ROAD: AUTONOMOUS DRIVING DATASETS FOR ROAD EVENT AWARENESS

Sivasankar Chittoor[1], Ankipalli Yaswanth Babu[2], M Vilekha[2], M Venkata Bharath Chandra[2], P Vivek[2],
V Madan Mohan Reddy[2]

[1]Assistant Professor, Department of Computer Science & Information Technology, Siddharth Institute of Engineering & Technology, Andhra Pradesh, India
[2]Research Scholar, Department of Computer Science & Information Technology, Siddharth Institute of Engineering & Technology, Andhra Pradesh, India

ABSTRACT

Humans drive holistically by understanding dynamic road events and their evolution over time. Incorporating these capabilities into autonomous vehicles can enhance their situational awareness and decision-making to be closer to human-level performance. To enable this, we introduce the Road Event Awareness Dataset (ROAD) for Autonomous Driving - the first of its kind to our knowledge. ROAD tests an autonomous vehicle's ability to detect road events, defined as triplets of an active agent, their actions, and corresponding scene locations. It comprises videos from the Oxford robotcar Dataset with bounding box annotations showing the location of each road event. We highlight the challenges still faced in situation awareness for autonomous driving. ROAD is designed to allow investigating exciting tasks like complex activity detection, future event anticipation, and continual learning.

Keyword: - *ROAD, Driving, Location, Anticipation*

1. 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 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 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. The information so extracted is then fused to suggest how the vehicle should move. Some authors, however, maintainthat 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 we consider the context of vision-based autonomous driving from video sequences captured by cameras mounted on the vehicle in a streaming, online fashion. While detector networks 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-drivingcapabilities require a deeper, more human-like understanding of dynamic road environments (andof 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 closerto 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, 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 situation

2. LITERATURE SURVEY

Various powerful people detection methods exist. Surprisingly, most approaches rely onstatic image features only despite the obvious potential of motion information for people detection. This paper systematically evaluates different features and classifiers in a sliding- window framework. First, our experiments indicate that incorporating motion information improves detection performance significantly. Second, the combination of multiple and complementary feature types can also help improve performance. And third, the choice of the classifier-feature combination and several implementation details are crucial to reach best performance. In contrast to many recent papers experimental results are reported for four different datasets rather than using a single one. Three of them are taken from the literature allowing for direct comparison. The fourth dataset is newly recorded using an onboard camera driving through urban environment. Consequently, this dataset is more realistic and more challenging than any currently available dataset. [1]

Although autonomous driving techniques have achieved great improvements, challengesstill exist in decision making for variety of different scenarios under uncertain and interactive environments. A good decision maker must satisfy the following requirements: (1) Be in a generic and unified form to cover as more scenarios as possible. (2) Be able to interact properly with other moving obstacles under the uncertainty of their motions. In this paper, the continuous decision making (CDM) framework is proposed to formulate different driving scenarios in a unified way, which encodes the high-level decision-making information into a continuous reference trajectory that can be naturally combined with a lower level trajectory planner. Withinthe framework, a maximum interaction defensive policy (MIDP) is proposed, which calculates the best action to interact with stochastic moving obstacles while guaranteeing safety. The method is applied to a ramp merging scenario and the stochastic behavior models of the surrounding vehicles are learned from the NGSIM dataset. Simulations are shown to visualize and analyze the results.[2]

Autonomous driving has attracted tremendous attention, especially in the past few years. The key techniques for a self-driving car include solving tasks like 3D map construction, self-localization, parsing the driving road, and understanding objects, which enable vehicles to reasonand act. However, large-scale data sets for training and system evaluation are still a bottleneck for developing robust perception models. In this project, present the ApolloScape datas and its applications for autonomous driving. Compared with existing public datasets from real scenes, e.g., , ApolloScape contains much large and richer labelling including holistic semantic dense point cloud for each site, stereo, per-pixel semantic labelling, lanemark labelling, instance segmentation, 3D car instance, high accurate location for every frame in various driving videos from multiple sites, cities and daytimes. For each task, it contains at lease 15x larger amount of images than SOTA datasets. To label such a complete dataset, develop various tools and algorithms specified for each task to accelerate the labelling process, such as joint 3D-2D segment labeling, active labelling in videos etc. Depend on ApolloScape, are able to develop algorithms jointly consider the learning and inference of multiple tasks. In this project, provide a sensor fusion scheme integrating camera videos, consumer-grade motion sensors (GPS/IMU), and a 3D semantic map in order to achieve robust self-localization and semantic segmentation for autonomous driving. Show that practically, sensor fusion and joint learning of multiple tasks are beneficial to achieve a more robust and accurate system. Expect our dataset and proposed relevant algorithms can support and motivate researchers for further development of multi-sensor fusion and multi-task learning in the field of computer vision.[3]

Study the question of feature sets for robust visual object recognition; adopting linear SVM based human detection as a test case. After reviewing existing edge and gradient based descriptors, show experimentally that grids of histograms of oriented gradient (HOG) descriptors significantly outperform existing feature sets for human detection. study the influence of each stage of the computation on performance, concluding that fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in overlapping descriptor blocks are all important for good results. The new approach gives near-perfect separation on the original MIT pedestrian database, so introduce a more challenging dataset containing over 1800 annotated human images with a large range of pose variations and backgrounds.[4]

Autonomous urban driving has a been rising problem since decades because of interactions with very complex environment. The traditional modular pipeline, rule-based algorithms, has not presented yet an efficient model to rely on, because it cannot cover the very large possible scenarios space. Machine Learning techniques like supervised learning or imitation learning and Reinforcement learning have initial promising results with better performance. propose an end-to-end Deep Conditional Imitation Learning model for autonomous driving inspired by both of Intel and Nvidia. The feature extraction part for Intel is replaced with Nvidia's model. Our proposed model outperforms Intel's architecture performance on CARLA Simulator, and overgeneralizes on various towns with different weather conditions.[5]

3. METHODOLOGY

3.1EXISTING SYSTEM

In Existing system, Single-Modality Datasets. Collecting and annotating RGB data only isrelatively less time-consuming and expensive than building multimodal datasets including range data from LiDAR or radar. Most single-modality datasets provide 2D bounding box and scene segmentation labels for RGB images. Examples include Cityscapes Mapillary Vistas, BDD100k and Apolloscape. To allow the studying of how vision algorithms generalize to different unseen data, collect RGB images under different illumination and weather conditions.

Other datasets only provide pedestrian detection. Recently, MIT and Toyota have releasedDriveSeg, which comes with pixel level semantic labeling for 12 agent classes. Multimodal Datasets. KITTI was the first-ever multimodal dataset. It provides depth labels from front-facing stereo images and dense point clouds from LiDAR alongside GPS/IMU (inertial) data. It also provides bounding-box annotations to facilitate improvements in 3D object detection. H3D and KAIST are two more examples of multimodal datasets. H3D provides 3D box annotations, using real-world LiDAR-generated 3D coordinates, in crowded scenes.

3.1.1DISADVANTAGES OF EXISTING SYSTEM

- The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets to road events.
- Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer

3.2 PROPOSED METHODOLOGY

A conceptual shift in situation awareness centred on a formal definition of the notion of road event, as a triplet composed by a road agent, the action(s) it performs and the location(s) of the event, seen from the point of view of the AV. A new ROad event Awareness Dataset for Autonomous Driving (ROAD), the first of its kind, TO designed to support this paradigm shift and allow the testing of a range of tasks related to situation awareness for autonomous driving: agent and/or action detection, event detection, ego-action classification.

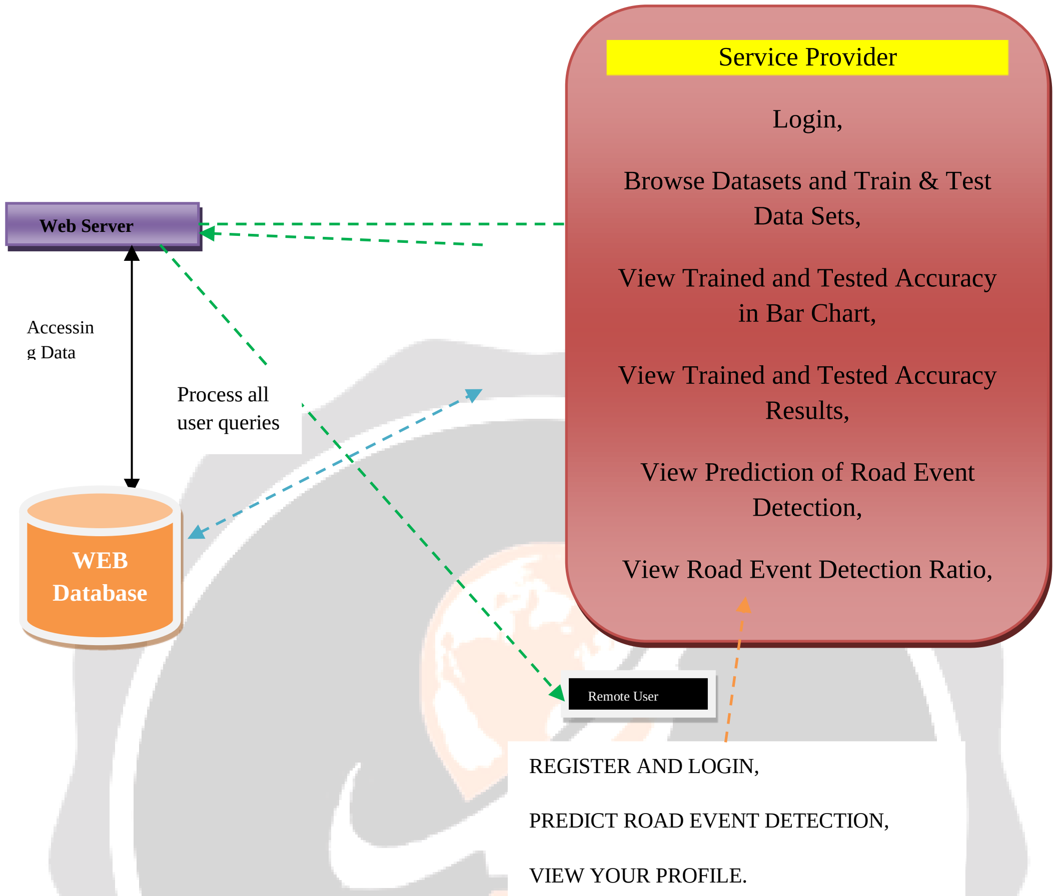
This work aims to propose a new framework for situation awareness and perception, departing from the disorganized collection of object detection, semantic segmentation or pedestrian intention tasks which is the focus of much current work. We propose to do so in a “holistic”, multi-label approach in which agents, actions and their locations are all ingredients in the fundamental concept of road event (RE).

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 thescene 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 modelthe causal relationships between what happens; these causality links can then be exploited to predict further future consequences.

4. SYSTEM DESIGN

It is a process of planning a new business system or replacing an existing system by defining its components or modules to satisfy the specific requirements. Before planning, you needto understand the old system thoroughly and determine how computers can best be used in order to operate efficiently.

4.1 SYSTEM ARCHITECTURE



4.2 MODULES

In this Proposed System, There are two Modules. They are:

1. Service Provider
2. Remote User

4.2.1 SERVICE PROVIDER

This system should provide the service provider with the convenience of providing trainingand testing of dataset

- Login
- Browse Dataset
- Train and Test Data sets
- Generate Trained and Tested Results Accuracy
- View Trained and Tested Results Accuracy in Bar Chart
- View Trained and Tested Accuracy Results
- View Prediction of Road Event Detection
- Download Prediction on dataset
- View All Remote Users
- Logout

4.2.2 REMOTE USER

This system should help the user by registering with his basic details that can be stored inthe database and it provides the following such as

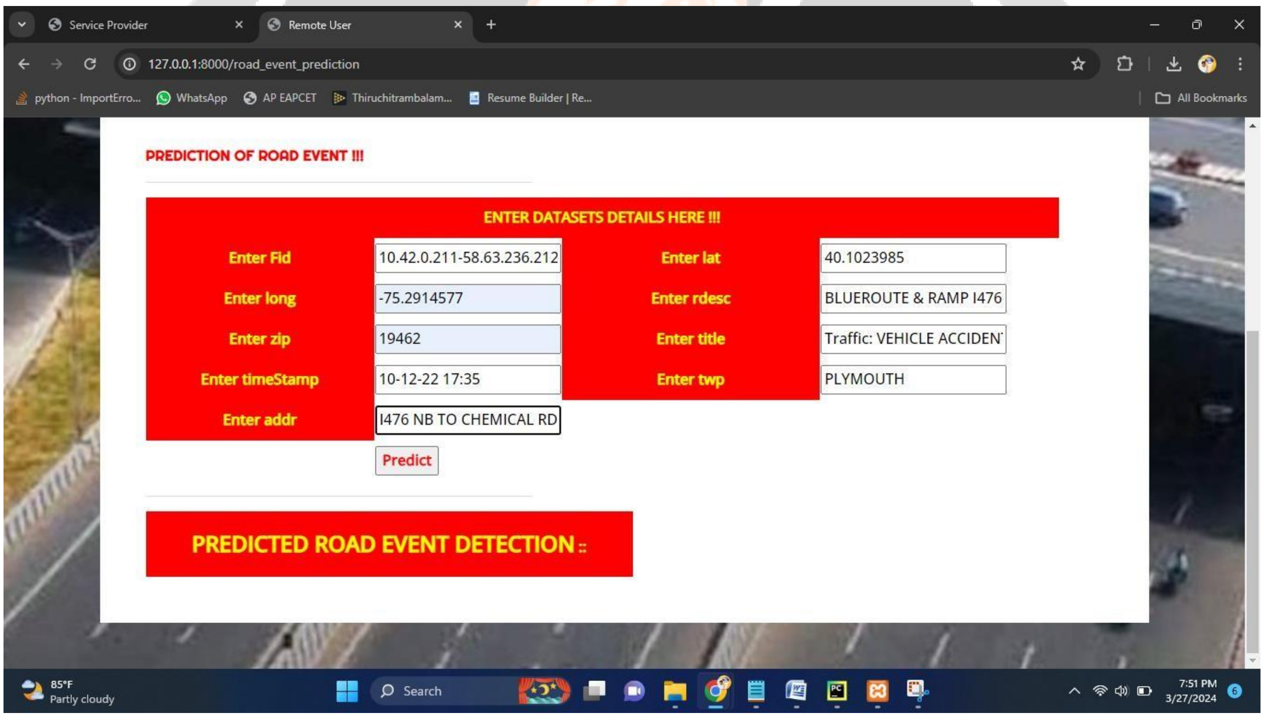
- Register
- Login
- Enter Details for Prediction
- Prediction Road Event Detection
- Logout

5. RESULTS AND PERFORMANCE

EXECUTION PROCEDURE

The Execution procedure is as follows:

1. In this research work with data with attributes are observable and then all of them are floatingdata. And there’s a decision class/class variable. This data was collected from Kaggle machine learning repository.
2. In this research 70% data use for train model and 30% data use for testing purpose.
3. Naïve Bayes is used as Classifier.



4. In the classification report we were able to find out the desired result
5. In this analysis the result depends on some part of this research. However, which algorithm gives the best true positive, false positive, true negative, and false negative are the best algorithmsin this analysis.

Fig. Enter Details for Prediction

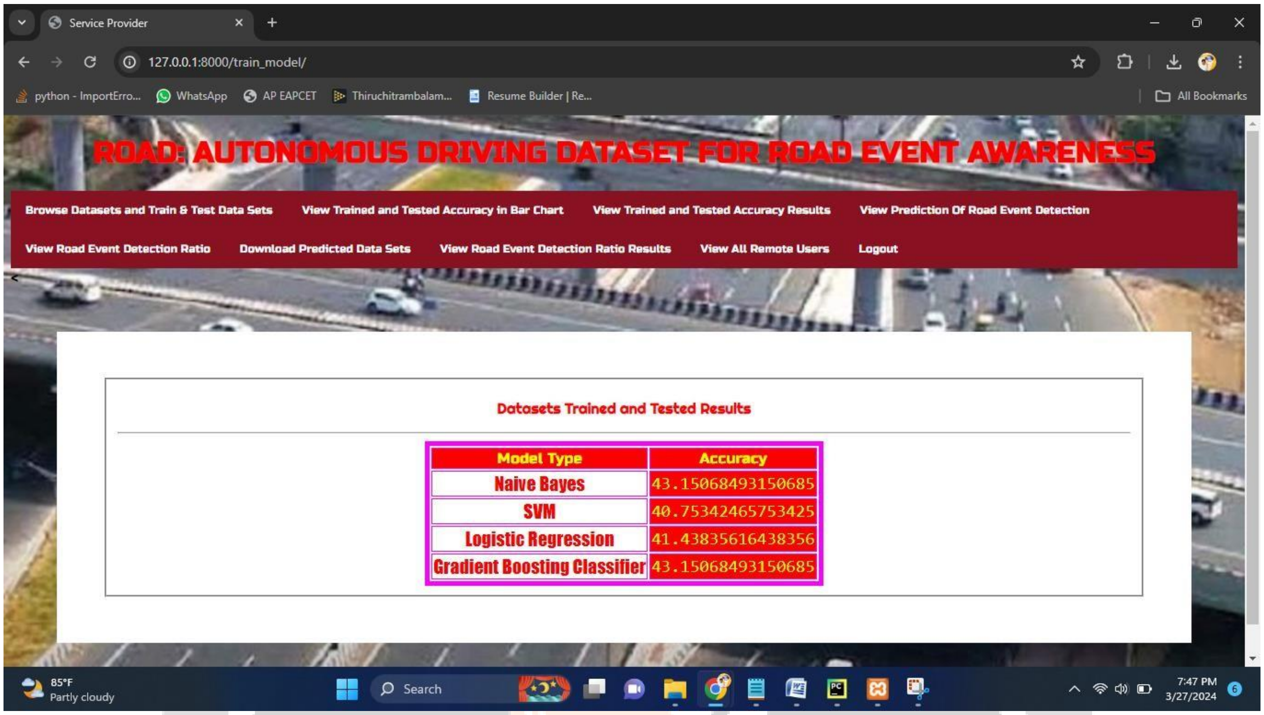


Fig. Datasets Trained and Tested Results

6. CONCLUSION

In this project, a strategy for situation awareness in autonomous driving based on the notionof 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, has unique features in the field. Its rich annotation follows amulti–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. 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.

7. REFERENCE

[1]. Basepaper: ROAD: The Road Event Awareness Dataset for Autonomous Driving

[2]. S. Azam, F. Munir, A. Rafique, Y. Ko, A. M. Sheri, and M. Jeon, “Object modeling from3D point cloud data for self-driving vehicles,” in Proc. IEEE Intell. Veh. Symp., 2018, pp.409–414.

[3]. 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.

[4]. M. Codevilla, Felipe Dosovitskiy, “End-to-end driving via conditional imitation learning,”in Proc. IEEE Int. Conf. Robot. Automat., 2018, pp. 1–9.

[5]. 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.

[6]. H. Caesar et al., “nuScenes: A multimodal dataset for autonomous driving,” in Proc.

IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 11 621–11 631.