Journal of Engineering, Management and Information Technology

ASPUR PUBLISHING

Vol. 03, No. 04 (2025) 311-318, DOI: DOI:10.61552/JEMIT.2025.04.009 - http://jemit.aspur.rs

DATASETS APPLIED INTO V2X 6G SYSTEM

Houda Chihi [1]

Received 27.06.2024.
Revised 03.08.2024.
Accepted 20.08.2024.

Keywords:

V2X, 6G, Reinforcement learning, Datasets, traffic management.

Original research

A B S T R A C T

Due to the great and continuous increase of traffic density, accurate data exchange between vehicles become more challenging and complex. In this regard, we require powerful tools able to improve the performance of the vehicular network. In this regard, real data collection about vehicular communication status is required for better traffic management. In addition, due to huge traffic density together with the great demand of efficient communication vehicular link the recent cellular technologies couldn't satisfy the required safety. For this artificial intelligence (AI) and machine learning (ML) application to the 6G network are considered as required tools for driving safety ensuring. The huge growth of traffic density together with the requirement of connectivity encourage 6G exploitation for V2X communication. In this survey we highlight the available datasets in literature exploited for vehicular to everything (V2X) applied in different scenarios.

1. INTRODUCTION

The great density of vehicular nodes leads to massive data generation which corresponds to a complexity towards vehicular network in particular regarding resource assignment. The communication between each vehicle and its ecosystem is a challenging problem which needs much reliability to ensure better quality of service together with high safety. Especially that essential problem of accident occurrence is related to congestion together with huge cars' density and unstable vehicular position.
Autonomous driving is recently a hot topic due to the great increase of cars' density leading to traffic jam. In this context, accurate data exchange between vehicles at real time could enhance driving safety.
For this reason, many researchers build datasets in order to record vehicle behavior. In this survey we present the state of the art of different proposed datasets for different applications into V2X scenarios.
V2X encloses the communication between vehicles, infrastructure and pedestrians.V2X communication is defined through a cooperative scenario between different actors participating in the same environment highlighted by an interaction between a vehicle and the surroundings.
Accurate description requires real exchange of awareness information about the behavior of each player in the same ecosystem. The communication is enabled by 6G providing advanced safety options.
Note that traditional vehicular communication methods have shown their limit, hence we need an advanced enabling technology which could satisfy the requirement of ultra high latency together with the huge need of high data rate supporting much coverage. In addition, due to the communication specificity of each node with everything, data exchange is a challenging task. According to the mobility particularity of cars leading to information interruption. AI enables much strength towards vehicular network leading to much reliability together with advanced services especially that some sensors show limitation. Real time data collection helps in traffic management and V2X communication improvement. In addition, ML integration ensures much smarter vehicles able to communicate reliably even in high complexity of environment to ensure much safety towards vehicular network interference and congestion management. ML integration into vehicular communication ensures driving experience upgrade. In

[1] Corresponding author: Houda Chihi
Email: houda.chihi@supcom.tn

this paper we present the state of the art of all datasets suggested for V2X communication using different types of ML algorithms. In addition, we suggest Reinforcement Learning (RL) as enabler for V2X 6G communication system. Artificial Intelligence (AI) integration leads to intelligent cars which are able to learn and investigate data and making the right decision by the end. Datasets definition is necessary for better traffic efficiency ensuring safety improvement. We need efficient datasets to explore the vehicular ecosystem especially due to the dynamic network variability. Datasets help in network prediction and traffic jam mitigation leading to much reliability especially that vehicular ecosystem changes dynamically and vehicles exchange notifications messages.

The paper is a survey about the datasets proposed for solving different problems towards the V2X network. The difference between the different datasets is the limitation towards the captured scenes since many factors impact driving quality of experience such as weather, roads status and car's density.

In this paper we present the state of the art of different datasets applied to V2X in different use cases. ML provides much intelligence to each car in the way that it is able to predict in advance the velocity of its neighbor which improves driving experience. In addition, ML allows each vehicle to make decisions with the help of a specific dataset resulting from real time collection of information. ML is an efficient tool which improves the quality of driving experience through data exchange between vehicles. ML allows much efficiency towards the communication between vehicles ensuring accurate monitoring especially that AI has demonstrated its power towards many services. In this survey we present a state of the art of the datasets applied in V2X context in different scenarios.

Authors in (Alablani & Arafah, 2022) presented a dataset for V2X related to Los Angeles city based on machine learning policy where the purpose is cell selection into 5G. The idea in (Yu et al., 2022) is to propose a dataset DAIR-V2X related to vehicular infrastructure where the scenario is in 3D schema towards the infrastructure together with cars. Authors in Ros et al. (2016) present a dataset of a scenario of a city with different states of roads and pedestrians. The presented datasets in (Cordts et al., 2016) is a set of recorded videos of vehicular behavior into different cities. The datasets proposed in Geiger et al. (2012) are based on a recording of different types of scenarios such as highway, rural and urban city. The aim in Caesar et al. (2020) is about a dataset collecting the sensors results related to vehicular scenarios. Another dataset in Li et al. (2022) is about a collaborative scenario of different vehicles. The report in Hernangómez et al. (2023) highlighted a dataset of Berlin related to a V2X communications which is about a collection resulting from GPS localization performed over 3 days into an urban scenario corresponding to sidelink and with the help of eNodeB or gNodeB. The paper in Xu et al. (2022) described a dataset of a vehicle to vehicle communication following a recording performed in different cities. The proposed datasets in Huang et al. (2018) was for an ApolloScape corresponding to a high accuracy of registration images towards driving status in different cities in 3D at different timing. Authors in Neuhold et al. (2017) exploited different tools to record different schemas around the word following an outdoor context. In Yu et al. (2020) authors proposed a dataset for driving scenarios taking into account the weather and environment following a set of videos. The paper in Sun at al. (2020) suggested a datasets helping about the driving way based on a prediction process and exploiting CARLA platform. A vehicle to vehicle dataset was proposed in Lee et al. (2019) based on a set of collections performed at different levels of communication. An overview about V2X communication enabled by 6G was described in Wang et al. (2023). Authors in Salehi et al. (2022) solved the challenge of beam assignment of vehicular scenarios based on millimeter wave communication via the dataset resulting from different sensors. Therefore, a commute dataset based on a recording of different localization together with the directivity of different cars was presented in Schafer et al. (2018). A prediction of roads status was defined in Santana and Hotz (2016) enable by a set of videos. Another proposed dataset in Houston et al. (2021) about driving behavior based on videos recording of the different actors participating in the driving environment. A recording of a driving proposal into an urban scenario was suggested in Sun et al. (2020).

Another datasets of different driving scenarios were introduced in Matuszka et al. (2022) considering the weather's status. A 3D concept of datasets for autonomous communication was the aim of the paper in Pham et al. (2020) covering all traffic conditions with accurate resolution of images. Another developed dataset was presented in Choi et al. (2018) taking into account the driving status at each time of the day with the help of different sensors. A 4D dataset was the focus of the paper in Zheng et al. (2022) where authors exploited radars in 4D status for vehicular networks. In Patil et al. (2019) an exploitation of Lidar in 3D context was proposed in order to build a dataset for vehicular tracking related to Honda research institute.

The contribution in Hernangómez et al. (2023) offered a dataset enabling different use cases with accurate resolution leading to better driving quality of experience. Industrial datasets for two kinds of testbeds related to vehicle to vehicle communication compared to another one related to vehicle to infrastructure scenario in which the control is performed by ML was the aim of the paper in Hernangómez et al. (2024). Authors in Ali et al. (2020) discussed the application of ML in the case of 6G network consideration. A datasets related to 6G communication was the focus of the paper in Alkhateeb et al. (2023) which is called DeepSence where the measure is performed with the help of sensors to enable deep learning exploitation suitable to many use cases of vehicular communication particularly following vehicle to infrastructure scenario. Further position prediction determination was the purpose of the stated datasets in Irio et al. (2021). Another dataset proposed in Seon et al. (2023) which helped in predicting the next generated message in V2X context avoiding congestion. The application of ML following the V2X 6G networks was the aim of the paper in Noor-A-Rahim et al. (2022). A dataset of i80 suggested in Colyar and Halkias (2006) where the aim was a simulation of a diving experience into a highway context in an accurate way.

Authors in Saikia et al. (2023) investigated a vehicular communication with the help of reconfigurable intelligent surface RIS in which the aim is data rate improvement.

The proposed datasets in Zhu et al. (2016) was a collection of a large amount of images taken in real scenarios about weather as well as signs related to traffic. In Stallkamp et al. (2011) the focus was based on a real collection of a large signs related to traffic status considering the variability of weather, position, and directivity. The datasets in Lee et al. (2017) took into account raining days together with road status and bad situations of illuminations which impacts the driving behavior. The survey in Balkus et al. (2022) was machine learning application for V2X scenario considering 5G as enabler. The Next part of the survey is about presenting an overview of the ML V2X 6G system. Then, section III is about system modeling and theoretical analysis investigation of reinforcement learning application for vehicular context. Next, final remarks are included in the conclusion.

2. OVRVIEW ABOUT ML V2X 6G SYSTEM

The continuous growth of internet users requires a huge spectrum offering much connectivity. In this regard, 6G is the powerful enabler of communication between vehicles allowing much coverage and huge data rates as accurate understanding of the vehicular ecosystem is important for safety improvement. The communication is based on an exchange of safety messages in a frequent way. Data are collected with the help of sensors and due to continuous growth of exchanged information which results in interference increase leading to continuous variability towards channel gain, advanced technology communication enabler is required. Driving experience improvement is with high priority into V2X communication.V2X communication stands for different scenarios between the vehicle and pedestrian, infrastructure, network and other vehicle participating in the same ecosystem. To better characterize information exchange between vehicles 6G is adopted as an enabler which allows advanced services especially that millimeter wave communication has shown pathloss issues. AI and ML integration into V2X 6G enhances the resource assignment in the way to guide the driver to find the efficient path especially in urban areas. 6G exploitation as an enabler for V2X communication is a requirement following technologies progress together with huge data rate demand which calls for advanced tools for better traffic prediction such as ML which allows proactive reaction towards roads issues. Since 6G already integrates AI policy which is an enabler of advanced services in which its application on V2X context enhances the performance of vehicular communication. In this regard, 6G is proposed as a solver for the requirement of a huge spectrum. 6G technology exploits terahertz (THz) spectrum and enables huge bandwidth exploitation. Vehicular traffic improvement relies on a specific dataset and to ensure reliable communication between vehicles accurate allocation of resources is mandatory. 6G offers advanced connectivity ways that could be maintained which is helpful for data exchange into a complicated scenario of V2X communication in which resource assignment is enabled by reinforcement learning process. To better characterize information exchange between vehicles 6G is adopted as an enabler which allows advanced services leading to a smart interaction between each car and the environment that surrounds it. 6G adopting AI policy represents a promising enabler for huge data rate and traffic efficiency satisfaction. AI enables self adaptation to the dynamic change of the ecosystem in the way that each car can take the efficient decision which improves the connectivity together with the possibility of advanced services. AI is a powerful tool for traffic management with accurate resource allocation leading to quality of service and connectivity improvement. AI is integrated in 6G technology which ensures an adaptation of each car to the vehicular environment leading to much safety and enhances the quality of experience. Each node is assimilated as a smart actor deciding about the position to occupy starting by a learning process of the vehicular environment. The vehicle is assimilated as a smart node predicting next localization of each neighbor which improves the awareness. In the way that each vehicle learns the ecosystem status and takes the appropriate decision especially in the context of vehicular spectrum occupancy. The aim is to forecast the possible occupied position by each node based on a neighbor's behavior due to the fact that velocity heterogeneity leads to unreliability and much complexity towards packet exchange. Information collection about vehicular behavior is required for better prediction of the cooperation between vehicles. Mobility management impacts the quality of service of the vehicular network. Vehicle's location proactive determination helps traffic management. AI redefines the communication between vehicles enhancing the efficiency of spectrum exploitation. It helps in informing in advance about the position of each car which is changing frequently. In the way that each car is assimilated as a smart device who is able to send and receive data with a prediction of the next position of its neighbor. 6G could handle the unstoppable increase of vehicular nodes. Using the potential of AI especially in urban areas which help in providing much safety. Following the continuous increase of traffic resource allocation is a challenging task for this a proactive assignment is a good tool for better traffic monitoring. The datasets are more accurate in the use of 6G as an enabler based on many sensors. Vehicular Data collection is required for better prediction function. AI improves the transmission of each packet at the specific time benefiting from the advanced features 6G offering huge communication spectrum corresponding to terahertz (THz) frequencies since although the efficiency of 5G it remains limited facing the huge increase of traffic density. A specific use case of artificial intelligence is V2X traffic prediction in the way to choose the efficient driving path which mitigates interference occurrence. A way of vehicular traffic improvement depends on velocity forecasting. In addition, due to different velocity levels of vehicular network resource allocation applied previously are no longer efficient. The variability of velocity levels leads to dynamic and instantaneous channel state information (CSI) change. Therefore, previous resource allocation methods depending on CSI are no longer efficient. Velocity change leads to dynamic variability of channel gain due to shift of Doppler effects. Due to massive information related to vehicular behavior, machine learning policy is an efficient tool to manage congestion in particular in urban areas. Mobility is a challenging issue towards vehicular communication in which accurate management helps in preventing accidents happening which relies on previous information about vehicular behavior which helps in improving road safety especially due to the challenging variability of traffic. Mobility is a challenging problem especially ensuring information's drop. Velocity optimization is performed with the help of AI which helps in interference mitigation especially considering heterogeneous network of nodes with different velocity level, providing much ability to vehicles in the way that it can manage different scenarios. Velocity prediction helps in accurate resource allocation. Proactive velocity estimation leads to better safety ensuring which helps in the modification of driving behavior dynamically based on the stored datasets. AI application ensures efficient interaction between each node and the surrounding environment. Vehicular communication link is impacted by many factors such as road status, cars' density and weather.

The system model, illustrated in Figure 1, is about a gNodeB communicating with a set of vehicles via Terahertz frequencies. On the other hand, THz communication exploitation allows wide access to the spectrum which improves the efficiency of the vehicular network. Data collection requires advanced sensors to generate accurate datasets helpful for resource allocation prediction. The prediction is based on a real collection of data from a real communication scenario. A vehicular scenario of different Velocities ensures connectivity issues. Proactive determination of a vehicle's position relies on ML integration in the way of providing the ability for cars to take efficient decision based on a specific dataset. ML allows new driving perspectives especially that the use of previous communication tools has shown vehicular communication drops impacting reliability and safety. Resource allocation forecasting helps in spectrum efficiency. AI is a powerful approach in the way to manage the complexity of vehicular networks. In addition, AI allows autonomous management of vehicular network. In addition, AI is able to follow dynamic variability of vehicular network and exploit the collected data in the way to ensure much reliability towards information exchange. AI tools have shown their efficiency especially considering broadcast transmission context. ML is a powerful tool in the way to manage vehicular network complexity enabling the exchange with the ecosystem which enables efficient resource assignment. ML provides much reliability towards V2X communication. Predictivity is enabled by ML policy which is recently a requirement due to the limitation of legacy methods. Reinforcement learning is a kind of ML tool proposed in this survey for resource assignment based on a real time registration of a vehicular scenario. ML is an efficient suggested tool to define further position of each vehicle. ML application helps in determining the best itinerary overcoming the challenging issues of traffic. ML provides advanced functions for the V2X network leading to much connectivity together with agility towards vehicular network. Mobility prediction is an important criterion to ensure efficient communication. A handover prediction leads to much reliability and it is a requirement due to huge increase of traffic density. A forecast of the vehicular position is essential for accurate management of vehicular traffic. Velocity profile of each vehicle prediction helps in determining the best path for interference overcoming. Whenever a vehicle knows in advance about the behavior and position of the other node it can make a quick decision which improves the safety. Previous information of vehicular behavior is required to forecast further one. Indeed, the proactive determined position of each car depends on the previous one. Whenever a vehicle knows in advance the position of the other one it can make a quick decision which improves the safety of vehicular environment. Previous information of vehicular behavior is required to forecast further one. The type of ML application relies on the features of the V2X scenario to define further coordinates occupied by each car. ML ensures much intelligence to vehicles in the way to improve the driving experience enhancing the cooperation and coordination. It helps each node to take the efficient decision regarding the traffic situation. Machine learning covers different brands of algorithms which could be reinforcement, unsupervised or supervised learning where the difference between them relies on running time. Proactive knowledge of the velocity the neighbor helps each car to manage its own mobility features; especially that interaction between cars is related to traffic velocity. The performance of a vehicular network depends on both the position and speed of each vehicle, especially that application of the efficient decision starts from Q-learning.

3. REINFORCEMENT LEARNING APPLICATION FOR 6G V2X SYSTEM

This study includes the designing of an airfraft hangar for airbus A-380 as a pre-engineered steel framed building using indian standard code (IS-code) willing to errect this structure in future if the airbus A-380 is bring to india as a own indian aircraft for indians travelling from own country to other countries

In this section, we investigate the driving strategy decision enabled by reinforcement learning policy which is an efficient tool that is able to handle handover and the distance separating vehicles in a frequent way, previously applied methods of resource allocation highlight a limit especially regarding a dynamic change of the vehicular network. Subchannel allocation is important to ensure vehicular communication.

Position forecasting depends on previous driving experience using the potential of AI especially in urban areas which help in providing much safety. Reinforcement learning efficient tool manages high complexity of vehicular network. RL is a heuristic tool considered as a problem solver for mobility management which is able to deal with huge amount of data especially following the continuous variability of vehicular nodes. In addition, ML tools enabling any possibility of hidden node detection.

In addition, RL has shown its efficiency especially where authors in Meng et al. (2020) described RL application for both power and data rate adaptation. Therefore, the velocity is optimized according to specific node's behavior. Based on RL policy each vehicular node is assimilated as an agent who changes its behavior after a learning process of vehicular ecosystem considering a set of a collected data related to the vehicular traffic. Each vehicular node is able to predict neighbor's assigned resource and acts based on this.

Figure 1 shows the application of RL into V2X 6G system in which each vehicular node is equipped by one antenna assuming Rayleigh distribution between vehicular nodes and gNodeB.

The interaction between different cars is enabled by different sensors such as Camera, Radar, Lidars, GPS...

The channel model is represented by :

$$\mathbf{Y} = \mathbf{H_v\, x} + \mathbf{n} \qquad (1)$$

being:

Y : the received signal

X : the transmitted signal

H_v: the communication channel

B: the corresponding bandwidth

n : noise

The data rate is expressed by : status. Note that both of the learning mechanisms start by a state leading to an action further. The set of vehicles denoted by N = (N_1, ….N_N) distributed following a Poisson process in which each vehicular node is equipped by one antenna. The V2X 6G communication scenarios are shown in figure 1 where two vehicular nodes are communicating with the help of a gNodeB.

$$\mathbf{R_v = B\,log_2\,(+\,SINR_v\,)} \qquad (2)$$

In which the SINR is defined by:

$$\mathbf{SINR_v} = \frac{\mathbf{Pe_v} + |\mathbf{h_v}|^2}{(\sum_{\mathbf{i \neq v}}(\mathbf{Pe_i}\,|\mathbf{h_v}|^2) + \boldsymbol{\partial}^2} \qquad (3)$$

where :

∂^2: represents the noise power

$\mathbf{h_v}$: channel separating the two vehicular nodes

$\mathbf{Pe_i}$: transmitted power of the ith vehicle different than vth one.

RL application process depends on many parameters varying over time in the way that we have a state e_t changing over time, action at and reward R due to environmental dynamic variability. The learning process is one instantaneously defined by a specific state e_t leading to an action a_t.

Hence, the reward corresponding to the global data rate is denoted by R(e_t,a_t,e_t+1).

The purpose is reward maximization by applying Q-Learning policy. The reward in our investigation corresponds to the data rate considering a downlink scenario. We assume similar reward for all participating nodes.

Therefore, the analysis is performed considering full channel state information (CSI). Each vehicular node denoted by V is defined by a position (x,y), a random velocity ve . The Vehicular decision to take facing another vehicular neighbor behavior is enabled with the help of Q-Learning. The purpose of RL application is to generate the optimal policy for reward maximization. The decided action is selected in the way to maximize the reward. Note that the goal is about data rate improving in which the policy P depends on the Q function such as:

$$\mathbf{P^*(e_t) = argmaxQ(\,e_t, a_t\,)} \qquad (4)$$

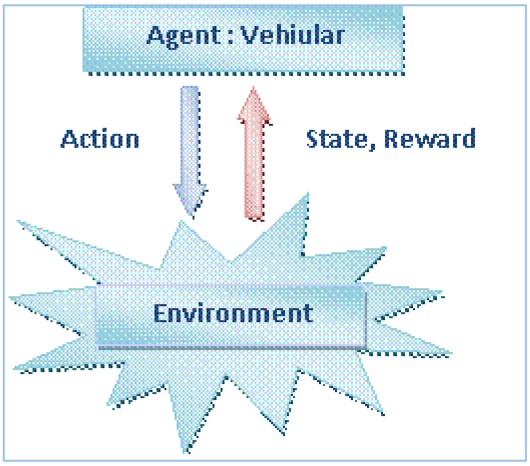

Figure 1. RL application into V2X 6G system

Q-Learning policy is proposed to determine the optimal policy for reward maximization such that following a state s_t and action at we have Q(s_t,a_t). Following RL process any agent behavior depends on environment reward (Figure 1). Thus the players of RL application are the vehicular ecosystem and the agent. The state gives an idea about the traffic status, spectrum availability enabling the decision which is based on a Q function.

4. CONCLUSION

Facing the great requirement of a huge communication spectrum together with much reliability we adopt 6G as enabler for advanced services ensuring the support of traffic density increase.

The increase of vehicular density ensures much safety requirement together with much reliability which encourages the adoption of 6G for communication between vehicles. Mobility represents a big challenge that threatens the efficiency of vehicular communication networks.

Mobility together with interference impacts information exchange between vehicles. Mobility management leads to accurate energy consumption. Vehicular behavior forecasting depends on the stored data collected from different sensors leading to efficient spectrum exploitation.

On the other hand, due to the huge demand for quality of service satisfaction AI and ML are necessities ensuring much reliability and safety. According to the efficiency of AI in many applications it becomes very necessary to integrate it in V2X communication. ML is an efficient tool enabling network connectivity enabled by 6G. Following the continuous progress of wireless communication, RL is a good candidate to manage mobility issue of vehicular nodes in proactive way. RL integration into V2X enables new communication scenarios. The combination between 6G features together with RL policy offers advanced driving way.

In this survey the focus was on the application of RL for vehicular network management in the way to enhance connectivity together with reliability. The idea is how to adapt RL policy into V2X context in the way to improve communication reliability especially with the successful increase of nodes' density. In this regard a specific dataset about the V2X ecosystem is required to predict further behavior of each car. The efficiency of a dataset depends on accurate collection of vehicular communication networks such as the car's position, direction, and velocity for this we shared some datasets applied in different V2X scenarios.

References:

Alablani, I. A., & Arafah, M. A. (2022). A new vehicle dataset in the city of Los Angeles for V2X and machine learning applications. *Applied Sciences*, *12*(8), 3751.

Ali, S., Saad, W., Rajatheva, N., Chang, K., Steinbach, D., Sliwa, B., ... & Malik, H. (2020). 6G white paper on machine learning in wireless communication networks. *arXiv preprint arXiv:2004.13875*.

Alkhateeb, A., Charan, G., Osman, T., Hredzak, A., Morais, J., Demirhan, U., & Srinivas, N. (2023). DeepSense 6G: A large-scale real-world multi-modal sensing and communication dataset. *IEEE Communications Magazine*, *61*(9), 122-128.

Balkus, S. V., Wang, H., Cornet, B. D., Mahabal, C., Ngo, H., & Fang, H. (2022). A survey of collaborative machine learning using 5G vehicular communications. *IEEE Communications Surveys & Tutorials*, *24*(2), 1280-1303.

Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., ... & Beijbom, O. (2020). nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11621-11631).

Choi, Y., Kim, N., Hwang, S., Park, K., Yoon, J. S., An, K., & Kweon, I. S. (2018). KAIST multi-spectral day/night data set for autonomous and assisted driving. *IEEE Transactions on Intelligent Transportation Systems*, *19*(3), 934-948.

Colyar, J., & Halkias, J. (2006). *Us highway i-80 dataset: Tech. rep. fhwahrt-06-137*. Tech. Rep.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3213-3223).

Geiger, A., Lenz, P., & Urtasun, R. (2012, June). Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition* (pp. 3354-3361). IEEE.

Hernangómez, R., Geuer, P., Palaios, A., Schäufele, D., Watermann, C., Taleb-Bouhemadi, K., ... & Stańczak, S. (2023, June). Berlin V2X: A machine learning dataset from multiple vehicles and radio access technologies. In *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)* (pp. 1-5). IEEE.

Hernangómez, R., Geuer, P., Palaios, A., Schäufele, D., Watermann, C., Taleb-Bouhemadi, K., ... & Stańczak, S. (2023, June). Berlin V2X: A machine learning dataset from multiple vehicles and radio access technologies. In *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)* (pp. 1-5). IEEE.

Hernangómez, R., Palaios, A., Watermann, C., Schäufele, D., Geuer, P., Ismayilov, R., ... & Stańczak, S. (2024). Toward an AI-Enabled Connected Industry: AGV Communication and Sensor Measurement Datasets. *IEEE Communications Magazine*, *62*(4), 90-95.

Houston, J., Zuidhof, G., Bergamini, L., Ye, Y., Chen, L., Jain, A., ... & Ondruska, P. (2021, October). One thousand and one hours: Self-driving motion prediction dataset. In *Conference on Robot Learning* (pp. 409-418). PMLR.

Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., ... & Yang, R. (2018). The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 954-960).

Irio, L., Ip, A., Oliveira, R., & Luís, M. (2021). An adaptive learning-based approach for vehicle mobility prediction. *IEEE Access*, *9*, 13671-13682.

Lee, S., Kim, J., Shin Yoon, J., Shin, S., Bailo, O., Kim, N., ... & So Kweon, I. (2017). Vpgnet: Vanishing point guided network for lane and road marking detection and recognition. In *Proceedings of the IEEE international conference on computer vision* (pp. 1947-1955).

Lee, T. K., Wang, T. W., Wu, W. X., Kuo, Y. C., Huang, S. H., Wang, G. S., ... & Tseng, Y. C. (2019, September). Building a V2X simulation framework for future autonomous driving. In *2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)* (pp. 1-6). IEEE.

Li, Y., Ma, D., An, Z., Wang, Z., Zhong, Y., Chen, S., & Feng, C. (2022). V2X-Sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving. *IEEE Robotics and Automation Letters*, *7*(4), 10914-10921.

Matuszka, T., Barton, I., Butykai, Á., Hajas, P., Kiss, D., Kovács, D., ... & Varga, B. (2022). aimotive dataset: A multimodal dataset for robust autonomous driving with long-range perception. *arXiv preprint arXiv:2211.09445*.

Meng, F., Chen, P., Wu, L., & Cheng, J. (2020). Power allocation in multi-user cellular networks: Deep reinforcement learning approaches. *IEEE Transactions on Wireless Communications*, *19*(10), 6255-6267.

Neuhold, G., Ollmann, T., Rota Bulo, S., & Kontschieder, P. (2017). The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision* (pp. 4990-4999).

Noor-A-Rahim, M., Liu, Z., Lee, H., Khyam, M. O., He, J., Pesch, D., ... & Poor, H. V. (2022). 6G for vehicle-to-everything (V2X) communications: Enabling technologies, challenges, and opportunities. *Proceedings of the IEEE*, *110*(6), 712-734.

Patil, A., Malla, S., Gang, H., & Chen, Y. T. (2019, May). The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes. In *2019 International Conference on Robotics and Automation (ICRA)* (pp. 9552-9557). IEEE.

Pham, Q. H., Sevestre, P., Pahwa, R. S., Zhan, H., Pang, C. H., Chen, Y., ... & Lin, J. (2020, May). A* 3d dataset: Towards autonomous driving in challenging environments. In *2020 IEEE International conference on Robotics and Automation (ICRA)* (pp. 2267-2273). IEEE.

Ros, G., Sellart, L., Materzynska, J., Vazquez, D., & Lopez, A. M. (2016). The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3234-3243).

Saikia, P., Pala, S., Singh, K., Singh, S. K., & Huang, W. J. (2023). Proximal policy optimization for RIS-assisted full duplex 6G-V2X communications. *IEEE Transactions on Intelligent Vehicles*.

Salehi, B., Reus-Muns, G., Roy, D., Wang, Z., Jian, T., Dy, J., ... & Chowdhury, K. (2022). Deep learning on multimodal sensor data at the wireless edge for vehicular network. *IEEE Transactions on Vehicular Technology*, *71*(7), 7639-7655.

Santana, E., & Hotz, G. (2016). Learning a driving simulator. *arXiv preprint arXiv:1608.01230*.

Schafer, H., Santana, E., Haden, A., & Biasini, R. (2018). A commute in data: The comma2k19 dataset. *arXiv preprint arXiv:1812.05752*.

Seon, H., Lee, H., & Kim, H. (2023). Predicting CAM generation times through machine learning for cellular V2X communication. *ICT Express*, *9*(5), 841-846.

Stallkamp, J., Schlipsing, M., Salmen, J., & Igel, C. (2011, July). The German traffic sign recognition benchmark: a multi-class classification competition. In *The 2011 international joint conference on neural networks* (pp. 1453-1460). IEEE.

Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., ... & Anguelov, D. (2020). Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2446-2454).

Wang, D., Nganso, Y. N., & Schotten, H. D. (2023, November). A Short Overview of 6G V2X Communication Standards. In *2023 International Conference on Intelligent Communication and Networking (ICN)* (pp. 20-26). IEEE.

Xu, R., Xiang, H., Xia, X., Han, X., Li, J., & Ma, J. (2022, May). Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. In *2022 International Conference on Robotics and Automation (ICRA)* (pp. 2583-2589). IEEE.

Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., ... & Darrell, T. (2020). Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2636-2645).

Yu, H., Luo, Y., Shu, M., Huo, Y., Yang, Z., Shi, Y., ... & Nie, Z. (2022). Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 21361-21370).

Zheng, L., Ma, Z., Zhu, X., Tan, B., Li, S., Long, K., ... & Bai, J. (2022, October). TJ4DRadSet: A 4D radar dataset for autonomous driving. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 493-498). IEEE.

Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., & Hu, S. (2016). Traffic-sign detection and classification in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2110-2118).

Houda Chihi
InnovCOM SupCOM,
Tunisia
houda.chihi@supcom.tn
ORCID: 0000-0002-2974-7680