

# Contents

<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xv</b>
<b>Abbreviations</b>	<b>xvii</b>
<b>Symbols</b>	<b>xix</b>
<b>Preface</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Vision-Based Architecture of Autonomous Vehicles . . . . .	3
1.2.1 Sensors . . . . .	4
1.2.2 Perception System . . . . .	4
1.2.2.1 Object Detection . . . . .	4
1.2.2.2 Object Tracking . . . . .	5
1.2.2.3 Trajectory Prediction . . . . .	5
1.2.3 Mapping and Localization . . . . .	6
1.2.3.1 Localization . . . . .	6
1.2.3.2 Map State Estimation . . . . .	6
1.2.4 Planning and Decision-Making . . . . .	6
1.2.4.1 Trajectory Planning . . . . .	7
1.2.4.2 Behaviour Planning . . . . .	7
1.2.4.3 Motion planning . . . . .	7
1.2.5 Safety Module . . . . .	7
1.2.6 Control Module . . . . .	8
1.3 Motivation . . . . .	8
1.4 Problem statement . . . . .	10
1.5 Thesis Objective . . . . .	11
1.6 Contributions to the Thesis . . . . .	12
1.7 Thesis Organization . . . . .	14
<b>2 Theoretical Foundation and Literature Survey</b>	<b>17</b>

2.1	Introduction . . . . .	18
2.2	Literature Review . . . . .	18
2.2.1	Literature Review of Object Detection . . . . .	19
2.2.1.1	Pioneer Methods . . . . .	19
2.2.1.2	Deep-Learning Methods . . . . .	20
2.2.1.3	Summary . . . . .	33
2.2.2	Literature Review of Multi-Object Tracking . . . . .	35
2.2.2.1	Traditional Multi-Object Tracking . . . . .	36
2.2.2.2	Stereo-Vision Based MOT . . . . .	37
2.2.2.3	Grid-Based MOT . . . . .	38
2.2.2.4	Sensor-Fusion Based MOT . . . . .	39
2.2.2.5	Deep-Learning Based MOT . . . . .	40
2.2.2.6	Summary . . . . .	41
2.2.3	Literature Review of Trajectory Prediction . . . . .	42
2.2.3.1	Feature Encoding based TP . . . . .	42
2.2.3.2	Interaction Modeling Based TP . . . . .	44
2.2.3.3	Prediction Head Based TP . . . . .	45
2.2.3.4	Generative Model-Based TP . . . . .	46
2.2.3.5	Summary . . . . .	46
2.2.4	Literature Review of Motion Planning . . . . .	49
2.2.4.1	Traditional Algorithms . . . . .	50
2.2.4.2	Machine-Learning and Deep-Learning Based Algorithm . . . . .	52
2.2.4.3	Motion Planning using simulator . . . . .	55
2.2.4.4	Summary . . . . .	57
2.3	Research Gaps . . . . .	59
2.3.1	Object detection . . . . .	59
2.3.2	Multi-Object Tracking . . . . .	60
2.3.3	Trajectory Prediction . . . . .	61
2.3.4	Motion Planning . . . . .	62
2.4	Benchmark Datasets and Simulator used for training and evaluation . . . . .	62
2.4.1	KITTI [1] . . . . .	63
2.4.2	Berkley Driving Dataset (BDD) [2] . . . . .	63
2.4.3	Waymo [3] . . . . .	64
2.4.4	Multi Object tracking (MOT) [4] . . . . .	65
2.4.5	ARGOVERSE . . . . .	66
2.4.6	APOLLOSCAPE . . . . .	66
2.4.7	LYFT . . . . .	66
2.4.8	CARLA Dataset . . . . .	67
2.5	Evaluation Metrics . . . . .	68
2.5.1	Precision . . . . .	68
2.5.2	Multi-Object Tracking Accuracy . . . . .	68
2.5.3	ID F1 Score . . . . .	68
2.5.4	Mostly Tracked Targets . . . . .	69

2.5.5	Mostly Lost Targets . . . . .	69
2.5.6	IDs Identity Switches . . . . .	69
2.5.7	FRAG . . . . .	69
2.5.8	Final Displacement Error . . . . .	69
2.5.9	ADE Average Displacement Error . . . . .	70
2.5.10	Infraction Management . . . . .	71
2.5.11	Driving Score . . . . .	71
2.5.12	Route Completion . . . . .	72
2.6	Conclusion . . . . .	72
<b>3</b>	<b>Single-Stage Attention based Object Detection for Autonomous Vehicles</b>	<b>73</b>
3.1	Introduction . . . . .	73
3.2	Proposed Method and Model . . . . .	76
3.2.1	Channel Spatial attention based Object Detector . . . . .	77
3.2.1.1	Filter Response Normalization . . . . .	78
3.2.1.2	Attention Module . . . . .	79
3.3	Result Analysis and Discussion . . . . .	82
3.3.1	Experimental Setup . . . . .	82
3.3.2	KITTI Dataset Result . . . . .	84
3.3.3	BDD Dataset Result . . . . .	84
3.3.4	Ablation Study . . . . .	86
3.4	Conclusion . . . . .	89
<b>4</b>	<b>An end-to-end Hybrid method for Multi-Object Tracking</b>	<b>91</b>
4.1	Introduction . . . . .	91
4.2	Proposed Method and Model . . . . .	93
4.2.1	A Hybrid method for Multi-Object Tracking . . . . .	93
4.2.1.1	Motion Estimation . . . . .	93
4.2.1.2	Re-identification of objects . . . . .	94
4.3	Results and Analysis . . . . .	96
4.3.1	Experimental Setup . . . . .	96
4.3.2	Loss function . . . . .	98
4.3.3	Ablation study . . . . .	102
4.4	Conclusion . . . . .	103
<b>5</b>	<b>Trajectory Prediction and Motion Planning</b>	<b>105</b>
5.1	Introduction . . . . .	105
5.2	Proposed Methods and Models . . . . .	108
5.2.1	Graph Neural Network-based Trajectory Prediction . . . . .	108
5.2.1.1	Pre-processing of Datasets . . . . .	109
5.2.1.2	The architecture of the model and its working . . . . .	110
5.2.1.3	Generation of the Graph . . . . .	111
5.2.1.4	Spatial Sampling of Adjacency Matrix for Spatial Features . . . . .	113

5.2.1.5	Convolutional Module . . . . .	113
5.2.1.6	Temporal Features . . . . .	114
5.2.1.7	Feature Fusion Module . . . . .	114
5.2.1.8	Path Fusion . . . . .	114
5.2.1.9	Trajectory Prediction Module . . . . .	115
5.2.2	Result Analysis and Discussions for TP . . . . .	116
5.2.2.1	Experimental setup . . . . .	116
5.2.2.2	Loss Function . . . . .	117
5.2.3	Motion Planning Using CARLA Simulator . . . . .	124
5.2.3.1	Architecture of the model and its working . . . . .	124
5.2.3.2	Algorithm for PID Controller . . . . .	127
5.2.4	Result Analysis and Discussion of Motion Planning . . . . .	128
5.2.4.1	Experimental Setup . . . . .	128
5.2.4.2	Loss Function . . . . .	128
5.2.4.3	Ablation Study . . . . .	133
5.3	Conclusion . . . . .	135
<b>6</b>	<b>Conclusion and Future Work</b>	<b>137</b>
6.1	Conclusions . . . . .	137
6.2	Suggestions for Future Research . . . . .	139
6.3	Future Work . . . . .	139
	<b>References</b>	<b>141</b>
	<b>List of Publications</b>	<b>175</b>

# List of Figures

1.1	Overview of Vision-based Architecture for Autonomous Vehicles . . .	5
1.2	The road accidents, deaths and injured people over the year in India .	9
2.1	Taxonomy of the Object Detection methods . . . . .	19
2.2	Architecture of (a) Two-Stage Object Detection and (b) Single-stage Object Detection . . . . .	26
2.3	Taxonomy of Multi-Object Tracking methods . . . . .	35
2.4	Taxonomy of the Trajectory Prediction Models . . . . .	42
2.5	Texonomy of the motion planning techniques . . . . .	50
2.6	Visualization of datasets in different conditions (a)The night vision of BDD dataset (b) cloudy weather of KITTI dataset . . . . .	64
3.1	Configuration of the attention modules . . . . .	77
3.2	CSA-SS Architecture . . . . .	78
3.3	Channel attention sub-module . . . . .	80
3.4	Association of attention module with ResNet network . . . . .	82
3.5	Comparison of different attention mechanism efficiency . . . . .	88
4.1	The framework of the proposed model for multi-object tracking that defines the two parallel process of detected and tracked object and match these two of two consecutive frame t-1 and t, with relative scale	94
4.2	Effect of visibility (a) and size of the object (b) of the proposed model	101
4.3	Qualitative Results of the proposed model on Waymo and MOT datasets	102
5.1	Architecture of temporal graph model for Trajectory Prediction . . .	110
5.2	The architecture of each spatial-temporal block for Graph convolu- tional network . . . . .	115
5.3	Basic architecture of backbone network . . . . .	116
5.4	(a) GT trajectory and PT of the road-agent with agent_ID = 337 of the Apolloscape Dataset (b) GT trajectory and PT of the road-agent with agent_ID = 659 of the Apolloscape Dataset . . . . .	122
5.5	(a)GT trajectory and PT of the road-agent with agent_ID = 2 of a scene with scene_ID = 127 of the Lyft Dataset (b) GT trajectory and PT of the road-agent with agent_ID = 16 of a scene with scene_ID = 141 of the Lyft Dataset . . . . .	123

5.6	(a)GT trajectory and PT of the road-agent with agent_ID = 11 of the Argoverse Dataset (b) GT trajectory and PT of the road-agent with agent_ID = 27 of the Argoverse Dataset . . . . .	123
5.7	Architecture of the Motion Planning Model . . . . .	124
5.8	Ego vehicle turning at road curve . . . . .	129
5.9	Variation of forward speed, steer and throttle parameters with time for the entire route of the scenario two of whose instances are shown in (a). (b) and (c). The variations of the forward speed of the vehicle and the vehicle and the values of throttle, brake and steer, with time are also shown up to the considered instance . . . . .	130
5.10	Vehicle Trajectory Map for the scenario one of whose instances are shown in Figure5.8. . . . .	130
5.11	Ego vehicle overtaking another vehicle . . . . .	132
5.12	Ego vehicle stopping at a STOP road-sign . . . . .	132
5.13	Variation of forward speed, steer and throttle parameters with time for the entire route of the scenario two of whose instances are shown in (a). (b) and (c). The variations of the forward speed of the vehicle and the vehicle and the values of throttle, brake and steer, with time are also shown up to the considered instance . . . . .	132
5.14	Vehicle Trajectory Map for the scenario one of whose instances are shown in Figure5.8. . . . .	133
5.15	Comparison of the average number of different types of infractions incurred during testing for the two models . . . . .	134

# List of Tables

2.1	Comprehensive details of few existing methods for object detection using Deep-learning. . . . .	34
2.2	Comparison details of few existing methods for Multi-Object Tracking using Deep-Learning Approaches . . . . .	43
2.3	Comprehensive details of few existing methods for Trajectory Prediction using deep-learning methods . . . . .	46
2.4	Comparison details of few existing methods for Motion Planning using Deep-Learning Approaches . . . . .	57
2.5	Details of Training, Testing and validation set for KITTI, BDD, Waymo, MOT variants, Argoverse, Apolloscape, Lyft Datasets . . . .	64
3.1	Comparison of performances on KITTI validation set . . . . .	85
3.2	Comparison of performances on a different backbone architectures . .	85
3.3	Comparison of performances on different backbone architectures . . .	87
3.4	Comparison of different attention mechanism efficiency in terms of number of parameters, floating point operations per second (FLOPs) and Top-1 accuracy . . . . .	88
4.1	Performance Comparison between the proposed method and other latest track association MOT approaches on the 2DMOT15 Dataset. The last column has frames/second that measure the speed of the model	99
4.2	Performance Comparison between the proposed method and other latest track association MOT approaches on the MOT16 Dataset. The last column has frames/second that measure the speed of the model . . . . .	100
4.3	Performance Comparison between the proposed method and other latest track association MOT approaches on the MOT17 Dataset. The last column has frames/second that measure the speed of the model . . . . .	100
4.4	Performance Comparison between the proposed method and other latest track association MOT approaches on the MOT20 Dataset. The last column has frames/second that measure the speed of the model . . . . .	101
4.5	Performance Comparison between the proposed method and other latest track association MOT approaches on the Waymo dataset . . .	101

4.6	Performance of the proposed model with the different backbone network through ablation study . . . . .	102
5.1	Comparison of the metrics values obtained for Apolloscape and Lyft datasets using for the model with the other existing models . . . . .	118
5.2	Comparison of the metrics values obtained for Argoverse dataset with prediction interval = 5 and the other existing models . . . . .	119
5.3	Change in time required for one-second-ahead prediction with increasing value of N . . . . .	119
5.4	Summarizes the mean values and standard deviation of the three metrics, viz. DS, RC and IM, obtained using the model and the other image-based existing models on the CARLA simulator . . . . .	131