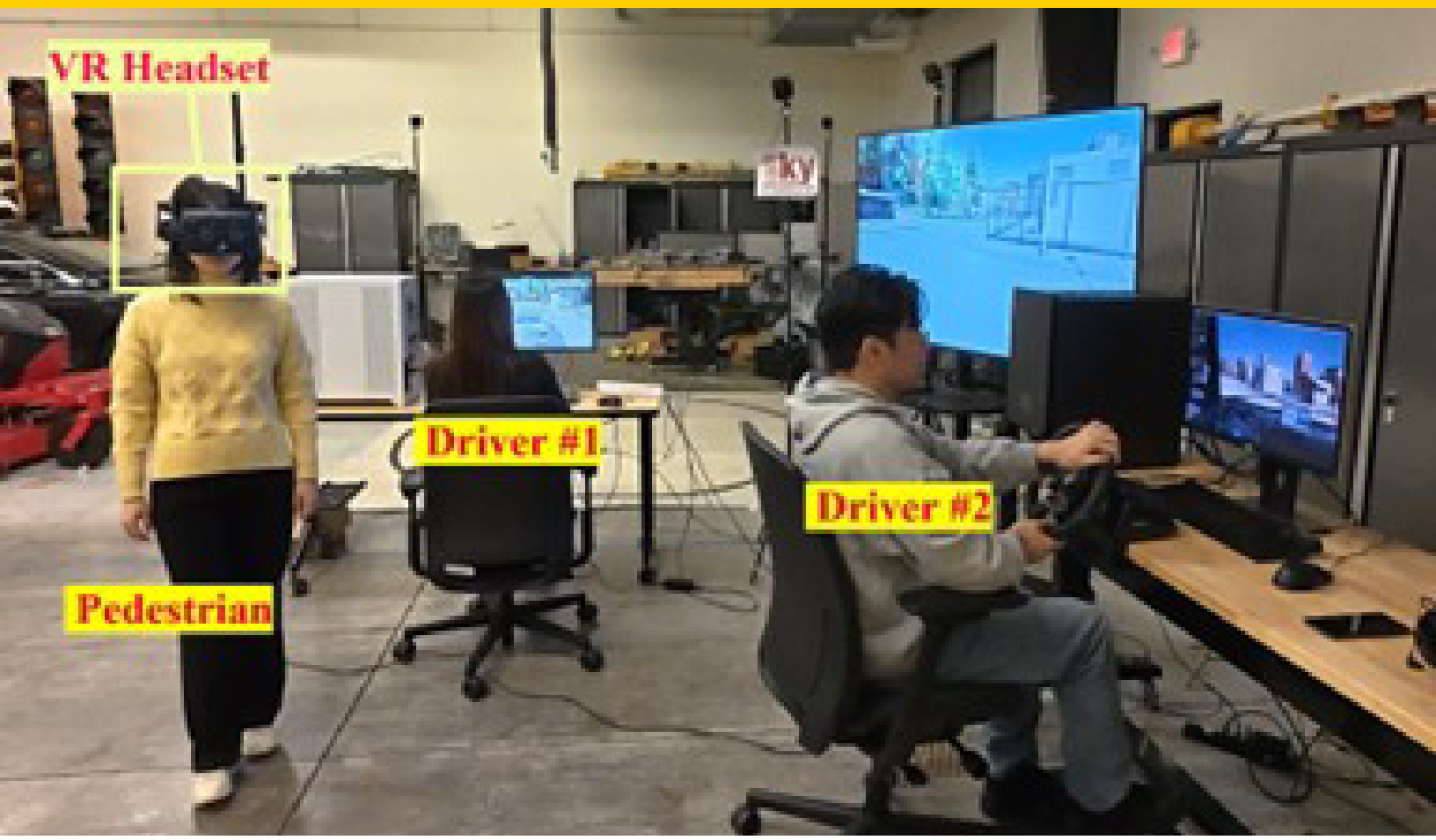
CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Final Report GR000034956
February 2026



Using Connected Intelligent Transportation to Enhance Vulnerable Road User Safety

Zilin Huang
Zhengyang Wan
Zihao Sheng
Sue Ahn
David A. Noyce
Sikai Chen

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Report No. GR000034956
Project Start Date: 01/01/2024
Project End Date: 12/31/2025

February 2026

Using Connected Intelligent Transportation to Enhance Vulnerable Road User Safety

by

Zilin Huang
Graduate Researcher

Zhengyang Wan
Graduate Researcher

Zihao Sheng
Graduate Researcher

Sue Ahn
Professor

David A. Noyce
Arthur F. Hawnn Professor

Sikai Chen
Assistant Professor

UMTRI | The University of Akron | Central State University | Illinois Center for Transportation | University of Minnesota | Northwestern | Purdue University | Washtenaw Community College | Wisconsin University of Wisconsin-Madison

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

DISCLAIMER

Funding for this research was provided by the Center for Connected and Automated Transportation under Grant No. 69A3552348305 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Suggested APA Format Citation:
Huang, Z., Wan, Z., Sheng, Z., Ahn, S., Noyce, D.A., Chen, S. (2025). Using Connected Intelligent Transportation to Enhance Vulnerable Road User Safety. CCAT Report # XGR000034956. The Center for Connected and Automated Transportation, The University of Wisconsin-Madison, Madison, WI.

Contacts

For more information:

Sikai Chen, Ph.D.
2266, Engineering Hall
1415 Engineering Dr
Madison, WI, 53706
Phone: (213) 806-0141
Email: sikai.chen@wisc.edu

CCAT
University of Michigan Transportation Research Institute
2901 Baxter Road
Ann Arbor, MI 48152
umtri-ccat@umich.edu
(734) 763-2498

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Technical Report Documentation Page

1. Report No. GR000034956	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Using Connected Intelligent Transportation to Enhance Vulnerable Road User Safety		**5. Report Date** December 2025 **6. Performing Organization Code** N/A
7. Author(s) Zilin Huang, Zhengyang Wan, Zihao Sheng, Sue Ahn, David A. Noyce, Sikai Chen		**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address Department of Civil and Environmental Engineering, University of Wisconsin-Madison Engineering Hall, 1415 Engineering Dr, Madison, WI 53706		**10. Work Unit No.** **11. Contract or Grant No.** Contract No. 69A3552348305
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE Washington, DC 20590		**13. Type of Report and Period Covered** Final Report, Jan 2024-Dec 2025 **14. Sponsoring Agency Code** OST-R

15. Supplementary Notes
Conducted under the U.S. DOT Office of the Assistant Secretary for Research and Technology's (OST-R) University Transportation Centers (UTC) program.

16. Abstract
This project leverages connected intelligent transportation technologies to enhance the safety of vulnerable road users (VRUs) within evolving urban road systems. This project targets three fundamental barriers: the lack of high-fidelity experimental platforms, localization failures in GNSS-denied areas, and insufficient trajectory prediction for heterogeneous agents. Our goal is to develop an integrated cooperative system that combines virtual reality, wireless communication, and physics-informed learning. To achieve this, we develop Sky-Drive, a distributed multi-agent simulation platform that enables human-in-the-loop interaction testing. Furthermore, the project implements a cellular Vehicle-to-Everything (C-V2X)-based cooperative localization framework to ensure lane-level accuracy and a kinematics-aware multigraph attention network for precise motion forecasting. The output of this project provides a comprehensive technological foundation for enabling the safe and equitable coexistence of autonomous vehicles (AVs) and VRUs in mixed traffic environments.

17. Key Words Connected and Automated Vehicles, Vulnerable Road Users, Traffic Safety, Real-world Test	**18. Distribution Statement** No restrictions.		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 72	**22. Price**

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLES

LIST OF ACRONYMS

VRUs	Vulnerable Road Users
AVs	Autonomous Vehicles
RSUs	Roadside Units
GNSS	Global Navigation Satellite System
IMUs	Inertial Measurement Units
RTK	Real-Time Kinematic
VR	Virtual Reality
C-V2X	Cellular Vehicle-to-Everything
CV2X-LOCA	C-V2X-Based Cooperative Localization Framework
KA-MGAT	Kinematics-Aware Multigraph Attention Network
MEMS	Micro-Electro-Mechanical Systems
EKF	Extended Kalman Filters
UKF	Unscented Kalman Filters
PDR	Pedestrian Dead Reckoning
SLAM	Simultaneous Localization and Mapping
V2V	Vehicle-to-Vehicle
V2I	Vehicle-to-Infrastructure
CSI	Channel State Information
RSSI	Received Signal Strength Indicator
RNNs	Recurrent Neural Networks
LSTM	Long Short-Term Memory
HVs	Human-Driven Vehicle
ASR	Automatic Speech Recognition
LAN	Local Area Network
VLAN	Virtual Local Area Network
RSSI	Received Signal Strength Indicator
GDOP	Geometric Dilution of Precision

SDP	Semi-Definite Programming
SDR	Semi-Definite Relaxation
C-RSUs	Cooperative RSUs
ALE	Average Localization Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
ADE	Average Displacement Error
FDE	Final Displacement Error
GRU	Gated Recurrent Unit
WSADE	Weighted Sum of ADE
WSFDE	Weighted Sum of FDE

CHAPTER 1 INTRODUCTION

1.1 Study Background

1.1.1 Urban Road Ecosystems

Urban transportation systems are complex and constantly changing, involving continuous interactions among various types of road users. These systems include not only motor vehicles but also vulnerable road users (VRUs), such as pedestrians, cyclists, and disabled people, as illustrated in Fig. 1.1. The safety and efficiency of urban transportation systems largely depend on the ability of all participants to express their intentions and predict the behaviors of others. Traditionally, such communication has been achieved through informal and implicit cues, including eye contact, facial expressions, hand gestures, and other visual signals, which have developed through long-term shared use of road environments.

Fig 1.1 Different vulnerable road users

However, the urban transportation system is experiencing a major shift with the development of autonomous vehicle (AV) technologies. While this technological advancement is expected to improve safety, efficiency, and accessibility, it also raises new concerns regarding the safety of VRUs. As automation levels increase and vehicles transition from human-driven to fully autonomous modes, the traditional human-to-human communication cues that VRUs rely on for safe movement become less effective or even unavailable. This situation highlights the urgent need to develop new communication methods and interaction systems to protect VRUs in mixed traffic settings, where human-driven vehicles (HVs), AVs with different levels of automation, and VRUs operate together.

1.1.2 Vulnerable Road Users

VRUs are a major concern in transportation safety. According to traffic accident statistics, VRUs account for 46% of all road traffic deaths globally (WHO, 2023), as shown in Fig. 1.2. In 2021, approximately 7,388 pedestrians were killed in traffic crashes in the United States (a 13% increase from 2020), and more than 60,000 pedestrians were injured nationwide (NHTSA, 2021). Similarly, bicyclist fatalities increased by 9.2% compared to 2019. The European Union reports even more concerning statistics, with VRUs-related collisions accounting for 46% of road traffic (European Commission, 2022). Projections suggest that by 2030, the number of VRUs accidents may approach the level of automobile accidents (Y. Ma et al., 2019), highlighting the critical importance of addressing VRUs safety in transportation system planning and design.

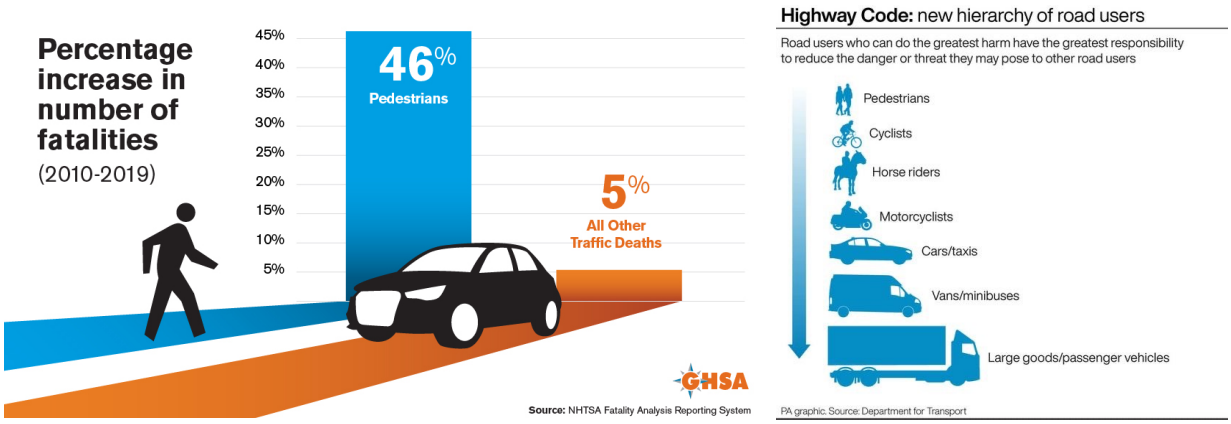


Fig 1.2 VRUs suffered seriously from traffic incidents

The vulnerability of these road users arises from several factors. First, VRUs do not have the physical protection provided by vehicle structures, which makes them more likely to suffer serious injuries or fatalities during collisions. Second, the behavior of VRUs is often less predictable than that of vehicles, since their movements can be affected by age, physical ability, distraction, and surrounding environmental conditions. Third, VRUs may have limited visibility in certain traffic situations, especially in dense urban areas with complex layouts and multiple conflict points. In addition, specific groups within the VRU population, such as children, older adults, and individuals with visual or hearing impairments, face further challenges related to perception, reaction speed, and mobility, which increase their overall risk.

With the emergence of AVs, these vulnerabilities become more evident. Traditional communication cues used by VRUs, such as eye contact with drivers to ensure they are noticed, interpreting drivers' facial expressions to understand intent, or using hand gestures to indicate crossing behavior, are no longer reliable when no human driver is present. This lack of direct communication creates a significant safety challenge that must be addressed to ensure that advances in vehicle automation do not undermine the safety and mobility of VRUs.

1.1.3 Autonomous Vehicles and Levels of Automation

The Society of Automotive Engineers (SAE) has established a widely adopted classification standard for vehicle automation, defining six levels ranging from Level 0 (no automation) to Level 5 (full automation) (SAE International, 2021). At Level 0, all driving tasks are performed by the human driver, with no automated support. Level 1 includes basic driver assistance functions, such as adaptive cruise

control or lane-keeping assistance, while the human driver remains responsible for the majority of driving activities. At Level 2, the vehicle is capable of controlling both steering and acceleration or braking at the same time under certain conditions, but the driver must continuously supervise the vehicle and the surrounding environment.

The transition to higher automation levels represents a qualitative shift in the driving paradigm. At Level 3 (conditional automation), the automated system can manage all driving tasks within specific scenarios, but the human driver must be ready to take over control when prompted. Level 4 vehicles are able to operate fully autonomously within predefined operational design domains, such as limited geographic regions or specific weather conditions, without requiring human intervention in those contexts. Level 5 represents complete automation, in which the vehicle can drive autonomously in all environments and situations without any involvement from a human driver (González et al., 2016).

In this research, we place particular emphasis on the challenges associated with highly automated vehicles, specifically those at Levels 4 and 5, which are collectively referred to as AVs. These vehicles function without continuous human oversight and do not include a human driver who can communicate directly with VRUs. As a result, the nature of interactions between VRUs and vehicles is fundamentally changed, since the informal and human-centered communication cues that VRUs have traditionally depended on are no longer present.

1.1.4 The Confluence of VRUs and Autonomous Vehicles

The interaction between VRUs and AVs in shared road environments introduces a range of complex technological and methodological challenges that must be resolved to achieve safe urban transportation systems (as shown in Fig. 1.3). Although AVs are equipped with advanced sensors and artificial intelligence (AI) techniques to perceive and interpret their surroundings, important questions remain regarding how to accurately test these systems, localize different agents, and predict the behaviors of both human participants and other AVs in mixed traffic conditions.

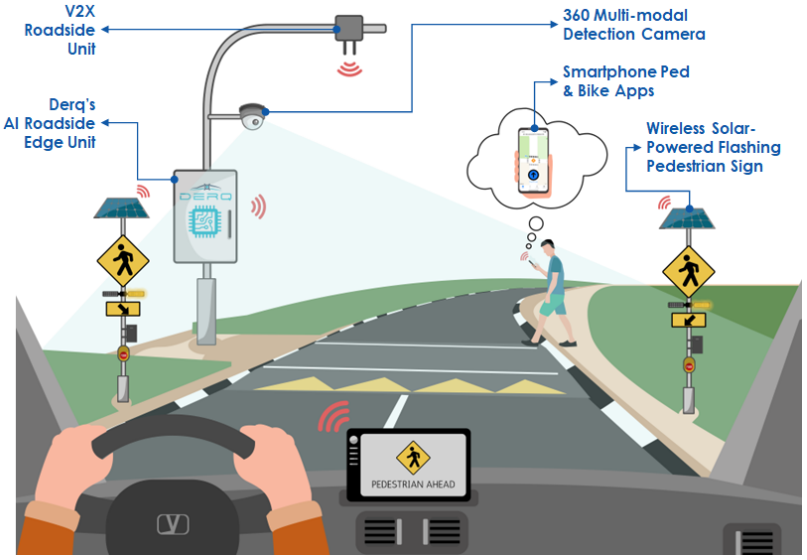


Fig 1.3 The interaction between AVs and VRUs in transportation system

From a research methodology perspective, studying interactions between VRUs and AVs involves

several unique difficulties. Conventional field tests are not able to safely reproduce rare yet critical events that are essential for assessing safety, such as emergency braking, unclear right-of-way situations, or near-collision encounters. Similarly, laboratory studies based on traditional driving simulators often lack sufficient immersion to generate realistic and natural behavior from human participants. In addition, modern traffic scenarios are becoming increasingly complex, involving multiple vehicles with different levels of automation, various types of VRUs with diverse abilities and limitations, and infrastructure enhanced with sensing and communication technologies. These factors require experimental platforms that can coordinate the actions of multiple agents across distributed systems while preserving a shared, high-fidelity virtual environment.

From the perspective of perception and localization, protecting VRUs requires accurate, real-time information about the positions of all traffic participants. Although vehicles can be equipped with advanced localization technologies such as global navigation satellite systems (GNSS), inertial measurement units (IMUs), and high-definition maps, VRUs usually depend on smartphones or other consumer-grade devices that provide much lower positioning accuracy. In addition, VRUs–AVs conflicts commonly take place in dense urban environments, such as city centers with tall buildings, where GNSS signals are often unreliable or unavailable. As a result, a critical technological challenge emerges: how to achieve lane-level positioning accuracy (approximately ±1.5 m or better) required for safety-critical applications when traditional GNSS-based localization approaches are insufficient.

From a prediction and decision-making perspective, AVs must be able to accurately anticipate the future motions of all nearby agents to support proactive collision avoidance and efficient traffic flow. However, the diversity of agents in real-world traffic environments makes trajectory prediction particularly challenging. Vehicle movements are generally more regular and constrained by road structure and physical limits, whereas pedestrian behavior is more variable and influenced by social interactions, individual intent, and surrounding context. Cyclists often display behaviors that fall between these two extremes. Many existing prediction approaches either model all agents in a uniform manner or focus on only a single type of road user, which limits their ability to capture the complex interaction patterns. In addition, purely data-driven methods may generate predictions that violate physical constraints or perform poorly in unseen situations, while purely physics-based models may fail to represent the nuanced and adaptive nature of human behavior.

Overall, ensuring safe VRUs–AVs interactions in future mixed urban transportation systems requires not only suitable road geometry and traffic control strategies but also the support of enabling technologies, including wireless communication, cooperative sensing, and shared perception of the environment.

1.2 Problem Statement

The fundamental challenges addressed by this research project center on three critical barriers to advancing VRUs safety in the era of AVs. These challenges involve experimental platform, localization accuracy, and motion prediction capabilities, all of which are essential for enabling safe VRUs-AVs interactions.

- **Challenge 1: Lack of High-Fidelity Experimental Platforms for VRUs-AVs Interaction Research**

One of the major barriers to advancing VRUs safety is the lack of high-fidelity experimental platforms that can realistically capture VRUs-AVs interactions without exposing participants to real-world danger. Field experiments involving rare but safety-critical situations, such as sudden pedestrian crossings, unclear right-of-way negotiations, or blind-spot conflicts, are difficult to conduct safely and are hard to reproduce at scale. Although traditional driving simulators provide a safe alternative, they often lack sufficient immersion and behavioral realism to elicit natural responses from VRUs. In addition, many existing simulation platforms do not support multi-user participation across distributed terminals, which limits their ability to model the complex, real-time interactions among heterogeneous agents, including AVs, HVs, pedestrians, and cyclists.

- To overcome these limitations, recent research has explored virtual reality (VR)-based simulation platforms that enable safe, repeatable, and data-rich investigations of human–vehicle interactions. However, current solutions still face several shortcomings. Many simulators lack key capabilities, such as (1) cross-terminal synchronization that allows multiple agents to be independently controlled within a shared virtual environment, (2) multi-modal sensor integration for collecting rich human behavioral data, including gaze behavior, physiological signals, and control inputs, (3) high-fidelity digital-twin representations of real-world traffic environments that accurately capture infrastructure and traffic dynamics. **Challenge 2: Limited Localization Capabilities in GNSS-Denied Environments**

Accurate localization and motion prediction are fundamental components of urban intelligent transportation systems. Lane-level positioning is especially important for enabling coordination among vehicles, infrastructure, and other road users, and it plays a critical role in VRU safety applications such as collision warnings and adaptive vehicle responses. However, traditional GNSS-based localization methods often suffer from severe accuracy degradation or complete signal loss in environments such as urban canyons, tunnels, areas beneath elevated roadways, and during adverse weather conditions. These are also the locations where conflicts between VRUs and vehicles are most likely to occur.

While vision-based and LiDAR-based localization techniques can supplement GNSS information, they introduce additional challenges, including occlusions caused by surrounding vehicles or infrastructure, limited sensing range, high computational costs, and sensitivity to lighting and weather conditions. Moreover, requiring VRUs to carry advanced sensing hardware is neither cost-effective nor practical in terms of user acceptance. As a result, there is a strong need for alternative localization solutions that can utilize existing or easily deployable infrastructure and communication technologies to provide lane-level positioning accuracy for VRUs without relying on expensive onboard sensors.

- **Challenge 3: Insufficient Trajectory Prediction Capabilities for Heterogeneous Traffic Agents**

Accurate prediction of vehicle and pedestrian trajectories is vital for anticipatory safety maneuvers and collision avoidance in mixed traffic environments. AVs must be able to anticipate not only the motions of other vehicles but also the more variable and less predictable behaviors of pedestrians, cyclists, and other VRUs. However, conventional trajectory prediction methods face several key limitations. Physics-based models are generally interpretable and computationally efficient, but they often depend on simplified assumptions that fail to represent the complex behaviors observed in real traffic. In contrast, purely learning-based approaches can achieve strong predictive performance, yet they may generate physically implausible results, require large-scale training data, generalize poorly to unseen scenarios, and offer limited interpretability, which raises concerns regarding safety and trustworthiness in real-world deployment.

In addition, many current methods do not sufficiently address several key factors: (1) the inherent heterogeneity of traffic agents with distinct motion characteristics and constraints, such as vehicles governed by kinematic limits versus pedestrians with greater behavioral flexibility; (2) the complex interaction dynamics among multiple agents that jointly shape future trajectories; and (3) the effective combination of prior physical knowledge with data-driven behavioral learning.

Addressing the above challenges requires a connected and cooperative system in which vehicles, roadside units (RSUs), and VRUs collaboratively build a shared understanding of the traffic environment. This can be enabled through technologies such as human-in-the-loop integration, Cellular Vehicle-to-Everything (C-V2X)-based communication localization, and graph-based trajectory prediction. In response, this research project targets the three challenges in an integrated manner by developing: (1) a distributed, multi-agent VR simulation platform that supports realistic human-in-the-loop evaluation of VRUs–AVs interactions; (2) a cooperative localization framework based on C-V2X that achieves lane-level positioning accuracy in GNSS-denied environments using only wireless signal measurements; and (3) a kinematics-aware trajectory prediction approach that combines the interpretability of physics-based models with the representational power of deep learning. Through the integration of these components, the project seeks to establish a unified technological foundation for improving VRU safety in connected and automated transportation systems.

1.3 Study Objectives

This research project was designed to address the critical challenges of VRUs-AVs interactions through the development of an integrated system combining VR simulation, connected vehicle technologies, and advanced AI algorithms. The specific objectives of this project were:

Objective 1: Develop a VR-based VRUs-AVs Interaction Experimental Platforms

Establish a comprehensive VR-based experimental platforms that enables realistic, immersive, and safe testing of VRUs-AVs interactions. This platform integrates:

- A distributed multi-agent simulation architecture that supports synchronized simulation across multiple terminals, allowing AVs, HVs, and VRUs to be controlled independently while maintaining a shared virtual environment
- Multi-modal human-in-the-loop capabilities that capture comprehensive behavioral data through integrated sensor systems including VR headsets with eye-tracking, steering wheels, cameras, and wearable devices
- A digital twin framework that creates high-fidelity virtual replicas of real-world transportation environments through integration of traffic simulation (SUMO) with 3D visualization (Unreal Engine)

Objective 2: Develop and Validate C-V2X-Based Localization for VRUs

Design and evaluate a C-V2X based cooperative localization framework that enables lane-level positioning accuracy for VRUs in GNSS-denied environments. This framework:

- Utilizes only C-V2X channel state information (received signal strength) without requiring expensive on-board sensors
- Implements novel RSUs-based cooperative positioning algorithms that can achieve accuracy

suitable for collision warning applications
- Incorporates environment parameter correction mechanisms to adapt to different road environments
- Provides real-time localization capabilities suitable for integration into VRUs warning systems

Objective 3: Develop Physics-Informed Trajectory Prediction for Heterogeneous Agents

Create an advanced trajectory prediction system that can accurately forecast future trajectories for heterogeneous traffic agents including vehicles, pedestrians, and cyclists. This framework:

- Integrates kinematic constraints explicitly into deep learning frameworks to ensure physically plausible predictions
- Employs multigraph attention mechanisms to model diverse interaction patterns among heterogeneous traffic agents
- Incorporates residual learning to refine predictions and address limitations of simplified kinematic assumptions
- Achieves state-of-the-art prediction accuracy while maintaining improved learning efficiency compared to purely data-driven approaches

Objective 4: Integrate Components into a Unified VRUs Safety System

Synthesize the developed technologies into an integrated system that demonstrates the feasibility of connected intelligent transportation approaches for VRUs safety. This integration:

- Combines VR simulation, C-V2X localization, and trajectory prediction into a cohesive system
- Demonstrates proof-of-concept for VRUs warning systems and AVs collision avoidance capabilities
- Generates data and insights that can inform future VRUs safety policies and standards

1.4 Organization of this Report

This final report documents the research activities, methodologies, findings, and outcomes of the project. The report is organized as follows: Chapter 1 provides an introduction to the research, including the study background, problem statement, research objectives, and organization of the report. Chapter 2 presents a comprehensive literature review. Chapter 3 describes the development of the Sky-Drive distributed multi-agent simulation platform. This chapter details the platform's architecture, including the VR-based human-in-the-loop framework, the distributed multi-terminal simulation capability, and the digital twin environment construction. Chapter 4 presents the C-V2X-based cooperative localization framework (CV2X-LOCA) for achieving lane-level positioning accuracy in GNSS-denied environments. Chapter 5 introduces the kinematics-aware multigraph attention network (KA-MGAT) for heterogeneous trajectory prediction. Chapter 6 summarizes project performance. Chapter 7 reports study outcomes and outputs. Finally, Chapter 8 concludes the report by summarizing key findings.

Throughout this report, we demonstrate how the integration of VR simulation, connected vehicle technologies, and physics-informed machine learning can create new opportunities for enhancing VRUs safety in the context of AVs. The findings presented here provide actionable insights for transportation agencies, vehicle manufacturers, and policymakers working to ensure that the benefits of vehicle automation are realized equitably across all road user populations.

CHAPTER 2 LITERATURE REVIEW

2.1 Multi-Agent Simulation Platforms for VRUs-AVs Interaction Research

2.1.1 Evolution of Simulation Approaches for Transportation Research

The study of human-vehicle interactions has evolved through several methodological paradigms, each with distinct advantages and limitations. Traditional field studies, while providing authentic behavioral data, face inherent challenges in safety, cost, and experimental control (Rasouli & Tsotsos, 2020). Closed-track testing facilities offer controlled environments but require substantial infrastructure investment and cannot easily replicate the complexity of real urban scenarios (Khoury & Kamat, 2009). These limitations have driven researchers toward simulation-based approaches that can balance realism, safety, and experimental flexibility. Early driving simulators focused primarily on driver behavior research, employing simplified visual displays and limited interactivity (Bella, 2008). As computational capabilities advanced, high-fidelity driving simulators emerged, incorporating motion platforms, wide field-of-view displays, and realistic vehicle dynamics (de Winter et al., 2009). However, these simulators were mainly designed to study driver behavior in HVs. As a result, they often fail to represent the unique operational characteristics of AVs or to capture the perspectives and behaviors of VRUs.

2.1.2 Virtual Reality for Transportation Research

The advent of VR technologies has created new opportunities for transportation research (Deb, Carruth, et al., 2017; Nezami et al., 2021). Recent studies have demonstrated VR's effectiveness for studying pedestrian behavior at crosswalks (Deb, Carruth, et al., 2017; Deb, Strawderman, et al., 2017), evaluating external human-machine interfaces for AVs (Colley et al., 2020; Holländer et al., 2019), investigating pedestrian trust in AVs, and analyzing gap acceptance behavior and crossing decisions (Feldstein et al., 2016; Schwebel et al., 2008). These applications have validated VR as a viable platform for eliciting authentic behavioral responses from participants in controlled, safe environments. However, existing VR-based platforms exhibit several limitations. First, most platforms support only single-user experiences, limiting their ability to capture multi-agent interactions that characterize real traffic environments (Deb, Carruth, et al., 2017). Second, existing systems typically lack integration with traffic simulation engines, preventing realistic representation of dynamic traffic scenarios (Holländer et al., 2019). Third, current platforms rarely incorporate bidirectional human-AI interaction mechanisms that would enable human participants to influence AV decision-making in real-time (Jayaraman et al., 2018). Fourth, most VR studies focus exclusively on pedestrian perspectives without capturing simultaneous data from both VRUs and vehicle perspectives in the same scenario (Colley et al., 2020).

2.1.3 Multi-Agent and Distributed Simulation Architectures

Traffic simulation platforms provide detailed modeling of individual vehicle movements and interactions, making them valuable tools for transportation planning and analysis, such as SUMO (Lopez et al., 2018) and VISSIM (PTV Group, 2020). While these platforms excel at modeling vehicular traffic, they typically employ simplified representations of VRUs behavior based on predetermined motion models (Rudenko et al., 2020). This limitation motivates the integration of traffic simulation with human-in-the-loop VR systems. Moreover, existing platforms typically lack distributed multi-agent capabilities

that support independent human control across separate terminals while maintaining shared environment.

2.2 Localization Technologies for Vulnerable Road Users

2.2.1 Positioning Requirements for VRU Safety Applications

For VRUs safety applications such as collision warning systems, crossing assistance, and adaptive vehicle behavior, positioning accuracy requirements are particularly stringent. Research indicates that lane-level positioning accuracy, typically defined as ±1.5 meters or better, is necessary for reliable collision risk assessment (Alam et al., 2013; J. Liu et al., 2020). Furthermore, VRUs localization systems must address several practical constraints. Solutions must be economically feasible for widespread deployment without requiring expensive onboard sensors (Kuutti et al., 2021). Systems should work reliably across diverse environmental conditions including urban canyons, tunnels, and adverse weather (Alam et al., 2013). Energy efficiency is critical for battery-powered devices carried by VRUs (Kamal et al., 2021). Privacy preservation is essential given the sensitivity of location data (Emara et al., 2015).

2.2.2 GNSS-Based Positioning Systems

Global Navigation Satellite Systems (GNSS) provide global coverage for outdoor positioning applications. Standard GNSS receivers in smartphones typically achieve positioning accuracy of 5-10 meters under open-sky conditions (Zandbergen & Barbeau, 2011). Various enhancement techniques have been developed to improve GNSS accuracy, such as Differential GNSS (DGNSS) (Kaplan & Hegarty, 2017), Real-Time Kinematic (RTK) (Takasu & Yasuda, 2009), and multi-constellation receivers (X. Li et al., 2015). Despite these advances, GNSS-based positioning faces fundamental limitations in urban environments. Signal blockage by buildings creates urban canyon effects that significantly degrade positioning accuracy (Groves, 2011). Multipath propagation occurs when satellite signals reflect off buildings before reaching the receiver (De Angelis et al., 2013). Non-line-of-sight (NLOS) reception happens when direct signals are blocked and only reflected signals are received (Hsu et al., 2015). Complete signal loss occurs in tunnels, underpasses, and under dense tree canopy (Groves, 2011).

2.2.3 Inertial Navigation and Sensor Fusion

Inertial Navigation Systems (INS) use accelerometers and gyroscopes to track position through dead reckoning, providing continuous positioning independent of external signals. Yet, it suffers from unbounded error growth due to sensor drift (Woodman, 2007). Consumer-grade MEMS (Micro-Electro-Mechanical Systems) inertial sensors in smartphones exhibit significant drift, with positioning errors accumulating to tens of meters within minutes (Niu et al., 2014). To address these limitations, sensor fusion approaches have been developed that integrate GNSS and INS through filtering techniques such as Extended Kalman Filters (EKF) or Unscented Kalman Filters (UKF) (Grewal et al., 2007). Pedestrian Dead Reckoning (PDR) represents a specialized application of inertial navigation for VRUs, estimating position through step detection and heading estimation using smartphone sensors (Harle, 2013). Advanced PDR algorithms incorporate machine learning for step detection and length estimation (H. Wang et al., 2012), map matching to constrain estimates to walkable areas (Zampella et al., 2013), and particle filters for improved robustness (T. Li et al., 2018). However, PDR accuracy degrades over time and distance, requiring periodic correction from absolute positioning sources.

2.2.4 Vision-Based Localization

Visual odometry estimates camera motion by tracking features across image sequences (Scaramuzza & Fraundorfer, 2011). Simultaneous Localization and Mapping (SLAM) builds maps while simultaneously localizing within them (Cadena et al., 2016). Visual place recognition matches camera views against geotagged image databases to determine location. Recent advances in deep learning have enabled monocular depth estimation (Eigen et al., 2014), semantic segmentation for scene understanding (L.-C. Chen et al., 2018), and end-to-end learned visual localization (Kendall & Cipolla, 2017). However, vision-based approaches face significant challenges including computational requirements that strain battery-powered mobile devices (Lategahn et al., 2013), sensitivity to lighting conditions and weather (Kendall & Cipolla, 2017), privacy concerns related to continuous camera operation, and limited applicability in visually non-distinctive (R. Zhao et al., 2025). For VRU applications specifically, the requirement to hold a smartphone in a consistent orientation and the inability to function when the device is in a pocket or bag limit practical applicability (Zampella et al., 2013).

2.2.5 C-V2X and Cooperative Positioning

C-V2X supports direct device-to-device communication for vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-pedestrian (V2P) connectivity with low latency, as well as network-based communication for broader connectivity (Molina-Masegosa & Gozalvez, 2017; Naik et al., 2019). For positioning applications, C-V2X offers several advantages including dedicated spectrum (5.9 GHz ITS band) for reliable communication, high penetration rate expected in future vehicle fleets, support for safety-critical latency requirements, and integration with cellular network infrastructure (S. Chen et al., 2017). C-V2X positioning can leverage channel state information (CSI) including received signal strength indicator (RSSI), time of arrival (ToA), angle of arrival (AoA), and Doppler shift (Wymeersch et al., 2017). Recent research has demonstrated that cooperative approaches can achieve decimeter-level accuracy even when individual agents have degraded positioning (Zampella et al., 2013; Zhou et al., 2024).

2.3 Trajectory Prediction for Heterogeneous Traffic Agents

2.3.1 Importance of Trajectory Prediction in VRU Safety

For AVs, trajectory prediction enables anticipatory decision-making that can maintain safety margins, optimize traffic flow, and provide comfortable passenger experiences (Lefèvre et al., 2014). For VRU safety applications specifically, early and accurate prediction of VRU movements allows AVs to adjust speed, trajectory, or yielding behavior before conflicts develop (Rasouli & Tsotsos, 2020). The prediction time horizon significantly impacts system performance. Shorter horizons (1-2 seconds) are sufficient for immediate collision avoidance but limit the smoothness of vehicle responses (Schöller et al., 2020). Longer horizons (3-5 seconds) enable more comfortable and efficient maneuvers but face increasing uncertainty (Rudenko et al., 2020). The optimal horizon depends on vehicle speed, traffic density, and specific application requirements. Prediction accuracy requirements vary by application. Safety-critical applications demand high reliability with low false-negative rates to avoid missing dangerous situations (Rasouli & Tsotsos, 2020). Comfort-oriented applications can tolerate more prediction uncertainty but benefit from probabilistic outputs that quantify confidence (Mozaffari et al., 2022).

2.3.2 Physics-Based Prediction Methods

For vehicles, common models include the Constant Velocity (CV) model (Schubert et al., 2008), Constant Acceleration (CA) model (X. R. Li & Jilkov, 2003), Constant Turn Rate and Velocity (CTRV) model (Schubert et al., 2008), and bicycle models (Rajamani, 2006). For pedestrians, the Social Force Model represents movement as a combination of attractive forces toward goals and repulsive forces from obstacles and other pedestrians (Helbing & Molnar, 1995). Variations include the Extended Social Force Model (Moussaïd et al., 2011) and the Pedestrian Interaction Force Model (Karamouzas et al., 2014). Physics-based models offer several advantages including interpretability, computational efficiency, and minimal data requirements for implementation (Lefèvre et al., 2014). However, they also face significant limitations including simplified assumptions that may not capture complex real-world behaviors, inability to model context-dependent social behaviors, and difficulty incorporating environmental constraints (Rudenko et al., 2020).

2.3.3 Deep Learning for Trajectory Prediction

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, excel at modeling temporal sequences and have become foundational for trajectory prediction (Alahi et al., 2016). The Social LSTM model introduced pooling mechanisms to capture interactions among nearby agents. Subsequent work incorporated attention mechanisms to selectively focus on relevant context (Vemula et al., 2018), generative models for multimodal prediction capturing multiple possible futures (Gupta et al., 2018), and convolutional networks for processing spatial scene context (Nikhil & Tran Morris, 2018). Graph Neural Networks (GNNs) provide powerful frameworks for modeling agent interactions. These approaches represent traffic scenarios as graphs where nodes correspond to agents and edges encode relationships (Mohamed et al., 2020; Salzmann et al., 2020). GNNs can naturally incorporate heterogeneous agent types, varying numbers of agents, and complex interaction patterns. Attention mechanisms within GNNs enable learned weighting of different interactions based on relevance (Mercat et al., 2020). Transformer architectures, originally developed for natural language processing, have been adapted for trajectory prediction with remarkable success (Giuliari et al., 2021; Ngiam et al., 2021).

2.3.4 Integration of Physics and Learning

Pure learning-based approaches, while powerful, can produce physically implausible predictions that violate kinematic constraints or collision avoidance requirements (Schöller et al., 2020). This has motivated research into physics-informed neural networks that integrate domain knowledge with data-driven learning. For example, incorporating physical constraints through network architecture design (Yin et al., 2019), soft constraint through physics-informed loss functions (Raissi et al., 2019), and residual learning where models predict corrections to physics-based baseline predictions (Bhattacharyya et al., 2018). Physics-informed approaches offer several advantages including improved sample efficiency, better generalization, physically plausible outputs, and enhanced interpretability (Karniadakis et al., 2021; Raissi et al., 2019). However, careful design is required to balance flexibility and constrain satisfaction, and poorly specified physics can limit model performance.

2.4 General Thoughts

Current simulation platforms excel in specific dimensions but lack the multidimensional integration required for realistic VRUs-AVs interaction research. VR systems provide immersion but typically lack

traffic simulation integration. Traffic microsimulation platforms model vehicle dynamics well but employ simplified VRUs representations. While VR can elicit authentic short-term behavioral responses, questions remain about whether decisions made in virtual environments accurately reflect real-world risk assessment and behavior, particularly for safety-critical scenarios.

GNSS-based systems suffer from blockage and multipath in urban canyons. Vision-based approaches face computational and privacy constraints. Inertial navigation exhibits unbounded drift. The emergence of C-V2X technology creates new opportunities for VRUs positioning through cooperative approaches leveraging infrastructure and vehicle support. However, research specifically addressing C-V2X-based VRU localization remains limited. Questions persist about achievable accuracy with channel state information alone, adaptability to different road environments, and robustness to communication failures and interference.

Learning-based models can produce predictions that violate basic kinematic constraints, vehicles accelerating beyond physical limits or pedestrians passing through obstacles. Physics-informed neural networks represent a promising direction but remain underexplored for heterogeneous traffic prediction. Most existing approaches incorporate physics through soft constraints in loss functions, allowing potential violations under challenging conditions. Furthermore, the majority of research focuses on vehicle-only or pedestrian-only scenarios without addressing the heterogeneity of real traffic environments where different agent types.

The present research project addresses these gaps through three synergistic contributions:

- First, the development of a distributed multi-agent VR simulation platform (Sky-Drive) that enables realistic human-in-the-loop testing of VRUs-AVs interactions through integration of immersive VR, traffic microsimulation, and digital twin frameworks. This addresses the experimental platform gap by enabling safe, repeatable, and data-rich studies of multi-agent interactions that cannot be conducted through field testing.
- Second, the creation of a C-V2X-based cooperative localization framework (CV2X-LOCA) that achieves lane-level positioning accuracy for VRUs in GNSS-denied environments using only wireless signal measurements. This addresses the localization gap by providing practical solutions that work with existing infrastructure without requiring expensive sensors.
- Third, the development of a kinematics-aware multigraph attention network (KA-MGAT) for heterogeneous trajectory prediction that combines physics-based kinematic constraints with data-driven learning to achieve both prediction accuracy and physical plausibility across diverse agent types. This addresses the prediction gap by explicitly incorporating domain knowledge while maintaining the flexibility to capture complex interaction patterns.

By integrating these three components, the project establishes a comprehensive foundation for advancing VRUs safety in the context of AVs, demonstrating how simulation, localization, and prediction technologies can be synergistically combined to address real-world safety challenges.

CHAPTER 3 DISTRIBUTED MULTI-AGENT SIMULATION PLATFORM FOR THE STUDY OF VULNERABLE ROAD USERS

To analyze interactions between VRUs and AVs, we established the Sky-Drive simulation platform. This distributed multi-agent simulator replaces dangerous field tests with a VR-based environment, allowing the safe experiments of VRUs-AVs interaction in real-time environments.

3.1 Distributed Multi-agent Simulation Platform Architecture

Sky-Drive introduces a novel distributed multi-agent architecture that enables the synchronized simulation of multiple independently operating agents across different computing devices, as shown in Fig. 3.1 (Huang et al., 2025).

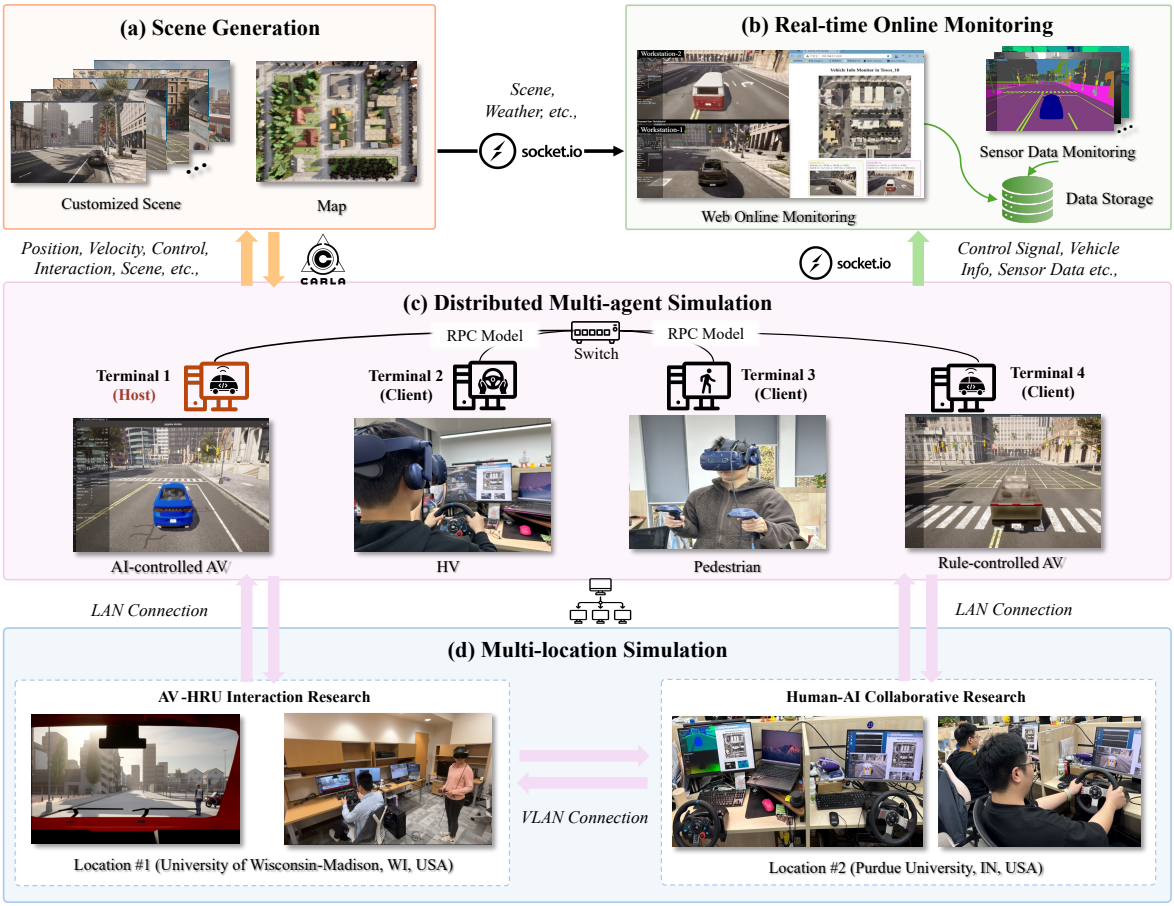


Fig 3.1 Sky Drive's Distributed Multi-agent simulation platform Architecture

1. System architecture: Built upon CARLA and the rpclib library, Sky-Drive utilizes an RPC-based framework to manage distributed vehicle control. The system architecture, as depicted in Fig. 3.1, assigns

the role of host server to Terminal 1 (the AI-controlled AV). Since the AI agent is the primary subject of evaluation, hosting the simulation locally significantly reduces synchronization latency and leverages the terminal's superior computational resources. Terminals 2 through 4 functions as clients, controlling HVs, pedestrians (VRUs), or AVs. While clients manage specific inputs, the host terminal retains authority over scene generation, constructing complex virtual environments, ranging from road topology to variable weather, and distributing this data to maintain consistency across the network.

2. Communication infrastructure: A dual-port TCP system underpins the communication layer, facilitating bidirectional traffic between the host and clients. Synchronization drift and data inconsistency are addressed through a strict centralized host-client model using time-aligned messaging. In local environments, dedicated Ethernet infrastructure reduces latency to 0.3 milliseconds. When extending the network to remote locations, such as the link between Purdue University and the University of Wisconsin–Madison, the architecture adopts VLAN tunneling to bridge the geographic gap without compromising signal reliability.

3. Real-time monitoring website: A key component of Sky-Drive's distributed architecture is its real-time monitoring and data management website. Parallel to the primary network layer, Sky-Drive incorporates a Socket.IO-based interface for real-time supervision. This module captures data from different clients in real-time. This website provides a visual feed for these activities, aggregating data into a central monitor. This design ensures data integrity across distributed nodes despite network delays or asynchronous actions.

3.2 VR-based human-in-the-loop framework

To capture human preferences and cognitive states of VRUs, Sky-Drive develops a multi-modal human-in-the-loop framework, which collects and synchronizes gaze patterns, voice commands, facial expressions, physiological signals, and control actions across multiple modalities, as shown in Fig 3.2.

1. Eye tracking: Sky-Drive provides an immersive experience through a custom-developed VR interface built on top of the Unreal Engine. Participants engage in the simulation using an HTC Vive Pro Eye headset, which supports full six degrees of freedom (6-DoF) head tracking via SteamVR and integrated eye tracking via the SRanipal SDK. The system captures high-frequency (up to 120 Hz) behavioral signals, including 3D gaze vectors, pupil positions and diameters, eye openness, and fixation points. These signals are critical for analyzing driver attention distribution, situational awareness, and cognitive workload during complex driving tasks.

2. Voice interaction: Sky-Drive supports voice commands as an explicit behavioral input modality. Spoken language is transcribed via Whisper, an OpenAI automatic speech recognition (ASR) model, and then interpreted by LLMs. The system extracts driver intent and sentiment from structured commands (“slow down at the next intersection”) and informal feedback (“too fast”), mapping them into semantic driving directives or policy preferences to guide AI behavior.

3. Facial expression recognition: A high-resolution in-cabin camera captures facial micro-expressions in real time. Sky-Drive employs expression classification models trained on affective datasets to recognize expressions such as stress, confusion, or satisfaction. These cues serve as implicit indicators of driver state and comfort, enabling real-time adaptation of AI behavior and intervention when necessary.

4. Physiological signal monitoring: Physiological states such as stress and alertness are inferred

through biometric signals collected by wearable devices. Sky-Drive integrates the Garmin vívoactive 5 smartwatch to continuously monitor heart rate and heart rate variability (HRV). These physiological signals are synchronized with other behavioral data streams, providing additional channels to model driver arousal, cognitive workload, and fatigue.

5. Steering wheel: The ego vehicle is equipped with a Logitech G920 racing wheel and pedal system, with force feedback enabled through the open-source Logitech Wheel Plugin. Steering, throttle, braking, and signaling inputs are logged in parallel with gaze and head pose data. This setup supports realistic driving control and is fully compatible with CARLA's ScenarioRunner for scenario-based experiments.

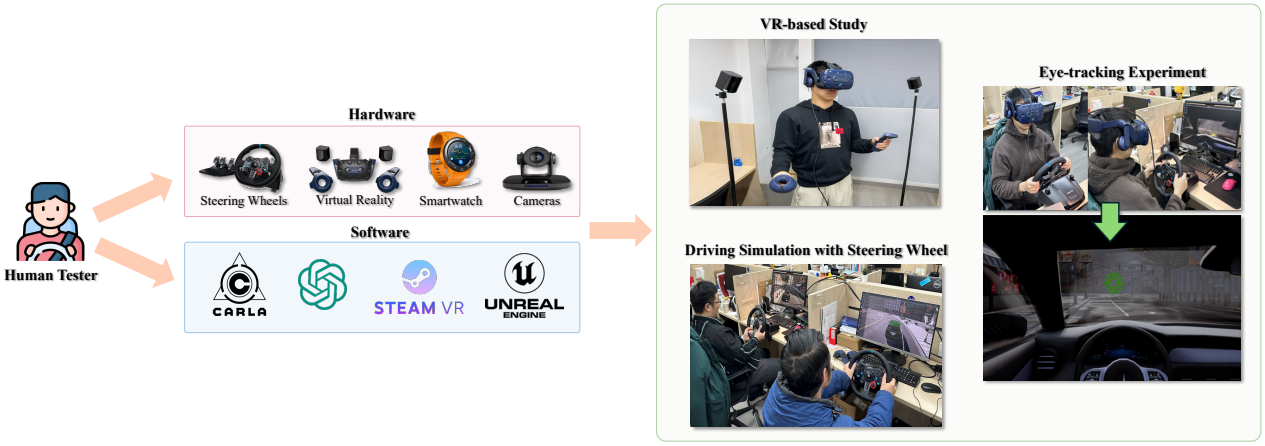


Fig 3.2 Human-in-the-loop workflow

3.3 Digital twin framework

To address this sim-to-real gap in the study of VRUs and ensure practical applicability, Sky-Drive introduces a digital twin framework that creates dynamic, high-fidelity replicas of real transportation systems.

As illustrated in Fig. 3.3, the digital twin framework consists of two core components: data integration and virtual environment construction. The multi-source data integration layer fuses static and real-time inputs from traffic cameras, loop detectors, connected vehicle telemetry, GPS traces, historical traffic records, and high-definition maps collected using lab-developed AVs equipped with LiDAR and radar. These inputs undergo temporal alignment, spatial correlation, and feature extraction to ensure semantic consistency across sources.

The simulation framework relies on CARLA and Unreal Engine to integrate real-time sensor streams with computer vision outputs. The framework reconstructs road user trajectories and projects them into a digital replica for risk assessment and forensic event replay. We validated this architecture through a pilot deployment along the Flex Lane on the Beltline (Dane County, Wisconsin). By synthesizing live feeds from WisDOT 511 with archival data from WisTransPortal, the digital twin moves beyond static modeling to achieve dynamic traffic state reconstruction and predictive analysis.

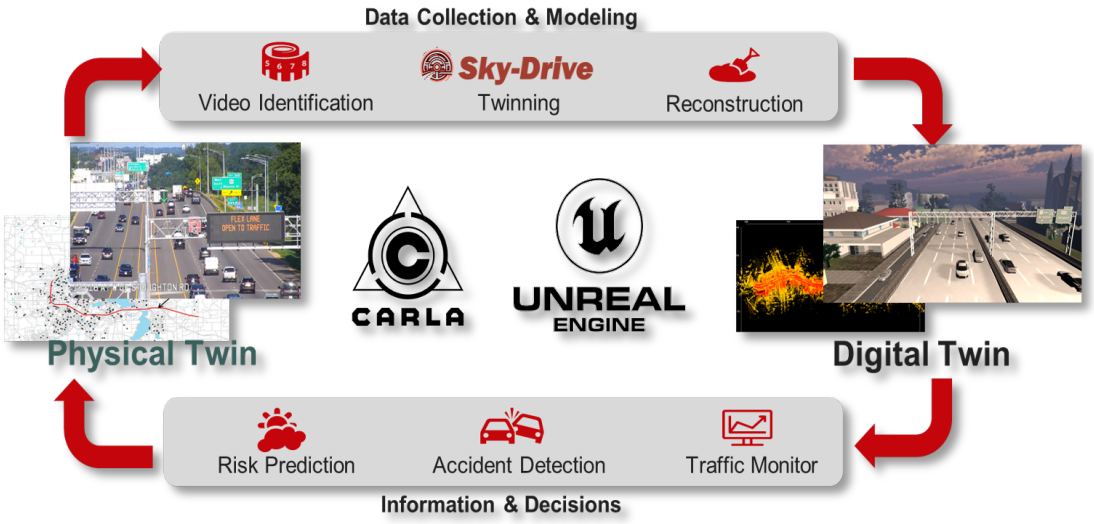


Fig 3.3 Digital twin framework

3.4 Case Study

3.4.1 Experiment Setup

Analyzing interactions between AVs and VRUs is important for traffic safety, yet real-world crash data remains scarce due to the risks of causing accidents. To overcome these physical limitations, this study employs Sky-Drive to reconstruct severe conflict scenarios. Unlike traditional methods, Sky-Drive uses a distributed multi-agent architecture to ensure real-time synchronization between multiple virtual agents. This setup is particularly valuable for modeling high-risk scenarios that are difficult to observe or replicate in the physical world.

To illustrate the experimental setup, we examine right-turn conflicts at unsignalized intersections arising from driver blind spots as an example. As shown in Fig. 3.4, we first formalize the safety-critical scenario and reconstruct it in Sky-Drive and then run simulations on the distributed multi-agent platform. In this study, we render both the driver's and the pedestrian's perspectives concurrently to analyze the mechanisms of the blind-spot conflict and its safety implications. This study leveraged Sky-Drive's synchronized multi-terminal architecture where human participants experienced the immersive scenario from the pedestrian's perspective through VR, while researchers monitored the decision-making of an AV from another terminal and intervened when necessary. During each interaction, Sky-Drive captured multimodal behavioral data from both the AV and the pedestrian. The VR recorded data like eye fixations and reaction times from the pedestrian, while Sky-Drive simultaneously logging the state of the AV. This configuration also allows researchers to analyze both the physical outcomes (e.g., successful yielding, near-misses, pedestrian hesitation) and the cognitive-emotional states of the VRUs, offering insight into how VRUs perceive and respond to AV behavior.

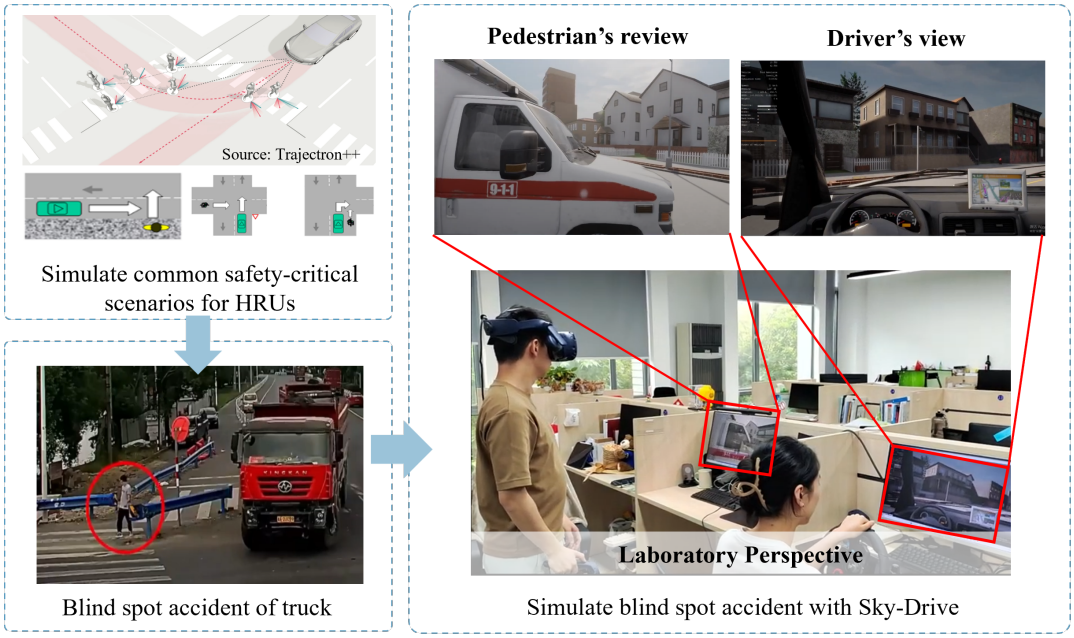
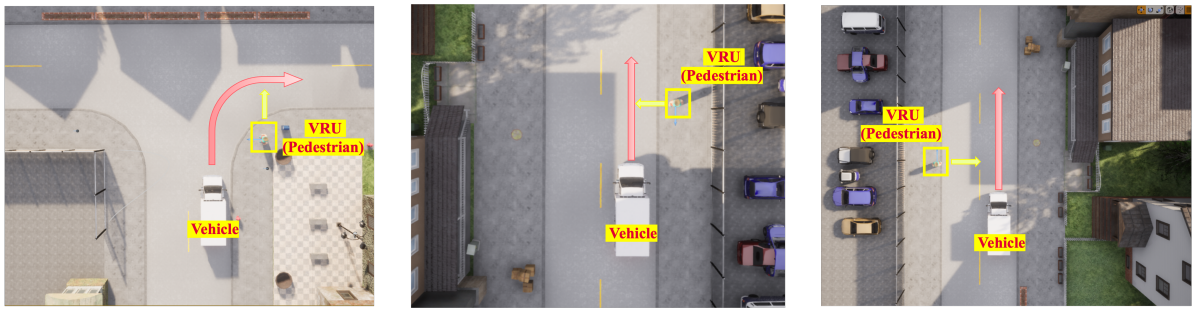


Fig 3.4 VR-based experimental setup for studying VRUs-AVs interactions at urban roads

3.4.2 Design of Safety- Critical Scenario

We investigate three safety-critical VRUs-AVs interaction patterns, selected from historical crash evidence: **(**i) Right-Turn Vehicle-to-Pedestrian**,** (ii) Straight Vehicle-to-Pedestrian, Nearside**,** and (iii) Straight Vehicle-to-Pedestrian, Farside (Fernández et al., 2022; Schmucker et al., 2010; Tan et al., 2021). Each scenario captures a distinct conflict geometry and common occlusion mechanism in urban settings. The simulation environments are implemented in CARLA and orchestrated via Sky-Drive's multi-agent simulation enabling synchronized control of the vehicles and the VRUs, with concurrent rendering of driver and pedestrian perspectives. In all figures, red arrows denote vehicle future trajectories and yellow arrows denote pedestrian future trajectories.

As shown in Fig. 3.5, right-turn conflicts model a vehicle crossing a crosswalk while a pedestrian steps off the near corner, emphasizing blind-spot and gaze-redirection effects. The nearside straight-crossing scenario represents a pedestrian entering from the vehicle's immediate curb side, stressing short preview and potential curbside occlusion; the farside case models a pedestrian initiating from the opposite curb and traversing toward the ego lane, reflecting multi-lane masking and delayed recognition. The framework readily extends to multi-vehicle/multi-pedestrian interactions without altering core scenario logic.



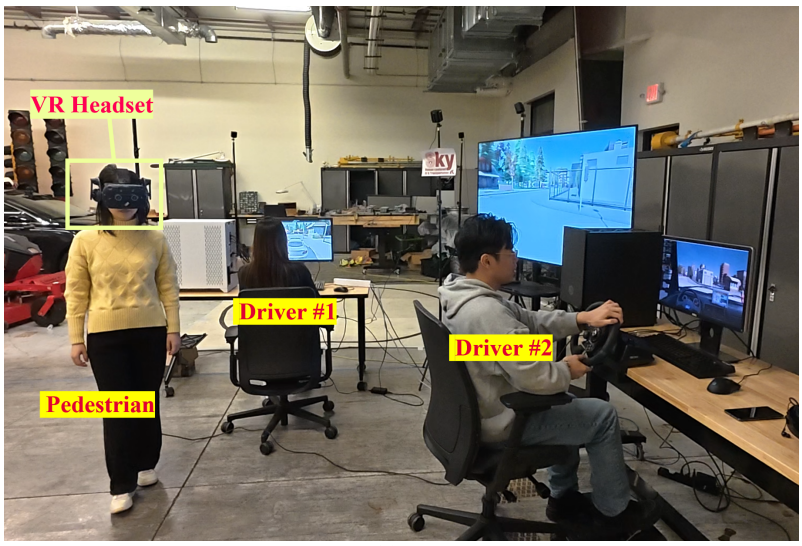
(a) (b) (c)

Fig 3.5 Design of safety-critical scenario

(a). Right turn car to Pedestrian (b). Straight car-to-Pedestrian, Nearside (c). Straight car-to-Pedestrian, Farside

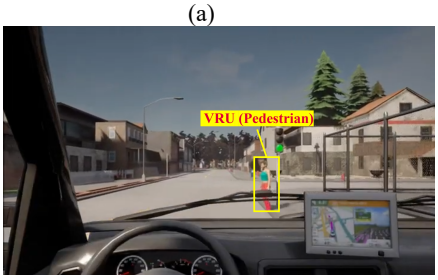
3.4.3 Experimental Results

All experiments were conducted in a controlled laboratory at the University of Wisconsin–Madison. Two operators controlled the vehicle through a force-feedback steering-wheel interface, and a third participant embodied the pedestrian using a VR headset. Scenarios were authored in CARLA and orchestrated via Sky-Drive's distributed multi-agent framework to synchronize inputs and viewpoints in real time; the testbed replicated the three conflict geometries specified earlier, with vehicle trajectories indicated in red and pedestrian trajectories in yellow.

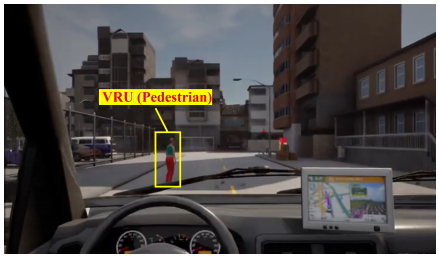


(a)

(b)



(c)



(d)

Fig 3.6 VR-based experiment setup for study of AVs-VRUs interactions

(a). Experiment setup (b). Right turn car to Pedestrian

(c). Straight car-to-Pedestrian, Nearside (d). Straight car-to-Pedestrian, Farside

1. Right turn scenario: in the right-turn vehicle-to-pedestrian configuration, trials consistently revealed a near-corner perceptual vulnerability: as drivers initiated the maneuver, gaze shifts from upstream traffic to the crosswalk, together with partial occlusion at the A-pillar/side-mirror, delayed recognition of pedestrian step-off. Yields, when present, were typically initiated after the turn apex,

indicating competition between turn execution and pedestrian search.

2. Straight road scenario: in the straight vehicle-to-pedestrian, nearside configuration, the pedestrian's emergence from the immediate curb side produced a compressed preview window; drivers frequently reported the pedestrian as "appearing suddenly," particularly with curbside clutter. Defensive stabilization followed by braking was common, and subtle VR cues (e.g., head orientation or a brief pause at the curb) shaped drivers' judgments of pedestrian intent. In the straight vehicle-to-pedestrian, farside configuration, timing and masking effects during multi-lane traversal dominated: attention was initially anchored to nearer lanes and re-prioritized only when the pedestrian entered the ego lane, with partial occlusion by opposing traffic or street furniture deferring recognition until mid-crossing. Compared with nearside trials, drivers more often assumed the pedestrian had sufficient time and delayed adjustments when the pedestrian's VR behavior appeared steady and confident.

Collectively, these simulation results indicate that right-turn conflicts are governed by gaze management and near-corner blind-spot effects, nearside crossings by short, clutter-sensitive preview and rapid intent inference, and farside crossings by mid-block masking and optimistic timing assumptions; rendering concurrent driver and pedestrian perspectives proved effective for diagnosing recognition lags while preserving safety and reproducibility.

CHAPTER 4 C-V2X-BASED COOPERATIVE LOCALIZATION FOR VULNERABLE ROAD USERS

Accurate localization of VRUs is fundamental to collision warning systems and autonomous vehicle decision-making in connected intelligent transportation systems. While vehicles can be equipped with sophisticated positioning systems including GNSS and IMUs, VRUs typically rely on consumer-grade smartphones with significantly limited positioning capabilities. This asymmetry creates a critical safety gap: the agents most vulnerable in traffic scenarios are precisely those with the least accurate positioning information. This chapter presents the adaptation of CV2X-LOCA (Huang et al., 2024) for VRUs positioning, leveraging C-V2X communication infrastructure to achieve lane-level positioning accuracy using only wireless channel state information (specifically, Received Signal Strength Indicator, RSSI), as shown in Fig. 4.1. The key advantage of this approach is that it does not require expensive onboard sensors, making it practical for widespread VRUs deployment through smartphone integration.

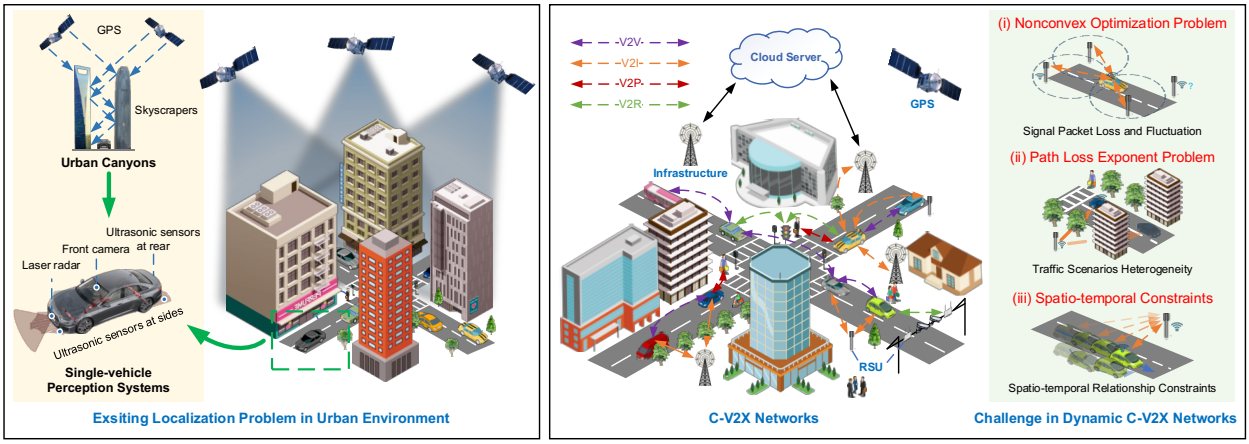


Fig 4.1 System model and problem formulation

4.1.1 RSU-Based Cooperative Network

The CV2X-LOCA system operates within a connected intelligent transportation infrastructure consisting of three primary agent types: Roadside Units (RSUs), connected vehicles (both autonomous and human-driven), and VRUs equipped with smartphones or other C-V2X-capable devices.

Roadside Units (RSUs) serve as anchor nodes with precisely known positions, typically obtained through professional surveying during installation. RSUs are equipped with C-V2X communication capabilities operating in the 5.9 GHz ITS band and are strategically deployed along roadways, particularly at intersections and other locations where VRUs-vehicle conflicts are likely to occur. Each RSU continuously broadcasts beacon messages containing its precise position coordinates, timestamp, and system information. The RSU placement strategy considers both communication coverage requirements and geometric dilution of precision (GDOP) for positioning accuracy.

Connected Vehicles serve as mobile reference nodes that can assist with VRUs positioning. Vehicles equipped with C-V2X capabilities broadcast their positions obtained from onboard high-accuracy positioning systems. While vehicle positions are subject to some uncertainty, they are generally

more accurate than VRU positions and can provide additional geometric diversity for cooperative positioning, particularly in scenarios where direct RSU-VRU communication may be obstructed.

Vulnerable Road Users are equipped with smartphones or wearable devices capable of C-V2X communication. VRUs receive beacon messages from RSUs and vehicles, measure received signal strength and apply cooperative positioning algorithms to estimate their own positions. The estimated positions can then be used for collision warning, crossing assistance, and other safety applications.

4.1.2 Communication Protocol

CV2X-LOCA leverages the C-V2X for direct device-to-device communication, which provides low-latency communication suitable for safety-critical applications. Communication follows a broadcast-based protocol where RSUs and vehicles periodically transmit beacon messages at 10 Hz (every 100 milliseconds), consistent with V2X safety message requirements.

Each beacon message contains the following information fields:

- Agent ID: Unique identifier for the transmitting agent (RSU or vehicle)
- Agent Type: Classification as RSU, autonomous vehicle, or human-driven vehicle
- Position Coordinates: Precise 2D coordinates (latitude and longitude)
- Position Accuracy: Estimated positioning uncertainty (horizontal and vertical)
- Timestamp: High-precision timestamp for message transmission
- Transmission Power: Transmit power level in dBm

4.1.3 Signal Propagation Model

The foundation of CV2X-LOCA is the relationship between received signal strength and distance. The received signal power (in dBm) at the j-th VRUs transmitted from the i-th RSU, denoted as $P_{i,j}$, follows the logarithmic path loss model (Huang et al., 2024; Q. Liu et al., 2021):

$$P_{i,j} = P_0 + 10\gamma \log_{10}\left(\frac{\boldsymbol{\theta}_j - \boldsymbol{\phi}_i}{d_0}\right) + m_{i,j}\,(\text{dBm}) \tag{4.1}$$

where:

- P_0 denotes the received signal power at reference distance d_0 (typically $d_0 = 1\,\text{m}$)
- $\| \boldsymbol{\theta}_j - \boldsymbol{\phi}_i \|$ is the Euclidean distance between the i-th RSU and j-th VRUs
- γ is the path loss exponent (PLE), which ranges from 2 to 6 depending on the environment
- $m_{i,j}$ represents the log-normal shadowing effect in multipath environments

The shadowing term $m_{i,j}$ is typically modeled as a zero-mean Gaussian random variable with shadowing standard deviation $\sigma_{i,j}^2$, i.e., $m_{i,j} \sim \text{N}\left(0, \sigma_{i,j}^2\right)$ (Huang et al., 2024). This model captures the random variations in signal strength due to obstacles, reflections, and other environmental factors.

The accuracy of positioning is fundamentally limited by how well the propagation model represents reality. In practice, the same VRUs location can experience different RSSI values due to

multipath fading from reflections off buildings and vehicles, signal attenuation from weather conditions, shadowing from large vehicles or infrastructure, and antenna orientation effects (Z. Wang et al., 2018). Our field experiments demonstrate that despite these challenges, the logarithmic path loss model provides sufficient accuracy for lane-level positioning when combined with appropriate filtering and environment adaptation mechanisms.

4.1.3 Problem Formulation

The vehicle localization problem can be formulated as a maximum likelihood (ML) estimation problem (Huang et al., 2024) Given N RSS measurements $\{P_{1,j}, P_{2,j}, \ldots, P_{N,j}\}$ from N RSUs, the ML estimator aims to find the VRUs position $\hat{\boldsymbol{\theta}}_j$ that maximizes the likelihood function:

$$\hat{\boldsymbol{\theta}}_j^{ML} = \arg\max_{\boldsymbol{\theta}_j} p\left(P_{1,j}, \ldots, P_{N,} \middle| \ \boldsymbol{\theta}_j\right) \tag{4.2}$$

Assuming independent shadowing effects, and taking the negative logarithm, this is equivalent to minimizing:

$$\hat{\boldsymbol{\theta}}_j^{ML} = \arg\min_{\boldsymbol{\theta}_j} \sum_{i=1}^{N} \frac{1}{\sigma_{i,j}^2} \left[P_{i,j} - P_0 - 10\gamma \log_{10}\left(\frac{\boldsymbol{\theta}_j - \boldsymbol{\phi}_i}{d_0}\right)\right]^2 \tag{4.3}$$

This formulation presents three fundamental challenges for practical implementation:

Challenge 1: Non-Convex Objective Function

The objective function in Eq. (4.3) is non-convex and highly nonlinear due to the logarithm and Euclidean norm terms (S. Ma et al., 2019) Traditional iterative methods like Gauss-Newton or gradient descent require good initial points and can easily get stuck in local minima, leading to poor positioning accuracy or convergence failure.

Challenge 2: Traffic Scenarios Heterogeneity

Wireless signal attenuation is affected by diverse environmental factors. The path loss exponent γ and shadowing standard deviation σ vary significantly across different road environments:

- Open highways: $\gamma \approx 2.0 - 2.5$, low shadowing
- Urban streets with moderate buildings: $\gamma \approx 2.5 - 3.5$
- Urban canyons (tall buildings): $\gamma \approx 3.5 - 4.5$, high shadowing
- Tunnels and parking structures: $\gamma \approx 4.5 - 6.0$, extreme attenuation

Using a single fixed propagation model across all environments leads to systematic positioning errors.

Challenge 3: Spatio-Temporal Constraints

The ML formulation in Eq. (4.3) treats each positioning instance independently, ignoring the temporal continuity of VRUs motion (Page & Wickramarathne, 2019; Y. Zhang et al., 2021). In reality, VRUs positions at consecutive time steps are highly correlated and constrained by physical motion models. Incorporating these spatio-temporal constraints can significantly improve positioning accuracy and robustness.

4.2 CV2X-LOCA Framework Overview

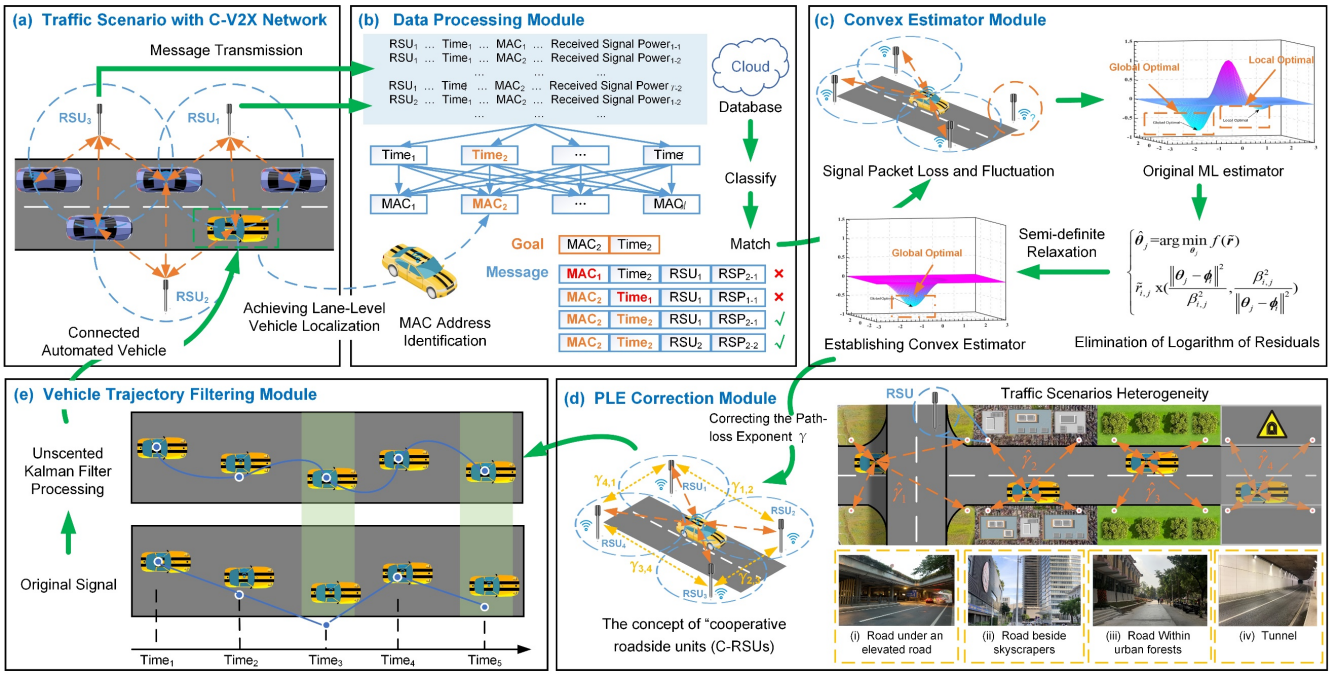


Fig 4.2 CV2X-LOCA framework

To address the three challenges identified above, we propose the CV2X-LOCA framework which consists of four key modules working in sequence: Data Processing Module, Coarse Positioning Module, Environment Parameter Correcting Module, and Trajectory Filtering Module. Fig. 4.2 illustrates the overall architecture.

4.2.1 Data Processing Module

The data processing module collects and organizes RSSI measurements from multiple RSUs. In C-V2X networks, RSUs periodically broadcast beacon messages (typically at 10 Hz) containing RSU position coordinates, transmission timestamp, transmit power level, and unique MAC address identifier (Naik et al., 2019).

The VRUs device (smartphone) receives these beacons and extracts the RSSI value from the physical layer. The key challenge is synchronizing measurements from different RSUs that correspond to the same VRUs position at the same time instant. This is achieved through MAC address and timestamp matching.

To mitigate signal packet loss and short-term fading, sliding window averaging (window size = 5) is applied to the RSSI measurements. This reduces measurement noise while maintaining sufficient temporal resolution for tracking moving VRUs.

4.2.2 Coarse Positioning Module

The coarse positioning module addresses Challenge 1 (non-convex objective function) by transforming the maximum likelihood estimation problem into a convex semi-definite programming (SDP) problem that can be solved efficiently and globally Wang et al., 2018).

The traditional ML estimator for RSS-based positioning is highly nonlinear due to the logarithmic relationship between signal strength and distance. This nonlinearity creates a non-convex optimization landscape with multiple local minima, making it difficult for iterative solvers to find the global optimum.

CV2X-LOCA overcomes this challenge through a two-step transformation process:

1. Logarithm Elimination

First, the algorithm eliminates the logarithm term by converting RSSI measurements into approximate distance estimates using the inverse of the path loss model. This creates a simpler non-convex problem involving only Euclidean distances.

2. Semi-Definite Relaxation

Second, semi-definite relaxation (SDR) techniques are applied to transform the non-convex distance-based problem into a convex SDP that can be solved using standard interior-point methods (Z. Wang et al., 2018). The key innovation is lifting the problem into a higher-dimensional space where the non-convex constraints become convex semi-definite constraints.

4.2.3 Environment Parameter Correction Module

The environment parameter correction module addresses Challenge 2 (traffic scenarios heterogeneity) by adapting the signal propagation model parameters to local road conditions (Huang et al., 2024).

The C-RSU Concept

A key innovation of CV2X-LOCA is the introduction of Cooperative RSUs (C-RSUs) (Huang et al., 2024). Rather than treating each RSU independently, the C-RSU concept aggregates information from multiple physical RSUs to characterize the local propagation environment. This enables environment-specific parameter adaptation without requiring per-RSU calibration.

Environment Classification

CV2X-LOCA classifies road environments into four categories based on signal propagation characteristics (Huang et al., 2024), as shown in Table 4.1.

Table 4.1 Environment classification and path loss exponents

Environment Class	Description	Path Loss Exponent (γ)	Typical Scenarios
Class 1	Open/Semi-Open	1.0-2.0	Under elevated roads, open suburban streets
Class 2	Urban Forests	2.0-3.0	Tree-lined residential streets parks
Class 3	Urban Canyons	3.0-4.0	Downtown streets, dense commercial districts
Class 4	Severely Obstructed	4.0-6.0	Tunnels, parking structures, underpasses

For each environment class, a set of correction parameters is maintained and applied to the distance estimates from the coarse positioning module. These parameters are learned offline from training data collected in each environment type.

Correction Process

The correction process works as follows:

1. Environment Recognition: When a VRU enters a new region, the system identifies the appropriate environment class based on map data or real-time signal characteristics

2. Parameter Application: The distance estimation uses environment-specific path loss exponents and shadowing variances

3. Position Refinement: The coarse position estimate is refined by re-solving the positioning problem with corrected parameters

4.2.4 Trajectory Filtering Module

The trajectory filtering module addresses Challenge 3 (spatio-temporal constraints) by incorporating VRU motion models to smooth position estimates and improve accuracy.

Motion Model Integration

VRUs' positions at consecutive time steps are highly correlated and constrained by physical motion dynamics. The trajectory filtering module exploits this temporal structure by modeling VRU motion using a discrete-time kinematic model that captures position and velocity (Page & Wickramarathne, 2019).

The state vector includes 2D position coordinates and velocity components. The motion model assumes constant velocity between updates, with process noise accounting for acceleration variability.

Unscented Kalman Filter

The UKF operates in two phases:

1. Prediction Phase: Projects the current state forward in time using the motion model to predict the VRU's position at the next time step.

2. Update Phase: When a new position measurement arrives, the UKF fuses it with the predicted position, weighting each based on their respective uncertainties.

4.3 Experimental Evaluation

4.3.1 Simulation Setup

Network Configuration: We simulated a two-way four-lane road segment of 2000m length with standard lane width of 3.5m. RSUs were deployed on both sides of the road with varying deployment spacing (dr_1) from 30m to 210m. The RSU-to-road-edge distance was set to $dr_2 = 5$ m. The parameter setting in the simulation experiments are shown in Table 4.2.

Table 4.2 Parameter setting in the simulation experiments

Symbol	Parameters	Value
$d0$	Reference distance	1m
v_x	Vehicle speed	25km/h
N	Number of hearable RSUs	3
σ^2_{dB}	Shadowing standard deviation	2dB
Δt	Detection frequency	0.1s
$dr1$	Deployment spacing	60m

$dr2$	Distance	1m
$dr3$	Width of road	14m
L	The number of anchor nodes	4

Environment Settings: Four road environments were simulated corresponding to the classification in (Huang et al., 2024): semi-open ($\gamma = 1$), urban forests ($\gamma = 2$), urban canyon ($\gamma = 3$), and tunnel ($\gamma = 4$).

Baseline Methods: We compared CV2X-LOCA against state-of-the-art localization methods such as LS-TS (J. Zhu et al., 2014), WLLS (S. Ma et al., 2019), WCL ,(Magowe et al., 2019) LS-EKF (Page & Wickramarathne, 2019), SDP-LSRE (Z. Wang et al., 2018), SDP-ML-KF (Y. Zhang et al., 2021), and ML-True (performance upper bound). More detail information can be found in (Huang et al., 2024).

4.3.2 Performance Comparison Across Environments

Table 4.3 presents the localization performance of different methods across the four road environments at two speed levels (25 km/h and 100 km/h for vehicles; proportionally scaled for VRUs).

Table 4.3 Average localization error (m) by environment and speed

Methods	**ALE**	**RMSE**	**MAE**	**MAPE**	**ALE**	**RMSE**	**MAE**	**MAPE**	**ALE**	**RMSE**	**MAE**	**MAPE**	**ALE**	**RMSE**	**MAE**	**MAPE**
	Road Environment (a)				Road Environment (b)				Road Environment (c)				Road Environment (d)			
25km/h																
ML-True	2.48	3.7	3.25	0.52	5.25	7.76	7.07	1.61	7.93	11.21	10.19	2.72	10.78	14.83	13.04	3.84
LS-TS	3.69	5.93	4.67	0.76	4.23	6.6	5.8	1.1	5.89	8.81	7.94	1.85	5.35	8.75	7.34	1.33
WCL	4.33	6.07	5.05	0.44	5.92	7.83	6.54	0.36	9.42	11.68	9.93	0.28	15.54	18.77	16.04	0.27
LLS	5.42	10.11	6.42	1.46	4.26	7.83	4.89	0.77	4.65	8.74	5.57	1.01	12.47	18.58	15.39	4.82
WLLS	4.73	8.37	5.71	1.18	4.04	6.84	4.68	0.68	4.4	7.98	5.31	0.9	12.31	18.33	15.24	4.75
WCL-TS	4.16	6.06	5.38	1.4	6.54	10.47	8.54	2.1	10.84	23.77	13.82	3.67	28.81	95.31	36.66	9.67
LS-EKF	3.52	6.97	4.15	0.71	3.18	6.61	3.63	0.56	3.84	7.57	4.54	0.93	12.99	17.82	14.81	5
SDP-LSRE	3.77	4.87	4.1	0.21	4.14	5.28	4.5	0.23	4.28	5.88	4.59	0.18	4.84	7.12	5.16	0.18
GRNN-UKF	2.36	3.5	3.08	0.53	2.66	4.23	3.44	0.83	95.84	115.85	104.25	38.05	2756	3584.55	3001.42	1092.89
CF-LS-UKF	2.26	3.33	2.91	0.46	6.23	7.96	7.34	2.44	167.72	175.94	170.61	67.13	4255	4683.48	4344.25	1691.75
SDP-ML-KF	3.28	4.74	3.84	0.38	3.43	4.55	3.76	0.23	3.49	4.57	3.79	0.2	4.47	6.68	5.78	0.92
CV2X-LOCA	1.47	2.26	1.91	0.34	1.4	2.36	1.69	0.22	1.33	2.58	1.59	0.16	1.28	2.06	1.54	0.19
100km/h																
ML-True	3.12	4.55	3.78	0.39	6.13	9.25	7.96	1.21	8.05	12.16	10.52	2.21	10.88	16.46	13.93	3.45
LS-TS	8.58	14.95	11.17	2.1	4.82	8.34	6.46	1.24	5.47	8.31	7.31	1.85	5.44	8.48	7.41	1.43
WCL	8.53	11.11	9.27	0.51	6.74	8.87	7.4	0.48	8.97	11.24	9.5	0.38	14.89	17.69	15.38	0.34
LLS	16.94	33.81	20.8	5.87	7.38	13.74	8.8	2.15	4.58	8.49	5.42	1.18	11.63	17.13	14.34	4.55
WLLS	16.29	33.44	20.09	5.57	6.9	13.31	8.44	2	4.03	7.07	4.81	0.93	11.85	17.25	14.8	4.65
WCL-TS	5.55	9.23	7.16	1.96	8.91	14.74	11.21	3.28	11.3	18.6	13.55	4.26	25.02	59.53	31.23	8.83
LS-EKF	35.65	53.66	39.51	2.66	33.27	47.77	35.06	1.3	33.17	47.4	35.19	1.24	36.91	58.98	45.62	5.18
SDP-LSRE	6.05	9.16	6.99	0.77	4.64	6.87	5.39	0.52	3.84	5.59	4.51	0.45	4.43	7.32	5.29	0.62
GRNN-UKF	13.57	19.24	15.86	4.82	5.56	7.88	7.12	2	88.71	126.13	105.87	32.86	2015	3293.06	2397.23	766.26
CF-LS-UKF	13.46	19.07	15.59	4.73	10.27	14.9	13.01	3.97	270.2	324.72	289.12	108.38	5124	6462.46	5466.99	2070.19
SDP-ML-KF	6.72	9.91	7.66	0.66	4.33	6.34	5.34	0.75	4	5.95	5.15	0.81	5.41	8.58	6.78	1.4

CV2X-LOCA	3.17	4.73	3.93	0.48	2.54	3.85	3.37	0.67	2.56	3.81	3.46	0.71	3.18	4.8	4.23	0.78

Key Observations: CV2X-LOCA consistently achieves the lowest positioning error among practical methods. The environment parameter correction module provides error reduction over SDP-ML-KF (Y. Zhang et al., 2021). Traditional least-squares methods (S. Ma et al., 2019; J. Zhu et al., 2014) show significantly degraded performance in challenging environments.

4.3.3 Impact of RSU Deployment Spacing

As shown in Fig. 4.3, RSU deployment spacing involves a critical trade-off between positioning accuracy and infrastructure cost. The experimental results reveal that for practical deployment, 120-150m spacing provides an optimal balance, with CV2X-LOCA maintaining ALE < 5m even at 150m spacing. Overall, the deployment spacing of 120~150m is a feasible solution.

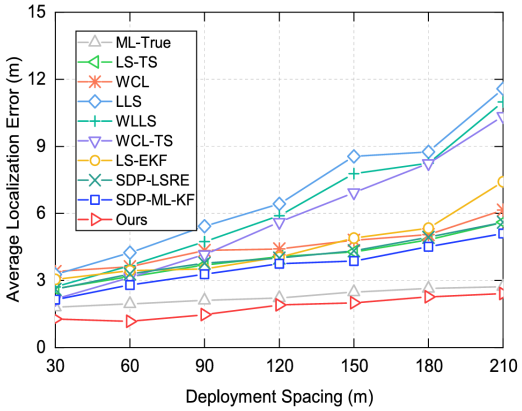
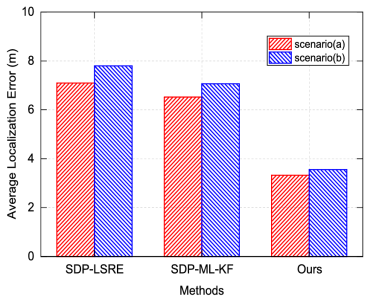


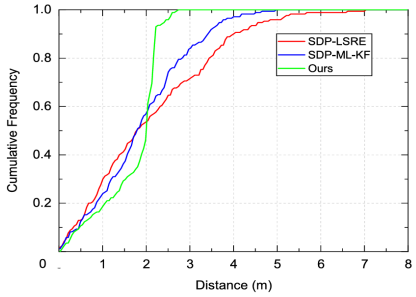
Fig 4.3 The change of ALEs under different deployment spacing (between 30 ~ 210m).

4.3.4 Field Experiment Results

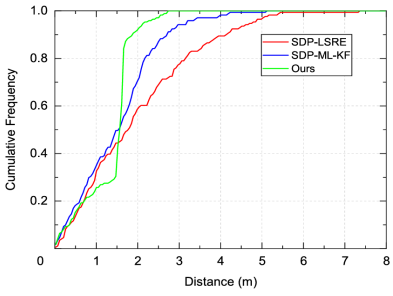
This section focuses on the localization performance of different real-word road environments under different speed (25km/h and 60km/h). In real-world road environment (a), the test vehicle drives in lane one and then changes to lane two. In real-world road environment (b), the test vehicle drives in lanes one/two at the default speed. To reveal the actual positioning accuracy of the compared positioning algorithms, we conducted multiple repetitions of field tests for both road environments with different vehicle speed (25 km/h and 60 km/h). In both experiments, the vehicle traveled forward from about 200m away from the first RSU and then forward after passing the last RSU about 200m.



(a)



(b)



(c)

Fig 4.4 The localization performance of different methods under real-world road environments

(a) The change of ALEs under two real-world road environments. (b) the CDFs of longitudinal coordinates under real-world road environment (a). (c) the CDFs of longitudinal coordinates under real-world road environment (b).

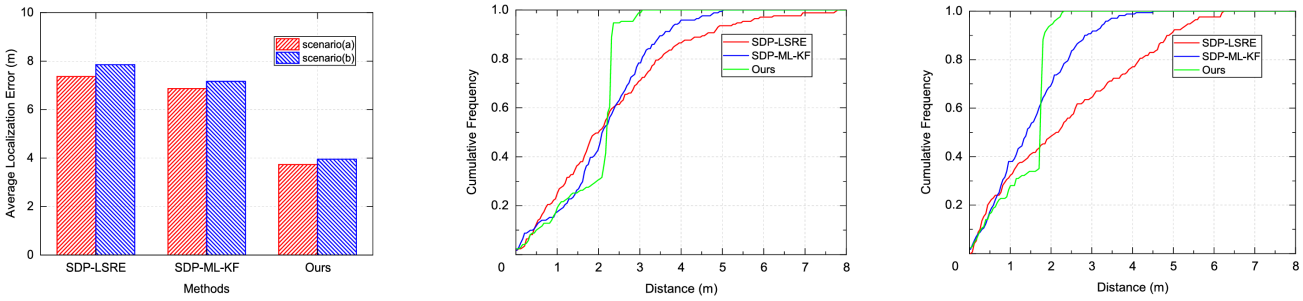


Fig 4.5 The localization performance of different methods under real-world road environments

(a) The change of ALEs under two real-world road environments. (b) the CDFs of longitudinal coordinates under real-world road environment (a). (c) the CDFs of longitudinal coordinates under real-world road environment (b).

The results of the field test with a speed set to 25km/h are shown in Fig. 4.4 (a), (b), and (c). From Fig. 4.4 (a), it can be concluded that our CV2X-LOCA achieves the best localization performance (ALE of 3.33m, and 3.56m, respectively) in both real-world road environments (a) and (b). The cumulative distribution functions (CDFs) of longitudinal coordinates are shown in Fig. 4.4 (b) and (c). Our CV2X-LOCA achieves a 90th percentile error of 2.3m and 2.1m in real-world road environments (a) and (b). In comparison, the SDP-LSRE achieves a 90th percentile error of 4.5m and 5.1m, while the SDP-ML-KF achieves a 90th percentile error of 3.4m and 2.8m.

In the second field test, the positioning accuracy was tested at 60km/h (i.e., the urban road speed limit), and the results are given in Fig. 4.5. A strong consistency between the two sets of results can be found, i.e., CV2X-LOCA (our) achieves the best localization performance (ALE of 3.73m and 3.96m) even with 60 km/h. From Fig. 4.5 (b) and (c), CV2X-LOCA (our) achieves a 90th percentile error of 2.34m and 1.84m in real-world road environments (a) and (b). In comparison, the SDP-LSRE achieves a 90th percentile error of 4.63m and 4.93m, and SDP-ML-KF commits a 90th percentile error of 2.73m and 2.84m, respectively. To summarize, our CV2X-LOCA maintains the best localization performance in real-world road environments (a) and (b) regardless of the speed of 25km/h or 60km/h.

4.4 Discussion

This chapter presents the adaptation of CV2X-LOCA for VRU localization in GNSS-denied environments. The positioning capabilities demonstrated by CV2X-LOCA have important implications for enabling advanced VRU safety applications. Lane-level accuracy enables reliable collision warning systems that can accurately determine whether a VRU is in a vehicle's path. For crossing assistance at intersections, accurate VRU positioning enables detection of crossing intent and coordination with approaching vehicles. For VRUs with visual or hearing impairments, accurate positioning enables assistive technologies.

CHAPTER 5 KINEMATICS-AWARE MULTIGRAPH ATTENTION NETWORK FOR HETEROGENEOUS TRAJECTORY PREDICTION

Accurate trajectory prediction for heterogeneous traffic agents, including autonomous vehicles, human-driven vehicles, pedestrians, and cyclists, is fundamental to ensuring safety in connected and automated transportation systems. In safety-critical scenarios involving vulnerable road users (VRUs), reliable motion forecasting allows proactive collision avoidance and smooth human–AI interaction within the connected network.

Traditional physics-based methods have offered interpretable representations of motion dynamics and physical feasibility, yet they often fall short in capturing the complex, context-dependent interactions among heterogeneous agents. Conversely, purely learning-based approaches powered by deep neural networks achieve higher prediction accuracy but frequently produce physically implausible trajectories due to the lack of explicit kinematic constraints and generalization to unseen environments.

To bridge this gap, we proposes a Kinematics-Aware Multigraph Attention Network (KA-MGAT) that integrates physical motion models into a deep learning framework for heterogeneous trajectory prediction. The model explicitly encodes agent-specific kinematic properties (such as velocity, acceleration, and steering angle constraints) into a multigraph attention architecture, which captures the complex spatial–temporal interactions among multiple heterogeneous agents. In addition, a residual prediction module is introduced to refine the outputs of the kinematics-guided branch, correcting systematic errors introduced by simplified assumptions in the underlying motion equations.

Unlike prior work focusing solely on vehicles or pedestrians, KA-MGAT is designed for mixed traffic environments, aligning with the human-in-the-loop simulation framework developed in the Sky-Drive platform. This allows the model to leverage behavior data from immersive VR-based experiments and integrate contextual cues (e.g., interaction intent, head orientation, and relative positioning) into its learning process.

Comprehensive evaluations were conducted using both public trajectory datasets (ApolloScape and NGSIM) and heterogeneous interaction scenarios that can be seamlessly integrated into the Sky-Drive digital-twin simulation environment. Experimental results demonstrate that KA-MGAT achieves superior performance compared to kinematics-agnostic baselines in terms of average displacement error (ADE), final displacement error (FDE), and physical plausibility of predicted trajectories. The model also exhibits improved data efficiency and generalization capability, making it a promising foundation for real-time VRUs safety applications and cooperative motion planning in connected intelligent transportation systems.

5.1 Framework Overview

In this section, we introduce our proposed kinematics-aware multigraph attention network. The overall framework is illustrated in Fig 5.1. To aggregate the interaction information of heterogeneous traffic agents, we first propose a multigraph attention module to model social interactions. Subsequently, we integrate kinematic models with decoder networks for each type of traffic agent. Finally, a novel residual

prediction module is proposed to further refine the predictions.

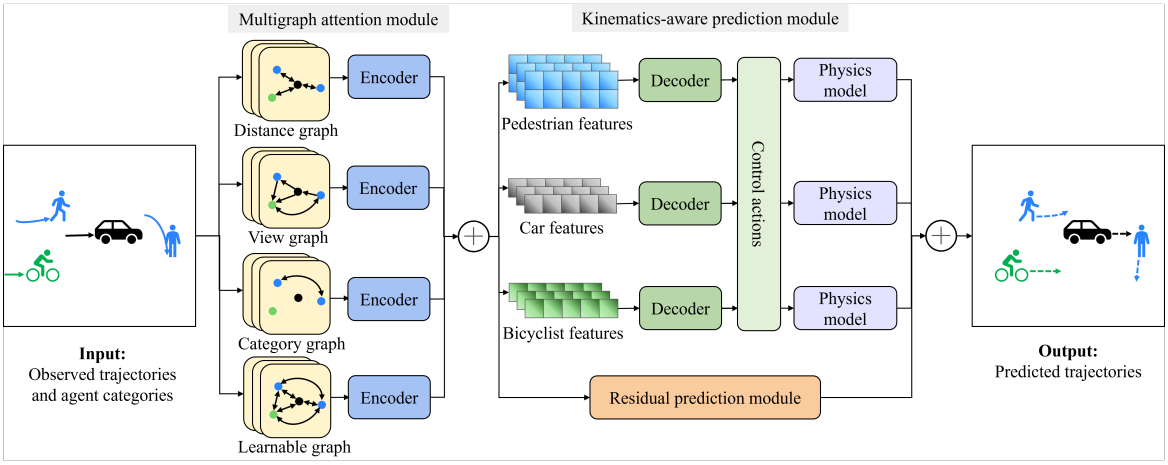


Fig 5.1 Architecture of the proposed kinematics-aware multigraph attention network

5.1.1 Problem Formulation

The goal of heterogeneous trajectory prediction is to generate accurate future trajectories for traffic agents in different categories based on their historical states. At time t t, the input of a prediction system is the historical states of N N traffic agents in a scene over an observation horizon T_{obs}, which can be written as $\boldsymbol{X}_t = \{\boldsymbol{x}_t^i \mid i = 1, \ldots, N\}$, where $\boldsymbol{x}_t^i = \left[\boldsymbol{s}_{t-T_{\text{obs}}}^i, \ldots, \boldsymbol{s}_t^i\right]$ is the observed state of traffic agent i i at time t t, and $\boldsymbol{s}_t^i$ $\mathbf{s}_t^i$ can include the location, speed, category, etc. Similarly, the output is the predicted future trajectories over a prediction horizon T_{pred}, denoted as $\boldsymbol{Y}_t = \{\boldsymbol{y}_t^i \mid i = 1, \ldots, N\}$, where $\boldsymbol{y}_t^i = \left[\boldsymbol{s}_{t+1}^i, \ldots, \boldsymbol{s}_{t+T_{\text{pred}}}^i\right]$. In practical scenarios, the current states of heterogeneous traffic agents can be captured using onboard sensors such as LiDAR, cameras, and roadside units (S. Chen et al., 2023; Guan et al., 2023; Huang et al., 2019, 2020, 2024, p. 202)

5.1.2 Multigraph attention module

To model the social interactions among traffic agents, we propose a multigraph attention network. As shown in Fig 5.1, we build three prior knowledge-based fixed graph attention topologies, i.e., distance graphs, view graphs, and category graphs. In addition, a learnable attention graph is constructed to explore and capture interaction patterns that might not have been covered by prior knowledge. Each node in a graph topology represents an agent, and the directed edge connecting two agents indicates that an interaction exists between them. The multigraph attention topology (G) is denoted as

$$G = \left(V, E_{\text{d}}, E_{\text{v}}, E_{\text{c}}, E_{\text{l}}\right), \tag{5.1}$$

where $V = \{v^i \mid i = 1, \ldots, N\}$ is the set of all nodes, $\boldsymbol{x}_t^i$ is the feature vector of node v^i at time t, and E_{d}, E_{v}, and E_{c} E_c represent the edge sets of distance, view, and category, respectively. E_{l} denotes the edge set of the learnable attention graph. Now, the question is how to define the attention weights in the adjacency

matrix of each edge set. We use the adjacency matrix from the last observed time step, i.e., the current time t.

First, to construct the attention weights of the distance graph, we regard any two agents with a relative distance less than 10m as connected by an edge. Considering that closer traffic participants have stronger interactions, we use this prior knowledge to calculate the edge attention weights, where the edge connecting two near neighbors is assigned a higher value. The attention weight of the distance graph is calculated as

$$e_d^{ij} = \begin{cases} \dfrac{1}{\| p_t^i - p_t^j \|_2}, & \text{if} \| p_t^i - p_t^j \|_2 \le 10 \\ 0, & \text{otherwise} \end{cases} \tag{5.2}$$

where p_t^i p_t^i and p_t^j p_t^j are the location coordinates of agents i i, j j at time t t, respectively, and $\|\cdot\|_2$ $\| \cdot \|_2$ denotes the L2 norm.

Next, we establish the view graph based on the assumption that road users are predominantly influenced by objects within their visual proximity. Accordingly, if a traffic agent falls within the range of $\pm\pi/2$ in the view of another agent, the edge attention weight between them in the view graph is designated as 1; conversely, it is set to 0 if outside the specified range, and vice versa. This approach ensures that the view graph captures the relevant interactions among traffic agents, emphasizing the impact of proximate objects on their behaviors.

$$e_v^{ij} = \begin{cases} 1, & \text{if} -\pi/2 \le \alpha \le \pi/2 \\ 0, & \text{otherwise} \end{cases} \tag{5.3}$$

where α is the angle of agent i's view with respect to the relative location vector of agents i and j.

Finally, traffic agents belonging to the same category exhibit similar motion patterns and adhere to shared social conventions. To elaborate, agents within the same category are likely to exhibit motion patterns that are similar due to shared physical capabilities and constraints. For example, vehicles generally have faster speeds and different acceleration/deceleration patterns than pedestrians. By categorizing agents, our model can more accurately predict their movements based on the typical behavior of their category. Different categories of agents also follow category-specific social conventions and rules. For instance, vehicles follow road lanes and traffic signals, while pedestrians may use sidewalks and crosswalks. The category graph in our model incorporates these social conventions, which are critical in predicting how agents interact with each other and with the environment.

To leverage this inherent similarity, we establish a category graph that consolidates interaction information for agents of the same type. In the category graph, two traffic agents sharing the same category are considered connected, and their attention weights are uniformly set to 1. This construction facilitates the focused aggregation of relevant interaction details within distinct agent types, enhancing the model's understanding of category-specific behaviors.

$$e_c^{ij} = \begin{cases} 1, & \text{if } c^i = c^j \\ 0, & \text{otherwise} \end{cases} \tag{5.4}$$

where c^i and c^j denote the categories of agents i and j, respectively.

The data-learned attention mechanism serves as a complementary component, aiming to capture interactions not explicitly covered by prior knowledge. This integration empowers our network to overcome the constraints of fixed attention, offering the flexibility to autonomously learn the relevance and importance of interactions from the raw data.

The data-learned attention mechanism in our proposed multigraph attention module is inspired by the graph attention network (GAT) (Velickovic et al., 2017) The edge attention weights e_l^{ij} in the learnable attention graph are calculated using a self-attention mechanism:

$$e_l^{ij} = \text{LeakyReLU}\left(\lambda \cdot \left[Wv^i \| Wv^j\right]\right), \tag{5.5}$$

where v^i and v^j are the feature vectors of nodes i i and j j, respectively; W W is a weight matrix; λ is a learnable parameter vector; and $\|$ denotes concatenation. The attention coefficients are further normalized using the softmax function to ensure that the contributions from neighboring nodes are appropriately scaled:

$$e_l^{ij} = \frac{\exp(e_l^{ij})}{\sum_{k \in N_i} \exp\left(e_l^{ij}\right)} \tag{5.6}$$

where N_i represents the set of neighbors of node i i.

The multigraph attention module consists of multiple encoders. Each encoder takes a graph attention topology as input and contains layers of graph convolutions and standard convolutions. The convolution operations in each layer are defined as

$$\begin{aligned} H_g^{(l)} &= \sigma\left(E_i H^{(l-1)} W^{(l-1)}\right) \\ H^{(l)} &= \sigma\left(\Phi \otimes H_g^{(l)}\right) \end{aligned} \tag{5.7}$$

where $H_g^{(l)}$ is the output of the graph attention convolution, $\sigma(\cdot)$ denotes an activation function, E_i represents an adjacency matrix of a graph, $H^{(l-1)}$ is the node feature with $H^0 = V$, $W^{(l-1)}$ is the trainable parameter of the graph attention convolution, $H^{(l)}$ denotes the output, Φ is the kernel parameter, and $\otimes$ denotes a standard convolution operation. After each graph topology passes through several layers of graph attention and temporal convolutions, we obtain interaction features from both spatial and temporal dimensions. The final interaction features are obtained by fusing these interaction features. We implement a weighted fusion mechanism, which is achieved by a dense layer. This mechanism involves learning weights through the training process, which allows the model to dynamically adjust the influence of each graph based on the current traffic context. The weights are learned in an end-to-end manner, ensuring that the model adaptively emphasizes the most relevant graph(s) for a given situation.

5.1.3 Kinematics-aware prediction module

Considering the distinct motion patterns exhibited by heterogeneous agents, our approach incorporates two fully differentiable kinematic models, i.e., the bicycle model (Rajamani, 2006) and the unicycle model (LaValle, 2006), as the physics components to construct the kinematics-aware prediction module. As depicted in Fig 5.2, the bicycle model, which considers the relationship between the front and rear wheels, is well suited for describing the motion of bicycles and vehicles. On the other hand, the unicycle model,

which simplifies the motion to a single wheel, is more suitable for characterizing pedestrian movement. The bicycle model is given by

$$\begin{cases} \dot{x} = v \cdot \cos(\psi + \beta) \\ \dot{y} = v \cdot \sin(\psi + \beta) \\ \dot{\psi} = \frac{v}{l_r} \cdot \sin\beta \\ \dot{v} = a \\ \beta = \arctan(\frac{l_r}{l_r + l_f} tan\delta) \end{cases} \tag{5.8}$$

where (x, y) is the coordinate of the center of mass, ψ denotes the heading, v represents the speed, and β is the angle of the current speed with respect to the longitudinal axis of the agent. l_f and l_r are the lengths from the center of mass to the front and rear axles, respectively. a and δ denote the acceleration and steering, respectively. The unicycle model is given by

$$\begin{cases} \dot{x} = v \cdot \cos\theta \\ \dot{y} = v \cdot \sin\theta \\ \dot{\theta} = \omega \\ \dot{v} = a \end{cases} \tag{5.9}$$

where θ is the orientation and ω denotes the angular velocity. The control actions for the bicycle model and unicycle model are $[a, \delta]$ and $[a, \omega]$, respectively. For clarity, we denote the kinematic model as $\dot{s} = f(u, s)$. With these two kinematic models, our task is to predict the control actions for heterogeneous traffic agents such that we can generate their future trajectories with kinematic constraints.

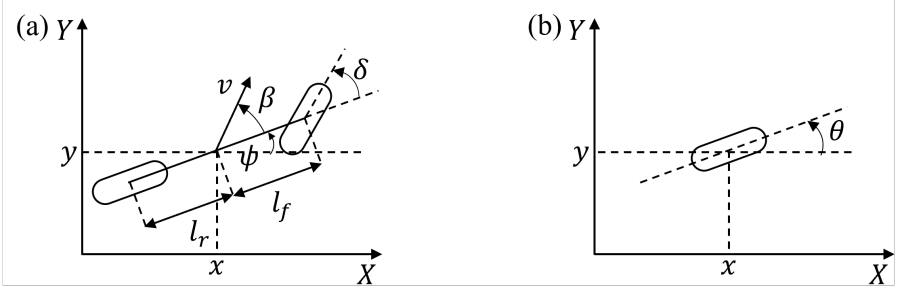


Fig 5.2 Illustration of the kinematic models

(a) bicycle model and (b) unicycle model.

To address the heterogeneity of traffic agents, we assign a specific decoder to each type of traffic agent. This will enhance the learning ability of our proposed network since agents of the same category share similar motion patterns. We consider three agent categories, namely, bicycles, pedestrians, and cars, so there are three category-specific decoders. The number of decoders can be scaled based on the actual

traffic conditions. As shown in Fig 5.1, each decoder takes the features of one type of agent as input. We first apply a Gated Recurrent Unit (GRU) to extract motion features for each traffic agent:

$$o^i, h^i = \mathrm{GRU}_1^c\left(g^i\right) \tag{5.10}$$

where GRU_1^c represents a GRU for agent category c; g^i is agent i's features extracted from the multigraph feature extractor; and o^i and h^i are the output and hidden features, respectively. Next, as shown in Fig 5.3, another GRU takes the hidden feature as the initial hidden feature and then generates control actions, which are sent to the corresponding kinematic model in an iterative manner:

(5.11)

where o_t^i and h_t^i are the initial input and hidden features for GRU_2^c, $o_t^i = \boldsymbol{s}_t^i$ is agent i's current state, and $h_t^i = h^i$ is the hidden feature generated by GRU_1^c. o_{t+1}^i is the output of GRU_2^c, which is clipped by $U(\cdot)$ to ensure that the control actions u_t u_t are in the feasible range. $\hat{f}(\cdot)$ denotes the discretized kinematic model. At time $t+2$, the inputs of GRU_2^c are o_{t+1}^i and h_{t+1}^i generated at the last time step, and the outputs are o_{t+2}^i and h_{t+2}^i, in which o_{t+2}^i is sent to the corresponding kinematic model. By repeating this process, the final output of this module is the predicted trajectories of all agents in the prediction horizon.

During training, the role of the physics model is to use these control actions to generate predictions based on physical principles. The output of the physics model, which includes the predicted trajectories of the traffic agents, is then compared to the actual trajectories in the dataset to compute the loss. This loss is used to update the parameters of the neural network layers to improve the prediction accuracy. Therefore, the physics model plays a critical role in the training process by providing physically plausible trajectory predictions that guide the learning of the neural network components of our model.

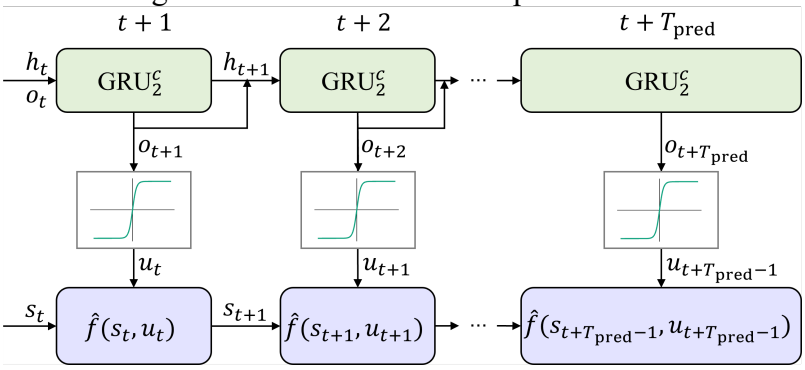


Fig 5.3 Kinematics-aware trajectory prediction.

5.1.4 Residual prediction module

The residual prediction module is designed to enhance the accuracy and realism of trajectory predictions produced by the kinematics-aware prediction module. As shown in Fig 5.4, although the kinematics-aware prediction module generates trajectory predictions that adhere to kinematic constraints, discrepancies with the actual ground truth trajectories can still arise due to inherent simplified assumptions in kinematic models. To address this limitation, we introduce the residual prediction module to further refine the

kinematics-constrained prediction. This module adopts a similar decoder architecture to that of the kinematics-aware prediction module.

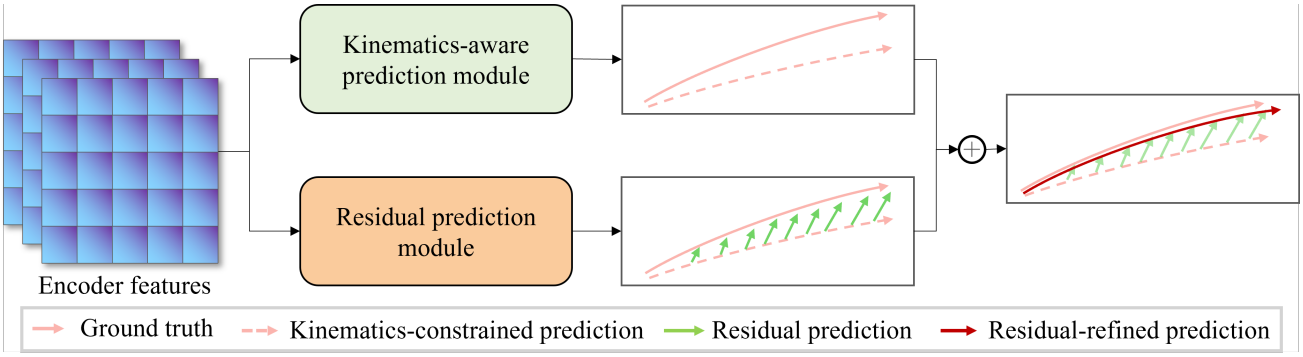


Fig 5.4 Illustration of residual prediction

A two-phase strategy is adopted for the training of the residual prediction module. In the first phase of training, the parameters of this module remain frozen, and the trajectory predictions from the kinematics-aware prediction module are directly used to calculate the loss and update the parameters. This frozen phase allows the model to initially focus on learning kinematically constrained trajectories. After the initial epochs, the parameters of the residual prediction module are unfrozen, and its primary objective is to correct the discrepancies between the predicted trajectories and the ground-truth trajectories. During this phase, the residual predictions are added to the outputs of the kinematics-aware prediction module, resulting in the final refined predictions.

5.2 Experimental Evaluation

5.2.1 Datasets

To evaluate the performance of our proposed KA-MGAT, we conduct comprehensive experiments on two challenging trajectory datasets: ApolloScape (S. Ma et al., 2019) and NGSIM. The ApolloScape dataset was collected in urban settings across various conditions and traffic densities. With 53 min of trajectory sequences annotated at 2 Hz, ApolloScape contains detailed annotations, including position, object size, heading, and category for vehicles, bicycles, and pedestrians. The NGSIM dataset comprises two subdatasets, I-80 and US-101, focusing on vehicle trajectories on highways. Each subdataset includes 45 min of vehicle trajectory data across diverse traffic conditions. All trajectories are recorded at 10 frames per second.

Following existing works (Carrasco et al., 2021; Fang et al., 2020; Y. Zhu et al., 2019), all approaches observe trajectories of 3 s to predict the subsequent 3 s in the ApolloScape dataset. For the NGSIM dataset, we follow existing works (Deo & Trivedi, 2018; T. Li et al., 2018; T. Zhao et al., 2019), downsampling the raw trajectory to 5 Hz, observing 3 s and predicting the next 5 s. The evaluation metrics are the average displacement error (ADE) and final displacement error (FDE). To balance the different speed ranges of cars, bicyclists, and pedestrians, ApolloScape also suggests a weighted sum of ADE (WSADE) and a weighted sum of FDE (WSFDE) as metrics, which are given by

$$\begin{aligned} \mathrm{WSADE} &= 0.2\times \mathrm{ADE_v} + 0.58\times \mathrm{ADE_p} + 0.22\times \mathrm{ADE_b} \\ \mathrm{WSFDE} &= 0.2\times \mathrm{FDE_v} + 0.58\times \mathrm{FDE_p} + 0.22\times \mathrm{FDE_b} \end{aligned} \tag{5.12}$$

where 'v', 'p', and 'b' represent the agent categories of cars, pedestrians, and bicyclists, respectively.

5.2.2 Implementation details

To augment the training dataset, we employ coordinate system rotation. This process involves rotating the coordinate axis with the origin as the center, resulting in an equivalent coordinate sequence in a new coordinate system. During training, we employ a 50% probability of using the original data axis. For the remaining 50% probability, we randomly select an angle within the range of 0 to 2π and rotate the coordinate axis accordingly. To handle the dynamic numbers of traffic agents in trajectory prediction, we set a large number as one of the dimensions in our feature tensor. This length is fixed and is designed to accommodate the maximum number of agents we expect to encounter. For each timestep, we allocate the first N positions of this dimension to fill the features of the observed traffic agents, where N is the number of agents present at that moment. The remaining positions are padded with zeros. We set the last frame of the historical frames as the observation frame. The set of agents contained in this observation frame is denoted as O. The set of agents contained in all frames is denoted as S. The set of agents present in the k-th frame is denoted as F_k. We categorize each agent $x \in S$ into three categories: (1) $x \in F_k$ and $x \in O$: For agents that are both present in frame k and visible in the observation frame, we append a 1 to their features as an indicator bit to indicate their presence and visibility. (2) $x \in F_k$ and $x \notin O$: Agents that are present in frame k but not visible in the observation frame are given a 0 after their features as an indicator bit, indicating that the data are invalid for prediction purposes. (3) $x \notin F_k$: Agents not present in frame k are entirely represented with zeros, including the indicator bit, signifying their nonexistence in this frame.

To apply convolutional operations in the temporal dimension, a 3×1 kernel size is used for the convolutional layer. Additionally, we set the hop of graphs to 2, which enables the model to consider interactions and dependencies among traffic agents across a broader range in the graph structure. For the data-learned attention mechanism, we set the number of attention heads to 3 to capture diverse and fine-grained interaction patterns among heterogeneous traffic agents. Utilizing multiple attention heads empowers the model to efficiently focus on various aspects of the input data, enabling it to acquire more informative and contextually relevant attention patterns directly from the data.

The proposed KA-MGAT aligns with the inherent uncertainty in predicting the movements of heterogeneous traffic agents in dynamic environments. Therefore, we use a negative log-likelihood loss function, which is given by

$$\mathrm{Loss} = -\sum_{i=1}^{N}\sum_{t=1}^{T_{\mathrm{pred}}} \log(\mathrm{P}(\boldsymbol{s}_t^i \mid \mu(\hat{\boldsymbol{s}}_t^i), \Sigma(\hat{\boldsymbol{s}}_t^i))) \tag{5.13}$$

This loss function is pivotal in guiding the model to learn and estimate the probability distribution of potential future trajectories rather than constraining it to predict a single deterministic path. It essentially enables the model to consider a spectrum of plausible trajectories, each grounded in historical data and contextual nuances. The proposed model is implemented with the PyTorch framework and executed on an NVIDIA GeForce 4090 GPU. We employ a batch size of 512 to update the model parameters. We train

the model using the Adam optimizer with an initial learning rate of 0.001, which is multiplied by 0.1 until loss converges.

5.2.3. Performance comparison

To verify the effectiveness and superiority of our proposed KA-MGAT, we compare it with representative baselines, including TrafficPredict (Y. Ma et al., 2019), S-LSTM (Alahi et al., 2016), S-GAN (Gupta et al., 2018), StarNet (Y. Zhu et al., 2019), GRIP++ (X. Li et al., 2019), SCOUT (Carrasco et al., 2021), TPNet (Fang et al., 2020), UNIN (Zheng et al., 2021), D2-Tpred (Y. Zhang et al., 2022), CS-LSTM (Deo & Trivedi, 2018), MATF (T. Zhao et al., 2019), MTDA (Guo et al., 2022), and AI-TP (K. Zhang et al., 2022).

Table 5.1 Comparison with baselines on the ApolloScape dataset (best results highlighted in bold (m))

Model	ADEv	ADEp	ADEb	WSADE	FDEv	FDEp	FDEb	WSFDE
TrafficPredict	7.94	7.18	12.88	8.59	12.77	11.12	22.79	24.23
S-LSTM	2.95	1.29	2.53	1.89	5.28	2.32	4.54	3.40
S-GAN	3.15	1.33	1.86	1.96	5.66	2.45	4.72	3.59
StarNet	2.39	0.78	1.86	1.34	4.28	1.51	3.46	2.49
GRIP++	2.24	0.71	1.80	1.25	4.07	1.37	3.41	2.36
SCOUT	2.21	0.73	1.82	1.26	3.93	1.41	3.37	2.35
TPNet	2.21	0.74	1.85	1.28	3.86	1.41	3.40	1.91
UNIN	—	—	—	1.09	—	—	—	1.55
AI-TP	—	—	—	1.15	—	—	—	2.13
D2-Tpred	—	—	—	1.02	—	—	—	1.69
KA-MGAT (Ours)	**1.58**	**0.57**	**1.21**	**0.91**	**2.55**	**0.91**	**1.87**	**1.45**

Table 5.1 shows the comparison of our method against all baselines in terms of the automated driving environment (ADE), FDE, WSADE, and WSFDE on the ApolloScape dataset. Overall, our method exhibits superior performance compared to all baseline approaches on all the metrics. In particular, our proposed method outperforms D2-TPred (the second best) by 10% in WSADE and 8% in WSFDE. We speculate that our method can leverage multigraph attention topologies to explicitly model social interactions in a more comprehensive way and apply prior physics models to explicitly learn the kinematic constraints of agents in different categories.

According to the prediction results in Table 5.1, our proposed approach demonstrates superior performance compared to LSTM-based methods, such as S-LSTM, SGAN, and StarNet, in both the WSFDE and WSADE metrics. This advantage can be attributed to the explicit consideration of the heterogeneity of traffic agents, allowing our model to capture more comprehensive interactions. Furthermore, compared to graph-based methods such as GRIP++, our approach outperforms it by 27% and 38% in the WSADE and WSFDE metrics, respectively. The improved performance can be attributed to the multigraph attention mechanism in our model, which enables the modeling of complex and diverse interactions among traffic agents by utilizing both prior knowledge-based fixed attention and data-learned attention. Additionally, our method surpasses TPNet by 28% and 24% in the WSFDE and WSADE

metrics, respectively. This performance improvement can be attributed to our use of physics models to guide trajectory prediction, ensuring compliance with kinematic constraints in the predicted trajectories. In contrast, the reliance on polynomial curves of TPNet may lead to inaccuracies in certain cases, impacting the overall trajectory quality.

Table 5.2 Comparison with baselines on the NGSIM dataset (best results highlighted in bold (m))

Model	FDE@1 s	FDE@2 s	FDE@3 s	FDE@4 s	FDE@5 s	ADE
CS-LSTM	0.58	1.32	2.22	3.26	4.40	2.36
MATF	0.66	1.34	2.08	2.97	4.13	2.24
MTDA	0.50	1.11	1.78	2.69	3.93	2.00
GRIP++	**0.38**	**0.89**	1.45	2.14	2.94	1.56
AI-TP	0.47	1.05	1.53	1.93	2.31	1.46
KA-MGAT (Ours)	0.44	0.98	**1.44**	**1.92**	**2.24**	**1.40**

Table 5.2 compares the prediction results of each model on the NGSIM dataset, evaluating the ADE and FDEs at various prediction time steps. Notably, our proposed KA-MGAT demonstrates superior performance across both the ADE and most of the FDEs. Compared with CS-LSTM, MATF, and MTDA, our network exhibits better predictive results in terms of the ADE and all FDEs, showing over 30% performance improvement in the ADE. This underscores the efficacy of graph convolution in capturing social interactions over LSTM. In contrast to GRIP++ and AI-TP, our KA-MGAT achieves a performance improvement of 10% and 4%, respectively, highlighting the effectiveness of our kinematics-aware prediction and multigraph attention approach. Notably, GRIP++ outperforms KA-MGAT when the prediction horizon is less than 2 s, but KA-MGAT exhibits better performance beyond this threshold. This observed trend in performance can be attributed to KA-MGAT's incorporation of kinematic models, which enhance the stability of predictions at longer horizons. The kinematic model enables KA-MGAT to capture and infer complex trajectory dynamics more effectively, resulting in improved accuracy and outperformance over GRIP++ as the prediction horizon increases. This strategic integration of physics models demonstrates KA-MGAT's adaptability and reliability in handling extended prediction scenarios, contributing to its overall enhanced predictive capabilities.

5.2.4 Discussion of proposed components

In this subsection, we conduct ablation experiments to investigate the individual contributions of the proposed modules in enhancing heterogeneous trajectory prediction. As presented in Table 3, we train and evaluate four variant models, each incrementally incorporating our proposed modules to systematically assess specific aspects of our approach:

- V1 Model: This baseline model relies solely on a distance graph (DG), excluding the incorporation of kinematic models or the residual prediction module.
- V2 Model: Building upon V1, the second variant model introduces the proposed prior knowledge-based fixed attention multigraph topologies (denoted as PKG).
- V3 Model: In this variant, we integrate the proposed data-learned attention mechanism (DLG) into the model. However, it remains kinematics-agnostic and residual-agnostic.

- V4 Model: The fourth variant model incorporates the kinematics-aware (KA) prediction module, but it remains residual-agnostic, excluding the contribution of the residual prediction (RP) module.

As presented in Table 5.3, it is evident that our fully proposed model stands out as the top performer among all variant models. This result demonstrates the effectiveness of our multigraph attention topology, kinematics-aware prediction module, and residual prediction module for advancing trajectory prediction accuracy. The combination of these well-designed modules enables our model to capture the complex interactions among diverse traffic agents. Moreover, the incorporation of the residual prediction module contributes significantly to refining trajectory predictions, further enhancing the overall predictive capabilities of our model.

Table 5.3 Comparison of variant models on the ApolloScape dataset (Unit: m)

Index	DG	PKG	DLG	KA	RP	WSADE
V1	☑	×	×	×	×	1.34
V2	✓	☑	×	×	×	1.25
V3	✓	✓	☑	×	×	1.03
V4	✓	✓	✓	☑	×	0.95
Full model	✓	✓	✓	✓	☑	**0.91**

To further illustrate the impact of the kinematics-aware prediction module, we provide trajectory prediction visualizations in Fig. 5.5. It is evident that both models perform well in accurately predicting trajectories for agents moving in a straight line. However, when faced with the more challenging task of predicting curved trajectories, such as lane change and turning behaviors, the kinematics-aware model exhibits a substantial performance improvement over its kinematics-agnostic counterpart. For instance, we can see that agent 3 in Fig. 5.5a, agent 1 in Fig. 5.5b, agent 2 in Fig. 5.5c, and agent 6 in Fig. 5.5d all engage in turning maneuvers. Notably, the kinematics-aware model demonstrates a remarkable ability to accurately predict the future trajectories of these agents engaged in turning maneuvers.

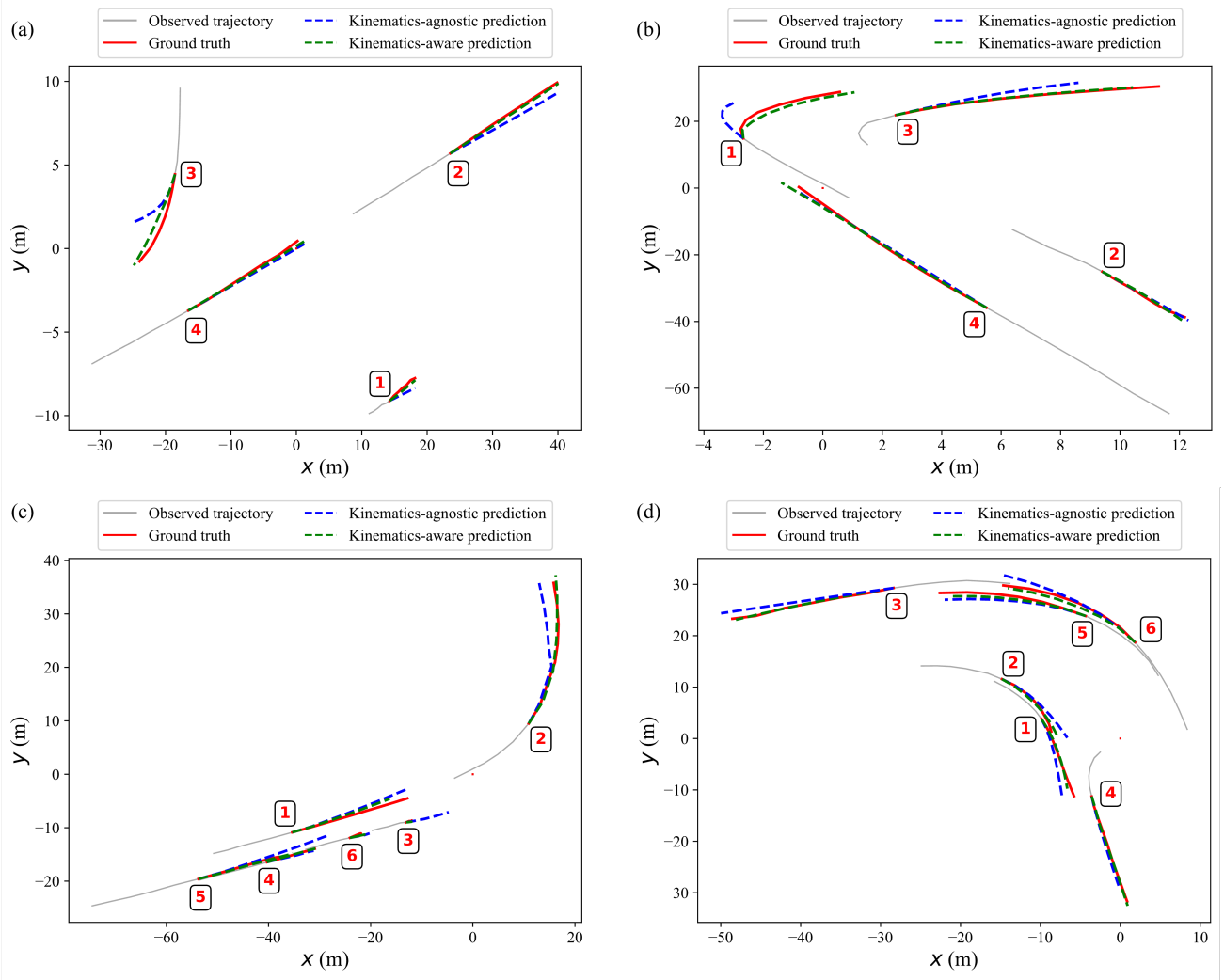


Fig 5.5 Visualization of the predicted trajectories of kinematic-agnostic/aware predictions.

The numbers represent the indices of the corresponding traffic agents.

The enhanced prediction performance of the kinematics-aware model for turning behaviors can be attributed to its explicit integration of physics-based constraints. Turning behaviors involve complex kinematic patterns, such as changes in velocity, acceleration, and angular motion. Unlike kinematics-agnostic models, the kinematics-aware model, which is explicitly aware of these physical constraints, can better capture and model the complexities of turning trajectories. By explicitly incorporating these constraints, the model provides insights into the expected behavior of turning trajectories, facilitating improved generalizability and predictive accuracy when faced with new, unseen turning scenarios. This adaptability is crucial in handling a diverse range of turning radii, from sharp turns to gradual curves. Consequently, the kinematics-aware model demonstrates superior predictive accuracy and realism in scenarios involving turning maneuvers, demonstrating the efficacy of explicitly incorporating physics constraints into trajectory prediction frameworks.

As mentioned before, the PIML paradigm is expected to improve learning efficiency, so we conducted two experiments to verify its effectiveness. The first experiment aims to compare the convergence speed of the prediction accuracy between our proposed kinematics-aware model and other variant models. As shown in Fig. 5.6, our kinematics-aware model exhibits the fastest convergence among all the models. More specifically, our model only requires approximately 120 epochs to reach the best prediction performance of the V3 model, saving 60% of the training epochs needed. The rapid convergence of our fully proposed model can be attributed to its ability to effectively leverage the underlying physics-based constraints encoded in the kinematic models. By doing so, our model can better align the predictions with real-world constraints and converge more efficiently than models that rely solely on implicit learning from raw data.

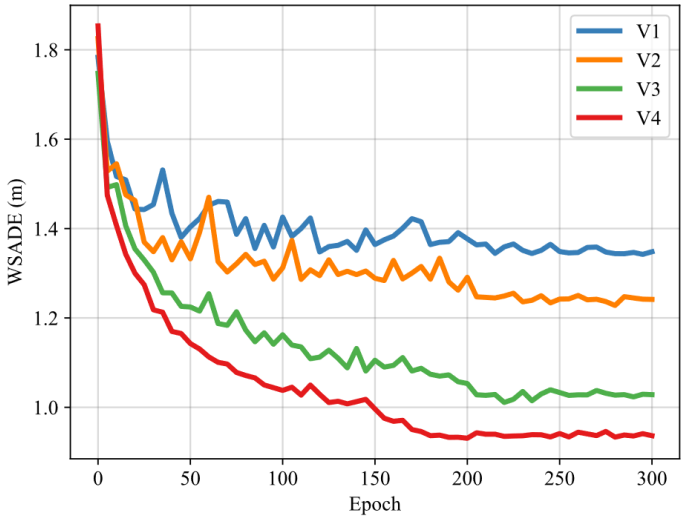


Fig 5.6 Training curves of all variant models

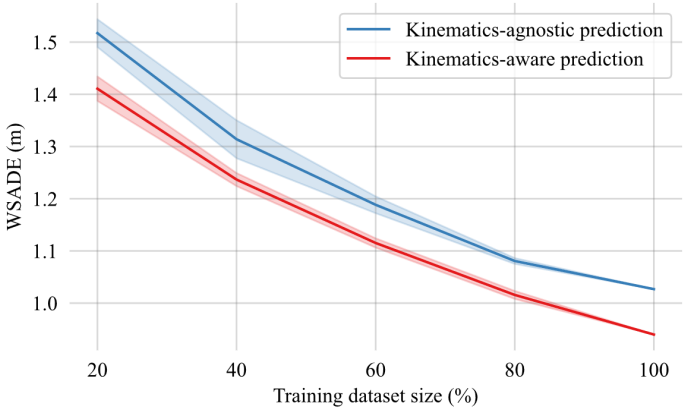


Fig 5.7 Prediction performance versus training dataset size

The second experiment investigated the data efficiency of both the kinematics-aware and kinematics-agnostic models. In this analysis, the models were trained with 20%, 40%, 60%, and 80% randomly selected data from the entire training set, followed by evaluation on the entire test set. This

process was repeated three times for each training dataset. The results of the data efficiency experiments are presented in Fig 5.7, where the solid lines denote the mean values and the shaded regions depict the standard deviations. Notably, Fig 5.7 illustrates that the kinematics-aware model demonstrates remarkable efficiency, requiring 25% less data to achieve an equivalent level of prediction accuracy compared to its kinematics-agnostic counterpart. This efficiency gain can be attributed to the innovative kinematics-aware module in our model, which explicitly integrates prior knowledge of kinematic models to guide the learning process. In contrast, the kinematics-agnostic model relies on implicit learning from raw data, necessitating a larger dataset to achieve comparable performance. This finding underscores the effectiveness of our proposed kinematics-aware approach in enhancing data efficiency and predictive capabilities.

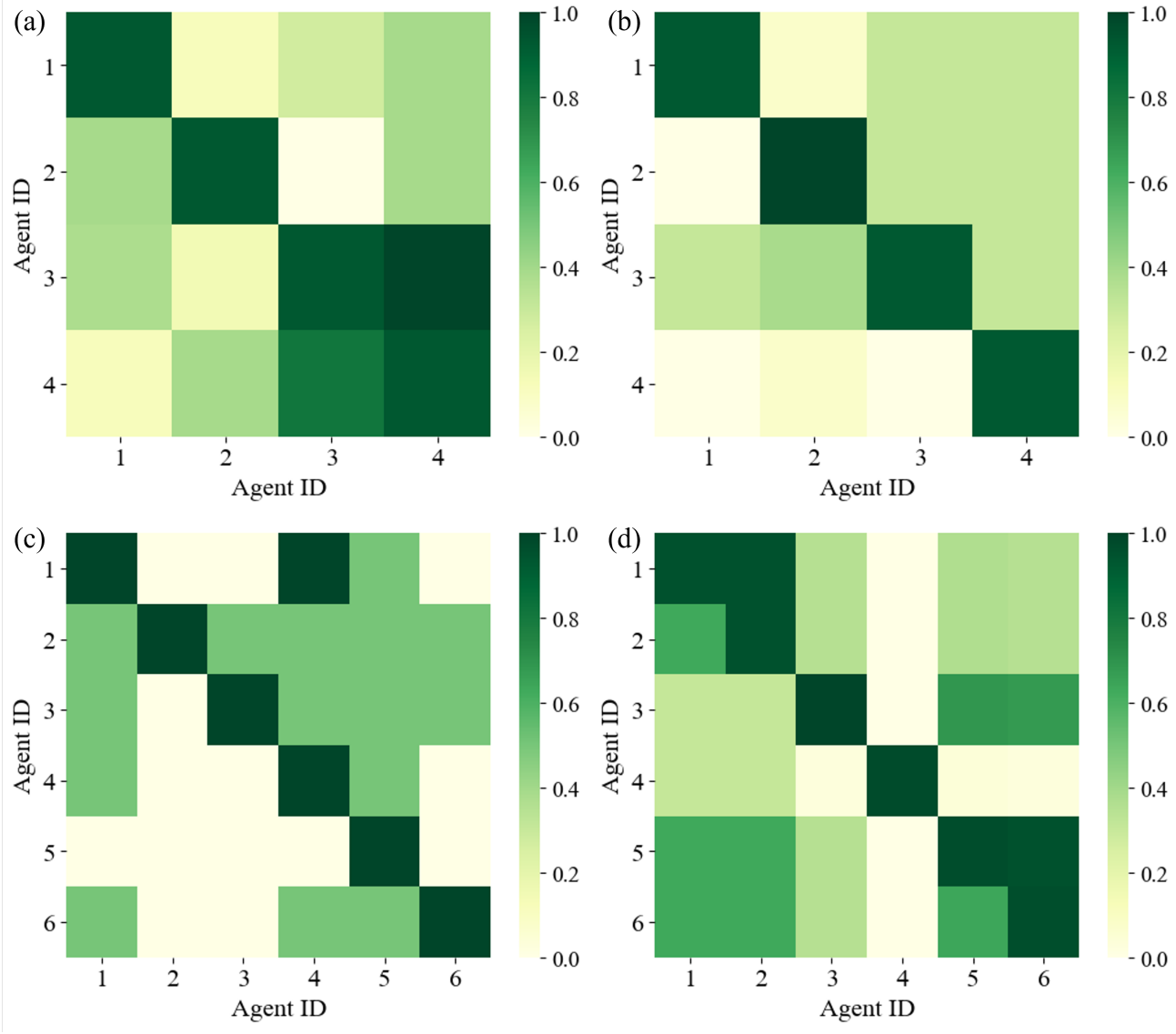


Fig 5.8 Visualization of attention matrices among traffic agents

where the color bar represents the normalized attention scores of the multigraph attention module.

Next, we analyze the attention matrices generated by our model to understand how the network allocates attention across different traffic agents during trajectory prediction. In safety-critical applications, particularly in the context of automated driving, interpretability is crucial because decisions made by automated systems directly impact human safety (Dong et al., 2022, 2023). The visualization of attention matrices enhances the interpretability of the decision-making process of our proposed multigraph attention module. By visualizing the attention distribution across the input space, we can obtain a more granular understanding of how the model processes and prioritizes information.

Fig 5.8 illustrates the normalized attention matrices of traffic agents, with subfigures a, b, c, and d corresponding to the scenarios depicted in Fig 5.5a, b, c, and d, respectively. Within each attention matrix, the i-th column represents how the model allocates attention weights to the traffic agents surrounding agent i. It is worth noting that each traffic agent holds the highest attention weight on itself, highlighting the significance of an agent's historical trajectories as the most informative feature for predicting its future trajectories. As shown in Fig 5.8a, we can see that the attention weights between agents 3 and 4 are greater than those between agents 3 and 4. This observation aligns with the traffic scene in Fig. 5.5a, where these two traffic agents are close to each other, and agent 3 is turning right toward the direction of agent 4. Thus, their attention weights signify increased mutual awareness, demonstrating the model's capacity to adapt attention based on spatial interactions. Additionally, we also notice that agent 2 in Fig 5.8c and agent 4 in Fig. 5.8d allocate minimal attention to surrounding traffic agents. This finding corresponds with the traffic scenes in Figs 5.5c and 5.5d, where these two agents are moving in front of each other. Their reduced attention to surrounding agents is consistent with the expectation that agents at the forefront experience less impact from those behind them.

These observations demonstrate that the attention matrices could offer a transparent view of which agents or regions most influence its outputs and providing interpretable insights into the model's decision-making process. Such interpretability is particularly valuable for understanding interactions between automated vehicles and VRUs, enabling safer and more trustworthy behavior prediction in connected transportation environments.

5.3 Discussion

The proposed KA-MGAT framework confirms that fusing physical motion constraints with graph-based attention optimizes trajectory prediction for heterogeneous agents. Unlike purely data-driven approaches, this model explicitly encodes kinematic properties. The result is a prediction output that is both accurate and physically plausible, resolving the interpretability issues common in baseline methods.

Architecturally, KA-MGAT functions as a core module within the project's connected ecosystem. We designed it for direct integration into the Sky-Drive simulation. This compatibility supports real-time interaction studies, allowing the system to process behavioral data derived from virtual VRU–AV experiments

Nevertheless, several challenges remain. First, the model currently relies on offline trajectory datasets and does not yet leverage real-time C-V2X signals or cooperative perception data. Future work should focus on coupling the model with the CV2X-LOCA localization framework to achieve fully connected prediction and warning capabilities. Second, while the residual learning module effectively

refines short-horizon forecasts, long-horizon prediction under uncertain interaction intent still presents difficulties. Extending the architecture to probabilistic or diffusion-based formulations could further improve robustness.

Overall, KA-MGAT provides a scalable, physically consistent, and interpretable foundation for trajectory forecasting in mixed traffic environments, contributing to the vision of proactive and equitable safety for all vulnerable road users in connected and automated transportation systems.

CHAPTER 6 SYNOPSIS OF PERFORMANCE INDICATORS

6.1 USDOT Performance Indicators I

Two (2) transportation-related courses were offered during the study period that were taught by the PI and teaching assistants who are associated with the research project. These courses include CEE 679 (AI & Data Science in Transportation), and CEE 370 (Transportation Engineering). Three (3) graduate students participated in the research project during the study period. One (1) transportation-related advanced degree program (a doctoral program) utilized the CCAT grant funds from this research project during the study period to support the graduate students.

6.2 USDOT Performance Indicators II

Leadership Development Performance Indicators: This research project generated 3 academic engagements and 2 industry engagements. The PI held positions in 3 national organizations that address issues related to this research project, including ASCE national committees: Connected & Autonomous Vehicle Impacts, and Economics & Finance; and IEEE Emerging Transportation Technology Testing Technical Committee

Education and Workforce Development Performance Indicators: The methods, data, and/or results from this study were incorporated in the class content of the following courses at the University of Wisconsin-Madison: (a) CEE 679: AI & Data Science in Transportation, a graduate-level course in the transportation engineering program, and (b) CEE 370: Transportation Engineering, an undergraduate-level course. The students in these classes will soon be entering the workforce. Thereby, the research helped enlarge the pool of people trained to develop knowledge and utilize at least a part of the technologies developed in this research, and to put them to use when they enter the workforce.

CHAPTER 7 STUDY OUTCOMES AND OUTPUTS

7.1 Publications and Conference Papers

The research findings from this project have been disseminated through multiple peer-reviewed publications and conference presentations, ensuring broad reach to both the academic community and transportation practitioners. Two journal articles have been published or accepted in high-impact journals. The first paper (Huang et al., 2025) is accepted by Journal of Intelligent and Connected Vehicles, describes the distributed Sky-Drive multi-agent simulation platform architecture and demonstrates its application to VRU-AV interaction research. The second paper by Sheng (Sheng et al., 2024) in the Journal of Intelligent and Connected Vehicles introduces the KA-MGAT model that combines physics-based kinematic constraints with deep learning for accurate and physically plausible trajectory prediction of heterogeneous traffic agents. Please refer to the appendix for details. We also presented our findings at key conferences, such as Transportation Research Board (TRB) Annual Meeting, International Conference on Transportation & Development (ICTD). These sessions attracted a broad mix of researchers, practitioners, and government officials. Engaging with this diverse group allowed our team to gather feedback and identify opportunities for collaborative follow-up.

7.2 Open-Source Software and Data Products

To maximize the impact and enable replication of research findings, the project team has released multiple open-source software platforms and datasets. The Sky-Drive platform represents a distributed multi-agent simulation framework integrating VR, traffic microsimulation (SUMO), and 3D visualization (Unreal Engine). For the details of Sky-Drive, please refer to our website. (https://sky-lab-uw.github.io/Sky-Drive-website/). Furthermore, we consolidated the project's output into several distinct datasets. The localization dataset benchmarks C2X-LOCA across eight scenarios (3.6–125 km/h), providing granular metrics. For trajectory prediction, separate files quantify KA-MGAT accuracy (ADE/FDE) using ApolloScape and NGSIM. These records validate the kinematic approach through ablation studies and visual comparisons against baselines during complex maneuvers.

7.3 Educational Materials and Resources

Research findings have been integrated into course curricula at the University of Wisconsin-Madison. CEE 679 (AI & Data Science in Transportation) now includes modules on physics-informed machine learning, VR simulation for transportation research, and cooperative localization technologies. CEE 370 (Transportation Engineering) features updated content on VRU safety in intelligent transportation

environments. Course materials including lecture slides, laboratory exercises, and project assignments have been developed and will be made available to other educators through the CCAT network.

Comprehensive technical documentation has been prepared to facilitate technology adoption and replication. This documentation includes system architecture documents for the Sky-Drive platform, experimental protocols for VR-based VRU-AV interaction studies, and best practices guidelines for digital twin development in transportation applications.

7.4 Technology Demonstrations and Prototypes

The project has produced several working technology demonstrations and will conduct several pilot deployments in the future. We have established a prototype of the Beltline Digital for Dane County, WI and will deploy it on a major corridor in Dane County, in coordination with WisDOT. This system will fuse real-time WisDOT 511 feeds with WisTransPortal archives, creating a dynamic replica used for incident detection and scenario simulation. We also established a distributed multi-agent simulation site at UW-Madison which supports continued research on VRU-AV interaction and other human-centered research. Key equipment include: HTC Vive Pro Eye headsets (for eye tracking), Logitech G920 wheels, and Garmin vívoactive 5 watches. Additionally, we integrated C-V2X modules to support field testing in Madison, WI.

7.5 Workforce Development and Human Capital

Three PhD students acquired specialized skills in VR simulation and cooperative positioning through this project. For undergraduate and graduate level exposure, research outcomes were integrated into UW-Madison courses (CEE 679/CEE 370). This ensured that future engineers encountered VRU safety challenges during their formal education. Knowledge transfer also reached the professional sector; through workshops and presentations, practitioners gained exposure to the project's interaction methodologies

7.6 Impacts

- **Impacts on Transportation Systems and Safety:** This project improves the safety and operation of mixed traffic environments involving autonomous vehicles, human-driven vehicles, and vulnerable road users. By enabling lane-level VRU localization in GNSS-denied urban areas and improving heterogeneous trajectory prediction, the proposed system enhances conflict detection and proactive safety responses. These capabilities are particularly relevant at intersections and other urban locations with high VRU crash risk.
- **Impacts on Knowledge and Technology Advancement: This** research advances connected and automated transportation through new platforms, algorithms, and validated methodologies. The Sky-Drive platform enables safe, repeatable, human-in-the-loop evaluation of safety-critical scenarios. The CV2X-LOCA framework and kinematics-aware trajectory prediction approach

address key gaps in VRU localization and heterogeneous motion forecasting, expanding the technical foundation for future research and development.

- **Impacts on Policy, Practice, and Deployment Readiness:** The project provides evidence-based insights for transportation agencies and policymakers. Results from C-V2X-based cooperative localization support infrastructure planning and connected safety system deployment. The VR-based evaluation framework offers a practical tool for assessing AV–VRU interaction strategies and safety policies prior to real-world implementation.
- **Impacts on Workforce Development and Education:** The project contributed to training graduate researchers in connected vehicle systems, simulation, and AI-based safety analysis. Students gained hands-on experience with distributed simulation, wireless communication, and human-in-the-loop experimentation. The developed software tools and educational materials support future workforce development in intelligent transportation systems.

CHAPTER 8 CONCLUDING REMARKS

8.1 Summary of Key Findings

8.1.1 Distributed Multi-Agent Simulation Platform

The Sky-Drive platform successfully demonstrated the feasibility of creating realistic, immersive, and safe experimental environments for studying VRU-AV interactions. The distributed multi-agent architecture enabled synchronized simulation across multiple terminals, allowing independent control of AVs, human-driven vehicles, and VRUs while maintaining a shared virtual environment. The multi-modal human-in-the-loop framework successfully captured comprehensive behavioral data including gaze patterns, physiological signals, facial expressions, voice commands, and control actions. This rich dataset enables deeper understanding of human decision-making processes in safety-critical scenarios that cannot be studied through field testing. Case studies on right-turn conflicts at unsignalized intersections revealed critical insights into VRU perception of AV behavior. The VR-based experiments captured scenarios where pedestrians hesitated or exhibited increased cognitive workload when interacting with AVs lacking clear communication signals. These findings underscore the importance of explicit communication mechanisms in VRU-AV interactions, particularly at locations with ambiguous right-of-way.

8.1.2 CV2X-Based Cooperative Localization

The CV2X-LOCA framework demonstrated that lane-level positioning accuracy for VRUs is achievable in GNSS-denied environments using only wireless channel state information. Simulation results across four environment types (semi-open, urban forests, urban canyon, and tunnel) showed that CV2X-LOCA consistently outperformed baseline methods, achieving 20-30% error reduction compared to state-of-the-art approaches such as SDP-ML-KF. Analysis of RSU deployment spacing revealed that 120-150m spacing provides an optimal balance between positioning accuracy and infrastructure cost, with CV2X-LOCA maintaining acceptable accuracy even at 150m spacing. This finding has important implications for transportation agencies planning C-V2X infrastructure deployment.

8.1.3 Physics-Informed Trajectory Prediction

The KA-MGAT framework successfully integrated kinematic constraints explicitly into deep learning architectures for heterogeneous trajectory prediction. Experimental results on the ApolloScape and NGSIM datasets demonstrated that the physics-informed approach achieved state-of-the-art prediction accuracy while exhibiting improved learning efficiency compared to purely data-driven methods. Data efficiency analysis revealed a key advantage of the physics-informed approach: KA-MGAT required only 25% of the training data to achieve accuracy comparable to kinematics-agnostic baselines. This finding has practical implications, as it suggests that physics-informed models can be deployed effectively in data-scarce scenarios or adapted to new environments with limited training data.

8.2 Implications for Practice

8.2.1 Transportation Agency Applications

VRU Safety Assessments: The Sky-Drive platform is expected to provide transportation agencies with a powerful tool for conducting VRU Safety Assessments. Agencies can use the platform to evaluate proposed infrastructure modifications, test different intersection designs, and assess the effectiveness of VRU warning systems before committing to expensive physical implementations.

Infrastructure Planning and Design: The CV2X-LOCA findings regarding optimal RSU deployment spacing (120-150m) provide concrete guidance for agencies planning C-V2X infrastructure investments. The environment-specific propagation characteristics identified in this research can inform deployment strategies tailored to local road conditions (urban canyons vs. open highways).

8.2.2 Autonomous Vehicle Development

Testing and Validation: AV developers can leverage the Sky-Drive platform for scenario-based testing involving human VRUs, complementing conventional simulation approaches that rely on pre-recorded or synthetic VRU behaviors. The ability to expose human participants to identical scenarios repeatedly enables statistically robust evaluation of AV decision-making.

Perception System Development: The C-V2X localization capabilities demonstrated by CV2X-LOCA provide AV developers with an alternative or complementary approach to vision-based and LiDAR-based VRU detection. By fusing C-V2X-based position information with onboard sensors, AVs can achieve more robust VRU detection, particularly in occlusion scenarios where direct line-of-sight is unavailable.

Motion Planning Enhancement: The KA-MGAT trajectory prediction framework can be integrated into AV motion planning systems to improve anticipation of VRU movements. The physics-informed approach ensures predictions remain plausible even in novel scenarios, addressing concerns about safety-critical prediction failures.

8.2.3 Policy and Regulatory Considerations

Communication Standards: The research findings underscore the importance of standardized communication protocols for VRU-AV interactions. Policymakers should encourage development and adoption of standards for V2P (vehicle-to-pedestrian) communication, building on existing V2X frameworks.

Privacy and Data Governance: As cooperative localization and connected VRU technologies become more prevalent, clear policies regarding location data privacy, retention, and use are essential. The local computation approach employed by CV2X-LOCA represents a privacy-preserving design that can serve as a model for policy development.

Infrastructure Investment Priorities: Research findings regarding RSU deployment density and environment-specific challenges can inform prioritization of infrastructure investments, focusing initially on high-risk locations (urban intersections with high VRU volumes) and challenging environments (urban canyons, tunnels) where benefits are greatest.

8.3 Concluding Statement

As autonomous vehicle technology continues to advance and deployment expands, ensuring the safety and equitable treatment of VRUs must remain a paramount priority. The traditional communication channels that VRUs have relied upon for over a century, such as eye contact, gestures, facial expressions,

become obsolete in the absence of a human driver. New paradigms for VRU-AV interaction are essential, and these paradigms must be grounded in rigorous research using appropriate methodologies and validated through careful experimentation.

This research project demonstrates that the convergence of virtual reality simulation, cooperative wireless positioning, and physics-informed machine learning creates powerful new capabilities for enhancing VRU safety in the era of connected and automated vehicles. The Sky-Drive platform enables safe, repeatable, and data-rich studies of VRU-AV interactions that were previously impossible. The CV2X-LOCA framework provides practical solutions for lane-level VRU positioning in challenging environments where conventional GNSS fails. The KA-MGAT approach combines the interpretability of physics-based models with the predictive power of deep learning for heterogeneous trajectory prediction.

Importantly, these three components are not merely independent technical contributions, but synergistic elements of a comprehensive approach to VRU safety. High-fidelity simulation enables development and testing of localization and prediction algorithms under diverse conditions. Accurate localization provides essential inputs for trajectory prediction and collision risk assessment. Reliable prediction enables proactive safety interventions and comfortable, efficient vehicle control.

The research presented in this report provides a foundation for new paradigm, but much work remains. Continued research, thoughtful policy development, infrastructure investment, and collaboration among researchers, industry, government, and affected communities will be necessary to realize the vision of a transportation system that provides both the efficiency benefits of automation and the safety and mobility that all road users deserve.

REFERENCES

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social lstm: Human trajectory prediction in crowded spaces. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 961–971.

Alam, N., Tabatabaei Balaei, A., & Dempster, A. G. (2013). Relative Positioning Enhancement in VANETs: A Tight Integration Approach. *IEEE Transactions on Intelligent Transportation Systems*, *14*(1), 47–55. https://doi.org/10.1109/TITS.2012.2205381

Bella, F. (2008). Driving simulator for speed research on two-lane rural roads. *Accident Analysis & Prevention*, *40*(3), 1078–1087. https://doi.org/10.1016/j.aap.2007.10.015

Bhattacharyya, A., Schiele, B., & Fritz, M. (2018). Accurate and diverse sampling of sequences based on a "best of many" sample objective. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8485–8493.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., & Leonard, J. J. (2016). Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, *32*(6), 1309–1332. https://doi.org/10.1109/TRO.2016.2624754

Carrasco, S., Llorca, D. F., & Sotelo, M. (2021). SCOUT: Socially-consistent and understandable graph attention network for trajectory prediction of vehicles and VRUs. *2021 IEEE Intelligent Vehicles Symposium (IV)*, 1501–1508.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, S., Hu, J., Shi, Y., Peng, Y., Fang, J., Zhao, R., & Zhao, L. (2017). Vehicle-to-Everything (v2x) Services Supported by LTE-Based Systems and 5G. *IEEE Communications Standards Magazine*, *1*(2), 70–76. https://doi.org/10.1109/MCOMSTD.2017.1700015

Chen, S., Zong, S., Chen, T., Huang, Z., Chen, Y., & Labi, S. (2023). A Taxonomy for Autonomous Vehicles Considering Ambient Road Infrastructure. *Sustainability*, *15*(14), 11258. https://doi.org/10.3390/su151411258

Colley, M., Walch, M., & Rukzio, E. (2020). Unveiling the Lack of Scalability in Research on External Communication of Autonomous Vehicles. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3334480.3382865

De Angelis, G., Baruffa, G., & Cacopardi, S. (2013). GNSS/Cellular Hybrid Positioning System for

Mobile Users in Urban Scenarios. *IEEE Transactions on Intelligent Transportation Systems*, *14*(1), 313–321. https://doi.org/10.1109/TITS.2012.2215855

de Winter, J. C. F., de Groot, S., Mulder, M., Wieringa, P. A., Dankelman, J., & Mulder, J. A. (2009). Relationships between driving simulator performance and driving test results. *Ergonomics*, *52*(2), 137–153. https://doi.org/10.1080/00140130802277521

Deb, S., Carruth, D. W., Sween, R., Strawderman, L., & Garrison, T. M. (2017). Efficacy of virtual reality in pedestrian safety research. *Applied Ergonomics*, *65*, 449–460. https://doi.org/10.1016/j.apergo.2017.03.007

Deb, S., Strawderman, L., Carruth, D. W., DuBien, J., Smith, B., & Garrison, T. M. (2017). Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *84*, 178–195. https://doi.org/10.1016/j.trc.2017.08.029

Deo, N., & Trivedi, M. M. (2018). Convolutional social pooling for vehicle trajectory prediction. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 1468–1476.

Eigen, D., Puhrsch, C., & Fergus, R. (2014). Depth Map Prediction from a Single Image using a Multi-Scale Deep Network. *Advances in Neural Information Processing Systems*, *27*. https://proceedings.neurips.cc/paper_files/paper/2014/hash/91c56ce4a249fae5419b90cba831e303-Abstract.html

Emara, K., Woerndl, W., & Schlichter, J. (2015). CAPS: Context-aware privacy scheme for VANET safety applications. *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, 1–12. https://doi.org/10.1145/2766498.2766500

European Commission. (2022). *Annual Accident Report 2022*. Directorate-General for Mobility and Transport.

Fang, L., Jiang, Q., Shi, J., & Zhou, B. (2020). Tpnet: Trajectory proposal network for motion prediction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6797–6806.

Feldstein, I., Dietrich, A., Milinkovic, S., & Bengler, K. (2016). A Pedestrian Simulator for Urban Crossing Scenarios. *IFAC-PapersOnLine*, *49*(19), 239–244. https://doi.org/10.1016/j.ifacol.2016.10.531

Fernández, P. D., Lindman, M., Isaksson-Hellman, I., Jeppsson, H., & Kovaceva, J. (2022). Description of same-direction car-to-bicycle crash scenarios using real-world data from Sweden, Germany, and a global crash database. *Accident Analysis & Prevention*, *168*, 106587.

Giuliari, F., Hasan, I., Cristani, M., & Galasso, F. (2021). Transformer networks for trajectory

forecasting. *2020 25th International Conference on Pattern Recognition (ICPR)*, 10335–10342.

González, D., Pérez, J., Milanés, V., & Nashashibi, F. (2016). A Review of Motion Planning Techniques for Automated Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *17*(4), 1135–1145. https://doi.org/10.1109/TITS.2015.2498841

Grewal, M. S., Weill, L. R., & Andrews, A. P. (2007). *Global Positioning Systems, Inertial Navigation, and Integration*. John Wiley & Sons.

Groves, P. D. (2011). Shadow Matching: A New GNSS Positioning Technique for Urban Canyons. *The Journal of Navigation*, *64*(3), 417–430. https://doi.org/10.1017/S0373463311000087

Guan, Q., Sheng, Z., & Xue, S. (2023). HRPose: Real-time high-resolution 6D pose estimation network using knowledge distillation. *Chinese Journal of Electronics*, *32*(1), 189–198.

Guo, H., Meng, Q., Cao, D., Chen, H., Liu, J., & Shang, B. (2022). Vehicle trajectory prediction method coupled with ego vehicle motion trend under dual attention mechanism. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–16.

Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., & Alahi, A. (2018). Social gan: Socially acceptable trajectories with generative adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2255–2264.

Harle, R. (2013). A Survey of Indoor Inertial Positioning Systems for Pedestrians. *IEEE Communications Surveys & Tutorials*, *15*(3), 1281–1293. https://doi.org/10.1109/SURV.2012.121912.00075

Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, *51*(5), 4282.

Holländer, K., Colley, A., Mai, C., Häkkilä, J., Alt, F., & Pfleging, B. (2019). Investigating the Influence of External Car Displays on Pedestrians' Crossing Behavior in Virtual Reality. *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–11. https://doi.org/10.1145/3338286.3340138

Hsu, L.-T., Gu, Y., & Kamijo, S. (2015). NLOS Correction/Exclusion for GNSS Measurement Using RAIM and City Building Models. *Sensors*, *15*(7), 17329–17349. https://doi.org/10.3390/s150717329

Huang, Z., Chen, S., Pian, Y., Sheng, Z., Ahn, S., & Noyce, D. A. (2024). Toward C-V2X Enabled Connected Transportation System: RSU-Based Cooperative Localization Framework for Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *25*(10), 13417–13431. https://doi.org/10.1109/TITS.2024.3410185

Huang, Z., Sheng, Z., Wan, Z., Qu, Y., Luo, Y., Wang, B., Li, P., Chen, Y.-J., Chen, J., Long, K., &

others. (2025). Sky-Drive: A Distributed Multi-Agent Simulation Platform for Human-AI Collaborative and Socially-Aware Future Transportation. *Journal of Intelligent and Connected Vehicles*.

Huang, Z., Xu, L., & Lin, Y. (2020). Multi-stage pedestrian positioning using filtered WiFi scanner data in an urban road environment. *Sensors*, *20*(11), 3259.

Huang, Z., Zhu, X., Lin, Y., Xu, L., & Mao, Y. (2019). A novel WIFI-oriented RSSI signal processing method for tracking low-speed pedestrians. *2019 5th International Conference on Transportation Information and Safety (ICTIS)*, 1018–1023.

Jayaraman, S. K., Creech, C., Robert Jr., L. P., Tilbury, D. M., Yang, X. J., Pradhan, A. K., & Tsui, K. M. (2018). Trust in AV: An Uncertainty Reduction Model of AV-Pedestrian Interactions. *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 133–134. https://doi.org/10.1145/3173386.3177073

Kamal, M., Srivastava, G., & Tariq, M. (2021). Blockchain-Based Lightweight and Secured V2V Communication in the Internet of Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), 3997–4004. https://doi.org/10.1109/TITS.2020.3002462

Kaplan, E. D., & Hegarty, C. (2017). *Understanding GPS/GNSS: Principles and Applications, Third Edition*. Artech House.

Karamouzas, I., Skinner, B., & Guy, S. J. (2014). Universal power law governing pedestrian interactions. *Physical Review Letters*, *113*(23), 238701.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440.

Kendall, A., & Cipolla, R. (2017). *Geometric Loss Functions for Camera Pose Regression With Deep Learning*. 5974–5983. https://openaccess.thecvf.com/content_cvpr_2017/html/Kendall_Geometric_Loss_Functions_CVPR_2017_paper.html

Khoury, H. M., & Kamat, V. R. (2009). Evaluation of position tracking technologies for user localization in indoor construction environments. *Automation in Construction*, *18*(4), 444–457. https://doi.org/10.1016/j.autcon.2008.10.011

Kuutti, S., Bowden, R., Jin, Y., Barber, P., & Fallah, S. (2021). A Survey of Deep Learning Applications to Autonomous Vehicle Control. *IEEE Transactions on Intelligent Transportation Systems*, *22*(2), 712–733. IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2019.2962338

Lategahn, H., Schreiber, M., Ziegler, J., & Stiller, C. (2013). Urban localization with camera and inertial measurement unit. *2013 IEEE Intelligent Vehicles Symposium (IV)*, 719–724.

https://doi.org/10.1109/IVS.2013.6629552

LaValle, S. (2006). *Planning Algorithms Cambridge, UK: Cambridge Univ*. Press.

Lefèvre, S., Vasquez, D., & Laugier, C. (2014). A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH Journal*, *1*(1), 1. https://doi.org/10.1186/s40648-014-0001-z

Li, T., Zhang, H., Gao, Z., Chen, Q., & Niu, X. (2018). High-Accuracy Positioning in Urban Environments Using Single-Frequency Multi-GNSS RTK/MEMS-IMU Integration. *Remote Sensing*, *10*(2), 205. https://doi.org/10.3390/rs10020205

Li, X., Ge, M., Dai, X., Ren, X., Fritsche, M., Wickert, J., & Schuh, H. (2015). Accuracy and reliability of multi-GNSS real-time precise positioning: GPS, GLONASS, BeiDou, and Galileo. *Journal of Geodesy*, *89*(6), 607–635. https://doi.org/10.1007/s00190-015-0802-8

Li, X. R., & Jilkov, V. P. (2003). Survey of maneuvering target tracking. Part I. Dynamic models. *IEEE Transactions on Aerospace and Electronic Systems*, *39*(4), 1333–1364.

Li, X., Ying, X., & Chuah, M. C. (2019). Grip: Graph-based interaction-aware trajectory prediction. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 3960–3966.

Liu, J., Gao, K., Guo, W., Cui, J., & Guo, C. (2020). Role, path, and vision of "5G + BDS/GNSS". *Satellite Navigation*, *1*(1), 23. https://doi.org/10.1186/s43020-020-00024-w

Liu, Q., Liang, P., Xia, J., Wang, T., Song, M., Xu, X., Zhang, J., Fan, Y., & Liu, L. (2021). A highly accurate positioning solution for C-V2X systems. *Sensors*, *21*(4), 1175.

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., & Wiessner, E. (2018). Microscopic Traffic Simulation using SUMO. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2575–2582. https://doi.org/10.1109/ITSC.2018.8569938

Ma, S., Wen, F., Zhao, X., Wang, Z.-M., & Yang, D. (2019). An efficient V2X based vehicle localization using single RSU and single receiver. *IEEE Access*, *7*, 46114–46121.

Ma, Y., Zhu, X., Zhang, S., Yang, R., Wang, W., & Manocha, D. (2019). TrafficPredict: Trajectory Prediction for Heterogeneous Traffic-Agents. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 6120–6127. https://doi.org/10.1609/aaai.v33i01.33016120

Magowe, K., Giorgetti, A., & Sithamparanathan, K. (2019). Closed-form approximation of weighted centroid localization performance. *IEEE Sensors Letters*, *3*(12), 1–4.

Mercat, J., Gilles, T., El Zoghby, N., Sandou, G., Beauvois, D., & Gil, G. P. (2020). Multi-head attention for multi-modal joint vehicle motion forecasting. *2020 IEEE International Conference on*

Robotics and Automation (ICRA), 9638–9644.

Mohamed, A., Qian, K., Elhoseiny, M., & Claudel, C. (2020). Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14424–14432.

Molina-Masegosa, R., & Gozalvez, J. (2017). LTE-V for Sidelink 5G V2X Vehicular Communications: A New 5G Technology for Short-Range Vehicle-to-Everything Communications. *IEEE Vehicular Technology Magazine*, *12*(4), 30–39. https://doi.org/10.1109/MVT.2017.2752798

Moussaïd, M., Helbing, D., & Theraulaz, G. (2011). How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, *108*(17), 6884–6888.

Mozaffari, S., Al-Jarrah, O. Y., Dianati, M., Jennings, P., & Mouzakitis, A. (2022). Deep Learning-Based Vehicle Behavior Prediction for Autonomous Driving Applications: A Review. *IEEE Transactions on Intelligent Transportation Systems*, *23*(1), 33–47. https://doi.org/10.1109/TITS.2020.3012034

Naik, G., Choudhury, B., & Park, J.-M. (2019). IEEE 802.11bd & 5G NR V2X: Evolution of Radio Access Technologies for V2X Communications. *IEEE Access*, *7*, 70169–70184. https://doi.org/10.1109/ACCESS.2019.2919489

Nezami, F. N., Wächter, M. A., Maleki, N., Spaniol, P., Kühne, L. M., Haas, A., Pingel, J. M., Tiemann, L., Nienhaus, F., Keller, L., König, S. U., König, P., & Pipa, G. (2021). Westdrive X LoopAR: An Open-Access Virtual Reality Project in Unity for Evaluating User Interaction Methods during Takeover Requests. *Sensors*, *21*(5), 1879. https://doi.org/10.3390/s21051879

Ngiam, J., Caine, B., Vasudevan, V., Zhang, Z., Chiang, H.-T. L., Ling, J., Roelofs, R., Bewley, A., Liu, C., Venugopal, A., & others. (2021). Scene transformer: A unified architecture for predicting multiple agent trajectories. *arXiv Preprint arXiv:2106.08417*.

NHTSA. (2021). *Traffic Safety Facts: 2021 Data—Pedestrians. National Highway Traffic Safety Administration*. U.S. Department of Transportation.

Nikhil, N., & Tran Morris, B. (2018). Convolutional neural network for trajectory prediction. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 0–0.

Niu, X., Chen, Q., Zhang, Q., Zhang, H., Niu, J., Chen, K., Shi, C., & Liu, J. (2014). Using Allan variance to analyze the error characteristics of GNSS positioning. *GPS Solutions*, *18*(2), 231–242. https://doi.org/10.1007/s10291-013-0324-x

Page, M., & Wickramarathne, T. L. (2019). Enhanced situational awareness with signals of opportunity: RSS-based localization and tracking. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 3833–3838.

PTV Group. (2020). *PTV Vissim 2020 User Manual*.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707.

Rajamani, R. (2006). *Vehicle dynamics and control*. Springer.

Rasouli, A., & Tsotsos, J. K. (2020). Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice. *IEEE Transactions on Intelligent Transportation Systems*, *21*(3), 900–918. https://doi.org/10.1109/TITS.2019.2901817

Rudenko, A., Palmieri, L., Herman, M., Kitani, K. M., Gavrila, D. M., & Arras, K. O. (2020). Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, *39*(8), 895–935. https://doi.org/10.1177/0278364920917446

SAE International. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (SAE Standard J3016_202104)*.

Salzmann, T., Ivanovic, B., Chakravarty, P., & Pavone, M. (2020). Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. *European Conference on Computer Vision*, 683–700.

Scaramuzza, D., & Fraundorfer, F. (2011). Visual Odometry [Tutorial]. *IEEE Robotics & Automation Magazine*, *18*(4), 80–92. https://doi.org/10.1109/MRA.2011.943233

Schmucker, U., Beirau, M., Frank, M., Stengel, D., Matthes, G., Ekkernkamp, A., & Seifert, J. (2010). Real-world car-to-pedestrian-crash data from an urban centre. *Journal of Trauma Management & Outcomes*, *4*(1), 2.

Schöller, C., Aravantinos, V., Lay, F., & Knoll, A. (2020). What the constant velocity model can teach us about pedestrian motion prediction. *IEEE Robotics and Automation Letters*, *5*(2), 1696–1703.

Schubert, R., Richter, E., & Wanielik, G. (2008). Comparison and evaluation of advanced motion models for vehicle tracking. *2008 11th International Conference on Information Fusion*, 1–6.

Schwebel, D. C., Gaines, J., & Severson, J. (2008). Validation of virtual reality as a tool to understand and prevent child pedestrian injury. *Accident Analysis & Prevention*, *40*(4), 1394–1400. https://doi.org/10.1016/j.aap.2008.03.005

Sheng, Z., Huang, Z., & Chen, S. (2024). Kinematics-Aware Multigraph Attention Network with Residual Learning for Heterogeneous Trajectory Prediction. *Journal of Intelligent and Connected Vehicles*, *7*(2), 138–150. https://doi.org/10.26599/JICV.2023.9210036

Takasu, T., & Yasuda, A. (2009). *Development of the low-cost RTK-GPS receiver with an open source*

program package RTKLIB. *1*, 1–6.

Tan, Z., Che, Y., Xiao, L., Hu, W., Li, P., & Xu, J. (2021). Research of fatal car-to-pedestrian precrash scenarios for the testing of the active safety system in China. *Accident Analysis & Prevention*, *150*, 105857.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., & others. (2017). Graph attention networks. *Stat*, *1050*(20), 10–48550.

Vemula, A., Muelling, K., & Oh, J. (2018). Social attention: Modeling attention in human crowds. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 4601–4607.

Wang, H., Sen, S., Elgohary, A., Farid, M., Youssef, M., & Choudhury, R. R. (2012). No need to war-drive: Unsupervised indoor localization. *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, 197–210. https://doi.org/10.1145/2307636.2307655

Wang, Z., Zhang, H., Lu, T., & Gulliver, T. A. (2018). Cooperative RSS-based localization in wireless sensor networks using relative error estimation and semidefinite programming. *IEEE Transactions on Vehicular Technology*, *68*(1), 483–497.

WHO. (2023). *Global status report on road safety*. World Health Organization.

Woodman, O. J. (2007). *An introduction to inertial navigation* (No. UCAM-CL-TR-696). University of Cambridge, Computer Laboratory. https://doi.org/10.48456/tr-696

Wymeersch, H., Seco-Granados, G., Destino, G., Dardari, D., & Tufvesson, F. (2017). 5G mmWave Positioning for Vehicular Networks. *IEEE Wireless Communications*, *24*(6), 80–86. https://doi.org/10.1109/MWC.2017.1600374

Yin, W., Liu, Y., Shen, C., & Yan, Y. (2019). Enforcing geometric constraints of virtual normal for depth prediction. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5684–5693.

Zampella, F., Jiménez R., A. R., & Seco, F. (2013). Robust indoor positioning fusing PDR and RF technologies: The RFID and UWB case. *International Conference on Indoor Positioning and Indoor Navigation*, 1–10. https://doi.org/10.1109/IPIN.2013.6817857

Zandbergen, P. A., & Barbeau, S. J. (2011). Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. *The Journal of Navigation*, *64*(3), 381–399. https://doi.org/10.1017/S0373463311000051

Zhang, K., Zhao, L., Dong, C., Wu, L., & Zheng, L. (2022). AI-TP: Attention-based interaction-aware trajectory prediction for autonomous driving. *IEEE Transactions on Intelligent Vehicles*, *8*(1), 73–83.

Zhang, Y., Gong, X., Liu, K., & Zhang, S. (2021). Localization and tracking of an indoor autonomous vehicle based on the phase difference of passive UHF RFID signals. *Sensors*, *21*(9), 3286.

Zhang, Y., Wang, W., Guo, W., Lv, P., Xu, M., Chen, W., & Manocha, D. (2022). D2-TPred: Discontinuous dependency for trajectory prediction under traffic lights. *European Conference on Computer Vision*, 522–539.

Zhao, R., Zhang, Y., Wang, T., Wen, W., Xiang, Y., & Cao, X. (2025). Visual Content Privacy Protection: A Survey. *ACM Computing Surveys*. https://doi.org/10.1145/3708501

Zhao, T., Xu, Y., Monfort, M., Choi, W., Baker, C., Zhao, Y., Wang, Y., & Wu, Y. N. (2019). Multi-agent tensor fusion for contextual trajectory prediction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12126–12134.

Zheng, F., Wang, L., Zhou, S., Tang, W., Niu, Z., Zheng, N., & Hua, G. (2021). Unlimited neighborhood interaction for heterogeneous trajectory prediction. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13168–13177.

Zhou, B., Liu, Z., & Su, H. (2024). 5G Networks Enabling Cooperative Autonomous Vehicle Localization: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, *25*(11), 15291–15313. https://doi.org/10.1109/TITS.2024.3420084

Zhu, J., Haiyong, L., Zili, C., & Zhaohui, L. (2014). RSSI based Bluetooth low energy indoor positioning. *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 526–533.

Zhu, Y., Qian, D., Ren, D., & Xia, H. (2019). StarNet: Pedestrian trajectory prediction using deep neural network in star topology. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 8075–8080.

APPENDIX

CCAT Project: Using Connected Intelligent Transportation to Enhance Vulnerable Road User Safety

Published Related Work

Paper 1: Huang, Z., Sheng, Z., Wan, Z., Qu, Y., Luo, Y., Wang, B., Li, P., Chen, Y.-J., Chen, J., Long, K., & others. (2025). Sky-Drive: A Distributed Multi-Agent Simulation Platform for Human-AI Collaborative and Socially-Aware Future Transportation. *Journal of Intelligent and Connected Vehicles*

Abstract:

Recent advances in autonomous system simulation platforms have significantly enhanced the safe and scalable testing of driving policies. However, existing simulators do not yet fully meet the needs of future transportation research—particularly in enabling effective human-AI collaboration and modeling socially-aware driving agents. This paper introduces Sky-Drive, a novel distributed multi-agent simulation platform that addresses these limitations through four key innovations: (a) a distributed architecture for synchronized simulation across multiple terminals; (b) a multi-modal human-in-the-loop framework integrating diverse sensors to collect rich behavioral data; (c) a human-AI collaboration mechanism supporting continuous and adaptive knowledge exchange; and (d) a digital twin framework for constructing high-fidelity virtual replicas of real-world transportation environments. Sky-Drive supports diverse applications such as autonomous vehicle–human road users interaction modeling, human-in-the-loop training, socially-aware reinforcement learning, personalized driving development, and customized scenario generation. Future extensions will incorporate foundation models for context-aware decision support and hardware-in-the-loop testing for real-world validation. By bridging scenario generation, data collection, algorithm training, and hardware integration, Sky-Drive has the potential to become a foundational platform for the next generation of human-centered and socially-aware autonomous transportation systems research.

Paper 2: Sheng, Z., Huang, Z., & Chen, S. (2024). Kinematics-Aware Multigraph Attention Network with Residual Learning for Heterogeneous Trajectory Prediction. *Journal of Intelligent and Connected Vehicles*, *7*(2), 138–150. https://doi.org/10.26599/JICV.2023.9210036

Abstract:

Trajectory prediction for heterogeneous traffic agents plays a crucial role in ensuring the safety and efficiency of automated driving in highly interactive traffic environments. Numerous studies in this area have focused on physics-based approaches because they can clearly interpret the dynamic evolution of trajectories. However, physics-based methods often suffer from limited accuracy. Recent learning-based methods have demonstrated better performance, but they cannot be fully trusted due to the insufficient incorporation of physical constraints. To mitigate the limitations of purely physics-based and learning-based approaches, this study proposes a kinematics-aware multigraph attention network (KA-MGAT) that incorporates physics models into a deep learning framework to improve the learning process of neural networks. Besides, we propose a residual prediction module to further refine the trajectory predictions and address the limitations arising from simplified assumptions in kinematic models. We evaluate our proposed model through experiments on two challenging trajectory datasets, namely, ApolloScape and NGSIM. Our findings from the experiments demonstrate that our model outperforms various kinematics-agnostic models with respect to prediction accuracy and learning efficiency.