

International Journal of
Engineering Research and Science & Technology



ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

ROAD: ROAD EVENT AWARENESS DATASET FOR INTELLIGENT VEHICLE PERCEPTION

¹Boya Sreenivasa Naidu, MCA Student, Department of MCA

²Dr.C.Gulzar, Ph.D, Associate Professor, Department of MCA

¹²Dr KV SubbaReddy Institute of Technology, Dupadu, Kurnool

ABSTRACT

Humans drive holistically, which involves understanding dynamic road events and how they evolve over time, among other things. These characteristics may help self-driving vehicles make more human-like decisions and have higher situational awareness. We provide the first-ever ROAD event Awareness Dataset (ROAD) for Autonomous Driving to achieve this. Evaluation of an autonomous vehicle's ability to identify road events—triplets consisting of an active agent, the action or acts it does, and the scenes that correspond to those actions—is the goal of ROAD. ROAD consists of films originally extracted from the Oxford RobotCar Dataset, annotated with bounding boxes that show the position of each road event in the image plane. As a starting point for various detection tasks, we suggest 3D-RetinaNet, a unique incremental approach for online road event awareness.

In order to illustrate the challenges situation awareness in autonomous driving confronts, we also provide the results of the ICCV2021 ROAD challenge winners as well as the performance of the Slowfast and YOLOv5 detectors on the ROAD tasks. The purpose of ROAD is to allow academics to investigate intriguing problems such as advanced (road) activity detection, future event prediction, and continuous learning.

I. INTRODUCTION

Autonomous driving, often known as robot-assisted driving, has been a rapidly expanding field of study in recent years.

Many major corporations, like Google, Toyota, and Ford, developed their own robot-car idea as a result of the race towards completely autonomous cars [1], [2], and [3]. There are still significant concerns about safety, ethics, cost, and dependability, even though self-driving vehicles are generally regarded as a significant area of study and testing for the practical application of artificial intelligence [4]. Particularly from a safety perspective, intelligent vehicles must be able to reliably comprehend the actions of the people (drivers, pedestrians, or cyclists) with whom they are in close proximity in order to make appropriate judgements. For autonomous vehicles (AVs) to be deployed safely, situation awareness and the capacity to comprehend the behaviour of other road users are thus essential.

To offer information about what is occurring on the road, the most recent generation of robot-cars is outfitted with a variety of sensors, including laser rangefinders, radar, cameras, and GPS [5]. The vehicle's movement is then suggested by fusing the information that has been gathered in this way [6, 7, 8, 9]. However,

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>

other writers contend that, given humans' capacity for environmental navigation, vision is a sufficient sense for AVs to use. In this study, we examine the background of vision-based autonomous driving [10] using video sequences recorded by cameras installed on the car in a streaming, online form, without claiming to be proponents of the latter viewpoint.

Even while detector networks [11] are often trained to help with actor and object detection in traffic scenarios, this just enables the car to "see" its surroundings. The idea behind this work is that in order to predict intentions and make automated decisions, strong self-driving capabilities necessitate a more profound, human-like comprehension of dynamic road environments (as well as the changing behaviour of other road users over time) in the form of semantically meaningful concepts. One benefit of this strategy is that it enables the autonomous car to learn how to make judgements by concentrating on a far smaller quantity of relevant data, which is maybe more similar to how people make decisions.

In contrast, end-to-end reinforcement learning is at the other extreme. There, an autonomous automobile is trained to react to road circumstances more "human-like" by mimicking the behaviour of a human driver in an imitation learning environment [12]. However, this requires an incredible volume of data from many road conditions. Friedman et al. in [13] had to deploy an entire fleet of cars to capture 45 million frames for highway driving alone, which is a comparatively easy effort compared to city driving. Perhaps more significantly, this

method does not try to represent the essential facts occurring in the scene or the thinking of the agents inside it; instead, the network learns a mapping from the scene to control inputs. Many writers [15], [16] have lately emphasised the shortcomings of models that directly map observations to actions [17], particularly in the context of self-driving automobiles, as was covered in [14].

ROAD: An idea for a multi-label, multi-task dataset. By moving away from the disjointed collection of object identification, semantic segmentation, or pedestrian intention tasks that are the focus of most of the present work, this study attempts to provide a new framework for situation awareness and perception. We suggest doing so using a "holistic," multi-label method where the core idea of a road event (RE) is composed of agents, actions, and their locations. As viewed from the perspective of an autonomous vehicle, road events are defined as triplets $E = \langle Ag; Ac; Loc \rangle$, which are made up of an active road agent Ag , the action(s) Ac it performs (possibly more than one at a time), and the location(s) Loc in which this takes place (which may vary from the beginning to the end of the event itself).

This raises the issue to a higher conceptual level, where AVs are evaluated on their comprehension of a dynamic scene rather than their ability to describe its appearance, enabling them to utilise that knowledge to inform choices and plot points. We may also describe the causative ties between occurrences by modelling dynamic road scenes in terms of road events.

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>

We can then use these causality links to forecast further future effects.

This paper presents ROAD, the first Road event Awareness in Autonomous Driving Dataset, a completely new kind of dataset created to enable researchers working on autonomous vehicles to test the situation awareness capabilities of their stacks in a way that hasn't been feasible up to this point. In contrast to all other standards now in use, ROAD offers ground truth for actions taken by all road agents—not just humans. In this regard, ROAD stands out due to the depth and complexity of its annotation, which is intended to facilitate the suggested conceptual change. We are sure that our work will be highly helpful in the future for the computer vision and autonomous driving communities.

characteristics. In order to build ROAD, 22 carefully chosen, rather long-duration films were meticulously annotated using the Oxford Robot Car Dataset [18]. "Tubes," or time series of frame-wise bounding box detections, are used to depict road events. In terms of the depth and richness of its annotation, rather than the quantity of video frames, ROAD is a large dataset. A total of 122K video frames are labelled for 560K detection bounding boxes, which are then linked to 1:7M distinct individual labels. These labels are divided into 499K location labels, 640K action labels, and 560K agent labels.

The following guidelines guided the construction of the dataset.

_ A multi-label benchmark: the label of the (moving) agent in charge, the label or labels of the action type being carried out, and the labels indicating the action's location make

up each road event. When appropriate, each event may be given more than one instance of the same label type (for example, a RE may represent both turning left and going away).

Labelling is done from the AV's perspective, with the ultimate objective being for the autonomous car to utilise this data to make the right choices.

_ The goal of the meta-data is to include all the information needed to completely describe a road situation; Fig. 1 provides an example of this idea. The collection of labels linked to the current video frame should be enough to allow one to mentally reproduce the road scenario after shutting their eyes (or, equivalently, enough for the AV to be able to make a judgement).

In an attempt to bring action detection into the real world, ROAD shifts the paradigm from human body actions to agent-caused events, focussing almost exclusively on human body actions and taking into account human actions as drivers of different kinds of vehicles, aside from pedestrian behaviour. The intricacy of road events occurring in actual, non-choreographed driving settings, the quantity of active agents present, and the range of weather variables covered make ROAD more difficult than existing action detection benchmarks, as shown by our research.

Activities. Several tasks related to situation awareness for self-driving can be validated with ROAD. These tasks are associated with a label type (agent, action, or location) or a combination of these label types: spatiotemporal (i) agent detection, (ii) action detection, (iii) location detection, (iv) agent-action detection, (v) road event

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>

detection, and (vi) temporal segmentation of AV actions. Both frame-level detection, which independently outputs the bounding box(es) (BBs) of the instances present for each video frame along with the pertinent class labels, and video-level detection, which entails regressing the entire sequence of temporally-linked bounding boxes (or, as it is now called, a "tube") associated with an instance along with the pertinent class label, can be evaluated for each task. We test both of these in this study. As explained in Section 4, each job includes the required annotation as well as a common baseline.

II. LITERATURE SURVEY

In Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, J. Winn and J. Shotton discuss "The layout consistent random field for recognising and segmenting partially occluded objects."

We provide a technique for identifying and separating several partly obscured items in pictures. The object class has a specified component hierarchy. Shape-oriented local image characteristics are enhanced in order to train whole-object segmentors and portion detectors. The input picture is subjected to the part detectors throughout the detection process. Every edge pixel in the picture that enhances part detection results is taken off. The object edges and portion detection responses are used to construct a combined probability of many objects. The inter-object occlusion logic used to calculate the joint probability is based on the object silhouettes that were recovered using the whole-object segmentor. Part detection responses are combined, sorted, and allocated to several object hypotheses by

optimising the joint likelihood. Two public test sets are used to assess the suggested method in the pedestrian class. The experimental findings demonstrate that our approach works better than the earlier ones.

K. Korosec, "Toyota is relying on this firm to further its aspirations for self-driving cars, In order to address action detection issues with autonomous driving, we provide the new Road Event and Activity Detection (READ) dataset, which was developed from the viewpoint of autonomous vehicles. With the ultimate goal of assisting autonomous decision making, READ will provide researchers in computer vision, smart cars, and machine learning in general the chance to investigate fascinating new issues like comprehending complex (road) activities, identifying the behaviour of sentient agents, and forecasting the label and the location of future actions and events.

"Ford campus vision and lidar data set," by G. Pandey, J. R. McBride, and R. M. Eustice
The data set gathered by an autonomous ground vehicle testbed based on a modified Ford F-250 pickup truck is described in this research. The car is equipped with a Velodyne 3D-lidar scanner, two push-broom forward-looking Riegl lidars, a Point Grey Ladybug3 omnidirectional camera system, and a professional (Applanix POS-LV) and consumer (Xsens MTi-G) inertial measurement unit (IMU). Here, we display the time-registered data from various onboard sensors that was gathered during November and December 2009 while the car was being driven throughout the Ford Research campus and downtown Dearborn, Michigan. Several big and small-scale loop closures may be found in the vehicle travel

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>

trajectory of these data sets, which should be helpful for evaluating several state-of-the-art computer vision and simultaneous localisation and mapping (SLAM) techniques.

Autonomous Driving: Technical, Legal, and Social Aspects, M. E. A. Maurer

This book examines fully automated, self-driving cars and addresses a number of unanswered questions: With a variety of users and human drivers, how might autonomous cars be included into the present transportation system? Under what legal frameworks do automated cars now fall? What are the dangers of automation, and how will society react to them? What adjustments would businesses need to make, and how will the market respond to autonomous vehicles? Key sociological, technical, and transportation challenges pertaining to vehicle automation are defined by experts from the US and Germany. They talk about the choices that programmers of autonomous cars have to make in order for them to sense their surroundings, communicate with other drivers, and make judgements that might have moral repercussions. The writers also point out the worries and expectations that will serve as the foundation for both social and personal acceptance of autonomous driving. Although such cars provide enormous safety advantages, the authors show that these benefits can only be realised if a suitable safety concept is included into the design of the vehicle. Similar consideration must be given to the design of networks and vehicles in order to realise the promise of autonomous vehicles to reorganise traffic and revolutionise the movement of people

and commodities. The book attempts to provide a contemporary, thorough, and scientifically sound account of the developing area of "autonomous driving" by addressing each of these subjects.

III. SYSTEM ANALYSIS AND DESIGN

EXISTING SYSTEM

one-dimensional datasets. It takes less time and money to gather and annotate RGB data alone than it does to create multimodal datasets that include range data from LiDAR or radar. For RGB photos, the majority of single-modality datasets [23], [24], [25], [26], [27], and [28] include 2D bounding box and scene segmentation labels. Cityscapes [24], Mapillary Vistas [25], BDD100k [26], and Apolloscape [27] are a few examples. In order to investigate how vision algorithms generalise to other unseen data, RGB photos are collected in various lighting and weather situations [25], [26], and [28].

Only pedestrian detection annotation is available in other datasets [29], [30], [31], [32], [33], [34], and [35]. DriveSeg, recently published by Toyota and MIT, includes pixel-level semantic tagging for 12 agent types [36]. datasets with several modes. The first multimodal dataset was KITTI [37]. Along with GPS/IMU (inertial) data, it offers depth labels from dense point clouds from LiDAR and front-facing stereo pictures. Additionally, it offers bounding-box annotations to help with 3D object detection. Multimodal datasets also include KAIST [39] and H3D [38]. In congested settings, H3D offers 3D box annotations

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>

based on actual LiDAR-generated 3D coordinates.

KAIST offers thermal camera data in addition to RGB, stereo, GPS/IMU, and LiDAR-based range data, whereas H3D, in contrast to KITTI, includes object recognition annotations in a complete 360° view. [18], [40], and other prominent multimodal datasets simply include raw data without semantic labels, whereas [41] and [42] provide labels for driving conduct and location category, respectively. In terms of both data (also recorded in different weather circumstances, such as at night or in the rain) and annotations (RGB, LiDAR/radar, 3D boxes), the most current multimodal large-scale AV datasets [43], [44], [45], [46], [47], and [48] are much bigger. Argoverse [43], for example, provides 3D bounding boxes with tracking information for 15 items of interest and doubles the number of sensors compared to KITTI [37] and nuScenes [49]. In a similar vein, Lyft [44] offers 3D boundary boxes for automobiles along with location annotation for speed bumps, lane segments, pedestrian crosswalks, stop signs, parking zones, and humps. Similar to KITTI's [37], KITTI-360 [48] adds a pushbroom laser scanner and two fisheye cameras to provide a complete 360-degree field of vision.

Semantic and instance annotations for 2D pictures, including 19 items, and 3D point clouds are included in KITTI-360. For the purpose of localisation, IMU/GPS sensors are incorporated. Waymo offers 2D annotation on camera data for four item types in addition to 3D bounding boxes based on LiDAR data [45]. Similar 3D annotation for five object types is used in

[46], where the authors include additional nighttime situations with a faster driving automobile to create a more difficult dataset. The nuScenes [49], Lyft L5 [44], Waymo Open [45], and A*3D [46] datasets are the most prevalent among large-scale multimodal datasets in terms of the quantity of instances, the utilisation of high-quality sensors with various data types (such as point clouds or 360° RGB videos), and the richness of the annotation that provides both semantic information and 3D bounding boxes. Furthermore, human-annotated rich semantic maps—a crucial prerequisite for scene understanding—are used by nuScenes [49], Argoverse [43], Lyft L5 [44], and KITTI-360 [48] to give contextual information.

Predicting a trajectory. Another area of research, which is based on a number of significant RGB-based datasets, examines the issue of pedestrian trajectory prediction in the context of autonomous driving. RGB data was recorded using either drone-mounted surveillance cameras [53] for an aerial perspective or stationary cameras [50], [51], and [52] to construct these datasets. In order to estimate future trajectories, [54], [55] use RGB photographs that capture an egocentric perspective from a moving automobile. The multimodal 3D point cloud-based datasets [37], [38], [43], [44], [45], and [49] have recently been used for trajectory prediction as well. These datasets were first presented for the benchmarking of 3D object identification and tracking. Numerous fascinating new studies [56], [57], [58], and [59] do provide datasets to investigate the intents and behaviours of agents utilising cameras installed on

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>

automobiles. They only cover a small number of action labels, such as walking, standing, gazing, and crossing, which is completely inadequate for a comprehensive analysis of road agent activity. Perhaps the most promising of them is TITAN [59].

DISADVANTAGES

- **Data complexity:** The majority of machine learning models now in use need to be able to correctly read large and intricate datasets in order to predict events.
- **Data availability:** In order to provide precise predictions, the majority of machine learning models need a lot of data. The accuracy of the model may degrade if data is not accessible in large enough amounts.
- **Inaccurate labelling:** The accuracy of the machine learning models that are now in use depends on how well the input dataset was used for training. Inaccurate labelling of the data prevents the model from producing reliable predictions.

PROPOSED SYSTEM

A formal description of a road event, as a triplet consisting of a road agent, the action or actions it takes, and the location or locations of the event as seen from the AV's point of view, was the focal point of a conceptual change in situation awareness.

The first-of-its-kind ROad event awareness dataset for autonomous driving (ROAD) was created to facilitate this paradigm shift and enable testing of various situation awareness tasks for autonomous driving, such as ego-action classification, event detection, and agent and/or action detection. By moving away from the disjointed collection of object identification, semantic segmentation, or pedestrian intention tasks

that are the focus of most of the present work, this study attempts to provide a new framework for situation awareness and perception. We suggest doing so using a "holistic," multi-label method where the core idea of a road event (RE) is composed of agents, actions, and their locations.

This raises the issue to a higher conceptual level, where AVs are evaluated on their comprehension of a dynamic scene rather than their ability to describe its appearance, enabling them to utilise that knowledge to inform choices and plot points. We may also describe the causative ties between occurrences by modelling dynamic road scenes in terms of road events. We can then use these causality links to forecast further future effects.

This paper presents ROAD, the first ROad event Awareness in Autonomous Driving Dataset, as a brand-new dataset type intended to enable researchers studying autonomous vehicles to test their stacks' situation awareness capabilities in a way that has not been feasible up to this point. In contrast to all other standards now in use, ROAD offers ground truth for actions taken by all road agents—not just humans. In this regard, ROAD stands out due to the depth and complexity of its annotation, which is intended to facilitate the suggested conceptual change. We have no doubt that this work will be very helpful in the future for the computer vision and autonomous driving communities.

ADVANTAGES

_ A multi-label benchmark: each road event is made up of labels identifying the sort of action being done, the label of the (moving) agent in charge, and labels indicating the

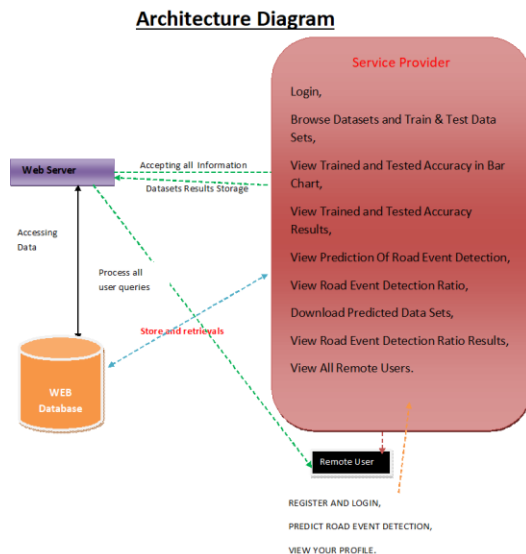
<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>

action's location. When appropriate, each event may be given more than one instance of the same label type (for example, a RE may represent both turning left and going away).

_ The labelling process is carried out from the AV's perspective, with the ultimate objective being for the autonomous vehicle to use this data to make the right choices.

_ All of the information needed to completely describe a road situation should be included in the meta-data; this system provides an example of this idea. The collection of labels linked to the current video frame should be enough to allow one to mentally reproduce the road scenario after shutting their eyes (or, equivalently, enough for the AV to be able to make a judgement).

IV. SYSTEM ARCHITECTURE



V. IMPLEMENTATION

Modules Description

Service Provider

The Service Provider must use a working user name and password to log in to this module. He can do many tasks after successfully logging in, including browsing datasets and training and testing datasets. View Road Event Detection Prediction, View Road Event Detection Ratio, Download Predicted Data Sets, View Road Event Detection Ratio Results, View All Remote Users, and View Trained and Tested Accuracy in Bar Chart.

View and Authorize Users

The administrator may see a list of all registered users in this module. Here, the administrator may see the user's information, like name, email, and address, and they can also grant the user permissions.

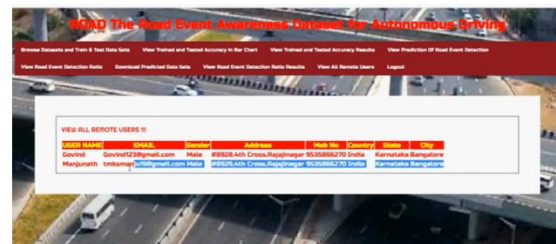
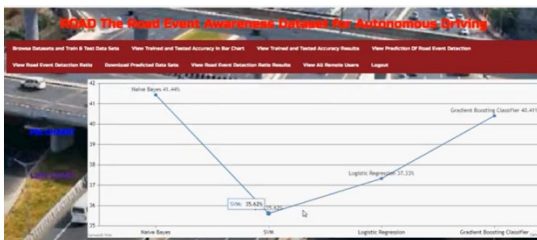
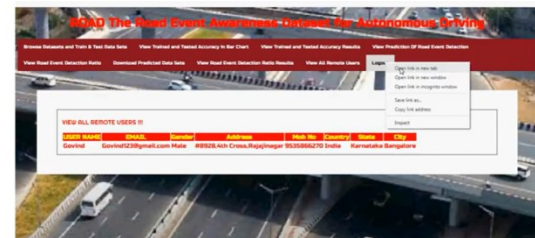
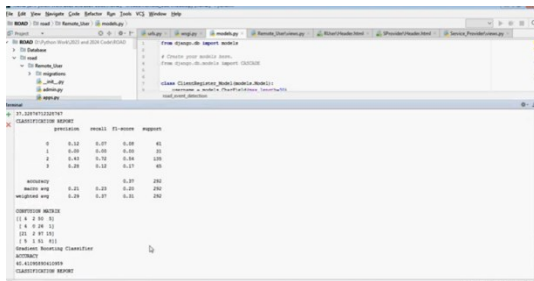
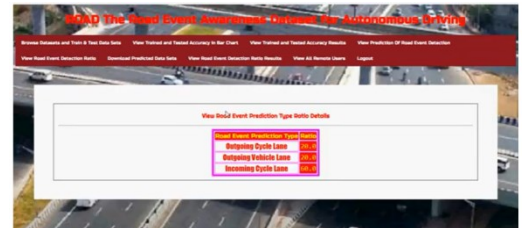
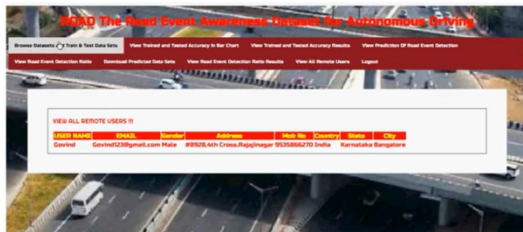
Remote User

A total of n users are present in this module. Before beginning any actions, the user needs register. Following registration, the user's information will be entered into the database. Following a successful registration, he must use his password and authorised user name to log in. Following a successful login, the user may do tasks including registering and logging in, predicting road event detection, and seeing their profile.

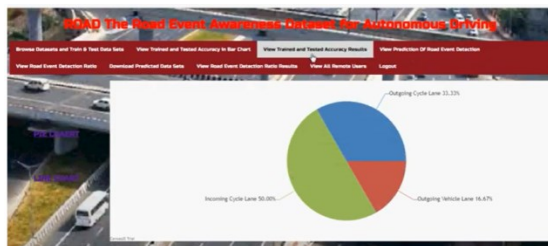
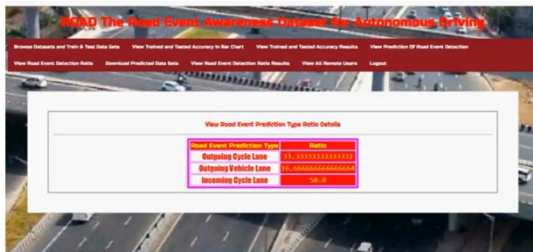
VI. SCREEN SHOTS



<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>



<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>



VII. CONCLUSION

This research proposed a situation awareness strategy for autonomous driving based on the concept of road events and introduced a new Road event Awareness Dataset for Autonomous Driving (ROAD) as a standard for this area of study. Because the dataset was created using videos that were collected as part of the Oxford Robot Car dataset, it contains unique features in the field [18]. Its comprehensive annotation, which follows a multi-label philosophy, labels road actors (including the AV), their

locations, and the action or acts they do. Labels of the three types may be assembled to create road events. The collection consists of 122K annotated video frames from 22 films with 560K detection bounding boxes associated with 1.7M unique labels.

A new 3DRetinaNet architecture, a Slow fast backbone, and a YOLOv5 model (for agent identification) were used in the ROAD baseline testing. Both frame-MAP and video-MAP were evaluated. Our first findings show the difficulties of ROAD, with the Slow fast baseline achieving a video-MAP on the three main tasks, which include between 20% and 30%, at poor localisation accuracy (20% overlap). However, YOLOv5 was able to achieve a notably superior performance. The results of the ROAD @ ICCV 2021 challenge validated these findings, highlighting the need for a more comprehensive study and the main challenges specific to situation awareness in road environments.

In the near future, we pledge to enhance our dataset by including ROAD-like annotation into important datasets such as Waymo and PIE. Our dataset may be expanded to include a range of challenging situation awareness tasks in autonomous driving, such as machine theory of mind, event prediction, trajectory prediction, and ongoing learning.

REFERENCES

- [1] J. Winn and J. Shotton, "The layout consistent random field for recognizing and segmenting partially occluded objects," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2006, pp. 37–44.
- [2] K. Korosec, "Toyota is betting on this startup to drive its selfdriving car plans

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>

forward,” [Online]. Available: <http://fortune.com/2017/09/27/toyota-self-driving-car-luminar/>

1050 IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 45, NO. 1, JANUARY 2023

[3] G. Pandey, J. R. McBride, and R. M. Eustice, “Ford campus vision and lidar data set,” *Int. J. Robot. Res.*, vol. 30, no. 13, pp. 1543–1552, 2011.

[4] M. E. A. Maurer, *Autonomous Driving: Technical, Legal and Social Aspects*. Berlin, Germany: Springer, 2016.

[5] A. Broggi et al., “Intelligent vehicles,” in *Springer Handbook of Robotics*. Berlin, Germany: Springer, 2016, pp. 1627–1656.

[6] S. Azam, F. Munir, A. Rafique, Y. Ko, A. M. Sheri, and M. Jeon, “Object modeling from 3D point cloud data for self-driving vehicles,” in *Proc. IEEE Intell. Veh. Symp.*, 2018, pp. 409–414.

[7] Z. Fang and A. M. Lopez, “Is the pedestrian going to cross? Answering by 2D pose estimation,” in *Proc. IEEE Intell. Veh. Symp.*, 2018, pp. 1271–1276.

[8] P. Wang, C. Chan, and A. D. L. Fortelle, “A reinforcement learning based approach for automated lane change maneuvers,” in *Proc. IEEE Intell. Veh. Symp.*, 2018, pp. 1379–1384.

[9] J. Chen, C. Tang, L. Xin, S. E. Li, and M. Tomizuka, “Continuous decision making for on-road autonomous driving under uncertain and interactive environments,” in *Proc. IEEE Intell. Veh. Symp.*, 2018, pp. 1651–1658.

[10] M. Bertozzi, A. Broggi, and A. Fascioli, “Vision-based intelligent vehicles:

State of the art and perspectives,” *Robot. Auton. Syst.*, vol. 32, no. 1, pp. 1–16, 2000.

[11] J. Redmon and A. Farhadi, “Yolo9000: Better, faster, stronger,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6517–6525.

[12] M. Codevilla, Felipe Dosovitskiy, “End-to-end driving via conditional imitation learning,” in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 1–9. [13] L. Fridman et al., “Arguing machines: Perception-control system redundancy and edge case discovery in real-world autonomous driving,” 2017, arXiv:1710.04459.

[14] F. Cuzzolin, A. Morelli, B. Cirstea, and B. J. Sahakian, “Knowing me, knowing you: Theory of mind in AI,” *Psychol. Med.*, vol. 50, no. 7, pp. 1057–1061, May 2020.

[15] A. Rasouli and J. K. Tsotsos, “Autonomous vehicles that interact with pedestrians: A survey of theory and practice,” *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 900–918, Mar. 2020.

[16] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, “Human motion trajectory prediction: A survey,” 2019, arXiv:1905.06113.

[17] S. Armstrong and S. Mindermann, “Occam’s razor is insufficient to infer the preferences of irrational agents,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 5603–5614.

[18] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 year, 1000 km: The oxford robotcar dataset,” *Int. J. Robot. Res.*, vol. 36, no. 1, pp. 3–15, 2017.

[19] G. Singh, S. Saha, M. Sapienza, P. Torr, and F. Cuzzolin, “Online real-time multiple spatiotemporal action localisation



<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1232-1243>

and prediction,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 3637–3646.

[20] S. Saha, G. Singh, M. Sapienza, P. H. Torr, and F. Cuzzolin, “Deep learning for detecting multiple space-time action tubes in videos,” 2016, arXiv:1608.01529.