Next-Level Enhanced Collaborative Scene Understanding

Ehsan Moradi-Pari [1] **Hossein Nourkhiz Mahjoub** [1] **Ryoji Igarashi** [1]

1) Honda Research Institute, US Inc.

E-mail: emoradipari@honda-ri.com hnourkhizmahjoub@honda-ri.com rigarashi@honda-ri.com

ABSTRACT: Accurate *"Situational Awareness"* is a key component for reliable decision-making in autonomous driving. Relying on only onboard sensor suites of individual AVs is not sufficient for this purpose, due to its inherent restrictions, like limited detection range and non-line-of-sight limitations. Therefore, to have a safe AV design having a framework that facilitates the collaboration among multiple connected AVs (CAVs) and jointly build up their scene understanding seems to be essential. In this work, we propose a novel framework by utilizing the capabilities of advanced 5G communication and edge-computing technologies. Additionally, our framework is capable of risk assessment and quantification that enables a clear characterization of the collaboration among CAVs and its impact on collision risk reduction and uncertainty. We further investigate the effectiveness of the proposed approach by utilizing real world data and postprocessing analysis.

KEY WORDS: Collaborative Sensing, Scene understanding, Risk Estimation, Wireless Technology, Situational Awareness, and 5G-Mobile Edge Computing (MEC)

1. INTRODUCTION

Higher levels of autonomy in the context of ground vehicles are expected to gradually enhance the safety and comfort of drivers and passengers. It is well understood from various studies in the human factor domain that lowering the amount of human intervention and managing task sharing between humans and the automated systems result in a reduction of imposed cognitive workload on humans hence improved safety. To realize this, an automated agent first needs to have a reliable and accurate understanding of its surroundings and be able to precisely predict the scene evolution, due to the future actions of the effective actors, and then utilize this *situational awareness* to design its automated actions. Historically, the dominant approach in the leading automated vehicle (AV) research and industrial communities to build up and continuously update this situational awareness, which is essential for automated decision-making, has been based on designing sophisticated sensor suites to be installed on each autonomous vehicle. Although fused sensory information coming from cameras, lidars, and radars can cover the requirements to a good extent, this approach has some intrinsic limitations which decreases the accuracy of prediction and decision-making modules. These limitations are mostly due to the limited sensing range of onboard sensors, sensor blindness, and their non-line-of-sight restrictions. Therefore, only relying on sensory information of individual AVs cannot cover all critical situations and therefore, this information needs to be augmented with other additional information streams especially in urban areas in which obstructed views are likely to happen frequently. To address these challenges, the concept of *collaborative scene understanding* has been proposed and investigated as a promising alternative framework. In this framework the intelligent vehicles, which are referred to as CAVs (Connected and Autonomous Vehicles), can communicate their information, coordinate their maneuvers with each other, and fuse their on-board sensory data with the information they receive from the other CAVs' through communication media. This will eventually result in a more comprehensive and accurate scene understanding. Any stream of information that comes from other CAVs can reveal some parts of the scene that might be hidden from the recipient if it only utilizes its onboard sensors.

However, the collaborative scene understanding has its own challenges that need to be addressed. This paper is devoted to two major challenging aspects of this solution. First, we need to have a consistent framework to be able to measure the benefits of collaboration in terms of risk and uncertainty reduction. The second point is related to the communication aspect of the problem and tries to answer the question that how the information exchange procedure among a group of CAVs could be facilitated by utilizing a meaningful combination of communication and edge-computation technologies. Due to potential high information demands of collaborative scene understanding methods, this is an extremely important and challenging point that needs to be investigated.

In this paper, we briefly explain our solutions for each of these two aspects of the problem and try to clarify our system-level

perspective for a realizable collaborative scene understanding framework. Some parts of the proposed framework have been individually investigated in the past by our team at Honda Research Institute, US, and here we leverage those findings and lessons we have learned to propose a cohesive solution. Fig. *1* illustrates a high-level schematic overview of collaborative sensing and risk mapping based on the 5G network. As it is shown in this figure, each CAV shares its sensory information through a 5G network. Smart infrastructure could also share its sensory information in addition to its signal phase and timing.

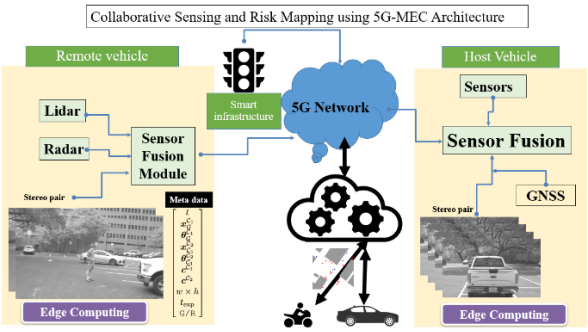

Fig. 1 System Architecture Schematic.

2. COLLABORATIVE SENSING AND RISK ESTIMATION

AVs are expected to have a precise understanding of their surroundings and be able to accurately estimate the risk of collision and continuously update these estimates, especially in challenging dynamic driving environments, to be able to surpass human drivers' capabilities and provide higher levels of safety and comfort. The notion of information exchange among CAVs, which is also referred to as Collaborative Sensing, could be a promising approach to reduce both collision risk and also risk uncertainty. This section provides a mathematical framework to assess and quantify these aspects of collaborative sensing schemes.

Two main dominant risk assessment methods in the literature are based on either the worst-case assumption of full occupancy on the hidden zones [1], or the idealistic assumption of exact identification and tracking of all other actors in the scene [2, 3]. However, our method for risk assessment does not purely follow any of these approaches. Instead, we aim to come up with a more accurate risk estimation in the hidden zones, using a priori probabilities coming from the sensing history of the CAV's sensor suite and what it is receiving from other CAVs. Our framework also provides the probability distribution of the estimated risk, which enables us to evaluate the risk uncertainty. This will further formalize the quantitative evaluation of the collaborative sensing impact on risk uncertainty reduction. To give the reader an idea about our method, our risk representation could be formulated as follows (Eq. 1):

$$R_k = \sum_c p_{occ}(c,k).p_{ego}(c,k).L(\mathbf{X}_{ego},\mathbf{X}_c) \quad (1)$$

Here, we are dividing the environment into multiple cells and c is the cell index, $p_{occ}(c,k)$ is the occupancy probability of cell c at time step k, $p_{ego}(c,k)$ is the probability that ego car occupies cell c at epoch k, and $L(\mathbf{X}_{ego},\mathbf{X}_c)$ is the loss function, while $\mathbf{X}_{ego}$ and $\mathbf{X}_c$ represent the ego vehicle and cell states, respectively (each one is a four-dimensional random vector, modeling two-dimensional position and two-dimensional speed distributions). In this formulation, we assume different cells to be independent in terms of their occupancy probabilities. It is also important to emphasize that Eq. 1 is based on a commonly accepted definition of risk which is equivalent to the expected loss. We are defining our loss function based on kinetic energy. More specifically, the loss we are considering for cell c due to collision could be formulated as:

$$L(\mathbf{X}_{ego},\mathbf{X}_c) = C_1\|\boldsymbol{v}_{ego} - \overline{\boldsymbol{v}_c}\|^2 + C_2E[(\boldsymbol{v}_c - \overline{\boldsymbol{v}}_c)^2]$$

where $\boldsymbol{v}_{ego}$ and $\overline{\boldsymbol{v}_c}$ are the expected velocity of the ego vehicle and the weighted mean of particle velocities in cell c, respectively. Also, $\boldsymbol{v}_c$ in the second term, is the random variable of velocity in cell c, which is the state of this cell $(\boldsymbol{X}_c)$. The constants

C_1 and C_2 are design choices and in our case, they should be formed based on kinetic energy. So:

$C_1 = m_{ego}mc/2(m_{ego} + m_c)$ and $C_2 = m_c/2$

It is also noteworthy that the cell loss has to be normalized with respect to the discretization time interval and also cell area to make sure that the calculated risk isn't sensitive to the discretization parameters choice. Following explanations also help to better clarify this formulation.

In our framework, we inherited an object-free approach representation scheme. In this method, unlike the object-based method that assumes a specific class and appropriate category of motion model for each traffic actor, there is no need to determine the motion-model class for each actor, but instead, we need to estimate the cells occupancy in a grid map for short future time horizons. In this framework, an acceptable (safe) navigation exists only if we can find at least one trajectory passing through only unoccupied cells [4]. It is noteworthy that although we utilized the object-free scheme in our modeling, it is not a restrictive assumption and our method could be further expanded to object-based representations, too. The mathematical tool we employ to model the occupancy of the environment is called the Probabilistic

Occupancy Map (POM), which, as we mentioned before, discretizes the map into a grid of cells and is capable of modeling the fused data received from multiple information sources. Note that in our framework this notion could be translated as if the sensory information is coming from multiple collaborating CAVs. For each cell in the grid, the occupancy is modeled via a binary Bernoulli random variable. Therefore, each cell is assumed to either be occupied with probability p or empty with probability $(1-p)$. POMs in general could be applied to dynamic environments. In this case, at each snapshot, they represent the instantaneous estimated occupancy status of each cell, conditioned on the measurement history. Bayesian Occupancy Filter (BOF) is then utilized to calculate the occupancy maps, which could be simpler than other methods, such as multi-target tracking. BOFs do not try to explicitly associate each sensor measurement to a specific detected object, but instead, they use measurements for updating the occupancy probabilities of cells. It should be noted here that BOFs assume that the grid cells are statistically independent and although this assumption might slightly reduce the accuracy, but its advantages, such as allowing the analytical representation and parallel updates of Bayesian update equations, can justify this accuracy reduction. In order to update the occupancy grid maps, a variety of particle filter tracking solutions are utilized based on the method proposed by Nuss, *et al* in [4], *i.e.,* Probability Hypothesis Density/Multi-Instance Bernoulli (PHD/MIB) filter, which connects the dynamic grid cells occupancy state estimation problem to the well-established notion of finite-set statistics.

Fig. 2 Equipped Honda Vehicles for Data Collection [5].

By explaining our techniques for collision risk evaluation and risk uncertainty modeling, which are capable of incorporating multiple data streams potentially coming from different collaborative CAVs, it becomes clear that how collaborative sensing reduces the risk uncertainty. For more details on our mathematical framework one can refer to our previous works such as [5]. We prototyped this claim by data collected using two equipped vehicles (Fig. *2*) in an example scenario (Fig. *3*).

In this sample scenario the ego vehicle sensor suite is not able to directly detect the pedestrian due to the occlusion caused by the bus which is next to it. On the other hand, another vehicle (called collaborator vehicle) which is traveling on the other direction has much better view of the intersection area and is able to detect and track the pedestrian most of the time. Fig. *3* shows a birds-eye view snapshot from this scenario. Also, Fig. *4* shows two snapshots taken at the same moment, from ego vehicle and collaborator vehicle perspectives.

Now, we consider two cases to compare in order to show the advantage of collaborative sensing concept. First, when ego vehicle can only access its own sensor suite information, and second when it can also benefit from the collaborator vehicle sensor information. A snapshot of the dynamic occupancy maps that could be built overtime using the sensor information of ego and collaborator vehicles in our collaborative sensing structure is shown in Fig. *5*. Also, Fig. *6* depicts the actual and predicted accumulated risks under two different cases, i.e., with and without using the collaborative vehicle information. The risk and its uncertainty are clearly less under collaborative sensing set up.

Fig. 3 Assessment of Collaborative Sensing in an Intersection Scenario [5].

Fig. 4 Scenario from ego vehicle and Collaborator vehicle points of view

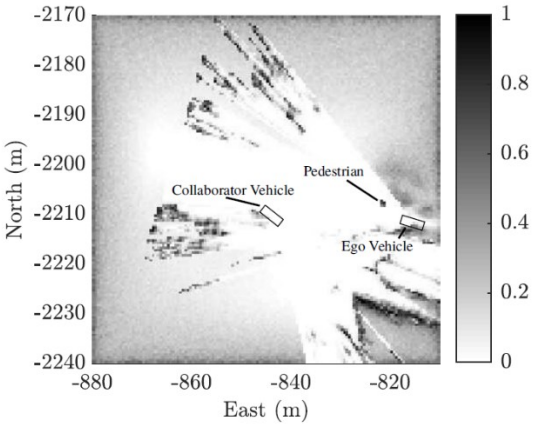

Fig. 5 Snapshot of dynamic occupancy map built with sensor information of both vehicles [5].

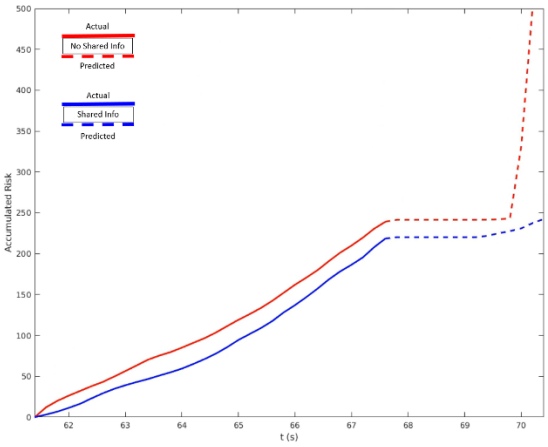

Fig. 6 Accumulated risk (actual and predicted) with and without information sharing between ego and collaborative vehicles

3. COLLABORATIVE SENSING AND 5G-MEC ARCHITECTURE

As mentioned in the previous sections, in a collaborative sensing setup the sensory information is shared among a group of vehicles, either directly or via a smart middle layer infrastructure which potentially could also have computing capabilities. This shared multi-source information stream enables the vehicles and infrastructure to build up a more accurate estimate of the scene, especially, in dynamic scenes and for the zones that are in the blind spots of the sensor suites of individual vehicles. This more accurate scene understanding in turn facilitates the implementation of more advanced adaptive and cooperative applications, such as adaptive platooning [6], remote maneuvering, and cloud-based fully automated driving. However, the challenge here is that sharing the raw sensor information is a high-demand solution in terms of both required communication resources (such as bandwidth and latency) and, also computational power, especially in dense multi-agent scenes.

There are several resource management strategies and solutions introduced in the literature (*e.g.,* [7], [8]). [7] proposes novel solutions for communication resource demand reduction while [8] focuses on efficient methods of computation resource allocation, for instance by preprocessing the sensory data at the transmitter and then sharing this processed information (*e.g.,* the final output of the perception stack). Also, as an alternative way to approach the problem of collaborative scene understanding, infrastructure technologies have been investigated. This concept relies on smart infrastructures with sensing and communication capabilities, instead of sharing only the vehicles' onboard sensory data. [9] introduces an early deployment solution to improve situational awareness and information exchange among road users and discusses the technical challenges associated with this concept and provides solutions to tackle these challenges. The technological solution introduced in this paper, however, is mainly based on utilizing capable (high bandwidth and low latency) communication technologies, such as 5G, along with integrating the edge computing solutions in the framework, to be able to address the requirements of these high demand use cases.

Currently, 5G is the most advanced deployed communication technology which offers the potential to address the demanding requirements of the emerging use cases for the collaborative scene understanding framework [10]. During the last twenty years, Wi-Fi-based (IEEE 802.11p) and cellular-based (C-V2X) RATs (Radio Access Technologies) have been two main enablers of V2X communications in the US, each with its advantages and disadvantages. 5G NR-V2X, which has been proposed by 3GPP as part of its Release 16 specifications, could be categorized as part of the C-V2X framework and is the rational evolution of its predecessors in the family of vehicular communications technologies. However, it has some essential uniqueness compared to previous C-V2X technologies such as a) enhanced Mobile Broadband (eMBB), b) massive Machine Type Communications (mMTC), and c) Ultra-Reliable and Low-Latency Communications (URLLC). The last item in this list is the core enabler of the advanced high-demand V2X use cases, but nevertheless, the other two factors are also very critical for this technology. Previous variants of C-V2X have not been considered seriously for automotive safety-critical use cases mostly due to the high bandwidth, reliability, and latency requirements. However, 5G technology, especially in the C-Band, potentially can address the bandwidth, and reliability demands, and also to some extent provide reasonable latency, and therefore, we see it as a feasible option to be included in our framework.

In addition to 5G as our choice for the radio-level technique, the framework could also be further improved by adding advanced network-level technologies such as Edge-computing techniques. Mobile Edge Computing (MEC), which is the evolutionary variant of Mobile Cloud Computing (MCC) technology, offloads the computational demands from the vehicles to the edge devices. Therefore, it is a promising technology to enhance the overall performance of the 5G-enabled application and also decrease the in-vehicle resource requirements, which in turn would potentially lead to a notable cost reduction in scale, an attractive element for AV manufacturers and an important consideration for the AD generalization roadmap.

In our framework, we have tried two different configurations in terms of the data flow between the vehicles and the MEC. In the first configuration, the computation burden is shared between the vehicles and the MEC to further reduce the uplink communication latency. More specifically, each CAV tries to conduct some post-processing to detect the non-CAV objects (including vehicles, motorcycles, and pedestrians) that are being captured by its sensor suite and then shares these detected objects with the MEC. MEC receives this information from different CAVs, runs its risk estimation logics, and then shares the outputs, in the form of direct warnings and/or more abstract pieces of information such as risk estimations, with CAVs. This configuration reduces the uplink latency but needs more computational power on the CAV side. In the second configuration, CAVs directly share their raw sensory information, *e.g.,* lidar point clouds with the MEC, MEC then runs the post-processing for object detection and afterward runs its risk estimation logic, similar to the previous configuration. Obviously, this option adds more latency to the uplink but reduces the required computational capabilities for CAVs. Fig. 7 shows a schematic of these two configurations, along with the latency results we observed in our tests for these two configurations. Following paragraphs provide more details on the hardware and software parts of our experiments.

In terms of the hardware, we used a Jackal Autonomous Mobile Robot (AMR) platform, developed by Clearpath Robotics in our tests. This Jackal robot was representing the vehicle. It should be mentioned here that since this Jackal platform had the identical on-board sensor-suite as a normal test car, replacing the test car with this robot did not affect the validity of our experiments. The sensor suite installed on the Jackal included a 5G UW modem (Telit FT980m) in addition to a top-mounted 360-degree 3D LiDAR (Velodyne HDL-32e). In order to simulate the other vehicle, we utilized a high performance gaming laptop, equipped with a 5G UW modem (Inseego MiFi M2100), as a part of the collaborative perception environment. This laptop was also used to visualize the incoming streams from the AMR and MEC.

The software set up to facilitate the stream of sensor information between the AMR, MEC and visualization PC was based on the Robot Operating System (ROS). In addition, a modified OEM built Velodyne's LiDAR processing packages utilized in our tests. The modifications were mostly applied in order to reduce the required bandwidth over a ROS topic, and also to leverage MEC's compute resources more efficiently. Also, it is noteworthy that I order to reduce the LIDAR bandwidth requirements, only the raw, unconstructed data was uploaded and instead of sending the data-rich PointCloud2 ROS messages. In this way we were able to reduce the LiDAR bandwidth requirements by a factor of 10. Another important point is that since initiating connections over multiple ports on the client by MEC is disabled on Verizon's public network by default to prevent spam callers and unwanted data usage, and having these connections among MEC, Jackal, and visualization PC was a requirement in our set up, we had to establish a VPN server hosted on the MEC to enable individual clients to connect to the MEC through secure tunnels. An algorithm based on center point [11] with SECOND [12] was chosen in our framework for LiDAR object detection. This choice was due to its faster execution than the required 100ms time frame on the MEC's GPUs. We tuned this algorithm in our set up to be able to recognize pedestrians and vehicles. It aslo should be mentioned that we utilized Verizon's public MEC for this trial which is located in Wall, NJ.

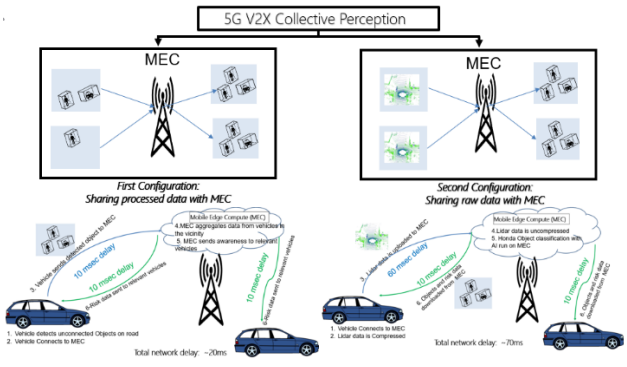

Fig. 7 Schematic Diagram for the Two CAV-MEC Connection Configurations

In Fig. 8, one can see a more clear break down of tbe delays in different stages of our set up, for the case that LIDAR raw data was being sent to the cloud. This set up

has higher uplink latency compared to our other configuration where processsed LIDAR data was being sent to the MEC. The main conclusion from this study is that the proposed architecture in this work seems to be feasible to be deployed using commercial 5G/MEC infrastructures.

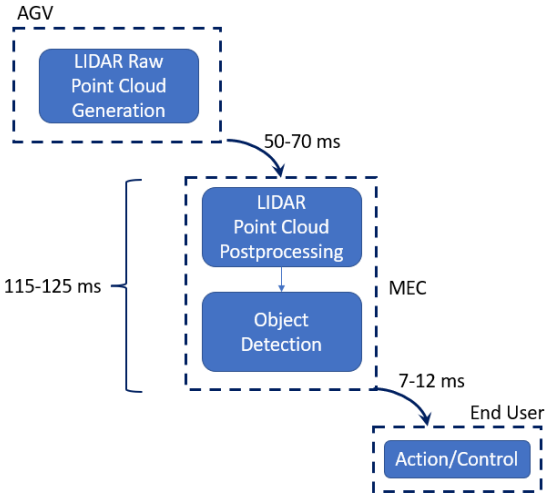

Fig. 8 Latency break down for different stages in our proposed architecture

4. CONCLUSION

In this paper, we propose a novel collaborative scene understanding scheme which incorporates the 5G communications and Mobile Edge Computing technologies as its two main pillars to enhance situational awareness. In addition, a risk quantification method has been integrated in this framework which works based on Bayesian Occupancy Filters concept, augmented with a collision loss function. Incorporating this risk estimation block in our framework, allows us to study and quantify the effect of collaborative sensing on both risk reduction and risk uncertainty reduction. Real world data and risk analysis are used to support the claims and quantify the value and feasibility of this solution.

REFERENCES

(1) Stefan Hoermann, Felix Kunz, Dominik Nuss, Stephan Renter, and Klaus Dietmayer, “Entering crossroads with blind corners. a safe strategy for autonomous vehicles,” in 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2017, pp. 727–732.

(2) Adrian Broadhurst, Simon Baker, and Takeo Kanade, “Montecarlo road safety reasoning,” in IEEE Proceedings. Intelligent Vehicles Symposium, 2005. IEEE, 2005, pp. 319–324.

(3) Matthias Althoff, Olaf Stursberg, and Martin Buss, “Modelbased probabilistic collision detection in autonomous driving”, IEEE Transactions on Intelligent Transportation Systems, vol. 10, no. 2, pp. 299–310, 2009.

(4) D. Nuss, *et al.*, “A Random Finite Set Approach for Dynamic Occupancy Grid Maps with Real-Time Application,” *The International Journal of Robotics Research*, vol. 37, no. 8, pp. 841–866, 2018.

(5) D. LaChapelle, T. Humphreys, L. Narula, P. Iannucci and E. Moradi-Pari, "Automotive Collision Risk Estimation Under Cooperative Sensing," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 9200-9204, doi: 10.1109/ICASSP40776.2020.9053745.

(6) Meier, J.N.; Kailas, A.; Adla, R.; Bitar, G.; Moradi-Pari, E.; Abuchaar, O.; Ali, M.; Abubakr, M.; Deering, R.; Ibrahim, U.; et al. “Implementation and evaluation of cooperative adaptive cruise control functionalities”, IET Intell. Transport. Syst. 2018, 12, 1110–1115.

(7) E. Emad, *et al.*, “Feature Sharing and Integration for Cooperative Cognition and Perception with Volumetric Sensors.” *ArXiv* abs/2011.08317 (2020).

(8) B. Dai, *et al.*, "Hybrid Sensing Data Fusion of Cooperative Perception for Autonomous Driving with Augmented Vehicular Reality," *IEEE Systems Journal*, vol. 15, no. 1, pp. 1413-1422, March 2021

(9) E. Moradi-Pari, *et al.*, "The Smart Intersection: A Solution to Early-Stage Vehicle-to-Everything Deployment," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 5, pp. 88-102, Sept.-Oct. 2022

(10) W. Zheng, A. Ali, N. González-Prelcic, R. W. Heath, A. Klautau and E. M. Pari, "5G V2X communication at millimeter wave: rate maps and use cases," 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 2020, pp. 1-5, doi: 10.1109/VTC2020-Spring48590.2020.9128612.

(11) Tianwei Yin, Xingyi Zhou, Philipp Krahenbuhl, “Center-Based 3D Object Detection and Tracking”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 11784-11793

(12) Yan, Yan, Yuxing Mao, and Bo Li. "Second: Sparsely embedded convolutional detection." Sensors 18.10 (2018): 3337.