VISION-BASED WORK ZONE SAFETY ALERT SYSTEM IN A CONNECTED-VEHICLE ENVIRONMENT

Haibo Cui
Graduate Research Assistant
Department of Electrical and Computer Engineering
University of Alberta, Edmonton, Canada, T6G 1H9
Email: haibo@ualberta.ca

Kaizhe Hou
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Alberta, Edmonton, Canada, T6G 1H9
Email: kaizhe@ualberta.ca

Jiarui Zhang
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Alberta, Edmonton, Canada, T6G 1H9
Email: jiarui6@ualberta.ca

Siqi Yan
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Alberta, Edmonton, Canada, T6G 1H9
Email: syan3@ualberta.ca

Dr. Mudasser Seraj
Innovation and Applied Research Manager
ATS Traffic, Edmonton AB Canada
Email: rafis@atstraffic.ca

Yingke Wang
Undergraduate Student
Department of Computer Science
University of Toronto, 40 St George St, Toronto, ON M5S 2E4
Email: yingkewang.wang@mail.utoronto.ca

Professor Tony Z. Qiu, Corresponding Author
[1]Department of Civil and Environmental Engineering
University of Alberta. 6-271 DICE. Edmonton. Alberta. Canada T6G 1H9
[2]Intelligent Transportation Systems Research Center
Wuhan University of Technology, Mailbox 125, No. 1040 Heping Road

Wuhan, Hubei, China 430063
1 Primary Affiliation
2 Secondary Affiliation
Email: tony.qiu@ualberta.ca

Professor Mahdi Tavakoli, Corresponding Author
Department of Electrical and Computer Engineering
University of Alberta. 13-356 DICE. Edmonton. Alberta. Canada T6G 1H9
Email: mahdi.tavakoli@ualberta.ca

Word Count: 9824 words + 5 table(s) $\times$ 250 = 11074 words

Submission Date: December 13, 2022

ABSTRACT

Work zones, being a critical component of roadway transportation systems, can benefit greatly from computer vision-enabled roadway infrastructures, specifically in connected vehicle (CV) environments. Connected infrastructures, such as roadside units (RSU) and on-board units (OBU), can greatly improve the environmental awareness and safety of CVs driving through a work zone. In this regard, the contribution of this thesis lies in developing a vision-based approach to generate work zone safety messages in real-time, utilizing video streams from roadside monocular traffic cameras that can be used by CV work zone safety APPs on mobile devices to reliably navigate through a work zone. A monocular traffic camera calibration method is proposed to establish the accurate mapping between the image plane and Global Position System (GPS) space. Real test scenarios show that our algorithm can precisely and effectively locate work zone boundaries within 1 meter from a monocular traffic camera in real-time. We demonstrate the capabilities and features of our system through real-world experiments where the driver cannot see the work zone. End-to-end latency analysis reveals that the vision-based work zone safety warning system satisfies the active safety latency requirements, which is below 100 milliseconds. This vision-based work zone safety alert system ensures the safety of both the worker and the driver in a CV environment.

Keywords: Connected infrastructures, Connected vehicles, Deep learning, Work zone safety, Work zone safety messages

INTRODUCTION

Transportation authorities and the public are concerned about safety in work zones on highways. According to the National Highway Traffic Safety Administration (NHTSA) [1], there was at least one traffic-related injury in 70 work zones daily. Additionally, NHTSA show that work zone collisions have greater mortality rates than crashes outside of them. It's common for employees on highway construction and maintenance projects to be in close proximity to moving traffic. Although many safety precautions are routinely taken to protect workers, these precautions may be insufficient owing to a variety of environmental and human variables, such as distracted driving, bad weather, and poor road conditions.

Work zones typically feature advanced warning zones with visual warning signs to alert oncoming vehicles. Static signs are the most prevalent type of warning system. Furthermore, dynamic warning systems are frequently utilized to improve traffic flow in work zones [2]. Because diverse operating characteristics of arriving vehicles and their relative positioning near a work zone are rarely considered in the design phase of these alerts, they typically fail to adequately aid drivers. Larger vehicles, such as trucks or buses, usually require more space to combine, endangering adjacent vehicles and pedestrians [3, 4]. Furthermore, vehicles behind trucks lose their ability to discern the precise location of lane merges and the proper speed to maintain, increasing the likelihood of rear-end collisions. Rear-end collisions are the most prevalent kind of crash in a work zone's advanced warning area, according to Garber et al. (2002) [5]. According to Garber et al. (2002) [5] and Nemeth et al. (1978) [6], the advanced warning region accounts for 10%-35% of all work zone collisions. Furthermore, comparative accident evaluations conducted by Hall and Lorenz (1989) and Rouphail et al. (1988) [7, 8] demonstrated that rear-end collisions are more likely in work zones than in non-work zones.

Given that the majority of work zone-related events can be avoided with early alerts to vehicles, safety research relating to work zones in a CV environment is an emerging area of research [9]. Through vehicle-to-infrastructure (V2I) communication, enabling a low latency communication channel, such as dedicated short-range communication (DSRC) [10] and upcoming 5G technology, may dramatically improve work zone safety. The drivers of CVs are alerted to potential collision hazards using work zone early warning messages provided by the DSRC or 5G-enabled V2I communication. Through communicated proactive decision-making aids using in-vehicle displays, also known as the Human Machine Interface (HMI), drivers can benefit from increased situational awareness about upcoming hazards or conditions by utilizing wireless communications such as: V2I, vehicle-to-vehicle (V2V), and vehicle-to- everything (V2X). The transmission of early warnings is one benefit that CVs can use. For instance, early in-vehicle lane closure alerts can be used to meet the demands of heavy truck drivers who need to move to the available lane in the work zone, well in advance of the lane closure area [11]. Significantly, a survey conducted by Benekohal et al. (1995) [12] revealed that over half of truck drivers preferred that warning signs for work zones be put up to 3 to 5 miles in advance of the work zone.

In current research studies of work zone alert systems in a CV environment, special equipment including smart traffic cones (traffic cones mounted with a GPS sensor) and wearable localization devices (mobile tags that can be placed on the worker with a localization sensor) are required to localize the work zone boundaries [13, 14]. Such specialized equipment are difficult to deploy and expensive. For instance, all workers or traffic cones are unlikely to attach a localization device similar to a GPS sensor.

To solve the research question of how connected infrastructures can help vehicles to navi-

gate through a work zone, the objective of this study is to provide CVs with work zone safety alerts in real-time. We provide a system for generating work zone safety alerts, utilizing real-time video feeds gathered from traffic cameras, in order to get around these restrictions and improve work zone safety. The contribution of this study is the real-time generation (every 100 milliseconds) of safety warnings and work zone safety messages in accordance with the Society of Automotive Engineers (SAE) J2945 standards [15]. The DSRC-enabled roadside infrastructure generates and broadcasts work zone safety alerts to approaching CVs using work zone information (such as location). No localization device attached to workers or work zone boundaries (traffic cones) are necessary in our work zone boundaries localization method. We evaluate the accuracy of generated work zone safety messages by comparing with field-collected ground truth data. Furthermore, we validate our vision-based approach at the system level in the real-world road environment field test.

The remainder of the paper is organized as follows. We begin by discussing the present research on work zone detection systems, work zone safety-related studies in a CV environment, and work zone safety message standards. We then discuss the real-time vision-based work zone safety message generating method, which is not dependent on work zone localization devices. The evaluation of the vision-based work zone safety message generation is then presented. Then, we discuss the system- level validation utilizing the vision-based system's generated work zone safety alerts. Finally, we go over the study's findings and potential directions for further research.

LITERATURE REVIEW

In this section, we will look at the current literature on work zone detection methods, work zone safety-related research in a CV environment, and work zone safety message requirements.

Work Zone Detection Methods

Existing work zone detection research is primarily concerned with identifying the existence or absence of a work zone. Their detection results are generally vague and lack work zone boundary information. Abodo et al. (2018) finds work zones by employing a convolutional neural network to identify work zone photos [16]. Seo et al. (2015) uses the vehicle's camera to recognize traffic signs [17]. This approach then identifies work zone signs and utilizes them to indicate the starting and ending point of a work zone road segment. Mathibela et al. (2012) also identifies traffic signs and traffic cones [18]. This method employs the detection findings as characteristics to calculate the likelihood that the vehicle is in a work zone. Kunz et al. (2017) creates a Bayesian network for detecting a work zone [19]. The Bayesian network uses identified traffic objects and vehicle statuses as input to forecast the likelihood of a work zone at various distance bins. None of these studies infer detailed geometric properties like work zone areas and borders. Knowing the work zone geometry is crucial for CVs to drive safely.

Graf et al. (2012) investigates a more limited example in which temporary lane markers assist drivers through a work zone [20]. This method monitors the lanes even if both temporary and original lane markers are present. As a result, it allows the vehicle to follow the lanes in a work zone. However, such a strategy is reliant on lane marking regulations and is unable to deal with regular work zones that lack temporary lane markers. Shi et al. (2021) [21] develops work zone detection, which detects and locates the boundaries of a work zone. They give many baseline implementations utilizing various sensor combinations, such as camera and LiDAR. Several cutting-edge deep learning-based object detection strategies, such as the Region-Convolutional Neural Network

(R-CNN) [22], Fast RCNN [23], Faster R-CNN [24], Single Shot MultiBox Detector (SSD) [25], and You Only Look Once - Version 5 (YOLOv5) [26], are used for real-time applications, taking advantage of high-performance GPU-enabled computing devices. The YOLOv5 outperforms all previous state-of-the-art object detection deep learning models [26, 27], with higher detection accuracy and lower detection time. Unfortunately, all of the studies mentioned in this subsection don't have the capability of determining the precise location of a work zone in terms of latitude and longitude in real-time.

In order to address these research gaps in work zone detection methods, we introduce a YOLO-based work zone detection and localization method in real-time. We contribute to the existing work zone detection methods by precisely positioning a work zone area, and then broadcasting work zone safety messages to nearby CVs in a CV environment.

Work Zone Safety-Related Studies in a CV Environment

Scholars began to think about how CVs may be combined with work zone safety as they became more prevalent. Vehicles that are connected can communicate with the driver, other vehicles on the road (V2V), roadside infrastructure (V2I), and the cloud server (V2C) via various communication methods through DSRC, Wi-Fi, and cellular communication technologies [28, 29]. DSRC technology has reduced connection latency less than existing Wi-Fi and Cellular LTE technologies, making it a far faster two-way communication alternative for information sharing [30].

Vehicles may now be viewed as integrated components of a system rather than as independent actors on the road, thanks to the incorporation of communication technologies. When the effects of CVs on safety performance are investigated in a work zone context, Abdulsattar et al. (2018) [31] reveals that V2V/V2I communication can increase work zone safety performance at low traffic flow rates. Genders et al. (2015) [29] also investigated how employing CVs in a network with work zones affected traffic safety. The Michigan Department of Transportation and 3M [32] constructed the first connected work zone in the United States. In this connected work zone setting, orange barrels with 2D barcodes were supplied by 3M, and the CV's infrared devices sent information to the vehicle and the driver by reading the barcode. Han et al. (2019) [13] designs a connected work zone alert system by wearable localization devices that can be placed on workers in work zones. This method monitors the potential danger between CVs and workers by calculating the collision risks from both CVs and workers trajectory. Schonroack et al. (2015) [14] developed a traffic cone with GPS and communication sensors that can be placed at a work zone boundary to give the location of a work zone in a CV environment. In this approach, the location of this special traffic cone is sent to a central server and then broadcasted to nearby CVs. Mishra et al. (2021) [33] developed work zone alert systems from work zone intrusion technologies and Qiao et al. (2017) [34] used cell phone APP to give drivers early warnings of work zones. Development of location sensing technologies and CVs has made it possible to gather data from all parties involved in a work zone, including vehicles, equipment, and workers on foot, and use it to ensure the safety of the work zone. However, current research in the localization of work zones for a CV environment usually requires localization devices, including GPS sensors attached to construction equipment, traffic cones, or workers, which makes it costly and difficult to set up.

In summary, Table 1 shows comparisons of different warning systems in a CV environment and three commercially available warning systems analyzed by Mishra et al. [33] are compared and explained in Table 2. These methods utilize pressure, radar sensors, or additional equipment to recognize incoming traffic. Unfortunately, such equipment usually costs a lot and has poor

mobility. In comparison, our method uses a camera for object detection, which has two main advantages. First, it can recognize various types of objects than traffic, including cones, barrels, vehicles, people, etc. Due to the nature of computer vision, we can feed our YOLOv5 models with as many object types as desired. Second, Our method has a lower expense and higher mobility than other proposed method devices. The only device we need onsite is the camera. The device is easy to obtain, move around, and adapt to work zones with different shapes and environments. For example, most pressure sensors can only adopt in a single-lane work zone.

To overcome limitations of current research in localizing work zones, we provide a framework to increase work zone safety using a vision-based deep learning technique, assuming that workers, construction equipment, and traffic cones do not have a localization device.

Author	**Real-Time**	**Lane Level Localization**	**Special Localization Device**	**Summary**
Han et al[13]	Yes	Separate a single line into four zones for detection.	Wearable localization devices that can be placed on workers in work zones.	Using risk score to calculate the collision risks from both CVs and workers trajectory. The algorithm can also be used in vehicle turning and workers from different categories have their own risk score.
Schonroack et al[14]	Yes	By extending STC using GNSS model, the position accuracy is in the sub-meter range.	Traffic cone with GPS and communication sensors.	The location of these 16 special traffic cones is sent to a central server and then broadcast to nearby CVs.
Qiao et al[34]	Yes	NA	No extra device needed.	Pre-installed application detects the approaching construction zone based on geolocation in phone to give drivers early warnings of work zones.
Islam et al[35]	Yes	Not at lane level.	Site camera and personal safety messages(PSMs) devices carried by roadside users.	Generate PSMs using real-time video streams collected from traffic camera. Train the YOLOv3 model with the video data from a roadway section to detect pedestrian.
Genders et al[29]	NA	NA	NA	Build a control simulation of a network with a work zone to simulate various market penetrations of 20%, 40%, 60%, 80%, and 100% connected vehicles to determine their effect on the safety of the network.

TABLE 1: Comparisons of Recent Studies

Standards for Work Zone Safety Messages

The SAE J2945 standard [15] defines the work zone safety messages for safety data communication between the work zone and other associated components (e.g., vehicle and traffic signals). Although the work zone safety messages standard is defined by SAE J2945, the format and structure of the message, data frames, and data components for sharing data between work zones and vehicles, as well as between work zones and infrastructure, are defined by SAE J2735 [36]. In contrast, the SAE J2945 considers all of the data items established in the SAE J2735. Work zone safety messages data items are shown in Table 3.

Method	Real-Time	Lane Level Localization	Special Localization Device	Note
Intellicone	Yes	Not at lane level.	Need traffic cone mounted sensor and portable site alarm.	Results are inconsistent and not available in US.
Advanced Warning And Reisk Easion	Yes	Lane intrusion system (not yet available for evaluation).	Onsite sentry needed.	Warning equipment worn by personal and sentry in front of work zone.
Worker Alert System	Delayed	Not at lane level.	A pneumatic trip hose senor with a signal transmitter, a Portable Alarm case, personal safety device.	Workers and pedestrians are alerted when sensor on the ground detect vehicle's pressure.

TABLE 2: Commercially Available Methods

Data Element	Purpose
Road Segment	Store complete description including road Geometry, allowed navigational paths, and any current disruptions such as a work zone or incident event.
Road Sign ID	Used to provide a precise location of one or more roadside signs.
Traveler Data	Used to send a single message in traveler information message. It uses the ITIS encoding system to send well-known phrases but allows limited text for local place names.
Public Safety and Road Worker Activity	Used to describe the type of activity a worker or workers are engaged in.
Speed Limit Type	Relates the type of speed limit to which a given speed refers.
Personal Safety Message	Used to broadcast safety data regarding the dynamic state of various types of Vulnerable Road Users (VRU), such as pedestrians, cyclists, or road workers.

TABLE 3: Data Elements for Work Zone Safety Messages Based on SAE J2945 and SAE J2735.

Work Zone Activity Data (WZAD) – Data Dictionary Report [37], developed as part of the Federal Highway Administration's Work Zone Data Initiative Project, defines and standardizes digital descriptions of work zone activities, allowing transportation authorities and third-party providers to describe and communicate work zone information. Work zone activities data structure based on WZAD is shown in Figure 29 in Appendix A.

WORK ZONE SAFETY ALERTS SYSTEMS

We created a system that used a real-time camera feed to produce work zone safety messages by YOLOv5 and provided a safety warning in the event of a probable vehicle-work zone collision. The primary data elements of the work zone safety messages were the work zone geometry, including the work zone starting point, ending point, and lane closures as given in Figure 29. Furthermore, positional accuracy was determined by the precision of work zone positioning information (i.e., longitude and latitude). Thus, after correctly identifying a work zone, the localization of the work zone must also be correct in order to create work zone safety messages. To locate a work zone, we

created a mathematical technique for converting an image's pixel coordinates to global coordinates. After determining a work zone's position, we built the work zone safety messages in accordance with the SAE J2945 standard and WZAD outlined in the preceding section.

In our system, the monocular traffic camera served as our primary data source. The ethernet, which is a wired communication method, transmitted real-time video data to the central server. The video data was then processed by the central server using YOLOv5 to identify traffic cones, which are commonly used to define work zone boundaries, and to convert pixel coordinates to global coordinates. The RSU then received the work zone position information by ethernet and transmitted it to the CV's OBU by DSRC. The HMI developed in this system was a work zone warning APP that could be used on a driver's cellphone or tablet. This APP, connected to OBU's WIFI, was meant to show the driver real-time work zone safety messages. The communication topology in this system and the experiment setup are shown in Figure 30 in Appendix A and Figure 1 respectively.

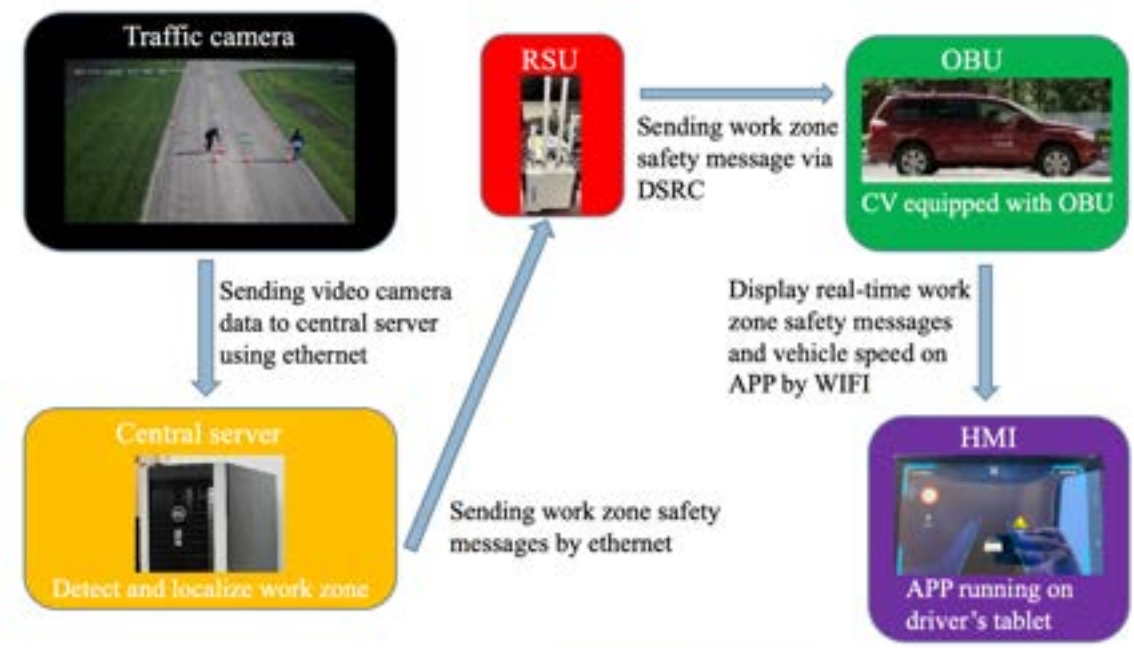
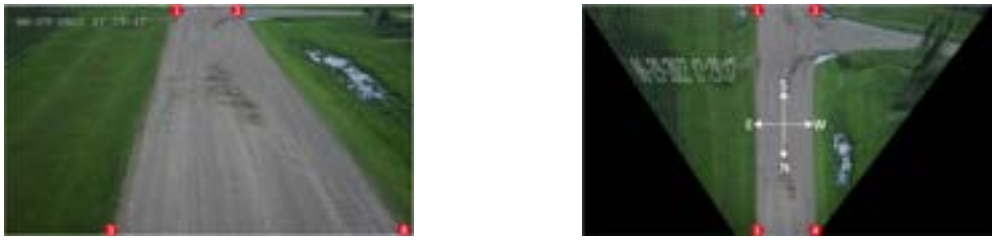


FIGURE 1: Experiment Setup for the Vision-Based Work Zone Alert System in CV Environment.

Localization for work zone boundaries (traffic cones) is vital to generate work zone safety alerts. But normally, monocular traffic cameras have poor localization ability compared to LiDAR sensors. Inspired by other research including Islam et al. and Wen et al. on object localization from monocular traffic cameras [35, 38], we developed a low-cost monocular traffic camera calibration method for work zone boundaries localization without knowing the camera's intrinsic parameters. First, the traffic camera image plane was converted into top-down view by perspective transformation (PT) in Equation 1. PT is a technique to obtain a different view from an image. We developed Equation 2 to convert from pixel coordinates into GPS coordinates by linear transformation (LT). The methodology to localize a work zone from monocular traffic cameras is shown in Figure 31 in Appendix A.

Perspective transformation (PT): convert from pixel coordinates (p_x,p_y) in traffic camera image plane into pixel coordinates (p'_x,p'_y) in top-down view image plane by transformation ma-



(a) The traffic camera image plane before PT. (b) The traffic camera image plane after PT.

FIGURE 2: Perspective Transformation (PT).

trix M shown in Equation 1. Where M is calculated from four pairs of matching points (shown in Figure 2) by python using the function cv2.getPerspectiveTransform.

$$[p_x^{'}, p_y^{'}, 1]^T = M \cdot [p^x, p^y, 1]^T \quad (1)$$

Linear transformation (LT): convert from pixel coordinates ($p^{'}_x$,$p^{'}_y$) in top-down view image plane into GPS coordinates (longitude,latitude) by transformation matrix L shown in Equation 2. Where L is calculated from calibration points with measured ($p^{'}_x$,$p^{'}_y$) in top-down view image plane and GPS coordinate (longitude,latitude) by python using the function np.linalg.lstsq. The detailed steps to calculate L will be covered in data collection section.

$$[longitude, latitude] = [p_x^{'}, p_y^{'}, 1] \cdot L \quad (2)$$

The road touching point of a traffic cone can be approximated by the bottom center of the detected bounding box from YOLOv5 illustrated by the black point in Figure 3.

FIGURE 3: Road Touching Point of a Traffic Cone.

To localize a traffic cone, we took the road touching point (p_x,p_y) from YOLOv5 to obtain the pixel coordinate in top-down view ($p^{'}_x$,$p^{'}_y$) by PT shown in Equation 1, and then used ($p^{'}_x$,$p^{'}_y$) to get the GPS coordinate (longitude,latitude) by LT shown in Equation 2. Then work zone safety alerts were generated from GPS coordinates of traffic cones used to form work zone boundaries.

DATA COLLECTION

Field data is collected for: 1) training and testing of YOLOv5 to ensure a high level of traffic cone detection accuracy; 2) calibrating the distortion in top-down view from PT; 3) calibrating the linear transformation matrix L in Equation (2); and 4) localization error analysis on vision-based work zone items localization method.

Training and Testing YOLOv5

Phase One

In phase one of the experiments, we focused on training and testing YOLO to detect small traffic cones shown in Figure 3.

According to official documentation of YOLOv5 [26], the recommended number of training images to detect an object is 1200 images. We placed traffic cones randomly in the vision range of traffic cameras at 118 Street NW, Edmonton under different weather conditions. The traffic cameras captured 1500 images to form a dataset to train and test YOLOv5. We annotated each traffic cone in every image of the dataset by a bounding box to generate ground truth data. In this dataset, 1200 images were used for training and 300 image were used for testing. Then, some dataset preprocessing techniques were applied to decrease training time and increase performance by applying image transformations to all images in this dataset. The preprocessing techniques [26] used in this dataset were: 1) auto-orientation, to correct a mismatch between the annotation and the image; and 2) reducing the size of the image to train faster throughout the training phase.

Dataset augmentation techniques were then used to create new examples for YOLOv5 to learn from, by generating augmented versions of each image in the training dataset. Figure 4 shows the dataset augmentations used in this dataset. The dataset augmentation techniques [26] used in this dataset were: 1) horizontal flip: to help YOLOv5 be insensitive to subject orientation; 2) crop: to add variability to positioning and size to help YOLOv5 be more resilient to subject translations and camera position; 3) brightness: to add variability to image brightness to help YOLOv5 be more resilient to lighting and camera setting changes; 4) blur: to add random Gaussian blur to help YOLOv5 be more resilient to camera focus; 5) noise: add noise to help YOLOv5 be more resilient to camera artifacts; 6) cutout: add cutout to help YOLOv5 be more resilient to object occlusion; and 7) bounding box brightness: add variability to bounding box brightness to help YOLOv5 be more resilient to lighting on traffic cone.

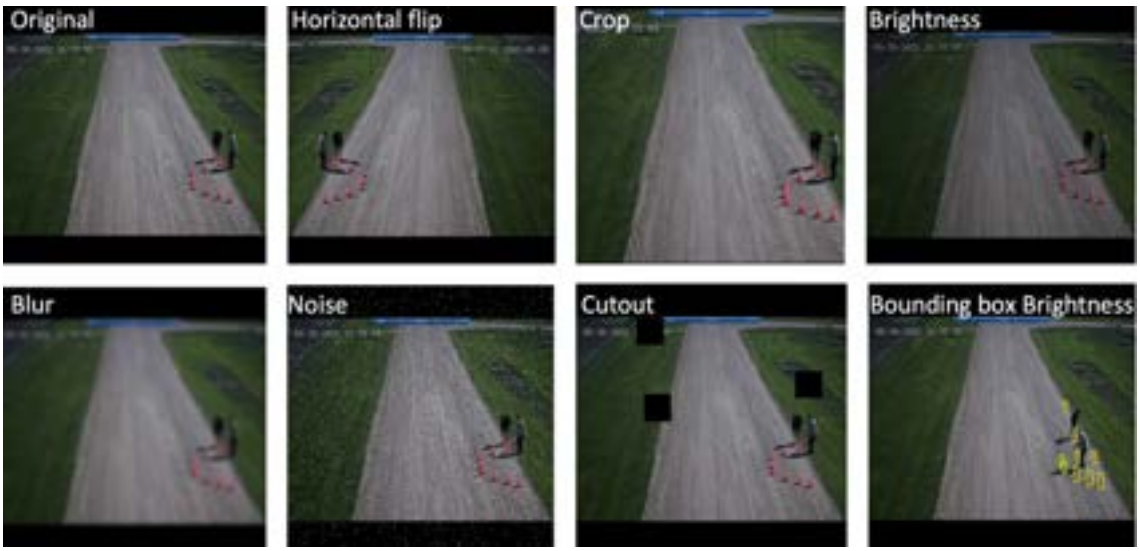


FIGURE 4: Dataset Augmentations Used in This Dataset.

The work zone items detection accuracy was measured using the following parameters and metrics: 1) True Positive (TP): YOLOv5 successfully recognized the presence of a traffic cone; 2)

False Positive (FP): YOLOv5 incorrectly recognized the presence of a traffic cone; and 3) False Negative (FN): YOLOv5 failed to recognize the presence of a traffic cone. Using the above definitions, two key parameter values were calculated, and the following definitions were used to determine the parameter values: 1) precision in Equation 3 is the fraction of correct recognition instances out of total successful recognitions; and 2) recall in Equation 4 is the fraction of correct recognition instances retrieved over total expected recognitions. The training result regarding precision and recall is shown as Figure 5.

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

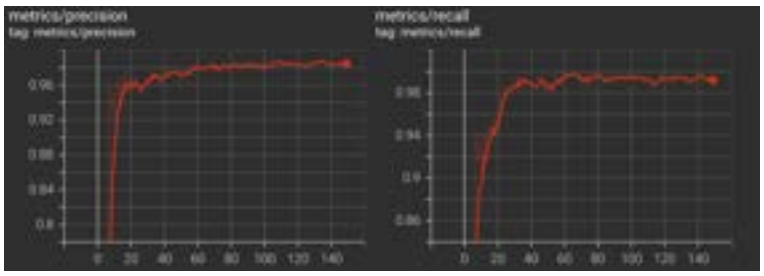


FIGURE 5: Training Result on YOLOv5 to Detect Traffic Cones.

The YOLOv5 uses the 416×416 input images and reaches 98% precision and 99% recall on our validation dataset. We trained the network from a checkpoint that is pre-trained in the MS-COCO dataset [39] using our training dataset. All the detection results indicated the overall result is above 90%, which is an adequate detection accuracy for any safety-critical work zone detection applications.

Phase Two

In phase two of the experiments, we repeated the training and testing procedure in phase one, but this time YOLO is trained on our improved dataset with labeling (Figure 32a in Appendix A) on not only small traffic cones, but also all common work zone objects like traffic barrels, traffic barricades, construction workers, as well as vehicles. We used drones to capture 2,000 images of real-world work zone setup by ATS traffic shown in Figure 32b in Appendix A. The drones were controlled to fly over from the starting point to the ending point of the work zone so that the camera on the drones could see everything within the work zone. We only labeled objects that weren't too small on the images since YOLO has difficulty detecting and distinguishing extremely small objects in images according to YOLO official release documents [26]. For example, we didn't label some traffic cones that were too far away from the camera where its bounding box length or width is less than 3 pixel on a 640 pixel by 640 pixel image. The whole dataset was split into three parts including training set, validation set, and test set. Then all images in the dataset were resized to 640 pixels by 640 pixels before some data augmentation techniques including horizontal flip and brightness variation introduced in phase one were applied. Yolo was later trained on these data using 16 batch and 300 epochs with Stochastic Gradient Descent (SGD) optimizer of 0.01 learning rate, and took 2 hours to complete training. Google Colab, which offers no-cost

access to potent GPUs and doesn't need configuration, was used to train the model. We utilized a Roboflow.ai [40] notebook that is based on YOLOv5 and employed pre-trained COCO weights. Figure 6 shows various performance measures for both the training and validation sets in each epoch during the whole training process and depicts three forms of loss: box loss, objectness loss, and classification loss. The box loss quantifies how effectively the algorithm can detect an item's center and how well the anticipated bounding box covers an object. Objectness loss is simply a measure of the likelihood of finding an item in a particular zone of interest. Classification loss indicates how successfully the algorithm predicts the proper class of an item. Around epoch 100, accuracy, recall, and mean average precision stop improving rapidly and the box, objectness, and classification losses evaluated in validation set stop declining dramatically. We chose the best weights by the evaluation in the validation set. Confusion matrix in Figure 33 in Appendix A is also an important metric to measure performance, where the values in the diagonal show that the model predicted correctly while other values indicate that model made wrong predictions. We can see that the model made more wrong predictions in detecting traffic cones than other objects, which is expected since small object detection is one of the difficulties that YOLO has.

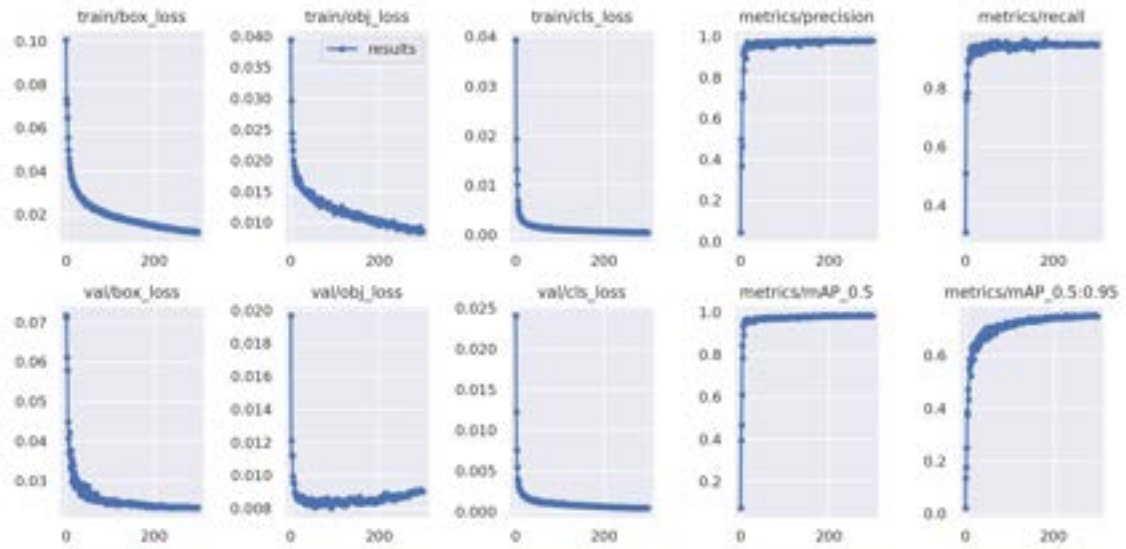


FIGURE 6: YOLO Training Results.

Calibration of the Distortion in the Top-Down Image

We calibrated the distortion in the top-down image by using perceptual hash algorithms. The function of the perceptual hash algorithm is to build a 64-bit hash value for each picture and then compare the hash values of different images. The more similar the images are, the closer the results are. We first reduced the image to a size of 8 pixels by 8 pixels. The goal of this stage is to eliminate picture features and maintain just essential information such as structure. Then we converted the reduced image to grayscale. Then a grayscale average of all 64 pixels was calculated. Next, we compared the gray level of each pixel to the average value; if it is larger than or equal to the average value, it is 1, otherwise, it is 0. Finally, we combined the results of the preceding comparison to generate a 64-bit binary representation of the hash of the image. We defined the distortion as the hamming distance calculated from two 64-bit binary representations of the hashes

(a) The grayscale ground truth image. (b) Our grayscale image.

FIGURE 7: Distortions.

of the ground truth image and our top-down image. The hamming distance increases by one for each difference between the hashes. If two photos are identical, they will have the same hash and will have a hamming distance of zero. A satellite image from Google Maps was used as ground truth to compare with our top-down image as shown in Figure 7. The distortion value was 23%, by dividing the hamming distance by 64 pixels. The distortion from our top-down image was mostly came from the horizontal road (top-right corner of the image).

Calibration of the Linear Transformation Matrix

The accuracy of GPS sensors is important to establish the high precision localization from the image plane. However, a normal commercial GPS sensor is unsuitable for our use, as it has a position error of about 20m. In our experiment, we used the differential GPS (dGPS) sensor from the OBU as GPS ground truth. The GPS data from the OBU is corrected by RSU as a base station using GNSS [41] to achieve the accuracy within 1m. The calibration procedure for calibration of the linear transformation (LT) matrix L in Equation 2 is as follows. We measured each point's GPS coordinate (longitude,latitude) with the dGPS sensor on the road in Figure 2, and this point's pixel coordinate (p_x,p_y) in Figure 2. Then we performed PT to get the transformed pixel coordinate (p'_x,p'_y) from (p_x,p_y). We recorded each point's (p'_x,p'_y) and (longitude,latitude) to calculate LT matrix L by python using the function np.linalg.lstsq. The calibration results are shown in Figure 8 where the blue points are GPS ground truth data and the orange points are GPS coordinates calculated from Equation 2 using LT matrix L.

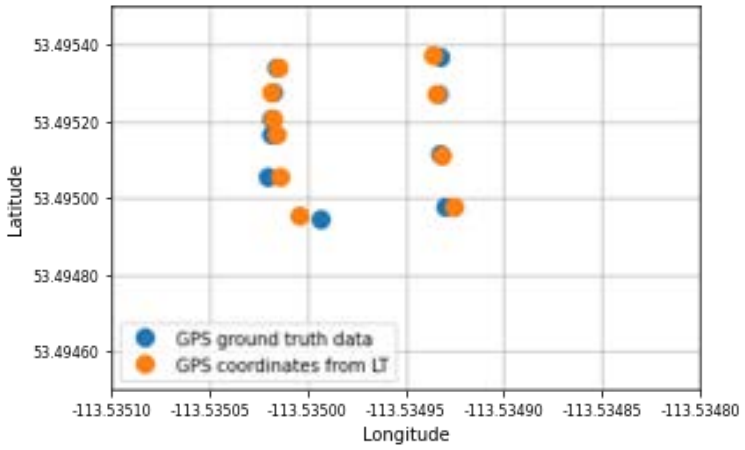


FIGURE 8: Calibration Results of LT.

Localization Error Analysis

The performance of a vision-based work zone alert system depends on the accuracy localization of work zone items from the traffic camera. The experiment setup is shown in Figure 9.

FIGURE 9: Experiment Setup to Measure Localization Error.

We used the OBU's dGPS data as GPS ground truth data in this experiment to calculate the localization error of our vision-based system. We also compared our localization performance with a GPS APP on a cellphone that the work zone site manager would use to provide the work zone location.

On each section of the road, we placed a traffic cone and moved the OBU's dGPS sensor next to the traffic cone to obtain GPS ground truth data for 15 minutes under good reception condition. From the vision-based work zone items localization method, we can obtain the GPS data estimated by the proposed localization. Our team member stood next to the traffic cone recording the data from the GPS APP on a cell phone. We then calculated the root mean square error (RMSE) based on Equation 5. Where G_i is the GPS ground truth data, and L_i is the location estimated by the proposed localization, N is the number of data points.

$$RMSE = \sqrt{\sum_{i=1}^{N}(G_i - L_i)^2/N} \tag{5}$$

Testing points used in this experiment are green points shown in Figure 10. We calculated RMSE on each testing point and obtained localization error analysis results shown in Figure 11. RMSE is the localization performance evaluation metric in this experiment.

From Figure 11, we can see our proposed localization method outperformed the GPS APP on the cellphone. Our proposed method had an average RMSE of 0.40m and maximum RMSE of 1.1m, which means a work zone item can be accurately located within half of a lane width (1.75m).

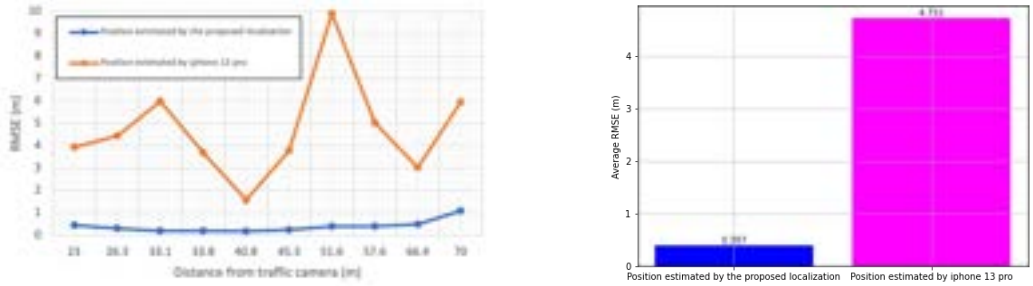
ALTERNATIVE LOCALIZATION METHOD

In this section, an alternative localization method was proposed and followed by detailed localization error analysis with visualizations on Google Maps and work zone boundary tests. The experiments were conducted at a different traffic camera at the same street.

Perspective Transformation and Linear Transformation Matrix Calculations

Google Maps was used to calculate the PT and LT matrix. First, the pixel coordinates of four pairs of matching points on the traffic camera image plane (Figure 12a) and the Google Maps satellite image plane (Figure 12b) were recorded. Then python was used to calculate the PT matrix. To solve the equation of PT matrix, at least four pairs of matching points were required. More pairs

FIGURE 10: Localization Error Analysis Testing Points.



(a) The RMSE comparison between our method and iPhone 13pro on each testing point.

(b) The average RMSE comparison calculated from all testing points.

FIGURE 11: RMSE Comparison.

of matching points would help us to reduce the distortions. The image plane after applying the PT matrix was shown in Figure 12c. Visually compared with Figure 2, where no ground truth top-down view image was used, calculating the PT matrix with the Google Maps satellite image as a ground truth (Figure 12) did reduce some distortions, although distortions were more obvious at a far distance from traffic camera.

The LT matrix was calculated by collecting the pixel coordinates on the Google Maps satellite image plane and the GPS coordinates from Google Maps of those matching points we chose when we solved the PT matrix. Again, at least four pairs of matching points were required to solve the LT matrix. The matching points used to calculate the LT matrix didn't have to be the same matching points used to solve the PT matrix. To show an example, Figure 34 in Appendix A shows the pixel coordinate in the traffic camera image plane after PT and the GPS coordinates from Google Maps of a matching point (shown by a green circle point) used to calculate the LT matrix. The GPS coordinate was collected by placing the pointer on the matching point and right clicked the mouse. This localization method was based on the assumptions that there were at least four distinguishable points that were not on the same line from the Google Maps satellite image. Otherwise, we would have to collect GPS data onsite.

This alternative localization method had a few advantages. First, onsite GPS data collection was not necessary to calculate the LT matrix since we could collect GPS data directly from Google

(a) The traffic camera image plane. (b) The Google Maps satellite image plane. (c) The traffic camera image plane after PT.

FIGURE 12: Process of Satellite Image Based PT

(a) The grayscale ground truth image. (b) Our grayscale image.

FIGURE 13: Distortions.

Maps. Second, this method was very low cost and fast. It required no GPS or GNSS sensor, and the process of collecting data from Google Maps was much faster than collecting data using GNSS sensor from OBU. It usually took half of an hour to collect all the data needed to calculate the LT matrix, whereas we could normally finish data collection within ten minutes, including five minutes of finding matching points plus five minutes of collecting GPS coordinates for these matching points, with the help of Google Maps. Third, since we used the Google Maps satellite image as a ground truth to calculate the PT matrix, the distortions were smaller, which would benefit the localization accuracy. This localization method is recommended to use on highways where onsite data collection is difficult or dangerous, and the roads have many lane markings.

Distortion Calculations

Again, we used perceptual hash algorithms to systematically calculate the distortion value in our top-down image. Figure 13 shows the ground truth image and our top-down image. The distortion value was 17% for this case. Compared with previous perspective transformation method which had a distortion value of 23%, this satellite image based perspective transformation method had smaller distortion value. The distortion from our top-down image was mostly came from the horizontal road (bottom-left corner of the image).

Localization Error Analysis with Visualizations

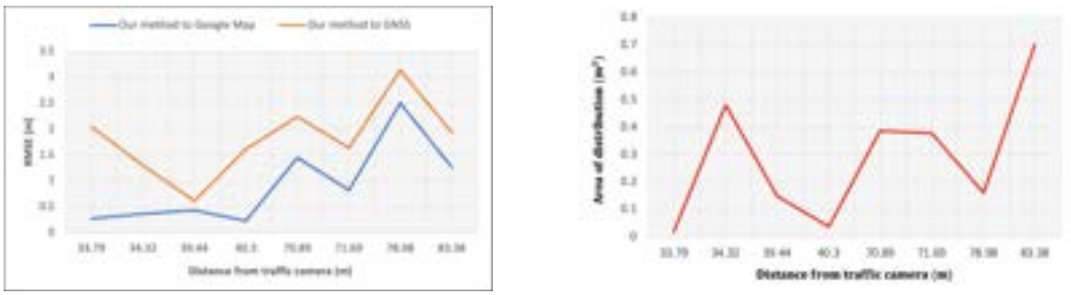
As we did the localization error analysis before, we chose eight points on the traffic camera image plane to be the testing points ranging from 30 meters to 80 meters from the traffic camera shown in Figure 14a. At the beginning of this experiment, we opened the live video stream from the traffic camera and clicked the record button so that localization error analysis could be conducted offline. Then we placed a traffic barrel on each testing point (Figure 14b) and our localization method would output the GPS coordinate of the traffic barrel. We also used GNSS sensor on OBU

(a) Eight testing points for localization error analysis are shown in green circle points.

(b) This is one example of a traffic barrel placed on a testing point with YOLO detection.

FIGURE 14: Testing Points for Localization Error Analysis

to collect GPS coordinates of those testing points in addition to collect GPS data from Google Maps, since we wanted to explore the difference between these data sources at the same point. By Equation 5, RMSE can be calculated. We calculated the RMSE between our method and GPS data from GNSS sensor and also obtained the RMSE between our method and GPS data shown in Figure 15a. We found out that RMSE had a trend of increasing as the testing points were further away from the traffic camera, which indicated our localization method generally worked well at a relative near distance and didn't perform as well at a far distance. To be more specific, we can see that testing points that were within 40 meters from the traffic camera had a RMSE of less than 0.5 meters when comparing our method and GPS data from Google Maps. RMSE certainly increased on testing points that were 70 meters further away from the traffic camera. The average RMSE of all testing points is 0.9 meters using Google Maps as a ground truth and 1.8 meters using GNSS sensor as a ground truth. We also calculated the localization noise produced from our localization method. The localization noise was defined by finding the area of distribution of all points generated by our localization method on each testing point. We defined the area of distribution as the area of the smallest circle that could covered all points from our method. The localization noise plot is shown in Figure 15b, showing all points had less than 0.5 square meters area of distribution, except for the furthest point that had 0.7 square meters. This experiment showed that longer distances from the traffic camera had some negative impacts on our localization method in terms of both RMSE and localization noise. Figure 16 shows a few visualizations by plotting all GPS data on Google Maps, where red dots were from our localization method, blue dots were from Google Maps, and purple dots were from the GNSS sensor. We can clearly see the area of distribution from our method is larger on the furthest testing point than other testing points. The area of distributions from GNSS sensor were basically the same. We also plotted all GPS data from all testing points in a scatter plot shown in Figure 35 in Appendix A and found out that the GPS data from GNSS sensor generally had about 0.5 meters offset in the west direction compared with GPS data from Google Maps. Since the lane information, including which lane was closed by the work zone, was generated by a map matching algorithm with GPS data of the work zone and GPS data of the traffic lanes as an input, this localization method is recommended to use when the GPS data of the traffic lanes is collected from Google Maps instead of GNSS sensor. The lane information may not be very accurate when using the GPS data of the traffic lanes collected from GNSS sensor. For example, the work zone warning APP may display two traffic lanes are closed by the work zone when there is actually one traffic lane closed by the work zone.



(a) The RMSE plot.

(b) Localization noise.

FIGURE 15: Results of Localization Analysis

(a) Visualization on the nearest testing point.

(b) Visualization on the furthest testing point.

(c) Visualization on the testing point that both RMSE values were less than 1 meter.

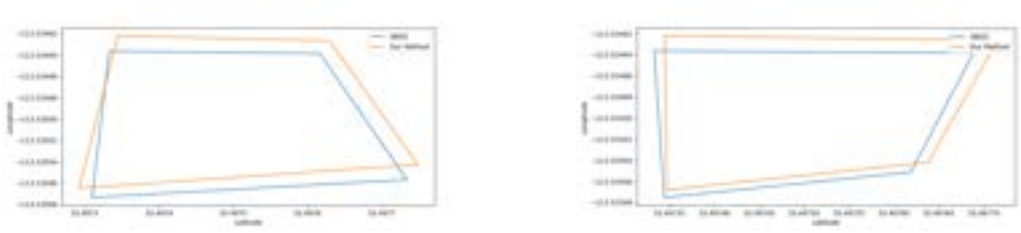
(d) Visualization on the testing point that both RMSE values were greater than 2 meters.

FIGURE 16: Visualization on Testing Points.

Work Zone Boundary Tests

In this section, we followed the pipeline designed by Shi et al. [21] to test the performance of our method in terms of work zone boundary localization. We placed many work zones of different layouts in 118 Street, NW, Edmonton, Canada and recorded work zone objects GPS coordinates by the GNSS sensor, the Google Maps and our localization method. There were 8 work zones in total and 6 of them completely showed up in the field of view of the traffic camera and the remaining 2 partly showed up. Then we used the equations developed by Shi et al. [21] to calculate results.

The equation 'Intersection over Union (IoU) at a Distance (Equation 6)' was used to calculate the overlap percentage between our work zone region and the ground truth work zone region



(a) Example 1. (b) Example 2.

FIGURE 17: Different Work Zone Boundaries.

at different distance thresholds:

$$IoU@d_t(S_o, S_g) = \frac{|(S_o \cap S_g) \cap S_r|}{|(S_o \cup S_g) \cap S_r|} \quad (6)$$

where S_o is the work zone area by our method, S_g is the work zone area by ground truth (GNSS sensor or Google Maps) and S_r is a circle of radius d_t centered at the traffic camera. We used three different distance thresholds d_t = 40m, 60m, 90m. The IoU result was valid when there was work zone area within the distance threshold. The average of valid IoU values of different work zones was used as the final result.

Precision and recall (Equation 3 and Equation 4) were used to quantify the detection rate and detection accuracy. We defined the ground truth work zone was accurately detected when IoU reached 0.7. We redefined the following parameters and metrics: 1) True Positive (TP): our method accurately detected a work zone boundary; 2) False Positive (FP): there was no work zone but our method detected a work zone boundary; and 3) False Negative (FN): our method failed to accurately detect a ground truth work zone. The results were shown in Table 4 and Figure 17 shows a few work zones of different arrangements.

Range	**Ground Truth**	**IoU**	**Precision**	**Recall**
40m	GNSS	69%	75%	60%
40m	Google Maps	91%	86%	75%
60m	GNSS	80%	86%	75%
60m	Google Maps	97%	86%	75%
90m	GNSS	76%	84%	71%
90m	Google Maps	94%	86%	75%

TABLE 4: Work Zone Boundary Detection Accuracy.

REAL WORLD TESTING

We conducted a case study on 118 street, NW, located in Edmonton, Alberta, Canada, to evaluate the performance of our vision-based work zone safety alert system in a CV environment.

Experiment Setup

In this experiment, we used a video camera that connected to the central server with NVIDIA RTX3090 and 24GB video memory. The central server ran the work zone items detection algorithm and sent work zone static information including the start point GPS coordinates, the end

point GPS coordinates, as well as which lane was closed by the work zone, to the DSRC-enabled RSU. After receiving work zone static information from the central server, the DSRC-enabled RSU broadcasted this information to the nearby CV equipped with DSRC-enabled On-Board Unit (OBU). Then each CV within the range of DSRC-enabled RSU (400 meters) received work zone safety alerts.

Experiment Design

In this experiment, we used a normal vehicle to block the CV, so that the driver in the CV could not see the work zone in front of the normal vehicle. The two vehicles are approaching the work zone. Figure 18 shows the experiment design. Since the driver in the CV cannot see the work zone, the work zone warning APP plays a vital role to inform the driver of the position of the work zone.

(a) Experiment Design and Planning (b) The actual picture at experiment site.

FIGURE 18: Experiment Design for Real World Testing.

Experiment Results

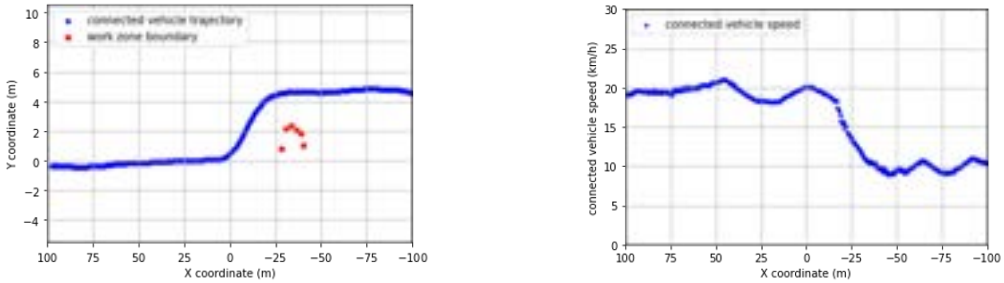
The work zone items localization results in this experiment are shown in Figure 19. From these results we generated work zone safety alert messages to the CV driven through the work zone.

(a) The work zone in the traffic camera view with virtual lane marking added from Google Maps (b) The localization results showing the precise location of the work zone.

FIGURE 19: Work Zone Localization

In the first experiment, we recorded the CV trajectory by OBU without the work zone warning APP enabled. The CV trajectory, speed and work zone location estimated by the proposed localization are shown in Figure 20.

From Figure 20, we can see that the CV performed a lane change 25 meters in front of the work zone starting point. Changing lanes at such a close distance can cause potential danger to



(a) How CV driven through work zone (from positive x axis to negative x axis) without the work zone warning APP enabled.

(b) CV's speed before and after work zone as well as where CV slowed down.

FIGURE 20: CV's Behavior Without the Work Zone Warning APP

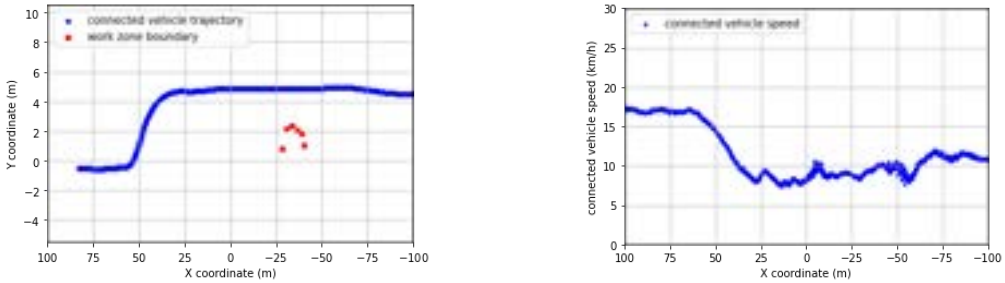
both the driver and the worker in the work zone. The lane changing action of the CV depended on the front vehicle, since the driver of the CV can only see the work zone when there is no vehicle in front of the CV.

In the second experiment, we repeated the same experiment but with the work zone warning APP enabled. Since the traffic camera can see the work zone, the central server ran the work zone detection and localization algorithm, and then sent the work zone static information to the RSU. The RSU then broadcasted the work zone static information through the DSRC channel to the OBU on the CV. The OBU then calculated the real-time distance between the CV and the work zone and displayed the distance on the work zone warning APP through OBU's Wi-Fi.

FIGURE 21: The Work Zone Warning APP.

Figure 21 shows the moment when the driver performed a lane change based on the work zone warning APP's safety alert. At this moment, the driver could not see the work zone, but still knew the precise location of the work zone with the help of the work zone warning APP. In the work zone warning APP, we set the work zone warning sound to trigger when the distance between the CV and work zone starting point is less than 100m, since the speed limit on this testing road is 40km/h. The driver can change the setting in the work zone warning APP to trigger the warning sound at a farther distance (for example, 1000m for highway). The CV trajectory, speed and work zone location estimated by the proposed localization are shown in Figure 22.

From Figure 22, we can see that the CV performed a lane change at 80 meters in front of the work zone starting point. With the help of the work zone warning APP, the driver in the CV



(a) How CV driven through work zone (from positive x axis to negative x axis) with the work zone warning APP enabled.

(b) CV's speed before and after work zone as well as where CV slowed down.

FIGURE 22: CV's Behavior With the Work Zone Warning APP

changed lanes and slow down much earlier, which ensured the safety of both the driver and the workers in the work zone.

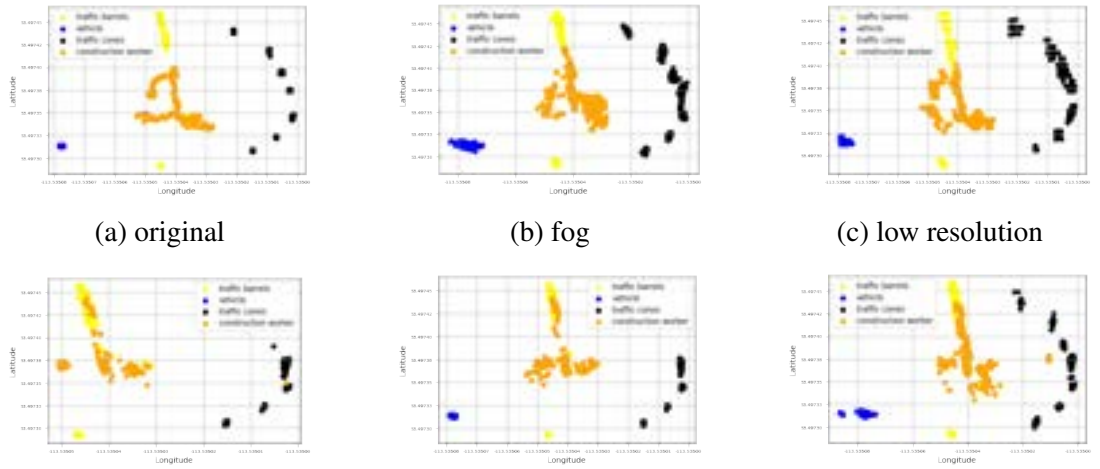
EFFECT OF DIFFERENT WEATHER CONDITIONS AND LOW RESOLUTION

To explore the effect of different weather conditions and low camera resolution, we used a 1280 pixel by 720 pixel test video (Figure 23a) where a construction worker was walking inside the work zone, placed at 50 meters from the traffic camera, as a input to YOLO and our localization method. Shown in Figure 23b, Figure 23d, Figure 23e and Figure 23f, negative 100% exposure and negative 100% brightness were applied to the test video before some weather effects were added by a video editing software. This step was to test how well YOLO can detect work zone items under nighttime fog, rain, snow, and thunderstorm. There were several reasons we didn't conduct this experiment under the real-world conditions. For example, it is unsafe for both drivers and our team members to collect work zone data during a thunderstorm at night. We also tested YOLO's performance on an extreme low camera resolution setting (254 pixel by 144 pixel) shown in Figure 23c. The localization results for different test videos are shown in Figure 24 by using the framework developed by Mike et al. [42]. It is obvious to see that localization noise increased compared with the original test video. The version to simulate nighttime rain had the largest impact on YOLO's detection, where YOLO missed 3 traffic cones and 1 vehicle completely, and missed the detection of the walking construction worker many times (Figure 24d). Looking into the annotation videos on nighttime fog, rain, snow, and thunderstorm, where each object was annotated on the videos when detected by YOLO, we found out that YOLO generally can't obtain a stable detection on small objects such as a traffic cone. Lowering the resolution of the test video didn't seem to largely effect YOLO's performance. YOLO still could detect and localize all objects within lane level accuracy although noise certainly increased for small objects. This indicated our method can be applied to most low-resolution traffic cameras in Edmonton, where the live traffic camera only supports 600 pixels by 400 pixels video stream. Lowering the resolution of traffic cameras generally would increase the stability of live video stream, for instance, more stable video frames, and decrease the latency since most traffic networks in Edmonton don't support high resolution video transmission.

(a) original (b) fog (c) low resolution

(d) rain (e) snow (f) thunderstorm

FIGURE 23: Test Videos from Different Weather Conditions and Camera Resolution.



(a) original (b) fog (c) low resolution

(d) rain (e) snow (f) thunderstorm

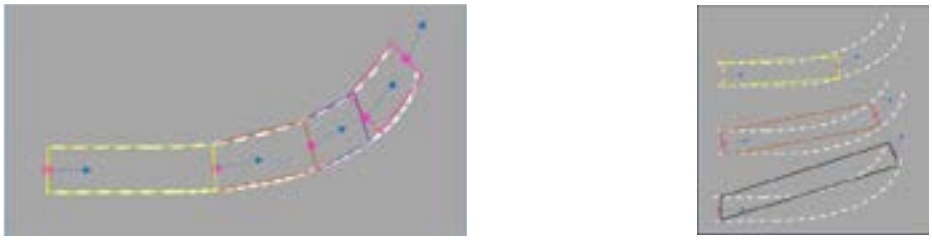
FIGURE 24: Tracking Results from Different Weather Conditions and Camera Resolution.

LANE CLOSURE INFORMATION BROADCASTING

We used standard map data message (MAP) and Traveler Information Message (TIM) defined by SAE J2735 standard to represent the work zone lane geometry and map matching algorithm was used to determine the lane closure information based on the coordinate of each traffic cone, which will be explained in this section.

MAP and TIM Definitions

As described in SAE J2735, the necessary fields shown in Figure 36a in Appendix A are required to serve as the basic reference to other messages such as Signal Phase and Timing Message (SPAT) and TIM which will reference back to Intersection/lane IDs in the MAP message. By sending a MAP message, it is possible to broadcast the static work zone lane geometry to a CV nearby. The CV would be able to know which lane it is driving, but not able to know if the lane is open or closed.



(a) Lane geometry defined by center line segments and road width.

(b) Displacement between bounding box affects lane geometry accuracy.

FIGURE 25: Lane Geometry Represented by MAP.

SAE J2735 also defines traveler information message (TIM) which contains information related to road conditions. Figure 36b in Appendix A shows the essential fields for TIM. It can be used to broadcast work zone location with respect to the lanes defined in the MAP message. The work zone definition in TIM is similar to Work Zone Data Exchange (WZDx) format. WZDx Specification [43] allows infrastructure owners and operators (IOOs) to make standardized work zone data available to third parties. The goal is to make public road traffic safer and more efficient by providing ubiquitous access to data on work zone activities. It is possible to choose a path manually/automatically in GPS coordinate and indicate road closure and generate the corresponding TIM to be broadcast by RSU. TIM messages can be broadcast simultaneously with the MAP message.

Shown in Figure 25, lane geometry can be defined by center line segments and road width based on SAE J2735 MAP message. To be more specific, the center line can be collected by: 1) road survey; 2) GNSS sensor waypoints; 3) manually select points from Google Maps; and 4) our localization method. Lane-level localization can be achieved using map matching algorithm by determining if the ego vehicle is located within any bounding box that belongs to a certain lane. Displacement between bounding box affects accuracy of lane geometry representation as we can see in Figure 25b.

Map Matching and Projecting

It is important to determine which lane is affected by road construction, and what is the exact starting and ending point. Using map matching algorithm to find the lane ID, and project the cone to the center line allows us to know the exact start/ending point of a work zone as shown in Figure 26. This is a naive approach assuming lane-level localization accuracy of traffic cone shown in Equation 7 to Equation 10, where p_{start}, and p_{end} as well as lane width (l) are known information about lane geometry. So the purpose of these equation is to find if traffic cones are in any specific traffic lanes.

Shown in Figure 37 in Appendix A, in TIM message the work zone geometry is represented as a series of points on the center lines projected by the cone's GPS coordinates (red). The starting point (green) of each work zone is an anchor point, and the remaining points (purple) are the offsets to the anchor point. For each segment, the Minimum projection length ($d_{0,1}$) is used for computing the starting point, and Maximum projection length ($d_{1,2}$) is used for computing the ending point. By rendering the MAP and TIM message data, Figure 38 in Appendix A could be obtained showing corresponding legends.

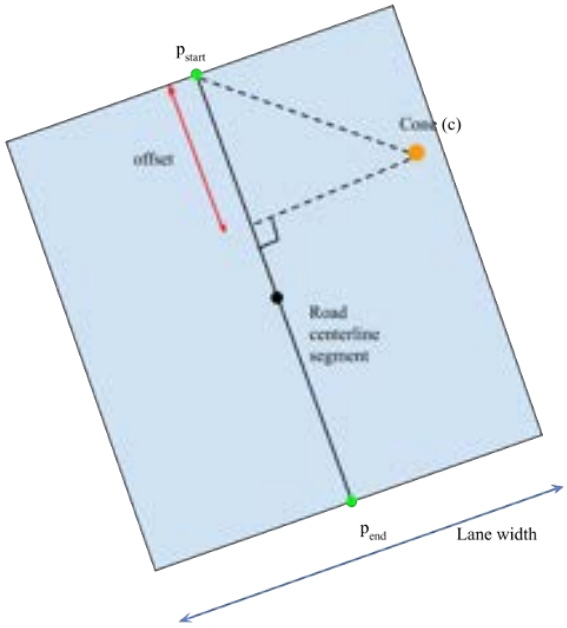


FIGURE 26: Map Matching.

$$\theta = arctan(\frac{p_{end_y} - p_{start_y}}{p_{end_x} - p_{start_x}}) \tag{7}$$

$$\begin{bmatrix} p'_{start_x} & p'_{end_x} & c'_x \\ p'_{start_y} & p'_{end_x} & c'_y \end{bmatrix} = \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix} \begin{bmatrix} p_{start_x} & p_{end_x} & c_x \\ p_{start_y} & p_{end_y} & c_y \end{bmatrix} \tag{8}$$

$$\text{Match iff. } p'_{start_x} < c'_x < p'_{end_x} \tag{9}$$

$$\text{AND } -\frac{l}{2} < c'_x < \frac{l}{2} \tag{10}$$

Field Test

We conducted a field test in a CV environment using C-V2X communication technology, as the new RSU and OBU device just arrived. We used a OBU to function as a RSU to broadcast MAP and TIM regarding work zone and lane information because the RSU could not be installed at that time. Figure 27 and Figure 39 in Appendix A show the communication topology and devices used in this field test respectively. Currently the map shown in Figure 40 in Appendix A is created using manually chosen points from OpenStreetMap [44] for simplicity. This method cannot be used for generating long distance work accurately. Resulting GPS coordinates of each node in Extensible Markup Language (XML) format is used to generate MAP message. MAP and TIM were broadcast at 10 hertz to nearby CVs. We developed a prototype of work zone warning APP with lane information running on a laptop placed in a CV to serve as an HMI for the driver. This work zone warning APP could show speed limit information, traffic lanes information and the work zone location as well as which lane was closed by the work zone. We set the speed limit from 40km/h to 10km/h when the work zone was less than 100 meters from the driver. We

recorded the data, and the results are shown in Figure 41, Figure 42 and Figure 43 in Appendix A, where we can see YOLO detection from the traffic camera, MAP and TIM visualization, the work zone warning APP, as well as the driver's vision. Since one of the amazing features in our system is real-time detection, localization, and warning, we also recorded results when removing traffic cones from the road. The results are shown in Figure 44 and Figure 45 in Appendix A. We can clearly see that the work zone region shown in MAP and TIM were decreasing and the work zone warning APP wouldn't give any warning when all traffic cones were removed. We repeated the experiments a few times to test stability. The mobile version of the work zone warning APP is being developed now. We repeated the experiments four times to show effectiveness. Figure 28 shows the trajectories of these experiments, where we didn't remove the work zone in the first three experiments and we removed the work zone in the fourth experiment to demonstrate real-time warning changes.

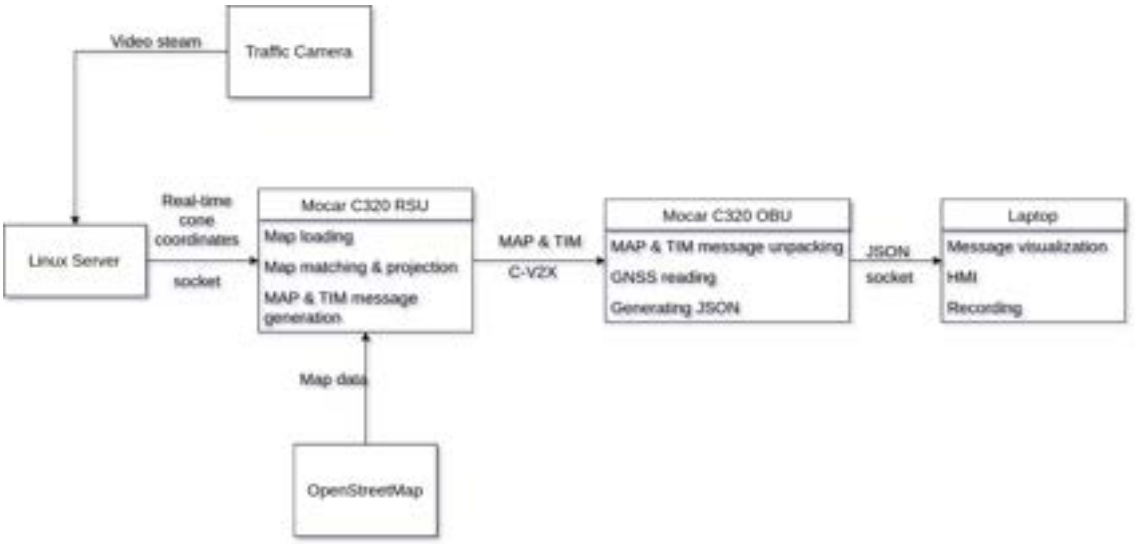


FIGURE 27: System Diagram.

System Level End-to-End Latency Tests

The data flow in this work zone warning system is shown in Table 5. The latency includes computational delay and communication delay. The computation delay depends on the hardware specification of the central server used to run the work zone detection and localization algorithm. For our experimental setup, we used a workstation with NVIDIA RTX3090 having 24GB GPU memory to run the work zone detection and localization algorithm. The communication network delay is the time difference between sending a message from one device and receiving the same message on another device. The latency was measured by Packet Internet or Inter-Network Groper (PING) function to determine Round-Trip Time (RTT) of a message. Each latency test was repeated 9 times and 100 packages were used to calculate RTT in each latency test to reduce error and improve accuracy. The tests were used Cellular Vehicle to Everything (C-V2X) RSU and OBU to broadcast work zone information.

Table 5 shows the end-to-end latency results in every part of the work zone alert system and the total latency that combines both computational and communication latency. The average latency is below 100ms, which satisfies the latency requirement for safety-critical applications.

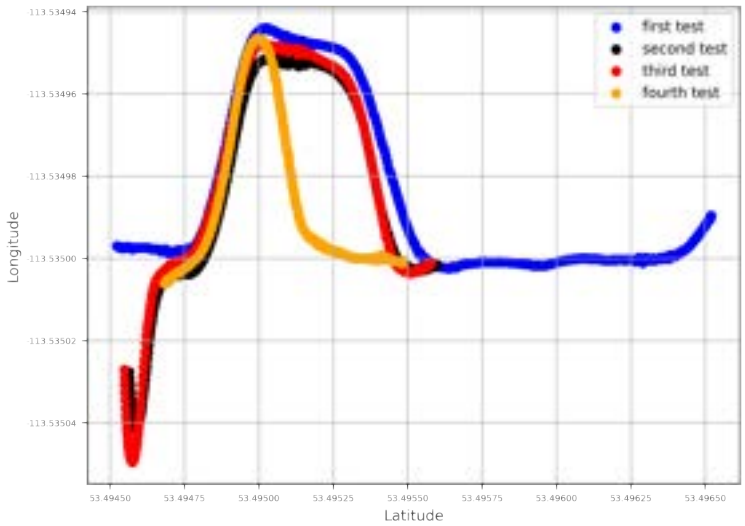


FIGURE 28: CV Trajectories.

End To End Latency	Latency Type	Average Latency	Standard Deviation
Traffic camera sending live video data to central server	Communication latency on wired network	2.1ms	0.3ms
Work zone detection and localization algorithm results generation	Computation latency on central server	5.4ms	1.2ms
Central server sending work zone information to RSU	Communication latency on wired network	1.5ms	0.2ms
RSU broadcasting work zone information to OBU	Communication latency on C-V2X	14.0ms	4.3ms
OBU sending vehicle's speed and work zone information to Samsung pad running work zone warning APP	Communication latency on OBU's Wi-Fi	17.7ms	6.6ms
Overall latency	Overall latency	40.7ms	12.6ms

TABLE 5: End-to-End Latency Test Results.

CONCLUSIONS

The goal of this research is to create a real-time vision-based work zone recognition and localization approach, that will increase both the driver and worker's safety in work zones. By broadcasting the safety alert in real-time (every 100ms) from RSUs to CVs within its communication range, our work zone safety messages were used to alert the driver. When compared to commercially available smart phones that the work zone site manager utilized to give work zone location, our localization error study demonstrates that the vision-based work zone localization approach can estimate the position more precisely in terms of RMSE. We found that the average RMSE using the vision-based work zone localization method is less than half a lane width (1.75m), which means we can accurately locate the work zone within a lane. Furthermore, we tested the work zone alert system in a scenario where the driver cannot see the work zone. The work zone warning APP generates safety alerts based on the potential forward collision risks between the work zone and the CVs. The results from the CV trajectory and work zone location demonstrate that the

work zone warning APP can inform the driver of the work zone location even if the driver cannot see the work zone and let the driver slow down much earlier. This cooperative perception ensures the safety of both the worker and the driver. The overall work zone alert system latency is below 100ms, which satisfies the latency requirement in a CV environment. Effect of different weather conditions and low resolution was studied and we found out our method can work even through the camera resolution was extremely low. MAP and TIM message was used to broadcast lane closure information of the work zone.

FUTURE WORK

For future works and plans, we will first complete the development of mobile versions of the work zone warning APPs. For the case study on 118 street, we will repeat the experiments as many times as possible to ensure effectiveness and stability. For kilometers-long work zones, we will use drones and multiple traffic cameras to localize work zones. We had used DSRC and C-V2X communication technologies in the CV environment, and 5G will be considered and utilized in the future to further reduce the latency. In our work zone warning applications, we used server-based structures to transmit data from the traffic camera to the central server then to the RSU. This structure can be simplified by a Multi-Edge Computing (MEC) device such as Islam et al. [35] used in their paper. They placed the computing device right next to the traffic camera and the RSU. As a result, this MEC device dramatically reduced the latency. We will also explore the possibilities of using cooperative perspective messages (CPM) to detect and localize moving objects, such as construction workers from on board cameras mounted on CVs. A few researches are being studied and will be reproduced in the future including Rauch et al. [45–47], Gunther et al. [48] and Shan et al. [49].

ACKNOWLEDGMENTS

This research work was jointly supported and funded by Steve Ennis and ATS Traffic Canada. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The data that support the findings of this study are available from the corresponding author, upon reasonable request. The contents do not necessarily reflect the official views or policies of Steve Ennis and ATS Traffic Canada. The authors declare that there is no conflict of interest regarding the publication of this paper.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study idea and scope definition: Haibo Cui, Dr. T. Z. Qiu; system development and training: Haibo Cui; Kaizhe Hou; Siqi Yan; Jiarui Zhang; M. Seraj; data collection and analysis: Haibo Cui; Kaizhe Hou; Siqi Yan; Jiarui Zhang; Yingke Wang; draft manuscript preparation: Haibo Cui, M. Seraj, Yingke Wang. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] N. H. T. S. Administration *et al.*, "2015 motor vehicle crashes: Overview," *Traffic safety facts: research note*, vol. 2016, pp. 1–9, 2016.

[2] E. Radwan, R. Harb, S. Ramasamy, *et al.*, "Evaluation of safety and operational effectiveness of dynamic lane merge system in florida: Final report, april 2009.," 2009.

[3] S. D. Schrock, G. L. Ullman, A. S. Cothron, E. Kraus, and A. P. Voigt, "An analysis of fatal work zone crashes in texas," *Report FHW A/TX-05/0-4028*, vol. 1, 2004.

[4] J. Daniel, K. Dixon, and D. Jared, "Analysis of fatal crashes in georgia work zones," *Transportation Research Record*, vol. 1715, no. 1, pp. 18–23, 2000.

[5] N. J. Garber and M. Zhao, "Distribution and characteristics of crashes at different work zone locations in virginia," *Transportation Research Record*, vol. 1794, no. 1, pp. 19–25, 2002.

[6] Z. A. Nemeth and D. J. Migletz, "Accident characteristics before, during, and after safety upgrading projects on ohio's rural interstate system," *Transportation Research Record*, vol. 672, pp. 19–23, 1978.

[7] J. Hall and V. Lorenz, "Characteristics of construction-zone accidents," *Transportation Research Record*, vol. 1230, pp. 20–27, 1989.

[8] N. M. Rouphail, Z. S. Yang, and J. Fazio, "Comparative study of short-and long-term urban freeway work zones," *Transportation Research Record*, vol. 1163, pp. 4–14, 1988.

[9] A. Dehman and B. Farooq, "Are work zones and connected automated vehicles ready for a harmonious coexistence? a scoping review and research agenda," *Transportation research part C: emerging technologies*, vol. 133, p. 103 422, 2021.

[10] X. Wu *et al.*, "Vehicular communications using dsrc: Challenges, enhancements, and evolution," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 399–408, 2013.

[11] O. Raddaoui, M. M. Ahmed, and S. M. Gaweesh, "Assessment of the effectiveness of connected vehicle weather and work zone warnings in improving truck driver safety," *IATSS research*, vol. 44, no. 3, pp. 230–237, 2020.

[12] R. F. Benekohal, E. Shim, and P. T. Resende, "Truck drivers' concerns in work zones: Travel characteristics and accident experiences," *Transportation Research Record*, vol. 1509, pp. 55–64, 1995.

[13] W. Han, E. White, M. Mollenhauer, and N. Roofigari-Esfahan, "A connected work zone hazard detection system for roadway construction workers," in *Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience*, American Society of Civil Engineers Reston, VA, 2019, pp. 242–250.

[14] R. Schönrock, F. Wolf, and T. Russ, "Smart traffic cone-dynamic detection and localization of traffic disruptions," in *FAST-zero'15: 3rd International Symposium on Future Active Safety Technology Toward zero traffic accidents, 2015*, 2015.

[15] V. V. A. T. Committee, *Vulnerable road user safety message minimum performance requirements*, Mar. 2017.

[16] F. Abodo, R. Rittmuller, B. Sumner, and A. Berthaume, "Detecting work zones in shrp 2 nds videos using deep learning based computer vision," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2018, pp. 679–686.

[17] Y.-W. Seo, J. Lee, W. Zhang, and D. Wettergreen, "Recognition of highway workzones for reliable autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 708–718, 2014.

[18] B. Mathibela, M. A. Osborne, I. Posner, and P. Newman, "Can priors be trusted? learning to anticipate roadworks," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, IEEE, 2012, pp. 927–932.

[19] P. Kunz and M. Schreier, "Automated detection of construction sites on motorways," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2017, pp. 1378–1385.

[20] R. Graf, A. Wimmer, and K. C. Dietmayer, "Probabilistic estimation of temporary lanes at road work zones," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, IEEE, 2012, pp. 716–721.

[21] W. Shi and R. R. Rajkumar, "Work zone detection for autonomous vehicles," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2021, pp. 1585–1591.

[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[23] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[25] W. Liu *et al.*, "Ssd: Single shot multibox detector," in *European conference on computer vision*, Springer, 2016, pp. 21–37.

[26] G. Jocher, *Yolov5. code repository https://www. github. com/ultralytics/yolov5*, 2022.

[27] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[28] C. for Advanced Automotive Technology. "Connected and automated vehicles." (2018), [Online]. Available: http : / / autocaat . org / Technologies / Automated _ and _ Connected _ Vehicles/ (visited on 09/12/2022).

[29] W. Genders and S. N. Razavi, "Impact of connected vehicle on work zone network safety through dynamic route guidance," *Journal of Computing in Civil Engineering*, vol. 30, no. 2, p. 04 015 020, 2016.

[30] K. C. Dey, A. Rayamajhi, M. Chowdhury, P. Bhavsar, and J. Martin, "Vehicle-to-vehicle (v2v) and vehicle-to-infrastructure (v2i) communication in a heterogeneous wireless network–performance evaluation," *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 168–184, 2016.

[31] H. Abdulsattar, A. Mostafizi, and H. Wang, "Surrogate safety assessment of work zone rear-end collisions in a connected vehicle environment: Agent-based modeling framework," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 8, 2017.

[32] J. Lombardo. "How connected vehicles make work zones safer." (2017), [Online]. Available: https://www.forconstructionpros.com/asphalt/article/20867242/how-connected-vehicles-make-work-zones-safer (visited on 09/12/2022).

[33] S. Mishra, M. M. Golias, D. Thapa, *et al.*, "Work zone alert systems," Tennessee. Department of Transportation, Tech. Rep., 2021.

[34] F. Qiao, R. Rahman, Q. Li, and L. Yu, "Safe and environment-friendly forward collision warning messages in the advance warning area of a construction zone," *International journal of intelligent transportation systems research*, vol. 15, no. 3, pp. 166–179, 2017.

[35] M. Islam, M. Rahman, M. Chowdhury, G. Comert, E. D. Sood, and A. Apon, "Vision-based personal safety messages (psms) generation for connected vehicles," *IEEE transactions on vehicular technology*, vol. 69, no. 9, pp. 9402–9416, 2020.

[36] V. C. T. Committee, *Dedicated short range communications (dsrc) message set dictionary*, Mar. 2016.

[37] O. Polly, D. Stephens, and R. Ostroff, "Work zone activity data (wzad) – data dictionary report (version 2)," 2019.

[38] T. Wen, Z. Xiao, K. Jiang, M. Yang, K. Li, and D. Yang, "High precision target positioning method for rsu in cooperative perception," in *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, IEEE, 2019, pp. 1–6.

[39] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *European conference on computer vision*, Springer, 2014, pp. 740–755.

[40] *How to train yolov5 on custom objects*. [Online]. Available: https://colab.research.google.com/drive/1gDZ2xcTOgR39tGGs-EZ6i3RTs16wmzZQ.

[41] T. U. goverment. "Other global navigation satellite systems (gnss)." (2021), [Online]. Available: https://www.gps.gov/systems/gnss/ (visited on 09/12/2022).

[42] M. Brostrom, *A collection of real-time multi-camera multi-object sota trackers using yolov5*. [Online]. Available: https://github.com/mikel-brostrom/Yolov5_StrongSORT_OSNet.

[43] *Usdot-jpo-ode/wzdx: The work zone data exchange (wzdx) specification aims to make harmonized work zone data provided by infrastructure owners and operators (ioos) available for third party use, making travel on public roads safer and more efficient through ubiquitous access to data on work zone activity.* [Online]. Available: https://github.com/usdot-jpo-ode/wzdx.

[44] [Online]. Available: https://www.openstreetmap.org/#map=15/53.4961/-113.5359.

[45] A. Rauch, F. Klanner, and K. Dietmayer, "Analysis of v2x communication parameters for the development of a fusion architecture for cooperative perception systems," in *2011 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2011, pp. 685–690.

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

[47] A. Rauch, S. Maier, F. Klanner, and K. Dietmayer, "Inter-vehicle object association for cooperative perception systems," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, IEEE, 2013, pp. 893–898.

[48] H.-J. Günther, B. Mennenga, O. Trauer, R. Riebl, and L. Wolf, "Realizing collective perception in a vehicle," in *2016 IEEE Vehicular Networking Conference (VNC)*, IEEE, 2016, pp. 1–8.

[49] M. Shan *et al.*, "A novel probabilistic v2x data fusion framework for cooperative perception," *arXiv preprint arXiv:2203.16964*, 2022.

APPENDIX A

FIGURE 29: Work Zone Activities Data Structure Based on WZAD.

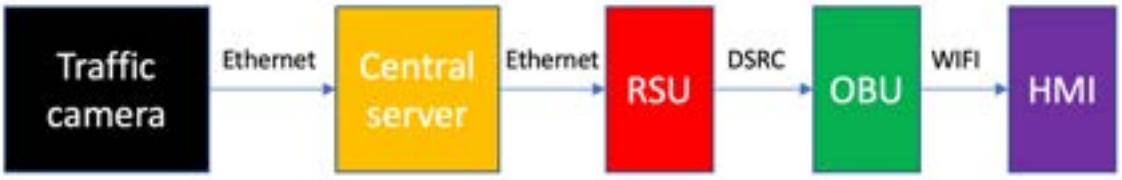
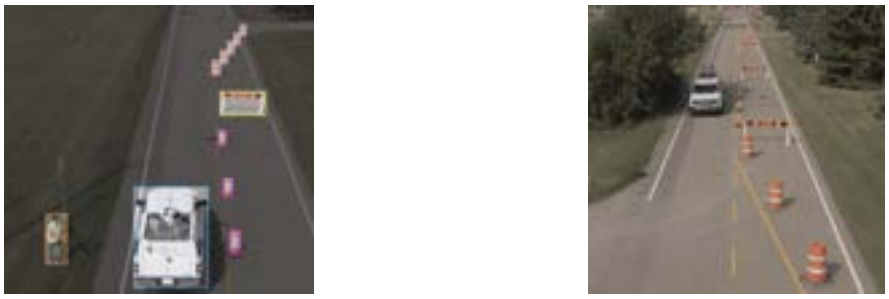


FIGURE 30: Communication Topology in the Vision Based Work Zone Alert System.

FIGURE 31: Method Used to Transform Pixel Coordinates to GPS Coordinates.



(a) Example of data labeling. (b) Example of work zone image.

FIGURE 32: Dataset Images.

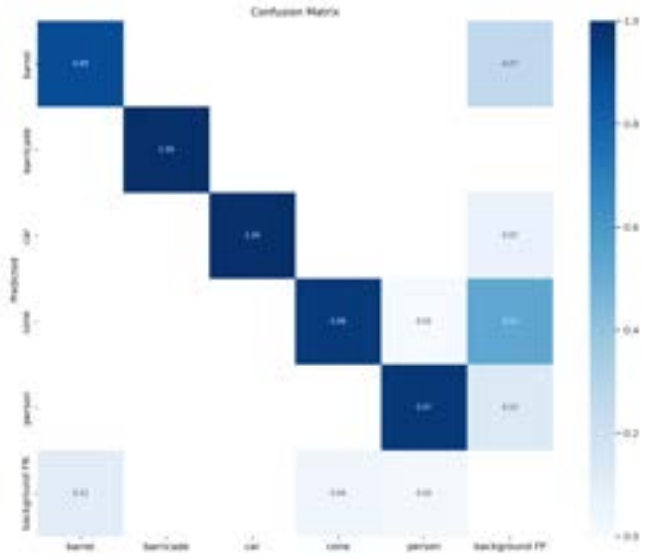
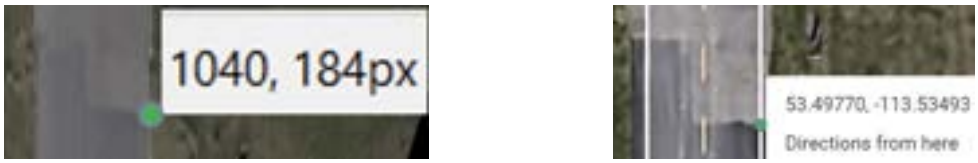


FIGURE 33: Confusion Matrix.



(a) The pixel coordinate of a matching point in the traffic camera image plane after PT. (b) The GPS coordinate of the same matching point from the Google Maps.

FIGURE 34: Example Data to Calculate LT

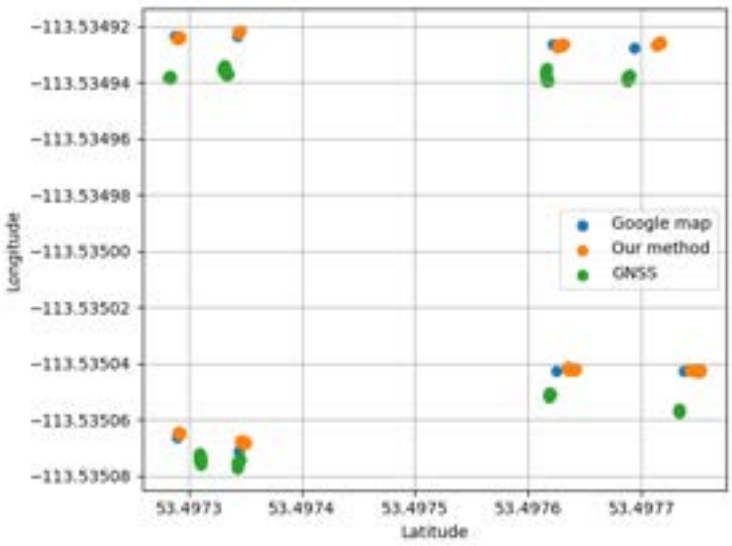
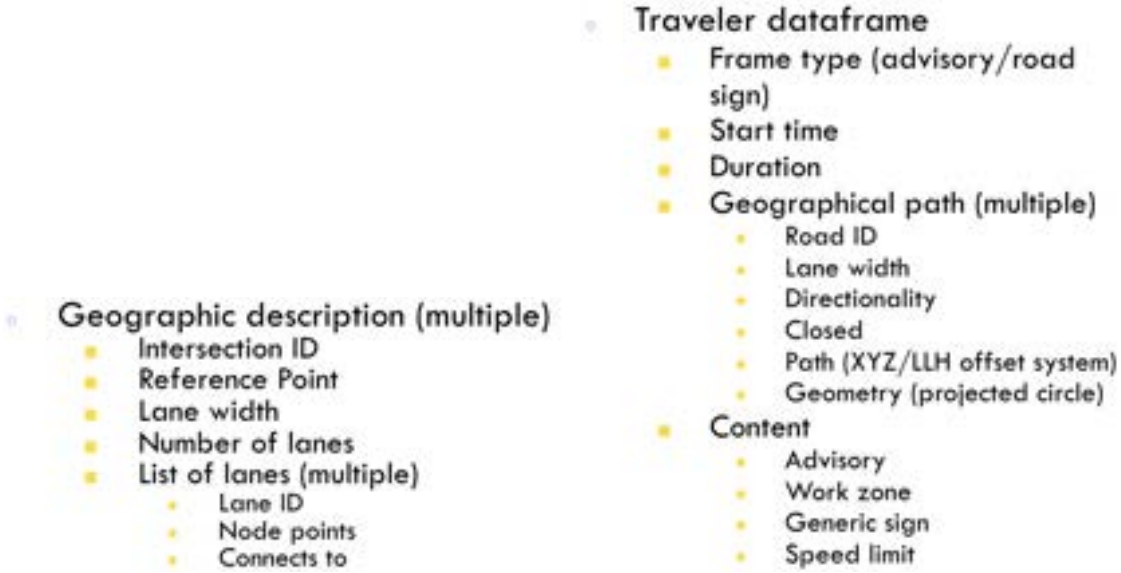


FIGURE 35: The Scatter Plot of All GPS Data.



(a) MAP Data Frame. (b) TIM Data Frame.

FIGURE 36: MAP and TIM.

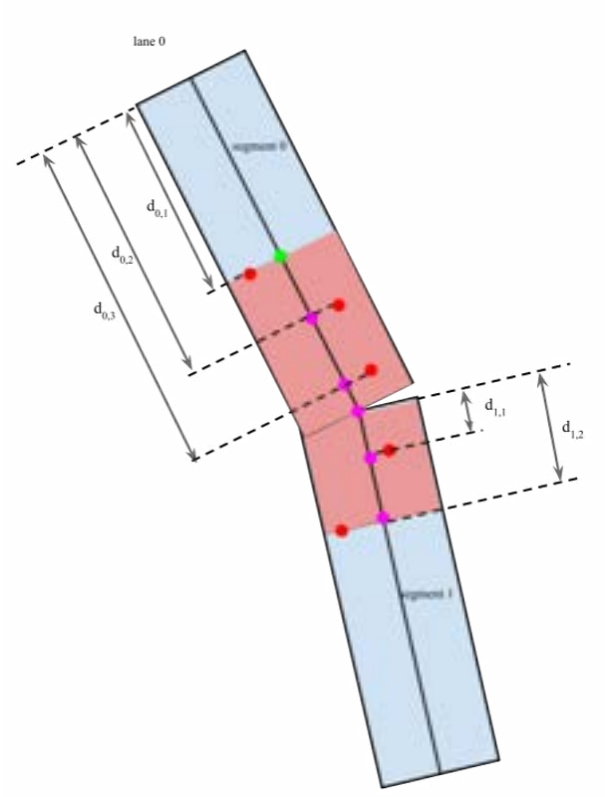


FIGURE 37: TIM Work Zone.

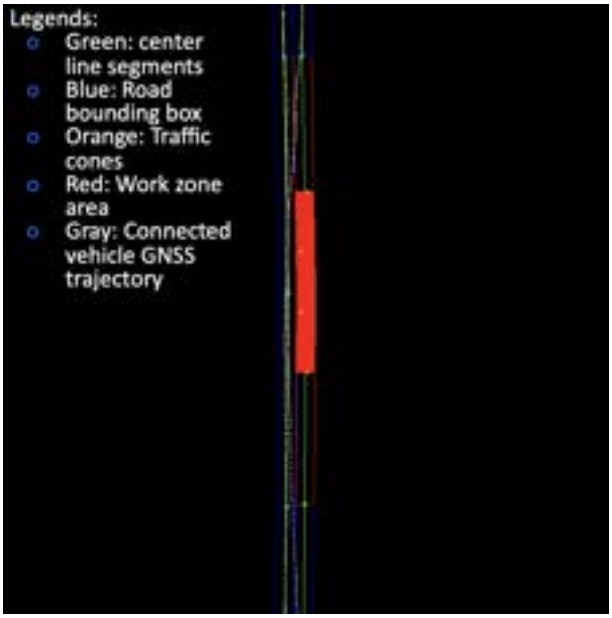
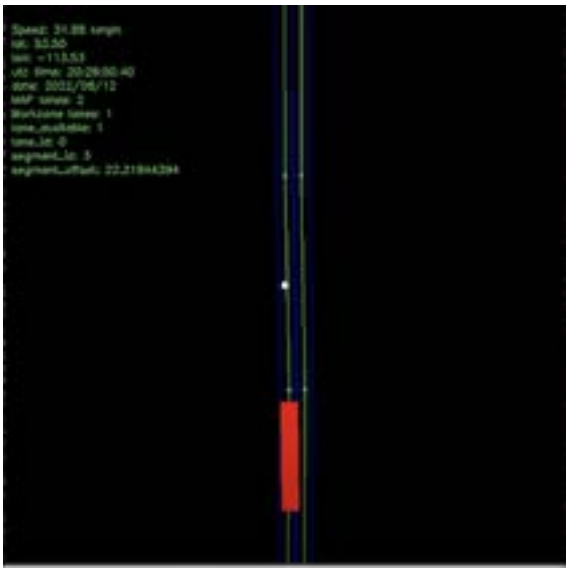


FIGURE 38: Render MAP and TIM.

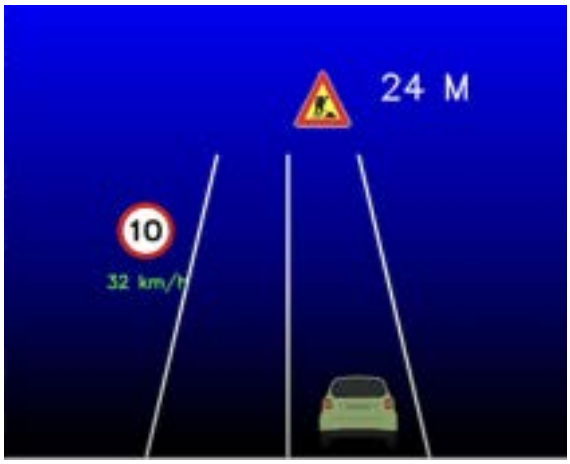
FIGURE 39: Hardware.

FIGURE 40: MAP Data Collection.

(a) YOLO detection.



(b) MAP and TIM.

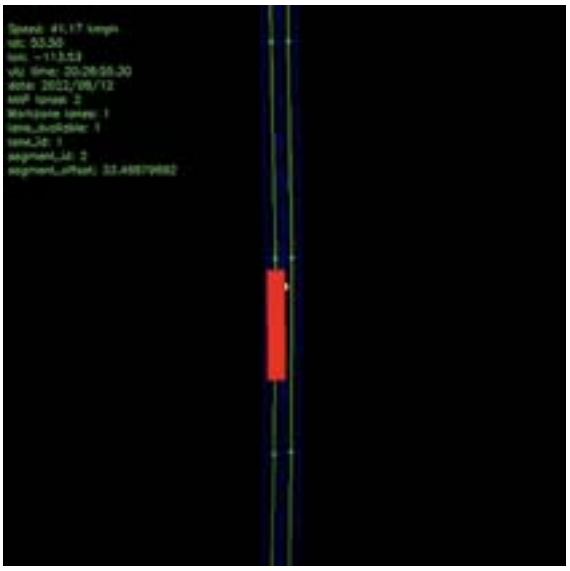


(c) Warning APP.

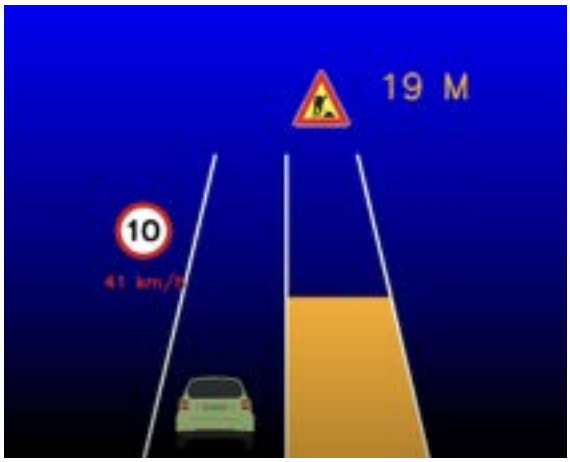
(d) Driver's vision.

FIGURE 41: Approaching the Work Zone.

(a) YOLO detection.



(b) MAP and TIM.

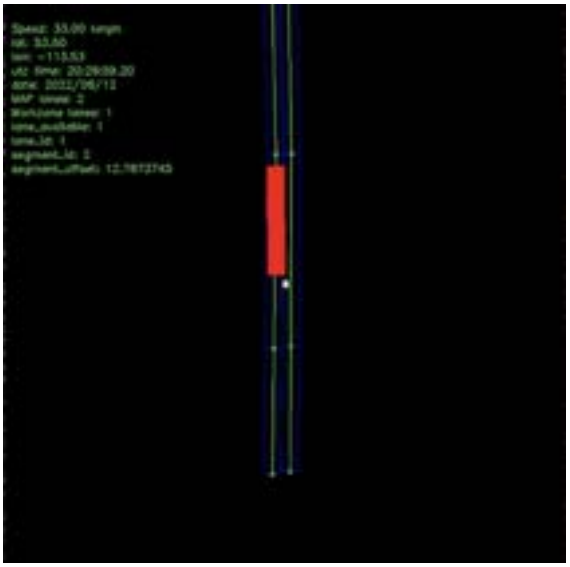


(c) Warning APP.

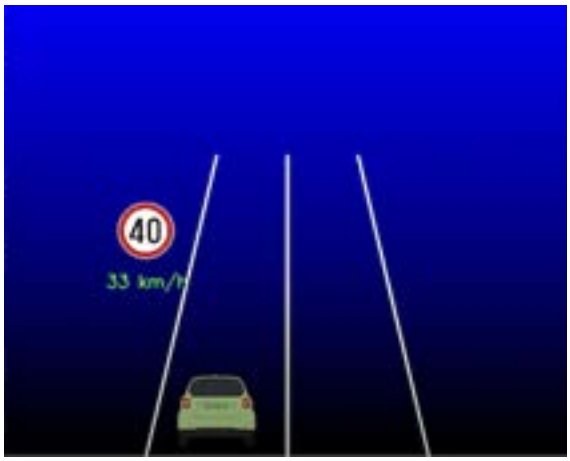
(d) Driver's vision.

FIGURE 42: Entering the Work Zone.

(a) YOLO detection.



(b) MAP and TIM.

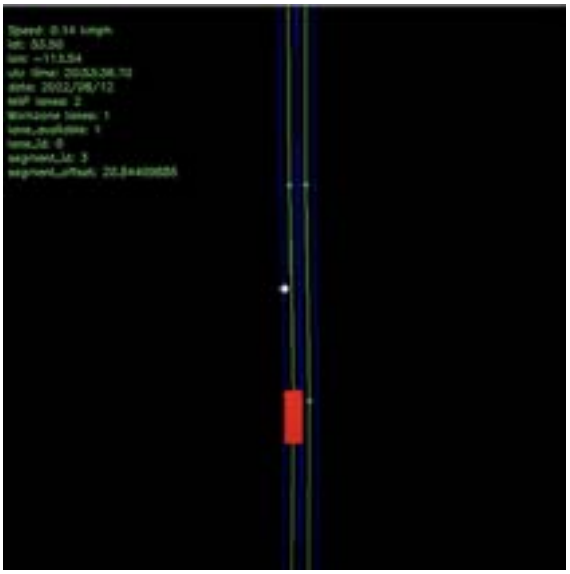


(c) Warning APP.

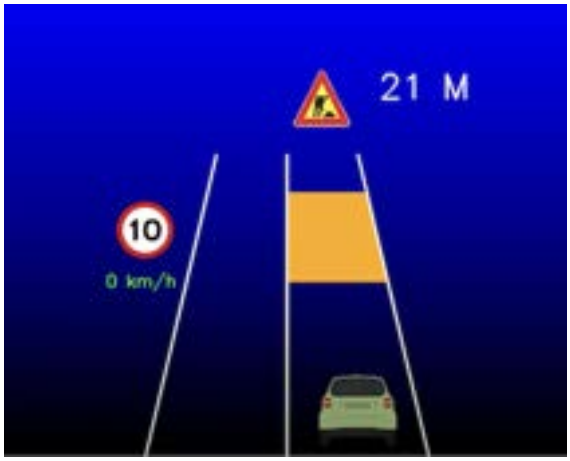
(d) Driver's vision.

FIGURE 43: Leaving the Work Zone.

(a) YOLO detection.



(b) MAP and TIM.

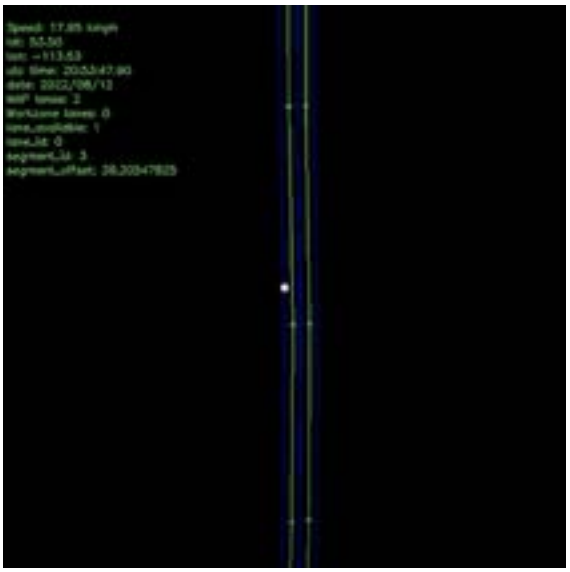


(c) Warning APP.

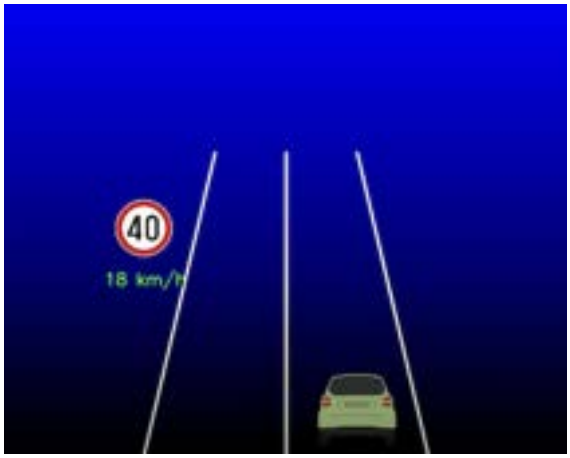
(d) Driver's vision.

FIGURE 44: Half of Traffic Cones were Removed.

(a) YOLO detection.



(b) MAP and TIM.



(c) Warning APP.

(d) Driver's vision.

FIGURE 45: All Traffic Cones were Removed.