CERTIFICATE

It is certified that the work contained in this thesis entitled " DESIGN AND DEVELOPMENT OF DEEP LEARNING BASED APPROACHES FOR PERCEPTION AND PLANNING MODULES OF THE AUTONOMOUS VEHICLES" by "DIVYA SINGH" has been carried out under my supervision and that it has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive, Candidacy and SOTA.

Prof. RAJEEV SRIVASTAVA

Professor

Department of Computer Science and Engineering

Indian Institute of Technology

(Banaras Hindu University)

Varanasi-221005

DECLARATION BY THE CANDIDATE

I, DIVYA SINGH , certify that the work embodied in this thesis is my own bonafide work and carried out by me under the supervision of Prof. RAJEEV SRIVASTAVA from JULY-2018 to JULY-2022, at the DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully lifted up any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and included them in this thesis and cited as my own work.

Date : **Signature of the Student**

Place : Varanasi **(DIVYA SINGH)**

CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my/our knowledge.

Signature of Supervisor

(Prof. RAJEEV SRIVASTAVA)
Department of Computer Science and Engineering
Indian Institute of Technology
(Banaras Hindu University)
Varanasi-221005

Signature of Head of Department

(Prof. Sanjay Kumar Singh)

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: DESIGN AND DEVELOPMENT OF DEEP LEARNING BASED APPROACHES FOR PERCEPTION AND PLANNING MODULES OF THE AUTONOMOUS VEHICLES

Name of the Student: DIVYA SINGH

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University) Varanasi, all rights under copyright that may exist in and for the above thesis submitted for the award of the DOCTOR OF PHILOSOPHY.

Date :

(DIVYA SINGH)

Place :

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

To

My Beloved Parents

Mrs. Sangeeta Singh

and

Mr. Santosh Kumar Singh

ACKNOWLEDGEMENTS

I want to take this opportunity to express my deep sense of gratitude to all who helped me directly or indirectly during this thesis work. Firstly, I would like to thank my supervisor, Prof. Rajeev Srivastava, for being a great mentor and the best adviser I could ever have. His advice, encouragement, and critics are a source of innovative ideas, inspiration, and causes behind the successful completion of this Thesis work. The confidence shown on me by him was the most significant source of inspiration for me. It has been a privilege working with him for several years. I am highly obliged to all the faculty members of the Computer Science and Engineering Department for their support and encouragement. I express my sincere thanks to Prof. K. K. Shukla and Dr. Pratik Chattopadhyay of the department of Computer Science and Engineering and Prof. Subir Das, Department of Mathematical Sciences IIT (BHU), for providing continuous support, encouragement, and advice. I express my sincere thanks to all the Professors, Deans, office staff, supporting staff, and Ph.D. Research Scholars of Indian Institute of Technology (BHU) Varanasi India. I express my gratitude to Director, Registrars, Deans, Heads, and Student Alumni of the Indian Institute of Technology (BHU) Varanasi.

My memory of the study period at IIT (BHU) can never be complete without mentioning my fellow research scholars. Special thanks to Dr. Alok Kumar Singh Kushwaha, Dr. Vibhav Praksh Singh, Dr. Roshan singh, Dr. Pratishtha Verma, Dr. Gargi Srivastava, Dr. Ankit Jaiswal, Dr. Jani Kuntesh Ketan, Mr. Santosh Tripathy, Ms. Shweta Singh and Mr. Devendra Sharma for their great help and cooperation. I extend special thanks to the non-teaching staff in the department, particularly the late Mr. A. N. Yadav, Mr. Manoj Kumar Singh, Mr. Ravi Kumar Bharti, Mr. Prakhar Kumar, Mr. Ritesh Singh, and Mr. Shubham Pandey for their consistent support.

My in-laws, Mrs. Vimla Singh and Mr. Hawaldar Singh, who gave me the power and brain to work out on this research and their help at every level, made me see this success. I owe thanks to my dearest husband, Dr. Ashutosh Kumar Singh for his continued and unfailing love, support and understanding during my pursuit of

Ph.D. degree that made the completion of thesis possible. You were always around at times I thought that it is impossible to continue, you helped me to keep things in perspective. I greatly value his contribution and deeply appreciate his belief in me. Words are insufficient to express my profound sense of gratitude to my friends Rakesh Yadav, Dr. Rajkumar Saini. My sisters and sister-in-law Smita singh, Rashmi Singh, Varsha Singh, Shreya Singh and Kiran Singh whose encouragement gave me physical and moral strength throughout my career as well in the present research. I extend my thanks to my uncle Rahul Singh, Satya Sheel Singh, Yashwant Singh and my brother-in-law Susheel Singh, Amit singh, brothers Bhawesh Singh, Prashant Singh, Siddhant Singh, Yash Vardhan singh who are part of my inspiration. Finally, I would like to wind up by paying my heartfelt thanks and prayers to the Almighty for their unbound love and grace.

- Divya Singh

Contents

List of Figures

List of Tables

Abbreviations

AVs	**A**utonomous **V**ehicles
NHTSA	**N**ational **H**ighway **T**raffic **S**afety **A**dministration
CNN	**C**onvolutional **N**eural **N**etwork
CARLA	**C**ar **L**earning **T**o **A**ct
C2ES	**C**enter **F**or **C**limate **A**nd **E**nergy **S**olutions
WHO	**W**orld **H**ealth **O**rganization
RPN	**R**egion **P**roposal **N**etwork
RNN	**R**ecurrent **N**eural **N**etwork
LSTM	**L**ong **S**hort **T**erm **M**emory
ResNet	**R**esidual **N**etwork.
TCN	**T**emporal **C**onvolutional **N**etwork
GRU	**G**ated **R**ecurrent **U**nit
RBPF	**R**ao **B**lackwellized **P**article **F**ilter
SSPF	**S**caling **S**eries **P**article **F**ilter
LBP	**L**ocal **B**inary **P**attern
CRR	**C**orrect **R**ecognition **R**ate
SVM	**S**upport **V**ector **M**achine
FTP	**F**ourier **T**emporal **P**yramid
SGD	**S**tochastic **G**radient **D**escent
HBRNN	**H**ierarchically **B**idirectional **R**ecurrent **N**eural **N**etworks
IDT	**I**mproved **D**ense **T**rajectory
MEI	**M**otion **E**nergy **I**mage

MHI	**M**otion **H**istory **I**mage
ReLU	**R**ectified **L**inear **U**nit
FC	**F**ully **C**onnected
BN	**B**atch **N**ormalization
NF	**N**umber of **F**eature **M**aps
DMMs	**D**epth **M**otion **M**aps
PID	**P**roportional **I**ntegral **D**erivative
DARPA	**D**efense **A**dvanced **R**esearch **P**rojects **A**gency
UBER	**U**nified **B**est **E**conomical **R**ide
LIDAR	**L**ight **D**etection **A**nd **R**anging
RADAR	**R**adio **D**etection **A**nd **R**anging
GPS	**G**lobal **P**ositioning **S**ystem
DARPA	**D**efense **A**dvanced **R**esearch **P**rojects **Agency**
DNN	**D**eep **N**eural **N**etwork
GNN	**G**raph **N**eural **N**etwork
FPN	**F**eature **P**yramid **N**etwork
YOLO	**Y**ou **O**nly **L**ook **O**nce
NMS	**N**on **M**axima **s**uppression
SSD	**S**ingle **S**hot **M**ultiBox **D**etector
DPM	**D**eformable **P**arts **M**odel
BERT	**B**idirectional **E**ncoder **R**epresentation **F**rom **Transformer**
NLP	**N**atural **L**angauge **P**rocessing
MSA	**M**ulti **H**eaded **S**elf **A**ttention
DATMO	**D**etector **A**nd **T**racker **O**f **M**oving **O**bstacles
RANSAC	**R**ANdom **S**ample **C**onsensus
GOTURN	**G**eneric **O**bject **T**racking **U**sing **R**egression **N**etworks

Symbols

t_f	Last Time Step
P_j	Infraction Penality coefficient for every j infraction instance
R_i	Percentage of the i^{th} route completed
A_c	Channel Attention Feature Map
A_s	Spatial Attention Feature Map
$\bigotimes$	Element-wise Multiplication
σ	Sigmoid function
δ	Spatial Context
$\mathbf{I}$	Identity Matrix
$\mathbf{A}$	Adjacency Matrix
μ	Threshold Parameter
$\mathbf{d}$	Euclidean distance
$\theta_r^{'}$	Chebyshev Coefficient
ξ_i^t	Encoder Vector
p^t	Positional Encoding Vector
$M_{x,y}$	Heatmap at location (x,y)
λ	Eigen Value
$\delta(M_{i,j})$	Local Spatial Context
O_t^*	Motion Vector
$\|V\|$	Number of Vertices
$\|E\|$	Number of Edges

$\mathbb{R}$	Real Number
L_p	Laplacian of Graph
ω	Waypoint
γ	Desired Speed
α	Steering angle

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.