Pegem Journal of Education and Instruction, Vol. 13, No. 3 , 2023 (pp. 512-523)

RESEARCH ARTICLE WWW.PEGEGOG.NET

ROAD: THE ROAD EVENT AWARENESS DATASET FOR AUTONOMOUS DRIVING

[1]Dr. M. Ramasubramanian, [2]Anisha Kapoor, [3]Sara Amreen, [4]K. S. Swathi

[1]Professor, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India

Email: ramanmass01@gmail.com

[2,3,4,]B.Tech Student, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India

Abstract:

Humans drive in a holistic fashion which entails, in particular, understanding dynamic road events and their evolution. Injecting these capabilities in autonomous vehicles can thus take situational awareness and decision making closer to human-level performance. To this purpose, we introduce the ROad event Awareness Dataset (ROAD) for Autonomous Driving, to our knowledge the first of its kind. ROAD is designed to test an autonomous vehicle's ability to detect road events, defined as triplets composed by an active agent, the action(s) it performs and the corresponding scene locations. ROAD comprises videos originally from the Oxford RobotCar Dataset, annotated with bounding boxes showing the location in the image plane of each road event. We benchmark various detection tasks, proposing as a baseline a new incremental algorithm for online road event awareness termed 3D-RetinaNet. We also report the performance on the ROAD tasks of Slowfast and YOLOv5 detectors, as well as that of the winners of the ICCV2021 ROAD challenge, which highlight the challenges faced by situation awareness in autonomous driving. ROAD is designed to allow scholars to investigate exciting tasks such as complex (road) activity detection, future event anticipation and continual learning.

Keywords: Awareness Dataset (ROAD) for Autonomous Driving, YOLOv5 detectors.

INTRODUCTION:IN recent years, autonomous driving (or robot-assisted driving) has emerged as a fast-growing research area. The race towards fully autonomous vehicles pushed many large

Corresponding Author e-mail: : ramanmass01@gmail.com

How to cite this article: 1Dr. M. Ramasubramanian, 2Anisha Kapoor, 3Sara Amreen, 4K. S. Swathi. ROAD: THE ROAD EVENT AWARENESS DATASET FOR AUTONOMOUS DRIVING.Pegem Journal of Education and Instruction, Vol. 13, No. 3, 2023, 512-523

Source of support: Nil **Conflicts of Interest:** None.

DOI: 10.47750/pegegog.13.03.50

Received: 12.09.2023

Accepted: 22.10.2023 **Published:** 24.11.2023

companies, such as Google, Toyota and Ford, to develop their own concept of robot-car [1], [2], [3]. While self-driving cars are widely considered to be a major development and testing ground for the real-world application of artificial intelligence, major reasons for concern remain in terms of safety, ethics, cost, and reliability [4]. From a safety standpoint, in particular, smart cars need to robustly interpret the behavior of the humans (drivers, pedestrians or cyclists) they share the environment with, in order to cope with their decisions. Situation awareness and the ability to understand the behavior of other road users are thus crucial for the safe deployment of autonomous vehicles (AVs). The latest generation of robot–cars is equipped with a range of different sensors (i.e., laser rangefinders, radar, cameras, GPS) to provide data on what is happening on the road [5]. The information so extracted is then fused to suggest how the vehicle should move [6], [7], [8], [9]. Some authors, however, maintain that vision is a sufficient sense for AVs to navigate their environment, supported by humans' ability to do just so. Without enlisting ourselves as supporters of the latter point of view, in this paper we consider the context of vision-based autonomous driving [10] from video sequences captured by cameras mounted on the vehicle in a streaming, online fashion. While detector networks [11] are routinely trained to facilitate object and actor recognition in road scenes, this simply allows the vehicle to 'see' what is around it. The philosophy of this work is that robust self-driving capabilities require a deeper, more human-like understanding of dynamic road environments (and of the evolving behavior of other road users over time) in the form of semantically meaningful concepts, as a stepping stone for intention prediction and automated decision making. One advantage of this approach is that it allows the autonomous vehicle to focus on a much smaller amount of relevant information when learning how to make its decisions, in a way arguably closer to how decision making takes place in humans. On the opposite side of the spectrum lies end-to-end reinforcement learning. There, the behavior of a human driver in response to road situations is used to train, in an imitation learning setting [12], an autonomous car to respond in a more 'human-like' manner to road scenarios. This, however, requires an astonishing amount of data from a myriad of road situations. For highway driving only, a relatively simple task when compared to city driving, Friedman et al. in [13] had to use a whole fleet of vehicles to collect 45 million frames. Perhaps more importantly, in this approach the network learns a mapping from the scene to control inputs, without attempting to model the significant

facts taking place in the scene or the reasoning of the agents therein. As discussed in [14], many authors [15], [16] have recently highlighted the insufficiency of models which directly map observations to actions [17], specifically in the self-driving cars scenario.

LITERATURE REVIEW

ROAD: A Multi-Label, Multi-Task Dataset Concept. This work aims to propose a new framework for situation awareness and perception, departing from the disorganized collection of object detection, semantic segmentation or pedestrian intention tasks which is the focus of much current work. We propose to do so in a "holistic", multi-label approach in which agents, actions and their locations are all ingredients in the fundamental concept of road event (RE). Road events are defined as triplets E ¼ ðAg; Ac; LocÞ composed by an active road agent Ag, the action(s) Ac it performs (possibly more than one at the same time), and the location(s) Loc in which this takes place (which may vary from the start to the end of the event itself), as seen from the point of the view of an autonomous vehicle. This takes the problem to a higher conceptual level, in which AVs are tested on their understanding of what is going on in a dynamic scene rather than their ability to describe what the scene looks like, putting them in a position to use that information to make decisions and a plot course of action. Modeling dynamic road scenes in terms of road events can also allow us to model the causal relationships between what happens; these causality links can then be exploited to predict further future consequences. To transfer this conceptual paradigm into practice, this paper introduces ROAD, the first Road event Awareness in Autonomous Driving Dataset, as an entirely new type of dataset designed to allow researchers in autonomous vehicles to test the situation awareness capabilities of their stacks in a manner impossible until now. Unlike all existing benchmarks, ROAD provides ground truth for the action performed by all road agents, not just humans. In this sense ROAD is unique in the richness and sophistication of its annotation, designed to support the proposed conceptual shift. We are confident this contribution will be very useful moving forward for both the autonomous driving and the computer vision community. Features. ROAD is built upon (a fraction of) the Oxford Robot Car Dataset [18], by carefully annotating 22 carefully selected, relatively long-duration videos. Road events are represented as 'tubes', i.e., time series of frame-wise bounding box detections. ROAD is a dataset of significant size, most notably in terms of the richness and complexity of its annotation rather than the raw number of video frames. A total of 122K video frames

are labeled for a total of 560K detection bounding boxes in turn associated with 1:7M unique individual labels, broken down into 560K agent labels, 640K action labels and 499K location labels. In an effort to take action detection into the real world, ROAD moves away from human body actions almost entirely, to consider (besides pedestrian behavior) actions performed by humans as drivers of various types of vehicles, shifting the paradigm from actions performed by human bodies to events caused by agents. As shown in our experiments, ROAD is more challenging than current action detection benchmarks due to the complexity of road events happening in real, non-choreographed driving conditions, the number of active agents present and the variety of weather conditions encompassed. Tasks. ROAD allows one to validate manifold tasks associated with situation awareness for self-driving, each associated with a label type (agent, action, location) or combination thereof: spatiotemporal (i) agent detection, (ii) action detection, (iii) location detection, (iv) agent-action detection, (v) road event detection, as well as the (vi) temporal segmentation of AV actions. For each task one can assess both frame-level detection, which outputs independently for each video frame the bounding box(es) (BBs) of the instances there present and the relevant class labels, and video-level detection, which consists in regressing the whole series of temporally-linked bounding boxes (i.e., in current terminology, a 'tube') associated with an instance, together with the relevant class label. In this paper we conduct tests on both.

S. Armstrong and S. Mindermann. Occam's razor is insufficient to infer the preferences of irrational agents. In Advances in Neural Information Processing Systems, volume 31, pages 5603—-5614, 2018. Inverse reinforcement learning (IRL) attempts to infer human rewards or preferences from observed behavior. Since human planning systematically deviates from rationality, several approaches have been tried to account for specific human shortcomings. However, the general problem of inferring the reward function of an agent of unknown rationality has received little attention. Unlike the well-known ambiguity problems in IRL, this one is practically relevant but cannot be resolved by observing the agent's policy in enough environments. This paper shows (1) that a No Free Lunch result implies it is impossible to uniquely decompose a policy into a planning algorithm and reward function, and (2) that even with a reasonable simplicity prior/Occam's razor on the set of decompositions, we cannot distinguish between the true decomposition and others that lead to high regret. To address this, we

need simple 'normative' assumptions, which cannot be deduced exclusively from observations. In today's reinforcement learning systems, a simple reward function is often hand-crafted, and still sometimes leads to undesired behaviors on the part of RL agent, as the reward function is not well aligned with the operator's true goals4 . As AI systems become more powerful and autonomous, these failures will become more frequent and grave as RL agents exceed human performance, operate at time-scales that forbid constant oversight, and are given increasingly complex tasks — from driving cars to planning cities to eventually evaluating policies or helping run companies. Ensuring that the agents behave in alignment with human values is known, appropriately, as the value alignment problem [Amodei et al., 2016, Hadfield-Menell et al., 2016, Russell et al., 2015, Bostrom, 2014, Leike et al., 2017]. One way of resolving this problem is to infer the correct reward function by observing human behaviour. This is known as Inverse reinforcement learning (IRL) [Ng and Russell, 2000, Abbeel and Ng, 2004, Ziebart et al., 2008]. Often, learning a reward function is preferred over imitating a policy: when the agent must outperform humans, transfer to new environments, or be interpretable. The reward function is also usually a (much) more succinct and robust task representation than the policy, especially in planning tasks [Abbeel and Ng, 2004]. Moreover, supervised learning of long-range and goal-directed behavior is often difficult without the reward function [Ratliff et al., 2006]. Usually, the reward function is inferred based on the assumption that human behavior is optimal or noisily optimal. However, it is well-known that humans deviate from rationality in systematic, non-random ways [Tversky and Kahneman, 1975]. This can be due to specific biases such as timeinconsistency, loss aversion and hundreds of others, but also limited cognitive capacity, which leads to forgetfulness, limited planning and false beliefs. This limits the use of IRL methods for tasks that humans don't find trivial. Some IRL approaches address specific biases [Evans et al., 2015b,a], and others assume noisy rationality [Ziebart et al., 2008, Boularias et al., 2011]. But a general framework for inferring the reward function from suboptimal behavior does not exist to our knowledge. Such a framework needs to infer two unobserved variables simultaneously: the human reward function and their planning algorithm5 which connects the reward function with behaviour, henceforth called a planner. The task of observing human behaviour (or the human policy) and inferring from it the human reward function and planner will be termed decomposing the human policy.

This paper will show there is a No Free Lunch theorem in this area: it is impossible to get a unique decomposition of human policy and hence get a unique human reward function. Indeed, any reward function is possible. And hence, if an IRL agent acts on what it believes is the human policy, the potential regret is near-maximal. This is another form of unidentifiability of the reward function, beyond the well-known ones [Ng and Russell, 2000, Amin and Singh, 2016]. The main result of this paper is that, unlike other No Free Lunch theorems, this unidentifiability does not disappear when regularising with a general simplicity prior that formalizes Occam's razor [Vitanyi and Li, 1997]. This result will be shown in two steps: first, that the simplest decompositions include degenerate ones, and secondly, that the most 'reasonable' decompositions according to human judgement are of high complexity. So, although current IRL methods can perform well on many well-specified problems, they are fundamentally and philosophically incapable of establishing a 'reasonable' reward function for the human, no matter how powerful they become. In order to do this, they will need to build in 'normative assumptions': key assumptions about the reward function and/or planner, that cannot be deduced from observations, and allow the algorithm to focus on good ways of decomposing the human policy. Future work will sketch out some potential normative assumptions that can be used in this area, making use of the fact that humans assess each other to be irrational, and often these assessments agree. In view of the No Free Lunch result, this shows that humans must share normative assumptions. One of these 'normative assumption' approaches is briefly illustrated in an appendix, while another appendix demonstrates how to use the planner-reward formalism to define when an agent might be manipulating or overriding human preferences. This happens when the agent pushes the human towards situations where their policy is very suboptimal according to their reward function.

S. Azam, F. Munir, A. Rafique, Y. Ko, A. M. Sheri, and M. Jeon. Object modeling from 3d point cloud data for selfdriving vehicles. In 2018 IEEE Intelligent Vehicles Symposium (IV), pages 409–414, June 2018. Humans drive in a holistic fashion which entails, in particular, understanding dynamic road events and their evolution. Injecting these capabilities in autonomous vehicles can thus take situational awareness and decision making closer to human-level performance. To this purpose, we introduce the ROad event Awareness Dataset (ROAD) for Autonomous Driving, to our knowledge the

first of its kind. ROAD is designed to test an autonomous vehicle's ability to detect road events, defined as triplets composed by an active agent, the action(s) it performs and the corresponding scene locations. ROAD comprises videos originally from the Oxford RobotCar Dataset, annotated with bounding boxes showing the location in the image plane of each road event. We benchmark various detection tasks, proposing as a baseline a new incremental algorithm for online road event awareness termed 3D-RetinaNet. We also report the performance on the ROAD tasks of Slowfast and YOLOv5 detectors, as well as that of the winners of the ICCV2021 ROAD challenge, which highlight the challenges faced by situation awareness in autonomous driving. ROAD is designed to allow scholars to investigate exciting tasks such as complex (road) activity detection, future event anticipation and continual learning. The dataset is available at https://github.com/gurkirt/road-dataset ; the baseline can be found at https://github.com/gurkirt/3D-RetinaNet .

In recent years, *autonomous driving* (or *robot-assisted driving*) has emerged as a fast-growing research area. The race towards fully autonomous vehicles pushed many large companies, such as Google, Toyota and Ford, to develop their own concept of *robot-car* [1], [2], [3]. While self-driving cars are widely considered to be a major development and testing ground for the real-world application of artificial intelligence, major reasons for concern remain in terms of safety, ethics, cost, and reliability [4]. From a safety standpoint, in particular, smart cars need to robustly interpret the behaviour of the humans (drivers, pedestrians or cyclists) they share the environment with, in order to cope with their decisions. *Situation awareness* and the ability to understand the behaviour of other road users are thus crucial for the safe deployment of autonomous vehicles (AVs). The latest generation of robot–cars is equipped with a range of different sensors (i.e., laser rangefinders, radar, cameras, GPS) to provide data on what is happening on the road [5]. The information so extracted is then fused to suggest how the vehicle should move [6], [7], [8], [9]. Some authors, however, maintain that vision is a sufficient sense for AVs to navigate their environment, supported by humans' ability to do just so. Without enlisting ourselves as supporters of the latter point of view, in this paper we consider the context of *vision-based* autonomous driving [10] from video sequences captured by cameras mounted on the vehicle in a streaming, online fashion. While detector networks [11] are routinely trained to facilitate object and actor recognition in road scenes, this simply allows the vehicle to 'see' what is around it.

The philosophy of this work is that robust self-driving capabilities require a deeper, more human-like understanding of dynamic road environments (and of the evolving behaviour of other road users over time) in the form of semantically meaningful concepts, as a stepping stone for intention prediction and automated decision making. One advantage of this approach is that it allows the autonomous vehicle to focus on a much smaller amount of relevant information when learning how to make its decisions, in a way arguably closer to how decision making takes place in humans.On the opposite side of the spectrum lies end-to-end reinforcement learning. There, the behaviour of a human driver in response to road situations is used to train, in an imitation learning setting [12], an autonomous car to respond in a more 'human-like' manner to road scenarios. This, however, requires an astonishing amount of data from a myriad of road situations. For highway driving only, a relatively simple task when compared to city driving, Fridman *et al.* in [13] had to use a whole fleet of vehicles to collect 45 million frames. Perhaps more importantly, in this approach the network learns a mapping from the scene to control inputs, without attempting to model the significant facts taking place in the scene or the reasoning of the agents therein. As discussed in [14], many authors [15], [16] have recently highlighted the insufficiency of models which directly map observations to actions [17], specifically in the self-driving cars scenario.

Existing System

In recent years, autonomous driving (or robot-assisted driving) has emerged as a fast-growing research area. The race towards fully autonomous vehicles pushed many large companies, such as Google, Toyota and Ford, to develop their own concept of robot-car. While self-driving cars are widely considered to be a major development and testing ground for the real-world application of artificial intelligence, major reasons for concern remain in terms of safety, ethics, cost, and reliability. From a safety standpoint, in particular, smart cars need to robustly interpret the behaviour of the humans (drivers, pedestrians or cyclists) they share the environment with, in order to cope with their decisions. Situation awareness and the ability to understand the behaviour of other road users are thus crucial for the safe deployment of autonomous vehicles (AVs). The latest generation of robot–cars is equipped with a range of different sensors (i.e., laser rangefinders, radar, cameras, GPS) to provide data on what is happening on the road. The information so extracted is then fused to suggest how the vehicle

should move Some authors, however, maintain that vision is a sufficient sense for AVs to navigate their environment, supported by humans' ability to do just so. Without enlisting ourselves as supporters of the latter point of view, in this paper we consider the context of vision-based autonomous driving [10] from video sequences captured by cameras mounted on the vehicle in a streaming, online fashion.

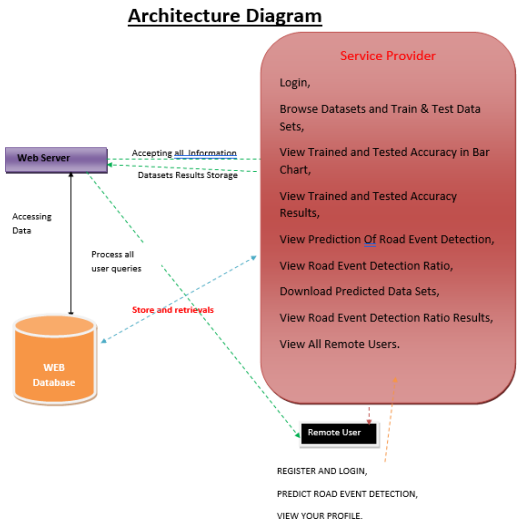
Drawback in Existing System

- **Limited diversity:** Datasets might lack diversity in certain scenarios, such as extreme weather conditions, unusual road layouts, or uncommon events. This limitation could impact the model's ability to handle rare or novel situations.
- **Labeling errors or biases:** Human labeling can sometimes be subjective or error-prone, leading to inaccuracies in the dataset. Additionally, biases in the dataset might reflect biases in the labeling process, affecting the model's performance.
- **Scale and volume:** Autonomous driving requires massive amounts of data to capture the variability of real-world driving conditions. The dataset's scale might be insufficient for training models robustly across all scenarios.
- **Temporal relevance:** With rapidly evolving technology and changing traffic conditions, datasets can become outdated. Newer challenges or events might not be adequately represented in older versions of the dataset.

Proposed System

- We propose to do so in a "holistic", multi-label approach in which agents, actions and their locations are all ingredients in the fundamental concept of road event (RE).
- The proposed conceptual shift. We are confident this contribution will be very useful moving forward for both the autonomous driving and the computer vision community.
- The proposed 3D-RetinaNet baseline, and recalls the ROAD challenge organised by some of us at ICCV 2021
- The first to propose an online, real-time solution to action detection in untrimmed videos, validated on UCF-101-24, and based on an innovative incremental tube construction method.

IMPLEMENTATION



Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Road Event Detection, View Road Event Detection Ratio, Download Predicted Data Sets, View Road Event Detection Ratio Results, View All Remote Users..

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT ROAD EVENT DETECTION, VIEW YOUR PROFILE.

CONCLUSION

This paper proposed a strategy for situation awareness in autonomous driving based on the notion of road events, and contributed a new Road event Awareness Dataset for Autonomous Driving (ROAD) as a benchmark for this area of research. The dataset, built on top of videos captured as part of the Oxford Robot Car dataset [18], has unique features in the field. Its rich annotation follows a multi–label philosophy in which road agents (including the AV), their locations and the action(s) they perform are all labeled, and road events can be obtained by simply composing labels of the three types. The dataset contains 22 videos with 122K annotated video frames, for a total of 560K detection bounding boxes associated with 1.7M individual labels. Baseline tests were conducted on ROAD using a new

3DRetinaNet architecture, as well as a Slow fast backbone and a YOLOv5 model (for agent detection). Both frame–MAP and video–MAP were evaluated. Our preliminary results highlight the challenging nature of ROAD, with the Slow fast baseline achieving a video-MAP on the three main tasks comprised between 20% and 30%, at low localization precision (20% overlap). YOLOv5, however, was able to achieve significantly better performance. These findings were reinforced by the results of the ROAD @ ICCV 2021 challenge, and support the need for an even broader analysis, while highlighting the significant challenges specific to situation awareness in road scenarios. Our dataset is extensible to a number of challenging tasks associated with situation awareness in autonomous driving, such as event prediction, trajectory prediction, continual learning and machine theory of mind, and we pledge to further enrich it in the near future by extending ROAD-like annotation to major datasets such as PIE and Waymo.

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 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. L_opez, "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.