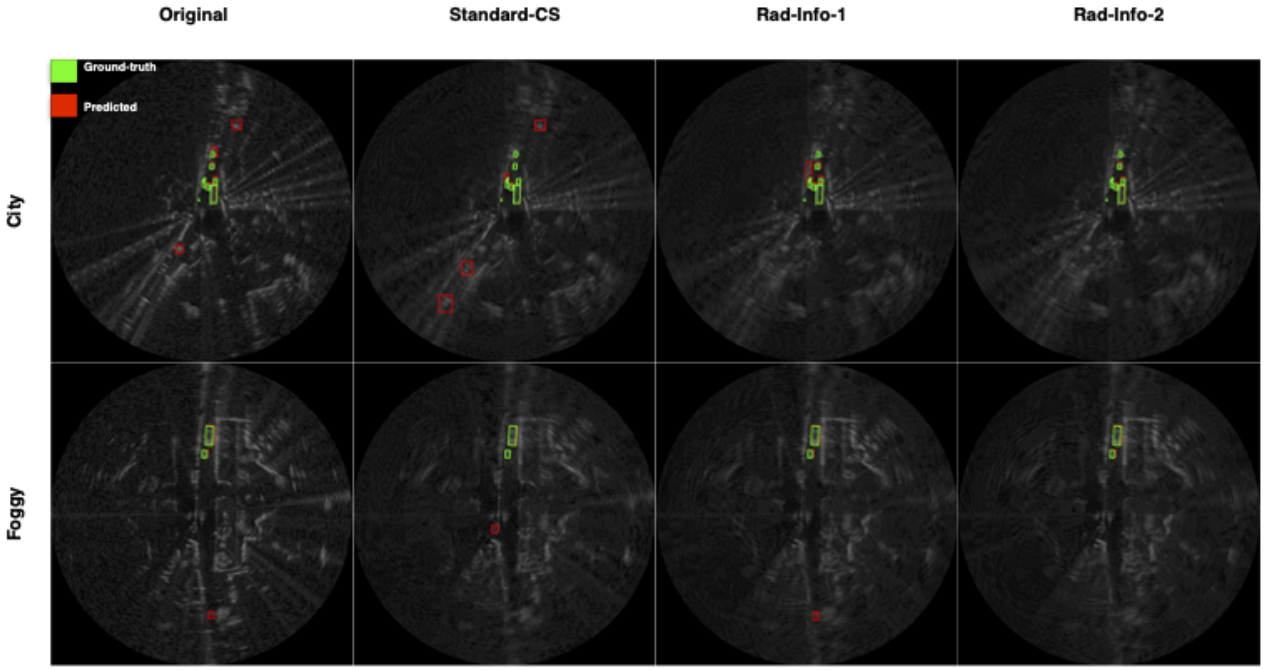

Fig. 3. Object detection on original radar, standard-CS, and on data sampled at 20% rate and reconstructed using Rad-Info-1 and Rad-info-2.

TABLE I
RADAR RECONSTRUCTION RESULTS [AP/AP50]. ORIGINAL RADAR AP50: 60.6 AND AP: 23.4

Sampling rate	10%	20%	30%
Standard-CS	6.3/2.4	34.5/13.1	47.6/18.9
Radar Info - 1	29.3/12.0	47.9/20.9	56.7/23.3
Radar-Info - 2	34.6/14.9	54.2/22.5	57.6/23.8

TABLE II
FINE-TUNING RESULTS [AP/AP50] FOR 20% SAMPLED RADAR DATA. FT: FINE-TUNING

Train-set	Overall	city	foggy	snow	night	motorway
Before FT	54.2 22.5	54.8 22.7	72.8 46.0	58.9 28.5	85.0 47.6	39.9 11.4
Original	53.5 22.2	54.8 22.7	72.7 46.0	58.5 27.4	81.2 46.6	39.8 11.4
20% radar	55.3 22.3	55.8 23.0	72.4 46.4	57.9 26.2	85.1 46.5	42.9 11.9

Since only the first radar frame is fully sampled, we also analyse the AP/AP50 results averaged across different scenes for each frame to ensure no propagation error has affected the reconstruction quality of the subsequent frames. Figure 4 shows AP and AP50 for each frame on the original, 10%, 20%, and 30% sampled radar frames. The first frames yield the same AP since they are fully sampled in all cases. In the subsequent frames, at 20% and 30% sampling rates, the AP curve follows the performance of the original radar AP curve until the 20th frame. However, AP performance for 10% sampled radar data deteriorates, especially from the 7th to the 11 th frame. Similarly, in the case of 20% and 30% sampled radar frames, AP50 closely follows AP50 of the full radar data scheme, i.e., there is no evidence of propagation error.

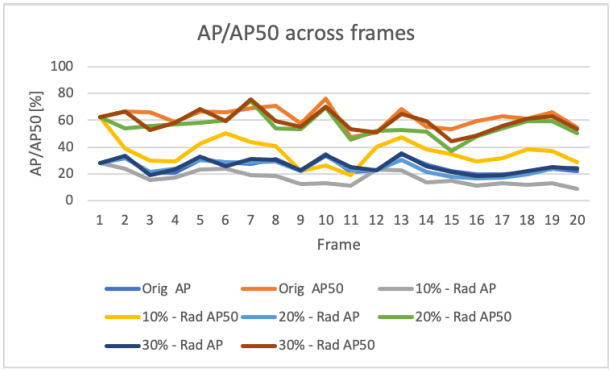

Fig. 4. AP/AP50 evaluated per frame across scenes.

Finally, we fine-tune the network on the 20% sampled reconstructed radar data. We hypothesize that tuning the network to predict objects on the sub-sampled radar data should aid in improving detection accuracy since the network learns to detect objects better in regions sampled at a higher rate than in other sub-sampled regions. The object detection results on the fine-tuned network are shown in Table II. In the first row, we show the detection results of the 20% sampled radar data reconstructed using the Rad-Info-2 algorithm. Without fine-tuning, the achieved AP50 and AP were 54.2 and 22.5, respectively. We use 200 images for fine-tuning, set the learning rate to 10^{-6}, and fine-tune for 100 iterations, with a batch size of 2. This results in an overall AP50 of 55.3 and AP of 22.3. Although there is a slight drop in AP, the overall AP50 increases by more than 1%; in the case of the motorway, AP50 increases by 3% while AP also slightly improves. As

an ablation study, we also fine-tune the network with the fully sampled 200 radar frames to show the benefit of fine-tuning using the 20% sampled radar data. Fine-tuning with the original radar data decreases the overall AP50 to 53.5 and AP to 22.2, indicating rapid overfitting. Therefore, in the case of a populated environment with many objects such as a city or motorway, where the sampling budget is shared across many objects and a wider region in the radar frame, fine-tuning with sub-sampled data improves the detection performance.

IV. Conclusion

In this paper, we present an end-to-end radar acquisition and radar-based object detection pipeline. The sub-sampled radar data is processed to identify important objects under multiple extreme weather conditions such as snow and fog; the location/region of the important object is used as prior knowledge for adaptive CS in the subsequent radar frame. We tested our algorithm on the RADIATE dataset across 5 scene conditions with about 20 frames each. In our experiments, the reconstruction based on data sampled at 20% rate shows an overall AP50 of 54.2 and AP of 22.5 detection performance, while the original data yields AP50 of 60.6 and AP of 23.4. By increasing the sampling rate to 30% our method results in AP50 of 57.6 and AP of 23.8, i.e., it achieves the original AP while AP50 that is lower than the original by 3%. When using the 20% sampling rate, the achieved AP is 1% lower than the original while AP50 is 6% lower. However, across all sampling rates, the proposed algorithm consistently outperforms the standard-CS baseline. Finally, we generated 200 radar frames sampled at 20% and fine-tuned the object detection system, demonstrating 1.1% improvement in overall AP50 and 3% AP50 gain in case of motorway scene.

References

[1] Ramin Nabati and Hairong Qi, "Centerfusion: Center-based radar and camera fusion for 3d object detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2021, pp. 1527–1536.

[2] M. Meyer and G. Kuschk, "Deep learning based 3d object detection for automotive radar and camera," in *2019 16th European Radar Conference (EuRAD)*, 2019, pp. 133–136.

[3] Shuo Chang, Yifan Zhang, Fan Zhang, Xiaotong Zhao, Sai Huang, Zhiyong Feng, and Zhiqing Wei, "Spatial attention fusion for obstacle detection using mmwave radar and vision sensor," *Sensors*, vol. 20, no. 4, 2020.

[4] Ramin Nabati and Hairong 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.

[5] Yizhou Wang, Zhongyu Jiang, Yudong Li, Jenq-Neng Hwang, Guanbin Xing, and Hui Liu, "Rodnet: A real-time radar object detection network cross-supervised by camera-radar fused object 3d localization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 954–967, 2021.

[6] Ana Stroescu, Liam Daniel, and Marina Gashinova, "Combined object detection and tracking on high resolution radar imagery for autonomous driving using deep neural networks and particle filters," in *2020 IEEE Radar Conference (RadarConf20)*, 2020, pp. 1–6.

[7] Chih-Chung Hsu, Chieh Lee, Lin Chen, Min-Kai Hung, Yu-Lun Lin, and Xian-Yu Wang, "Efficient-rod: Efficient radar object detection based on densely connected residual network," in *Proceedings of the 2021 International Conference on Multimedia Retrieval*, New York, NY, USA, 2021, ICMR '21, p. 526–532, Association for Computing Machinery.

[8] Pengliang Sun, Xuetong Niu, Pengfei Sun, and Kele Xu, "Squeeze-and-excitation network-based radar object detection with weighted location fusion," New York, NY, USA, 2021, ICMR '21, p. 545–552, Association for Computing Machinery.

[9] Shoaib Azam, Farzeen Munir, and Moongu Jeon, "Channel boosting feature ensemble for radar-based object detection," in *2021 IEEE Intelligent Vehicles Symposium (IV)*, 2021, pp. 762–769.

[10] Bo Ju, Wei Yang, Jinrang Jia, Xiaoqing Ye, Qu Chen, Xiao Tan, Hao Sun, Yifeng Shi, and Errui Ding, "Danet: Dimension apart network for radar object detection," in *Proceedings of the 2021 International Conference on Multimedia Retrieval*, New York, NY, USA, 2021, ICMR '21, p. 533–539, Association for Computing Machinery.

[11] Ana Stroescu, Liam Daniel, Dominic Phippen, Mikhail Cherniakov, and Marina Gashinova, "Object detection on radar imagery for autonomous driving using deep neural networks," in *2020 17th European Radar Conference (EuRAD)*, 2021, pp. 120–123.

[12] Felix Nobis, Maximilian Geisslinger, Markus Weber, Johannes Betz, and Markus Lienkamp, "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.

[13] Emmanuel J Candès and Michael B Wakin, "An introduction to compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[14] Seongwook Lee, Yunho Jung, Myeongjin Lee, and Wookyung Lee, "Compressive sensing-based sar image reconstruction from sparse radar sensor data acquisition in automotive fmcw radar system," *Sensors*, vol. 21, no. 21, 2021.

[15] Fabian Roos, Philipp Hügler, Christina Knill, Nils Appenrodt, Jürgen Dickmann, and Christian Waldschmidt, "Data rate reduction for chirp-sequence based automotive radars using compressed sensing," in *2018 11th German Microwave Conference (GeMiC)*, 2018, pp. 347–350.

[16] Z. Slavik, A. Viehl, T. Greiner, O. Bringmann, and W. Rosenstiel, "Compressive sensing-based noise radar for automotive applications," in *2016 12th IEEE International Symposium on Electronics and Telecommunications (ISETC)*, 2016, pp. 17–20.

[17] A. Correas-Serrano and M. A. González-Huici, "Experimental evaluation of compressive sensing for doa estimation in automotive radar," in *2018 19th International Radar Symposium (IRS)*, 2018, pp. 1–10.

[18] S. K. Sahoo and A. Makur, "Signal recovery from random measurements via extended orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2572–2581, 2015.

[19] A. M. Assem, R. M. Dansereau, and F. M. Ahmed, "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.

[20] 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.

[21] J. Zhang, D. Zhu, and G. Zhang, "Adaptive compressed sensing radar oriented toward cognitive detection in dynamic sparse target scene," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, 2012.

[22] Madhumitha Sakthi and Ahmed Tewfik, "Automotive radar data acquisition using object detection," in *2021 29th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1770–1774.

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

[24] Ali Akbari and Maria Trocan, "Robust image reconstruction for block-based compressed sensing using a binary measurement matrix," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 1832–1836.

[25] Irfan Mehmood, Ran Li, Xiaomeng Duan, Xiaoli Guo, Wei He, and Yongfeng Lv, "Adaptive compressive sensing of images using spatial entropy," *Computational Intelligence and Neuroscience)*, 2017.

[26] Kaiming He et al., "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv*, 2015.

[27] Kaiming He et al., "Deep residual learning for image recognition," *arXiv*, 2015.