North Carolina
Department of Transportation
Research Project No. 2019-28

Enhancing AV Traffic Safety through Pedestrian Detection, Classification and Communication

November 2022

This page is intentionally blank.

1. Report No. **FHWA/NC/2019-28**	overnment Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle **Enhancing AV Traffic Safety through Pedestrian Detection, Classification and Communication**		5. Report Date **November 2022**
2.		6. Performing Organization Code
7. Author(s) R. Thomas Chase, Jing Feng, Seth Hollar, Ali Karimoddini, Waugh Wright		8. Performing Organization Report No.
9. Performing Organization Name and Address **Institute for Transportation Research and Education** **North Carolina State University** **Centennial Campus Box 8601** **Raleigh, NC**		10. Work Unit No. (TRAIS) 11. Contract or Grant No. NCDOT RP 2019-28
12. Sponsoring Agency Name and Address **North Carolina Department of Transportation** **Research and Analysis Group** **104 Fayetteville Street** **Raleigh, North Carolina 27601**		13. Type of Report and Period Covered **Final Report** **August 2018 – July 2021** 14. Sponsoring Agency Code **NCDOT/NC/2019-28**
Supplementary Notes: Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration		

16. Abstract

Safety for all road users is a key concern as Connected and Autonomous Vehicle technologies develop and reach the testing phase. One key concern is the change to two-way communication that often occurs in traditional pedestrian-vehicle interaction. This project focused on three key aspects to this issue. First, a prototype autonomous shuttle system was expanded to include additional communication features in a lightbar as well as more advanced pedestrian detection systems. Secondly, multiple detection methods were trained and tested using traditional datasets as well as a new dataset including occluded pedestrians. Finally, a survey was conducted to determine how well pedestrians understand specific static or dynamic lightbar patterns as an additional communication tool for CAVs.

This project tested multiple pedestrian detection methods and developed improved methods with increased accuracy and reduced latency. The EcoPRT vehicle was able to incorporate the improved detection method, however the training image set included multiple camera perspectives and the method could likely be applied to infrastructure-based detection systems. Additionally, the project developed a body part-based method which detects head, arms and legs of pedestrians in order to improve the overall detection of pedestrians when they are partially occluded. The project also developed a database of occluded pedestrian images which can be used for training or testing other new methods addressing this issue. Finally, the project examined multiple methods for signaling the CAV intent to pedestrians using fixed or moving lightbars. Respondents struggled to correctly identify the message communicated by the lightbar in cases where multiple movements are expected (such as locations with potential turning movements) but identification improved in more constrained environments.

17. Key Words CAV, Pedestrian-Vehicle Interaction, Autonomous Shuttles, Pedestrian Detection, Pedestrian Occlusion	18. Distribution Statement No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 46	22. Price $199,728

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Disclaimer

The contents of this document reflect the views of the authors and are not necessarily the views of the Institute for Transportation Research and Education or North Carolina State University. The authors are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the North Carolina Department of Transportation or the Federal Highway Administration at the time of publication. This report does not constitute a standard, specification, or regulation.

Acknowledgments

The research team thanks the North Carolina Department of Transportation for supporting and funding this project. We are particularly grateful to the Steering and Implementation Committee members and key stakeholders for the exceptional guidance and support they provided throughout this project:

Executive Summary

NCDOT has several Connected and Autonomous Vehicle (CAV)-related projects ongoing throughout the state including vehicle testing, connected infrastructure and multiple research projects sponsored at universities. Safety has been a critical concern for all of these projects, and one key gap identified in the literature was the change to two-way communication that often occurs in traditional pedestrian-vehicle interaction. Traditional pedestrian-vehicle communication of intent relies on vehicle/pedestrian dynamics, signaling and non-verbal communication with drivers (typically eye contact).

This project brings together an interdisciplinary team including the Transportation Human Factors group members at ITRE and NCSU Psychology, the EcoPRT Autonomous Vehicle team from NCSU and the TECHLAV Center from NCAT. The overall goal of the project was to examine how pedestrians and CAVs can effectively communicate their intent where these modes conflict. This project focused on three key aspects to this issue. First, a prototype autonomous shuttle system was expanded to include additional communication features in a lightbar as well as more advanced pedestrian detection systems. Secondly, multiple detection methods were trained and tested using traditional datasets as well as a new dataset including occluded pedestrians. Finally, a survey was conducted to determine how well pedestrians understand specific static or dynamic lightbar patterns as an additional communication tool for CAVs.

This project tested multiple pedestrian detection methods and developed improved methods with increased accuracy and reduced latency. The EcoPRT vehicle was able to incorporate the improved detection method, however the training image set included multiple camera perspectives and the method could likely be applied to infrastructure-based detection systems. Additionally, the project developed a body part-based method which detects head, arms and legs of pedestrians in order to improve the overall detection of pedestrians when they are partially occluded. This issue is less severe in infrastructure-based detection systems with elevated cameras that avoid most obstructions, but is very important in CAV pedestrian detection. The project also developed a database of occluded pedestrian images which can be used for training or testing other new methods addressing this issue. Finally, the project examined multiple methods for signaling the CAV intent to pedestrians using fixed or moving lightbars. Respondents struggled to correctly identify the message communicated by the lightbar in cases where multiple movements are expected (such as locations with potential turning movements) but identification improved in more constrained environments.

This page is intentionally blank.

Table of Contents

List of Tables

List of Figures

Introduction

NCDOT has several CAV-related projects ongoing throughout the state including vehicle testing, connected infrastructure and multiple research projects sponsored at universities. Safety has been a critical concern for all of these projects, and one key gap identified in the literature was the change to two-way communication that often occurs in traditional pedestrian-vehicle interaction. Traditional
1. pedestrian-vehicle communication of intent relies on vehicle/pedestrian dynamics, signaling and non-verbal communication with drivers (typically eye contact).

This project focused on three key aspects to this issue. First, a prototype autonomous shuttle system was expanded to include additional communication features in a lightbar as well as more advanced pedestrian detection systems. Secondly, multiple detection methods were trained and tested using traditional datasets as well as a new dataset including occluded pedestrians. Finally, a survey was conducted to determine how well pedestrians understand specific static or dynamic lightbar patterns as an additional communication tool for CAVs.

This report presents a literature review on these areas in Chapter 2, then summarizes the EcoPRT vehicle development in Chapter 3. Chapter 4 discusses the challenges and solutions for pedestrian detection, followed by the AV-Pedestrian communication study in Chapter 5. Finally, Chapter 6 summarizes the findings, potential application areas and future research.

Literature Review

2.1. Autonomous Vehicle Types and Development Progress

As autonomous vehicles continue to progress and evolve, it is important to note that autonomous vehicles can take on differing flavors and offer different capabilities and usages. Traditionally, autonomous vehicles have conceptualized as typical passenger car platforms that have autonomous technology integrated on them. These vehicles are restricted to typical uses on highways and other roadways. Recently, there have been additional advances that point out other solutions. A small variant is the driverless shuttle that holds 10 to 12 people still operating on roadways but going along a fixed path to move people in more urban situations. Vehicles such as EZ Mile and Optimus Ride are two examples that fit this form factor. SB 739 also highlights that there is ongoing development of delivery vehicles much smaller not intended to move passengers but instead to deliver cargo and goods. Cities as well are planning for a multifaceted deployment of autonomous vehicles for their transit solutions. With Singapore planning for completely autonomous transit by 2030, Ongel points out the upfront costs for driverless shuttles compared to buses but their TCO is much less (Ongel, 2019). The current development of EcoPRT falls in line with this evolving landscape. Operating much like a driverless shuttle, it initially follows a fixed route. It's weight and width may fall within the large size of vehicles SB 739 guidelines, however, and the farther-reaching goal is to look at autonomous transit vehicle solutions that can not only operate on existing roadways but also operate on smaller pedestrian or bicycle paths. Others have recognized the advantage of these smaller vehicles to interact in a mixed mode setting. Woodman examines the human factors element of such small vehicles platooning in these mixed mode settings (Woodman, 2019). The vehicles they use are being developed by RDM Group which currently has a contract to build and deploy 40 of these small autonomous vehicles within a city in UK (RDM).

2.

Furthermore, the lightweight, small format of the vehicle allows one to explore the use of elevated guideway to move these vehicles around, and given the size and weight, the associated cost of the elevated guideway could be a substantial cost savings when compared to separated infrastructure of other transit solutions. In this project, with the continued development of these 5 autonomous vehicles, this provides a platform for multi-vehicle autonomous testing that can really examine a number of facets yet to be explored including vehicle interactions, passenger throughput, and some farther-reaching research endeavors such as information sharing for safety and alerts. Critical to the underlying goal of autonomous vehicle research, access to the underlying detection and control provides unique insight not available to DOTs when contracting or leasing privately developed autonomous vehicles. The benefits of AV research are not limited to situations where public research is further developed than private technology, as private firms developing autonomous vehicle platforms are highly protective of algorithms and provide no open access to vehicle data. Open data and algorithms allow for transparent policy or regulatory decision making and has been a highly successful approach for connected vehicle technologies where USDOT and ITS JPO are able to independently verify data, cybersecurity and protection of personal information and develop informed standards.

2.2. Pedestrian Interaction with AVs

Despite the fast growing body of literature (for a review, see Rasouli & Tsotsos, 2019), and the implementation of LED stripe light in commercialized models of some highly automated vehicles (e.g., Mercedes-Benz F015, S500), it remains unclear how to effectively communicate the vehicle's intent to the pedestrian. There remain large differences among the implementations. For example, the Mercedes-Benz models use LED stripes to indicate the sensed direction of a pedestrian walking in front of the vehicle to show that the vehicle knows the existence of the pedestrian. Some car models (e.g., Mitsubishi electric, 2015) use LED lights to indicate their intended direction of travel. Others use LED lights to instruct pedestrians to perform certain actions (e.g., Matthews et al., 2017, "Cross Now").

This lack of standardization in vehicle pedestrian communication could cause confusion of pedestrians and impair road safety. For example, rightward flowing arrows could be interpreted as the vehicle will turn right (if communicating the vehicle's intended move) or the pedestrian can move to the right of the vehicle (if communicating an instruction). As the context is very important in the interpretation of abstract and ambiguous symbols/signals, it is possible that pedestrians' interpretation could be influenced by the vehicle's motion, road layout, and other contextual or situational information (e.g., pedestrians' understanding/experience, other road users' action), which is valuable to understand. Although there is an increasing awareness in the field to understand pedestrian interaction with autonomous vehicles in various contexts and under a range of traffic situations (Kaβ, et al., 2020), very little research has been done about these contextual effects. In addition, it is also important to consider how policy within a community, city, or state can shape LED communication standards to enhance the safety of pedestrians.

Another important aspect of vehicle pedestrian communication which has been relatively overlooked in the literature is the attentional state of a pedestrian. A large proportion of pedestrians walk in areas such as a campus (or population dense residential areas) while being distracted both visually and auditorily. This brings challenge to the effectiveness of communication and begs the question of whether communicating specific intention/instruction could be useful or it is simply more effective to alert pedestrians whenever the vehicle changes its path or speed.

2.3. Pedestrian Detection

The Decision-making process in an AV is ideally a multi-modal sensing mechanism that receives and fuses information from multiple sensors to perceive the environment and react accordingly (Rasouli & Tsotsos, 2019). Practically, however, we are limited by computation resources; accuracy, range, and update rate of sensors; high-computation and imperfect performance of object classification techniques which are still at their infancy. Therefore, existing detection mechanisms in AVs primarily rely on radar or Lidar sensors which can reliably detect an obstacle but cannot reliably classify whether it is a pedestrian or not (Flores et al, 2019). Challenges specific to pedestrian detection include varying appearance of pedestrians (in different cloths, color, size) and their dynamic shape, as well as the sensitivity of cameras to environmental noises (weather, shadow, illumination) (Gerónimo et al., 2010). In addition, EcoPRT is supposed to be driven in population-dense environments such as downtown or university campuses, which are considered dynamic cluttered environments, exposing specific challenges regarding pedestrian detection, amongst them occlusion is one of the most challenging problems (Zhang et al., 2018).

In recent years, numerous approaches for detecting and tracking pedestrians in sequential images have grown steadily. With the recent advancement in deep learning, we can utilize machine learning models to accurately detect and classify pedestrians in complex scenarios. In this section, we will begin with a brief overview of learning-based pedestrian detection, then some existing fusion techniques for combining multimodal sensor data, and finally a brief overview of pedestrian tracking.

Learning-based Pedestrian Detection

Pedestrian detection from RGB images is an important yet difficult task. Recent works focus on improving the robustness and accuracy using deep neural networks. (Tian et al., 2015; Zhao et al., 2016; Zhang et al., 2018; Zadobrischi and Negru, 2020; Zhang, Yange and Schiele, 2018) Though these methods exhibit satisfactory performance in well-lit environments, they struggle to detect pedestrians in low light conditions such as nighttime, dawn, sunrises, and sunsets. This is because it is hard to generate shape information from images in ill-lit environments.

2.3.1.

On the other hand, the LiDAR can provide comparatively better shape features under these scenarios. LiDAR-based feature extraction for pedestrian detection is studied in early research (Premebida, Ludwig and Nunes, 2009). As LiDAR can provide the only geometric feature of pedestrians, inferring context-aware relations of pedestrians' body parts is one way to distinguish among multiple pedestrians in a complex scenario (Oliveira and Nunes, 2010). While using the LiDAR sensor data —the distance, intensity, and width of the received pulse signal in each scanning direction, pedestrians can be classified using a clustering algorithm (Ogawa et al., 2011). To improve the classification performances of pedestrians, hand-crafted features such as slice feature and distribution of reflection intensities are explored (Kidono et al., 2011). The slice provides human body information based on body height and width ratio. Some works are focused on the density enhancement method for improving the sparse point cloud of LiDAR and they provide an improved shape feature for long-distance pedestrian detection (Li et al., 2015; Lin et al., 2019). Pioneering work on the conversion of a 3D point cloud of LiDAR into the 2D plane extracts both hand-crafted features and learned features, and then trains a support vector machine (SVM) classifier to detect pedestrians (Chen et al., 2020). Later, 3D point clouds are converted into 2D panoramic depth maps and these depth maps are used in pedestrian detection (Premebida, Ludwig and Nunes, 2009). Even though the LiDAR provides better results in the nighttime while it is difficult to get shape features using the camera or compare to a distorted image frame, camera-based methods perform better for long- distance pedestrians in the daytime where they appear in small sizes. The best result can be achieved by fusing both of these sensors to jointly predict pedestrians.

2.3.2.

Pedestrian Detection by Fusing Sensors

Since using the LiDAR or the camera independently un- veils their own limitations, it becomes an interesting research direction to fuse different sensor modalities. In this setting, the improvement can be achieved from the use of multiple views of the pedestrian by learning a strong classifier that accommodates both different 3D points of view and multiple flexible articulations. In order to integrate multiple sensor modalities, several fusion mechanisms are investigated (Premebida, Ludwig and Nunes, 2009; Premebida and Nunes, 2013; Premebida et al., 2014; Gonzalez et al., 2015; Schlosser, Chow and Kira, 2016; Matti, Ekenel and Thiran, 2017; Kim et al., 2018). These sensor fusion techniques mostly focus on either

combining feature information from different sensors or generating candidate regions from one sensor and map these candidate regions to other sensor information. For instance, a deformable part detector is trained using optical images and depth images generated from 3D point clouds using upsampling technique (Premebida et al., 2014). Some fusion techniques cluster the LiDAR point cloud to generate candidate regions and map these regions on an image frame for detecting pedestrians (Matti, Ekenel and Thiran, 2017; Lahmyed and Ansari, 2016). Most of these methods sacrifice runtime performance while improving detection accuracy. Therefore, a balanced fusion mechanism is needed to deal with the trade-off between accuracy and speed.

Pedestrian Tracking

Recent pedestrian trackers are designed mostly based on end-to-end deep learning networks. A common approach is adding recurrent layers with the detector module. For example, the ROLO (Ning et al., 2017) is the combination of the convolutional layers of YOLO and the recurrent unit of LSTM. TrackR-CNN (Voigtlaender et al., 2019) is considered as a baseline method of multi-object tracking that adds instance segmentation along with multi-object tracking. Tractor++ is an efficient multiple object tracking that utilizes the bounding box regression on predicting the position of an object in the next frame where there is no train or optimization on tracking data (Bergmann, Meinhardt and Leal-Taixe, 2019). Besides, a single object tracking along with semi-supervised video object segmentation based on siamese neural network is introduced in (Wang et al., 2019). On the other hand, the machine learning pipeline-based methods such as the Deep SORT which integrates appearance information along with Simple Online and Realtime Tracking (SORT) technique, adopts a single hypothesis tracking methodology with the recursive Kalman filter and the frame-by-frame data association. This technique focuses on an offline pre-training stage where the model learns a deep association metric on a large- scale person re-identification dataset (Wojke, Bewley and Paulus, 2017). A single-stage efficient multi-object tracking is introduced in (Wang et al., 2019), where target detection and appearance are embedding to be learned in a shared way, and a Kalman filter is used for predicting the locations of previously detected objects in the current frame. While considering the LiDAR data for pedestrian tracking, a stochastic optimization method is introduced in (Granstrom et al., 2017) that merges the clustering and assignment task in a single stage. Inspired by these works, we use both LiDAR and camera sensors to complement individual sensor limitations on detection and tracking performances. Thus, our solution can be applied to a wide variety of complex scenarios.

2.4. Public Agency Pilot Research

NCDOT developed the Connected Autonomous Shuttle Supporting Innovation (CASSI) program to learn more about how this technology can be safely and effectively used in the future to offer additional mobility solutions, to help familiarize people with new transportation technologies, and to encourage environmentally-friendly transportation solutions (NCDOT/NCSU 2020). While the EcoPRT vehicle and CASSI are both forms of autonomous transit, they are differently sized and the research/program scopes and objectives diverge significantly. CASSI operates only on roadways with a fixed route and human operator carrying up to 12 passengers. EcoPRT is designed to operate on roadways or mixed-use paths at least 10’ in width currently on a fixed route with no human operator and up to two passengers. CASSI prior to COVID was operating at up to 8mph with plans to potentially reach 12mph during the NCSU deployment, while EcoPRT operates at up to 15mph on the roadway. More recently, NCDOT has deployed CASSI at the Wright Brothers National Memorial in Kill Devil Hills, where it makes two stops along its

approximately 1-mile fixed route between the Wright Brothers National Memorial museum and the First Flight bronze sculpture. NCDOT has used this deployment to test a new roadway environment as well as survey riders on their experience.

It is noted that pilot deployments of autonomous transit vehicles have rapidly progressed across the US with similar scope and operational limitations to CASSI. In addition to agency pilot tests, universities often partner with deployments to provide additional research into the shuttles with the higher profile programs funded in the tens to around $100 million. At the University of Florida, a partnership with the City of Gainesville and FDOT called I-STREET is testing a comprehensive CAV system with autonomous shuttles as well as connected signals and additional detection technologies. The University of Michigan's Mcity Test Facility provides a controlled environment built to model many urban facilities and research has evaluated public opinion of AVs and recommended a safety evaluation framework for AV technologies. VTTI has a past research project testing participants reactions in virtual reality to different passenger car light bar patterns, as well as developed an ontology and evaluation framework for vehicle and pedestrian interaction although it was only tested using simulated scenarios.

Autonomous Vehicle Development

3.1. EcoPRT Summary

3.

Figure 1 EcoPRT System Vision

This report is a follow on to the development of microTransit vehicles listed in NCDOT Project 2017-025 "Feasibility and Demonstration of Small Automated Vehicles as a Viable Transit Solution in NC." We are currently in development of a system named EcoPRT. As a continuation to the aforementioned project, we continue to build and develop five microTransit vehicles for a system targeting mobility in closely congregated areas such as college campuses, corporate campuses, large shopping centers, airports, fair grounds, sports complexes, amusement parks, etc. Such environments being too far to walk while at the same time too short to drive are not easily addressed with current transmit options.

The key characteristics of EcoPRT include (Hollar *et al*., 2017):

- *Flexibility*. Vehicles can run on existing paths or on dedicated guideways. Compared to existing solutions that a) rely solely on dedicated guideways or b) rely exclusively on existing infrastructure, EcoPRT is a unique hybrid of the two. As a rubber-tired vehicle, it can be operated on existing concrete roadways as a low-speed automated vehicle, and, as a light-weight vehicle, the cost and load requirements of elevated dedicated roadways is substantially less when compared to other vehicles.
- *Low cost*. Light-weight, small footprint vehicles reduce infrastructure costs. A two-person, fully laden EcoPRT vehicle weighs 1,000 lbs., much lighter than conventional automobiles or other PRT systems. Consequently, elevated guideways would require less support loads and therefore could be built at less cost.
- *Convenience*. Automated vehicles would be on demand, allowing point-to-point travel without stopping, all hailed by a smart phone.
- *Organic growth*. EcoPRT's flexibility allows a system to be installed quickly at low cost (even using a single vehicle). Adding additional vehicles or expanding the routes is still a relatively low cost/short term effort allowing EcoPRT to grow incrementally as demand grows.

This research helped to accelerate development and deployment of a working at-grade EcoPRT demonstration system on NCSU's campus. In all, this the following sections discuss the following items:

- Status of vehicle builds
- Future Pilot Testing
- Suspension Evaluation for Generation 2.5 Vehicles
- Pedestrian Detection
- AV LED and Sound Design
- Planned future interaction studies

3.2. Status of vehicle builds

Though the goal of the build was to complete five fully functional vehicles, Covid-19 slowed down the development. To date, we have one fully functional vehicle (version 2.0) followed by another partially functional vehicle (version 2.5). Three other vehicles are in later stages of construction since the last update. Figure 2 and Figure 3 show the vehicles as they are now.

Figure 2 Completed (nearly) two vehicles currently being tested with graduate students

Figure 3 Three vehicles currently being constructed (third one not shown)

Through the course of the project, we made a number of notable improvements and advances on the manufacturing of the vehicles. We worked with machine shop at the Mechanical and Aerospace Department to build the components for the suspension. Further, we worked with the Biological and Agricultural Engineering Research Shop for the construction of the frames and welding tasks.

In all, we had four steel/aluminum hybrid frames built. The steel allowed for a stiffer frame improving riding dynamics, and the aluminum upper frame reduced the overall weight. In addition to the frames, the suspension, steering, and locomotion linkages for all four vehicles were designed and build. The steering was improved form the Gen 2.0 vehicle by including a stronger motor within a rack and pinion steering linkage. A redesigned motor controller increased the motor current amperage rating from 5 amps to 15 amps allowing a quicker and stronger steering turning torque. Suspension was designed with a tighter turning radius in mind. With such vehicles, pedestrian bicycle paths could have sharp turns, so the EcoPRT vehicles were designed to have turning radii on the order of 10 feet. Given their narrow frames, however, careful attention was paid to the suspension, especially, while turning. Our Gen 2.0 vehicle had a shared axle frame and was relatively wobbly on turns. The newer Gen 2.5 vehicles included a lower center of gravity, independent suspension, tuned shock absorbers, and an anti-rollbar which all contributed to reduced rolling component.

Initial field testing of the version 2.5 vehicle was conducted and affirmed the design choices for a smoother riding experience. Changes to the steering joints needed to be made stronger so we migrated from a ball joint to a “C” joint. Figure 4 shows a picture of the improve steering linkage.

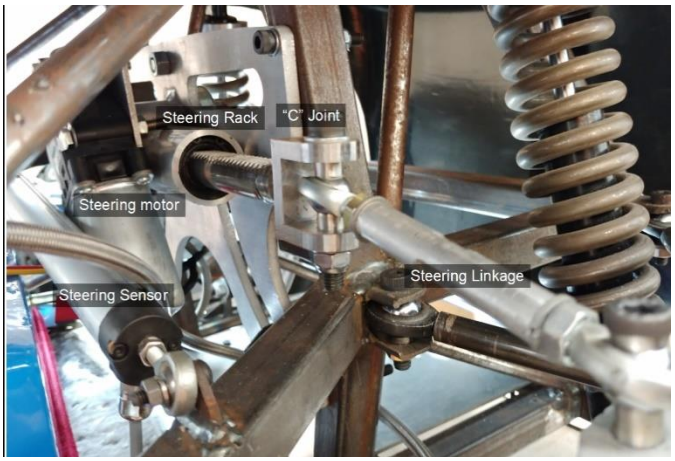

Figure 4 EcoPRT of steering assembly

Battery improvements have been made as well. Initially, we used two 36V 35Amp-hour Lithium batteries in series, but we found in initial testing that the current max was too limited. We upgraded to a single 72V 50 Amp-hour Lithium battery pack with a max sustained current of 120amps with peak currents well above that. As a very compact design, the battery dimensions are 350 x 266 x 150 mm. Figure 5 shows the battery within the vehicle. Currently, we are designing a battery enclosure for improved protection.

Figure 5 Battery within vehicle

Currently, the two of the vehicles have the most extensive wiring done are the gen 2.0 and one of the newer gen 2.5 vehicles. We have manually field tested both of them. Notable improvement in power and torque is seen in the gen 2.5 vehicles. As opposed to the gen 2.0 vehicle, at full power, the gen 2.5 vehicle can easily lose traction on the concrete and spin the wheels.

In Figure 6, the various electrical components are situated under the seat in the back of the vehicle. The brake actuators have been upgraded to supply more force (150 lbs) in the gen 2.5 vehicles. Further, the motor controller board has been modified to recharge the auxiliary batter during operation. Power relays were chosen to handle upwards of 400 amps at a relatively low control voltage of 12 volts. Brake actuators were placed in a parallel arrangement to further reduce space. There is a USB Hub to handle the

multiple USB devices including the inertial measure unit, the motor controller board, GPS unit, and additional sensors.

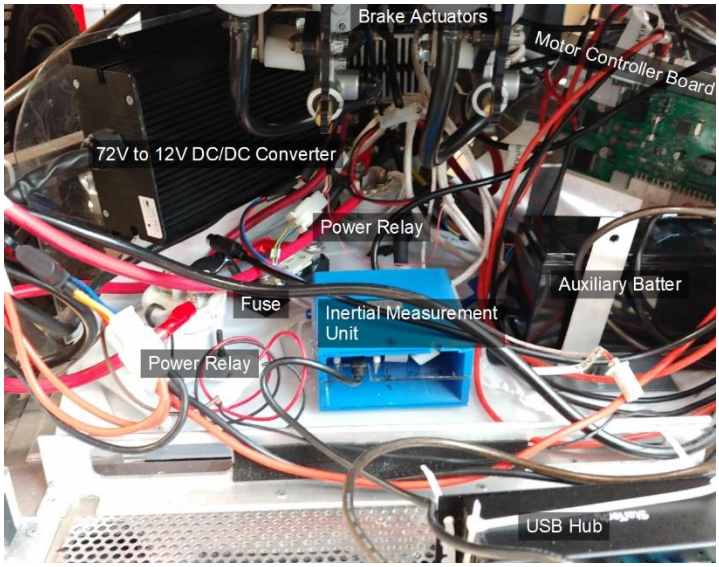

Figure 6 Electronics layout within gen 2.5 vehicles.

We have upgraded our ROS system to the latest version ROS Melodic. This required that we also upgrade our OS to Ubuntu 18.04 LTS. The upgrade allowed us to take advantage of latest driver and library versions and additionally migrate from the 16.04 LTS version that is no longer being officially supported. As part of the migration, we further finalized our shift to the Swift Navigation's differential GPS, the Piksi Multi. With centimeter level accuracy with a steady output even within an urbanized landscape, the GPS unit has proven to be reliable. Currently, we use a base reference station for the differential GPS reference, but it is expected that Swift Navigation could be providing a reference difGPS signal that may obviate the need for our own base station. We have further upgraded our sensors used for object recognition including the industrial ZED stereo camera and Velodyne's VLP-16. Both sensors are mounted on the front of the vehicle and provide continuous sensor data for object recognition and collision avoidance as seen in Figure 7.

Figure 7 Front of vehicle showing Stereo Camera and 2D Lidar.

3.3. Future Pilot Testing

The goal with the build of the five vehicles is multifold. As part of this report, the vehicles provide a testbed for evaluating human AV interactions. Further, once vehicles are completed, a pilot test will be performed at NC State to better understand the utility of microTransit solutions on campus. Working with the Institutional Review Board (IRB), testing of the vehicles was divided up into three phases. In the first phase tests the vehicle on a test track located at the loading dock near the vehicle lab. Phase 2 tests the vehicle on an isolated parking lot area. The goal of the second phase is to run identical tests but in in a different environment. Phase 3 is the final test where we take passengers on the vehicle to capture ridership information. Initially for the university's IRB approval (Institutional Review Board), we simplified the approval process by seeking approval of just Phase 1, and with success, we would then seek approval for phases 2 and 3.

3.4. Riding Dynamics for Generation 2.5 Vehicles

Contributing members:

Graduate Students: Nikhil Patil

As part of the vehicle development, we examined the process of designing and optimizing the suspension system for the prototype vehicle. The objectives of the prototype development are building a small, low cost, lightweight, and comfortable vehicle. The version 2.0 build of the vehicle lacks enough roll stiffness or a smooth ride. As such, a complete redesign of the suspension system for the version 2.5 build of prototypes was desired. The Short-Long Arm (SLA) double wishbone suspension with outboard coil was

the design of choice for the new prototype. To evaluate the ride and safety, a quarter car model was evaluated for suspension travel, body acceleration, and dynamic wheel load over a pseudo-random road profile. The results from these models showed a comparison between the two prototype vehicles in relation to their ride comfort and safety. For lateral stability, a few performance metrics were evaluated, and the two designs are compared by their body roll angle against steady state lateral acceleration. The design is validated by comparing the yaw rate and roll rate data from the simulation and road tests. As shown below, the chassis has increased floor space in Version 2.5, and the overall vehicle body has increased torsional stiffness.

Figure 8 Front and back suspension geometries for generation 2.5 vehicle.

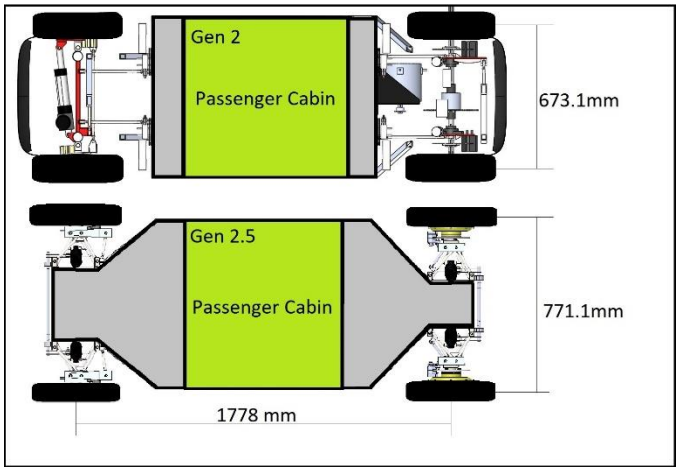

Figure 9 Chassis floor space (colored gray) comparison between Gen 2 (top) and Gen 2.5 (bottom) vehicles

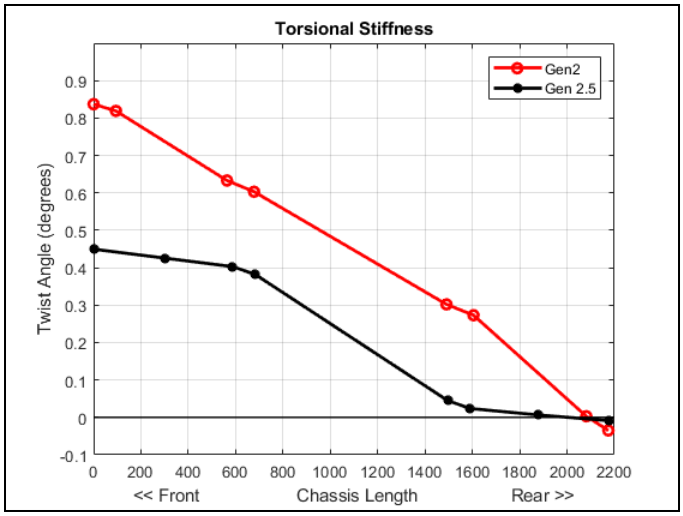

Figure 10 Torsional Rigidity Comparison between Gen 2 and Gen 2.5 Chassis

3.5. Pedestrian Detection and Path Estimation

Contributing members:

Graduate Students: Srinivas Gopalakrishnan, Deveshwar Hariharan

As part of the effort in the project, we evaluated and tested the pedestrian detection algorithm supplied by NC A&T University partners as part of the detection element in the pedestrian communication. Furthermore, graduate student Srinivas Gopalakrishnan worked on pedestrian detection and pedestrian path prediction in his Master's Thesis entitled, "Application of Neural Networks for Pedestrian Path Prediction

in a Low-cost Autonomous Vehicle." (Gopalakrishnan, 2020) Here he worked on pedestrian detection in addition to applying novel algorithms for path prediction. Though not directly used in this project, the path prediction of pedestrians can be used in bi-directional communication between the vehicle and pedestrian as a part of the collision avoidance.

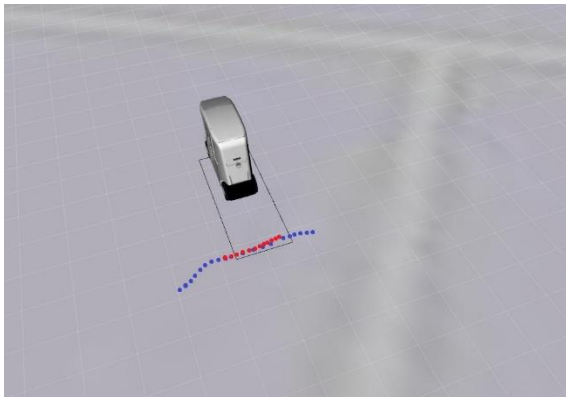

Figure 11 Vehicle simulation of path prediction

Figure 11 is the prediction by the algorithm called Social-General Adversial Network(S-GAN) (Gupta et al., 2021), where it recognizes pedestrians, keeps track of them over successive frames, and makes predictions on where they are possibly headed. In this image, the blue dots represent the ground truth values of the pedestrian's movements, while the red dots represent the predictions of the pedestrian movement from the algorithm.

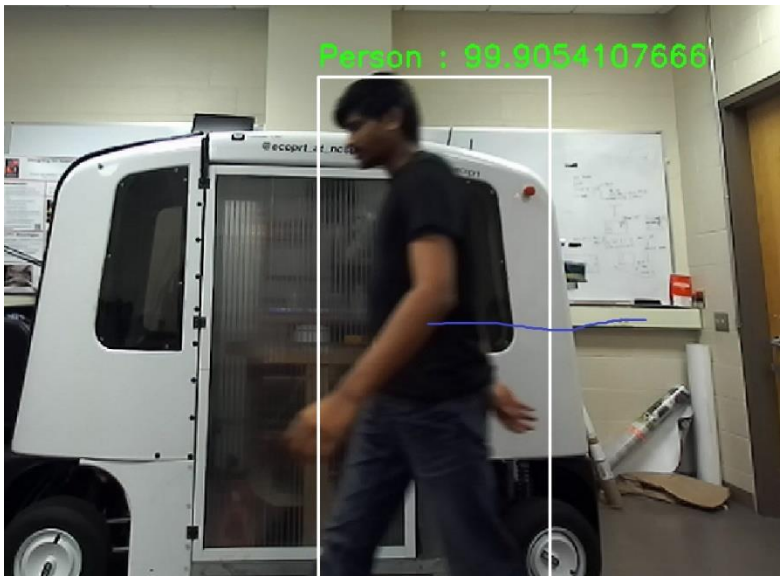

Figure 12 Camera capture, recognition, and path plotting

Both "You only look once" (YOLO) (Redmon, 2021) and Single Shot MultiBox Detector (SSD) (Liu et al., 2021) algorithms for object detection were implemented. In Figure 12, the camera is able to recognize the object in the image as a person with about 99% confidence.

Additionally, this was captured directly from a camera in real-time. The blue path represents the trajectory of the person over time, effectively being able to track them. This information is used in the tracking and prediction algorithms to estimate the location of the pedestrian over a given period of time.

Figure 13 was captured by the vehicle within the ROS framework. With the camera on the vehicle, the ROS system is able to track the person in real-time. These pedestrian detection algorithms will be used as the vehicle uses LED's and sound as a means of alerting the pedestrians to the intent of the vehicle.

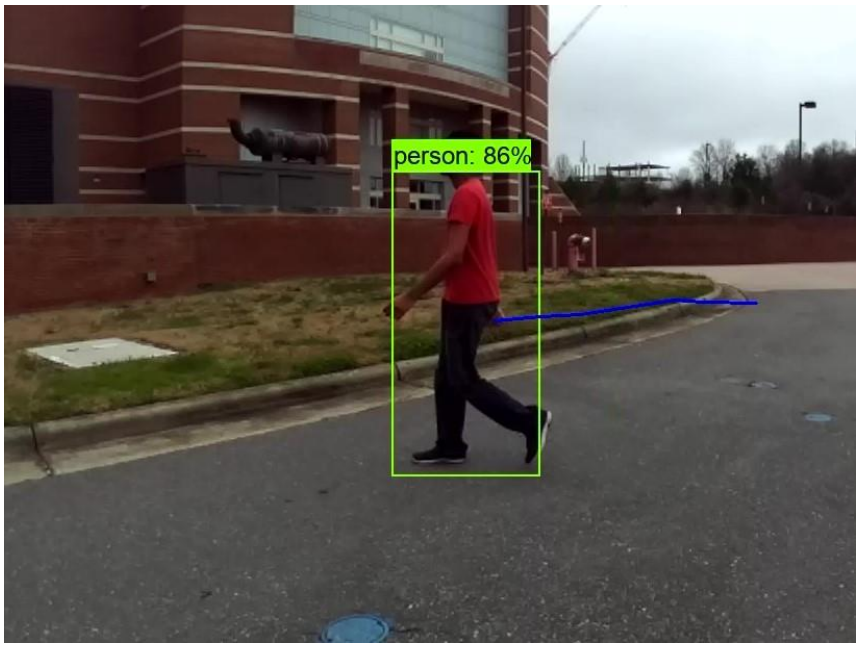

Figure 13 Capturing through the ROS framework along with path estimation

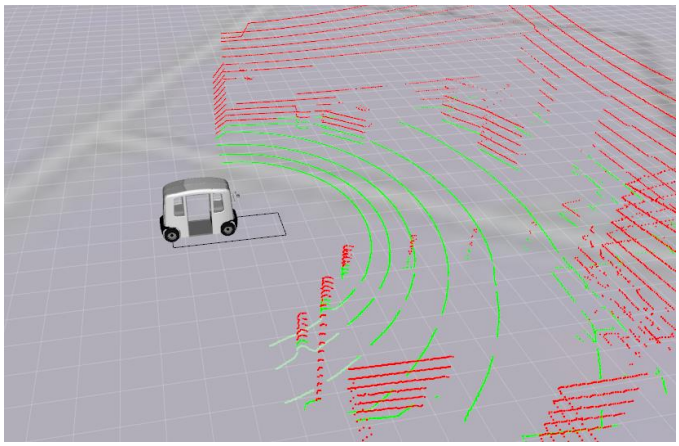

Figure 14 Lidar ground and obstacle detection.

Figure 14 contains the image captured through the visualizer of the ROS framework. Here the lines represent the sensor input from the Lidar to the computer. The green lines represent the ground and the red lines represent the potential obstacles to the vehicle. This is a demonstration of ground plane detection on the vehicle. The EcoPRT vehicles contain both camera data and Lidar data as part of the object detection sensors. Further research is ongoing to blend the sensor data together for improved accuracy.

3.6. AV LED Light and Sound Design

Contributing members:

Senior design students: Jiachen Zhao, Wenqi Jiang, Zhen Chen, Wuge Wang

Graduate students: Deveshwar Hariharan, Srinivas Gopalakrishnan, Abhishek Singh

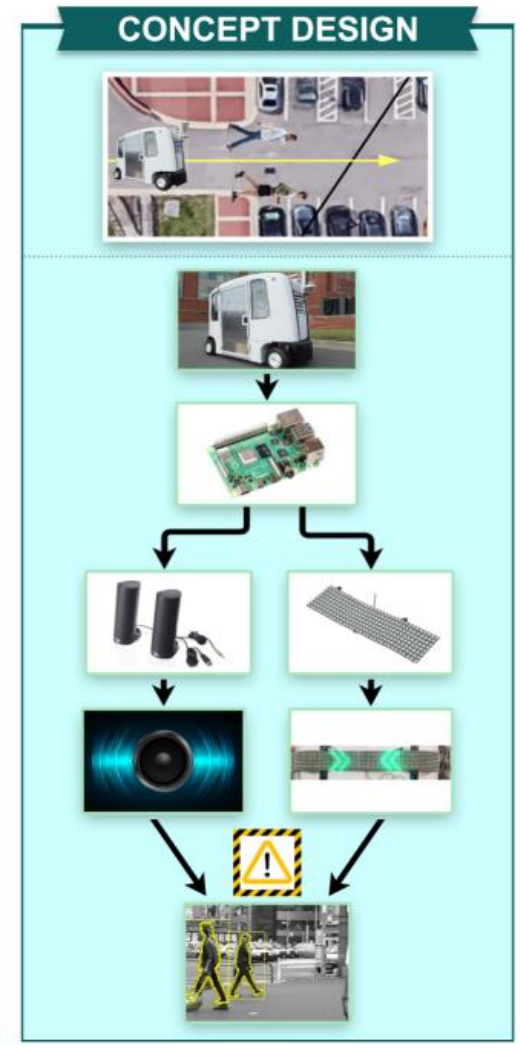

Figure 15 Concept Design of LEDs and Sound

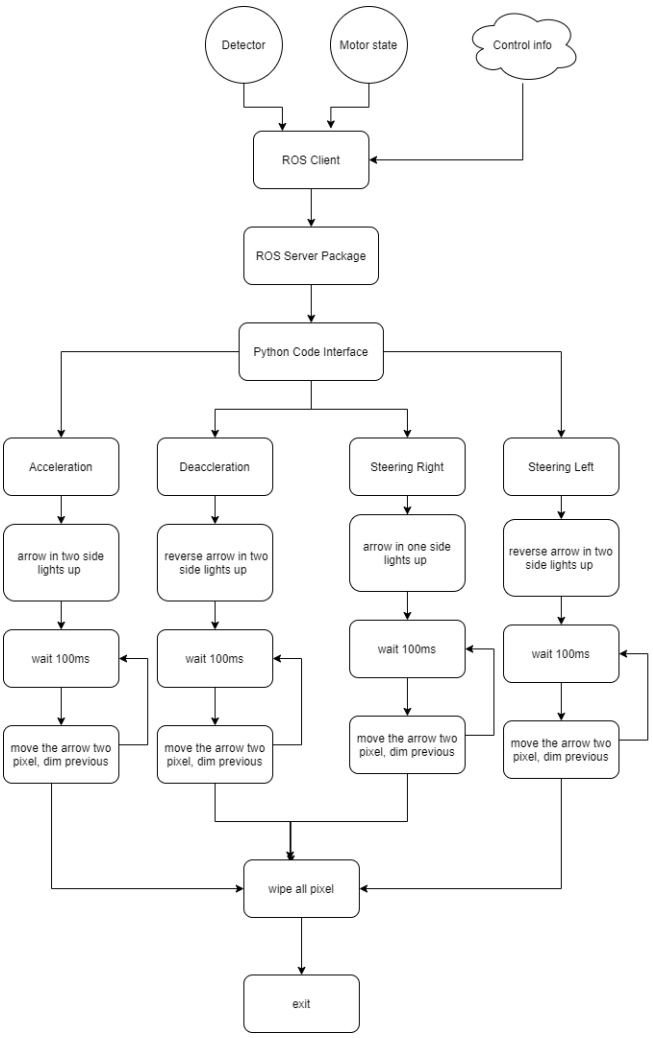

Figure 16 System Operational Flowchart

The electrical engineering senior design team worked on implementing the hardware for the pedestrian visualization and sound. Our initial experiments focused solely on the LED animated display to communicate to the pedestrian, and further experiments will also incorporate sound and automated speech from the vehicle to communicate to the Pedestrian.

Figure 15 Concept Design of LEDs and Sound

Figure 16 System Operational Flowchart

shows the concept design of the solution with a Raspberry PI interfacing to the ROS package and controlling both the LED lighting and the sound.

Figure 16 shows the flow diagram as originally implemented to automatically determine the type of signal to use. The four cases showing acceleration, deceleration, turning right, and then turning left. Additional cases could be easily incorporated as well.

Initial demonstrations of the LED display capability were successful in Spring of 2021. Videos were recorded for the purposes of the online survey that is shown in the results of this report. Additional investigations will continue with an in-person study as described below.

3.7. Future In-Person Study Pedestrian Interactions with Autonomous Vehicle

As part of the final report, we investigated people's interpretations of signals from autonomous vehicles through a survey. In addition to the survey, we intend to intend to continue the effort of doing live testing of human pedestrian interactions as well. This is currently unfinished, but the follow section details the experiment and what is involved.

During the experiment, the participant will act as a pedestrian walking along the sidewalk and cross a small intersection (see map below). Sometimes, an autonomous vehicle will be driving toward the intersection where the participant will be crossing. The participant will need to determine when to cross the intersection safely. The participant should behave as they normally would crossing a typical intersection. In some experimental trials, the participant will be required to interact with the phone text messaging function while walking, just as a pedestrian on a phone would do.

The autonomous EcoPRT vehicle will not have a rider but instead be programmed to travel one of two paths marked in the map below (Path #2 and Path #1).

The EcoPRT vehicle will follow a trajectory autonomously going at 5mph. The vehicle will be programmed to cross the path of the participant in the parking lot when taking path #1. In addition, the vehicle includes obstacle detection so that it can automatically stop in the presence of an oncoming pedestrian.

Figure 17 The study area for the pedestrian / AV interactions

Pedestrian Detection and Intent Analysis

Pedestrian tracking is a challenging problem as pedestrians need to be firstly detected in the current frame and then associated frames. The success in deriving a good tracker is mainly governed by a superior detector. In this work, we use YOLOv5 (Jocher et al., 2021) as a base module of the pedestrian detector. Therefore, we will briefly explain the working principle of YOLOv5 as a pedestrian detector. Since the

4. YOLOv5 architecture is the same as the YOLOv4 (Bochkovskiy, Wang and Liao, 2020) except the training procedure, we will describe the architecture of the YOLOv4 and then will highlight the training differences.

Figure 18 illustrates the architecture of YOLOv4 which can be segmented into four major parts: input, backbone, neck, and output. In the input section, the network takes an image and completes a data augmentation procedure that uses a data loader for scaling, color space adjustments, and mosaic augmentation. Among these augmentation techniques, mosaic augmentation firstly introduces in YOLOv4. The mosaic augmentation combines four training images into one in certain ratios to simulate four random crops which help to detect small-scale and partially occluded pedestrians.

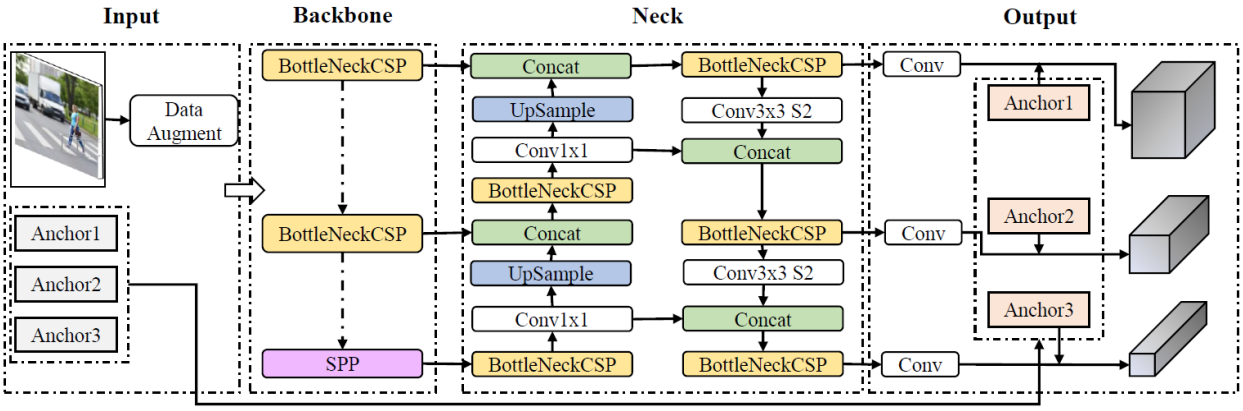

Figure 18 Architecture of YOLOv5

After data augmentation, the augmented image is feed into the backbone of the network. In the backbone section, a BottleNeckCSP is used which is a modification of DenseNet (Huang et al., 2017). Using BottleNeckCSP different shallow features like edges, colors, etc., are extracted. During training, the backbone module learns these features. Besides, an additional Spatial Pyramid Pooling (SPP) block is used to increase the receptive field and separate the most important features from the feature maps of the BottleNeckCSP. The next part of the network is the neck part where the network enhances the understanding and extraction of the shallow features adopted in the backbone part. To do that a Path Aggregation Network (PANet) is used that includes a bottom- up augmentation path in conjunction with the top-down path used in Feature Pyramid Network (FPN). The PANet processes combine and analyzes the extracted features and finally optimizes based on the target of the model. The last part of the network is output where the model yields the detection results using dense predictions. Dense predictions provide a vector by combining predicted bounding boxes and confidence scores for the classified pedestrians.

During the training process of YOLOv5, the floating-point precision is set to 16 bit instead of the 32-bit precision used in YOLOv4. Therefore, YOLOv5 exhibits higher performances than Yolov4 under certain circumstances.

4.1. Methodology

In this section, we will explain our proposed methodology. At the high level, we propose Fused-YOLO which is an integrated framework of multi-modal sensor information to track pedestrians at a real-time speed. At the core of our proposed solution lies the distinction between improving detection accuracy and limiting the computational complexity.

Conversion from LiDAR Scans to Depth Images

4.1.1.

Camera-based pedestrian detection systems suffer from either low illumination or over-exposed images. It is a better idea to complement the system with a LiDAR which acts as the primary depth sensor due to its high accuracy and long sensing range. LiDAR scan produces sparse point clouds, albeit this representation of data is rather challenging to incorporate as an input to neural networks. Instead, depth images are better correspondents of point clouds that are easy to manipulate constructively. Therefore, we convert a 3D LiDAR scan to the depth image in 2D image space. Formally, the LiDAR stream consists of a sequence of registered 3D scans $\{S_1, S_2, ..., S_t\}$ arriving at time points $t_1, t_2, ..., t_t$. Each scan S_t is a point cloud, i.e., a set of 3D points, $S_t = \{p_1, p_2, ..., p_i\}$ and $p_i := \{x, y, z\}$ represents the Euclidean coordinate. Due to the huge amount of memories that are required over time, it is inefficient to work upon the raw point clouds. Instead, we can convert the 3D scan to a 2D depth map. A depth image can be thought of as a 2D grid map comprised of un cells. To generate the depth image we need to compute the distance of the scan objects from a viewpoint in such a way that maps pi to u. Then, we transform each point in point clouds from the Euclidean coordinate (x, y, z) to Spherical coordinate (θ, φ, r). This way we can map each point to the corresponding grid cell such that $u : \{\theta, \varphi\} \rightarrow r$. The pixel values of depth images lie in either gray or RGB color spaces. For the grayscale image, we normalize each cell value in the grid map to $0 \rightarrow 255$ to the known maximum depth value and thus the intensity of the gray image represents the depth information. On the other hand, For the RGB scale image, we assign a distinct color from the RGB space to each cell value in the grid map based on the r parameter.

Fusion with Depth Image

4.1.2.

Our goal is to predict pedestrians in a joint space that combines both the RGB and the depth spaces. Although detecting pedestrians in RGB images is a common practice, there are few ways to incorporate depth images to detect pedestrians in a joint space. An end-to-end deep learning network takes an RGB image and corresponding depth image as input to generate joint predictions over pedestrians. In another setting, an RGB image and a depth image can be processed sequentially using a single network. However, in the former case, the network architecture becomes very complex to be able to process depth and RGB images in an end-to-end fashion. In the latter case, the network requires to process sequential call which causes a huge runtime overhead in a long run.

Our solution utilizes the Kalman filter along with parallel processing of RGB and depth images. To predict in a joint space, first, we project the LiDAR scan as a depth image to the RGB camera space. Let x_r and x_d be the RGB and depth images, respectively. Since (in common settings) the positions of camera and LiDAR are fixed but the resolution of camera image and the depth image from the LiDAR scan varies in size, we can project the depth image to the RGB camera space with either zero padding to the smaller image or cropping each of them into a same size. We denote this synchronized depth image by x_s. Second, we vertically concatenate the RGB image x_r and the zero-padded depth image x_f by resizing each of them into a fixed size such that $x = \{x_s, x_r\}$. Although it is possible to use an image classification network to predict pedestrians directly over x_s, it requires multiple calls for the joint prediction, i.e., the x_r and the x_s need to be fed to the pedestrian detection and pedestrian classification models, respectively. Thus, concatenating the x_r and the x_s into x reduces the number of calls to different models and significantly improves the runtime efficiency. Finally, we feed this concatenated image x to the pedestrian detector f to obtain bounding boxes and scores over fused images such that $\hat{y} = f(x)$. From the prediction $\hat{y}$, we can also separate individual predictions the $\hat{y}_r$ and the $\hat{y}_s$ for the x_r and the x_s correspondents, respectively. Since we vertically concatenate the x_s with the x_r, we can calculate an offset o based on the height of x_r. Then, we translate each bounding box $b_s \in \hat{y}_s$ to the down by o.

Overlaying $\hat{y}_s$ to $\hat{y}_r$ may raise three distinct types of scenarios. Firstly, $\hat{y}_s$ reduces the miss detection by accurately detecting pedestrians. Secondly, $\hat{y}_s$ provides redundant inference with respect to $\hat{y}_r$. Finally, $\hat{y}_s$ does not improve detection accuracy since it cannot detect any pedestrians. To overcome these scenarios, we utilize a Kalman filter to evaluate the joint predictions systematically. In our next subsection, we will describe the proposed Kalman filter in detail.

4.1.3. Integrated Framework for Pedestrian Tracking

The Kalman filter has been extensively applied in pedestrian tracking from the camera stream. Our framework uses such a technique to predict and update the pedestrian trajectories from the continuous camera and LiDAR streams. Our integrated framework augments the capability of the existing pedestrian tracking method by fusing depth information. To track multiple pedestrians in a frame, our framework uses three important information, i.e., bounding boxes from the RGB images, optical flow between consecutive RGB image frames, and bounding boxes from the depth images.

One of the important properties of the Kalman filter is that the state vector is a hidden parameter and the observation provides useful information to update the state vector. Therefore, in our setting while using the Kalman filter, the observations, i.e., bounding boxes, from the detector are not directly useful for tracking pedestrians. Basically, the proposed Kalman filter-based tracking has two stages: the prediction and the update stages. In the prediction stage, the bounding boxes for pedestrians are predicted using the corresponding state of the bounding boxes in the previous frames. In the update stage, the observation of pedestrians in the current frame is used to update the predicted states of pedestrians.

Let s_t^i be the state vector of i^{th} bounding pedestrian window in frame t. To track multiple pedestrians, it is convenience to have multiple Kalman filters, e.g., one for each pedestrian detected in the frame as follows:

$$\mathbf{s}_t^i = \mathbf{F}_{t-1}^i \mathbf{s}_{t-1}^i + w_{t-1},$$
$$\mathbf{z}_t^i = \mathbf{H}_t^i \mathbf{s}_t^i + v_t,$$

where F_{t-1}^i and H_t^i denote the state transition and the measurement matrices for the *i*th pedestrian, respectively. The vectors w_{t-1} and v_t are noise terms which are assumed to be Gaussians with zero mean and covariance matrices Q_t and R_t. The prediction stage involves reasoning about the state vectors and their associated error covariance matrices at time t given the measurements up to t − 1 as follows:

$$\mathbf{s}_{t|t-1}^i = \mathbf{F}_{t-1}^i \mathbf{s}_{t-1}^i + w_{t-1},$$
$$\mathbf{P}_{t|t-1}^i = \mathbf{F}\mathbf{P}_{t-1|t-1}^i \mathbf{F}^T + Q_t.$$

Next, in the update stage updates, the state vectors and their error covariance matrices with the current observations are as follows:

$$\mathbf{K}_t^i = \mathbf{P}_{t|t-1}^i \mathbf{H}^T \left(\mathbf{H}\mathbf{P}_{t|t-1}^i \mathbf{H}^T + R_t\right)^{-1},$$
$$\mathbf{s}_{t|t}^i = \mathbf{s}_{t|t-1}^i + \mathbf{K}_t^i \left(\mathbf{z}_t^i - \mathbf{s}_{t|t-1}^i\right),$$
$$\mathbf{P}_{t|t-1}^i = \left(I - \mathbf{K}_t^i \mathbf{H}\right) \mathbf{P}_{t|t-1}^i,$$

where K_t^i is the Kalman gain which emphasizes how pre- diction and measurement are intimately related. Therefore, the process of fusion begins with identifying observation models and associated measurement noises for each observation modality. For instance, in our settings, the bounding boxes from the RGB and depth images are considered as positional information whereas the optical flow provides the velocity information only. This way we can assign a separate observation model for updating the joint prediction state.

4.1.4. Detecting Occluded Pedestrians

In this section, we address the problem of detecting body parts of pedestrians using deep neural networks. In particular, we consider the occluded pedestrian detection problem in autonomous driving settings. While state-of-the-art deep neural models perform reasonably well for detecting full-body pedestrians, their performances are not satisfactory for occluded pedestrians. Introducing a new training strategy along with a fusion mechanism, we enhance the performance of the SSD-Mobilenet and the Faster R-CNN by utilizing body parts information to handle occluded pedestrians.

In public datasets, we have found it is challenging to find detailed labeling of body parts, e.g., the Caltech dataset or the CityPerson dataset do not have any body part label for detecting occluded pedestrians. On the other hand, it is convenient to label different body parts on the Penn-Fudan dataset (Wang et al., 2007). However, the downside of Penn-Fudan dataset is small. Therefore, we create a dataset with detailed body parts for 1500 images. We segregate the full-body pedestrian into three parts: head, arm, and leg. Thus, covering the most region of a pedestrian's body. We introduce variation in our dataset by collecting data from different conditions, e.g., variation in illuminations and sizes, occlusion by objects, indoor and outdoor environments. Although our dataset is small, we introduce variation by collecting data from different places around the world, i.e., India, Bangladesh, Malaysia, Dubai, and USA.

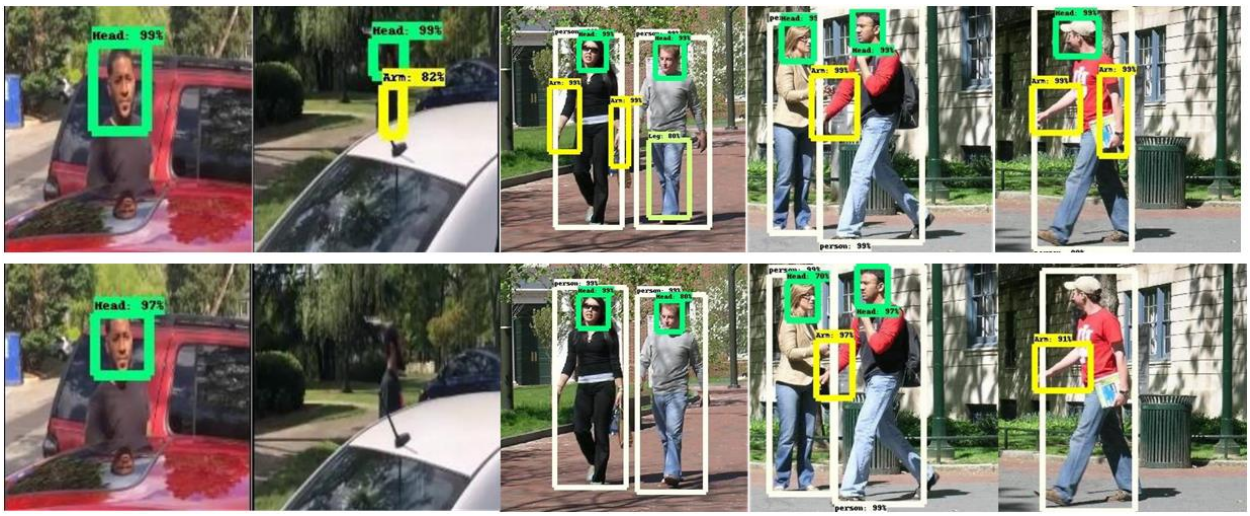

Figure 19 Body part detection

In Figure 19, the first row shows the detection results of Faster R-CNN and the second row shows the detection results of the SSD-Mobilenet. Yellow, green, light green and white boxes represent arm, head, leg and person, respectively. It is evident from the figure that the Faster R-CNN exhibits higher detection rate in contrast with the SSD- Mobilenet.

4.2. Experiment Results

We evaluate the pedestrian detection performances in terms of Miss Rate (MR) vs False Positive Per Image (FPPI) and also provide accuracy, precision, recall, and run-time efficiency of the model. We conduct our experiments on a 64-bit Ubuntu 18.04 server that has an Intel(R) Core(TM) i9-7900XCPU @ 3.30GHz with 64GB memory. In our setup, we also have an NVIDIA GeForce RTX 2080 GPU with 8GB memory.

4.2.1. Dataset

We perform our experiments on the Waymo open dataset which contains a wide range of diverse examples since data are collected among Phoenix, Mountain View, and San Francisco cities in the USA, plus it contains daytime and nighttime driving data (Sun et al., 2020). This dataset is recently released and comprising of large-scale multimodal sensor data, i.e., high-resolution camera and LiDAR data. In particular, the dataset is collected using five LiDAR sensors and five high- resolution pinhole cameras and contains four object classes: Vehicles, Pedestrians, Cyclists, Signs. It has 12.6M high- quality 3D bounding box labels in total for 1,200 segments for LiDAR data. On the other hand, it has 11.8M 2D tightly fitting bounding box labels in total for 1,000 segments of camera data. In our setup, we use front camera images and project LiDAR data onto their corresponding camera images.

Performance Analysis

We evaluate our proposed Fused-YOLO and the baseline YOLOv5 in terms of Miss Rate vs False Positive Pes Image (FPPI) curve in Figure 20. Testing on 456 numbers of images from the Waymo dataset, the proposed Fused-YOLO shows the miss rate of 33.435% whereas the YOLOv5 has the miss rate of 41.945%. This is because fusion helps more for accurately detecting pedestrians in ill-lit conditions. In low illumination conditions, the shape of pedestrians given by depth images entails useful features for pedestrian detection, which are more challenging to detect from camera images. Especially, we notice that the Fused-YOLO achieves significantly less miss rate in low illumination conditions in contrast to the baseline YOLOv5 model. On the other hand, fusion without the Kalman filter exhibits a 38.602% miss rate. Because when naively fusing the bounding boxes from the depth and RGB images, it increases the false detection. On contrary, the Kalman filter provides a systematic approach to reduce false detection and achieves the lowest miss rate.

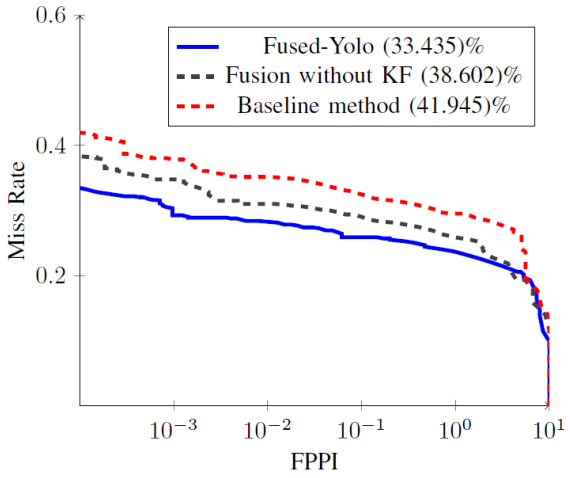

Figure 20 The comparison among Fused-YOLO, Fusion without applying Kalman Filter, and Baseline YOLOv5

Table 1 Evaluation Metrics

Model	FP	TP	FN	Accuracy	Precision	Recall
YOLOv5	12	105	75	0.546	0.897	0.583
Fused-YOLO	12	117	63	0.609	0.900	0.650

Table 1 represents that our proposed Fused-YOLO method has better accuracy, precision, and recall compared to the YOLOv5. Tabular data shows that the YOLOv5 shows the accuracy of 0.546, precision of 0.897, and recall of 0.583, whereas we obtain the accuracy of 0.609, the precision of 0.900 and, recall of 0.650 using the proposed Fused-YOLO. The downside of Fused-YOLO is of course increased false detection Rate when combining with the prediction on depth images. One of the possible reasons is the fact that YOLOv5 did not train on depth images. Therefore, our pre-trained YOLOv5 struggles to accurately detect pedestrians on depth images.

We find that our tracking method performs very well even in the cases where there are some miss detections in sequential frames. Figure 21 illustrates the performance of our tracker on a sequence of images where four pedestrians are detected in four consecutive frames and our method can track all of them. The green and cyan bounding boxes in the top row represent detected and tracked pedestrians, respectively. Figure 22 shows that there are two pedestrians detected in the first frame and then in the second frame model failed to detect one of the pedestrians. However, with the help of Kalam filter in the next two consecutive frames that pedestrian is detected and tracked again which represents the efficient performances of our method. Our tracking method uses a Kalman filter to predict multiple pedestrian bounding boxes. Then, fusing the detection results from RGB images and depth images, the Kalman filter update pedestrians' state estimation. Correlating the previous bounding boxes to the current estimation, the Kalman filter can track pedestrians even if the detector might fail to detect pedestrians on the current frame. Furthermore, additional detected bounding boxes from depth images help the Fused-YOLO to track the pedestrians robustly. We observe that the Fused-YOLO achieves negligible runtime performance overhead (28 FPS) in contrast to the baseline YOLOv5 model (30 FPS).

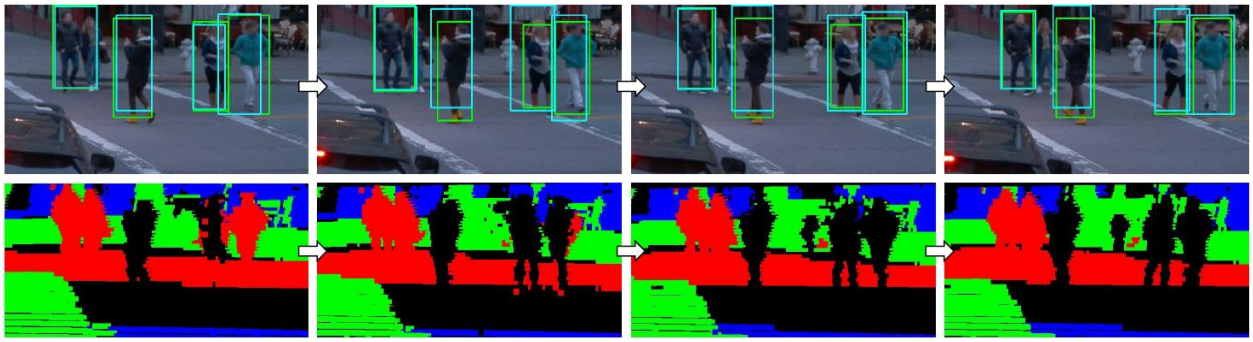

Figure 21 Multiple Pedestrian Tracking

Figure 22 Robust Pedestrian Tracking

In conclusion, we developed a real-time accurate pedestrian detection and tracking framework by fusing camera and LiDAR sensor data. The developed framework is integrated with the Kalman filter to detect and track multiple pedestrians accurately and robustly. The novelty of our framework lies in the adoption of the Kalman filter for both sensor fusion and tracking applications while minimizing the overall runtime

overhead. Experimental results demonstrated the improvement over the baseline YOLOv5 model. Our fusion method outperforms YOLOv5 in terms of detection and tracking accuracy with a negligible amount of runtime overhead. The difference becomes even more pronounced in ill-lit conditions when pedestrians are hard to find in camera images. Finally, our implementation of the Kalman filter along with the optical flow algorithm reduced the detection miss rate and improved the overall performance. Future work includes extending our tracking method for pedestrians' behavior/intention analysis.

Autonomous Vehicle and Pedestrian Communication

5. In this study, we examined the factors affecting pedestrians' interpretations of LED lighting on autonomous vehicles. Because of the importance of vehicle motion which influence pedestrians' crossing behavior, we aimed to explore whether motion will affect the participants' interpretation of each LED light pattern. In addition, signals have two types, meaningful type and attention capturing type. We also explored whether meaningful LED signal will affect the participants' interpretation of vehicle intent. This study adopted mTurk platform to carry out experiments. Participants watched videos of an authentic vehicle with LED stripes, which defined as "an autonomous vehicle". And they should finish a LED light interpretation task by using descriptive words or sentences. In this study, we coded participants' answer and calculated the interpretation correctness as the index of the communication quality. We hypothesized that AVs' intention in motion condition would be better interpreted; participants would be better at interpreting the intentions of vehicles with meaningful LED light pattern.

5.1. Methodology

This study was a 4 (group: (1) motion groups including group 1, group 2 and group 3; (2) motionless group including static group) × 5 (LED light pattern: (1) meaningful patterns including "arrows going in", "arrows going out", "arrows to the left" and "arrows to the right"; (2) attention capture pattern including "flashing squares") mixed design, the design for each group is shown in Figure 1. Group was a between-subject factor and LED light pattern was a within-subject factor. In this study, participants completed an LED light interpretation task and a demographic survey.

5.1.1. Participants

Participants were recruited using Amazon Mechanical Turk (mTurk), compensated with 1 dollar given completion of the study. Eighty participants (72.5% male, 2 participants didn't report; range: 23 – 71 years, M_{age} = 35.98 years, SD_{age} = 9.92 years) completed the experiment.

Materials

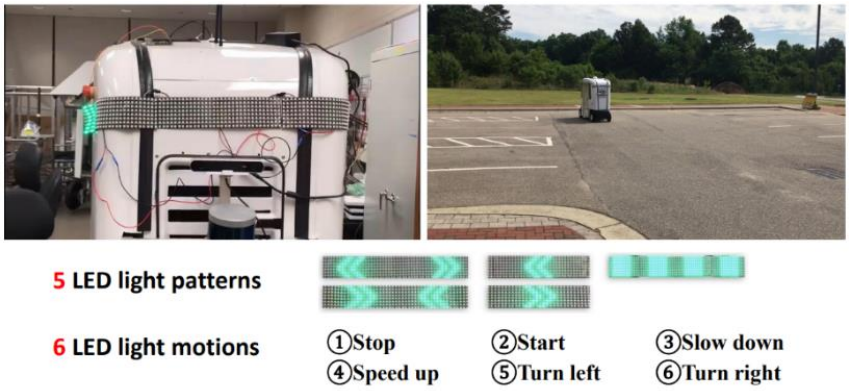

Figure 23 An example of a motionless condition (a) and a motion condition (b).

5.1.2.

The LED light interpretation task. Participants needed to watch videos and complete interpretation questions. The videos were recorded by an authentic autonomous vehicle with an LED screen. In motionless condition (static group), participants could only see the pattern on the LED screen and could not observe the motion of the autonomous vehicle, as shown in Figure 23a. In motion condition (group 1, group 2, and group 3), participants could both see the pattern on the LED screen and the motion of the car, as shown in Figure 23b. Furthermore, the groups in motion condition are separated in two conditions related to LED display, meaningful display (vehicle perspective signal and pedestrian perspective signal) and attention capture display. For participants in static group, after watching a video, they needed to complete one question that was describe the perceived meaning of each LED light pattern and were encouraged to write down all possible meanings. There were 5 trials in static group. For participants in other three groups, after watching a video, they needed to complete two questions. First, participants should describe the LED light pattern first to make sure they saw the signal with the vehicle in motion clearly. After that, they needed to describe the perceived meaning of each LED light pattern and were encouraged to write down all possible meanings. There were 6 trials in each motion group.

The demographic survey. The demographic survey aimed to get basic demographic information of participants, including age, gender, education, living area, frequency of public transportation use, self-reported general health, vision, hearing and memory conditions.

5.2. Procedure

Participants were directed to the Qualtrics page after they selected the study on the mTurk. Participants were required to read the general instruction of the experimental procedure and then completed a consent form. Once consent was received, they first read the specific and detailed instruction of the study. This instruction included size of the vehicle, purpose of the LED light pattern and content of tasks. Then participants accomplished *the LED light interpretation task*. One trial of the LED light interpretation task covered one video viewing (almost 15 seconds) and one or two question(s) answering (e.g., please describe your perceived meaning of each LED light pattern. If you think there is more than one possible meaning, please describe all in the order of most strongly perceived meaning to the least.), the number of questions and trials varied in different groups (details are in the materials section). In the same group, the order of the problems was random among the participants. Following the completion of the LED light interpretation task, participants are required to describe the factors which made the signal's meaning clear or ambiguous, answer the size of display and finish *a demographic survey*. The detailed procedure was presented in Figure 24.

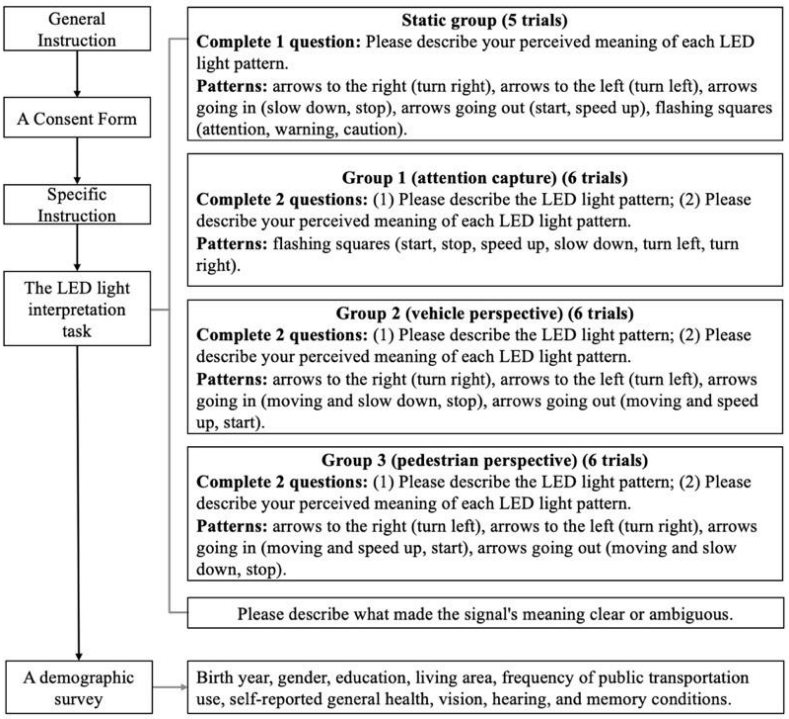

Figure 24 The experimental procedure and description for the experimental design of each group.

5.3. Results

Participants' level of communicating quality with LED signals in each driving video was measured by the rate of correct interpretation. The interpretations they answered were coded as 0 and 1 according to the standard meaning. Answers that were close to/consistent with the standard meaning were coded as 1, and those unrelated to the standard meaning were coded as 0. The mean and standard error of the code value were calculated and the code value was defined as interpretation correctness. Figure 25 summarizes participants' overall correctness in interpreting the vehicle's intention under various motion conditions.

4(group) × 5(LED light pattern) mixed design.	
Within-group variable	LED light pattern
Between-group variable	group

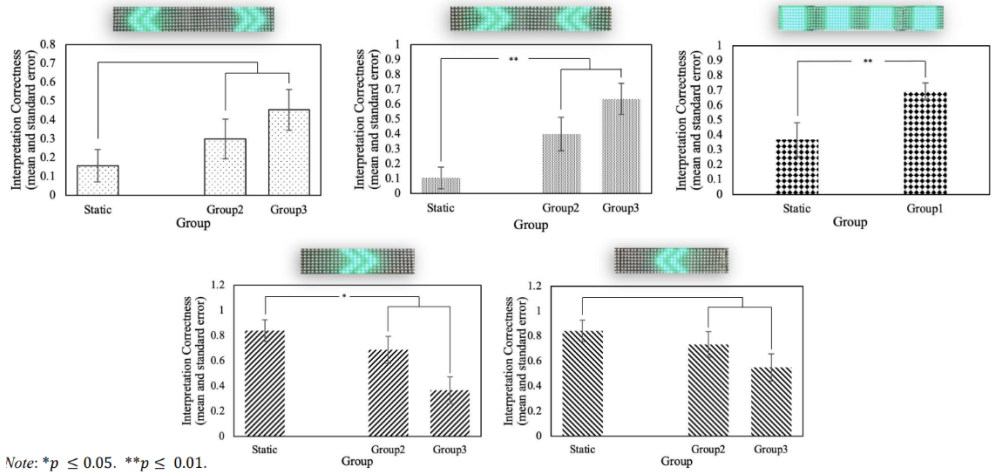

Figure 25 Interpretation correctness of 5 LED light patterns in static and with vehicle motion from either vehicle perspective (Group 2) or the pedestrian perspective (Group 3).

The flashing squares (top right pattern) stayed the same in both perspective conditions thus the contrast was only between static and Group 1 conditions.

5.3.1. Impact of motion condition on interpretation

Difference of interpretation correctness of static group, group 2 and group 3 in different LED light patterns was compared. The results were shown in Figure 26. It indicated that the main effect of group was not significant ($\chi^2(2) = .368$, $p = .832$), motion state did not affect participants' interpretation of the signal. However, to be specific, in "arrows going in" ($t(59) = 3.320$, $p = .002$) and "flashing squares" ($t(112) = 2.772$, $p = .007$) LED light pattern, participants were better at interpreting vehicle intention in the motion condition. In contrast, in "arrows to the right" LED light pattern, participants were better at interpreting vehicle intention in the motionless condition ($t(52) = -2.471$, $p = .017$). The main effect of pattern was significant ($\chi^2(3) = 18.928$, $p < .001$), participants interpreted better in a clear turning direction. The interaction effect between group and pattern was significant ($\chi^2(6) = 18.198$, $p = .006$).

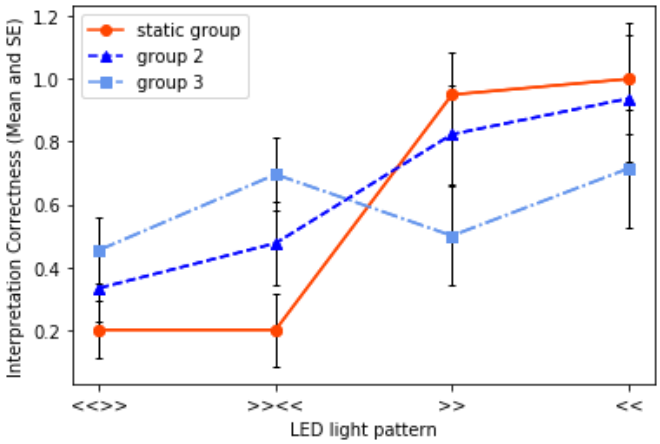

Figure 26 Interpretation correctness of 4 LED light patterns in motion and motionless conditions

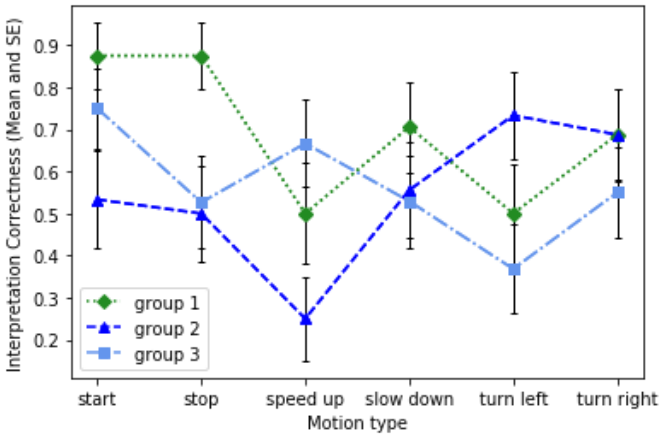

Figure 27 Interpretation correctness of 6 motion types in meaningful and attention capture conditions

5.3.2. Impact of meaningful condition on interpretation

As shown in Figure 27, interpretation correctness of group 1, group 2 and group 3 in different motion types was compared. The results showed that the main effect of group and pattern were not significant (group: $\chi^2(2) = 3.936$, $p = .140$; pattern: $\chi^2(5) = 9.411$, $p = .094$). The interaction effect between the group and the pattern was close to a marginally significant level ($\chi^2(10) = 17.207$, $p = .070$). In "stop" motion type, participants better interpreted in meaningless condition than the attention capturing condition($t(53) = 2.620$, $p = .011$), however, in other motion types, there was no significant difference in interpretation correctness between the two conditions.

5.3.3. Participants' special interpretations for specific conditions

We also summarized participants' special and interesting interpretations for specific conditions. For example, some participants interpreted the "arrows going in" in the static group as "The car is charging", although the standard meaning is "slow down and stop" (shown in Table 1). These are interesting observations showing that participants may be more open to new meanings of various signal patterns for an autonomous vehicle.

Table 2 Participants' special interpretations for specific conditions

LED light pattern	Standard meaning	Participants special and interesting interpretations
Flashing squares (Group 1)	speed up	This vehicle is not an ordinary car and to use caution when near it.
		It is to inform anyone walking in front of the machine.
	slow down	It was coming close to other vehicles.
	turn left	The vehicle is on duty and patrolling the area.
Arrows going in (Group 2)	moving and slow down	These look like emergency vehicle lights to me.
	stop	Feels like a "follow me" signal.
Arrows going out (Group 2)	moving and speed up	The vehicle is trying to part a crowd.
		Blue led lights are mostly used for medical services.
		Like this signal is telling me to back off or get out of the way.
Arrows going in (Group 3)	moving and speed up	It is going slow to find its destination.
Arrows going out (Group 3)	moving and slow down	To get out of the way of the moving vehicle.
Arrows to the right (Group 3)	turn left	The car is going to its destination. Searching for a passenger to pick up.
		It is most commonly associated with police vehicles.
Arrows going out (Static group)	start, speed up	The vehicle is out of order.
		People should go around the sides.
Arrows going in (Static group)	slow down, stop	The vehicle is not functioning properly.
		The car is charging.
Flashing squares (Static group)	attention, warning, caution	The car stopped moving.

5.4. Summary of Findings

This study aimed to explore the factors affecting LED communication from autonomous vehicles to pedestrians. Due to the pandemic, it was difficult to carry out this study offline. We experimented with

recorded videos of real vehicles. In this study, participants were asked to use statements (not choices) to finish the questions. After coding the answer of each participant, we calculated the interpretation correctness in order to do statistical analysis.

From this study, we concluded that when pedestrians were not familiar with the meaning of LED patterns, interpretation correctness of the vehicle's intention in the motion condition is better than the motionless condition; meaningful and attention capture conditions had no effect on interpretation correctness of vehicle intention, however, when the vehicle had no apparent turning action, participants were better at interpreting vehicle intention in the attention capture condition.

As shown in Figure 25 and Figure 26, in "arrows to the left" and "arrows to the right" LED light patterns, participants were better interpreting in group 2. It suggested that pedestrians, despite being observers of autonomous vehicles, were still used to using the vehicle perspective.

5.5. Limitations and Future Studies

Due to COVID-19, the study adopted an online data collection method, the experimental environment for participants was difficult to control. Participants might be disturbed by factors such as noise and computer networks when completing tasks. We also could not control the distance between the participant and the screen, video viewing time, audio information displayed from videos, etc. These factors would affect participants' interpretation of the LED light pattern.

Communication between pedestrians and autonomous vehicles needs to be studied in other detailed aspects in the future. Initially, how a pedestrian interprets information from an autonomous vehicle directly influences his behavior on the road, so hidden safety issues of autonomous vehicles should be noticed. Also, it is necessary to discuss how to use signals to improve the safety of pedestrians around autonomous vehicles. Furthermore, when crossing the road, pedestrians' visual behavior (e.g., the point of gaze) affects the communication between pedestrians and vehicles. The information indicator should be placed where it can be noticed most quickly to help pedestrians make appropriate and efficient decisions. In addition, estimating speed and distance of the vehicle might cause different interpretation. The accuracy of pedestrians' estimation of a vehicle speed is affected by the weather (Sun et al., 2015). If information such as vehicle speed, distance, arrival time, crossing safety were estimated by autonomous vehicles and presented on the screens, it may reduce the difficulty for pedestrians to understand the intention of the vehicle. Finally, pedestrians’ response to signals in different states (e.g., using mobile phones, communicating with companions, etc.) also affect their interpretation of the vehicle.

Recommendations and Conclusions

The research project identified several key findings and recommendations in cross-cutting categories related to Connected and Automated Vehicles as well as Multimodal Safety:

Pedestrian Detection

6. Pedestrian detection methods are used in CAVs as well as emerging infrastructure-based safety applications such as crosswalk clearance detection. These systems may include detection based on single source or fusion of multiple sources. This project tested multiple detection methods and developed improved methods with increased accuracy and reduced latency. The EcoPRT vehicle was able to incorporate the improved detection method, however the training image set included multiple camera perspectives and the method could likely be applied to infrastructure-based detection systems.

Pedestrian Occlusion

Occlusion is a common issue for detection in complex environments. This project developed a body part-based method which detects head, arms and legs of pedestrians in order to improve the overall detection of pedestrians when they are partially occluded. This issue is less severe in infrastructure-based detection systems with elevated cameras that avoid most obstructions, but is very important in CAV pedestrian detection. The project also developed a database of occluded pedestrian images which can be used for training or testing other new methods addressing this issue.

Communicating Intent

Traditional pedestrian-vehicle communication of intent relies on vehicle dynamics, signaling and non-verbal communication with drivers (typically eye contact). This project examined multiple methods for signaling the CAV intent to pedestrians using fixed or moving lightbars. Respondents struggled to correctly identify the message communicated by the lightbar in cases where multiple movements are expected (such as locations with potential turning movements) but identification improved in more constrained environments.

Future Study

The research team was not able to complete the planned in person experiment with communicating CAV intent due to delays to the project from COVID. The findings of the online survey can be used to better select lightbar patterns and test scenarios with varying complexity of vehicle movements to gauge pedestrian understanding in the field. In addition, the detection methods developed have promise for application in vehicle-based and infrastructure-based detection systems. Especially the work addressing occluded pedestrians is a key safety concern for CAV detection, and the dataset created for training and testing the method is a great resource for any future work in this domain.

References

Gopalakrishnan, Srinivas , “Application of Neural Networks for Pedestrian Path Prediction in a Low-cost Autonomous Vehicle,” Thesis/Dissertation, [Raleigh, North Carolina] : North Carolina State University, 2020.

7. Gupta, A. et al. “Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks”. CoRR abs/1803.10892 (2018). arXiv: 1803.10892. URL: http://arxiv.org/abs/1803.10892 accessed 10/13/2021

Redmon, J. URL: https://pjreddie.com/darknet/yolo/ accessed 10/13/2021

Liu,W. et al. SSD: Single ShotMultiBox Detector. 2016. URL: https://arxiv.org/abs/1512.02325 accessed 10/13/2021

Clamann, M., Aubert, M., & Cummings, M. (2017). *Evaluation of Vehicle-to-Pedestrian Communication Displays for Autonomous Vehicles*.

Dey, D., & Terken, J. (2017). *Pedestrian Interaction with Vehicles: Roles of Explicit and Implicit Communication* Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany. https://doi.org/10.1145/3122986.3123009

Farraher, B., Weinholzer, R., & Kowski, M. (1999). THE EFFECT OF ADVANCED WARNING FLASHERS ON RED LIGHT RUNNING--A STUDY USING MOTION IMAGING RECORDING SYSTEM TECHNOLOGY AT TRUNK HIGHWAY 169 AND PIONEER TRAIL IN BLOOMINGTON, MINNESOTA.

Habibovic, A., Lundgren, V. M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., Fagerlönn, J., Edgren, C., Fredriksson, R., Krupenia, S., Saluäär, D., & Larsson, P. (2018, 2018-August-07). Communicating Intent of Automated Vehicles to Pedestrians [Original Research]. *Frontiers in psychology*, *9*(1336). https://doi.org/10.3389/fpsyg.2018.01336

Klugman, A., Boje, B., & Belrose, M. (1992). *A STUDY OF THE USE AND OPERATION OF ADVANCE WARNING FLASHERS AT SIGNALIZED INTERSECTIONS. FINAL REPORT* (No. MN/RC-93/01).

Lagström, T., & Lundgren, V. (2016). AVIP - Autonomous vehicles' interaction with pedestrians - An investigation of pedestrian-driver communication and development of a vehicle external interface.

Matthews, M., Chowdhary, G., & Kieson, E. (2017). Intent Communication between Autonomous Vehicles and Pedestrians. *ArXiv, abs/1708.07123*.

Pant, P. D., & Xie, Y. (1995). Comparative study of advance warning signs at high speed signalized intersections. *Transportation research record*, *1495*, 28-35.

Rasouli, A., & Tsotsos, J. (2019, 03/15). Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice. *IEEE transactions on intelligent transportation systems*, *PP*, 1-19. https://doi.org/10.1109/TITS.2019.2901817

Sun, R., Zhuang, X., Wu, C., Zhao, G., & Zhang, K. (2015, 2015/04/01/). The estimation of vehicle speed and stopping distance by pedestrians crossing streets in a naturalistic traffic environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *30*, 97-106.

https://doi.org/10.1016/j.trf.2015.02.002

M. M. Islam, A. A. R. Newaz, B. Gokaraju, and A. Karimoddini, "Pedestrian detection for autonomous cars: Occlusion handling by classifying body parts," in 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2020, pp. 1433–1438.

A. Homaifar, A. Karimoddini, B. A. Erol, M. A. Khan, E. Tunstel, R. L. Roberts, R. F. Young, K. Snyder, R. S. Swanson, M. Jamshidi et al., "Operationalizing autonomy: A transition from the innovation space to real-world operations," IEEE Systems, Man, and Cybernetics Magazine, vol. 5, no. 4, pp. 23–32, 2019.

M. K. M. Rabby, M. M. Islam, and S. M. Imon, "A review of iot application in a smart traffic management system," in 2019 5th Inter- national Conference on Advances in Electrical Engineering (ICAEE). IEEE, 2019, pp. 280–285.

P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger, and B. Leibe, "Mots: Multi-object tracking and segmen- tation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7942–7951.

P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking without bells and whistles," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 941–951.

Z. Wang, L. Zheng, Y. Liu, and S. Wang, "Towards real-time multi- object tracking," arXiv preprint arXiv:1909.12605, vol. 2, no. 3, p. 4, 2019.

A. Milan, S. H. Rezatofighi, A. Dick, I. Reid, and K. Schindler, "Online multi-target tracking using recurrent neural networks," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1, 2017.

D. Frossard and R. Urtasun, "End-to-end learning of multi-sensor 3d tracking by detection," in 2018 IEEE international conference on robotics and automation (ICRA). IEEE, 2018, pp. 635–642.

A. Asvadi, P. Girao, P. Peixoto, and U. Nunes, "3d object tracking using rgb and LIDAR data," in 2016 IEEE 19th International Confer- ence on Intelligent Transportation Systems (ITSC). IEEE, 2016, pp. 1255–1260.

N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in 2017 IEEE international conference on image processing (ICIP). IEEE, 2017, pp. 3645–3649.

Y. Tian, P. Luo, X. Wang, and X. Tang, "Deep learning strong parts for pedestrian detection," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1904–1912.

X. Zhao, W. Li, Y. Zhang, T. A. Gulliver, S. Chang, and Z. Feng, "A faster RCNN-based pedestrian detection system," in IEEE Vehicular Technology Conference (VTC-Fall). IEEE, 2016, pp. 1–5.

W. Zhang, L. Tian, C. Li, and H. Li, "A SSD-based crowded pedestrian detection method," in International Conference on Control, Automation and Information Sciences. IEEE, 2018, pp. 222–226.

E. Zadobrischi and M. Negru, "Pedestrian detection based on tensor- flow yolov3 embedded in a portable system adaptable to vehicles," in 2020 International Conference on Development and Application Systems (DAS). IEEE, 2020, pp. 21–26.

S. Zhang, J. Yang, and B. Schiele, "Occluded pedestrian detection through guided attention in cnns,"

in Proceedings of the IEEE Con- ference on Computer Vision and Pattern Recognition, 2018, pp. 6995–7003.

C. Premebida, O. Ludwig, and U. Nunes, “Exploiting LIDAR-based features on pedestrian detection in urban scenarios,” in IEEE Interna- tional Conference on Intelligent Transportation Systems. IEEE, 2009, pp. 1–6.

L. Oliveira and U. Nunes, “Context-aware pedestrian detection using LIDAR,” in 2010 IEEE Intelligent Vehicles Symposium. IEEE, 2010, pp. 773–778.

T. Ogawa, H. Sakai, Y. Suzuki, K. Takagi, and K. Morikawa, “Pedes- trian detection and tracking using in-vehicle LiDAR for automotive application,” in IEEE Intelligent Vehicles Symposium (IV). IEEE, 2011, pp. 734–739.

K. Kidono, T. Miyasaka, A. Watanabe, T. Naito, and J. Miura, “Pedes- trian recognition using high-definition LIDAR,” in IEEE Intelligent Vehicles Symposium (IV). IEEE, 2011, pp. 405–410.

K. Li, X. Wang, Y. Xu, and J. Wang, “Density enhancement-based long-range pedestrian detection using 3-d range data,” IEEE Transac- tions on Intelligent Transportation Systems, vol. 17, no. 5, pp. 1368–1380, 2015.

K. Liu, W. Wang, and J. Wang, “Pedestrian detection with LiDAR point clouds based on single template matching,” Electronics, vol. 8, no. 7, p. 780, 2019.

T.-C. Lin, D. S. Tan, H.-L. Tang, S.-C. Chien, F.-C. Chang, Y.-Y. Chen, W.-H. Cheng, and K.-L. Hua, “Pedestrian detection from LiDAR data via cooperative deep and hand-crafted features,” in IEEE International Conference on Image Processing (ICIP). IEEE, 2018, pp. 1922–1926.

G. Chen, Z. Mao, H. Yi, X. Li, B. Bai, M. Liu, and H. Zhou, “Pedestrian detection based on panoramic depth map transformed from 3d-LiDAR data,” Periodica Polytechnica Electrical Engineering and Computer Science, vol. 64, no. 3, pp. 274–285, 2020.

C. Premebida, O. Ludwig, and U. Nunes, “LIDAR and vision-based pedestrian detection system,” Journal of Field Robotics, vol. 26, no. 9, pp. 696–711, 2009.

C. Premebida and U. Nunes, “Fusing LIDAR, camera and semantic information: A context-based approach for pedestrian detection,” The International Journal of Robotics Research, vol. 32, no. 3, pp. 371–384, 2013.

C. Premebida, J. Carreira, J. Batista, and U. Nunes, “Pedestrian detection combining RGB and dense LIDAR data,” in IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2014, pp. 4112–4117.

A. Gonza´lez, G. Villalonga, J. Xu, D. Va´zquez, J. Amores, and A. M. Lo´pez, “Multiview random forest of local experts combining RGB and LIDAR data for pedestrian detection,” in 2015 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2015, pp. 356–361.

J. Schlosser, C. K. Chow, and Z. Kira, “Fusing LiDAR and images for pedestrian detection using convolutional neural networks,” in IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2016, pp. 2198–2205.

D. Matti, H. K. Ekenel, and J.-P. Thiran, “Combining LiDAR space clustering and convolutional neural networks for pedestrian detection,” in 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2017, pp. 1–6.

T. Kim, M. Motro, P. Lavieri, S. S. Oza, J. Ghosh, and C. Bhat, "Pedes- trian detection with simplified depth prediction," in International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018, pp. 2712–2717.

R. Lahmyed and M. E. Ansari, "Multisensors-based pedestrian de- tection system," in IEEE/ACS International Conference of Computer Systems and Applications (AICCSA), 2016, pp. 1–4.

G. Ning, Z. Zhang, C. Huang, X. Ren, H. Wang, C. Cai, and Z. He, "Spatially supervised recurrent convolutional neural networks for visual object tracking," in 2017 IEEE International Symposium on Circuits and Systems (ISCAS), 2017, pp. 1–4.

Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P. H. Torr, "Fast online object tracking and segmentation: A unifying approach," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1328–1338.

K. Granstro¨m, S. Renter, M. Fatemi, and L. Svensson, "Pedestrian tracking using velodyne data—stochastic optimization for extended object tracking," in 2017 ieee intelligent vehicles symposium (iv). IEEE, 2017, pp. 39–46.

G. Jocher, A. Stoken, J. Borovec, NanoCode012, A. Chaurasia, TaoXie, L. Changyu, A. V, Laughing, tkianai, yxNONG, A. Hogan, lorenzomammana, AlexWang1900, J. Hajek, L. Diaconu, Marc, Y. Kwon, oleg, wanghaoyang0106, Y. Defretin, A. Lohia, ml5ah, B. Milanko, B. Fineran, D. Khromov, D. Yiwei, Doug, Durgesh, and F. Ingham, "ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations," Apr. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.4679653

A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Op- timal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.

P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine et al., "Scalability in perception for autonomous driving: Waymo open dataset," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2446–2454.

L. Wang, J. Shi, G. Song, and I.-F. Shen, "Object detection combining recognition and segmentation," in Asian conference on computer vision. Springer, 2007, pp. 189–199.