PREFACE

Due to technological advancements, massive multimedia data is typically available in image and video formats. Currently, images and videos are used in many complex applications such as human-computer interaction, autonomous security systems, 3D scene understanding, sports performance analysis, etc. The autonomous vehicle is one application that uses images to analyze its surroundings. Autonomous Vehicles (AVs) have various components for execution. The Sensors are used to capture the data from the surroundings. The perception module converts this raw data into meaningful information. The mapping and localization module is used to localize the vehicles in real-world coordinates and destination locations through the Global Positioning System (GPS). The Planning module uses this meaningful information to plan their actions. The control module actuates the control through the steering, brake and accelerators.

This thesis focuses on the perception and planning tasks of the autonomous vehicle using computer vision and deep learning techniques. The various tasks associated with the perception module include Object Detection, Object Tracking and Trajectory Prediction. The various tasks associated with the planning module include Trajectory Planning, Motion Planning, and Behavior planning. The significant challenges associated with these modules include highly dynamic background, gradual and abrupt illumination changes, camera jitter, shadows, reflections, and weather conditions that may cause false detection and may become a big reason for

wrong decisions in the navigation of AVs. Most of the existing methods have been reported for the different modules of AVs. This thesis addresses some of the challenges and issues arising from these sub-modules of AVs. The problem statement of the research is defined as the Study of existing methods, analyzing their merits and demerits, implementation of the algorithm and proposed new methods and models for perception and planning modules using deep learning-based approaches for providing reliable navigation of the AVs.

First, this thesis presents a detailed literature survey, including a survey of various modules and sub-modules of AVs and a study on the evaluation of modern datasets for object detection, object tracking, trajectory prediction and motion planning. Further, the hierarchy of different approaches has been discussed and research gaps have been identified. Finally, a detailed list of the dataset used for training and evaluation proposed models has been presented, followed by a discussion on performance measures used in this thesis.

The first proposed model of this thesis is for object detection, which is based on the attention mechanism. High accuracy ensures the vehicle for collision-free navigation tasks, while the faster detection speed helps make decisions quickly. In this thesis, the proposed model is a single-stage object detection that provides faster detection. The channel attention mechanism provides more fine-grained features and emphasizes that 'what' is a semantic part of a given input. Apart from the channel attention mechanisms, spatial attention emphasizes 'where' is meaningful information that is working to boost the performance of the attention block for accurate

detection. The experimental result shows that the proposed model surpasses the state-of-the-art techniques for the KITTI and BDD datasets.

Further, the research reported in this thesis has been extended for Multi-Object Tracking as a second contribution. The proposed model, An end-to-end Hybrid model for Multi-Object Tracking, involves detection-based tracking, which generally requires a scale-up of two subtasks: motion estimation and re-identification. The proposed model utilized dense-optical flow for motion estimation. 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 another trajectory (trajectories not assigned in current frames). The detection that this process cannot match is initialized as a new trajectory. The achieved state-of-the-art results of the tasks allow for high accuracy of tracking with detection and surpasses existing state-of-the-art methods by a considerable margin on MOT and Waymo publicly available datasets (Multi-Object Tracking, Waymo).

Finally, in this thesis, two methods have been proposed for trajectory prediction and motion planning. A Graph Neural Network with RNNs-based Trajectory Prediction of dynamic Agents for Autonomous Vehicles is proposed for trajectory prediction. A Semantic Supervision Guided Image-based Motion Planning of the Autonomous Vehicles method is proposed for motion planning. The trajectory prediction model extracts the spatial-temporal features using a graph neural network and predicts the long-term trajectory using LSTM (Long-short term memory). Experiments show that the proposed model effectively captures comprehensive

Spatio-temporal correlations through modeling GNN with temporal features for TP and consistently surpasses the existing state-of-the-art methods on three publicly available datasets (Lyft, Argoverse, Apolloscape) for trajectory. Compared to prior methods, The proposed model performs better for sparse datasets than for dense datasets.

The motion planning task utilized multi-view images and a CNN (Convolutional Neural Network) model for feature extraction and the number of GRUs (Gated Recurrent Units) to generate the waypoints. The ego vehicle generated a sequence of coordinates representing the waypoints in the predicted path of the vehicle for the upcoming few time steps. The model has generated the waypoints using GRUs, which are used as input for the PID (Proportional Integral Derivative) controller. The PID controller is used an inverse dynamic algorithm to derive the values of the driving parameters, like steering angle, throttle, and brake value, from the coordinates of the waypoints. The proposed model has shown an improvement in the Route Completion and Driving Score metrics that outperform state-of-the-art methods on this simulator's dataset.