ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue.2 April 2024

ROAD THE ROAD EVENT AWARENESS DATASET FOR AUTONOMOUS DRIVING

Vardelli Rakesh[1], Prushotham Sharath [2],
Rade Ravi Teja [3], Dr. T. Anilkumar
[1,2,3] UG Student, Dept. of ECE, CMR Institute of Technology, Hyderabad
[4]Professor, Dept. of ECE
CMR Institute of Technology, Hyderabad

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.

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 largecompanies, such as Google, Toyota and Ford, to develop their own concept of robot-car [100, 50, 65]. While selfdriving 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 [59]. 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 [6]. The information so extracted is then fused to suggest how the vehicle should move [2, 28, 95, 14]. 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 [4] from video sequences captured by cameras mounted on the vehicle in a streaming, online fashion. While detector networks [76] are routinely trained to facilitate object and actor recognition in road scenes, this simply allows the vehicle to 'see' what is around it, without any real understanding of the scene context. Our position is that robust self-driving capabilities require a deeper, more human-like understanding of the road environment and of the evolving behaviour of other road users over time. Behavioural inference has been proposed as an option [25], as the historical behaviour of road users can be used to predict possible future events in accordance with a discrete set of policies, defined at programming time.

Literature survey

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.

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.

Harkirat S Behl, Michael Sapienza, Gurkirt Singh, Suman Saha, Fabio Cuzzolin, and Philip HS Torr. Incremental tube construction for human action detection. arXiv preprint arXiv:1704.01358, 2017.

Current state-of-the-art action detection systems are tailored for offline batch-processing applications. However, for online applications like human-robot interaction, current systems fall short. In this work, we introduce a real-time and online joint-labelling and association algorithm for action detection that can incrementally construct space-time action tubes on the most challenging untrimmed

action videos in which different action categories occur concurrently. In contrast to previous methods, we solve the linking, action labelling and temporal localization problems jointly in a single pass. We demonstrate superior online association accuracy and speed (1.8ms per frame) as compared to the current state-of-the-art offline and online systems. Detecting human actions has been defined as the task of automatically predicting the start, end and spatial extent of various actions [10, 21, 23] by predicting sets of connected windows in time (called tubes) in which each action is enclosed, as illustrated in Fig.1. Human action detection has gained huge popularity in the computer vision community due to its broad range of exciting applications

Massimo Bertozzi, Alberto Broggi, and Alessandra Fascioli. Vision-based intelligent vehicles: State of the art and perspectives. Robotics and Autonomous Systems, 32(1):1 – 16, 2000.

Recently, a large emphasis has been devoted to Automatic Vehicle Guidance since the automation of driving tasks carries a large number of benefits, such as the optimization of the use of transport infrastructures, the improvement of mobility, the minimization of risks, travel time, and energy consumption. This paper surveys the most common approaches to the challenging task of Autonomous Road Following reviewing the most promising experimental solutions and prototypes developed worldwide using AI techniques to perceive the environmental situation by means of artificial vision.

EXISTINGSYSTEM

Single-Modality Datasets. Collecting and annotating RGB data only is relatively less time-consuming and expensive than building multimodal datasets including range data from LiDAR or radar. Most single-modality datasets [23], [24], [25], [26], [27], [28] provide 2D bounding box and scene segmentation labels for RGB images. Examples include Cityscapes [24], Mapillary Vistas [25], BDD100k [26] and Apolloscape [27]. To allow the studying of how vision algorithms generalise to different unseen data, [25], [26], [28] collect RGB images under different illumination and weather conditions.

Other datasets only provide pedestrian detection annotation [29], [30], [31], [32], [33], [34], [35]. Recently, MIT and Toyota have released DriveSeg, which comes with pixellevel semantic labelling for 12 agent classes [36]. Multimodal Datasets. KITTI [37] 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 [38] and KAIST [39] are two more examples of multimodal datasets. H3D provides 3D box annotations, using real-world LiDAR-generated 3D coordinates, in crowded scenes.

Disadvantages

• 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.

• Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

PROPOSED SYSTEM

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, 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 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.

Advantages

_ A multi-label benchmark: each road event is composed by the label of the (moving) agent responsible, the label(s) of the type of action(s) being performed, and labels describing where the action is located.

_ Each event can be assigned multiple instances of the same label type whenever relevant (e.g., an RE can be an instance of both moving away and turning left).

_ The labeling is done from the point of view of the AV: the final goal is for the autonomous vehicle to use this information to make the appropriate decisions.

_ The meta-data is intended to contain all the information required to fully describe a road scenario: an illustration of this concept is given in this system. After closing one's eyes, the set of labels associated with the current video frame should be sufficient to recreate the road situation in one's head (or, equivalently, sufficient for the AV to be able to make a decision).

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 asBrowse 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 likeREGISTER AND LOGIN, PREDICT ROAD EVENT DETECTION, VIEW YOUR PROFILE.

ALGORITHMS

DECISION TREE CLASSIFICATION ALGORITHM

- Decision Tree is a **Supervised learning technique** that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where **internal nodes represent the features of a dataset, branches represent the decision rules** and **each leaf node represents the outcome.**
- In a Decision tree, there are two nodes, which are the **Decision Node** and **Leaf Node.** Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

- The decisions or the test are performed on the basis of features of the given dataset.
- ***It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.***
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

K-NEAREST NEIGHBOR(KNN) ALGORITHM FOR MACHINE LEARNING

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K-NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
- It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data,then it classifies that data into a category that is much similar to the new data.

LOGISTIC REGRESSION —

DETAILED OVERVIEW

In statistics, the logistic model (or logit model) is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. This can be extended to model several classes of events such as determining whether an image contains a cat, dog, lion, etc. Each object being

ISSN2321-2152
www.ijmece .com
Vol 12, Issue.2 April 2024

detected in the image would be assigned a probability between 0 and 1, with a sum of one.

NAVIE BAYES

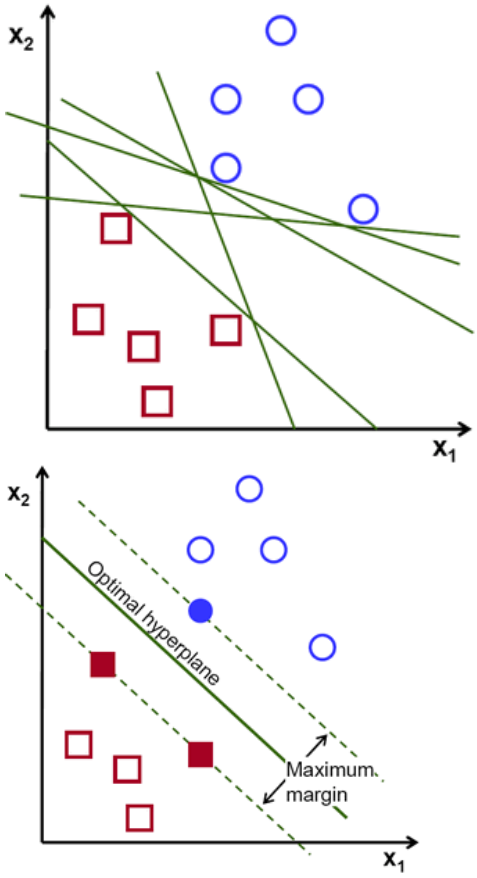
Naive Bayes classifiers are a collection of classification algorithms based on **Bayes' Theorem**. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

To start with, let us consider a dataset.

Consider a fictional dataset that describes the weather conditions for playing a game of golf. Given the weather conditions, each tuple classifies the conditions as fit("Yes") or unfit("No")

SUPPORT VECTOR MACHINE?

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.



CONCLUSION

This paper proposed a strategy for situation awareness inautonomous driving based on the notion of road events, and contributed a new Road event Awareness Dataset forAutonomous Driving (ROAD) as a benchmark for this areaof research. The dataset, built on top of videos captured aspart of the Oxford RobotCar dataset [18], has unique featuresin the field. Its rich annotation follows a multi–

labelphilosophy in which road agents (including the AV), theirlocations and the action(s) they perform are all labeled, androad events can be obtained by simply composing labels ofthe three types. The dataset contains 22 videos with 122Kannotated video frames, for a total of 560K detection boundingboxes 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.

[12] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Design of a Dual Doping Less Double Gate Tfet and Its Material Optimization Analysis on a 6t Sram Cells."

[13] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Low power process, voltage, and temperature (PVT) variations aware improved tunnel FET on 6T SRAM cells." Sustainable Computing: Informatics and Systems 21 (2019): 143-153.

[14] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Survey on improvement of PVT aware variations in tunnel FET on SRAM cells." In 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), pp. 703-705. IEEE, 2017

[15] Karne, R. K. ., & Sreeja, T. K. . (2023). PMLC- Predictions of Mobility

and Transmission in a Lane-Based Cluster VANET Validated on Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5s), 477–483. https://doi.org/10.17762/ijritcc.v11i5s.7109

[16] Radha Krishna Karne and Dr. T. K. Sreeja (2022), A Novel Approach for Dynamic Stable Clustering in VANET Using Deep Learning (LSTM) Model. IJEER 10(4), 1092-1098. DOI: 10.37391/IJEER.100454.

[17] Reddy, Kallem Niranjan, and Pappu Venkata Yasoda Jayasree. "Low Power Strain and Dimension Aware SRAM Cell Design Using a New Tunnel FET and Domino Independent Logic." International Journal of Intelligent Engineering & Systems 11, no. 4 (2018).

[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 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.

[21] G. Gkioxari and J. Malik, "Finding action tubes," in Proc. IEEE
Int. Conf. Comput. Vis. Pattern Recognit., 2015, pp. 759–768.

[22] C. Feichtenhofer, H. Fan, J. Malik, and K. He, "Slowfast networks
for video recognition," in Proc. IEEE Int. Conf. Comput. Vis., 2019,
pp. 6202–6211.

[23] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla,
"Segmentation and recognition using structure from motion
point clouds," in Proc. Eur. Conf. Comput. Vis., 2008, pp. 44–57.

[24] M. Cordts et al., "The cityscapes dataset for semantic urban scene
understanding," in Proc. Conf. Comput. Vis. Pattern Recognit.,
2016, pp. 3213–3223.

[25] G. Neuhold, T. Ollmann, S. Rota Bulo , and P. Kontschieder, "The
mapillary vistas dataset for semantic understanding of street
scenes," in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 4990–4999.