Chapter 1

Introduction

This section of the thesis first highlights the background in section 1.1, the vision-based architecture of the autonomous vehicles (AVs) in section 1.2, the motivation behind the present work 1.3, an introduction to the problems discussed in section 1.4 and objectives of this thesis in section 1.5. Finally, the chapter discussed a list of contributions of this thesis in section 1.6 in the field of autonomous vehicles. At last, the chapter concludes with thesis organization in section 1.7.

1.1 Background

The first AVs were conceived as Remote-control automobiles in the early 1920s. The development of AVs will be prolonged in the next 50 years as pre-directing control has stalled. The AVs developments rose in 1980 over the next 20 years, and academics and industries began using vision control. Today life is become easy due to computers, control algorithms and sensors that make AVs more realistic for the common public. The Defense Advanced Research Projects Agency (DARPA) grand challenge from 2007 to till date, some technological companies like Unified Best Economical Ride (Uber), Tera electron volt Energy Superconducting Linear Accelerator (Tesla) and Google reported their vehicles, predicting that even though we may not see AVs on the road yet, we can expect them within two decades. Therefore, there is a need to foresee AVs possible benefits for the public and their influence on traffic networks.

For comfortable and safe traveling, the AVs could be more helpful because of their pre-programmed driving skills/conditions. AVs can eliminate human factors like driving while calling or drinking, which may cause accidents. In addition, sensor-based computers can take better decisions to avoid obstacles and are easy to perform to avoid collisions. The faster decision-making of AVs than humans and the number of sensors make the vehicles more capable of detecting and handling more complex situations and sometimes, they may be able to avoid dangerous situations rather than human prediction. AVs drive better than human drivers, which motivates people to have their own AVs.

Apart from comfort and safety, AVs also gives the liberty of free hand. A driver can get more privileged of in-vehicle travel drive time to focus on other activities, using a cell phone, enjoying entertainment, working, and taking a nap. There are very benefits of AVs that make the AVs more demanding and attract passengers to travel more. The distance and number of traveling will increase the utilization of

AVs. They are more interested in taking additional trips- trips but they did not have time to take- if they could use this in-vehicle travel time to get other things done.

The more road-trips may cause more traffic, but the AVs help reduces the crowd on the road. The sensors and computers are benefitted the AVs that can operate with less decision time and follow a shorter distance compared to other road agents. AVs can also quickly deal with acceleration and deceleration smoothly.

1.2 Vision-Based Architecture of Autonomous Vehicles

Over the previous years, Computer vision has achieved a new level of sophistication. The deep-learning development, low rate with high processing and large datasets availability are the factors that motivated us to use computer vision-based work that surpasses human performance. Considering the advanced computer vision methods and the camera sensor can effectively address all the perception scenarios. In this thesis, the vision-based architecture of the AVs has illustrated in Fig-1.1. The vision-based autonomous vehicle architecture is typically categorized into many parts.

The automatic system for the vehicles needs to understand its surroundings and localize itself, take decisions for vehicles, plan the route and actuate the controls. In this thesis, like other robotic systems, the distribution of the solution for the AVs system is in the following manner as mentioned in Fig-1.1.

1.2.1 Sensors

The AVs is used various types of sensors to visualize the external surroundings and collect information about the surroundings. Various types of sensors like camera, light detection and ranging (LiDAR). Radio detection and ranging (RADAR). The fusion of data collected by these sensors may produce more informative information. The GPS sensor is used for localization and route mapping.

1.2.2 Perception System

The perception subsystem is used to convert the raw information into meaningful information. The essential computational block of the AVs is the perception subsystem. The perception system helps to extract semantic information for the AVs from their surroundings without any involvement of a human driver. This information is very useful for producing a collision avoidance path and guaranteeing that the autonomous vehicle systematizes with traffic and road regulations. The perception system of AVs has various subsystems to perform many functions like object detection, object tracking and trajectory prediction (TP) etc.

1.2.2.1 Object Detection

The object detection subsystem is responsible for detecting moving objects. The object detection task is responsible for recognizing the object class and localizing the object to get the object's position in a 2D image.

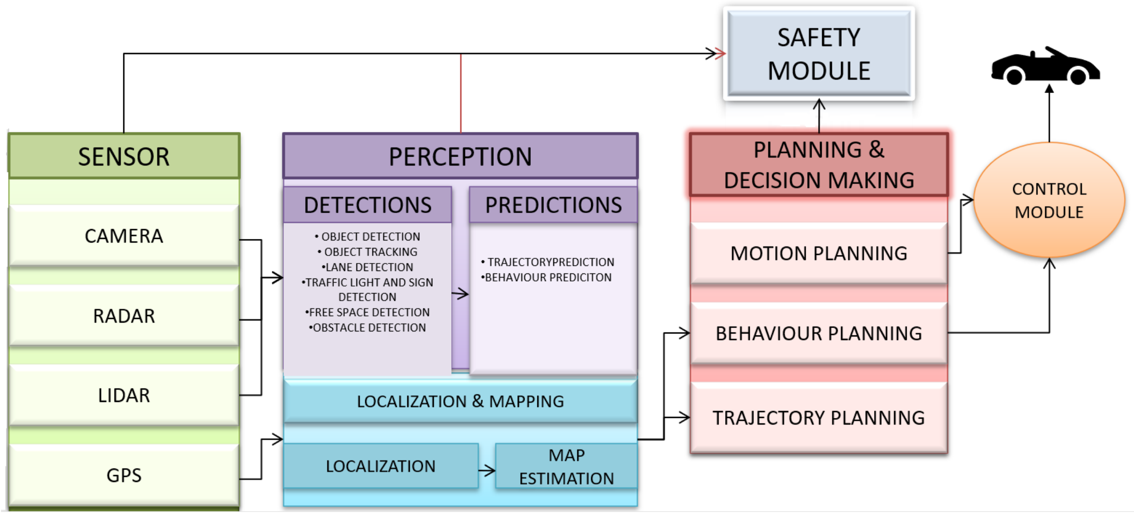

FIGURE 1.1: Overview of Vision-based Architecture for Autonomous Vehicles

1.2.2.2 Object Tracking

Object tracking subsystem is responsible for predicting the next location of the moving object. Multi-object tracking aims to track multiple objects in digital images. It is also known as Multi-target tracking, as it aims to analyze videos to recognize the objects ("targets") that belong to different predetermined classes.

1.2.2.3 Trajectory Prediction

The TP subsystem of the AVs is responsible for estimating the probability of the other objects with time-stamps present in surrounding areas. This subsystem is worked to deduce possible trajectories for the surrounding objects. TP of autonomous vehicles predicts other vehicle trajectories to take action for collision-free routes and abide by the traffic rules.

1.2.3 Mapping and Localization

This system acts as a global planner for the vehicle.

1.2.3.1 Localization

In the vision-based architecture, the localization subsystem's task is to process the GPS sensor data to get the exact estimation of the current location. The responsibility of the localization submodule is assumed to obtain the GPS coordinates of the vehicles. The vehicles need to provide their own location into a world coordinate frame.

1.2.3.2 Map State Estimation

The Map State Estimation subsystem provides the current location and the location of the destination, which helps find all the required actions to reach the destination. This sub-module is used to provide the map and different route for the destination that helps to select the best route.

1.2.4 Planning and Decision-Making

The sensory input is the combination of perception, mapping and localization used to understand the surroundings and the planner works as a brain of an autonomous vehicle to process the information. The task of the planning subsystem is to plan the local task and behavior of the autonomous vehicle and the control system with the final waypoints that the vehicle is interested in following. Since we are assuming vision-based driving, the reference frame for the planning will be the ego-vehicle.

1.2.4.1 Trajectory Planning

The task of trajectory planning is to provide a valid set of trajectories such that the AVs avoid collision with other road objects. The task of the trajectory planning subsystem is to provide a valid set of trajectories such that the AVs avoid collision with other road agents. The trajectory planning subsystem help to provide the control system of the AVs with the final set of waypoints that the vehicle should follow.

1.2.4.2 Behaviour Planning

The behavior planning subsystem is responsible for deciding to go left, right, and straight from the intersection based on its position, traffic signs and traffic signals.

1.2.4.3 Motion planning

The motion planning task breaks down the complete motion into small discrete motions that need to satisfy the motion constraints and possibly optimize some aspects of the movement.

1.2.5 Safety Module

The task of the safety module is to ensure that all system and subsystem is working as needed. The safety module system is usually created to involve the list of possible malfunctions by documenting a safety plan like risk assessment, a hazard analysis, a technical concept of safety and a safety requirement software.

1.2.6 Control Module

The control module's task is to produce control signals like acceleration, brake, steering torque, etc., to ensure that the vehicle accompanies the trajectory planning system. The proportional integral derivative (PID) controller is used to minimize the crossroad of the trajectory of the vehicle and selected waypoints.

1.3 Motivation

The main cause of road accidents (approximately 95-96%) is human mistakes. The number of road accidents, deaths and injured people in India over the years is mentioned in Fig- 1.2. The use of AVs will help to reduce this with large margins. The human driver even used smartphones, distraction and fatigue for other reasons. [5] The AVs will help prevent most accidents. In addition, the AVs are designed and programmed to handle road conditions. Therefore, if the failure of programming is detected that has not been found in the tests, the vehicles will be able to notify an update on how to correct the failure. The autonomous vehicle will update itself and covey all other vehicles to conduct themselves in a similar case [6].

The autonomous vehicle will be trained with the experience of all other AVs

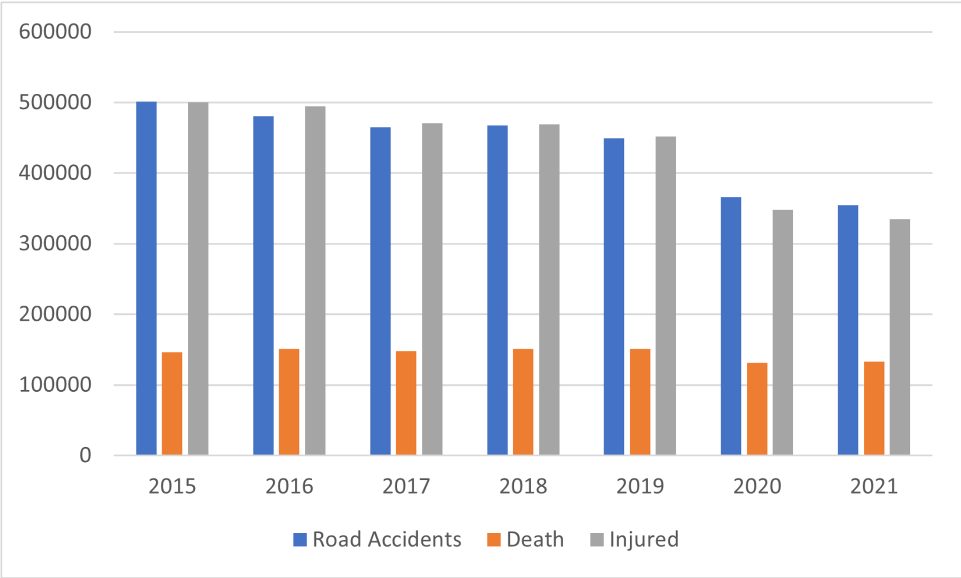

FIGURE 1.2: The road accidents, deaths and injured people over the year in India

while the human driver learns by their own experience. AVs also be helpful for disabled people (Divyang People) [7]. For example, the visual impairments, people can safely travel from one place to another place safely and comfortably in autonomous vehicles. In addition, busy people with exhausting jobs or drunken people can safely travel without a driver. Moreover, AVs also release people from the burden of driving and will free them for other activities during traveling like work, enjoying the view of the journey and rest [8].

Some other issues are air pollution and traffic, which will be well moderated. The smart junctions, where traffic is routed according to the slots allocation of the junction, will help to reduce the delays and traffic jams [9]. These junctions are expected to have a strong effect on the road system of each city. The travel and waiting times will be considerably reduced [10]. The fuel consumption will also be decreased and the crowd will flow freely and smoothly, with fewer accelerations and stopping that help to reduce the air pollution [11] significantly.

Another issue is addressed, which is air pollution. Traffic congestion and other delays will be considerably reduced at smart junctions where traffic is routed based on the allotment of slots to pass the intersection [12]. Such intersections are projected to impact each city's road infrastructure significantly. Travel and waiting times will be significantly reduced [13]. Fuel consumption will be reduced, and traffic will flow more smoothly with fewer stops and accelerations, reducing air pollution dramatically [14].

Because a significant quantity of gas is spent when driving at a very high speed (too high for the vehicle), accelerating excessively or braking without reason, autonomous cars will consume far less energy and gas while traveling than a regular vehicle operated by a human. AVs will remove these inefficient modes of transportation, resulting in lower air pollution.

1.4 Problem statement

A lot of literature has been reported on the different modules of the AVs. This thesis addresses challenges and issues arising for different modules and sub-modules of AVs. The problem statement of the research can be defined as.

"Design and development of algorithms for perception and planning sub-modules, design and implementation of new methods and models

by using machine-learning/deep-learning based approaches for providing reliable navigation and path determination of autonomous vehicles"

1.5 Thesis Objective

This thesis aims to introduce different deep-learning approaches for perception and planning sub-module with better performance in terms of time and accuracy compared to the other state-of-the-art techniques. Firstly, a deep learning-based approach has been introduced for object detection. Then an approach for multi-object tracking has been introduced i.e. detection-based tracking approach to scale-up two tasks into an end-to-end model. The third is proposed for TP using graph neural network spatial-temporal features with RNN and the last is CARLA simulator-based motion planning for the AVs. These models have been evaluated on publicly available datasets like KITTI, BDD, MOT, Waymo, Lyft, Argoverse and Apolloscape datasets for AVs. The objectives of this thesis are as

- An extensive study of the existing literature on experiments and research performed under various modules of AVs using deep-learning approaches to identify the research gap.
- An extensive study of literature on the evolution of modern datasets for autonomous vehicles.
- To study and implement object detection method based on a single-stage method. Further design and implementation of an object detection method

can efficiently detect objects for real-time systems.

- To design a deep learning-based approach for multi-object tracking, their implementation and performance evaluation to address the limitations of the existing methods.
- To propose some efficient TP techniques by using a multi-scale graph neural network for the uncertainty of road-agents.
- To design CARLA simulator-based motion planning for the autonomous vehicle to get collision-free and abide by the traffic rules.

1.6 Contributions to the Thesis

This section provides the essential contributions made in this thesis. The critical contributions to the thesis, development, implementation and comparative analysis of the following proposed methods for addressing the problems of submodules of the perception and planning modules of the AVs.

A Literature review on submodules of perception and planning modules of the Autonomous Vehicles: In the past two decades, much literature has been detailed on both modules of AVs. This thesis discusses issues related to these techniques, challenges and gaps. Various research problems are identified using the discussion.

Channel spatial attention based single-shot object detector for Autonomous Vehicles: considering the research issues and challenges discussed in chapter 2 (literature review of object detection), an approach for detection and localization of objects has been developed. This approach takes the advantages of attention-based channel spatial features that help for accurate and faster object detection.

An end-to-end trained hybrid model for multi-object tracking: Motivated to design an end-to-end trained DNN for MOT that scale-up two tasks (motion estimation and Re-identification) together. The proposed model utilizes a matching technique that utilizes the relative scale between the boundary boxes and relative position calculates the relative distance between the objects for MOT. The proposed model has a matching technique that poses two subtasks to efficiently scale up a single shot DNN tracking approach for a flexible number of objects in the video frames. The proposed method uses a relative scale and position to match the detected and targeted objects.

Graph Neural Network (GNN) with Recurrent Neural Networks (RNNs) based Trajectory Prediction of Dynamic Agents for autonomous vehicle: The trajectory for each dynamic object (or road agent) is described as a sequence of states within a time interval, with each state representing the object's spatial coordinates under the world coordinate frame. In the TP problem, given the trajectory of each object between intervals of time, The proposed model predicted their trajectories between these intervals of time and designed a Multi-Scale GNN with temporal features architecture for this prediction problem.

Semantic Supervision guided image-based Motion Planning of the

Autonomous Vehicles: presented an image-based approach that considered three images from different angles to understand the ego vehicle's surroundings. The proposed model decreased the occurrence of infractions and reduced the violation of traffic signals. The high-level semantic information available in the dataset is used as an input with extracted features to plan the ego vehicle's motion. The proposed model has generated a sequence of coordinates. These sequences of coordinates represent the waypoints in the predicted path of the vehicle for upcoming time steps. The PID controller has also used an inverse dynamic algorithm to derive the values of the driving parameters, like steering angle, throttle, and brake value, from the generated coordinates of the waypoints.

1.7 Thesis Organization

The overall thesis is organized into six chapters as follows

Chapter 1 briefs about the background knowledge on AVs and its importance. This chapter also discusses the vision-based architecture of AVs and describes the functionality of modules and sub-modules. Further, challenges in the existing sub-modules of AVs, motivation of the research and problem statement are also mentioned in this chapter. This chapter focuses objective of the research contribution of the thesis and concludes with the organization of the thesis.

Chapter 2 presents the theoretical background and literature review related to the identified research problem and research gaps are also identified. This chapter also presents the description of datasets, evaluation metrics and a comparative study of the proposed methods and models in the thesis.

Chapter 3 titled "Single-Stage Attention Based Object Detection" presents a method namely "Channel Spatial Attention-based single shot object detector for autonomous vehicles". This approach aims to make a simple and efficient system for object detection with two modules. The single-shot object detection is provided faster results and the attention module helps to provide more accurate detection.

Chapter 4 titled "An End-to-End method for Multi-Object Tracking" presents an object tracking method, "An end-to-end trained hybrid CNN model for Multi-object tracking." Further, the research reported in this thesis has been extended for Multi-Object Tracking. The model involves detection-based tracking, which generally requires a scale-up of two subtasks, motion estimation and re-identification. The proposed model utilized Lucas-Kanade optical flow for the motion estimation and the relative scale of boundary boxes is formulated to find the maximum likelihood of a couple of correct matches. The model repeats this for unmatched detection to match with another trajectory (trajectories not assigned in current frames). The detection that this process cannot match is initialized as a new trajectory.

Chapter 5 titled "Trajectory Prediction and Motion Planning" presents a prediction technique for the trajectory and a motion planning method named "Graph Neural Network with RNNs based Trajectory Prediction of dynamic agents for Autonomous Vehicles" and "Semantic Supervision Guided Image-based Motion Planning of the Autonomous Vehicles."

The first method is used a data preprocessing strategy for selecting the semantic road links effectively, which is more efficient for TP. The proposed model has a backbone network with a spatial-temporal graph convolutional network for the TP. The proposed model overcomes the existing issues of the state-of-the-art techniques to extract the semantic information for the TP accurately.

The second has presented a multi-view image-based approach that considers three images from different angles to understand the ego vehicle's surroundings. To decrease the occurrence of infractions by the proposed model and significantly reduce the violation of traffic signals, the high-level semantic information available in the dataset can be explicitly used combined as input for planning the motion of the ego vehicle to generate a sequence of coordinates representing the waypoints in the predicted path of the vehicle for upcoming few time-steps.

Chapter 6 concludes the research work and summarizes the main findings of the thesis. This chapter also discussed the future directions for the research.