Thesis for The Degree of Doctor of Philosophy

Pedestrian Behavior Prediction Using Machine Learning Methods

Chi Zhang

UNIVERSITY OF GOTHENBURG

Department of Computer Science and Engineering
Division of Interaction Design and Software Engineering
University of Gothenburg
Gothenburg, Sweden, 2024

Pedestrian Behavior Prediction Using Machine Learning Methods

Chi Zhang

© Chi Zhang, 2024
except where otherwise stated.
All rights reserved.

ISBN 978-91-8069-817-7 (PRINT)
ISBN 978-91-8069-818-4 (PDF)

Division of Interaction Design and Software Engineering
Department of Computer Science and Engineering
University of Gothenburg
SE-412 96 Gothenburg,
Sweden
Phone: +46(0)31 772 1000

Printed by Chalmers Reproservice,
Gothenburg, Sweden 2024.

To my family.

Pedestrian Behavior Prediction Using Machine Learning Methods

CHI ZHANG

Department of Computer Science and Engineering
University of Gothenburg

Abstract

Background: Accurate pedestrian behavior prediction is essential for reducing fatalities from pedestrian-vehicle collisions. Machine learning can support automated vehicles to better understand pedestrian behavior in complex scenarios.

Objectives: This thesis aims to predict pedestrian behavior using machine learning, focusing on trajectory prediction, crossing intention prediction, and model transferability.

Methods: We identified research gaps by reviewing the literature on pedestrian behavior prediction. To address these gaps, we proposed deep learning models for pedestrian trajectory prediction using real-world data, considering social and pedestrian-vehicle interactions. We integrated spectral features to improve model transferability. Additionally, we developed machine learning models to predict pedestrian crossing intentions using simulator data, analyzing interactions in both single and multi-vehicle scenarios. We also investigated cross-country behavioral differences and model transferability through a comparative study between Japan and Germany.

Results: For trajectory prediction, incorporating social and pedestrian-vehicle interactions into deep learning models improved accuracy and inference speed. Integrating spectral features using discrete Fourier transform improved motion pattern capture and model transferability. For crossing intention prediction, neural networks outperformed other machine learning methods. Key factors that influence pedestrian crossing behavior included the presence of zebra crossings, time to arrival, pedestrian waiting time, walking speed, and missed gaps. The cross-country study revealed both similarities and differences in pedestrian behavior between Japan and Germany, providing insights into model transferability.

Conclusions: This thesis advances pedestrian behavior prediction and the understanding of pedestrian-vehicle interactions. It contributes to the development of smarter and safer automated driving systems.

Keywords: Pedestrian behavior, trajectory prediction, intention prediction, pedestrian-vehicle interaction, deep learning, machine learning

List of Publications

Appended publications

This thesis is based on the following seven publications:

[I] **Chi Zhang** and Christian Berger, *"Pedestrian Behavior Prediction Using Deep Learning Methods for Urban Scenarios: A Review"*
IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 10, pp. 10279-10301, 2023.

[II] **Chi Zhang**, Christian Berger, and Marco Dozza, *"Social-IWSTCNN: A Social Interaction-Weighted Spatio-Temporal Convolutional Neural Network for Pedestrian Trajectory Prediction in Urban Traffic Scenarios"*
In proceedings of the 2021 IEEE Intelligent Vehicles Symposium (IV), pp. 1515-1522. IEEE, 2021.

[III] **Chi Zhang** and Christian Berger, *"Learning the Pedestrian-Vehicle Interaction for Pedestrian Trajectory Prediction"*
In proceedings of 2022 the 8th International Conference on Control, Automation and Robotics (ICCAR), pp. 230-236. IEEE, 2022.

[IV] **Chi Zhang**, Zhongjun Ni, and Christian Berger, *"Spatial-Temporal-Spectral LSTM: A Transferable Model for Pedestrian Trajectory Prediction"*
IEEE Transactions on Intelligent Vehicles, vol. 9, no. 1, pp. 2836-2849. 2023.

[V] **Chi Zhang**, Amir Hossein Kalantari, Yue Yang, Zhongjun Ni, Gustav Markkula, Natasha Merat, and Christian Berger, *"Cross or Wait? Predicting Pedestrian Interaction Outcomes at Unsignalized Crossings"*
In proceedings of the 2023 IEEE Intelligent Vehicles Symposium (IV), pp. 1-8. IEEE, 2023.

[VI] **Chi Zhang**, Janis Sprenger, Zhongjun Ni, and Christian Berger, *"Predicting and Analyzing Pedestrian Crossing Behavior at Unsignalized Crossings"*
In proceedings of the 2024 IEEE Intelligent Vehicles Symposium (IV), pp. 674-681. IEEE, 2024.

[VII] **Chi Zhang**, Janis Sprenger, Zhongjun Ni, and Christian Berger, *"Predicting Pedestrian Crossing Behavior in Germany and Japan: Insights into Model Transferability"*
Manuscript. In submission to IEEE Transactions on Intelligent Vehicles. 2024.

Other publications

The following publications are not appended to this thesis due to their contents overlapping those of appended publications or their content not related to the thesis.

[a] **Chi Zhang** and Christian Berger, *"Analyzing Factors Influencing Pedestrian Behavior in Urban Traffic Scenarios Using Deep Learning"*
Transportation Research Procedia, 72, 1653-1660. Elsevier, 2023.

[b] **Chi Zhang**, Christian Berger, and Marco Dozza, *"Towards Understanding Pedestrian Behavior Patterns from LiDAR Data"*
In the 32nd Annual Workshop of the Swedish Artificial Intelligence Society (SAIS), 2020.

[c] Zhongjun Ni, **Chi Zhang**, Magnus Karlsson, and Shaofang Gong, *"A study of deep learning-based multi-horizon building energy forecasting"*
Energy and Buildings, 303, 113810. Elsevier, 2024.

[d] Zhongjun Ni, **Chi Zhang**, Magnus Karlsson, and Shaofang Gong, *"Leveraging Deep Learning and Digital Twins to Improve Energy Performance of Buildings"*
In 2023 IEEE 3rd International Conference on Industrial Electronics for Sustainable Energy Systems (IESES). IEEE, 2023.

[e] Zhongjun Ni, **Chi Zhang**, Magnus Karlsson, and Shaofang Gong, *"Edge-based Parametric Digital Twins for Intelligent Building Indoor Climate Modeling"*
In the 20th IEEE International Conference on Factory Communication Systems. IEEE, 2024.

Personal Contributions

Paper I: Collected and reviewed existing research studies, analyzed and discussed the strengths and weaknesses of the methods, established the overall framework, and addressed the progress and identified challenges and research gaps of existing works. Structured and wrote the majority of the paper. Responded to the journal reviewers and editors and revised the paper.

Paper II: Proposed the method, designed the network structure, built the model, pre-processed the data, conducted the model training and evaluation, quantitatively and qualitatively analyzed and interpreted the results. Structured and wrote the majority of the paper.

Paper III: Proposed the method, designed the network structure, built the model, pre-processed the data, conducted the model training and evaluation, quantitatively and qualitatively analyzed and interpreted the results. Structured and wrote the majority of the paper.

Paper IV: Proposed the method, designed the network structure, built the model, pre-processed the data, conducted the model training and evaluation, quantitatively and qualitatively analyzed and interpreted the results. Structured and wrote the majority of the paper. Responded to the journal reviewers and editors and revised the paper.

Paper V: Proposed and built the models, conducted the model training and evaluation, quantitatively and qualitatively analyzed the results, interpreted the outcomes. Structured and wrote the majority of the paper.

Paper VI: Proposed and built the models, conducted the model training and evaluation, quantitatively and qualitatively analyzed the results, interpreted the outcomes. Structured and wrote the majority of the paper.

Paper VII: Proposed and built the models, conducted the model training and evaluation, quantitatively and qualitatively analyzed the results, interpreted the outcomes. Structured and wrote the majority of the paper.

List of Abbreviations

3D	three dimensional
AD	automated driving
ADAS	advanced driver assistance systems
ADE	average displacement error
AI	artificial intelligence
AISS	Arnett inventory of sensation seeking
AV	automated vehicle
CNN	convolutional neural network
FDE	final displacement error
GAN	generative adversarial network
GNN	graph neural network
LSTM	long short-term memory
MAE	mean absolute error
MLP	multi-layer perceptron
NN	neural network
RF	random forest
RMSE	root mean squared error
RNN	recurrent neural network
SVM	support-vector machine
SVO	social value orientation
TCN	temporal convolutional network

Acknowledgment

This research was conducted at the University of Gothenburg and was funded by the European research project “SHAPE-IT - Supporting the Interaction of Humans and Automated Vehicles: Preparing for the Environment of Tomorrow” (https://www.shape-it.eu/). It was a Marie Skłodowska-Curie Action (MSCA) Innovative Training Network (ITN) project and received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 860410.

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Christian Berger. I am grateful for the opportunity to be one of his PhD students over the past four years. He has not only shared his technical knowledge and experience but has also provided mental support, encouraging both my academic and personal growth. His continuous guidance, encouragement, and support have been invaluable throughout my research. I would also like to thank my co-supervisor, Prof. Marco Dozza, for his insightful advice and constructive feedback.

My sincere appreciation goes to my colleagues in the SHAPE-IT project. The support and collaboration from all the supervisors and PhD students in SHAPE-IT have ensured that I have never felt alone during this journey. I am grateful to the Institute for Transport Studies at the University of Leeds and Asymptotic AI for hosting my secondment, and I deeply appreciate the wonderful collaborations with all my co-authors.

I want to thank the Wallenberg AI, Autonomous Systems and Software Program (WASP) for providing excellent courses, conferences, and other events. I am grateful to be an affiliated PhD student in the WASP program.

Many thanks to my colleagues at the Software Engineering Unit, the Division of Interaction Design and Software Engineering (IDSE), and the Department of Computer Science and Engineering (CSE) at the University of Gothenburg and Chalmers. Their support has made my PhD journey more wonderful.

I am deeply thankful to my family and friends. Many thanks go to my parents, Guoliang Zhang and Dewen Liu, for their unconditional love and support. I also thank my friends for their support whenever and wherever.

Finally, my heartfelt thanks to my husband, Zhongjun Ni, for his endless love and encouragement, and to my lovely son, Leo, for his energy and smiles that brighten my every day.

Contents

Chapter 1

Introduction

Pedestrian safety is a critical concern that attracts global attention. Intelligent driving systems that can predict pedestrian behavior offer promising solutions for reducing pedestrian-vehicle collisions, potentially reducing injuries and fatalities. Machine learning models, in particular, provide powerful tools for accurate prediction, enabling these intelligent systems to better predict and avoid hazardous situations. This chapter introduces the background, benefits, and challenges of using machine learning methods for pedestrian behavior prediction.

1.1 Background

1.1.1 Global Road Safety Evolution

Global road safety trends indicate an urgent need to reduce road fatalities. According to the 2023 global status report on road safety by the World Health Organization (WHO) [1], there were an estimated 1.19 million fatalities caused by road traffic crashes in 2021, representing a 5% decrease from the estimated 1.25 million fatalities reported in 2010. However, this number is still incredibly high. Road traffic injuries remain the main cause of death for children and young people aged between 5 and 29 years [1], [2], and rank as the 12th leading cause of death across all age groups as of 2019 [1].

The economic impact of road traffic injuries is also significant. While there is no global estimate, in general, traffic crashes are estimated to cost between 1 and 2% of gross domestic product (GDP) [3].

The rapid rise in the number of motor vehicles shows the necessity for improving traffic safety. From 2000 to 2016, the number of motor vehicles increased sharply from 0.85 billion to 2.1 billion [2]. This significant increase requires greater efforts to reduce death rates associated with road traffic accidents to mitigate the impact of the growing number of vehicles on road safety.

In response to these challenges, the Sustainable Development Goals (SDG) target 3.6 in the 2030 Agenda for Sustainable Development [4] aims to halve the number of global deaths and injuries from road traffic accidents. Therefore, there is an increasing demand for developing safer vehicles to prevent hazardous situations and reduce fatalities.

1.1.2 Pedestrian Safety Statistics

Pedestrians, as vulnerable road users, represent a significant proportion of road traffic fatalities. Approximately 273,700 pedestrians were fatally injured in road traffic crashes in 2021, making up about 23% of all road traffic fatalities, as shown in Figure 1.1. This makes pedestrians the second largest group of fatalities, second only to four-wheelers, which accounted for about 30%.

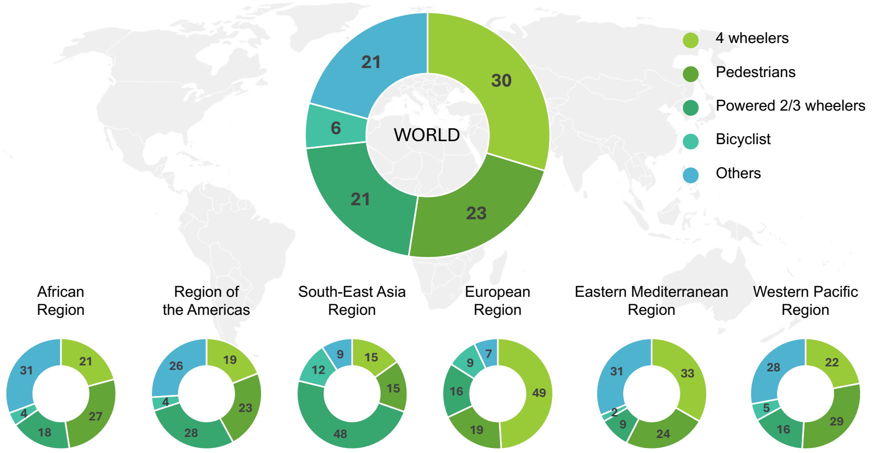


Figure 1.1: Percentage of deaths among road user categories (data from WHO's road safety report 2023 [1]).

While the global fatalities caused by road traffic have decreased over recent years, vulnerable road users, particularly pedestrians, remain dangerously exposed [1], [3]. Globally, pedestrian deaths have increased at nearly twice the rate of overall road crash deaths, with a 12.9% rise from 2013 to 2016 compared to a 6.6% increase for other road users [5]. The International Transport Forum (ITF) report on urban road safety highlights that while road crash death rates decreased from 2010 to 2018, reductions for pedestrians were slower [6]. Pedestrians are nine times more at risk of death than car occupants per kilometer traveled [7].

Pedestrian-vehicle collisions are predictable and preventable according to Peden et al.'s study [8], suggesting that accurately predicting pedestrian behavior could potentially reduce fatalities and injuries. The WHO emphasizes that vehicles can be designed to better protect pedestrians [2]. Understanding and predicting pedestrian behavior has the potential to prevent pedestrian-vehicle collisions, thereby facilitating the development of safer vehicles. Accurate and robust predictions of pedestrian behavior can reduce misunderstandings, provide more reaction time for pedestrians' unexpected movements, and thus prevent hazardous situations. This information is crucial for automated driving (AD) systems and advanced driver assistance systems (ADAS), enabling them to make better and safer decisions.

1.1.3 Locations of Pedestrian-Vehicle Collisions

In this section, we present the locations where pedestrian-vehicle collisions commonly occur. In the European Union, most pedestrian fatalities occur in urban areas. Most pedestrian-vehicle collisions happen when pedestrians are crossing the road. Therefore, these scenarios particularly require the development of intelligent safety measures to reduce accidents.

The distribution of pedestrian-vehicle collisions varies across different countries and regions. Within the European Union, approximately 70% of pedestrian fatalities occur in urban areas. A similar situation is observed in the United States, where 76% of all pedestrian deaths occur in urban areas [3]. Research in Sweden by Värnild et al. [9] from 2003 to 2014 showed that the distribution of seriously injured road users varied between rural and urban areas. Compared to rural areas, where pedestrians constituted only 7% of serious injuries, in urban areas, their proportion increased to 40%, representing a large portion of all serious injuries [9], as shown in Figure 1.2.

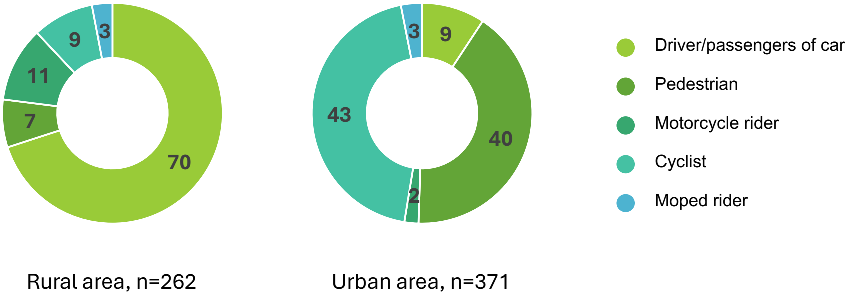


Figure 1.2: Percentage of serious injuries among road user categories in Sweden (Region Västmanland) in rural and urban areas 2003-2014, N=633, with 262 in rural areas and 371 in urban areas (data from Värnild et al.'s report [9]).

The ITF collected traffic safety data from 48 cities across different continents to monitor progress in urban road safety. Their report [6] shows that, in more densely populated cities, the proportion of pedestrians in road fatalities is higher, as illustrated in Figure 1.3. Therefore, our research focuses on urban scenarios instead of rural areas to address the higher risk in these densely populated environments.

Most pedestrian-vehicle collisions are likely to occur when pedestrians are crossing the road [10]. Research by Lane et al. [11] indicates that most fatal

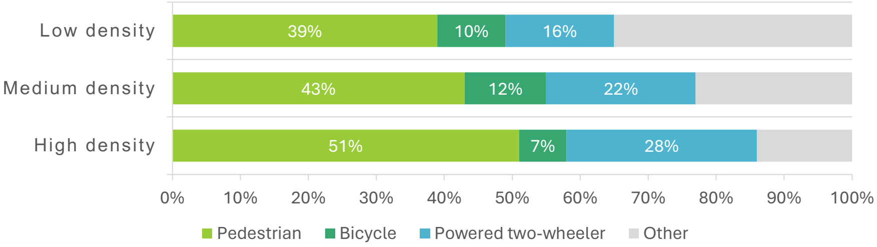


Figure 1.3: Distributions of road fatalities of road user types in different densities of the city, using the average values of figures available between 2014 and 2018. The low population density is less than 5,000 inhabitants per square kilometer. The medium density is less than 10,000, and the high density is 10,000 and above (data from ITF's report [6]).

collisions occurred when pedestrians were crossing, and most injury cases occurred at intersections. This highlights the importance of understanding pedestrian crossing intention and their interaction with the vehicles.

Unsignalized crossings have higher risks of pedestrian injury. Unsignalized crossings are crossings without signal displays or traffic lights. They can be either marked (zebra crossings) or unmarked (non-zebra crossings). Unlike signalized crossings, where traffic signals regulate pedestrian and vehicle movements, road user behavior at unsignalized crossings relies on the judgment and interactions of pedestrians and drivers. This lack of formal control can lead to increased uncertainty and potential conflicts. Olszewski et al. [12] revealed that almost 30% of pedestrian injury accidents took place at unsignalized zebra crossings. Rothman et al. [13] stated that crossing at unsignalized locations resulted in more severe injury compared to crossing at signalized intersections. Therefore, studying pedestrian behavior at unsignalized crossings is essential for designing intelligent systems that can better deal with these complex environments.

1.1.4 The Role of Machine Learning Methods in Improving Pedestrian Safety

The global road and pedestrian safety statistics suggest an urgent need to reduce road traffic fatalities. Human error is one of the main factors in traffic accidents [2], [14]. These errors include distracted driving, speeding, fatigue, and impaired judgment due to alcohol or drugs. Studies by the National Highway Traffic Safety Administration have shown that human errors were involved in approximately 94% of serious road accidents [15]. Similarly, Santoso and Maulina [16] reported that 70% to 90% of accidents can be attributed to human error, and Morgan et al. [17] stated that human error is a contributing factor in over 90% of collisions. Therefore, addressing human error is critical for improving road safety, indicating the need for systems designed to mitigate or eliminate these risks.

AD systems, including automated vehicles (AVs) and ADAS, offer opportunities for enhancing road safety. AVs are designed to perform driving tasks with little human intervention, using advanced sensors and algorithms to navigate and respond to the environment [18], while ADAS supports drivers in tasks such as driving and parking. These systems have the potential to reduce traffic accidents and fatalities [19], [20] by reducing the impact of human errors [19]. Statistical analysis indicates that crashes involving AVs in autopilot mode tend to be less severe than those with human drivers [21]. While AVs show promise in improving road safety, addressing the challenges through continued research and development is essential for fully unleashing their potential.

Machine learning (ML), including deep learning (DL), provides powerful tools for handling complex scenarios [22], [23], which equip AD systems with enhanced capabilities. These methods are subsets of AI, as illustrated in Figure 1.4, and are sometimes used interchangeably with AI [24]. These data-driven technologies build models to identify and predict human behavior patterns and benefit from large-scale datasets. By learning complex non-linear behaviors, ML methods enable accurate and robust predictions.

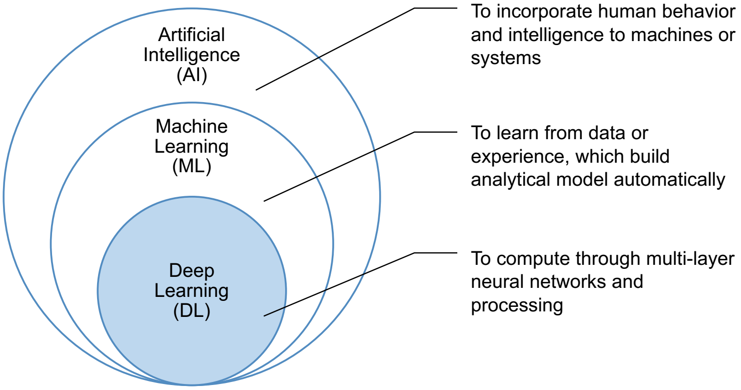

Artificial Intelligence (AI)
Machine Learning (ML)
Deep Learning (DL)
To incorporate human behavior and intelligence to machines or systems
To learn from data or experience, which build analytical model automatically
To compute through multi-layer neural networks and processing

Figure 1.4: The relationship of deep learning, machine learning, and artificial intelligence (cf. Sarker [24]).

Given the vulnerability of pedestrians compared to other road users [3], ML models play a vital role in reducing pedestrian-vehicle collisions, especially at locations where complex interactions occur. Accurate and robust prediction allows AD systems to better understand pedestrian behavior in complex scenarios, ultimately leading to safer decision-making. For instance, ML models can predict pedestrian trajectories [25]–[27] and crossing intentions [28]–[30], supporting AVs and ADAS to make more informed decisions.

1.2 Motivation

The demand for driving safety and automated driving has stimulated an increasing number of research studies in pedestrian behavior prediction [31]. However, research gaps and challenges still exist, including the need for faster and more accurate trajectory prediction methods, deeper analysis for pedestrian-vehicle interactions, and improved model transferability.

Predicting pedestrian behavior is challenging due to the agility of pedestrians, who can change speed and direction abruptly [32], [33]. Factors such as destination, age, gender [34], and interactions with other pedestrians [35] and vehicles [36]–[38] also increase the complexity.

Pedestrian trajectory prediction models should be both fast and accurate to be viable for implementation in AD systems. While incorporating pedestrian interactions has the potential to enhance prediction accuracy [31], previous studies used either traditional rule-based models (e.g., [39]) that lack precision, or overly complex structures that result in a slow prediction speed (e.g., [40], [41]). Many existing models, such as [40], [42]–[44], assume symmetric social interactions, using pooling methods to model them. Models such as [41] addressed this issue by using graph-based algorithms to learn interaction influences but relied on hand-crafted non-linear functions to represent these relationships. There is significant potential to improve accuracy and inference speed using deep learning networks. To address these limitations, we design sub-network structures that effectively consider interactions among pedestrians and between pedestrians and vehicles, improving both accuracy and prediction speed.

Pedestrian intention prediction models should integrate pedestrian-vehicle interactions with in-depth analyses to support AD systems in making more informed decisions. Pedestrian crossing behavior is complex as they are influenced by various factors, especially when interacting with vehicles. Previous studies using machine learning models to predict intentions did not explicitly consider the pedestrian-vehicle interaction and overlooked detailed pedestrian-vehicle analysis [45]–[50]. Those who investigated pedestrian-vehicle interactions mainly focused on liner relationships [51]–[54], missing the non-linearity and complexity of the interactions, and did not predict specific interaction outcomes. To address these gaps, we develop machine learning models that predict pedestrian-vehicle interaction outcomes at unsignalized crossings and analyze the key factors influencing pedestrian crossing behavior and their impact.

Predictive models for pedestrian behavior need transferability to be applied in various scenarios. Most existing models are based on data from specific datasets or single countries [28], [40]–[42], [53], limiting their applicability in new scenarios or different countries. To address these limitations, we focus on investigating the transferability of pedestrian behavior models.

1.3 Research Goals and Questions

This thesis aims to use machine learning methods, including deep learning and traditional machine learning algorithms, to understand and predict pedestrian behavior in complex traffic scenarios. To achieve this goal, the following sub-goals are addressed:

G1: To predict pedestrian trajectory in urban traffic scenarios using deep learning methods.

G2: To predict pedestrians' crossing intentions and understand their interactions with vehicles at unsignalized crossings using machine learning methods.

G3: To investigate the transferability of pedestrian behavior prediction models.

We derive the following research questions from the goals:

RQ1: What are the state-of-the-art deep learning algorithms for predicting pedestrian behavior and how do they perform in urban scenarios?

RQ2: How can deep learning methods improve the prediction of pedestrian trajectories in urban traffic scenarios?

RQ2-1: What are the improvements in pedestrian trajectory prediction when using deep learning methods to extract social interactions within pedestrians compared to existing methods?

RQ2-2: What are the improvements in pedestrian trajectory prediction when using deep learning methods to extract pedestrian-vehicle interactions compared to existing methods?

RQ2-3: What are the improvements in pedestrian trajectory prediction when considering spectral information compared to existing methods?

RQ3: How can machine learning methods improve the prediction of pedestrian intentions and enhance the understanding of pedestrian-vehicle interactions at unsignalized crossings?

RQ3-1: How can machine learning methods improve the prediction of pedestrian crossing intention and interaction outcomes when interacting with a single vehicle?

RQ3-2: How can machine learning methods improve the prediction of pedestrian crossing time gap selection and the use of zebra crossings when interacting with multiple vehicles?

RQ4: What is the transferability of our proposed deep learning and machine learning models?

RQ4-1: How does considering spectral information in deep learning models improve the transferability of trajectory prediction compared to existing methods?

RQ4-2: How transferable are machine learning models for predicting pedestrian crossing behavior between different countries?

The appended papers contribute to improving pedestrian behavior prediction using machine learning methods. We provide a literature review on pedestrian behavior prediction (Paper I). We focus on trajectory prediction (Papers II, III, and IV), intention prediction (Papers V, VI, and VII), and model transferability (Papers IV and VII). Table 1.1 outlines how these papers contribute to corresponding research goals and address the research questions.

Table 1.1: A summary of appended papers.

Paper	Goal	Research Question	Main countributions
I	G1, G2	RQ1	Reviewed current deep learning research studies for pedestrian trajectory and intention prediction, discussed their strengths and weaknesses, outlined a comprehensive framework, and highlighted research gaps.
II	G1	RQ2-1	Proposed a trajectory prediction method that improves performance by considering social interactions.
III	G1	RQ2-2	Proposed a trajectory prediction method that improves performance by considering pedestrian-vehicle interactions.
IV	G1, G3	RQ2-3, RQ4-1	Proposed a trajectory prediction method that improves performance and transferability by considering spatial, temporal, and spectral information.
V	G2	RQ3-1	Proposed machine learning methods for pedestrian behavior prediction when interacting with a single vehicle, and identified key factors and their impacts on pedestrian behavior.
VI	G2	RQ3-2	Proposed machine learning methods for pedestrian behavior prediction when interacting with multiple vehicles, and identified key factors and their impacts on pedestrian behavior.
VII	G2, G3	RQ3-2, RQ4-2	Proposed machine learning methods for pedestrian behavior prediction when interacting with multiple vehicles, identified key factors and their impacts on pedestrian behavior, and investigated the differences between countries.

1.4 Thesis Outline

The remaining chapters of this thesis are structured as follows:

Chapter 2: Methodology. This chapter introduces the problem formulation, outlines the evaluation metrics, introduces the deep learning algorithms for pedestrian trajectory prediction and the machine learning algorithms for crossing intention and interaction prediction, and presents the dataset in use.

Chapter 3: Summary of Appended Papers. This chapter outlines the objectives, methodologies, results, and contributions of the appended papers.

Chapter 4: Discussion. This chapter discusses the findings from the seven appended papers, highlighting their contributions. It also presents the limitations of the study and proposes potential future research.

Chapter 5: Conclusions. This chapter summarizes the key contributions and findings presented in this thesis.

Chapter 2

Methodology

The methodology chapter is organized as follows: First, we introduce the definition of the problems. Next, we present how the methods are evaluated. Following this, we provide details for the methods used in this thesis, including deep learning algorithms for trajectory prediction and machine learning methods for intention and interaction prediction. Finally, we introduce the datasets that we used in this thesis.

2.1 Problem Definition

This thesis focuses on the prediction of pedestrian behaviors, including trajectory prediction (Papers II, III, and VI), and crossing intention and interaction outcomes (Papers V, VI, and VII). Paper I reviews existing literature and provides the definition of trajectory and intention prediction problems.

2.1.1 Trajectory Prediction

We predict pedestrian future trajectories based on past trajectories in Papers II, III, and VI. The trajectory of a pedestrian consists of a sequence of positions in two-dimensional (2D) x-y coordinates $X = (x, y)$ with their temporal order. The positions in each frame are first pre-processed to x-y coordinates on a 2D map representation from the bird's-eye-view. This allows us to accurately capture pedestrians' spatial movement patterns. In each frame at time-step t with the number of pedestrians n_p, the i^{th} person at time-step t is represented by x-y-coordinate $X_t^i = (x_t^i, y_t^i)$, where $i \in \{1, \ldots, n_p\}$. The observed trajectories can be denoted as $X_t = [X_t^1, X_t^2, \ldots, X_t^{n_p}]$, with all observed time-steps $1 \leq t \leq T_{obs}$. Given this input, our goal is to predict the most likely future trajectories $\hat{Y}_t = [\hat{Y}_t^1, \hat{Y}_t^2, \ldots, \hat{Y}_t^{n_p}]$, where future time steps $T_{obs} + 1 \leq t \leq T_{pred}$. The ground truth of the future trajectories is denoted as $Y_t = [Y_t^1, Y_t^2, \ldots, Y_t^{n_p}]$, where $T_{obs} + 1 \leq t \leq T_{pred}$.

2.1.2 Intention and Interaction Prediction

Pedestrian interacting with a single vehicle: In Paper V, we focus on scenarios where a pedestrian interacts with a single vehicle at unsignalized crossings. Given observed variables, we predict the "cross or wait" decisions of all pedestrians, which can be considered as a classification problem. For those crossing cases, we are also concerned about the crossing initiation time and crossing duration, which are regression problems.

Pedestrian interacting with multiple vehicles: In Papers VI and VII, we focus on scenarios when pedestrians interact with multiple vehicles. Given the observed variables, we predict the time gap selected and accepted by the pedestrian for non-zebra crossing scenarios, which is a regression problem. For zebra crossing scenarios, we predict whether pedestrians use the zebra crossing, which is a classification problem.

2.2 Evaluation Metrics

2.2.1 Trajectory Prediction Problem

For the trajectory prediction problem, we use the average displacement error (ADE) and the final displacement error (FDE) for evaluation. In Paper I, existing algorithms are reviewed and compared, and the state-of-the-art algorithms are listed. In Papers II, III, and IV, in addition to ADE and FDE, which evaluate the displacement error, the average inference speed of different models is also evaluated to compare computational performance.

- ADE: the average distance between ground truth and prediction trajectories over all predicted time-steps, as defined in Eq. 2.1. It also refers to the mean square error over all estimated positions of every trajectory and the true positions.

$$ADE = \frac{\sum_{i \in n_p} \sum_{t=T_{obs}+1}^{T_{pred}} \|Y_t^i - \hat{Y}_t^i\|_2}{n_p \times (T_{pred} - T_{obs})} \tag{2.1}$$

- FDE: the average distance between ground truth and prediction trajectories for the final predicted time-step, as defined in Eq. 2.2:

$$FDE = \frac{\sum_{i \in n_p} \|Y_t^i - \hat{Y}_t^i\|_2}{n_p}, t = T_{pred} \tag{2.2}$$

2.2.2 Classification Problem

For the classification problems in Papers V, VI, and VII, we use prediction accuracy and F1 score for evaluation. The evaluation functions are shown below, where P and N denote the numbers of positives and negatives, respectively. TP, TN, FP, and FN are the numbers of true positives, true negatives, false positives, and false negatives, respectively.

$$Accuracy = \frac{TP + TN}{P + N} \tag{2.3}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{2.4}$$

2.2.3 Regression Problem

For the regression problems in Papers V, VI, and VII, we use mean absolute error (MAE) and root mean squared error (RMSE) for evaluation. The evaluation functions are defined as follows, where y_i denotes the ground truth for the i^{th} trial, and $\hat{y}_i$ denotes the corresponding prediction, n denotes the number of trials.

$$MAE = \frac{\Sigma|\hat{y}_i - y_i|}{n} \tag{2.5}$$

$$RMSE = \sqrt{\frac{\Sigma(\hat{y}_i - y_i)^2}{n}} \tag{2.6}$$

2.3 Deep Learning for Pedestrian Trajectory Prediction

2.3.1 Deep Learning Models for Sequence Prediction

Deep learning models can benefit more from large-scale datasets compared to traditional machine learning methods [24]. Many recent studies have explored the application of deep learning and neural networks for pedestrian behavior prediction [31]. In this section, we introduce deep learning algorithms that are utilized in our research.

Long Short-Term Memory Networks

Recurrent Neural Networks (RNNs) maintain a memory of previous inputs through internal states, making them effective for sequential prediction. While RNNs can capture temporal dependencies, they struggle with long-term dependencies due to gradient vanishing.

Long short-term memory (LSTM) networks are an improved version of RNNs. They have both feedforward and feedback connections to capture long and short-term information, making them especially effective for sequential data predictions. LSTMs have been successfully applied for tasks such as handwriting [55] and speech recognition [56].

Due to their capabilities in sequential prediction, LSTMs have been adopted by researchers for predicting pedestrian trajectories (e.g., [42]–[44], [57]). For instance, Alahi et al. [42] proposed Social-LSTM, which assumed the trajectories of pedestrians follow the bi-variate Gaussian distribution, and many researchers followed this uni-modal assumption. The drawback of the LSTM-based methods is that they cannot be parallelized as predictions at each time step depend on preceding time steps.

Generative Adversarial Networks

Generative Adversarial Networks (GANs), proposed by Goodfellow et al. [58], consist of two neural networks, a generator and a discriminator, that contest with each other. The generator generates multiple possible candidates, while the discriminator evaluates them. For pedestrian trajectory prediction, Gupta et al. [40] stated that assuming a uni-modal distribution may lead to learning the "average" trajectories rather than multiple "good behaviors". Instead, they introduced GANs to model pedestrian trajectories following a multi-modal distribution. Although GANs are able to predict multiple feasible outcomes, it is challenging to achieve convergence for training two deep networks within one structure.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are widely used for processing structured grid data such as images. They employ convolutional layers with filters that slide over the input to capture local patterns. CNNs have been successfully used in tasks like image classification and segmentation, due to their capability of extracting spatial features. Many studies have used CNNs for pedestrian intention prediction to extract appearance and behavioral features.

Additionally, convolutional networks in temporal space, also known as temporal convolutional networks (TCNs), can extract temporal features and be used for sequential prediction. Bai et al. [59] stated that RNNs' inefficient parameter usage can make training costly, and LSTMs' dependency on preceding time steps can lead to error accumulation. Nikhil and Morris [60] utilized CNNs for trajectory prediction, achieving competitive results with a faster inference speed. Mohamed et al. [41] proposed the Social-STGCNN method, combining spatial and temporal features using TCNs and CNNs with graph structures, achieving a 20% improvement in FDE and being 48 times faster compared to sequential models.

Transformers

In recent years, Transformers [61] have become popular due to their superior ability to memorize information in long sequences compared to RNNs. Their attention mechanism allows for shortcuts between the context vector and the entire input, instead of relying only on the last hidden state as in RNNs. Transformers achieve better performance compared to RNNs, and allow for parallelization, thus reducing training time. The original Transformers [61] are implemented with a fixed length, limiting their ability to model dependencies longer than that length. Enhanced versions, such as TransformerXL [62], address these limitations and allows for learning dependencies beyond a fixed length without disrupting temporal coherence. Later variations such as compressive Transformer [63], Longformer [64], and Reformer [65] improved the efficiency for processing long sequences.

Transformers have made groundbreaking progress in the Natural Language Processing (NLP) field and are now being adopted for predicting pedestrian trajectories. Giuliari et al. [66] demonstrated that Transformers achieved better performance than previous LSTM- and CNN-based models on individual trajectory prediction. Other studies, such as spatio-temporal graph Transformer networks proposed by Yu et al. [67], AgentFormer proposed by Yuan et al. [68], and STGT proposed by Syed et al. [69] have also used Transformers, considering interactions with other road users and context information, and achieved more accurate results.

In this thesis, we apply all aforementioned prediction structures for trajectory prediction. The appended Paper I introduces these networks, highlighting their advantages and drawbacks. Paper II utilizes CNNs for trajectory prediction, considering social interactions, and compares their performance with LSTMs and GANs. Paper III utilizes both CNNs and LSTMs separately for trajectory prediction, considering social interactions and pedestrian-vehicle interactions. Paper IV proposes an LSTM-based model considering spectral information, and compares the performance of LSTMs, CNNs, and Transformers.

2.3.2 Interactions between Pedestrians and Other Road Users

As pedestrians frequently interact with other road users in urban traffic scenarios, considering interaction information can potentially improve the accuracy when predicting pedestrian trajectory. In this thesis, we utilize deep learning networks to capture these interactions and investigate their influence on prediction. We investigate how different interactions influence trajectory prediction and identify relevant features for extracting these interactions. In Paper II, we consider social interactions with other pedestrians, while in Paper III, we also consider interactions between pedestrians and vehicles.

Social Interactions with Other Pedestrians

Moussaid et al. [35] stated that pedestrian behavior is influenced not only by individual factors but also by social interactions with nearby pedestrians. Many researchers have focused on modeling social interactions in pedestrian trajectory prediction. Alahi et al. [42] pioneered applying deep learning networks to trajectory prediction. They utilized LSTMs and proposed “social pooling” to capture social interactions, instead of using conventional handcrafted functions like the Social Force model [39]. Pooling layers are typically used in CNNs for dimensionality reduction. These layers commonly calculate either the average or maximum value within a specified pooling operation area. In the context of trajectory prediction, “social pooling”, as introduced by Alahi et al. [42], uses pooling operations to allow information sharing among neighboring pedestrians. Subsequent research further refined and improved the social pooling module with more complicated structures [40], [70].

Some researchers have stated that social interactions are not symmetric, so instead of using pooling methods, they introduced graph-based networks for learning social interactions [41], [71]. Graph neural networks (GNNs) construct a graph $< V, E >$ where vertices represent the states of each road user, and edges represent interaction relationships between them. Approaches like STGAT [72] and Social-BiGAT [71] used graph attention networks [73] to model interactions, while Mohamed et al. [41] utilized graph convolutional networks [74] to assign interaction weights of the target pedestrian's surrounding neighbors for social interaction modeling.

However, constructing GNNs and calculating non-linear edge values are time-consuming [25]. Multi-layer perceptrons (MLPs) are capable of learning interaction relationships with linear computation and activation functions, offering faster inference speeds. In Paper II, we use MLPs to learn interaction weights and propose a weighted sum aggregation function to aggregate the influence of neighboring pedestrians, avoiding graph convolutional operations and thereby accelerating the computation. Compared to the previous state-of-the-art Social STGCNN model, our proposed approach reduces prediction error by 1.8% and speeds up inference by a factor of 4.7.

Interactions between Pedestrians and Vehicles

In addition to social interactions with other pedestrians, pedestrian-vehicle interactions also influence pedestrian trajectory. Many researchers have attempted to include vehicle information in pedestrian behavior prediction models. Traditional approaches explicitly use hand-crafted features such as speed, orientation, distance to pedestrians, and time to collision (TTC) as inputs to neural networks to learn their impact on pedestrian trajectory [34], [48], [49], [75]–[77]. However, pedestrian-vehicle interactions are complex in scenarios involving multiple pedestrians and vehicles, making it challenging to generalize designed features to new scenarios. Therefore, an increasing number of studies use deep learning sub-networks to learn these interactions.

As pedestrians and vehicles exhibit different motion patterns, their interactions are inherently asymmetric. For example, consider a situation where a pedestrian waits to cross a street *without* zebra markings where the pedestrian does not clearly have the right of way. The pedestrian's motion is more agile and can freely adjust their trajectory based on the vehicle's behavior, changing their direction and speed as needed. In contrast, the car's trajectory is primarily determined by its original speed and direction, with minimal adjustments made for the waiting pedestrian. This scenario highlights how the car significantly impacts the pedestrian's trajectory, while the pedestrian has limited influence on the vehicle's trajectory.

GNNs are particularly suited for learning these asymmetric interaction relationships between pedestrians and vehicles. This approach has been adopted by several studies, such as [78]–[82]. Models proposed by Chandra et al. [83]–[85] predicted the trajectories of different types of road users simultaneously, but focused mainly on vehicles, with limited emphasis on pedestrians.

To address the drawbacks of GNNs as mentioned earlier, MLPs can be used for capturing pedestrian-vehicle interactions. Relative positions and velocities between pedestrians and vehicles are used as inputs of this sub-network. In Paper III, we use a separate multilayer perceptron (MLP) network to extract pedestrian-vehicle interactions in addition to social interactions.

We also identify optimal inputs for learning, and evaluate the influence of pedestrian-vehicle interactions on trajectory prediction accuracy.

2.4 Machine Learning for Crossing Intention and Interaction Prediction

As reviewed and summarized in Paper I, recent studies have made significant strides in predicting pedestrian crossing behavior using naturalistic data [45], [46], [48], [49], [86]. These studies mainly used deep learning networks to capture the features of pedestrian appearance, postures, and nearby environments without explicitly considering pedestrian-vehicle interactions. Furthermore, they failed to thoroughly analyze key influencing factors and their impact on crossing behavior.

The research by Völz et al. [75], [87], Zhang et al. [34], and Jayaraman et al. [88] investigated pedestrian crossing behavior at unsignalized crossings and addressed pedestrian-vehicle interactions. However, these studies either neglect to investigate pedestrian behavior at different crossing locations (with or without zebra crossings), overlook the influence of personality traits, or focus only on single-vehicle and single-pedestrian interactions, neglecting the complex interactions with multiple vehicles that are prevalent in real-world situations.

To address these limitations, in Paper V, we investigate pedestrian interactions with a single vehicle and consider the influence of crossing locations and personality traits. In Papers VI and VII, we address pedestrian interactions with multiple vehicles, predicting the selected and accepted time gaps for crossing, as well as the usage of zebra crossings.

Simulator data has become increasingly popular for studying pedestrian crossing behavior due to its controlled and safe environment. Our research (Papers V, VI, and VII) utilizes data collected from simulator studies. As these datasets are relatively small, deep learning methods may cause overfitting [89]. In such cases, traditional machine learning models, such as linear regression and logistic regression models, random forest (RF), support vector machine (SVM), are a suitable alternative. When predicting a pedestrian's intention and interaction outcomes, various machine learning algorithms can be employed. This section introduces the machine learning algorithms used in our research.

Linear Regression

Linear regression is a statistical method for modeling the relationship between a dependent variable and one or more independent variables [90]. It assumes a linear relationship and aims to determine the best-fit line that minimizes the sum of squared differences between the observed and predicted values. It is widely used for *regression* and predictive analysis.

Logistic Regression

Logistic regression is used for binary *classification* problems. It predicts the probability of an event by modeling the log-odds for the event as a linear combination of independent input variables. The output is a probability value between 0 and 1, which can be thresholded to classify the input into one of two classes.

Support-Vector Machine

Support-Vector Machine (SVM) aims to find a hyperplane in the feature space that best separates the data into different classes and can be used for *classification*. The hyperplane is chosen to maximize the margin between the classes, making SVM effective in high-dimensional spaces and cases where the classes are clearly separable. SVM can also handle non-linear classification using kernel functions. In this thesis, we use a linear kernel, so the SVM is also considered a linear-based model.

Random Forest

Random Forest (RF) is an ensemble learning method for both *classification* and *regression* tasks [91]. It combines multiple decision or regression trees to improve classification or regression performance. Each tree is trained on a random subset of the data and features. The final output is obtained by aggregating the predictions from all separate trees, using the most selected label for classification and the average prediction for regression. Random forests are robust to overfitting and can handle large datasets with high dimensionality.

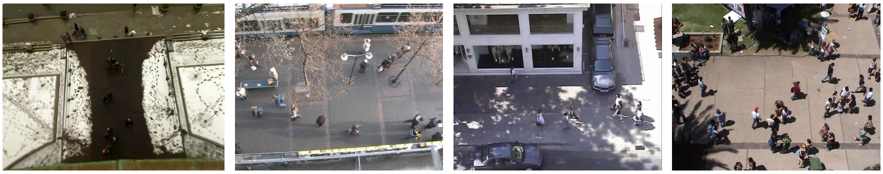
Neural Networks

Neural Networks (NNs) are based on a collection of artificial nodes and can be used for both *classification* and *regression*. NNs usually contain several node layers, consisting of an input layer, an output layer, and one or several hidden layers. When the number of models' input features is small, the multilayer perceptron (MLP) can be used. MLP is a fully connected feedforward NN. MLPs are trained by backpropagation, where the network adjusts the weights of the connections to minimize the difference between the predicted and actual outputs. This is typically done using gradient descent optimization.

In Papers V, VI, and VII, all aforementioned models are employed to predict pedestrian intention and interaction outcomes. These papers evaluate and compare machine learning algorithms to understand pedestrian crossing behavior in unsignalized crossings. The key factors for modeling are identified, and their impact on pedestrian crossing behavior is analyzed.

2.5 Datasets

This thesis utilizes both naturalistic and simulator data. In this section, we introduce the datasets in use. Papers II, III, and IV employ naturalistic data



(a) ETH-univ (b) ETH-hotel (c) UCY-zara (d) UCY-univ

Figure 2.1: Screenshots of the ETH [92] and UCY [93] datasets. These data are collected at different locations.

for trajectory prediction, while Papers V, VI, and VII utilize simulator data to investigate pedestrian-vehicle interaction.

2.5.1 Naturalistic Data

Naturalistic data are widely used for pedestrian trajectory prediction. Paper I provides a comprehensive list of publicly available datasets commonly used for training and evaluating pedestrian trajectory prediction models.

ETH and UCY Datasets

The ETH [92] and UCY [93] datasets are commonly used for evaluating pedestrian trajectory prediction models. These datasets consist of five scenes collected in crowded urban scenarios at fixed locations from the bird's-eye-view. The ETH dataset contains two scenes, namely university and hotel scenes, with annotations for 750 unique pedestrians. The UCY dataset comprises three scenes, including two street scenes and one university scene, with annotations for 786 unique pedestrians. Figure 2.1 provides snapshots of the scenarios in these datasets. The ETH and UCY datasets are used in Paper IV.

Waymo Open Dataset

The Waymo Open Dataset [94] contains 1,150 real-world road scenes collected in the United States from the vehicle's view, including 450 scenes from urban street scenarios. To investigate pedestrian behavior in urban scenarios, we focus on these 450 urban street scenes, consisting of 374 training records and 76 test records, each lasting 20 seconds. The training records are further divided into a training set containing 337 records and a validation set containing 37 records. The 76 test records are reserved for model evaluation. A snapshot of the urban traffic scenario in the Waymo Open Dataset is shown in Figure 2.2. The Waymo Open Dataset is used in Papers II, III, and IV.

Data Pre-processing

The basic information of the datasets used for trajectory prediction in this thesis is shown in Table 2.1. The statistics of these datasets is shown in Table 2.2. These datasets are collected from different countries using various

Figure 2.2: A snapshot of an urban traffic scenario in Waymo Open Dataset [94].

Table 2.1: Basic information about the datasets used in this thesis.

Dataset name	Labeled objects	Collected view	Collected location	Frequency
ETH	Pedestrians	Bird's-eye-view	Switzerland	2.5 Hz
UCY	Pedestrians	Bird's-eye-view	Cyprus	2.5 Hz
Waymo	Pedestrians, vehicles, cyclists, signs	Vehicle's view	United States	10 Hz

Table 2.2: Statistics of the datasets used in this thesis.

Dataset name	Number of frames (@2.5Hz)	Number of pedestrian sequences	Average number of targets per frame	Average speed (m/s)
ETH-univ	1,448	603	6.27	0.92
ETH-hotel	1,168	301	5.60	1.04
UCY-zara1	872	602	5.91	1.07
UCY-zara2	1,052	921	9.24	0.79
UCY-univ	985	947	40.37	0.63
Waymo (train)	17,127	8,328	29.33	0.91
Waymo (test)	3,570	1,978	30.91	0.95

sensors and views. We use only trajectories as input to reduce the influence of different sensors and calibration parameters of these datasets. We pre-process the naturalistic data into a consistent coordinate system and frequency to ensure uniform input information for the proposed models.

The ETH and UCY datasets are collected from the bird's-eye-view using cameras. Pedestrian labels are transformed from image coordinates (u, v) into center positions (x, y) of pedestrians in the real world.

The Waymo Open Dataset is collected with high-resolution cameras and LiDARs from the vehicle's view. It uses a local coordinate system with the ego-vehicle's center as the origin in each frame. Since using local coordinates introduces the ego-vehicle's movement into the pedestrians' movement, affecting prediction accuracy, we pre-process the data. We use the pedestrians and vehicles labeled in the LiDAR data with their real-world center positions (x, y, z). We include all labeled pedestrians and vehicles within the LiDAR scan range of 75 m. Each sequence of objects has a unique track ID. In this thesis, pedestrians and vehicles are considered as points without the size and shape information. We first pre-process them into 2D positions (x, y) sequences from a bird's-eye view. Then, to avoid the influence of the ego-vehicle's movement, we transform the coordinates into global coordinates, using the ego-vehicle's position at the first time step of the recording as the origin.

The labeled sequences we used from the Waymo Open Dataset have a frequency of 10 Hz, whereas many previous state-of-the-art models were evaluated on the ETH and UCY datasets with a frequency of 2.5 Hz. To ensure consistency and avoid the influence of different frequencies, we down-sample all sequences to 2.5 Hz.

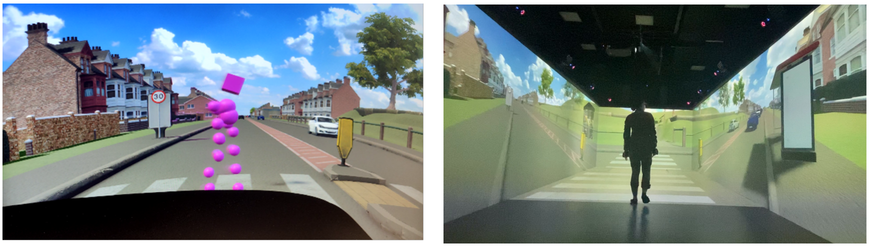
2.5.2 Simulator Data

Simulator data is becoming popular for investigating pedestrian crossing behavior, especially when exploring pedestrian-vehicle interaction [28]. Investigating near-crash scenarios in real-world settings is unsafe and unethical, as it may harm pedestrians. Ensuring pedestrian safety is important, and simulators provide a safe environment for collecting crossing behavior and interaction information.

Besides, when using naturalistic data, it is challenging to separate interaction factors from other potential latent variables influencing pedestrian behavior. Simulator studies, however, provide a controlled or semi-controlled environment, making it easier to investigate these interaction factors. Therefore, conducting simulator studies within safe and controlled environments is an effective way to explore pedestrian crossing behavior and pedestrian-vehicle interactions.

2.5.2.1 Pedestrians Interacting with a Single Vehicle

In Paper V, we investigate scenarios where the target pedestrian interacts with a single vehicle, using the distributed simulator study data collected by Kalantari et al. [53]. The distributed simulator study was conducted by connecting two high-fidelity simulators: the University of Leeds Driving Simulator (UoLDS) and the Highly Immersive Kinematic Experimental Research (HIKER) pedestrian lab. The UoLDS is a motion-based driving simulator with eight degrees of freedom and a 300-degree field of view, housed in a 4 m spherical dome. HIKER is a Cave Automatic Virtual Environment (CAVE)-based simulator with dimensions of $9m \times 4m$. Utilizing eight Barco F90 4k projectors, virtual scenes are projected at a frequency of 120 Hz onto the floor and walls. As shown in Figure 2.3 (a), fourteen body markers were attached to the pedestrian's body, represented as pink spheres visible to the driver. The vehicle was also observable as an entity to the pedestrian, as shown in Figure 2.3 (b).



(a) (b)

Figure 2.3: Illustration for the distributed simulator study. (a) A pedestrian from the driver's view: the pink spheres represent body markers attached to the pedestrian. (b) An interaction example from the third view: a pedestrian is at the zebra crossing and interacting with the vehicle to their right.

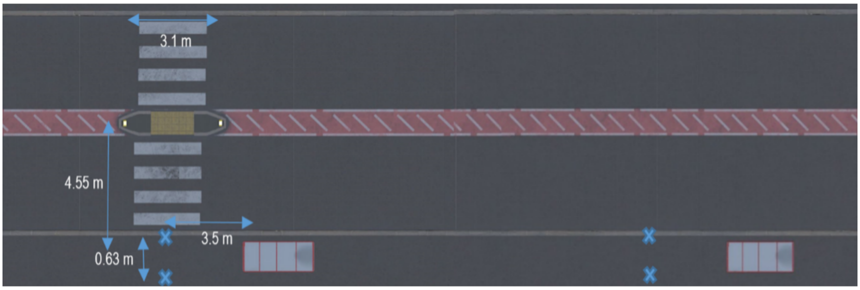


Figure 2.4: Bird's-eye-view of the zebra (left) and non-zebra crossings (right) with the designated standpoints (shown by the blue cross markers). The gray rectangles are visual obstructions (bus stops).

Sixty-four participants including 32 drivers and 32 pedestrians are paired and interacted with each other in various scenarios. The drivers are at age: mean $(M) = 31.53$, range $(R) = 21 - 50$, standard deviation $(SD) = 1.72$. The pedestrians at age: $M = 25.09$, $R = 19 - 34$, $SD = 0.87$. Both drivers and pedestrians are equally distributed by gender. The interaction scenarios include five different time to arrival and four crossing locations (two at zebra crossings and two at non-zebra crossings), as shown in Figure 2.4. This led to 20 conditions repeated in two separate experimental blocks, resulting in 40 randomized trials per participant pair. An 890 m two-way urban road with traffic on both lanes was created in Unity 3D to apply these settings. The data was collected after obtaining approval from the University of Leeds Ethics Committee.

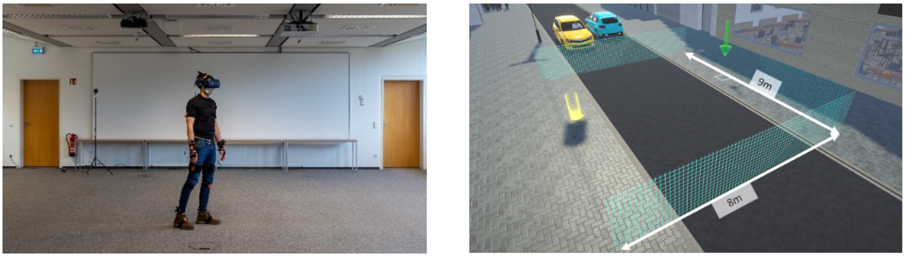
Both pedestrian and driver participants were instructed to act assuming they were in a hurry while prioritizing safety. Drivers were informed to maintain a speed limit of 30 mph and to yield to pedestrians only when necessary. In addition to objective variables, personality traits including social value orientation (SVO) slider measure [95] and Arnett inventory of sensation seeking (AISS) [96] were collected using questionnaires. The observed variables considered as candidate inputs are listed in Table 2.3.

Table 2.3: Candidate input variables for predicting pedestrian intentions when interacting with a single vehicle.

Variable [Unit]	Description (type)
T_a [s]	Time to arrival (continuous)
T_w [s]	Waiting time (continuous)
L	Crossing location type, including two categories: zebra and non-zebra (categorical)
A_d, A_p	Age for both pedestrians and drivers. (discrete)
G_d, G_p	Gender for both pedestrians and drivers (categorical)
SVO_d, SVO_p [degree]	SVO slider measure for both pedestrians and drivers, calculated from the questionnaire [95] (continuous)
$AISS_d$, $AISS_p$	AISS for both pedestrians and drivers, calculated from the questionnaire [96] (continuous)

2.5.2.2 Pedestrians Interacting with Multiple Vehicles

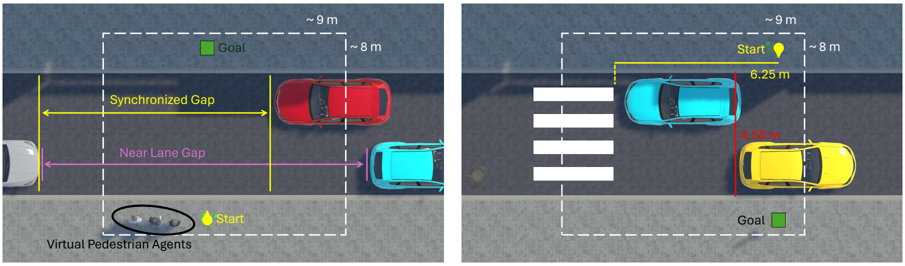
In Papers VI and VII, we investigate scenarios where the target pedestrian interacts with multiple vehicles, using the simulator data collected by Sprenger et al. [97]. The data was collected in a controlled experiment in a virtual street environment with a bi-directional one-lane road. Pedestrian participants, equipped with an untethered and head-mounted virtual reality headset, navigated freely in a $9m \times 8m$ space, selecting routes and physically crossing (walking or running) the virtual road to reach a goal on the opposite side. Figure 2.5 (a) shows the pedestrian wearing the headset. An example of the scenarios for the pedestrian participants are shown in Figure 2.5 (b).



(a) (b)

Figure 2.5: Illustration for the simulator data collection. (a) A pedestrian participant wearing the headset. (b) An example of the experimental setup.

This study included 120 pedestrian participants in Germany and Japan, with 60 participants in each country and an equal distribution of gender. In Paper VI, we used the data collected from Germany. In Paper VII, we used the data collected from both Germany and Japan. The experimental setup is illustrated in Figure 2.6. A total of 60 trials were conducted for each participant, including 15 trials without any crossing facility, 15 trials with a zebra crossing, and 30 trials without zebra crossings involving the presence of risky and safe virtual pedestrian avatars. Vehicles maintained a constant speed of 30 km/h and only stopped at zebra crossings if participants were nearby. The gaps between cars per lane were uniformly sampled between 2.5 and 8.5 seconds. The data was collected after obtaining approval from the local ethical review boards in each country.



(a) Baseline and virtual pedestrians

(b) Zebra environment

Figure 2.6: Overview of the experimental environment. Start (yellow) and goal (green) are visualized, and are alternated in every other trial. Cars were approaching from both directions with randomly sampled gaps per lane.

Table 2.4: Candidate input variables for predicting pedestrian intentions when interacting with multiple vehicles.

Variable (unit)	Description (Data type)
$T_w(s)$	Pedestrian waiting time before crossing (continuous)
$v_p(m/s)$	Pedestrian average walking speed (continuous)
N_{en}	Number of unused effective gaps at near lane (discrete)
N_{ef}	Number of unused effective gaps at far lane (discrete)
N_{eb}	Number of unused effective gaps for both lanes (discrete)
$M_{en}(s)$	Largest missed effective gap at near lane (continuous)
$M_{ef}(s)$	Largest missed effective gap at far lane (continuous)
$M_{eb}(s)$	Largest missed effective gap for both lanes (continuous)
N_{cn}	Number of unused car gaps at near lane (discrete)
N_{cf}	Number of unused car gaps at far lane (discrete)
N_{cb}	Number of unused car gaps for both lanes (discrete)
$M_{cn}(s)$	Largest missed car gap at near lane (continuous)
$M_{cf}(s)$	Largest missed car gap at far lane (continuous)
$M_{cb}(s)$	Largest missed car gap for both lanes (continuous)

Variables describing pedestrian behavior are collected from the experiment recordings, including the pedestrian's average crossing velocity and the selected and accepted gap for crossing. Besides, we also focus on the gap-related measures. We calculate the *car gap* in traffic from the perspective of vehicles by determining the temporal distance between them. We also consider the participant's ego-movement and observe the *effective gaps* using an automated stopwatch calculating the time between a first car passing and a second one arriving at the participant's spatio-temporal position. For each way of computation, gaps are computed for the near and far lanes separately, as well as the synchronized gap for both lanes. All calculations are in seconds, but conversion to meters is straightforward due to the vehicles' constant speed. The observed variables are listed in Table 2.4.

Chapter 3

Summary of Appended Papers

3.1 Paper I

Chi Zhang and Christian Berger, *"Pedestrian Behavior Prediction Using Deep Learning Methods for Urban Scenarios: A Review"*, *IEEE Transactions on Intelligent Transportation Systems, 2023*

Objective

This research aims to comprehensively analyze and categorize existing methods for pedestrian behavior prediction, focusing on both trajectory and intention prediction using deep learning models in urban scenarios.

Methodology

This literature review systematically analyzes existing papers on pedestrian behavior prediction using deep learning methods. Through direct searches on IEEE Xplore and Google Scholar, combined with snowballing techniques, we reviewed 92 papers from 2016 to 2021. The selection criteria focus on pedestrian behavior prediction and deep learning, excluding works on drivers, robots, detection, and tracking. We expanded existing taxonomies, and categorized studies based on prediction tasks (trajectory, intention, joint prediction), input data types, model features, and network structures. We compared methods using common evaluation metrics and datasets, identifying research gaps and suggesting future research directions.

Results and Contributions

This research has provided a comprehensive review of pedestrian behavior prediction models using deep learning algorithms, drawing from 92 papers. The original contributions included a detailed analysis and categorization of existing literature, an introduction to publicly available datasets and evaluation metrics, a comparison of state-of-the-art algorithms, and a discussion of the advantages and drawbacks of existing algorithms. The study also identified research gaps and suggested potential directions for future improvements in prediction algorithms.

3.2 Paper II

Chi Zhang, Christian Berger, and Marco Dozza, *"Social-IWSTCNN: A Social Interaction-Weighted Spatio-Temporal Convolutional Neural Network for Pedestrian Trajectory Prediction in Urban Traffic Scenarios", In proceedings of the 2021 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2021*

Objective

This research aims to improve the accuracy of predicting pedestrian trajectories and reducing computational costs, by considering social interactions between pedestrians.

Methodology

We proposed the Social Interaction-Weighted Spatio-Temporal Convolutional Neural Network (Social-IWSTCNN) to predict pedestrian trajectories in urban traffic scenarios. We used a 3.2-second observed trajectory to predict a 4.8-second future trajectory. The Social-IWSTCNN framework consists of three key components: (1) the Social Interaction Extractor, which extracts spatial features and interaction weights between pedestrians without constructing a graph representation, directly using spatial features extracted from the observed locations relative to the last frame; (2) Temporal Convolutional Networks (TCNs), which extract temporal features from the spatial and social features to generate spatio-temporal features; and (3) Time-Extrapolator Convolutional Networks, which predict future trajectory distributions by applying convolutional networks to the spatio-temporal features, sampling the predicted Gaussian distributions to obtain future trajectories.

Results and Contributions

In this research, we have proposed the Social-IWSTCNN model, which effectively learns social interactions among pedestrians using a sub-network. The key innovation was the Social Interaction Extractor, a novel structure that captures pedestrian interactions in a data-driven manner, instead of using a fixed non-linear function approach used by the previous state-of-the-art model Social-STGCNN. Compared with five baseline methods, including linear regression, Naive LSTM, Social-LSTM [42], Social-GAN [40], and Social-STGCNN [41], our model outperformed existing approaches on both ADE and FDE. Furthermore, our model achieved faster inference speed by avoiding graph construction and non-linear interaction weight computation. This novel approach achieved accuracy improvements on ADE and FDE by 1.50% and 1.82%, respectively. Furthermore, our proposed network demonstrated speed improvements, with a 4.7x increase in total inference speed and a 54.8x increase in data pre-processing speed compared to Social-STGCNN.

3.3 Paper III

Chi Zhang and Christian Berger, *"Learning the Pedestrian-Vehicle Interaction for Pedestrian Trajectory Prediction", In proceedings of 2022 the 8th International Conference on Control, Automation and Robotics (ICCAR). IEEE, 2022*

Objective

This research aims to propose a deep learning approach that can better understand the interaction between pedestrians and vehicles and more accurately predict pedestrian trajectories in urban traffic scenarios.

Methodology

This research improves pedestrian trajectory prediction by considering three types of features: spatial features, social interaction features, and pedestrian-vehicle interaction features. Spatial features are embedded from relative pedestrian positions between consecutive time steps to capture spatial information. Social interaction features are extracted from the relative positions between pedestrians using an MLP sub-network. The interactions between pedestrians and vehicles are extracted using the proposed pedestrian-vehicle interaction (PVI) extractor, considering pedestrian-vehicle relative positions and vehicle movement states. These features are aggregated and fed into LSTM-based models and CNN-based models. The models are trained on the Waymo Open Dataset to predict pedestrian trajectories for a 4.8-second horizon, based on 3.2 seconds of observed data.

Results and Contributions

This study has proposed a novel Pedestrian-Vehicle Interaction (PVI) extractor for pedestrian trajectory prediction, to capture the complex interactions between pedestrians and vehicles. Integrated into LSTM and CNN-based models, the PVI extractor led to performance improvements, outperforming previous leading models such as Social-LSTM [42], Social-GAN [40], Social-STGCNN [41] and Social-IWSTCNN [25]. Results have shown that when integrated with LSTM-based models, the proposed PVI extractor improved ADE and FDE by 7.46% and 5.24%, respectively. Similarly, when integrated with CNN-based models, it improved ADE and FDE by 2.10% and 1.27%, respectively. The proposed algorithms were trained and evaluated on the Waymo Open Dataset, which contains real-world urban traffic data, demonstrating their effectiveness in predicting trajectories in urban traffic scenarios.

3.4 Paper IV

Chi Zhang, Zhongjun Ni, and Christian Berger, *"Spatial-Temporal-Spectral LSTM: A Transferable Model for Pedestrian Trajectory Prediction"*, *IEEE Transactions on Intelligent Vehicles. 2023*

Objective

This research aims to explore the transferability of deep learning models and propose a transferable model that can be generalized to other new and unseen datasets.

Methodology

Our proposed "Spatial-Temporal-Spectral (STS) LSTM" model integrates spatial, temporal, and spectral information to improve trajectory prediction accuracy and model transferability. The framework comprises three key components: a spatial-temporal-spectral feature representation module that captures motion patterns, an LSTM encoder-decoder prediction structure, and a trajectory sampling module that uses negative log-likelihood (NLL) loss for probabilistic prediction, capturing uncertainty in trajectory predictions. The study investigates model transferability through non-transfer and transfer tasks, evaluating performance on both scenarios and using pre-processed 2D real-world coordinates to mitigate data source differences.

Contributions

This study has proposed the Spatial-Temporal-Spectral (STS) LSTM model, which captures general pedestrian motion patterns and shows strong performance on both non-transfer and transfer tasks. We have proposed a novel representation of pedestrian trajectory input features in spatial, temporal, and spectral domains to improve the model's ability to capture motion patterns and transferability. We have also investigated the transferability of LSTMs, CNNs, and Transformers, comparing the L2 loss and negative log-likelihood (NLL) loss for trajectory prediction, and conducted quantitative and qualitative analysis on model transferability. This study takes trajectories as input, and identifies effective components that can be integrated into more complex models. The STS LSTM model has shown better performance on source and target datasets with faster inference speeds compared to state-of-the-art methods. This research has advanced the understanding of pedestrian trajectory prediction and highlights the importance of model transferability in real-world applications. The proposed STS LSTM model has offered a promising approach across diverse datasets and scenarios.

3.5 Paper V

Chi Zhang, Amir Hossein Kalantari, Yue Yang, Zhongjun Ni, Gustav Markkula, Natasha Merat, and Christian Berger, *"Cross or Wait? Predicting Pedestrian Interaction Outcomes at Unsignalized Crossings", In proceedings of 2023 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2023*

Objective

This research aims to use machine learning algorithms to understand and predict pedestrian crossing behavior when interacting with a vehicle at unsignalized crossings. We focus on pedestrians' "cross or wait" decisions, their crossing initiation time, and crossing duration. We also analyze the influence of input factors on pedestrian behavior.

Methodology

This research uses data from a distributed simulator study conducted by Kalantari et al. [53] involving 64 participants (32 drivers paired with 32 pedestrians) interacting under different scenarios within a simulated urban road environment, resulting in 1279 collected trials. Input features for predictive models include time to arrival, pedestrian waiting time, the presence of zebra crossings, and demographic information like age and gender. We also considered personality traits such as social value orientation and sensation seeking. Machine learning models, including logistic regression, linear regression, support vector machine, random forest, and neural networks, are developed and compared for prediction. We used five-fold cross-validation for training and testing. We used accuracy and F1 score as evaluation metrics for classification tasks and evaluated regression tasks using MAE and RMSE.

Results and Contributions

In this study, we used distributed simulator data to predict pedestrian-vehicle interaction and established baseline models using logistic regression and linear regression as predictability benchmarks. We have developed machine learning models for prediction and investigate various input features. The neural network model we developed improved accuracy and F1 score by 4.46% and 3.23% for crossing decision prediction, and reduced MAE and RMSE by 30.84% and 21.56% for crossing initiation time prediction, and by 35.00% and 30.14% for crossing duration prediction, respectively. We have also conducted an ablation study to analyze pedestrian-vehicle interaction factors, providing insights into model selection in scenarios with partial input information and evaluating the performance when lacking information on drivers' and pedestrians' age, gender, SVO, and AISS.

3.6 Paper VI

Chi Zhang, Janis Sprenger, Zhongjun Ni, and Christian Berger, *"Predicting and Analyzing Pedestrian Crossing Behavior at Unsignalized Crossings"*, *In 2024 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2024*

Objective

The research aims to use machine learning algorithms to predict and analyze pedestrian crossing behavior at unsignalized crossings, considering complex scenarios involving multiple vehicles and pedestrians. We focus on gap selection and zebra crossing usage and analyze factors that influence pedestrian behavior.

Methodology

We used data from a virtual reality simulator experiment conducted by Sprenger et al. [97]. We focused on data from the study conducted in Germany, involving 60 participants (30 females) navigating a virtual street environment for crossing, consisting of 3585 trials. Input features include pedestrian waiting time, average walking speed, number of unused effective and car gaps, and largest missed effective and car gaps from near, far, and both lanes. Machine learning models are used and compared for prediction. We identified the three most important features of each model and analyzed their impact on pedestrian crossing behavior. Five-fold cross-validation is used, with input variables normalized into standard normal distributions for better convergence and stability during model training. For evaluation, we used MAE and RMSE for gap selection prediction, and prediction accuracy and F1 score for zebra crossing usage prediction.

Results and Contributions

We have proposed and evaluated machine learning models for predicting pedestrian gap selection behavior and zebra crossing usage. Neural network models achieved the best performance, with a mean absolute error of 1.07 seconds for gap selection prediction and 94.27% prediction accuracy for zebra crossing usage prediction. Additionally, the research identified the most important features of each model, including the number of unused car gaps, the largest missed car gap, pedestrian waiting time, and pedestrian average walking speed, and investigated their impacts on pedestrian crossing behavior. We have also investigated group behavior and found that pedestrians tend to follow the behaviors of leading agents. This research has improved the understanding of pedestrian crossing behavior predictability and identified key factors influencing these decisions, thereby contributing to the advancement of intelligent vehicle systems.

3.7 Paper VII

Chi Zhang, Janis Sprenger, Zhongjun Ni, and Christian Berger, *"Predicting Pedestrian Crossing Behavior in Germany and Japan: Insights into Model Transferability"*, *In submission to IEEE Transactions on Intelligent Vehicles. 2024*

Objective

This research aims to explore the differences in pedestrian crossing behavior at unsignalized crossings between Germany and Japan and to investigate the transferability of machine learning models.

Methodology

We used simulator data collected by Sprenger et al. [97] from Germany and Japan, involving 120 participants (60 from each country, equally distributed by gender) navigating a virtual street environment for crossing. We developed and evaluated machine learning models separately for each country to investigate the similarities and differences in road crossing behavior. This comparative analysis focuses on model predictability, identifying key features, and understanding their influence on pedestrian behavior. To explore model transferability, we evaluated the performance of models trained on data from one country when applied to data from the other. To enhance transferability, we used unsupervised learning, specifically clustering, to reduce intra-dataset variance. Evaluation metrics include MAE and RMSE for gap selection prediction and prediction accuracy and F1 score for zebra crossing usage prediction.

Results and Contributions

We have proposed and evaluated machine learning models to predict pedestrian gap selection behavior and zebra crossing usage using data from studies conducted in Germany and Japan, comparing the similarities and differences in their crossing behaviors. The results showed that pedestrians from the study in Japan selected and accepted larger gaps, waited longer, and walked faster when crossing compared to participants in Germany, indicating differences in crossing behavior between the two countries. Participants from both countries shared similar key factors influencing crossing behavior, suggesting the possibility of developing transferable models. We evaluated the transferability of models trained on data from one country to another and proposed methods using unsupervised learning to incorporate cluster information, thereby improving transferability and performance on test sets from both countries.

3.8 Connections Between Appended Papers

The appended papers collectively contribute to improving pedestrian behavior prediction using machine learning methods. Figure 3.1 shows how they

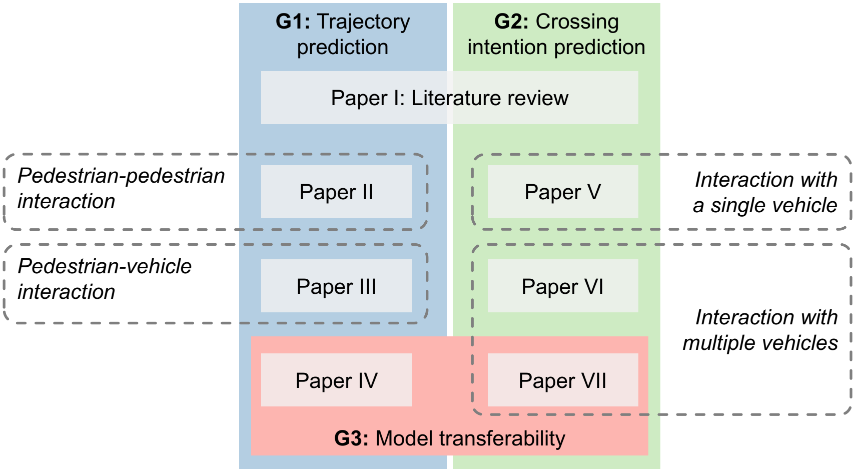

G1: Trajectory prediction
G2: Crossing intention prediction
Paper I: Literature review
Pedestrian-pedestrian interaction
Paper II
Paper V
Interaction with a single vehicle
Pedestrian-vehicle interaction
Paper III
Paper VI
Interaction with multiple vehicles
Paper IV
Paper VII
G3: Model transferability

Figure 3.1: Connections between appended papers and their contributions to research goals.

contribute to the different research goals, and the key differences between the papers.

Paper I provides a comprehensive literature review, covering the state-of-the-art models in pedestrian behavior prediction including both trajectory prediction (G1) and crossing intention prediction (G2).

Papers II, III, and IV focus on improving trajectory prediction models, contributing to G1. Paper II explores social interactions among pedestrians, introducing a new interaction modeling structure that improves accuracy and inference speed. Paper III builds upon this by incorporating pedestrian-vehicle interactions. Paper IV expands the focus by investigating the transferability of trajectory prediction models across different scenarios, contributing to both G1 and G3.

Papers V, VI, and VII focus on pedestrian crossing intention prediction, addressing G2. Paper V models pedestrian interaction with a single vehicle, while Papers VI and VII explore pedestrian interactions with multiple vehicles, reflecting more complex scenarios. Paper VII further explores model transferability (G3) by comparing crossing behavior in different countries.

Chapter 4

Discussions

The discussion chapter is organized as follows: We begin by discussing models for pedestrian trajectory prediction. Following this, we investigate methods for pedestrian intention prediction and analyze the key factors influencing pedestrian behavior. We then explore model transferability before outlining the key contributions. Finally, we address the limitations and propose directions for future research.

4.1 Trajectory Prediction in Urban Scenarios

Pedestrian trajectory prediction provides detailed spatial and temporal information, which can be used for collision avoidance or to assist intelligent vehicles in planning their future path. In Papers II, III, and IV, we have focused on trajectory prediction in urban scenarios. We compared the prediction results with previous state-of-the-art models using ADE and FDE metrics, as detailed in Table 4.1, where our models showed continued improvements.

This section discusses the prediction methods, the interaction features within the models, and the inference speeds to offer valuable insights for future research and establish general guidelines for developing pedestrian trajectory prediction models.

4.1.1 Prediction Methods for Trajectory Prediction

Paper I has categorized existing deep learning methods for pedestrian behavior prediction into sequential and non-sequential methods. Sequential methods include RNNs along with their improved version LSTMs, GANs, and Transformers. Non-sequential methods include CNNs, GNNs, and other artificial neural networks. In Papers II, III, and IV, we compared both sequential and non-sequential deep learning methods to find the most suitable ones for pedestrian prediction.

In both Papers II and III, CNN-based models, including our proposed model Social-IWSTCNN and the previous state-of-the-art model Social-STGCNN,

Table 4.1: Comparison of trajectory prediction models on the Waymo Open Dataset using Average Displacement Error (ADE) and Final Displacement Error (FDE) metrics. Upper section: models using L2 loss. Lower section: models using NLL loss. Units: meters. Smaller errors indicate better performance.

Model Name	ADE	FDE	Comments
Social-LSTM [42] (2016)	0.402	0.840	LSTM-based
Social-GAN [40] (2018)	0.386	0.826	LSTM-based
SI-PVI-LSTM [26] (Paper III)	0.372	0.796	LSTM-based
Transformer [66] (2021)	0.363	0.782	Transformer-based
Social-STGCNN [41](2020)	0.334	0.550	CNN-based
Social-IWSTCNN [25] (Paper II)	0.329	0.540	CNN-based
SI-PVI-Conv [26] (Paper III)	0.327	0.543	CNN-based
SLS-LSTM [27] (Paper IV)	**0.284**	**0.532**	LSTM-based

outperformed previous LSTM-based methods such as Social-LSTM and Social-GAN in pedestrian trajectory prediction tasks. This is because LSTM-based models tend to accumulate errors, whereas CNN-based models do not have this drawback. Furthermore, CNN-based models may better represent pedestrian motion states by directly embedding features from spatial and temporal information, unlike LSTMs that rely on hidden states. Besides, different evaluation methods contribute to differences in error. CNN-based models (e.g., [25], [26], [41]) use NLL loss and select the best result from 20 trials, while LSTM-based models (e.g., [40], [42]) use L2 loss and evaluate a single deterministic result, which can result in larger errors in LSTM evaluations.

In Paper IV, we explored trajectory prediction accuracy and model transferability by comparing various prediction backbones (LSTMs, CNNs, and Transformers) with two different loss functions (L2 and NLL). Our findings showed that LSTMs combined with NLL loss produced the best results. Therefore, we used LSTMs as the prediction backbone. We conducted six experimental settings, testing the combination of three aforementioned network structures with two loss functions. All models used the spatial positions of individual pedestrian trajectories in temporal order as input features.

For models using L2 loss, the Transformer outperformed other models, showing the least errors as shown in Table 4.1, and best transferability. This aligns with the findings of Giuliari et al. [66], which indicated that Transformers outperformed LSTM-based approaches due to their attention mechanism. CNNs performed poorest with L2 loss, possibly because they are not designed to capture time series dependencies and may fail to learn pedestrian motion patterns without considering randomness. LSTMs, while slightly behind Transformers, achieved competitive results due to their effectiveness in shorter sequence predictions (12 time steps for 4.8 seconds). In this thesis, we only predict short sequences, as in automated driving scenarios, pedestrian interactions last only a few seconds. Therefore, LSTMs are a good choice for prediction in this scenario.

Probabilistic models using NLL loss showed significantly lower errors compared to deterministic models using L2 loss. This is because probabilistic models evaluate the best of 20 samples, comparing the "upper-bound" of achievable performance, while L2 loss-based predictions provide deterministic "average" results that lose the randomness of pedestrian behavior. NLL loss, by allowing sample-based predictions, better preserves this randomness and improves prediction results. Among models using NLL loss, the LSTM model performed best, and demonstrated better transferability, while the Transformer model showed a comparatively lower performance. Therefore, we used the LSTM model with NLL loss and achieved impressive performance.

Besides, in Paper IV, we have integrated the spectral feature to capture the moving pattern of pedestrians and greatly improved the performance, as shown in Table 4.1. The improvements demonstrate the effectiveness of spectral features in capturing pedestrian motion patterns across both macro and micro scales. The strong performance of our proposed models demonstrates their potential for integration into automated driving systems.

Recent models, such as Trajectron++ [98], Y-Net [99], and NSP [100], have made significant advancements by incorporating richer inputs and more sophisticated architectures. These models address the multi-modality of pedestrian behavior, modeling interactions between pedestrians as well as interactions with features from the environment. These state-of-the-art models are evaluated on the ETH/UCY datasets.

Trajectron++ [98] focuses on forecasting dynamically-feasible trajectories for various traffic agents (e.g., cars, buses, and pedestrians) using heterogeneous input data such as camera images, LiDAR, and maps. The model employs a sophisticated spatiotemporal graph structure to represent agents and their interactions, and a Conditional Variational Autoencoder (CVAE) framework to account for multi-modality in predictions.

Y-Net [99] further improves multi-modal trajectory prediction by addressing both epistemic uncertainty, which related to long-term goals, and aleatoric uncertainty, which related to intermediate waypoints. It uses past trajectories and RGB images as inputs, resulting in enhanced prediction accuracy.

Yue et al. [100] proposed Neural Social Physics (NSP) models that combine traditional physics-based modeling with deep learning networks. It explicitly models goal attraction, inter-agent repulsion, and environmental repulsion. By integrating these factors into its neural network, NSP achieves state-of-the-art results on ETH/UCY datasets.

Although our models were not directly evaluated on the ETH/UCY datasets as we used the Waymo Open Dataset instead, we evaluated transferability in Paper IV by training on the Waymo Open Dataset and testing on ETH/UCY datasets. While this might not be a completely fair comparison due to differences in training datasets and scenarios, it provides a point of reference for comparing our models with the state-of-the-art algorithms. The results of the NSP [100], Y-net [99], and Trajectron++ [98] models from their original papers, along with comparisons to other state-of-the-art models, are listed in Table 4.2.

While models such as Trajectron++, Y-Net, and NSP achieve impressive accuracy, they come with a higher computational cost due to their complex structures and reliance on multiple data types. In contrast, our models focus on simplifying model architectures while exploring the impact of individual components, which could be potentially integrated into more complex models.

Table 4.2: The ADE/FDE metrics (in meters) of trajectory prediction models for model transferability. Upper section: results from the original papers without testing the transferability, where the models were trained directly on ETH/UCY datasets. Lower section: models retrained on the Waymo Open Dataset, and the transferability is evaluated on ETH/UCY datasets. Methods marked with the star *: used the best of 20 samples. Smaller number indicates better performance. **Bold**: best, underline: second best.

Model (Year)	Waymo test (Source data)	ETH/UCY (Target data)	Inter-actions	Scene
SoPhie* [70]	– / –	0.540 / 1.150	✓	✓
Social BiGAT* [71]	– / –	0.480 / 1.000	✓	✓
HSTA* [101]	– / –	0.400 / 0.790	✓	–
Trajectron++ [98]	– / –	0.370 / 0.910	✓	–
Trajectron++* [98]	– / –	0.190 / 0.410	✓	–
Y-net* [99]	– / –	0.180 / 0.270	✓	✓
NSP-SFM* [100]	– / –	**0.170** / **0.240**	✓	✓
Social LSTM [42]	0.393 / 0.841	0.486 / 1.013	✓	–
TF Individual [66]	0.363 / 0.782	0.448 / 0.945	–	–
Social STGCNN* [41]	0.335 / 0.550	0.346 / 0.584	✓	–
Social-IWSTCNN* (Paper II) [25]	0.328 / 0.538	0.332 / 0.558	✓	–
DMRGCN* [102]	**0.275** / **0.468**	**0.300** / **0.520**	✓	–
STS LSTM* (Paper IV) [27]	0.284 / 0.532	0.316 / 0.568	–	–

4.1.2 Trajectory Prediction Considering Interactions

Paper I has demonstrated that pedestrian behavior is influenced by interactions with other road users. Therefore, in Papers II and III, we consider pedestrians' interactions when predicting their trajectories.

By extracting the interaction features using our designed sub-network, our models outperformed existing models, as shown in Table 4.1. Paper II has considered social interaction features between pedestrians and proposes the Social-IWSTCNN [25] model. Unlike the previous state-of-the-art Social-STGCNN [41] model, which used a hand-crafted non-linear function based on distances to describe interaction relationships, our method uses pedestrians' velocities and relative positions as inputs to learn interaction weights with a deep learning sub-network. Our proposed Social-IWSTCNN model outperforms state-of-the-art methods, including Social-LSTM [42], Social-GAN [40], and Social-STGCNN [41], on ADE and FDE metrics. This demonstrates the effectiveness of using deep learning to learn interaction relationships, achieving better accuracy than hand-crafted weights or pooling layers.

In Paper II, we have compared algorithms across different traffic densities by dividing the Waymo Open Dataset into three groups based on the number of pedestrians. Compared with the Social-STGCNN [41] model, our model showed only marginal improvements in dense scenarios but significant improvements in less crowded scenarios, with a 17.3% improvement in ADE and 16.8% in FDE. One possible explanation is that the Social-STGCNN's hand-crafted function is tailored for densely populated scenarios in the ETH [92] and UCY [93] datasets, making it only effective for crowded scenarios. This highlights the adaptability of deep learning methods to various traffic scenarios compared to manually designed interaction weight functions.

In Paper III, we have introduced pedestrian-vehicle interaction (PVI) features in addition to social interaction (SI) features, each captured by separate sub-networks. The PVI features were extracted using vehicles' velocities and their relative positions to pedestrians as inputs. We applied the proposed sub-network to both LSTM-based and CNN-based prediction structures.

For LSTM-based models, our proposed SI-PVI-LSTM outperformed naive LSTM, Social-LSTM [42], and Social-GAN [40], demonstrating the importance of pedestrian-vehicle interactions. Compared to Social-GAN, our SI-PVI-LSTM achieved better results without the need for the complex and challenging-to-train GAN structure, suggesting that considering influencing factors can achieve improvements at a minimal computational cost. For CNN-based models, the SI-PVI-Conv model achieved better ADE compared to LR, Social-STGCNN [41], and Social-IWSTCNN [25] (as proposed in Paper II), though it did not improve FDE compared to Social-IWSTCNN. This could be due to two factors: first, using vehicle information only from the observation period potentially lacks sufficient information for long-term prediction, and second, considering all vehicles within sensor range without accounting for pedestrian orientation may introduce noises. Addressing these issues, such as updating vehicle information during the prediction horizon or incorporating pedestrian orientation and direction, could potentially improve performance. Besides, we found that using either social or pedestrian-vehicle interaction alone did not achieve optimal accuracy. The best performance was achieved by incorporating both social and pedestrian-vehicle information, highlighting the contributions of both interactions to performance enhancement.

4.1.3 Inference Speed

The algorithms should have the potential to be applied to automated vehicles, which demands considering real-time performance. While striving for higher prediction accuracy by incorporating more information and using more complex algorithms, it is also crucial to consider inference speed. In this section, we compare the inference speed of different models, as shown in Table 4.3. Our training and evaluation were conducted using an Nvidia GeForce RTX 2080 Ti GPU.

Table 4.3: Inference speed comparison. **Bold**: fastest.

Model (Year)	Average inference time per sequence (ms)
Social-STGCNN (2020) [41]	15.81
Social-IWSTCNN (2021) [25] (Paper II)	3.38
SI-PVI-Conv (2022) [26] (Paper III)	3.39
DMRGCN (2021) [102]	31.13
STS LSTM (2023) [27] (Paper IV)	**3.08**

Paper II has compared the inference speed between our proposed model Social-IWSTCNN and Social-STGCNN [41]. Our model showed an inference speed of 3.38 ms per sequence, 4.7 times faster compared to Social-STGCNN. The speed improvement is because of two key changes: first, we avoided non-linear calculations for attention weight computation, improving the inference speed by 2.7 times to 5.83 ms per sequence; second, we removed graph construction, further accelerating the inference speed.

Paper III has compared the inference speeds of Social-IWSTCNN [25] and SI-PVI-Conv. The inference speed of SI-PVI-Conv was measured at 3.39 ms per sequence, which closely matches Social-IWSTCNN's inference speed of 3.38 ms per sequence. This indicates that the computation of pedestrian-vehicle interaction features does not significantly increase inference time while contributing to accuracy improvement.

In Paper IV, we have compared the inference speeds of four competitive models: Social-STGCNN [41], Social-IWSTCNN [25], DMRGCN [102], and our proposed STS LSTM model. Our proposed model achieved the fastest inference speed of 3.08 ms per sequence. Although DMRGCN has better prediction accuracy, its complex structure resulted in a prediction time of 31.13 ms, which is 10 times slower than our proposed model.

These findings demonstrate the possibility of deploying our proposed models in real-world scenarios due to their efficient computational performance and competitive accuracy.

4.2 Intention and Interaction Prediction at Unsignalized Crossings

Accurate prediction of pedestrian crossing intention allows automated vehicles to make better decisions, reducing the risk of potential conflicts and collisions. As discussed in Section 1.1.3, pedestrians are most likely to collide or conflict with vehicles during crossings. Therefore, in addition to trajectory prediction, we also focus on predicting pedestrian crossing intention.

As proposed in Paper I, the terminology "intention" is often interchangeably used with "actual actions in the future" in many studies, due to the difficulty in distinguishing between them without the help of questionnaires [31]. In this thesis, we do not differentiate pedestrian intention and actual behavior and investigate pedestrian behavior during crossings.

Predicting pedestrian crossing intention is particularly challenging at unsignalized crossings where the right of way is ambiguous, leading to frequent interactions between pedestrians and vehicles. We have investigated these interactions and pedestrian crossing behavior in Papers V, VI, and VII. This section discusses the prediction methods and the key factors that influence pedestrian crossing behavior.

4.2.1 Prediction Methods for Intention and Interaction Prediction

Machine learning algorithms have been used to predict pedestrian crossing behavior in Papers V, VI, and VII. As described in Section 2.4, for classification tasks, logistic regression, SVM, RF, and NN models were evaluated and compared. For regression tasks, linear regression, RF, and NN models were used. Our studies showed that NNs outperformed other algorithms, with RF also being competitive.

In Paper V, we investigated pedestrian crossing behavior in interactions with a single vehicle. For crossing decision prediction (cross or wait), the NN model achieved the best results. Compared to the logistic regression baseline, the NN model we proposed improved the accuracy and F1 score by 4.46% and 3.23% on the whole dataset, and by 9.38% and 11.52% on non-zebra crossing data.

For regression tasks, we predicted crossing initiation time and crossing duration, with NNs achieving the best results. For crossing initiation time prediction, the NN reduced MAE and RMSE by 30.84% and 21.56%, respectively. Comparisons of the distributions between predicted crossing initiation time values and ground truth show that, although RF models showed smaller errors, the NN model's predictions aligned more closely with the ground truth distribution, demonstrating NN model's better generalizability. For crossing duration prediction, the NN model achieved the best results, with MAE and RMSE of 0.282s and 0.446s, respectively, improving by 35.00% and 30.14% compared to the linear regression model.

However, when only limited input variables are available, the model choice depends on the available input features. For crossing decisions, the linear model performed stably with fewer features, indicating its dependency on key basic features. RF and NN models showed significant performance decreases with fewer features, suggesting a broader range of input dependency. The NN outperformed with personality traits (SVO and AISS), indicating its abilities for handling non-linearity, while RF performs best with age and gender included. For crossing initiation time and duration predictions, the linear models' errors increased slightly with fewer features, relying primarily on time to arrival, waiting time, and crossing location type. RF and NN models showed obvious larger errors, indicating their dependence on additional features like personality traits, age, and gender. These results indicate the importance of various factors and highlight the different capabilities of linear and non-linear models in handling complex prediction tasks.

In Papers VI and VII, we have investigated scenarios where pedestrians interact with multiple vehicles, focusing on time gap selection and zebra crossing usage. For gap selection prediction, the NN outperformed others in predicting the accepted gap for individual crossing scenarios at non-zebra crossings. The NN achieved a mean absolute error of 1.07 seconds, and the distribution of its predictions closely resembles the ground truth.

Our studies reveal that pedestrians' gap acceptance is influenced by leading agents, indicating their following behavior. Within the risky group, non-linear models demonstrated smaller errors compared to the linear model, suggesting increased non-linearity in pedestrian behavior. The NN performed the best in this group. Conversely, in the safe group, errors were significantly smaller, indicating more predictable pedestrian behavior. The RF model performed the best in this group. This indicates NNs' capability in handling scenarios with complexity and non-linearity.

For zebra crossing usage prediction, the NN model achieved the best performance with 94.27% predicting accuracy. Compared to logistic regression, a commonly used analysis model, our proposed NN model showed a 4.02% improvement in accuracy, demonstrating its effectiveness.

4.2.2 Factors that Influence Pedestrian Crossing Behavior Prediction

In Papers V, VI, and VII, the key factors that influence pedestrian behavior have been identified and analyzed for each model. For linear models, feature importance was determined by the coefficients in the regression function. The RF models compute feature importance by the mean and standard deviation of impurity decrease across all trees. For the NN model, permutation importance was used.

Pedestrians Interacting with a Single Vehicle

In Paper V, we have investigated scenarios where pedestrians interact with a single vehicle, investigating key factors influencing crossing behavior, including the presence of zebra crossings and time to arrival.

We investigated pedestrian crossing behavior at both zebra and non-zebra crossings. The results showed that crossing decisions at zebra crossings were more predictable, with models achieving about 91% accuracy and a 95% F1 score. At non-zebra crossings, non-linear models (RF and NN) significantly outperformed linear models. The NN model improved accuracy and F1 score by 8.75% and 10.82%, respectively, achieving 88.91% accuracy and an 86.63% F1 score, compared to the linear baseline's 80.16% accuracy and 75.81% F1 score. The NN model's superior performance in non-zebra crossings indicates its ability to handle non-linearity.

Non-zebra crossings presented shorter crossing initiation times and narrower distributions compared to zebra crossings, indicating quicker decision-making for pedestrians at non-zebra crossings. NNs provided more distributed predictions compared to the baseline linear regression models, demonstrating their capacity to capture variability in pedestrian crossing behavior across different locations.

Regarding time to arrival, prediction accuracy decreased with a shorter time to arrival, reflecting the increased difficulty due to shorter reaction times and heightened interactions between pedestrians and drivers. Non-linear models outperformed linear models in capturing the non-linearity in pedestrian-driver interactions.

Additionally, an ablation study was conducted to identify important input features. In addition to the presence of zebra crossings and time to arrival, the AISS of pedestrians and pedestrian waiting time were also significant across all models. In linear models, objective properties such as age and gender were more influential, whereas in non-linear models (RF and NN), personality traits like AISS and SVO were more critical for prediction. This indicates that personality traits contribute to crossing decision prediction in a non-linear manner.

Pedestrians Interacting with Multiple Vehicles

In Paper VI, we have investigated scenarios where pedestrians interact with multiple vehicles, focusing on key factors influencing pedestrian behavior.

For time gap selection behavior, we identified and analyzed the most important features affecting the size of the time gap pedestrians choose for crossing. These factors include the number of unused gaps, the largest missed gap, pedestrian waiting time, pedestrian walking speed, and the influence of other pedestrians. As the number of unused car gaps increases, pedestrians tend to accept smaller gaps, indicating a willingness to take riskier choices. Similarly, when pedestrians miss larger gaps, they tend to accept smaller gaps, compromising safety for convenience. On the contrary, longer waiting times lead pedestrians to select larger, safer gaps. This finding is consistent with Yannis et al.'s [54] results, suggesting that pedestrians who wait longer are more cautious and risk-averse. Pedestrians with faster walking speeds tend to accept smaller gaps. This finding aligns with the study by Wan and Rouphail [103], which correlates increased walking speed with the need for shorter gaps for safe crossing. This behavior can be interpreted in two ways: faster walkers require smaller gaps to cross safely, or individuals in a hurry prefer shorter gaps, leading to faster walking speeds.

We also investigated group behavior, simulating scenarios with leading pedestrians using virtual avatars. Two types of group behavior were explored: the risky group, where the leading pedestrian crosses at a 4-second gap, and the safe group, where the leading pedestrian crosses at a 6.5-second gap. The analysis reveals that pedestrians tend to follow the behavior of leading pedestrians, with the distributions of accepted gaps shifting towards the gap chosen by the leading pedestrian. In the risky group, while many pedestrians maintain safer choices, a significant number follow the risky behavior of crossing at a 4-second gap. In the safe group, most pedestrians follow the safer behavior of crossing at a 6.5-second gap.

For predicting zebra crossing usage, important features include the number of unused gaps and pedestrian waiting time. As the number of unused gaps increases, prediction accuracy decreases, indicating the increased difficulty in predicting zebra usage. Similarly, as waiting time increases, the predictability of zebra crossing usage decreases. Besides, time gap selection is also influenced by the choice of zebra crossing usage. Pedestrians using zebra crossings tend to accept smaller gaps than those not using them. This behavior could be due to two factors: first, pedestrians may prefer not to use zebra crossings when larger gaps are available; second, the presence of zebra crossings may decrease pedestrians' risk aversion as they expect vehicles to yield.

4.3 Model Transferability and Generalizability

4.3.1 Trajectory Prediction

For trajectory prediction, Paper IV proposes the "Spatial-Temporal-Spectral (STS) LSTM" model, which uses spatial, temporal, and spectral domain information. This model shows better transferability and prediction accuracy on target datasets without prior knowledge, outperforming many state-of-the-art models with a faster inference speed.

Experiments were designed to evaluate the models' transferability. All models were trained on the Waymo Open Dataset, which was collected in the United States. They were evaluated on the ETH-univ, ETH-hotel, UCY-zara1, UCY-zara2, and UCY-univ datasets, which were collected in Switzerland and Cyprus, without prior information about these new datasets. While DMRGCN [102] achieved the best average results on target datasets, it has a more complex structure and slower inference speed. On four out of five datasets, the ADE of the STS LSTM model was only marginally worse than DMRGCN, but the STS LSTM model was ten times faster, demonstrating its efficiency. Its simple LSTM encoding-decoding structure with spectral domain information outperformed other baseline methods, showing better transferability to unseen cases.

An ablation study confirms that combining temporal and spectral information improves both prediction accuracy and transferability. Comparing different input representations (temporal, spectral, and spatial-temporal-spectral), the combination of temporal and spectral information yielded the best prediction results and transferability. The study reveals that low-frequency components reflect macroscopic trends, while high-frequency components capture microscopic adjustments, demonstrating the importance of both temporal and spectral features in learning pedestrian motion patterns.

4.3.2 Crossing Intention and Interaction Prediction

In Paper VII, we have explored pedestrian crossing intention and their interaction with vehicles in Germany and Japan to provide insights into model transferability. Our machine learning models predict gap selection behavior and zebra crossing usage. We evaluated and analyzed the transferability of models trained on simulator data from both countries.

We predicted the accepted gap for pedestrians at non-zebra crossings and compared the behavior in both countries. Pedestrians from the study conducted in Japan tend to accept larger gaps and safety thresholds than those in Germany. They also waited longer than those in Germany, indicating a higher level of caution.

While differences exist, pedestrian behavior was influenced by similar key factors in both countries. As the number of unused car gaps increases, pedestrians from both countries tend to accept smaller gaps, indicating riskier choices. When the missed gaps exceed five, the accepted gap size stabilizes, suggesting a safety threshold. Pedestrians in both countries tend to accept smaller gaps after missing larger ones.

Pedestrians with faster walking speeds in both countries tended to choose shorter gaps for crossing. Interestingly, despite walking slightly faster, pedestrians from the study conducted in Japan still selected larger gaps. This preference for larger gaps suggests a safety-conscious behavior, and may also reflect a consideration for reducing disruptions to traffic flow.

In zebra crossing scenarios, participants from the study in Japan waited longer than those in Germany, both when using and not using zebra crossings, indicating greater caution. Pedestrians from the study conducted in Germany showed more non-linearity in their behavior, with longer waiting times and more unused gaps.

Neural network models demonstrated the best transferability when tested on data from the other country, maintaining top performance and showing similar important features for both countries. Although model accuracy decreases when transferred, the increase in error is minimal, indicating its robustness.

Considering the behavior differences between countries, Paper VII proposes a method using clustering information through unsupervised learning and improves model transferability for gap selection models.

4.4 Contributions

4.4.1 Contribution to Vehicular Automation

This thesis contributes to enhancing AD systems by providing advanced predictive models and insights into pedestrian behavior. We have provided a comprehensive review of existing deep learning methods on pedestrian behavior prediction, evaluating their strengths and weaknesses, recommending suitable approaches for various predictive tasks. This foundational analysis guides the development of novel models and strategies for enhancing pedestrian behavior prediction and vehicle safety.

Our algorithms, with enhanced prediction accuracy and faster inference speeds, provide earlier warnings for AD systems, allowing more time for reaction and control. We have proposed novel methods that consider social and pedestrian-vehicle interactions, improving both prediction accuracy and inference speed. We have also addressed model transferability, proposing a transferable model that can adapt to different datasets. These advancements enable AVs and driver assistance systems to better avoid pedestrian-vehicle collisions, ultimately contributing to pedestrian safety.

Our studies have provided insights into pedestrian crossing behavior and their interaction with vehicles. We have developed machine learning models for predicting pedestrian-vehicle interaction outcomes. These models enable intelligent vehicles to predict pedestrian crossing behavior, allowing for smoother and safer interactions. Vehicles can adjust their speed and trajectories based on predictive information, leading to proactive measures that prevent conflicts and collisions.

4.4.2 Contribution to Software Engineering

This thesis has contributed to the field of software engineering by driving the development and refinement of machine learning algorithms for trajectory prediction, especially designed for real-time, complex multi-agent interactions. By integrating diverse data sources and addressing complex interaction patterns, our research has improved the accuracy and inference speed of trajectory prediction software systems.

Our research has improved the explainability of machine learning models for pedestrian behavior prediction software by providing in-depth analyses of key factors and their impact. By providing in-depth insights into how various factors influence predictions, our work enhances the transparency and interpretability of these models. This contributes to a better understanding and trust in software systems designed for intelligent driving, which is essential for their effective deployment and user acceptance.

Our research addresses the challenge of transferability in software systems, demonstrating the capability of our predictive models to generalize across different scenarios and regions. By applying and testing these models in diverse contexts, including cross-country comparisons, we have established their robustness and adaptability. This contribution paves the way for applying these software solutions to various fields and situations, ensuring their relevance and utility beyond the initial scope.

4.5 Limitations

The limitations of appended papers highlight several key areas for improvement and further research.

In Paper I, we noted that existing datasets are mainly captured in high-income and middle-income countries, neglecting low-income countries where pedestrian fatalities are disproportionately high (36%). This limitation indicates the need for future research to focus on developing datasets and studies concerning low-income regions.

While Papers II, III, and IV used real-world data, our data-driven methods rely heavily on the quality of data collection and annotation, which is often costly. Low-resolution images impede the extraction of useful features, and improperly labeled data compromises algorithm training.

In Paper IV, we addressed model transferability challenges but focused on individual trajectories, limiting our models' ability to handle pedestrian intentions and complex interactions. Failures occur with sudden movements or interactions with other pedestrians or objects.

Besides, Papers II, III, and IV focused solely on using trajectory information as input for a faster inference speed. This limits the models' ability to capture more complex environmental and contextual information. In contrast, research studies such as Trajectron++ [98], Y-net [99], and NSP [100] incorporated richer information and achieved higher accuracy.

Furthermore, for Papers V, VI, and VII, while simulation studies ensure safety, they may not fully capture real-world risk perceptions and behaviors. Our models reflect the behavior observed within the simulated environment. The simulator studies used an untethered head-mounted virtual reality (VR) headset to display realistic traffic scenarios, and motion capture equipment to record participants' crossing behaviors. To closely emulate real-world behavior, we included training sessions for participants to familiarize themselves with the environment. However, the intentional absence of facing *real* risk in a simulated environment may lead to riskier crossing behaviors compared to real-life scenarios. Real-world validation and consideration of risk perceptions could enhance the applicability of our models in real-world contexts.

Addressing these limitations will lead to more robust and applicable pedestrian behavior prediction models, which will contribute to safer urban environments worldwide.

4.6 Future Work

Future work on trajectory prediction can focus on integrating intention-based and interaction-based models. Although Paper IV addressed model transferability challenges, it focused on individual trajectories, which limited its ability to handle pedestrian intentions and complex interactions. This approach may fail in scenarios involving sudden movements or interactions with other pedestrians or objects. Improving our models by considering intention-based or interaction-based components could potentially improve performance in these complex scenarios. Furthermore, generative models based on multi-model distributions may also improve the prediction results.

Some recent models for trajectory prediction such as NSP proposed by Yue et al. [100], combined the neural social physics (NSP) model into deep learning network, which improves both prediction accuracy and model explainability. Future work may also consider combining the traditional physics model-based algorithm into deep learning networks to make them more accurate and explainable.

Additionally, given that most deep learning models require large datasets, future work can explore recent advancements in large language models (LLM) and few-shot learning to reduce dependency on extensive labeling. This approach could help to address challenges related to data quality and annotation.

Leveraging pre-trained off-the-shelf models could also enhance performance and efficiency while reducing labeling costs.

Papers VI and VII investigated pedestrian gap selection, but treated gap duration as a continuous value, overlooking its inherent uncertainty. Future research should explore stochastic models that represent gap duration as distributions to account for its randomness. Besides, generative models based on multi-modal distributions could also be developed for future work. When predicting pedestrian intention, the posture of pedestrians could also be considered to provide additional information.

Our current studies on pedestrian crossing behavior rely on simulator data. To ensure that models built on this data can generalize to real-world scenarios, future work should focus on validating pedestrian behaviors and risk perceptions in real-world environments.Besides, while our research has primarily focused on pedestrian crossing intention and trajectory separately, future studies could investigate the combination between the two, and to explore how a pedestrian's intention influences their trajectory.

Chapter 5

Conclusions

This research has contributed to enhancing pedestrian safety for AD systems by improving pedestrian behavior prediction using machine learning models and analyzing pedestrian-vehicle interactions. We first reviewed existing methods in pedestrian behavior prediction and identified key research gaps. Then we advanced the field through the development of deep learning models that improve trajectory prediction accuracy and speed. By creating machine learning models that predict and analyze pedestrian crossing behavior, we have provided deeper insights into the factors influencing pedestrian crossing decisions. We have also addressed the critical challenge of model transferability and generalizability, improving the ability of these models to perform accurately across diverse datasets and different countries.

For pedestrian trajectory prediction, Paper I shows that integrating pedestrian interactions into prediction models has the potential to improve performance. Papers II and III demonstrate that extracting interaction features using our designed sub-networks improves pedestrian trajectory prediction in both accuracy and speed. Paper IV further concludes that integrating spatial, temporal, and spectral information into the model not only improves prediction accuracy but also enhances transferability across different scenarios.

For pedestrian intention prediction and interactions with vehicles, neural networks show better performance in handling nonlinear scenarios, achieving improved results compared to other machine learning algorithms. Paper V demonstrates that when pedestrians interact with a single vehicle, key factors, including the presence of zebra crossings, time to arrival, and personality traits, influence pedestrian crossing decisions. Papers VI and VII highlight that when pedestrians interact with multiple vehicles, key factors, including waiting time, time gaps, pedestrian walking speed, and the behavior of other pedestrians, play a crucial role in gap selection decisions. These factors are crucial in understanding and predicting how pedestrians make crossing decisions, providing valuable insights for developing safer traffic systems.

Regarding model transferability, Paper IV reveals that combining temporal and spectral information enhances model transferability. It effectively captures both macroscopic trends and microscopic adjustments in pedestrian motion patterns. Paper VII shows that neural network models exhibit strong cross-country transferability when predicting gap selection behavior. The comparative study between Germany and Japan demonstrates that pedestrians from the study conducted in Japan selected larger gaps and showed more cautious behavior. While pedestrian behavior varies between countries, our study finds that key factors and their impacts on pedestrian behavior are similar across countries. This consistency suggests that the models developed in this research have the potential to be adapted for use in diverse geographical contexts, contributing to the broader applicability and generalization of the findings.

References

[1] WHO, “Global status report on road safety 2023,” World Health Organization, Report, 2023 (cit. on pp. 1, 2).

[2] WHO, “Global status report on road safety 2018: Summary,” World Health Organization, Report, 2018 (cit. on pp. 1, 2, 4).

[3] W. H. Organization *et al.*, “Pedestrian safety: A road safety manual for decision-makers and practitioners,” 2013 (cit. on pp. 1–3, 5).

[4] U. N. G. Assembly, “Transforming our world: The 2030 agenda for sustainable development,” 2015. [Online]. Available: `https://sdgs.un.org/2030agenda` (cit. on p. 1).

[5] R. F. S. Job, “Policies and interventions to provide safety for pedestrians and overcome the systematic biases underlying the failures,” *Frontiers in Sustainable Cities*, vol. 2, p. 30, 2020 (cit. on p. 2).

[6] A Santacreu, “Monitoring progress in urban road safety,” *International Transport Forum Policy Papers, No. 79, OECD Publishing*, 2020 (cit. on pp. 2–4).

[7] W. H. Organization *et al.*, *Pedestrian safety: a road safety manual for decision-makers and practitioners.* World Health Organization, 2023 (cit. on p. 2).

[8] M. Peden, R. Scurfield, D. Sleet *et al.*, *World report on road traffic injury prevention.* World Health Organization, 2004 (cit. on p. 2).

[9] A. Värnild, P. Larm and P. Tillgren, “Incidence of seriously injured road users in a swedish region, 2003–2014, from the perspective of a national road safety policy,” *BMC public health*, vol. 19, no. 1, pp. 1–10, 2019 (cit. on p. 3).

[10] A. H. Do, S. A. Balk and J. W. Shurbutt, “Why did the pedestrian cross the road?” *Public Roads*, vol. 77, no. 6, 2014 (cit. on p. 3).

[11] P. L. Lane, K. J. McClafferty and E. S. Nowak, “Pedestrians in real world collisions,” *Journal of Trauma and Acute Care Surgery*, vol. 36, no. 2, pp. 231–236, 1994 (cit. on p. 3).

[12] P. Olszewski, P. Szagała, M. Wolański and A. Zielińska, “Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in poland,” *Accident Analysis & Prevention*, vol. 84, pp. 83–91, 2015 (cit. on p. 4).

[13] L. Rothman, A. W. Howard, A. Camden and C. Macarthur, "Pedestrian crossing location influences injury severity in urban areas," *Injury prevention*, vol. 18, no. 6, pp. 365–370, 2012 (cit. on p. 4).

[14] K. Rumar, "Transport safety visions, targets and strategies: Beyond 2000," *1st European Transport Safety Lecture. European Transport Safety Council, Brussels, Tech. Rep*, 1999 (cit. on p. 4).

[15] U NHTSA, "Critical reasons for crashes investigated in the national motor vehicle crash causation survey," *DOT HS*, vol. 812, p. 115, 2015 (cit. on p. 4).

[16] G. P. Santoso and D. Maulina, "Human error in traffic accidents: Differences between car driver and motorcyclist experiences," *Psychological Research on Urban Society*, vol. 2, no. 2, p. 12, 2019 (cit. on p. 4).

[17] P. Morgan, C. Alford and G. Parkhurst, "Handover issues in autonomous driving: A literature review.(2016)," *Project Report. University of the West of England, Bristol, UK*, 2016 (cit. on p. 4).

[18] E. Papa and A. Ferreira, "Sustainable accessibility and the implementation of automated vehicles: Identifying critical decisions," *Urban Science*, vol. 2, no. 1, p. 5, 2018 (cit. on p. 5).

[19] P. Ghorai, A. Eskandarian, M. Abbas and A. Nayak, "A causation analysis of autonomous vehicle crashes," *IEEE Intelligent Transportation Systems Magazine*, 2024 (cit. on p. 5).

[20] T. S. Combs, L. S. Sandt, M. P. Clamann and N. C. McDonald, "Automated vehicles and pedestrian safety: Exploring the promise and limits of pedestrian detection," *American journal of preventive medicine*, vol. 56, no. 1, pp. 1–7, 2019 (cit. on p. 5).

[21] Á. Török, "Do automated vehicles reduce the risk of crashes–dream or reality?" *IEEE transactions on intelligent transportation systems*, vol. 24, no. 1, pp. 718–727, 2022 (cit. on p. 5).

[22] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015 (cit. on p. 5).

[23] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015 (cit. on p. 5).

[24] I. H. Sarker, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 6, pp. 1–20, 2021 (cit. on pp. 5, 13).

[25] C. Zhang, C. Berger and M. Dozza, "Social-iwstcnn: A social interaction-weighted spatio-temporal convolutional neural network for pedestrian trajectory prediction in urban traffic scenarios," in *2021 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2021, pp. 1515–1522. DOI: 10.1109/IV48863.2021.9575958 (cit. on pp. 5, 16, 29, 36, 38–40).

[26] C. Zhang and C. Berger, "Learning the pedestrian-vehicle interaction for pedestrian trajectory prediction," in *2022 8th International Conference on Control, Automation and Robotics (ICCAR)*, IEEE, 2022, pp. 230–236. DOI: `10.1109/ICCAR55106.2022.9782673` (cit. on pp. 5, 36, 40).

[27] C. Zhang, Z. Ni and C. Berger, "Spatial-temporal-spectral lstm: A transferable model for pedestrian trajectory prediction," *IEEE Transactions on Intelligent Vehicles*, pp. 1–14, 2023. DOI: `10.1109/TIV.2023.3285804` (cit. on pp. 5, 36, 38, 40).

[28] C. Zhang, A. H. Kalantari, Y. Yang *et al.*, "Cross or wait? predicting pedestrian interaction outcomes at unsignalized crossings," in *2023 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2023, pp. 1–8. DOI: `10.1109/IV55152.2023.10186616` (cit. on pp. 5, 6, 21).

[29] C. Zhang, J. Sprenger, Z. Ni and C. Berger, "Predicting and analyzing pedestrian crossing behavior at unsignalized crossings," in *2024 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2024, pp. 1–8 (cit. on p. 5).

[30] C. Zhang, J. Sprenger, Z. Ni and C. Berger, "Predicting pedestrian crossing behavior in germany and japan: Insights into model transferability," in *In submission to Transactions on Intelligent Vehicles*, IEEE, 2024 (cit. on p. 5).

[31] C. Zhang and C. Berger, "Pedestrian behavior prediction using deep learning methods for urban scenarios: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 10, pp. 10 279–10 301, 2023. DOI: `10.1109/TITS.2023.3281393` (cit. on pp. 6, 13, 40).

[32] S. Ferguson, B. Luders, R. C. Grande and J. P. How, "Real-time predictive modeling and robust avoidance of pedestrians with uncertain, changing intentions," in *Algorithmic Foundations of Robotics XI*, Springer, 2015, pp. 161–177 (cit. on p. 6).

[33] N. Schneider and D. M. Gavrila, "Pedestrian path prediction with recursive bayesian filters: A comparative study," in *German Conference on Pattern Recognition*, Springer, 2013, pp. 174–183 (cit. on p. 6).

[34] H. Zhang, Y. Liu, C. Wang, R. Fu, Q. Sun and Z. Li, "Research on a pedestrian crossing intention recognition model based on natural observation data," *Sensors*, vol. 20, no. 6, p. 1776, 2020 (cit. on pp. 6, 16, 17).

[35] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing and G. Theraulaz, "The walking behaviour of pedestrian social groups and its impact on crowd dynamics," *PloS one*, vol. 5, no. 4, e10047, 2010 (cit. on pp. 6, 15).

[36] M. S. Shirazi and B. Morris, "Observing behaviors at intersections: A review of recent studies & developments," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2015, pp. 1258–1263 (cit. on p. 6).

[37] E. Ohn-Bar and M. M. Trivedi, "Looking at humans in the age of self-driving and highly automated vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 90–104, 2016 (cit. on p. 6).

[38] A. Rasouli and J. K. Tsotsos, “Autonomous vehicles that interact with pedestrians: A survey of theory and practice,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 900–918, 2019, ISSN: 1524-9050 (cit. on p. 6).

[39] D. Helbing and P. Molnar, “Social force model for pedestrian dynamics,” *Physical review E*, vol. 51, no. 5, p. 4282, 1995 (cit. on pp. 6, 15).

[40] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese and A. Alahi, “Social gan: Socially acceptable trajectories with generative adversarial networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2255–2264 (cit. on pp. 6, 14, 15, 28, 29, 36, 38, 39).

[41] A. Mohamed, K. Qian, M. Elhoseiny and C. Claudel, “Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 14 424–14 432 (cit. on pp. 6, 14–16, 28, 29, 36, 38–40).

[42] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei and S. Savarese, “Social lstm: Human trajectory prediction in crowded spaces,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 961–971 (cit. on pp. 6, 13, 15, 28, 29, 36, 38, 39).

[43] P. Zhang, W. Ouyang, P. Zhang, J. Xue and N. Zheng, “Sr-lstm: State refinement for lstm towards pedestrian trajectory prediction,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019 (cit. on pp. 6, 13).

[44] H. Xue, D. Q. Huynh and M. Reynolds, “Ss-lstm: A hierarchical lstm model for pedestrian trajectory prediction,” in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2018, pp. 1186–1194, ISBN: 1538648865 (cit. on pp. 6, 13).

[45] A. Rasouli, I. Kotseruba and J. K. Tsotsos, “Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, 2017, pp. 206–213 (cit. on pp. 6, 17).

[46] Z. Fang and A. M. López, “Is the pedestrian going to cross? answering by 2d pose estimation,” in *2018 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2018, pp. 1271–1276, ISBN: 1538644525 (cit. on pp. 6, 17).

[47] M. Chaabane, A. Trabelsi, N. Blanchard and R. Beveridge, “Looking ahead: Anticipating pedestrians crossing with future frames prediction,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 2297–2306 (cit. on p. 6).

[48] B. Yang, W. Zhan, P. Wang, C. Chan, Y. Cai and N. Wang, “Crossing or not? context-based recognition of pedestrian crossing intention in the urban environment,” *IEEE Transactions on Intelligent Transportation Systems*, 2021 (cit. on pp. 6, 16, 17).

[49] S. Zhang, M. Abdel-Aty, Y. Wu and O. Zheng, “Pedestrian crossing intention prediction at red-light using pose estimation,” *IEEE Transactions on Intelligent Transportation Systems*, 2021 (cit. on pp. 6, 16, 17).

[50] A. Rasouli, I. Kotseruba, T. Kunic and J. K. Tsotsos, “Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6262–6271 (cit. on p. 6).

[51] A. Gorrini, L. Crociani, G. Vizzari and S. Bandini, “Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation,” *Transportation research part F: traffic psychology and behaviour*, vol. 59, pp. 269–285, 2018 (cit. on p. 6).

[52] J. P. N. Velasco, Y. M. Lee, J. Uttley *et al.*, “Will pedestrians cross the road before an automated vehicle? the effect of drivers’ attentiveness and presence on pedestrians’ road crossing behavior,” *Transportation research interdisciplinary perspectives*, vol. 12, p. 100 466, 2021 (cit. on p. 6).

[53] A. H. Kalantari, Y. Yang, J. G. de Pedro *et al.*, “Who goes first? a distributed simulator study of vehicle–pedestrian interaction,” *Accident Analysis & Prevention*, vol. 186, p. 107 050, 2023 (cit. on pp. 6, 21, 31).

[54] E. P. G. Yannis and A. Theofilatos, “Pedestrian gap acceptance for mid-block street crossing,” *Transportation Planning and Technology*, vol. 36, no. 5, pp. 450–462, 2013. DOI: 10.1080/03081060.2013.818274. eprint: https://doi.org/10.1080/03081060.2013.818274 (cit. on pp. 6, 43).

[55] A. Graves, “Generating sequences with recurrent neural networks,” *arXiv preprint arXiv:1308.0850*, 2013 (cit. on p. 13).

[56] A. Graves and N. Jaitly, “Towards end-to-end speech recognition with recurrent neural networks,” in *International conference on machine learning*, PMLR, 2014, pp. 1764–1772 (cit. on p. 13).

[57] T. Fernando, S. Denman, S. Sridharan and C. Fookes, “Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection,” *Neural networks*, vol. 108, pp. 466–478, 2018 (cit. on p. 13).

[58] I. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014 (cit. on p. 14).

[59] S. Bai, J. Z. Kolter and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” *arXiv preprint arXiv:1803.01271*, 2018 (cit. on p. 14).

[60] N. Nikhil and B. Tran Morris, “Convolutional neural network for trajectory prediction,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018 (cit. on p. 14).

[61] A. Vaswani, N. Shazeer, N. Parmar *et al.*, “Attention is all you need,” in *31st Conference on Neural Information Processing Systems (NIPS)*, 2017 (cit. on p. 14).

[62] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le and R. Salakhutdinov, “Transformer-xl: Attentive language models beyond a fixed-length context,” *arXiv preprint arXiv:1901.02860*, 2019 (cit. on p. 14).

[63] J. W. Rae, A. Potapenko, S. M. Jayakumar and T. P. Lillicrap, “Compressive transformers for long-range sequence modelling,” *arXiv preprint arXiv:1911.05507*, 2019 (cit. on p. 14).

[64] I. Beltagy, M. E. Peters and A. Cohan, “Longformer: The long-document transformer,” *arXiv preprint arXiv:2004.05150*, 2020 (cit. on p. 14).

[65] N. Kitaev, Ł. Kaiser and A. Levskaya, “Reformer: The efficient transformer,” *arXiv preprint arXiv:2001.04451*, 2020 (cit. on p. 14).

[66] F. Giuliari, I. Hasan, M. Cristani and F. Galasso, “Transformer networks for trajectory forecasting,” in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, 2021, pp. 10 335–10 342 (cit. on pp. 15, 36, 38).

[67] C. Yu, X. Ma, J. Ren, H. Zhao and S. Yi, “Spatio-temporal graph transformer networks for pedestrian trajectory prediction,” in *European Conference on Computer Vision*, Springer, 2020, pp. 507–523 (cit. on p. 15).

[68] Y. Yuan, X. Weng, Y. Ou and K. Kitani, “Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting,” *arXiv preprint arXiv:2103.14023*, 2021 (cit. on p. 15).

[69] A. Syed and B. Morris, “Stgt: Forecasting pedestrian motion using spatio-temporal graph transformer,” in *2021 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2021 (cit. on p. 15).

[70] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi and S. Savarese, “Sophie: An attentive gan for predicting paths compliant to social and physical constraints,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1349–1358 (cit. on pp. 15, 38).

[71] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, S. H. Rezatofighi and S. Savarese, “Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks,” *arXiv preprint arXiv:1907.03395*, 2019 (cit. on pp. 15, 16, 38).

[72] Y. Huang, H. Bi, Z. Li, T. Mao and Z. Wang, “Stgat: Modeling spatial-temporal interactions for human trajectory prediction,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6272–6281 (cit. on p. 16).

[73] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio and Y. Bengio, “Graph attention networks,” in *International Conference on Learning Representations (ICLR)*, 2017 (cit. on p. 16).

[74] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016 (cit. on p. 16).

[75] B. Völz, K. Behrendt, H. Mielenz, I. Gilitschenski, R. Siegwart and J. Nieto, "A data-driven approach for pedestrian intention estimation," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2016, pp. 2607–2612 (cit. on pp. 16, 17).

[76] B. Völz, H. Mielenz, I. Gilitschenski, R. Siegwart and J. Nieto, "Inferring pedestrian motions at urban crosswalks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 544–555, 2018 (cit. on p. 16).

[77] I. Kotseruba, A. Rasouli and J. K. Tsotsos, "Benchmark for evaluating pedestrian action prediction," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1258–1268 (cit. on p. 16).

[78] Y. Ma, X. Zhu, S. Zhang, R. Yang, W. Wang and D. Manocha, "Trafficpredict: Trajectory prediction for heterogeneous traffic-agents," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 6120–6127, ISBN: 2374-3468 (cit. on p. 16).

[79] B. Liu, E. Adeli, Z. Cao *et al.*, "Spatiotemporal relationship reasoning for pedestrian intent prediction," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3485–3492, 2020 (cit. on p. 16).

[80] S. Eiffert, K. Li, M. Shan, S. Worrall, S. Sukkarieh and E. Nebot, "Probabilistic crowd gan: Multimodal pedestrian trajectory prediction using a graph vehicle-pedestrian attention network," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5026–5033, 2020 (cit. on p. 16).

[81] Y. Hu, S. Chen, Y. Zhang and X. Gu, "Collaborative motion prediction via neural motion message passing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6319–6328 (cit. on p. 16).

[82] S. Carrasco, D. F. Llorca and M. A. Sotelo, "Scout: Socially-consistent and understandable graph attention network for trajectory prediction of vehicles and vrus," in *2021 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2021 (cit. on p. 16).

[83] R. Chandra, U. Bhattacharya, A. Bera and D. Manocha, "Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8483–8492 (cit. on p. 16).

[84] R. Chandra, U. Bhattacharya, C. Roncal, A. Bera and D. Manocha, "Robusttp: End-to-end trajectory prediction for heterogeneous road-agents in dense traffic with noisy sensor inputs," in *ACM Computer Science in Cars Symposium*, 2019, pp. 1–9 (cit. on p. 16).

[85] R. Chandra, T. Guan, S. Panuganti *et al.*, "Forecasting trajectory and behavior of road-agents using spectral clustering in graph-lstms," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4882–4890, 2020, ISSN: 2377-3766 (cit. on p. 16).

[86] M. Chaabane, A. Trabelsi, N. Blanchard and R. Beveridge, "Looking ahead: Anticipating pedestrians crossing with future frames prediction," in *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020 (cit. on p. 17).

[87] B. Völz, H. Mielenz, G. Agamennoni and R. Siegwart, "Feature relevance estimation for learning pedestrian behavior at crosswalks," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, IEEE, 2015, pp. 854–860 (cit. on p. 17).

[88] S. K. Jayaraman, D. M. Tilbury, X. J. Yang, A. K. Pradhan and L. P. Robert, "Analysis and prediction of pedestrian crosswalk behavior during automated vehicle interactions," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020, pp. 6426–6432 (cit. on p. 17).

[89] X. Ying, "An overview of overfitting and its solutions," in *Journal of physics: Conference series*, IOP Publishing, vol. 1168, 2019, p. 022 022 (cit. on p. 17).

[90] D. C. Montgomery, E. A. Peck and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2021 (cit. on p. 17).

[91] A Liaw, "Classification and regression by randomforest," *R news*, 2002 (cit. on p. 18).

[92] S. Pellegrini, A. Ess, K. Schindler and L. Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *2009 IEEE 12th International Conference on Computer Vision*, IEEE, 2009, pp. 261–268 (cit. on pp. 19, 39).

[93] A. Lerner, Y. Chrysanthou and D. Lischinski, "Crowds by example," in *Computer graphics forum*, Wiley Online Library, vol. 26, 2007, pp. 655–664 (cit. on pp. 19, 39).

[94] P. Sun, H. Kretzschmar, X. Dotiwalla *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2446–2454 (cit. on pp. 19, 20).

[95] R. O. Murphy, K. A. Ackermann and M. Handgraaf, "Measuring social value orientation," *Judgment and Decision making*, vol. 6, no. 8, pp. 771–781, 2011 (cit. on pp. 22, 23).

[96] J. Arnett, "Sensation seeking: A new conceptualization and a new scale," *Personality and individual differences*, vol. 16, no. 2, pp. 289–296, 1994 (cit. on p. 23).

[97] J. Sprenger, L. Hell, M. Klusch, Y. Kobayashi, S. Kudo and C. Müller, "Cross-cultural behavior analysis of street-crossing pedestrians in japan and germany," in *2023 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2023, pp. 1–7 (cit. on pp. 23, 32, 33).

[98] T. Salzmann, B. Ivanovic, P. Chakravarty and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*, Springer, 2020, pp. 683–700 (cit. on pp. 37, 38, 47).

[99] K. Mangalam, Y. An, H. Girase and J. Malik, "From goals, waypoints & paths to long term human trajectory forecasting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 233–15 242 (cit. on pp. 37, 38, 47).

[100] J. Yue, D. Manocha and H. Wang, "Human trajectory prediction via neural social physics," in *European conference on computer vision*, Springer, 2022, pp. 376–394 (cit. on pp. 37, 38, 47).

[101] Y. Wu, G. Chen, Z. Li *et al.*, "Hsta: A hierarchical spatio-temporal attention model for trajectory prediction," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 11 295–11 307, 2021 (cit. on p. 38).

[102] I. Bae and H.-G. Jeon, "Disentangled multi-relational graph convolutional network for pedestrian trajectory prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 911–919 (cit. on pp. 38, 40, 44).

[103] B Wan and N. Rouphail, "Simulation of pedestrian crossing in roundabout areas using arena," in *Paper Submitted for publication and presentation at the 83rd TRB Annual Meeting in January*, 2004 (cit. on p. 43).