

Design, Implementation, and Evaluation of a Roadside Cooperative Perception System

Rusheng Zhang¹, Zhengxia Zou¹ , Shengyin Shen² , and Henry X. Liu^{1,2} 

Abstract

Recently, with the advancement in autonomous driving, artificial intelligence, and vehicle-to-everything (V2X) communications, cooperative perception between roadway infrastructure and connected and automated vehicles via V2X communications has attracted increasing attention. However, most of the systems and prototypes reported are still at an initial stage. This paper introduces a newly developed and deployed roadside cooperative perception system with an edge-cloud structure and multiple kinds of sensors, including fisheye cameras, thermal cameras, and long-distance radars. The system is deployed at a roundabout at the intersection of State St. and W. Ellsworth Rd. in Ann Arbor, Michigan, U.S. With an edge-cloud structure, it can execute high-performance detection algorithms on edge devices at the field and collect data and analyze it on the cloud. The performance, in relation to detection rate, localization accuracy, and latency, is analyzed using the data collected from the field. These results show that the system has high potential for a wide range of applications, including infrastructure-vehicle cooperative perception, traffic data collection, and road safety studies. Two example applications—traffic volume monitoring and road safety warning—are introduced in the latter part of the paper to demonstrate the system's capability.

Keywords

autonomous driving perception system, roadside cooperative perception, cooperative autonomous driving system, proxy BSM, V2X communications

Recently, with the advancement in autonomous driving, artificial intelligence, and vehicle-to-everything (V2X) communications, cooperative perception between roadway infrastructure and connected and automated vehicles (CAVs) via V2X communications has attracted increasing attention. Such systems could potentially help CAVs to achieve more robust detection, especially in detecting occluded vehicles and pedestrians, solving the corner cases, and navigating through complicated scenarios.

V2X communications technology enables vehicles equipped with V2X radio (i.e., connected vehicles [CVs]) to share essential information with other road users, as well as roadside infrastructure. SAE J2735 standard gives one example of message sharing, where vehicles share information by periodically broadcasting basic safety messages (BSMs) that contain vehicle position and motion information to their surroundings (1). Such information could significantly help CVs to understand their surroundings. However, to date, only very few cars have V2X communications capability, which

substantially restricts the usage of BSMs. Meanwhile, the low penetration rate of CVs results in slow deployment of V2X communications technologies, as the benefit of equipping the radio is marginal at the current stage.

Such a dilemma is expected to exist for a relatively long time in the future. To overcome this difficulty, roadside-assisted infrastructure-to-vehicle (I2V) messages, in the form of proxy-BSM in the U.S., and proxy-cooperative awareness message (CAM) in Europe and Japan, is proposed (2). These messages are generated by infrastructure and broadcast for the detected vehicle as if the messages are broadcast by themselves. In such a way,

¹Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI

²University of Michigan Transportation Research Institute, Ann Arbor, MI

Corresponding Author:

Henry X. Liu, henryliu@umich.edu

Transportation Research Record
2022, Vol. 2676(11) 273–284
© National Academy of Sciences:
Transportation Research Board 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0361198122109240
journals.sagepub.com/home/trr


the penetration rate of CVs could be boosted by roadside infrastructure at critical locations without modifying the current V2V communications protocols.

To further utilize the infrastructure detections—that is, to shared roadside sensors detected vulnerable road users (e.g., pedestrians and bicycles) information—new V2X protocols are currently under development by the Telecommunications Standards Institute (ETSI) in Europe and the Society of Automotive Engineers (SAE J3224) in the U.S. (3, 4). These standards use new message types known as collective perception messages (CPMs) in Europe or sensor data sharing messages (SDSMs) in the U.S., allowing the presence of objects on the roads detected by onboard sensors or roadside sensors to be shared.

Research into different aspects of cooperative perception systems has been conducted recently (5–8). However, this research and the prototypes reported in the literature are still at an early stage. This paper takes one step further in relation to real-world implementation, and reports a newly developed and deployed roadside cooperative perception system. The system detects, localizes, and tracks vehicles and pedestrians, and shares the information with road users via proxy-BSMs broadcast by a dedicated short-range communication (DSRC) roadside unit (RSU). The system is designed in an edge-cloud structure that performs latency-sensitive tasks (such as detection and tracking) at an edge device, and uploads the data to the cloud for other latency-tolerant applications, both offline and online.

The main contributions of this paper are as follows:

1. Present a fully established, deep learning (DL) empowered roadside cooperative perception system, and implement it at a two-lane roundabout located at the intersection of State St. and W. Ellsworth Rd. in Ann Arbor, Michigan.
2. Evaluate the performance of the perception system in relation to delay, detection accuracy, and localization accuracy.
3. Present a new semi-automatic continuous model retraining scheme that updates the perception system during the deployment.
4. Introduce some preliminary applications that demonstrate the potential of this system.

This paper is organized as follows: in the next section, related works are discussed; system design and algorithms are then discussed in the section after that; followed by field experiments on the detection accuracy and system latency in the following section; in the penultimate section, the system's potential is demonstrated with two example applications, namely traffic volume monitoring and real-time road safety warning; in the

final section, the paper is concluded by discussion of future research directions.

Related Works

Roadside-sensor-based surveillance systems have a long history and can be traced back to as early as 1986. Early systems aim at traffic monitoring to detect abnormal behavior of vehicles (9). Since then, roadside surveillance/perception systems have been developed quickly. Many such systems are vision-based, by mounting one or multiple cameras at a highly elevated position on the roadside to detect and track moving objects on the road. To detect road objects, different methods have been used: background subtraction, frame difference, feature-based detection, Kanade-Lucas-Tomasi tracking, cascading classifiers, and many more (9–14). Recently, with the fast advancement in DL, many newly developed vehicle detection systems adopt DL-based vehicle detection and tracking algorithms. For instance, Zhang et al. introduce a vehicle detection method for unmanned aerial vehicles (UAV) based on the faster-RCNN method (15). Aboah introduces a method that detects vehicles from the roadside, based on the YOLOv5 detector, and finds anomaly behavior using a decision tree (7). In general, DL-based methodology for vehicle detection is at the initial stage, but has significant potential, thereby attracting increasing attention.

The aforementioned systems use regular cameras and have many disadvantages, such as low viewing angle, poor detection at night, and privacy violation. To address these concerns, systems with other types of camera are investigated, for instance, fisheye cameras to enhance per-camera coverage and thermal cameras to increase the robustness at night and in different weathers (16–18).

While most of the systems above are proposed for traffic surveillance—namely, offline usages—recently, with the advancement of V2X communications, roadside perception systems are proposed to help CAVs with cooperative perception in a real-time fashion. Rauch et al. provide a design of using a V2X communication system for cooperative perception, report experimental results conducted by 802.11p radios, and measure the transmission latency of CAMs (19). Their research has found that the transmission latency of CAMs is within 20 ms, but the overall latency can be as large as 200 ms considering the idle time between periodical transmission. In Rauch et al., a vehicle-roadside cooperative perception system is proposed with a high-level fusion strategy (20). In Tsukada et al., the authors report the design of a communication system that broadcasts proxy-CAM to CVs and some initial field-test results on communication reception rate and latency (21). The work is continued in

Tsukada et al., where a roadside perception unit is implemented with open-source autonomous driving software, inter-connected with RSU and sensors with high-speed networks (6, 22). Thorough latency analysis is carried out using OMNET++ and field experiments. In Shan et al., a roadside perception system with lidar and dual cameras is prototyped and demonstrated with CAVs self-driving using the roadside perception information only (23).

This paper reports a well-established cooperative perception and surveillance system that stands out from the aforementioned systems in several aspects. First of all, the system reported here is fully established and deployed, and has been robustly running for months. As opposed to normal cameras, the system is equipped with fisheye cameras and thermal cameras to increase coverage as well as robustness against weather and darkness. The choice of sensor combination is also unique. Finally, the system is cloud empowered, and capable of both online and off-line perception, data collection, and analysis.

System Design

Devices

The system consists of sensors for vehicle detection, edge devices for processing the raw sensor data, a V2X RSU for communication with CAVs, and upstream cloud services. How these components are integrated will be further introduced in later sections. In this subsection, these devices are briefly introduced.

Sensors: Figure 1 shows the sensors used in the system. Three different types of sensor are chosen to cooperatively achieve the best detection results: GRIDSMART Bell fisheye camera, FLIR Trafisense2 Dual thermal camera, and Econolite AccuScan EVO Radar.

The GRIDSMART Bell (Figure 1a) camera is a fisheye camera that is required to be pole-mounted at an elevated position and pointed down to the road surface.

The main benefit of the fisheye camera is that it covers a large area in a single distorted image. However, the further from the camera, the larger the distortion, which leads to lower performance in detection and localization results. Therefore, four GRIDSMART cameras are installed on four corners of the designated roundabout so that only the less distorted parts of each camera are used for detection and localization purposes.

FLIR Trafisense2 Dual (Figure 1b) combines thermal and visual imaging technology for vehicle detection at intersections. The sensor produces infrared images as opposed to RGB images, which make it more robust under different weather and light conditions. During the day time, the FLIR thermal camera will perform detection tasks together with GRIDSMART cameras to increase detection accuracy and localization accuracy. As the fisheye cameras perform poorly at night, the FLIR thermal cameras become the main sensor. Like GRIDSMART cameras, FLIR Trafisense2 Dual is also installed on all four corners of the roundabout.

The Econolite AccuScan EVO Radar sensor (Figure 1c) uses the latest forward-fire radar design and technology to achieve superior traffic detection accuracy and reliability. The radar can detect approaching traffic as far as 152 m. Unlike FLIR and GRIDSMART, whose primary task is to “see” vehicles and pedestrians in the detection area, the AccuScan radar faces the far side, aiming to detect approaching traffic from outside the roundabout. Two AccuScan radars are equipped, each for two approaches of the roundabout. At the time when this paper is written, the Accuscan radars are installed at the roadside, but the radar data is not yet utilized for detection. Therefore, the Accuscan device is only introduced in the device and system design sections. The rest of this paper will focus on discussing the results from the GRIDSMART and FLIR cameras. The radar data will be integrated in the near future and reported in a future publication.

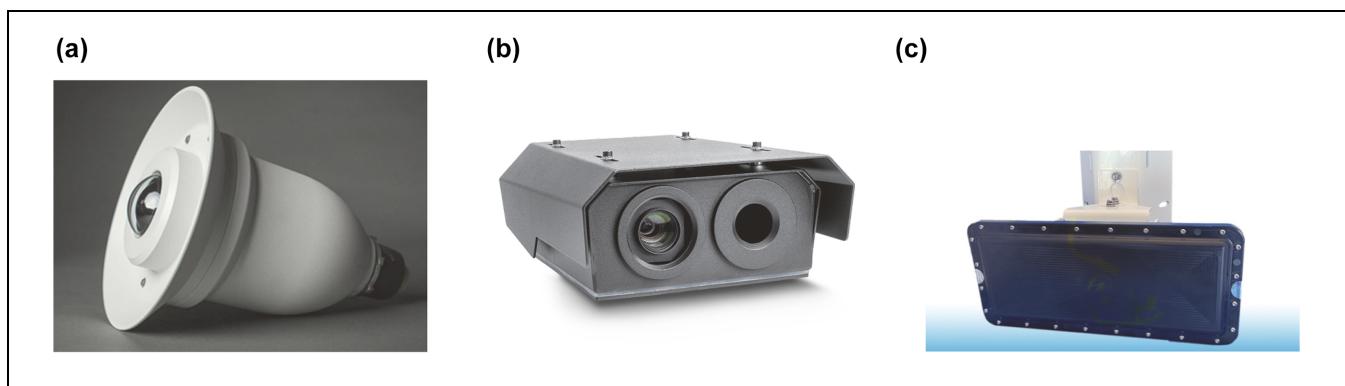


Figure 1. Sensors used on the cooperative perception system. (a) GRIDSMART Bell Camera, (b) FLIR Trafisense2 Dual 632, and (c) Econolite AccuScan EVO Radar.

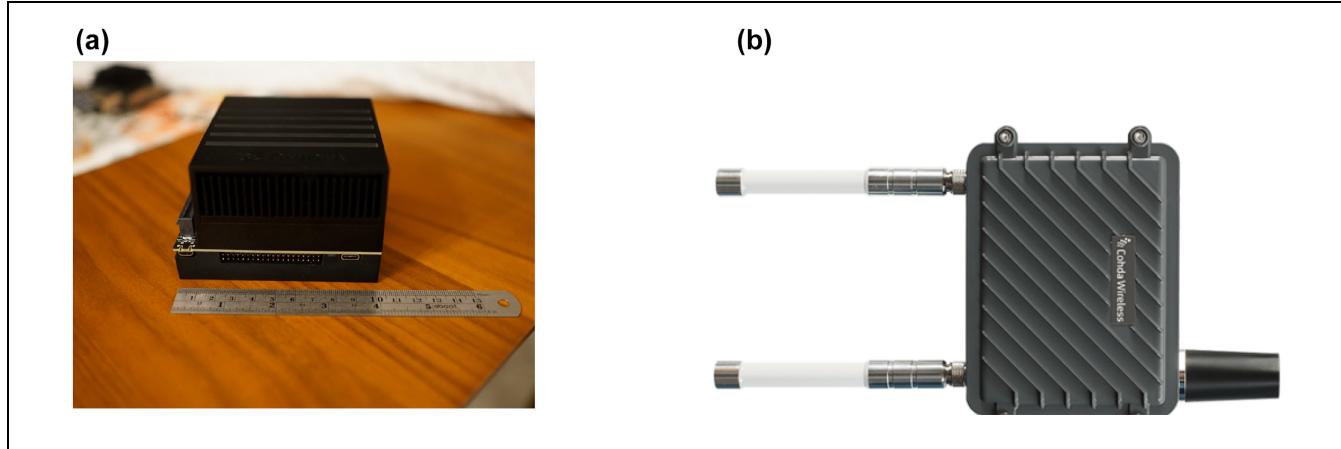


Figure 2. Devices used on the cooperative perception system. (a) Edge device, Nvidia Jetson AGX Xavier and (b) Cohda dedicated short-range communications (DSRC) roadside unit (RSU).

Edge Device: The raw images from the fisheye cameras and thermal cameras (and also the Accuscan radar data that will be ready in the near future) are forwarded to the edge device; the device performs DL-based object detection algorithms to detect, localize, and track vehicles and pedestrians. The edge device is also responsible for communicating with the cloud endpoint to upload images and detection results. Nvidia AGX Xavier (Figure 2a) is chosen for these tasks. It is equipped with a 512-core NVIDIA Volta™ GPU with 64 Tensor cores, optimized for DL algorithms. The per-frame processing time is within 20 ms. Two edge devices are equipped on the roadside, one for the four fisheye cameras and the other for the four thermal cameras.

DSRC RSU: One Cohda MK5 Wireless RSU is installed on the roadside to broadcast proxy BSMs containing perception results to CVs. This DSRC device is able to communicate with CVs on the roadside via 802.11p and SAE J2735 protocols. The MK5 RSU is a rugged outdoor unit with integrated dual antennas housed in a weatherproof enclosure. The MK5 RSU offers large coverage, which can cover all approaches to the roundabout. It uses power over ethernet (PoE) to gain power and ethernet communication with a single cable. The RSU is functional for immediate forward messaging (IMF), which allows an easy implementation of proxy-BSM by directly encoding the BSM outside the RSU and letting the RSU forward it and broadcast on DSRC channels. It is worth mentioning that it is planned to upgrade the V2X communication device to a dual-stack device that is capable of handling both DSRC and C-V2X technology, with the same IMF interface.

Upstream Data Service: A cloud server set up on Amazon Web Services (AWS) is deployed for data collection. An AWS Elastic Compute Cloud (EC2) instance

is deployed as the server that receives raw images and detection results forwarded from the edge device. The images will be stored in AWS S3 storage, and the data will be stored in a MySQL database set up on AWS Relational Database Service (RDS). Micro-services-based applications that interact with the images and database are set up on the cloud using AWS Lambda and operated in a serverless manner.

Installation

The system is installed at a roundabout at the intersection of State Street and W. Ellsworth Rd., Ann Arbor, Michigan, U.S. Figure 3a shows the roundabout. Sensors are installed on four light poles located at four corners of the roundabout: on each light pole, a GRIDSMART fisheye camera and a FLIR thermal camera are mounted. On the south-east and north-west poles, AccuScan long-distance radars are mounted. Figure 3b shows a photo of the light pole mounted with all the devices. The Cohda MK5 RSU is also mounted on this pole. Figure 3a shows the four installation locations and the coverage of the sensors. The AccuScan radar is facing the far side of each approach, to detect incoming traffic. The FLIR and GRIDSMART cameras cover the roundabout area. The sensor data are streaming via Ann Arbor city's traffic network/CV network to the edge device mounted in the traffic control cabinet.

System Integration

As shown in Figure 4, the overall system contains two main parts: the local part (residing on the roadside), and the cloud part. In general, the edge device on the roadside executes latency-sensitive applications, such as

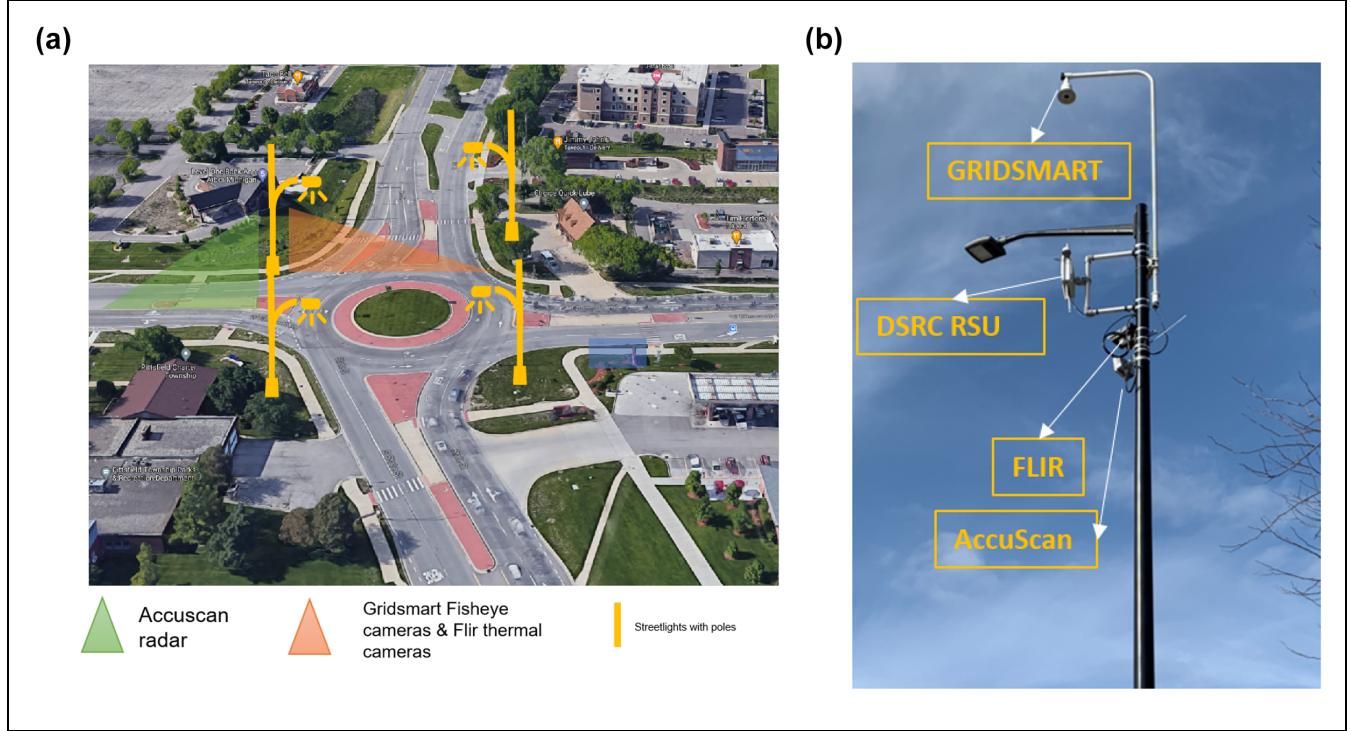


Figure 3. Illustration of how the sensors are installed. (a) The installation locates at a roundabout at the intersection of State Street and W. Ellsworth Rd., Ann Arbor, Michigan, U.S. Four poles at four corners are equipped with GRIDSMART camera, FLIR cameras, and AccuScan radar to achieve best coverage, and (b) Sensors installed on a pole at a corner of the roundabout. There are four poles like this at four corners of the roundabout.

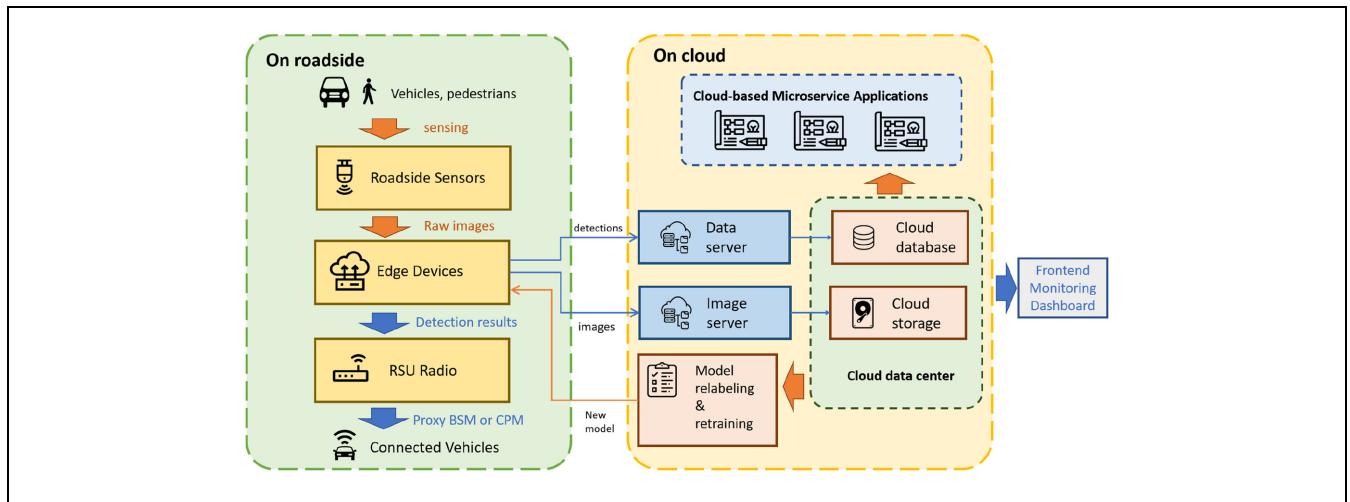


Figure 4. System design of the roadside perception system and its applications.

autonomous driving perception assistance, proxy BSM forwarding, and road safety alert; the sensor data is forwarded to the upstream cloud server, where the data gets further processed.

Roadside system: The local system consists of a roadside sensor, an edge device, and a V2X RSU radio. The sensors detect vehicles and pedestrians, the edge device processes the raw sensor data (here, including the

GRIDSMART fisheye camera images, FLIR thermal images and in the future, Accuscan radar data), and the RSU radio broadcasts the detection results to the CAVs. One of the possible ways to achieve such a task is by using **Proxy-BSM**, in which RSUs broadcast BSMs in the same format as if these BSMs are broadcast from the detected vehicles.

Roadside-cloud data flow: There are two data flows going from the roadside edge device to the server side. The raw images are forwarded to the image server endpoint that stores the images into a cloud storage service (referred to as image flow) and the detection results are forwarded to the cloud data server endpoint, which pushes the results into a cloud-based database (referred to as data flow). The image flow and data flow are correlated, meaning each detection is bounded with each image and vice versa. The image and detection results are further analyzed to relabel the objects in the misdetected images. The detection model will be retrained on the cloud and updated from the cloud to the edge device at the roadside to improve the detection performance.

Cloud system: A cloud-based data processing system is designed and implemented to process large amount of data. The cloud system consists of three main parts: the endpoint servers that receive data flow and image flow; the data center that stores images and detection results; and the cloud-based applications. The applications are implemented in a micro-service manner. Therefore, the system is highly maintainable, scalable, and robust.

Detection, Tracking and Fusion Algorithms

A fast one-stage detection algorithm is developed, inspired by Shan et al. (24). Figure 5 shows the main structure of the algorithm. The model uses **MobileNetV2 as backbone** (25). A decoder is used to decode the feature vector into the same pixel size as the input frame, in a **feature-pyramid-fusion (FPF)** manner, to predict the pixel-wise objectness of the bottom center for vehicles and pedestrians (26). The predicted bottom center's pixel coordinates are then mapped to real-world longitudes

and latitudes using a static mapping. The designed algorithm is specifically designed for accurate localization. The model is computationally light-weight and able to execute on the edge device (NVIDIA AGX Xavier) within 20 ms per frame, enabling the edge device to process four streams simultaneously. The localization results are then forwarded to a **SORT multi-object tracker (MOT)** to assign a unique object ID to the same object across frames (27). The perception system outputs all detected objects' geographic coordinates together with the ID assigned by the tracker periodically (every 100 ms). The output of the perception system can then be forwarded to CAVs via either proxy BSMs or CPMs.

The detection and tracking results are sent to the cloud, and an online application then automatically locates the misdetected objects in the images by analyzing the vehicle trajectories. Those objects are then re-labeled and the model is retrained and updated periodically. The model's performance is expected to improve as such iteration continues. The detection and localization results reported in this paper are on the model updated in one such iteration (namely, one retrain).

Experiments and Results

Detection Accuracy

The classification accuracy is evaluated over 160 randomly selected fisheye images. The detector is trained with 1,000 labeled fisheye images, with one iteration of negative sample retraining described in section System Design. Figure 6 shows the detection results of both fisheye images (Figure 6a) and thermal images (Figure 6b). The detection results on these 160 images are manually evaluated by counting the total vehicles, number of mis-detections, and false positives in these images. The mis-detection rate is defined as $\frac{\# \text{misdetections}}{\# \text{totalobjects}}$, and the false positive rate is defined as $\frac{\# \text{falsepositives}}{\# \text{totalobjects}}$.

There are, in total, 1,422 vehicles in all these 160 images. All images used in this evaluation are new images that the detector has never seen in the training. It was

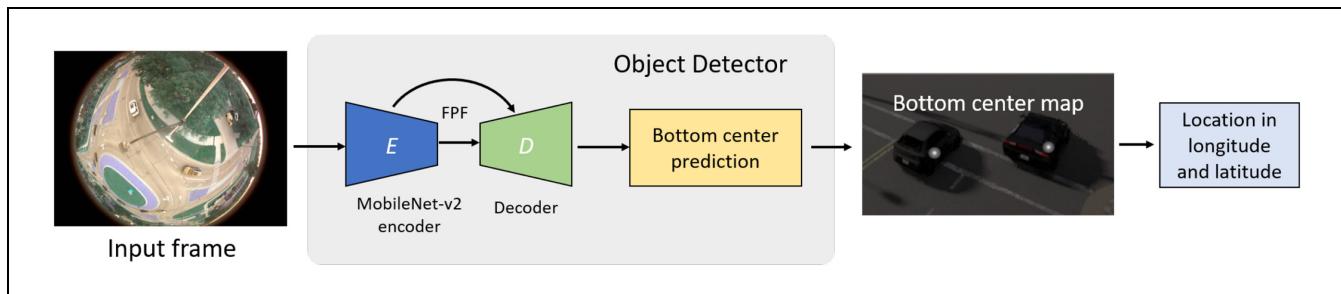


Figure 5. A simple flow chart showing the main structure of the detection and localization model.

Note: FPF = feature-pyramid-fusion.

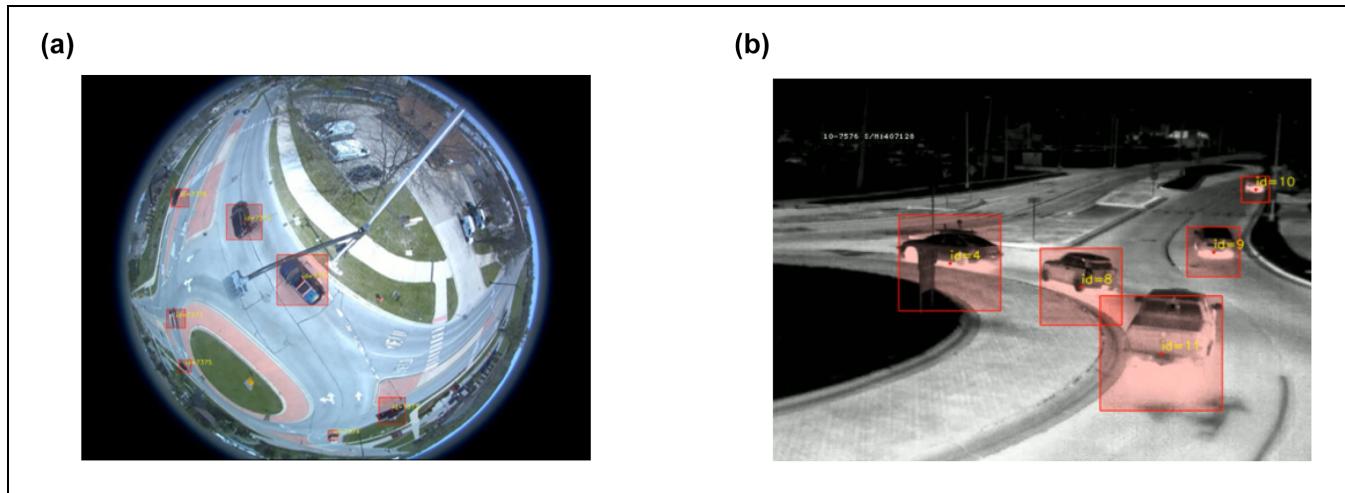


Figure 6. Examples of detection results. (a) Example of detection results on fisheye images and (b) example of detection results on thermal images.

found that there were 49 misdetection cases and 1 false positive. Therefore, the misdetection rate is 3.45%, and the false-positive rate is 0.07%.

There are also some interesting qualitative observations on these results. The false-positive cases happen very rarely—in fact, the only case in all of these 160 images happens when the detector detects a long truck as two vehicles (whether this should be considered as false positive is debatable). Therefore, the false-positive rate can be considered to be much lower than the false-negative (misdetection) rate. Another interesting observation is that most of the misdetection cases happen on the far side, where the fisheye image is significantly distorted. Such an observation indicates that the misdetection rate can be further reduced after fusion of all four fisheye cameras and the use of only the less distorted area of each camera images in future work.

Better results are expected with more retraining iterations in the future when more labeled images are available. Nevertheless, this preliminary model trained with only 1,000 images already shows its potential. With such a small training dataset, the model's performance is relatively satisfactory, which indicates the advantage of training a perception model on a stationary sensor.

Latency and Localization

The cooperative perception pipeline can be broken down into several steps:

1. **Sensor processing:** The sensor senses and prepares the data. This is from sensing the real-world until the data is ready (e.g., until the image is prepared).

2. **Sensor sending data to edge device:** After the data is ready, the sensor sends the data to the edge device through the network.
3. **Perception algorithm:** The edge device processes the sensed data, running perception algorithms such as detection, tracking, fusion, and so forth.
4. **Encoding message:** Encoding the perception result from the last step into V2X messages.
5. **Sending to vehicle:** V2X radio forwards the encoded message from the last step to the vehicle side.
6. **Decoding message:** On the vehicle side, the onboard unit decodes the message and gets perception results.

Figure 7 shows the detailed breakdown of the pipeline. As shown in the figure, three phases for this pipeline are defined. Phase 1 includes sensor processing time and network delay. This is the time elapsed from when the scenario happens until the perception algorithm receives the scenario. The latency caused by phase 1 is mainly from the sensor processing time, as the network delay is lower than 20 ms. Both GRIDSMART and FLIR sensors provide Real-Time Streaming Protocol (RTSP) as the sensor interface to obtain raw image data. Such protocol is used for generic real-time streaming, and, thus, will not annotate each frame's timestamp. Therefore, phase 1 latency will not be directly obtained by the system itself. A field test is carried out to measure this value, which will be further discussed in the following paragraphs. Phase 2 refers to the perception process on the edge device, which is the core algorithm running time. This latency can be easily obtained by comparing the timestamps before and after running the perception algorithm. Phase 3 includes encoding the perception results,

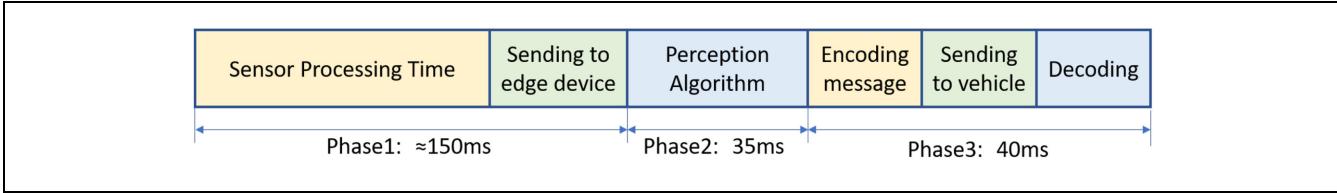


Figure 7. Delay measurement and breakdown.

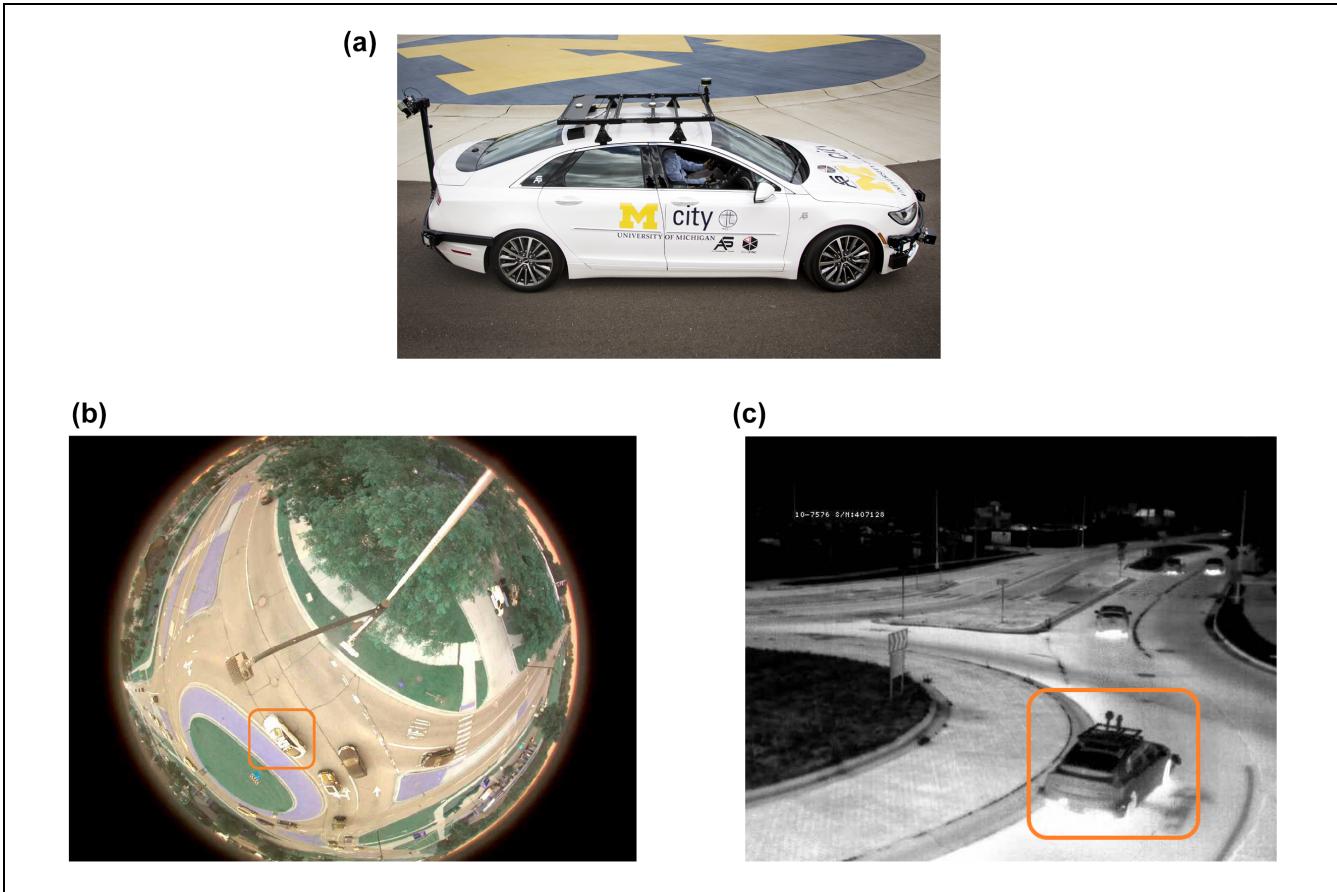


Figure 8. Experimental autonomous driving vehicle with RTK GPS and its detection image in GRIDSMART and FLIR sensors.
(a) Experimental autonomous driving car, (b) experimental autonomous driving car in GRIDSMART fisheye camera, and (c) experimental autonomous driving car in FLIR thermal image.

broadcasting to road users, and decoding at the road user's end. This latency can be measured by comparing the timestamp at the end user's end with the message's timestamp.

The field test is carried out using a Hybrid Lincoln MKZ autonomous vehicle, equipped with a high-precision RTK 3003 module from Oxford Technical Solutions and an inertial measurement unit (IMU). These sensors make it possible to read the vehicle's precise position in a timely manner. The RTK GPS readings are considered as ground truth in the localization and latency evaluation. Figure 8 shows the vehicle used

(Figure 8a), and its image in fisheye (Figure 8b) and thermal image (Figure 8c), in orange boxes.

Six driving trajectories were completed on two separate days: June 30, 2021, and July 19, 2021. All six trajectories are visualized in Figure 9. With the detected vehicle trajectories plotted in blue and ground truth obtained by RTK GPS in red, the localization accuracy and latency results are listed in Table 1. The localization accuracy has an average of 0.2 m, which indicates a lane-level accuracy most of the time.

Meanwhile, phase 2 and phase 3 latency are stable and low; the total time cost is 80 ms, at the same level as

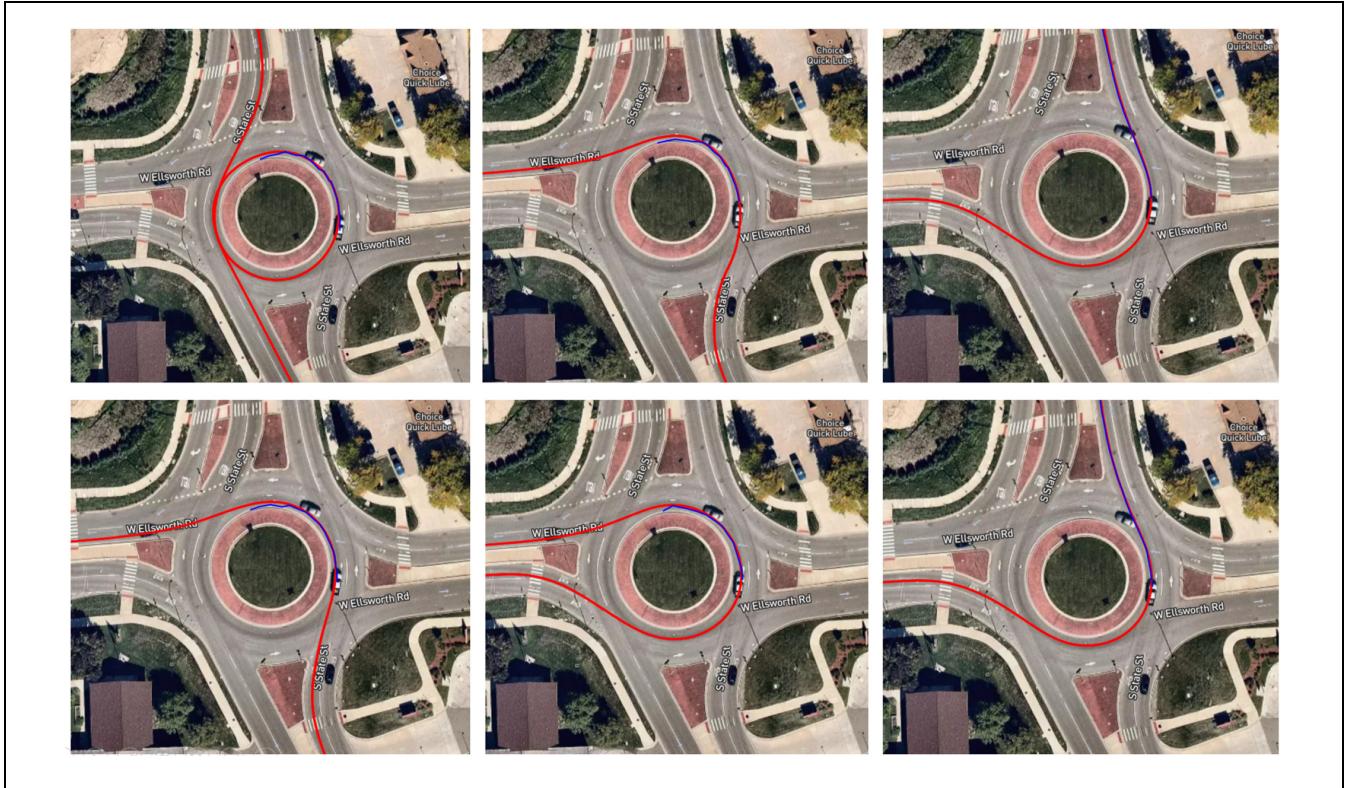


Figure 9. Plotted trajectories showing field-test results

Note: blue line = the detected trajectory; red line = the RTK GPS trajectory (ground-truth).

Table I. Localization Error and Latency in Each Field Test

	Traj. 1	Traj. 2	Traj. 3	Traj. 4	Traj. 5	Traj. 7	Overall average
Localization error (m)	0.121	0.103	0.596	0.155	0.139	0.501	0.269
Phase 1 latency (ms)	108.15	104.84	202.97	119.28	103.85	256.87	149.33
Phase 2 latency (ms)	32	35	33	42	39	32	35.5
Phase 3 latency (ms)	31	45	38	44	37	42	39.5
Total latency (ms)	171.15	184.84	288.97	205.28	179.85	330.87	224.33

Note: Traj. = trajectory.

an onboard perception process time. It is worth mentioning that phase 3 latency is the overhead between roadside perception and vehicle onboard perception, since an onboard perception system will also have phase 1 and phase 2. Phase 3 latency was managed as low as 30–40 ms, which shows the potential of the roadside perception system. The overall latency, on average, is around 224.33 ms. Such latency mainly comes from phase 1 sensor processing time. As one of the future works, the authors will work with the manufacturer to further reduce phase 1 latency.

Example Applications

This section introduces some applications based on the system reported in this paper to show the system's

potential. Two example applications are presented: traffic volume monitoring, and post-encroachment time (PET)-based safety alert.

Traffic Volume Monitoring

An application is created that collects traffic volume information at the roundabout, as traffic volume information is critical to many applications of traffic engineering. The volume information service is implemented on a cloud-based micro-service and is executed on a daily schedule.

Figure 10 shows two typical traffic volume patterns at the roundabout. Each bin in the histogram represents the number of vehicles crossing the roundabout during 15 min. Figure 10a shows typical weekday traffic with a

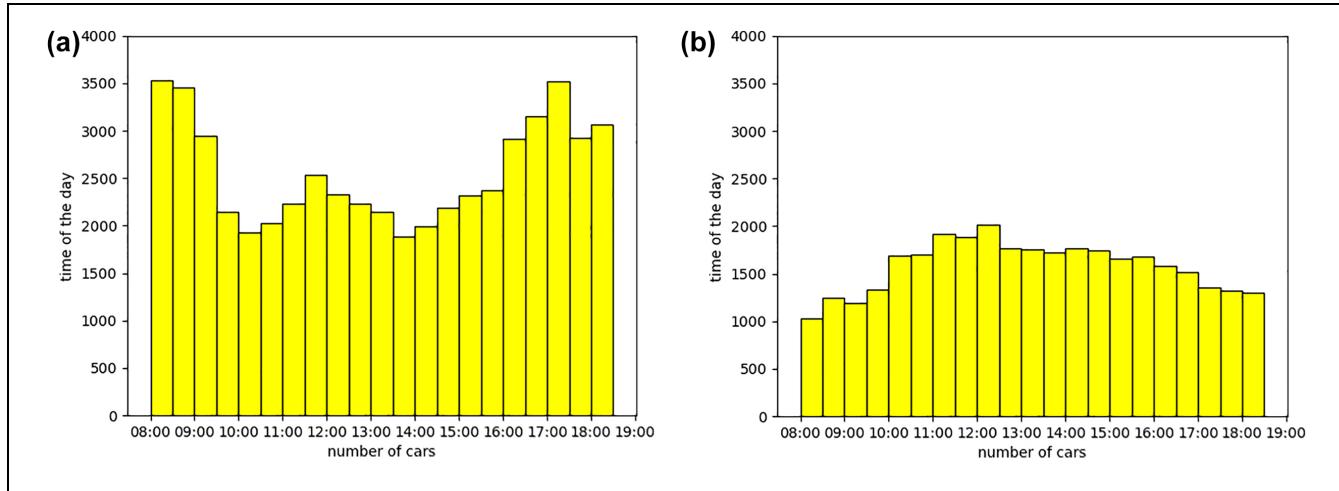


Figure 10. The traffic volume captured by the system. (a) Traffic volume at the roundabout on a weekday, with a total of 53,821 vehicles, and (b) traffic volume at the roundabout on a weekend day, with a total of 33,185 vehicles.

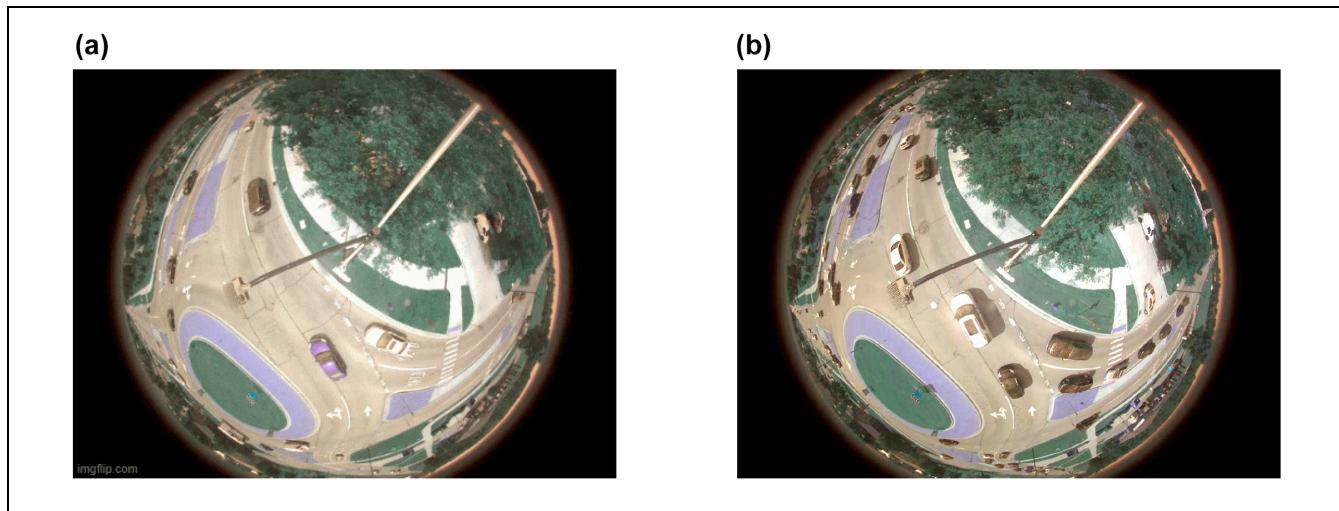


Figure 11. Near-crash scenario found from post-encroachment time (PET) alert. (a) A near-crash scenario that triggers PET alert, with PET of 0.6 s and (b) a near-crash scenario that triggers PET alert, with PET of 0.65 s.

total of 53,821 vehicles on Tuesday, July 20, 2021; Figure 10b shows typical weekend traffic with a total of 33,185 vehicles on Saturday, July 24, 2021. Thus, it can clearly be seen that the weekday traffic volume is 1.62 times larger than the weekend traffic volume, with a very different distribution over time of day.

Post-Encroachment-Time (PET) Detection and Safety Alert

PET is a safety measurement that has been frequently used to detect traffic accidents and near-crash scenarios (28). PET is defined as the time from the end of

encroachment of the lead vehicle to the beginning of encroachment of the second vehicle to the potential point of collision (of a conflict). A micro-service is deployed to calculate PET online at the critical merging area of the roundabout, where the accident rate is high; a violation occurs if the measured PET is below a pre-determined threshold.

Figure 11 shows the cases when the PET are low and triggers the near-crash alarm, the PET in Figure 11a is 0.6 s and the PET in Figure 11b is 0.65 s. In both cases, the vehicle from the merging side did not stop and yield to the vehicle already in the roundabout, creating a near-crash scenario. A PET alert will be able to capture such behavior in real time.

Conclusion

This paper reports the design, implementation, experiment results, and application of a newly developed cooperative perception system for real-time application and data collection at a roundabout at the intersection of State St. and W. Ellsworth Rd., Ann Arbor, Michigan, U.S. The system is designed to use edge devices to do the low-latency roadside perception and communicate the data gathered from the roadside to the cloud for more cloud-based applications. The hard negative samples are then automatically detected and manually relabeled; then, the DL-based detector is retrained and redeployed, making the system evolve and improve during the deployment.

Experimental results have shown the accuracy of the current detector is more than 97%, the localization error is near 0.2 m, the perception algorithm takes a time within 40 ms, and V2X takes a time of 30–40 ms, making it capable of several autonomous driving perception applications. Two example applications—traffic volume monitoring and road safety warning—are also presented to show the system's capability. The end-to-end delay of the system is around 200 ms, mainly coming from sensor processing time, and will be reduced in the future.

Several future works are planned to be carried out to further improve the system's detection accuracy, localization, and latency. A fusion system will be applied on top of the detection results of all the sensors on the roadside, as indicated in Section Detection Accuracy. This will further increase the detection accuracy and reduce misdetection. In addition, such a fusion system is also expected to improve localization accuracy. More applications will also be developed and tested, especially for CAV cooperative perception. The results of the fusion system performance will be included in a future publication. Another important future plan is to update the DSRC RSU to a dual-stack RSU that supports both DSRC and C-V2X communications. Last, but not least, the authors will work with manufacturers to improve the sensor processing time, to reduce end-to-end delay.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: R. Zhang, Z. Zou, S. Shen, H. Liu; data collection: R. Zhang, Z. Zou, S. Shen, H. Liu; analysis and interpretation of results: R. Zhang, Z. Zou, S. Shen, H. Liu; draft manuscript preparation: R. Zhang, Z. Zou, S. Shen, H. Liu. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was partially funded by the Mcity of the University of Michigan, Ann Arbor.

ORCID iDs

Zhengxia Zou  <https://orcid.org/0000-0003-1774-552X>

Shengyin Shen  <https://orcid.org/0000-0001-9946-0071>

Henry X. Liu  <https://orcid.org/0000-0002-3685-9920>

References

1. Kenney, J. B. Dedicated Short-Range Communications (DSRC) Standards in the United States. *Proceedings of the IEEE*, Vol. 99, No. 7, 2011, pp. 1162–1182.
2. Kitazato, T., M. Tsukada, H. Ochiai, and H. Esaki. Proxy Cooperative Awareness Message: An Infrastructure-Assisted v2v Messaging. *Proc., 9th International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*, Kaiserslautern, Germany, IEEE, New York, 2016, pp. 1–6.
3. Thandavarayan, G., M. Sepulcre, and J. Gozalvez. Generation of Cooperative Perception Messages for Connected and Automated Vehicles. *IEEE Transactions on Vehicular Technology*, Vol. 69, No. 12, 2020, pp. 16336–16341.
4. SAE Advanced Applications Technical Committee. *V2X Sensor-Sharing for Cooperative & Automated Driving*. SAE 3224. 2019. <https://www.sae.org/standards/content/j3224/>.
5. Liu, W., S. Muramatsu, and Y. Okubo. Cooperation of V2I/P2I Communication and Roadside Radar Perception for the Safety of Vulnerable Road Users. *Proc., 16th International Conference on Intelligent Transportation Systems Telecommunications (ITST)*, Lisboa, Portugal, IEEE, New York, 2018, pp. 1–7.
6. Tsukada, M., T. Oi, M. Kitazawa, and H. Esaki. Networked Roadside Perception Units for Autonomous Driving. *Sensors*, Vol. 20, No. 18, 2020, p. 5320.
7. Aboah, A. A Vision-Based System for Traffic Anomaly Detection Using Deep Learning and Decision Trees. *Proc., IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, IEEE, New York, 2021, pp. 4207–4212.
8. Chtourou, A., P. Merdignac, and O. Shagdar. Collective Perception Service for Connected Vehicles and Roadside Infrastructure. *Proc., IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, Helsinki, Finland, IEEE, New York, 2021, pp. 1–5.
9. Datondji, S. R. E., Y. Dupuis, P. Subirats, and P. Vasseur. A Survey of Vision-Based Traffic Monitoring of Road Intersections. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 10, 2016, pp. 2681–2698.
10. Furuya, T., and C. J. Taylor. *Road Intersection Monitoring from Video with Large Perspective Deformation*. Ph.D. thesis. University of Pennsylvania, 2014.
11. Messelodi, S., C. M. Modena, and M. Zanin. A Computer Vision System for the Detection and Classification of

- Vehicles at Urban Road Intersections. *Pattern Analysis and Applications*, Vol. 8, No. 1, 2005, pp. 17–31.
- 12. Saunier, N., and T. Sayed. A Feature-Based Tracking Algorithm for Vehicles in Intersections. *Proc., 3rd Canadian Conference on Computer and Robot Vision (CRV'06)*, Quebec, Canada, IEEE, New York, 2006, p. 59.
 - 13. Li, C., A. Chiang, G. Dobler, Y. Wang, K. Xie, K. Ozbay, M. Ghandehari, J. Zhou, and D. Wang. Robust Vehicle Tracking for Urban Traffic Videos at Intersections. *Proc., 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Colorado Springs, CO, IEEE, New York, 2016, pp. 207–213.
 - 14. Faisal, F., S. K. Das, A. H. Siddique, M. Hasan, S. Sabrin, C. A. Hossain, and Z. Tong. Automated Traffic Detection System Based on Image Processing. *Journal of Computer Science and Technology Studies*, Vol. 2, No. 1, 2020, pp. 18–25.
 - 15. Zhang, J. -S., J. Cao, and B. Mao. Application of Deep Learning and Unmanned Aerial Vehicle Technology in Traffic Flow Monitoring. *Proc., International Conference on Machine Learning and Cybernetics (ICMLC)*, Ningbo, China, IEEE, New York, 2017, Vol. 1, pp. 189–194.
 - 16. Wang, W., T. Gee, J. Price, and H. Qi. Real Time Multi-Vehicle Tracking and Counting at Intersections from A Fisheye Camera. *Proc., IEEE Winter Conference on Applications of Computer Vision*, Waikoloa, HI, IEEE, New York, 2015, pp. 17–24.
 - 17. Gasczak, A., T. P. Breckon, and J. Han. Real-Time People and Vehicle Detection From UAV Imagery. *Proc., Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques*, San Francisco, CA, International Society for Optics and Photonics, Vol. 7878, 2011, p. 78780B.
 - 18. Iwasaki, Y., M. Misumi, and T. Nakamiya. Robust Vehicle Detection Under Various Environmental Conditions Using an Infrared Thermal Camera and Its Application to Road Traffic Flow Monitoring. *Sensors*, Vol. 13, No. 6, 2013, pp. 7756–7773.
 - 19. Rauch, A., F. Klanner, and K. Dietmayer. Analysis of V2X Communication Parameters for the Development of a Fusion Architecture for Cooperative Perception Systems. *Proc., IEEE Intelligent Vehicles Symposium (IV)*, Baden-Baden, Germany, IEEE, New York, 2011, pp. 685–690.
 - 20. Rauch, A., F. Klanner, R. Rasshofer, and K. Dietmayer. Car2x-Based Perception in a High-Level Fusion Architecture for Cooperative Perception Systems. *Proc., IEEE Intelligent Vehicles Symposium*, Madrid, Spain, IEEE, New York, 2012, pp. 270–275.
 - 21. Tsukada, M., M. Kitazawa, T. Oi, H. Ochiai, and H. Esaki. Cooperative Awareness Using Roadside Unit Networks in Mixed Traffic. *Proc., IEEE Vehicular Networking Conference (VNC)*, Los Angeles, CA, IEEE, New York, 2019, pp. 1–8.
 - 22. Tsukada, M., T. Oi, A. Ito, M. Hirata, and H. Esaki. AutoC2X: Open-Source Software to Realize V2X Cooperative Perception Among Autonomous Vehicles. *Proc., IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, Victoria, BC, IEEE, New York, 2020, pp. 1–6.
 - 23. Shan, M., K. Narula, Y. F. Wong, S. Worrall, M. Khan, P. Alexander, and E. Nebot. Demonstrations of Cooperative Perception: Safety and Robustness in Connected and Automated Vehicle Operations. *Sensors*, Vol. 21, No. 1, 2021, p. 200.
 - 24. Zhou, X., D. Wang, and P. Krähenbühl. Objects as Points. *arXiv preprint arXiv:1904.07850*, 2019.
 - 25. Sandler, M., A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proc., IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, IEEE, New York, 2018, pp. 4510–4520.
 - 26. Lin, T.-Y., P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature Pyramid Networks for Object Detection. *Proc., IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, New York, 2017, pp. 2117–2125.
 - 27. Bewley, A., Z. Ge, L. Ott, F. Ramos, and B. Upcroft. Simple Online and Realtime Tracking. *Proc., IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, IEEE, New York, 2016, pp. 3464–3468.
 - 28. Wishart, J., S. Como, M. Elli, B. Russo, J. Weast, N. Altekar, E. James, and Y. Chen. Driving Safety Performance Assessment Metrics for Ads-Equipped Vehicles. *SAE International Journal of Advances and Current Practices in Mobility*, Vol. 2, 2020, pp. 2881–2899.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors.