Road Event Detection using using 3D-RetinaNet and ROAD-R 2023 Challenge

Srutanik Bhaduri and Anshuman Dangwal
Data Science and Engineering
Indian Institute of Science Education and Research Bhopal
srutanik21@iiserb.ac.in anshuman21@iiserb.ac.in

Abstract

Neural networks have proven to be very powerful at computer vision tasks. However, they often exhibit unexpected behaviors, acting against background knowledge about the problem at hand. This calls for models (i) able to learn from requirements expressing such background knowledge, and (ii) guaranteed to be compliant with the requirements themselves. Unfortunately, the development of such models is hampered by the lack of real-world datasets equipped with formally specifed requirements. In this paper, we introduce the ROad event Awareness Dataset with logical Requirements (ROAD-R) [1], the frst publicly available dataset for autonomous driving with requirements expressed as logical constraints. Given ROAD-R, we show that current state-of-the-art models often violate its logical constraints, and that it is possible to exploit them to create models that (i) have a better performance, and (ii) are guaranteed to be compliant with the requirements themselves.

1. Problem Statement

This project as well as the competition, ROAD-R 2023, consists of 2 tasks. The dataset provide comprises of 22 (about 8 minutes long) videos annotated with road events together with a set of 243 requirements expressing hard facts about the world that are commonly known by humans (e.g., "a traffic light cannot be red and green at the same time") and develop the best-performing model with only a subset of the annotated data. In the second task we create systems whose predictions are compliant with the requirements. The tasks are:

1. To develop the best-performing model with only a subset of the annotated data
2. To create systems whose predictions are compliant with the requirements

The main aim is to first create a model which predicts the labels correctly with minimum error. Furthermore if possible we would also take the challenge to modify the model to be compliant with the requirements.

2. Motivation

The motivation behind this project is the upcoming future of transportation. In a world full of Teslas and self driving cars, companies like Waymo are developing and building better systems for autonomous transportation. So we want to mimic just a portion of that and see how the whole thing works. Because in the future these systems will be prevalent all over the world and learning about them would give us a huge amount of exposure and skills required for developing even better systems. Another motivation is participating in an open world challenge like ROAD-R 2023.

3. Related Work and Limitations

Research on vehicle based object detection has been mostly done by industrial research labs like Meta, Google, Toyota etc. These companies are those who are actively developing systems for autonomous vehicles. Research on these topics in academia is very limited to only a few papers like this[3]. So I believe creating a project on this topic can result in a good research output.

4. Dataset

We will be using the dataset provided by ROAD-R themselves, called the ROAD dataset [2] which is specially designed from the perspective of self-driving cars which includes actions performed by humans, cars and other objects annotated as road events(REs). A road event corresponds to a tube which is a sequence of frame-wise bounding boxes linked in time. The dataset is publicly available and was built on top of the Oxford RobotCar Dataset which consists of 22 (about 8 minutes each) videos annotated with road events. There are three different labels, namely:

- Category of road agent involved(e.g. Pedestrian, Car, Cyclist, etc)

- Type of action performed by agent(e.g. Moving Away, Moving Towards, etc)
- Location of the agent relative to the autonomous vehicle perspective scene(e.g. In vehicle lane, On right pavement, etc)

Table 1. ROAD active agent classes with description.

Label name	Description
Autonomous-vehicle	The autonomous vehicle itself
Car	A car up to the size of a multi-purpose vehicle
Medium vehicle	Vehicle larger than a car, such as van
Large vehicle	Vehicle larger than a van, such as a lorry
Bus	A single or double-decker bus or coach
Motorbike	Motorbike, dirt bike, scooter with 2/3 wheels
Emergency vehicle	Ambulance, police car, fire engine, etc.
Pedestrian	A person including children
Cyclist	A person is riding a push/electric bicycle
Vehicle traffic light	Traffic light related to the AV lane
Other traffic light	Traffic light not related to the AV lane

Figure 1: Few of the annotations under ROAD-R

5. Algorithms

The 3D-RetinaNet baseline code uses the method of frame extraction and uses frames as an input. For checking peformance of the model, we have to do a trial and error method with several algorithms like:

- Semantic Segmentation:
  Fully Convolutional Networks (FCN)
- Instance Segmentation:
  Mask R-CNN
- Background Subtraction:
  Simple Background Subtraction
- Motion-Based Segmentation:
  Motion Detection Techniques
- Deep Learning-Based Approaches:
  3D CNNs Temporal Convolutional Networks (TCNs)

6. Work Plan

The work plan is still in progress but below is a very short description on how the project is going to move ahead.

6.1. Phase 1 - Annotation of Extracted Frames using Deep Learning Models

From the sets of videos, the extracted frames would be divided into the 243 annotations provided which would be useful in identifying the objects present on the road. This would also suffice for the Task 1 of the ROAD-R 2023 competition.

6.2. Phase 2 - Create systems whose predictions are compliant with the requirements:

In this phase we would focus on the task 2 of the ROAD-R 2023 competition. This part is still under consideration and would be done only if Phase 1 is completed successfully.

7. Collaborators

I am working with Anshuman Dangwal (21044) of DSE department on this project. We would work as a team and take up different parts of the project according to our liking. We would both contribute equally and make sure that this project ends succesfully.

8. Project Output

The desired output for this project would be a research paper in NeurIPS 2023 and to win the competition. Although a bit far fetched, but ideally this could be the output of this. But if not successful in doing so, we would like to propose for a paper for the project itself.

9. Conclusion

I believe this project has the capability to produce some good results in the field of object detection in moving vehicles. Also the idea is a very interesting one and by applying novel methods we can create a desired output with our machine learning and deep learning process. Overall it would a great opportunity to learn more about the subject and dive deeper into the field of computer vision.

References

[1] Eleonora Giunchiglia, Mihaela Cătălina Stoian, Salman Khan, Fabio Cuzzolin, and Thomas Lukasiewicz. Road-r: the autonomous driving dataset with logical requirements. *Machine Learning*, 112(9):3261–3291, Sep 2023.

[2] Gurkirt Singh, Stephen Akrigg, Manuele Di Maio, Valentina Fontana, Reza Javanmard Alitappeh, Suman Saha, Kossar Jeddisaravi, Farzad Yousefi, Jacob Culley, Tom Nicholson, et al. Road: The road event awareness dataset for autonomous driving. *IEEE Transactions on Pattern Analysis Machine Intelligence*, (01):1–1, feb 5555.

[3] Huansheng Song, Haoxiang Liang, Huaiyu Li, Zhe Dai, and Xu Yun. Vision-based vehicle detection and counting system using deep learning in highway scenes. *European Transport Research Review*, 11(1):51, Dec 2019.