This is the Pre-Published Version.
The following publication Y. Wang, V. Menkovski, I. W. -H. Ho and M. Pechenizkiy, "VANET Meets Deep Learning: The Effect of Packet Loss on the Object Detection Performance," 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), Kuala Lumpur, Malaysia, 2019 is available at https://doi.org/10.1109/VTCSpring.2019.8746657.

VANET Meets Deep Learning: The Effect of Packet Loss on the Object Detection Performance

Yuhao Wang[1], Vlado Menkovski[1], Ivan Wang-Hei Ho[2], Mykola Pechenizkiy[1]
[1]Department of Mathematics and Computer Science, Eindhoven University of Technology, Netherlands
[2] Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong
{y.wang9, v.menkovski, m.pechenizkiy}@tue.nl, ivanwh.ho@polyu.edu.hk

Abstract—**The Intelligent Transportation System with the integration of machine learning and inter-vehicle communications can enable various active safety measures in internet-of-vehicles. Specifically, the environmental perception is processed by the deep learning module from vehicular sensor data, and the extended perception range is achieved by exchanging traffic-related information through inter-vehicle communications. Under such condition, the intelligent vehicles can not only percept the surrounding environment from self-collected sensor data, but also expand their perception range through the information sharing mechanism of VANET (Vehicular Ad-hoc Network). However, the dynamic urban environment in VANET leads to a number of issues, such as the effect of packet loss on the real-time perception accuracy of the received sensor data. In this work, we propose a point cloud object detection module via an end-to-end deep learning system and enable the wireless communications between vehicles to enhance driving safety and facilitate real-time 3D mapping construction. Besides, we build a semi-realistic traffic scenario based on the Mong Kok district in Hong Kong to analyze the network performance of data dissemination under the dynamic environment. Finally, we evaluate the impact of data loss on the deep learning-based object detection performance. Our results indicate that data loss beyond 50% (which is a common scene based on our simulation) can lead to a rapid decline of the object detection accuracy.**

Index Terms—**3D Point Cloud, VANET, autonomous driving, deep learning, SUMO**

I. Introduction

Vehicular safety is improved in automated systems with accurate perception from sensors. Furthermore, the information shared through VANET (Vehicular Ad-hoc Network) can greatly increase vehicular perception range. Thus, if autonomous vehicles could cooperate with one another through inter-vehicle communications, and exchange sensor data for real-time mapping and deep learning-based safety alert, both the distant and hidden objects can be easily detected to enhance driving safety.

From the above two directions, the first is sensor data perception. Sensors relevant to autonomous driving include GPS (Global Positioning System), cameras, LiDAR (Light Detection and Ranging), etc. GPS data is widely used in both path planning and VANET Basic Safety Message (BSM) since it can provide the velocity and position coordinates in a straightforward way. However, the accuracy of GPS localization is highly related to the satellite signal condition. When comparing LiDAR data to the camera data (2D information that is constrained by lighting condition), LiDAR data contain 3D metric information and can directly measure the surrounding environment with accurate localization and 360-degree coverage.

The second direction is the integration of connected vehicles through wireless communications and deep learning-based intelligent driving. More specifically, with the installation of wireless units for Dedicated Short Range Communications (DSRC) on light-duty vehicles in the US [1], and the development of the vehicle-to-everything (V2X) communications for improving road safety, VANET can thus be a strong catalyst to enhance traffic efficiency through information sharing mechanism. Besides, intelligent driving requires the ability to identify navigable terrain through various sensing systems. The implementation of end-to-end deep learning modules from vehicular-sensor data for perception and decision making can enhance the on-road driving safety in an intelligent manner. With more sensor data coming in, some of the blind spots of vehicles can be eliminated for better safety.

Therefore, information sharing and intelligent processing can be achieved on-the-fly through the integration of VANET and the end-to-end deep learning perception system. This has several challenges. Despite the increasing interests in the separate world of either data dissemination in VANET [2], [3] or deep learning-based 3D object detection [4], [5], very few studies considered the multimodal framework of the integrated VANET with deep learning [6]. The consideration of either the real VANET communication protocols or adequately exploit the urban vehicular mobility patterns is absent. In addition, the information dissemination in vehicular networks seldom considers the delivery of both point cloud data and the deep learning-based perception results, let alone the analysis of the impact of VANET Packet Loss Ratio (PLR) under dynamic urban environment on the deep learning-based 3D object detection performance.

To address these challenges, we investigate the feasibility of integrating VANET data dissemination of both BSM and original point cloud data with the end-to-end deep learning module for 3D point cloud object detection. For analyzing the dynamic VANET environment and its impact on the deep learning-based object detection accuracy, we also developed a semi-realistic urban traffic scenario with the consideration of various traffic infrastructures.

Overall, the major contributions of this work are three-fold:

- We propose a system architecture that integrates vehicular

©2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

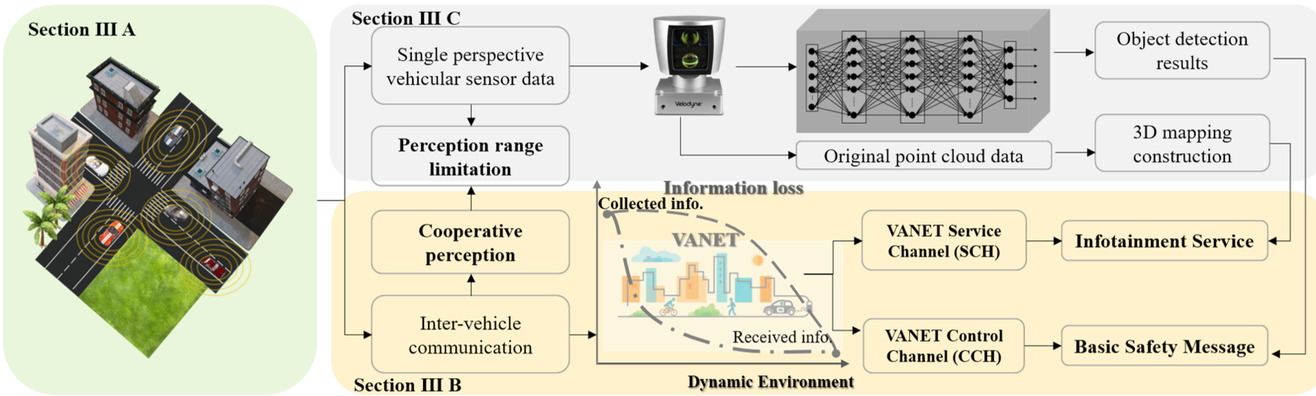

Fig. 1: System architecture.

communications and deep learning based object detection for analyzing the impact of communication loss on 3D object detection;

- We build a semi-realistic traffic scenario to evaluate the amount of packet loss due to fading and signal attenuation in dense city like downtown Hong Kong. We found that the amount of packet loss could be up to 90% depending on the transmission power and vehicular density in the surroundings;
- Through the integrated system framework, we identified that a packet loss of more than 50% can already lead to a rapid decline of object detection accuracy.

II. Related Work

Environmental sensing devices with the application of deep learning allow the autonomous driving vehicles to perceive the surrounding environment in an intelligent manner. There is considerable research on the accurate localization [7] and 3D object detection [5], [8] with LiDAR data and end-to-end deep learning. This shows 3D LiDAR can conquer the degrade GPS localization accuracy in urban scenarios, and its 360-degree sensing offers a solution to the decreased camera sensitivity in challenging lighting conditions. However, with the absence of vehicular sensor networks, both the perception range limitation of an individual vehicle and physical occlusion in the environment make it challenging to ensure traffic safety.

Works related to the collaborative perception through inter-vehicle communications [9], [10] aim to alleviate this problem by leveraging additional sensor information from other vehicles. For example, [11] shows that 3D sensor fusion is achieved with the aid of vehicle-to-vehicle (V2V) communications. In short, VANETs can help achieve a global view of the traffic environment through the benefits of wireless networks. However, information dissemination in VANETs under the real-scene, especially the urban scenario, has long being a problem [3]. The Quality of Service (QoS) will be highly affected due to the distributed, intermittently connected, and time-critical nature among high-mobility of vehicular networks. Therefore, with the merits of receiving extra sensor information in VANETs to expand the vehicular perception range, the demerits of the packet loss ratio with side effects to the deep learning perception performance are non-negligible.

III. Research Methodologies

Our work integrates inter-vehicle communications and deep learning perception in a whole system, as shown in Fig.1. Deep learning is responsible for the real-time object detection from LiDAR point cloud data. VANETs in urban traffic scenarios can expand the on-road perception range into a global view. Furthermore, we use different data dissemination services to analyze the dynamic VANET environment and its impact on the perception performance in deep learning. In detail, BSMs are broadcast through the DSRC control channel and the original point cloud data is transmit through the service channel. To evaluate the impact of data loss to the deep learning performance, we link the consequent packet loss ratios to various point cloud sparsity degrees. The following subsections explain each part in Fig.1.

A. Vehicular mobility modeling in VANET

Urban traffic is the communication carrier between vehicles. In VANET, the motion of vehicles vary greatly within time and space. It needs to be considered under dynamic road usage conditions (such as the sparse, moderate and congestion traffic) and real vehicle driving routes. Due to the differentiation of the traffic simulator and network simulator, in this work, the vehicular mobilities are generated from the traffic simulator of SUMO (Simulation of Urban Mobility) [12] with various traffic conditions. Then, we further export the mobility traces into the network simulator of NS3 [13] for the inter-vehicle communications.

In our experiments, Mong Kok road topology is extracted from OpenStreetMap [14]. Then, we use JOSM (Java Open Street Map Editor) to analyze the location data of bus stops, traffic lights, fixed bus routes and building topologies. The traffic signal duration and the buses' dwell time are determined according to the empirical data. These are combined into our vehicular mobility scenario. This process resulted in 23 bus stops, 29 traffic lights, and 5 fixed bus routes. The road usage condition is also an important factor of the VANET connectivity. Given the constant number of 60 buses, to generate different types of traffic flows, we vary the number of other generic vehicles, including sparse (with extra 14 vehicles), moderate (extra 55 generic vehicles) and congestion (139 generic vehicles). Table I provides an overview.

TABLE I: Different levels of traffic flows

Traffic Condition	Vehicular numbers (Bus/Total Vehicles)	Congestion rate
Sparse	60/74	—
Moderate	60/111	15%
Congestion	60/199	53%

B. Inter-vehicle communication

The dynamic environment with diverse road topologies and uneven vehicular density distributions leads to the instability of inter-vehicle communications. In this work, with the above traffic scenario, our VANET simulation implements the IEEE 802.11p standard on all vehicular nodes. The simulation parameters are listed in Table II. Specifically, we use the ITU-R1411Los propagation loss model, which is suitable for the short-range outdoor communication. Besides, routing protocol plays an important role to ensure the successful wireless communication between vehicles. We adopt the OLSR (Optimized Link State Routing)[1] routing protocol for the BSMs dissemination. In addition, the traffic simulation duration is 1200s within the scenario range of $274.441m \times 433.396m$. We also vary the wireless transmission power (from 16 dBm to 28 dBm) and the transmission range (from 50 m to 400 m) for the analysis of numerical vehicular communication conditions. Furthermore, in order to disseminate both the deep learning perception results and the original point cloud data, we propose two VANET scenarios, summarized below:

- Scenario 1: Deep learning-based 3D object detection results as the 200 Bytes BSMs and broadcast through the DSRC control channel to all other vehicles under the sparse, moderate and congestion traffic conditions;
- Scenario 2: The dissemination of the original point cloud data (4 MB) as the infotainment service through the service channel. We assume 30 user-pairs apply this customized service under the moderate traffic condition.

C. 3D point cloud object detection with deep learning

In order to assist intelligent vehicles with a clear understanding of the surrounding environment, the representation and extraction of useful features from high-volume point cloud data with deep learning is necessary. For this, we use VoxelNet [5] as our object detection benchmark. This end-to-end deep learning framework can detect 3D objects with an efficient implementation and includes three major parts, namely, feature learning network, convolutional middle layer and region proposal network. The feature learning network can organize the orderless point cloud data through voxel feature encoding layers. Meanwhile, the 3D convolutional middle layer is responsible for the spatiotemporal feature learning, while the region proposal network detects the vehicles.

The dataset we applied is the KITTI Velodyne 64E range scan data [15], which contains 7481 annotated training data. As we explained in the beginning of Section III, our major aim is to evaluate the effect of the dynamic environment with packet loss to the deep learning perception accuracy. Thus, in our experiments, we use the annotated training data only and randomly separate this data into 50% training, 30% validation and the rest 20% (1500 frames data) for the offline performance analysis. The VoxelNet was trained with a NVIDIA Geforce GTX1080 GPU, and we measured the performance on the bird's eye view with Average Precision (AP). Table III shows the results of the detection accuracy and the comparison with the original VoxelNet results.

[1]This routing protocol shows the best performance than AODV (Ad Hoc On-Demand Distance Vector), DSDV (Destination-Sequenced Distance Vector) and DSR (Dynamic Source Routing) through our experimental study.

TABLE II: VANET Simulation Parameters

Scenario size	$274.441m \times 433.396m$
Simulation duration	1200s
Transmission power	16dBm - 28dBm
Routing protocol	OLSR
Physical mode	OFDMRate6MbpsBW10MHz
80211mode	MAC:802.11p / 5.9GHz
Packet size	200 bytes (Basic Safety Message) & 4,000,000 bytes (3D point cloud data)
Transmission range	50m - 400m
Porpagation loss model	ITUR1411LosPropagationLossModel

TABLE III: Bird's eye view performance evaluation

	Easy	Moderate	Hard
Our repeated results (AP)	83.70%	74.22%	67.20%
Original results (AP) [10]	89.35%	79.26%	77.39%

D. The effect of the packet loss to the deep learning accuracy

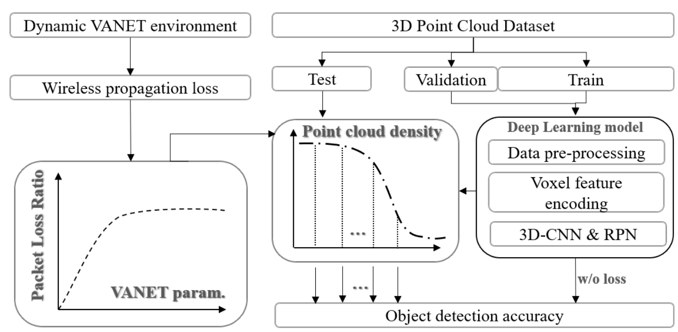

Fig. 2: VANET packet loss to the deep learning performance.

Training the deep neural network from high-volume historical data is necessary. But it's also important to use on-road data to ensure quality of real-time intelligent control. The problem is that with the pre-traind deep learning module from historical data, the information loss in VANET could degrade the detection performance. Fig.2 shows the VANET information loss to the different sparsity levels of point cloud data, and further to the effect of deep learning performance. The transmission of the point cloud data is based on the UDP protocol under the dynamic VANET scenario. For this experiment, we follow [16] and assumes the interleaving techniques are used to ensure that the packet loss will disperse uniformly in the point cloud, and will not affect the point-cloud-visualization at the receiver side.

Based on the prerequisite in [16], we assume a least serious condition that 50% data loss can only leads to 50% random sparsity of the original point cloud data, and we set the pre-trained deep learning module as the no-missing-data baseline,

TABLE IV: Packet Loss Ratio based on the BSMs data dissemination(%)

Packet Loss Ratios under various transmission power, different vehicular density and various transmission ranges.																					
Txp	16dBm			18dBm			20dBm			22dBm			24dBm			26dBm			28dBm		
Den. / Dis.	S	M	C	S	M	C	S	M	C	S	M	C	S	M	C	S	M	C	S	M	C
50m	22.20	28.11	54.53	20.71	25.88	51.46	27.67	23.97	48.46	18.29	21.95	45.74	16.56	19.00	41.62	14.71	16.55	37.96	13.05	13.93	33.69
100m	46.27	58.08	80.57	45.24	56.78	79.26	50.05	55.67	77.98	43.58	54.59	76.74	42.38	52.28	75.05	41.10	51.34	73.46	39.96	49.81	71.60
150m	56.32	71.25	88.24	55.49	70.35	87.44	59.40	69.59	86.64	54.13	68.79	85.59	53.16	67.61	85.93	53.16	66.62	83.93	51.19	65.58	82.80
200m	63.48	78.86	91.96	62.79	78.19	91.41	60.06	77.64	90.88	61.65	77.05	90.37	60.84	76.18	89.67	60.84	75.46	89.01	59.19	74.68	88.24
250m	68.46	82.03	92.64	67.86	81.47	92.64	70.68	80.99	92.18	66.89	80.49	91.74	66.18	79.76	91.15	65.43	79.15	90.58	64.76	78.49	89.92
300m	70.64	83.45	93.62	70.08	82.93	93.19	72.70	82.49	92.77	69.16	82.03	92.36	68.51	81.35	92.39	67.81	80.79	91.28	67.19	80.18	90.67
350m	78.22	86.31	94.07	77.80	85.88	93.67	79.75	85.52	93.28	77.13	85.14	92.93	76.64	84.58	92.90	76.13	84.11	91.90	75.66	83.61	91.33
400m	78.22	86.31	94.07	77.80	85.88	93.67	79.75	85.52	93.28	77.13	85.14	92.93	76.64	84.58	92.90	76.13	84.11	91.90	75.66	83.61	91.33

TABLE V: PLR based on the point cloud data dissemination

Transmission power	16dBm	24dBm	28dBm
Transmission distance	100m	150m	200m
Packet Loss Ratio	90.03%	89.68%	89.63%

with the performance in Table III. To evaluate the effect of packet loss on the deep learning-based perception accuracy, we mapped various VANET PLRs to different levels of point cloud sparsity in the test data. Table IV shows the BSMs' PLR and Table V shows the PLR on the original point cloud data.

IV. Simulation results and discussion

We now show our traffic scenario and VANET communications. Mobility traces are generated with the professional traffic simulator SUMO to better reflect the realistic vehicular traffic. The generated urban scenario offers a benchmark for inter-vehicle communications. Moreover, we consider two VANET communication scenarios, and one of them is the broadcasting of BSMs through VANET control channel. The PLR under this condition is shown in Table IV. As the results show, increasing the transmission distance[2] (i.e., marked as Dis.), the communication range will be enlarged. This can introduce more frequent contention and collision among neighboring vehicles, and thus a higher packet loss ratio. On the contrary, the adjustment of the transmission power (Txp in dBm) can provide better connectivity of nodes and improves the packet delivery ratio. With the various traffic scenarios in Table I, we also analyze the effect of traffic conditions (i.e., marked as Den.) to the PLR ('S' means sparse traffic, 'M' means moderate and 'C' represents congestion condition). Through our extensive simulation, the dense traffic condition is likely to suffer more packet loss than the sparse traffic due to message collision.

We also considered the original point cloud data (4 MB) dissemination. To cover the Non-Line-of-Sight condition in inter-vehicle communications, we added the real world building topology with extra 14 dBm [3] wall-penetration loss in VANETs. As we can see in Table V, with a 16 dBm transmission power (consistent with the 100 m transmission distance), the packet loss ratio can reach more than 90%. Increase the transmission power to 28 dBm only gives minor improvements.

[2]A simulation setting in NS3 for calculating different values of Packet Delivery Ratio.

Based on the above analysis, we can conclude that the PLR within urban area could range from 10% to 90% depending on the transmission power and vehicular density. Then, we re-arrange the point-cloud data with various information loss conditions into 10 different datasets (as shown in the x-axis in Fig.4), such that the data can have different sparsity degree due to the dynamic VANET communication. After that, these 10 datasets served as input to our pre-trained deep learning module for the object detection task. The detection accuracy is shown in Fig. 4 and further illustrated in Fig.3. It is clear to see that with less than 50% data loss, the detection accuracy only decreases 4.21% in the easy level and more than 3% in the hard level. This means that deep learning module is robust enough (only suffers a slight accuracy decline) to handle the on-road intelligent perception with even 50% VANET data loss. When data loss beyond 50% (which is more common based on our PLR simulation in both Tables IV&V), the rapid decline of the object detection accuracy occurs. With a 90% data loss, the detection accuracy in the moderate level is lower than 35%, while the hard level is even worse (31.90%). Note that the static historical dataset with more than 83% accuracy in Table III was proven to be efficient. However, under the dynamic VANET environment with information loss, the static pre-trained module is obviously not sufficient to be directly applied for the real-time use of on-road dynamic applications.

To further illustrate the insufficient use of deep learning, we analyze our detection results from Fig.3 with three representable conditions, namely, sparse traffic (2 vehicles under this example), dense traffic (9 vehicles) and the urban scenario (one vehicle under a complex scene). Specifically, under the sparse traffic condition (Fig.4 a, d, g), it is clear to see that 50% data loss can still generate an accurate detection. However, detection fails with a 90% information loss ratio. In addition, under the dense vehicle scenario in Fig. 4 (b, e, h), the 50% data loss can maintain the true positive detection but introduce more false positives at the same time. When data loss becomes larger (90%), the vehicles in the distance cannot be well detected. Besides, the urban scenario in Fig. 4 (c) makes the things worse due to the complexity of the background. The 90% data loss in Fig.4 (i) can mix the noise data with the object features together and generate numerical false detections. Notice that the corresponding minimal distance for the KITTI object detection of moderate and hard level is 47m (25px pixel height) [8]. The 31.90% detection accuracy due

(a) Sparse traffic (Without data loss) (b) Dense traffic (Without data loss) (c) Urban scenario (Without data loss)

(d) 50% data loss (e) 50% data loss (f) 50% data loss

(g) 90% data loss (h) 90% data loss (i) 90% data loss

Fig. 3: Deep learning-based object detection results with various packet loss degree. ((a, b, c) are qualitative results. We projected the point cloud detection results on RGB-images for better explanation.)

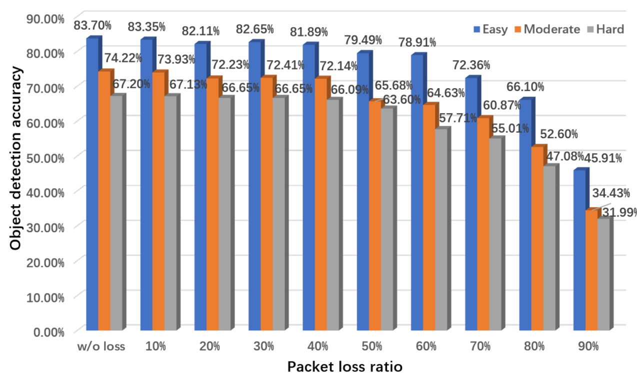

Fig. 4: Data loss to the effect of object detection accuracy.

to the 90% data loss is obviously not enough to ensure the driving safety.

V. Conclusion and Future Work

This paper proposes a novel system framework to integrate the inter-vehicle communication and deep learning-based object perception for the high-efficiency safety measurement in intelligent driving. We address the concern of the potential issue under such framework: the VANET packet loss to the effect of the deep learning-based object detection performance. As explained, under the dynamic urban environment, the packet loss can greatly affect the compliance of the received data. The vehicular perception range and deep learning accuracy will thus be affected. As future work, in order to tackle this issue, reducing the packet loss ratio through the global adjustment of inter-vehicle communications could be a straightforward way. Meanwhile, the interpretability of the deep neural network is also important to avoid the opacity of the decision-making process.

References

[1] S. Bayless, A. Guan, A. Shaw, M. Johnson, G. Pruitt, B. Abernathy *et al.*, "Recommended practices for dsrc licensing and spectrum management: a guide for management, regulation, deployment, and administration for a connected vehicle environment." United States. Dept. of Transportation. ITS Joint Program Office, Tech. Rep., 2015.

[2] F. Martelli, M. E. Renda, G. Resta, and P. Santi, "A measurement-based study of beaconing performance in ieee 802.11 p vehicular networks." *2012 Proceedings IEEE INFOCOM*, pp. 1503–1511, 2012.

[3] Y. Wang and I. W.-H. Ho, "On-road feature detection and fountain-coded data dissemination in vehicular ad-hoc networks," *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications*, pp. 1–6, 2017.

[4] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in Neural Information Processing Systems*, pp. 5099–5108, 2017.

[5] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," *arXiv preprint arXiv:1711.06396*, 2017.

[6] Y. Maalej, S. Sorour, A. Abdel-Rahim, and M. Guizani, "Vanets meet autonomous vehicles: A multimodal 3d environment learning approach," *2017 IEEE Global Communications Conference*, pp. 1–6, 2017.

[7] E. Javanmardi, M. Javanmardi, Y. Gu, and S. Kamijo, "Autonomous vehicle self-localization based on multilayer 2d vector map and multi-channel lidar," *IEEE Intelligent Vehicles Symposium*, pp. 437–442, 2017.

[8] B. Li, T. Zhang, and T. Xia, "Vehicle detection from 3d lidar using fully convolutional network," *arXiv preprint arXiv:1608.07916*, 2016.

[9] S.-W. Kim, B. Qin, Z. J. Chong, X. Shen, W. Liu, M. H. Ang, E. Frazzoli, and D. Rus, "Multivehicle cooperative driving using cooperative perception: Design and experimental validation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 663–680, 2015.

[10] A. Rauch, F. Klanner, R. Rasshofer, and K. Dietmayer, "Car2x-based perception in a high-level fusion architecture for cooperative perception systems," *IEEE Intelligent Vehicles Symposium*, pp. 270–275, 2012.

[11] Y. Ryan, C. Ellick, S. Carmine, C. Bin, and B. Gaurav, "Collaborative perception for automated vehicles leveraging vehicle-to-vehicle communications," *IEEE Intelligent Vehicle Symposium*, pp. 1099–1106, 2018.

[12] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumo–simulation of urban mobility," *The Third International Conference on Advances in System Simulation*, vol. 42, 2011.

[13] G. F. Riley and T. R. Henderson, "The ns-3 network simulator," *Modeling and tools for network simulation*, pp. 15–34, 2010.

[14] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervas Comput*, vol. 7, no. 4, pp. 12–18, 2008.

[15] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 3354–3361, 2012.

[16] P. F. (LUH), "Mobile robots with novel environmental sensors for inspection of disaster sites with low visibility," Mobile Robots with Novel Environmental Sensors for Inspection of Disaster Sites with Low Visibility, SmokeBot, Tech. Rep., 2015.