

UNIVERSITY OF TARTU
Institute of Computer Science
Computer Science Curriculum

Ankit Garg

**Machine learning solutions for the task of
pedestrian trajectory prediction – a systematic
literature review**

Master's Thesis (30 ECTS)

Supervisor: Naveed Muhammad
Co-Supervisor: Dmytro Zabolotnii

Tartu 2024

Machine learning solutions for the task of pedestrian trajectory prediction – A Systematic literature review

Abstract: This study aims to provide an in-depth overview of existing methodologies, trends, and challenges in human trajectory prediction. It analyzes diverse literature to examine various approaches involved in machine learning techniques.

This study categorizes these methodologies based on their foundational principles, delving into their strengths and limitations. Particular emphasis is placed on recent advances in machine learning mixed with psychological and environmental aspects for human trajectory prediction.

This study finds three significant categories: cognitive approaches, pattern-based approaches, and probabilistic approaches. These are then further divided into different sub-categories, thus forming a taxonomy. Categories at each level of the hierarchical taxonomy are compared, with information about their pros, cons, and where each category should be used. Furthermore, the research papers studied during this survey were split into categories based on their methods.

In conclusion, it was found that the “Behavioral Features Method” category performed the best among the other categories. Thus, more research should be done on combining machine learning methods with behavioral features.

Keywords:

Machine Learning, Pedestrian Trajectory Prediction, Average Displacement Error, Final Displacement Error

CERCS: T125 Automation, robotics, control engineering

Masinõppe lahendused jalakäijate trajektoori ennustamiseks - süsteemaatiline kirjanduse ülevaade

Lühikokkuvõte:

Selle uuringu eesmärk on anda põhjalik ülevaade olemasolevatest meetodikatest, suundumustest ja väljakutsetest inimese trajektoori ennustamisel. Selles analüüsitakse mitmekesist kirjandust, et uurida erinevaid masinõppetehnikatega seotud lähenemisviise.

Selles uuringus liigitatakse need meetodikad nende aluspõhimõtete alusel, süvenedes nende tugevatesse külgedesse ja piirangutesse. Inimese trajektoori ennustamisel pööratakse erilist rõhku hiljutistele edusammudele masinõppes, mis on segatud psühholoogiliste ja keskkonnaaspektidega.

Selles uuringus leitakse kolm olulist kategooriat: kognitiivsed lähenemisviisid, mustripõhised lähenemisviisid ja tõenäosuslikud lähenemisviisid. Seejärel jagatakse need erinevatesse alamkategooriatesse, moodustades nii taksonoomia. Hierarhilise taksonoomia iga taseme kategooriaid võrreldakse koos teabega nende plusside ja miinuste kohta ning kus iga kategooriat tuleks kasutada. Lisaks jagati selle uuringu käigus uuritud uurimistööd kasutatud meetodite alusel kategooriatesse.

Kokkuvõttes selgus, et kategooria „Käitumistunnuste meetod“ toimis teiste kategooriate seas kõige paremini. Seega tuleks rohkem uurida masinõppemeetodite kombineerimist käitumisomadustega.

Võtmesõnad:

Masinõpe, jalakäija trajektoori ennustamine, keskmise nihke viga, lõpliku nihke viga

CERCS: T125 Automatiseerimine, robotika, juhtimistehnika.

Acknowledgments

I sincerely thank my thesis supervisor, Dr. Naveed Muhammad, and my co-supervisor, Dmytro Zabolotnii, for their guidance and encouragement throughout this study. Their professional mentorship has been invaluable, and I am grateful for our shared, enriching collaboration. Thanks to the University of Tartu and the Government of Estonia for their ongoing support.

I extend heartfelt gratitude to my family for their support during the compilation of this study. I am genuinely grateful to all my friends, who occasionally motivated me and made this research possible.

A special acknowledgment is reserved for my wife, Mahima Singhal, whose support and motivation pulled me through challenging moments and inspired me to pursue my dream of obtaining this degree. I sincerely thank God for His blessings and guidance throughout this journey.

Table of Contents

1.	Introduction	7
1.1	Research Goal	7
1.2	Motivation	7
1.3	Objective	7
1.4	Structure of the Manuscript	8
1.5	Terminology	8
2	Background	10
2.1	Related Work	10
3	Methodology	14
3.1	Searching Research Papers	16
3.2	Maintaining a list of research papers	18
3.3	Understanding Research Papers	18
4	Result Analysis	19
4.1	Categorisation	19
4.1.1	Cognitive Approaches	19
4.1.1.1	Surrounding-based methods	20
4.1.1.1.1	Attention-Based Methods	21
4.1.1.1.1.a	Attention Methods	21
4.1.1.1.1.b	Location-Velocity Attention Methods	23
4.1.1.1.2	Context-aware methods	24
4.1.1.2	Behavioural Methods	26
4.1.1.2.1	Behavioural Features Method	26
4.1.1.2.2	Rational Behaviour Method	28
4.1.1.2.3	Social Interaction Method	28
4.1.1.2.4	Intention Estimation	32
4.1.1.2.5	Goal-Driven Method	34
4.1.2	Pattern-based Approaches	36
4.1.2.1	Spatio-Temporal Graphs	36
4.1.2.2	Sequence-to-Sequence Model	37
4.1.3	Probabilistic Approaches	39
4.1.3.1	Equilibrium-based Interaction Modeling (Nash Equilibrium)	39
4.1.3.2	Latent Belief Energy-Based Model	39
4.1.3.3	Latent Variational Model	40
4.2	Comparative Analysis for the Categories	40

4.2.1 Comparative study for the methods at the highest level of the Hierarchy tree presented in Fig1	41
4.2.1.1 Comparative analysis for cognitive approaches:	42
4.2.1.1.a Comparative analysis for surrounding based methods:	43
4.2.1.1.a.1 Comparative analysis for attention-based methods:	45
4.2.1.1.b Comparative analysis for behavioral methods:	46
4.2.1.1.b.1 Comparative analysis for social interaction methods:	47
4.2.1.2 Comparative analysis for pattern-based approaches	49
4.2.1.3 Comparative analysis for probabilistic methods	50
4.3 Summary Table	51
4.4 Dataset	51
4.5 Performances	52
5 Summary and Emerging Trends	55
5.1 Summary	55
5.2 Emerging Trends	55
5.3 Future Research Area	56
Reference	57
License	65

1. Introduction

In the last few years, there has been significant improvement in predicting human trajectory, mainly because of the need for safe and efficient interactions between humans and robots [1] especially for self-driving cars, handling crowds, and organizing cities' functions [2]. Scientists have developed various models to predict human trajectory. Each model addresses the challenges of understanding the complex and always-changing ways people act while walking around [3].

Different methods are being developed to predict human trajectory. Some methods use machine learning, and some use physics-based methods. This study mainly focuses on machine learning methods. This focused survey study of existing machine learning methods is a new and promising way to improve the models that predict human trajectory.

The objective is to provide a well-structured and insightful analysis of the various methodologies, techniques, and approaches in making predictive models. This study defines each trajectory prediction method and the conditions under which each technique should be used. Then, it examines the integration of contextual information, the consideration of social interactions, and the inclusion of cognitive and psychological factors in predictive models.

1.1 Research Goal

The research goal of this study is to thoroughly review the current literature and conduct a comparative analysis of the methods used in human trajectory prediction. By scrutinizing existing research, the intention is to provide insights into the strengths and limitations of different machine-learning approaches. Thus, identifying methods that have demonstrated effectiveness in real-world applications offers valuable direction for future work in the field.

1.2 Motivation

While human trajectory prediction has been widely explored in research works, a comparative and focused study of the machine learning methods is still lacking [4]. Most survey papers compare traditional machine learning methods with non-machine learning methods, thus leaving behind some essential factors, such as social interaction, intention, and the person's goal, while forming different methodologies/taxonomy. This study explores traditional machine learning methods with environmental, psychological, and contextual factors involved in human trajectory prediction. This study may help researchers improve autonomous vehicles' accuracy in forecasting human behavior while moving on the road.

1.3 Objective

Through carefully consolidating existing research, the study seeks to provide a detailed understanding of human trajectory prediction, explicitly focusing on machine-learning approaches. Additionally, it aims to critically assess and identify a taxonomy of different machine-learning methods and compare the categories mentioned at different taxonomy levels for potential future exploration and development.

1.4 Structure of the Manuscript

This manuscript includes four chapters, each containing different information about the study.

- Chapter 1 introduces the study's research goal, motivation, and objectives.
- Chapter 2 discusses the study's background and relevant prior work.
- Chapter 3 outlines the methodology used for the survey study, covering the formal approach to research, strategies for literature review, and methods for summarising research papers.
- Chapter 4 details the categorization formed after reviewing research papers, exploring the definitions of these categories, highlighting the contributions of different research papers within each category, providing an overview of where each category can be used, and conducting a comparative analysis.
- Chapter 5 describes the summary of the study, about the future works and trends in the area.

NOTE—This study was not written using an LLM or AI model. Grammarly was used to check the grammar.

1.5 Terminology

1. **Labeled Data** - Labeled Data is the data that comes with some label. For example, pictures of different animals are collected with their labels¹, which defines which animal they are. This type of data is used in Supervised machine learning.
2. **Unlabelled Data** - Unlabelled Data is the data that comes without any categorization, label, or grouping. For example, pictures of many animals without labels, features, or categories that can define the animal itself². This type of data is used in unsupervised machine learning, and the machine groups the animals by analyzing the similarities in the pictures of the same animal.
3. **Supervised machine learning** - Supervised machine learning is a machine learning technique using labeled data mapped to some output³. Here, the machine is trained using data tagged with the correct answer. The machine predicts the outcome of the new set of examples based on trained data.
4. **Unsupervised machine learning**- Unsupervised machine learning and unlabelled data predict the class for the newly introduced data⁴. This means algorithms would be left to learn based on the structures/features present in the data and then predict the class of freshly introduced data.
5. **Transformer**—A transformer is a deep learning model that adopts the self-attention mechanism, differentially weighting the significance of each input part (including the recursive output) data. It is used primarily in natural language processing (NLP) and computer vision (CV) [5].

¹<https://www.datacamp.com/blog/what-is-labeled-data>

²<https://livebook.manning.com/book/grokking-machine-learning/2-1-what-is-the-difference-between-labelled-and-unlabelled-data-v-4/>

³<https://kili-technology.com/data-labeling/machine-learning>

⁴ <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>

6. Graphormer - Graphormer is a deep learning package that helps operators train custom models, researchers, and developers for molecular modeling tasks.

7. Average Displacement Error (ADE) - The average l_2 distance between the predicted trajectories and the ground truth across all estimated time steps [6].

$$ADE = \frac{\sum_{i=1}^n \sum_{t=T_{\text{obs}}+1}^{T_{\text{pred}}} [(\hat{x}_i^t - x_i^t)^2 + (\hat{y}_i^t - y_i^t)^2]}{n(T_{\text{pred}} - (T_{\text{obs}} + 1))}$$

8. Final Displacement Error (FDE) - The average l_2 distance between the predicted trajectories and the ground truth for the final estimated time step [6].

$$FDE = \frac{\sum_{i=1}^n \sqrt{(\hat{x}_i^{T_{\text{pred}}} - x_i^{T_{\text{pred}}})^2 + (\hat{y}_i^{T_{\text{pred}}} - y_i^{T_{\text{pred}}})^2}}{n}$$

9. GAN - A generative adversarial network (GAN) is a machine learning (ML) model in which two neural networks compete using deep learning methods to become more accurate in their predictions⁵.

10. Stochastic- Stochastic is a description that refers to outcomes based on random probability. Its etymology traces to a Greek word, "stókhos," meaning "guess." Stochastic systems, stochastic analysis, and stochastic optimization can occur whenever a collection of random variables comes into play. Stochastic processes are widely used as mathematical models of systems and phenomena that appear to vary randomly⁶. Examples include the growth of a bacterial population, an electrical current fluctuating due to thermal noise, or the movement of a gas molecule.

11. Gaussian Mixture models - Gaussian Mixture models represent Normally Distributed subpopulations within an overall population. The advantage of Mixture models is that they do not require which subpopulation a data point belongs to. It allows the Model to learn the subpopulations automatically⁷.

⁵ <https://www.theinterline.com/2023/08/23/will-ai-bring-positive-change-to-plm/>

⁶ <https://www.rwmansiononpeachtree.com/what-type-of-process-is-described-by-the-ornstein-uhlenbeck-model/>

⁷ <https://scikit-learn.org/stable/modules/mixture.html>

2 Background

This chapter offers a concise overview of previous surveys in human trajectory prediction, emphasizing significant milestones and key breakthroughs. It focuses on what each survey discussed, what contributions it made, and what taxonomy it could derive.

2.1 Related Work

Many papers, methods, and models available on pedestrian trajectory prediction. However, there are significantly fewer surveys for “Human Trajectory Prediction” [4], while many existing surveys talk about trajectory prediction for autonomous vehicles or the interaction between service robots and humans. To the author’s knowledge, no survey studies on human trajectory prediction focused only on machine learning methods. Below, we present some survey papers that do not directly speak about the subject of interest but are still related to such a study.

The authors in [7] conducted a comprehensive review, analysis, and organization of diverse work from various communities for all the road agents, whether autonomous vehicles, pedestrians, or robots. A taxonomy is proposed to categorize existing methods, considering the motion modeling approach and the level of contextual information employed. An overview of current datasets and performance metrics is presented. Limitations of state of the art are discussed, and directions for further research are outlined. The categorization also covers all the spheres of prediction, including machine and non-machine learning. Figure 1 represents the overview of the categories authors presented in [7].

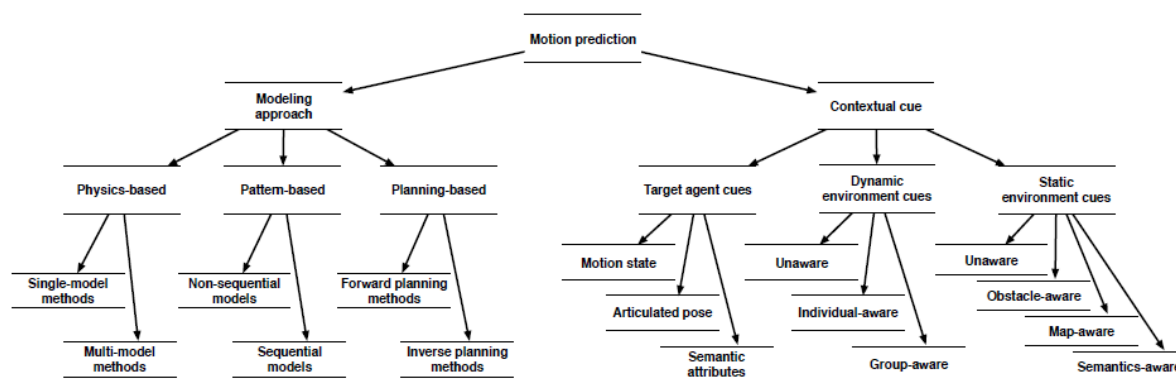


Figure 1. Overview of the categories presented by the authors in [7].

In [4], the authors focus on predicting the trajectories of road users, specifically pedestrians and vehicles. The objective is to organize recent literature into a novel taxonomy encompassing motion prediction techniques for pedestrians and vehicles. The survey reviews various motion prediction methods, evaluates their advantages and disadvantages, and assigns each method to a distinct category within the proposed taxonomy. The authors introduced a taxonomy based on three approaches: a modeling approach, output type, and situational awareness, as shown in Figure 2.

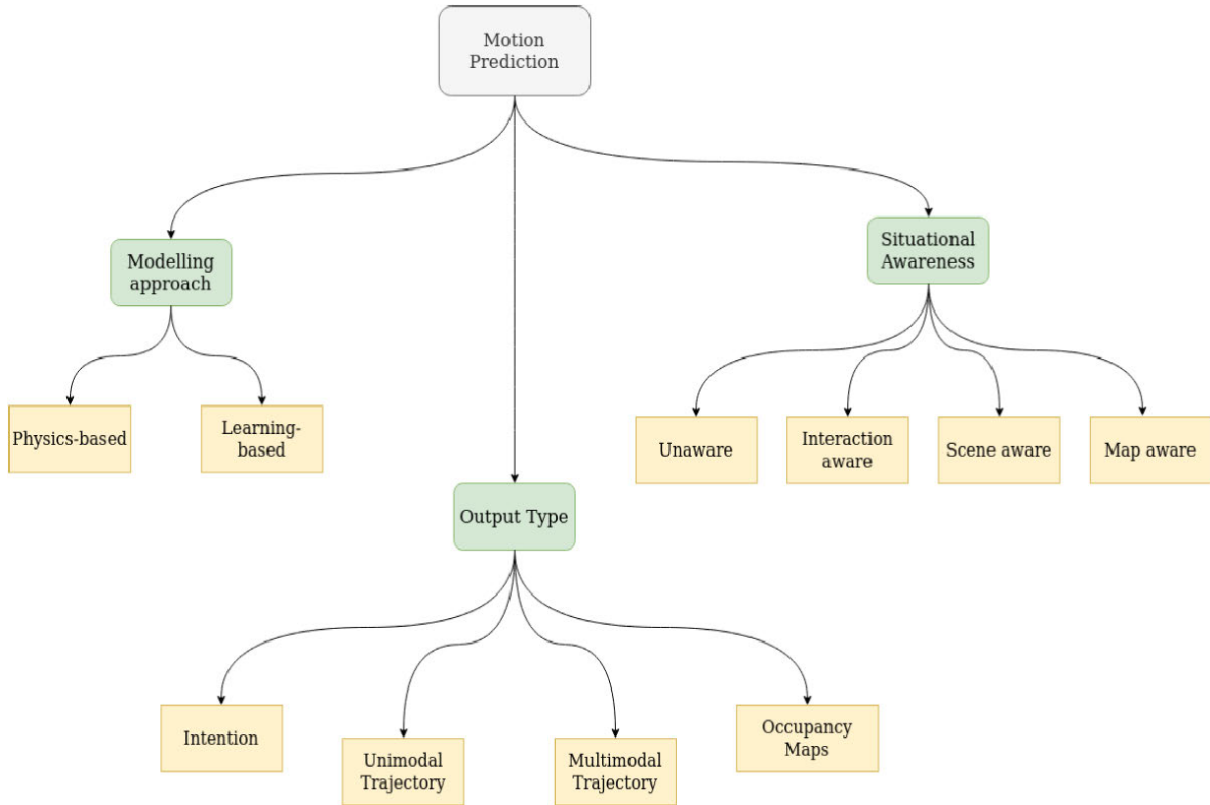


Figure 2, Proposed taxonomy of motion prediction for pedestrians and vehicles, presented in [4].

The study in [8] thoroughly examines common ways autonomous vehicles predict movements, including physics-based methods and various types of machine learning, such as classic, deep, and reinforcement learning. It also explains the measures and datasets used to assess how well these methods work. Lastly, it discusses each approach's strengths and weaknesses and suggests areas for future research. Figure 3 shows the taxonomy based on reviews of physics-based methods, classic machine learning-based methods, deep learning-based methods, and reinforcement learning-based methods.

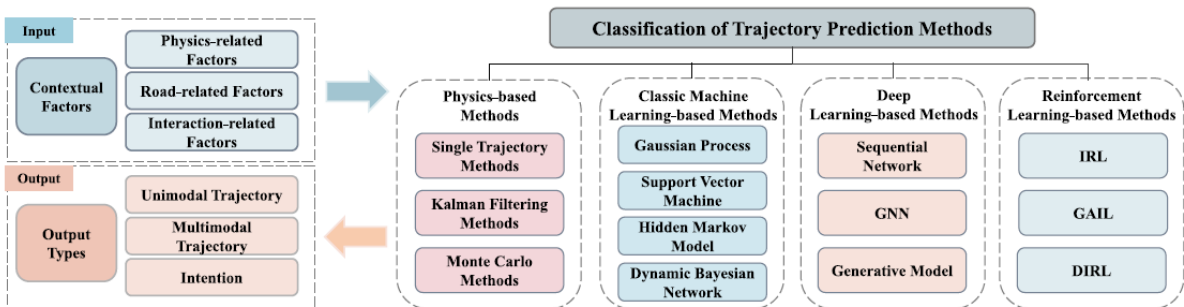


Figure 3, The taxonomy of trajectory-prediction models for AVs, presented in [8].

In [9], the authors discuss the kinds of sensors usually used with deep learning techniques to predict human trajectory. They look closely at how deep learning predicts pedestrians' trajectories. The authors also give an overview of the enormous datasets in this area, including

sensor types, size, number of pedestrians, and location. Lastly, they highlight what's missing in current research and suggest new places to explore.

In [10], the authors look at how Deep Learning can better predict trajectories for self-driving cars, especially for pedestrian safety. As far as they claim, no other studies have looked at recent deep-learning methods for this purpose as profoundly as they have. They examined how cars and pedestrians interact, determining the most critical information for successful prediction methods. They also talk about different ways predictions can be shown. Furthermore, it gives a detailed look at how well these methods work on different sets of data, analyzes the results carefully, and suggests new areas for research that look promising.

In [11], special attention was given to the main models used in composite-model approaches for 3d human motion prediction. Since some methods mix with smaller parts, authors concentrated on the big models and new ideas for organizing them. They stressed the importance of looking at different viewpoints. How they see what needs predicting and how they represent the human body affects how they solve problems. The authors divided the methods into three groups: how humans are defined, how the networks are designed, and what they try to predict. They have carefully reviewed all the essential papers from journals and conferences about predicting human motion since 2015. Then, they explained them according to the categories created in the survey. Figure 4 represents the categorization the authors came up with in [11] after reviewing the related papers published from 2015 till mid of 2021.

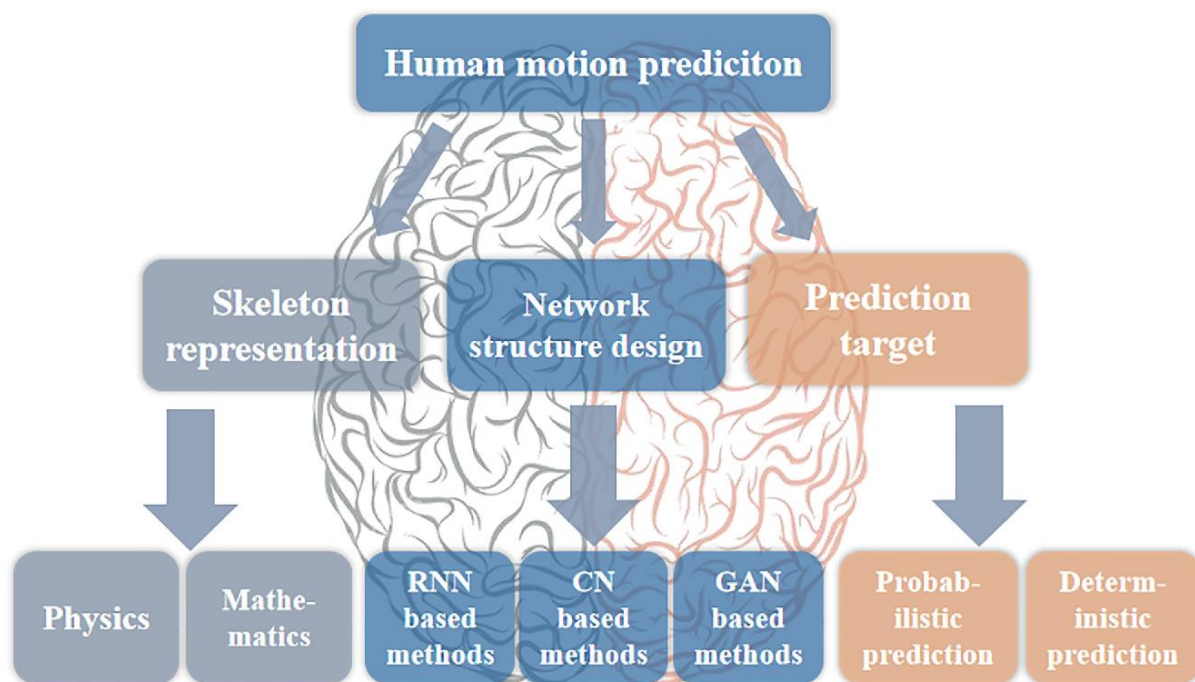


Figure 4, Overview of the categories, presented in [11].

In this article [12], the authors discuss a clear review of KB (Knowledge-Based) and DL (Deep Learning) approaches for modeling and predicting human trajectory in a crowded space. They carefully compare these two approaches regarding their technical details and where they're used. The authors pointed out what's similar and different between them and suggested ideas for how they can develop in the future. The comparison indicates that combining both approaches (the hybrid approach) looks hopeful in overcoming drawbacks, such as the lack of explanation in deep learning methods.

As discussed above, [7] covers both the spheres of prediction, machine learning, and non-learning, whereas in [4] and [8], physics-based methods and various types of machine learning are considered. [9], [10] and [12] focus on methods using deep learning, which also gave importance to contextual and physical methods or based on expert's experiences (KB methods) and thus compared DL with KB. [11] is a bit different in the sense that it considers the 3D motion of humans, which further tells about intentions, but they also rely on deep learning for better image processing.

In contrast to the survey studies covered above, the analysis presented in this study focuses on humans as road agents. All the methods are categorized from the study of papers that used machine learning methods.

3 Methodology

Doing a literature review demands time and discipline, requiring following specific principles throughout the process. In the initial stages of preparation, a comprehensive survey of previous systematic literature reviews across various research domains was undertaken. Detailed reviews distinguished themselves among them for their superior structure and thoughtful design. One widely endorsed best practice is using the PRISMA model, which has guided the execution of systematic literature reviews across numerous research fields⁸. Additionally, valuable insights can be gleaned from the guidance provided by the Cochrane Collaboration to prepare systematic literature reviews [13]. Given that the PRISMA model was jointly developed by the Cochrane Collaboration and a diverse array of field experts, it was deemed the most fitting set of principles to employ in the current attempt to conduct a systematic literature review. The PRISMA methodology, formulated by Moher et al. [14], has gained widespread recognition for enabling well-structured systematic literature reviews in many studies [7],[15],[16],[17].

Table 1 provides a detailed checklist for preparing systematic reviews based on the PRISMA model. This table outlines the steps that should be followed during the survey study. Although it is optional to follow all the steps to conduct good research, following this provides an excellent outline of the methodology.

Table 1: PRISMA checklist for preparing systematic reviews per the guidance presented in the paper⁹.

Sub-topic number	Topic and subtopic in PRISMA checklist	Description
Title		
1	Title	Publication is identified as systematic review, meta-analysis, or both.
Abstract		
2	Structured summary	Provide a structured summary including, as applicable, background, objectives, data sources, eligibility criteria, conclusions, and results.
Introduction		
3	Rationale	Describe the rationale for the publication.
4	Objectives	Provide the questions being addressed in the study.
Methods		
5	Protocol and registration	Indicate if a review protocol exists and if it is available. Provide registration information and registration number if they are available.
6	Eligibility criteria	Specify characteristics used as eligibility criteria.

⁸ <https://www.prisma-statement.org/prisma-2020>

⁹ <https://www.prisma-statement.org/prisma-2020-checklist>

7	Information sources	Describe the information sources where the publications were obtained, and the date searched.
8	Search	Present an electronic search strategy for at least one database.
9	Study selection	State the process of study selection, including screening and eligibility.
10	Data collection process	Describe the method of data extraction from the publications.
11	Data items	List and define all variables for which data were sought.
12	Risk of bias in individual studies	Describe assessing the risk of bias in individual studies.
13	Summary measures	State the summary measures.
14	Synthesis of the results	Describe the strategy of handling data and combining the results of studies.
15	Risk of bias across the studies	Specify any assessment of risk of bias that may affect the cumulative evidence.
16	Additional analyses	Describe methods of other analyses (if done).
Results		
17	Study selection	Provide several studies screened, assessed for eligibility, and included in the review.
18	Study characteristics	Present the characteristics for which data were extracted.
19	Risk of bias within studies	Data on the risk of bias of each study, if available (see also item 12).
20	Results of individual studies	For each study, present a simple summary of data for each intervention group and effect estimates (if applicable).
21	Synthesis of results	Present results of each meta-analysis done.
22	Risk of bias across studies	Present results of any assessment of risk of bias across studies (see also item 15).
23	Additional analysis	Give results of further analyses, if done.
Discussion		
24	Summary of evidence	Summarize the main findings.
25	Limitations	Discuss limitations at the study and outcome level (e.g., risk of bias).
26	Conclusions	Provide a general interpretation of the results and implications for future research.
Funding		
27	Funding	Describe sources of funding for the systematic review.

3.1 Searching Research Papers

Conducting a comprehensive survey of research papers demands a thoughtful and systematic approach. Initially, it is essential to compile a list of sources for searching, considering accessibility, especially in paid journals. Developing an effective method for filtering out relevant papers is vital to this strategy. These filters can vary and may include criteria such as the publication year, specific keywords in the title, or combinations of keywords. Additionally, researchers can use the methodology as a filter, allowing for a more focused search based on preferences like supervised machine learning, unsupervised machine learning, or non-machine learning methods. This strategic approach ensures that the survey is targeted, efficient, and aligned with the specific criteria and objectives of the researcher, promoting a more meaningful and insightful exploration of the available literature.

The search process was conducted on SpringerLink, Science Direct, and IEEE, focusing on human motion trajectory prediction. Here is a detailed account of the search methodology and results:

Google Scholars:

With the exact phrase – „human motion trajectory prediction."
Where my word occurs – „anywhere in the article."
Dated between - 1990-2023
Results - 613

SpringerLink:

Search 1:

The search was performed on SpringerLink with a year filter spanning 1990-2023.
Query parameters included "human motion trajectory prediction" with specific criteria:
With all of the words: "human motion trajectory prediction."
With the exact phrase: "human motion."
Where the title contains: "human motion"
The search yielded 108 papers, and the initial screening was based on the titles.

Search 2:

Another search was performed with the following criteria:
With all of the words: "trajectory prediction."
With the exact phrase: "human motion."
Where the title contains: "human motion"
This search resulted in 9 irrelevant papers after the initial screening.

Search 3:

The search query was modified to:
With all of the words: "human motion trajectory prediction."
With the exact phrase: "trajectory prediction."
This search produced 1063 articles, all deemed irrelevant after the first screening.

Search 4:

A separate search focused on pedestrian trajectory prediction:
With all of the words: "pedestrian trajectory prediction."
Where the title contains: "pedestrian"
The search yielded 264 research papers.

Science Direct:

Search 1:

Science Direct (2000-23)
Find articles with these terms: "human motion trajectory prediction."
Title, abstract, or author-specified keywords: "human motion trajectory prediction"
The search yielded 74 research papers.

Search 2:

Title, abstract, or author-specified keywords: human motion trajectory prediction
Title: [human motion]
The search yielded 13 research papers.

Search 3:

Find articles with these terms: "trajectory prediction."
Title, abstract, or author-specified keywords: human motion trajectory
Title: [human motion]
The search yielded 34 research papers.

IEEE:

A search on IEE utilized the "pedestrian trajectory prediction" query in "All Metadata."
The search yielded 472 articles.

These searches were carefully conducted across multiple platforms, employing specific search parameters to identify and collect research papers on human motion and pedestrian trajectory prediction. The screening process ensured the inclusion of relevant articles meeting the specified criteria.

A document is maintained to keep the records for the search, where several searches for each calendar day are held. So, in case it's required to check from which search this paper was found. Filtered paper names were individually searched, and further screening was conducted based on abstracts. Articles were excluded if the abstract and conclusion presented objectives unrelated to the study. Article details were recorded in an Excel sheet, and the relevant papers were downloaded.

3.2 Maintaining a list of research papers

Following the initial shortlisting based on the titles of papers related to human trajectory prediction using machine learning, each paper's name, link, year of publication, download status, date of download, and search number are recorded in an Excel sheet. Table 2 shows a small sample of the sheet for the research paper management.

Table 2. Sample sheet for research paper management.

Title	Link	Publication Year	Download status	Date downloaded	Search Number	Date Read
Egocentric Two-Frame Pedestrian Trajectory Prediction Algorithm Based on a Panoramic Camera	https://ieeexplore-ieee.org.ezproxy.utlib.ut.ee/stamp/stamp.jsp?tp=&ar_number=9964286	2022	yes	26022023	2	3092023
OD-network-based Pedestrian-path Prediction for People-flow Simulation	https://ieeexplore-ieee.org.ezproxy.utlib.ut.ee/stamp/stamp.jsp?tp=&ar_number=9006314	2019	yes	26022023	2	
Pedestrian Trajectory Prediction Combining Probabilistic Reasoning and Sequence Learning	https://ieeexplore-ieee.org.ezproxy.utlib.ut.ee/stamp/stamp.jsp?tp=&ar_number=8957246	2020	yes	26022023	2	3092023
Learning Trajectory-Conditioned Relations to Predict Pedestrian Crossing Behavior	https://ieeexplore-ieee.org.ezproxy.utlib.ut.ee/stamp/stamp.jsp?tp=&ar_number=9897655	2022	yes	26022023	2	4092023
Pedestrian Motion Model Using Non-Parametric Trajectory Clustering and Discrete Transition Points	https://ieeexplore-ieee.org.ezproxy.utlib.ut.ee/stamp/stamp.jsp?tp=&ar_number=8638524	2019	yes	26022023	2	

The provided list has a 'Date Read' column, allowing us to monitor when each paper was read and summarized. The color-coded system indicates the relevance of each paper, with sky blue representing a relevant paper and green marking irrelevant papers. This determination was made by applying a filter based on the abstract content. Papers were considered appropriate if the abstract focused on human trajectory prediction.

3.3 Understanding Research Papers

There is no standardized approach for reading and analyzing research papers, but the method that was found most comfortable involves the following steps:

- Read the abstract and conclusion before delving into the entire paper. This provides an initial understanding of the author's intentions and achievements.
- Read the introduction and identify the paper's main contributions. For studies like this, conceptual comprehension precedes an understanding of the practical implementation.
- If unfamiliar terms and methods exist, take the time to understand and document them separately.
- After thoroughly understanding the paper's content, create a personal summary for future reference, which will help you write the survey study.

Diligently following these steps for each research paper generated a document containing essential terminologies and summaries. This process facilitated a deeper understanding of the individual papers and led to the development of a structured categorization throughout the study.

4 Result Analysis

After carefully examining the shortlisted research papers, it was observed that numerous methods were recurrent across multiple papers, with significant overlap and similarities among them. Some papers presented methods that have similar features and can be consolidated into broader categories. Consequently, a systematic categorization of the methods was undertaken. This categorization aimed to streamline the understanding of the various methodologies by identifying commonalities, eliminating redundancies, and grouping similar approaches under broader categories. This study makes the research landscape easier to understand by organizing it better. It helps people see the different methods used in papers more clearly and shows the main themes and trends in the research. Furthermore, after the definition of each method, contributions from all the papers using the respective papers have been noted.

4.1 Categorisation

The categorization of approaches in human trajectory prediction serves as a valuable framework, allowing for the systematic organization and classification of the diverse methodologies employed in this field. This categorization not only aids in studying these approaches individually but also facilitates meaningful comparisons among them. The logic behind categorizing these approaches is driven by the conceptual understanding of the fundamental principles, methods, and considerations inherent in modeling and predicting human movement. Figure 5 depicts an organized hierarchy visually representing the relationships and dependencies among various trajectory prediction approaches. As a result, this categorization serves as a valuable resource for anyone delving into human trajectory prediction, providing a coherent framework for exploration and Analysis.

A hierarchical tree map has been formed to represent these categories better. This tree tries to cover most of the methods available widely in the industry. At the first level of division, methods have been divided into three categories:

- Cognitive Approaches
- Pattern-based Approaches
- Probabilistic Methods

Many research papers fall under each category, which might be repeated in different categories. Contributions from each paper under each category have been mentioned.

4.1.1 Cognitive Approaches

Predicting how people will walk using cognitive approaches takes ideas from how people think and act. This method indicates human trajectory by understanding and including the critical role of thinking in people's movements [18]. In simpler terms, thinking processes like making decisions, estimating intentions, and following goals play a big part in figuring out how people walk. Sorting cognitive approaches into categories is an organized way of understanding and using different methods in this always-evolving field. By understanding how people think and make decisions, researchers hope to figure out the complexities of human trajectory prediction, ultimately making predictions about their paths more accurate.

A noteworthy classification within cognitive approaches branches into two distinct categories:

- Surrounding-Based Methods
- Behavioral Methods

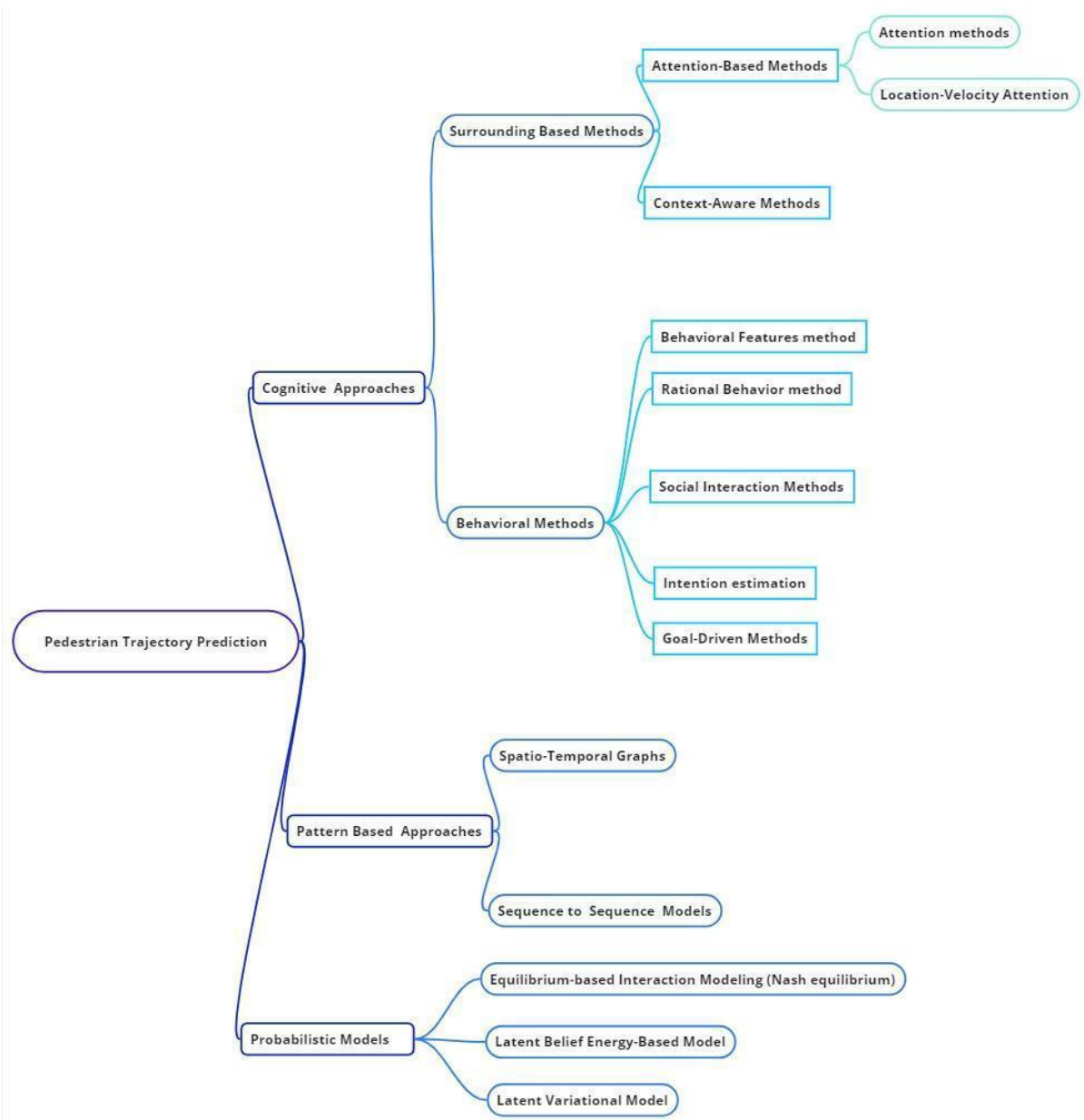


Figure 5. Overview of the categorization of methods available for Human trajectory prediction.

4.1.1.1 Surrounding-based methods

Surrounding-based methods for predicting pedestrian trajectories concentrate on the environment and external elements that impact human movement [4]. This classification is based on how vital (i) the environment and (ii) what's around people is in shaping how they move, and it tries to use this knowledge to make better predictions. The Surrounding-Based category can be divided into two main subcategories:

- Attention-Based Methods
- Context-Aware Methods.

4.1.1.1.1 Attention-Based Methods

Attention-based methods are made to understand the importance of specific places or speeds in predicting where someone will go next [19]. Models that use attention better predict movements by focusing on essential things in the surroundings. Following are the methods in which Attention-Based methods can be divided further:

- Attention Methods
- Location-Velocity Attention Methods

4.1.1.1.1.a Attention Methods

Attention methods involve identifying and examining a trajectory's vital locations or specific time intervals [20]. This method allocates differing levels of attention to distinct trajectory segments, offering a detailed perspective on human movement patterns [21]. The deliberate and targeted focus on different parts of the trajectory enables attention methods to effectively capture the complexities and dynamics of how individuals navigate through space and time. In essence, attention methods provide an advanced approach for dissecting and comprehending the intricate dynamics of human trajectories, delivering valuable insights for predictive modeling and analysis. Below are the contributions from the research papers that use the attention method in their model for human-trajectory prediction. Furthermore, this study compares them based on each paper's core concepts. As presented in Figure 6, attention methods adapt the interaction term from focusing on robot-human interactions to human-human interactions. In the local map, the concentration is on predicting the future trajectory of a specific individual. For this example, three attention scores corresponding to three nearby individuals are calculated [22]. Following are the contributions from the research papers falling under this category:

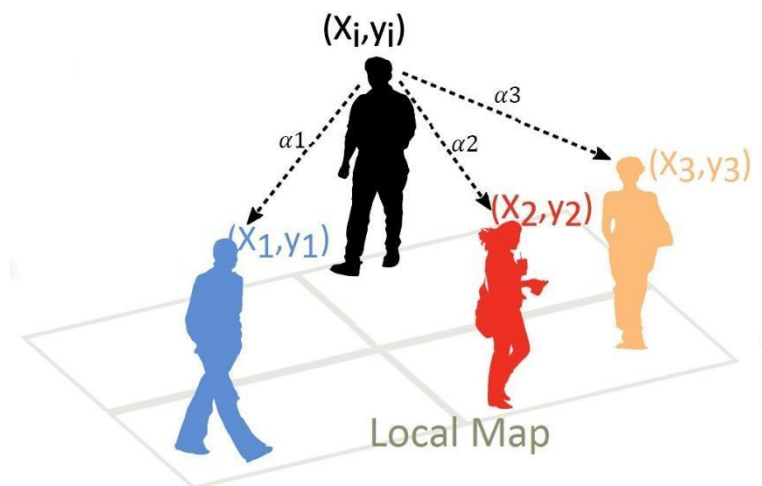


Figure 6: This image represents attention methods. Different attention values are calculated for each nearby person while predicting the trajectory for the next move. This image is taken from [22].

In [23], authors presented a method to predict the dynamic spatiotemporal behavior of people moving using multiple navigational modes in crowded scenes. The presented model involves studying various aspects such as activities, spatial elements, and the positions of stationary

objects within a pool via a pooling mechanism. This involves aggregating information from the surrounding elements better to understand their impact on the main subject of interest.

In contrast, the authors of [24] presented a straightforward yet effective model for predicting pedestrian trajectories. This model focuses on predicting pedestrian positions in urban-like environments, considering surroundings such as the map and nearby agents. It incorporates cross-attention and transformer blocks, commonly used in deep learning models. The objective is to enhance understanding of relationships within the complex data embeddings derived from historical events. Specifically, the method aims to better capture and represent the connections over time, contributing to more accurate predictions of human trajectories based on past patterns and events.

In [25], the model predicts human paths by considering observed trajectories and the evolving scene context. To achieve this, it incorporates a conditional 3D visual attention mechanism, which operates on input videos. This mechanism is designed to selectively emphasize certain aspects of the visual information, enabling the model to infer both environmental constraints and social interactions. In essence, the Introvert model leverages visual attention to understand and predict how pedestrians navigate based on the observed dynamics of their trajectories and the surrounding scene.

The approach outlined in [26] method uses past movement data and predicts future paths to understand and model how people interact with each other. It focuses on three social behaviors from observed and expected movements: (1) how close people are to each other, (2) the direction they're moving compared to each other, and (3) how they move compared to their distance. The model they provided uses these behaviors to understand how people interacted with their surroundings in the past and what might happen in the future. Then, it used an attention mechanism to emphasize the essential features.

In the referenced research [27], the authors presented a new model for predicting paths that use a detailed map to understand complex road elements and interactions. It considers moving obstacles like pedestrians, vehicles, and cyclists and fixed road features like lanes, traffic lights, and sidewalks. It utilizes MapNet and AgentNet, which are techniques for extracting spatiotemporal features using deep convolutional networks and Long Short-Term Memory (LSTM). This enables the automatic derivation of relevant information from the data. Following this, the study introduces Interaction-AttNet to aggregate features. This process involves using affine transformation and a multi-head attention mechanism to understand and capture the high-level interactions among all components within the dataset.

The authors in [28] believed predicting human trajectory requires knowledge about past motion, the environment, and likely destination areas. Thus, it is firmly analyzed and stated that multimodality is a fundamental aspect of architecture. Therefore, it puts forward a simplified recurrent backbone that specifically considers past observed positions as a basis for predicting future trajectories. Notably, the model's accuracy is shown to be enhanced when it is combined with a module that is adept at estimating goals by analyzing the scene context. This integration of attention mechanisms and goal estimation contributes to the overall effectiveness of the trajectory prediction model.

The approach in [29] focuses on developing a spatio-temporal graph method for human trajectory prediction. This method is designed to explicitly capture interactions occurring at both global and local scales within the scene. A novel spatio-temporal attention mechanism enhances the model's capabilities, specifically tailored for individual pedestrian trajectories.

The model described in [30] involves a collaborative effort between location and velocity attention mechanisms within the Long Short-Term Memory (LSTM) network. Additionally, a

specialized module is introduced to refine and optimize the functioning of the LSTM network. The attention mechanism plays a crucial role in combining location and velocity information, contributing to the model's effectiveness in predicting trajectories by leveraging the optimal integration of these critical components.

The model described in [31] can predict the paths of people and vehicles in various situations. It uses a unique method to understand the relationships between different agents and predict how they'll move. The authors also introduce a spatiotemporal dual-attention mechanism to focus on important information from other agents as time goes on. Additionally, the authors include a module that helps ensure the predictions for vehicle paths make physical sense. This module makes the predictions smoother and more accurate and allows the model to handle noisy data better.

In [32], authors use a new method to measure how people interact by looking at their positions, past movements, and directions. They study how individuals and groups interact in crowds, dividing interactions into different types. They introduce a graph attention network (GAT) to understand social interactions better. This network can easily handle group information. Additionally, they created a hierarchical graph attention network to focus on and study human interaction at different levels. They also use a vanilla GAN to generate potential paths directly and refine them using interaction details.

In [33], the authors introduced the social pyramid-based prediction method, which combines information from how people move and interact with others nearby. They tested their method on different datasets and compared it to other methods to see how well it works. From location and the velocity coordinates, their Motion encoders find further information. They also use a special attention method to combine these details for each person's path. The authors introduce a new structure called a social pyramid to understand how people relate to each other at different levels, starting from the whole scene and then focusing on individual people.

These various attention-based methods employ distinctive strategies, including cross-attention, transformer blocks, location-velocity attention, 3D visual attention, and spatiotemporal attention mechanisms. Each method advances human trajectory prediction by emphasizing different aspects of attention mechanisms, from modeling temporal dependencies to capturing complex spatial and social interactions.

4.1.1.1.1.b Location-Velocity Attention Methods

The Location-Velocity Attention method is an advanced approach for predicting human trajectories, leveraging spatial and velocity information [30]. Acknowledging the importance of integrating an individual's spatial coordinates with the velocity of movement [34]. Focusing on specific locations and their associated velocities within a trajectory aims to capture detailed patterns and dynamics in human motion. This dual focus on both spatial and temporal aspects enhances the precision of predicting future movements. Figure 7 represents how this method chooses the distance between pedestrians, the speed of surrounding pedestrians, and the velocity angle as motion features to construct a query [35].

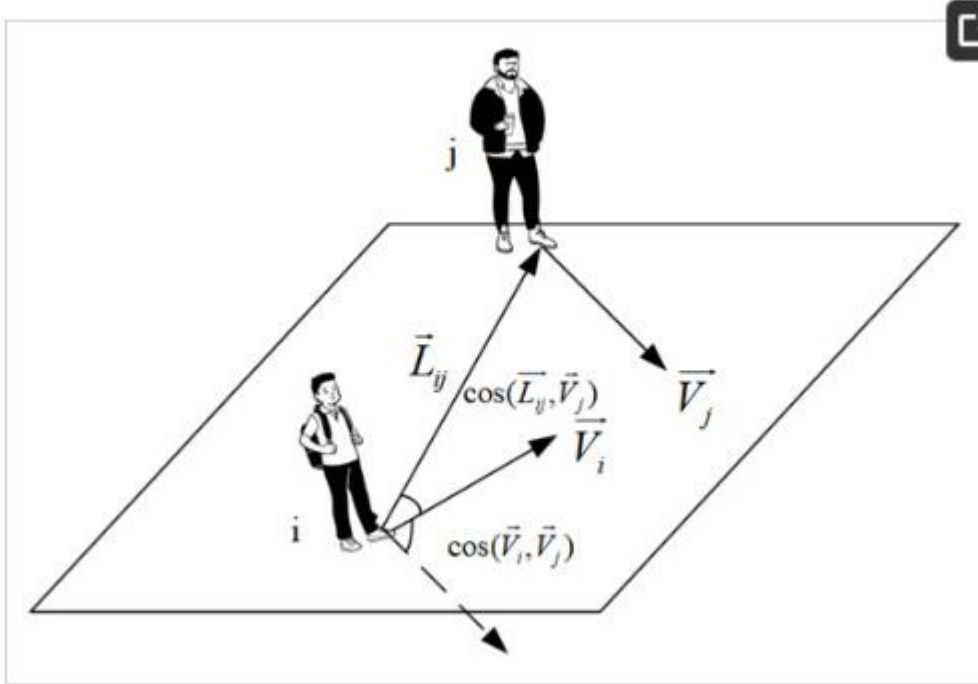


Figure 7 illustrates relative distance, speed, and velocity angle as motion features. In the LVTA method for trajectory prediction, people consider the location and velocity of other people nearby. This image is taken from [35].

Following are the contributions from the research papers falling under this category:

The model, detailed in [25], introduces a sophisticated architecture, a variant of recurrent neural networks designed to capture critical aspects of trajectories. The model enhances its predictive capabilities by incorporating attention mechanisms focusing on spatial (location and velocity) and temporal dimensions. Including an attention layer within a tweak module signifies a specialized component aimed at refining the accuracy of predicted location and velocity coordinates, contributing to more precise trajectory predictions.

On the other hand, [30] the method utilizes a combination of Location-Velocity Attention (LSTM), where LSTM stands for Long-Short-Term Memory, a type of recurrent neural network. The approach includes a specific module that adjusts or fine-tunes the LSTM network, ensuring it performs optimally for the task at hand. An attention mechanism is also integrated and trained to effectively merge information about pedestrians' location and velocity. This integration of location and velocity information is crucial for enhancing the accuracy of predictions regarding how pedestrians move over time.

When comparing the methods in [25] and [30], they both center around utilizing Location-Velocity Attention LSTM models for trajectory prediction. However, they exhibit differences in their implementation details. One uses recurrent neural networks, whereas another uses an LSTM.

4.1.1.1.2 Context-aware methods

Context-aware methods in human trajectory prediction present a specialized approach that considers the surrounding context and external elements influencing human movement [36].

These methods acknowledge the environment's vital role in shaping pedestrian trajectories and aim to utilize contextual information for more accurate predictions. Figure 8 represents the overview of context-aware trajectory prediction. The input of the LSTM is defined by the trajectory of the person under analysis (green), the grid of human-human interactions for the human-human pooling (red), and the context-aware pooling (blue) [37].

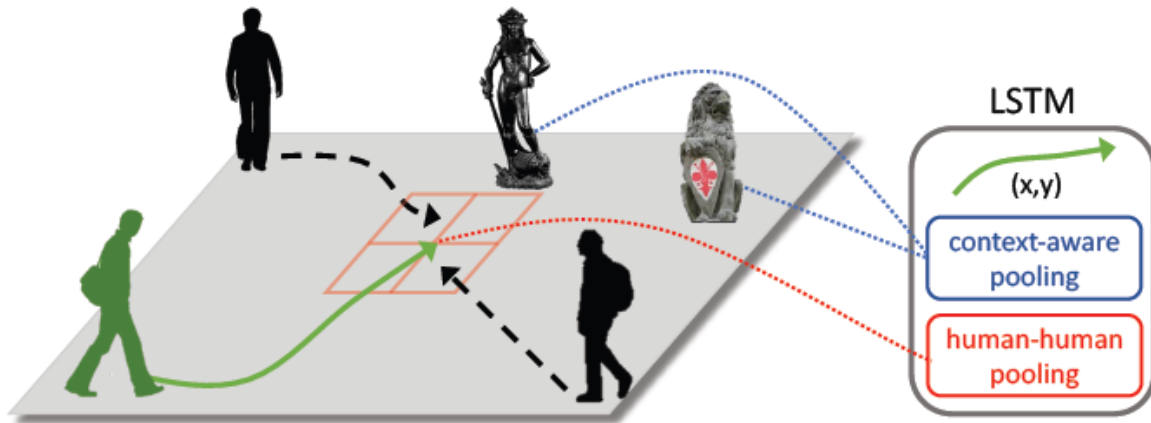


Figure 8: The image represents the trajectory prediction via context-aware methods. Here, the trajectory presented in green shows that it considers the nearby objects and another human. The image is taken from [37].

Following are the contributions from the research papers falling under this category:

In [32], authors introduce a "context-aware" pooling mechanism capable of learning and encoding human-human and human-space interactions. Their model consists of a recurrent neural network LSTM model that can learn and predict human movement in crowded areas like sidewalks, museums, or shopping malls. The authors tested their model using a public pedestrian dataset. Also, they introduced a new challenging dataset featuring videos of people navigating a busy real-world environment, such as a large museum.

In [38], the authors proposed the Introvert model, which predicts human paths based on observed trajectories and dynamic scene context. It infers environmental constraints and social interactions without direct human communication but through observing the dynamic scene. The model uses the observed path of the target person to guide its 3D attention model, helping it focus on important spatial and temporal details.

In [39], the model presented is designed to provide a probabilistic framework for forecasting human trajectories. It incorporates a well-defined cost function within a hidden or latent space. This cost function considers not only the historical movement patterns of individuals but also the social context, contributing to a more comprehensive and context-aware prediction of diverse human trajectories.

The authors in [40], discuss a study incorporating information about velocity and context to address the challenge of predicting human trajectories. They did not consider single-shot (one-frame observation) scenarios where velocity data is missing. As done in previous methods, reducing the observation period clearly lowers their performance. Even without temporal features, three critical pieces of information between two frames help predict trajectories: i) agent velocity, ii) social context, and iii) scene context. Velocity shows short-term movement, while context clues long-term trends and disturbances.

The algorithm described in [41], employs various components, including a panoramic camera, a free pedestrian prediction network (FPPN), a future social pooling module (FSPM), and a social pedestrian prediction network (SPPN). These elements work together to provide comprehensive predictions of human trajectories. The Social Pedestrian Prediction Network (SPPN) combines frame data with FSPM output for final predictions. A panoramic camera reduces blind spots, using only two frames for better real-time performance. They also recorded panoramic videos to create a dataset and tested FSP against two other algorithms on this dataset.

Research in [42], focuses on creating a structured approach that considers the connections between observed pedestrian movements and scene frames. By encoding joint information and global relations over time, the framework aims to improve the accuracy of predicting human trajectories.

In [43], the model emphasizes the influence of contextual information on pedestrian movements and employs a latent variational model to address uncertainty about future trajectories. The model also considers relative interpersonal influences among individuals in the scene and their interactions with the overall scene layout to improve trajectory predictions.

In [31], the model emphasizes its purpose in predicting trajectories for multiple agents. The model incorporates relational inductive biases to improve predictions and employs a dynamic graph representation that integrates trajectory and scene context information.

Researchers in [44], introduced a Dynamic Bayesian Network designed to predict pedestrian paths in the context of intelligent vehicles. The model incorporates latent states, including pedestrian situational awareness, criticality, and spatial layout. These states are considered on top of a Switching Linear Dynamical System (SLDS), enhancing the model's prediction ability.

In [45], the authors include other people nearby, unlike previous methods that focused only on the person walking. They found that considering these nearby people gives better results. To do this, they used simple but useful features and a learning process to improve them. They also used an Auto Encoder model to understand better how the person interacts with others nearby. These enhanced features helped them predict where the person would walk next using a Gaussian Mixture Model.

4.1.1.2 Behavioural Methods

Behavioural-based methods mean using what we know about how people behave to estimate better where they will go [46]. These methods understand that different behaviors affect how people move, and if we model or copy these behaviors, we can make models that are better at predicting paths [47]. There are different ways or approaches to this, and they each look at other parts of how people behave to make better predictions. We can split these behaviors into the following groups:

- Behavioural Features Method
- Rational Behaviour Method
- Social Interaction Method
- Intention Estimation
- Goal Driven Method

4.1.1.2.1 Behavioural Features Method

The main idea behind the Behavioural Features Method is to use information we can see from people's actions to guess where they will go next. By finding and measuring specific behaviors,

this method wants to discover patterns that help predict paths more accurately. An example of a behavioral feature is represented in Figure 9, which is taken from [4]. This shows that people often have consistent behavior patterns influencing their move [48]. By adding these behaviors to the prediction model, the technique makes predictions about paths much better overall.



Figure 9 is an example of gestures used by pedestrians; by analyzing such gestures/behavioral features, the person or vehicle in front can better understand the trajectory of the person showing gestures. Thus, this image represents the behavioral methods for human trajectory prediction. This image is taken from [4].

Following are the contributions from the research papers falling under this category:

The method in [49] is designed to understand and predict human movement in indoor environments. It adapts dynamically, utilizing a traditional linear motion model when behavioral data is scarce and combining it with a conventional linear motion model when sufficient data is present. They proved their methodology via mathematical experiments and presented their results.

The model presented in [50] discusses an alternative perspective. It introduces a pedestrian trajectory prediction method designed explicitly for onboard camera systems. The proposed method, Behaviour-Aware Pedestrian Trajectory Prediction (BA-PTP), relies on behavioral features like body orientation, head orientation, and pedestrian pose to enhance prediction accuracy.

In [40], the model emphasizes the importance of considering velocity and context information, as discussed. It suggests that understanding motion patterns is crucial for predicting short-term behavior, while context information provides valuable insights for anticipating long-term trends and potential disruptions in movement.

In [51], authors look at how pedestrians behave around vehicles using the ego-vehicle vision systems. Instead of predicting if a pedestrian will cross the road, they focused on understanding behaviors like eye contact or hand gestures, as presented in Figure 9. This helps self-driving cars understand pedestrian actions like a human driver would. Their main contributions are: (1)

They believed understanding detailed pedestrian behavior is more important than predicting if they'll cross the road. They created a new system to classify pedestrian behaviors and a large video dataset. (2) They used a new method to describe pedestrian movements using skeleton features, which is reliable even if some key points are missing, regardless of how long the video is.

In [52], authors used feature selection algorithms to determine the most important features for predicting and classifying postures in real time. They combined these features with robust classifiers to create a system that could predict postures with 94% accuracy, as they claimed. The authors used a Support Vector Machine (SVM) to classify different postures with an accuracy of 89.3%. The authors also improved trajectory tracking by filtering out noise from acceleration and angular velocity data and calibrating the gyroscope to reduce errors. They could accurately predict motion trajectories using this information and inertial navigation.

4.1.1.2.2 Rational Behaviour Method

The Rational Behaviour Method uses logical decision-making ideas to predict human trajectory [53]. Scientists study and model how people make rational and thoughtful decisions about where they want to go and what's around them [53]. This method goes beyond just looking at what people do; it considers the thinking processes behind their movements [54]. It believes people make logical decisions when deciding where to go, thinking about the shortest way to reach a destination, possible obstacles, and following social rules.

The method wants to make path predictions more accurate by including these logical decision-making ideas. It recognizes that how people move is not random but often guided by thoughtful decision-making processes. Adding these logical elements to the prediction model is meant to improve the overall process of predicting paths.

In [55], the presented model aims to understand how pedestrians decide where to walk, especially when avoiding collisions. Avoiding crashing into others is a complex process that involves conscious thought. Simply using mechanical rules to imitate this process can only achieve limited success. Their model has two main parts: one, where pedestrians take visual information about their surroundings and potential obstacles, and another, where they decide which path to avoid collisions.

4.1.1.2.3 Social Interaction Method

The Social Interaction Method examines how social dynamics, or how people interact, affect their paths [56]. This method understands that individual actions and interactions with others can influence where people walk. In this method, different ways or sub-methods help understand and predict human trajectory based on social situations. Figure 10 shows a pedestrian motion trajectory in a real scenario. Pedestrian trajectory prediction, considering social interaction information, shows that pedestrian trajectories are dynamic and multi-modal [57].

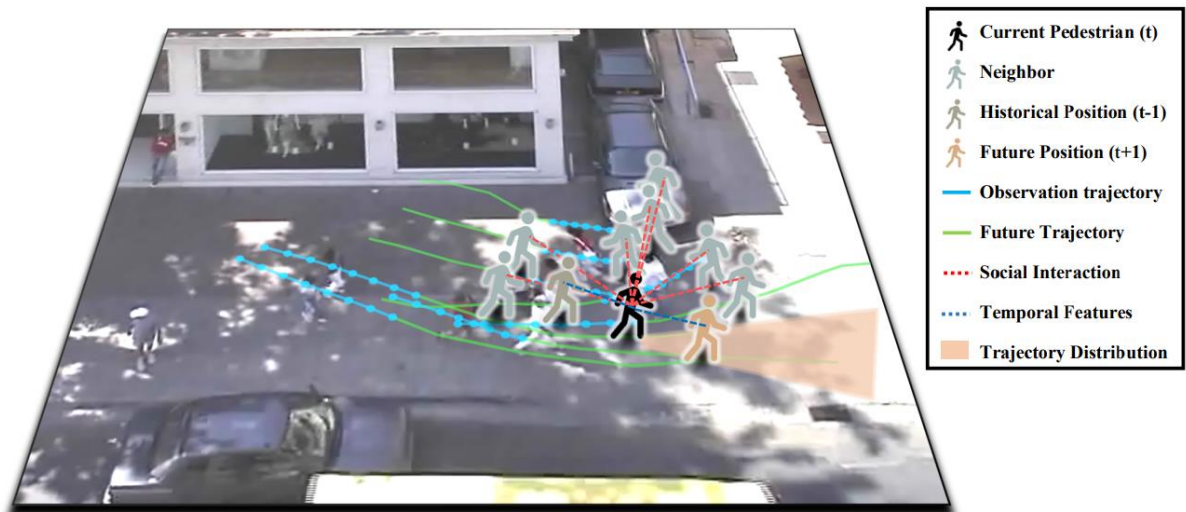


Figure 10, Image showing the social interaction method into consideration for human trajectory prediction. Here, the future trajectory is based on how people interact with each other socially. This image is taken from [57].

Following are the contributions from the research papers falling under this category:

The model in [58] creatively investigates interpretable social interactions in predicting pedestrian trajectories, an aspect often overlooked in existing research. The paper introduces a two-stage methodology involving "mode extraction" and "mode aggregation" and proposes an LSTM-based model designed explicitly for long-term trajectory prediction. The model naturally incorporates multiple types of social interactions, providing a comprehensive approach to understanding and predicting pedestrian movements.

In [38], the model employs a conditional 3D visual attention mechanism on input videos, allowing it to deduce environmental constraints and social interactions by observing the dynamic scene without relying on direct human communication.

On the other hand, [26] employs a unique approach by considering speculated trajectories in the future. This method extracts social interactive features and the models' interaction behaviors, presenting an alternative perspective on incorporating social interactions into the trajectory prediction process.

In [59], the authors introduce a novel method for training Deep Neural Networks (DNNs) in predicting pedestrian trajectories with social interaction. This approach involves a stochastic sampling-based simulation that produces many automatically annotated synthetic trajectories. These trajectories are realistic and naturalistic, generated from a small amount of accurate annotation.

In [60], the model explores how pedestrians interact with each other in groups, going beyond the usual consideration of their distance. The paper introduces a concept of social interaction modeling based on everyday movements, behaviors, goals, or destinations.

In [61], the model addresses a problem in human trajectory prediction using deep learning, specifically a generative adversarial network (GAN) called the Social Interaction GAN (SIGAN). The paper introduces a novel Social Interaction Module (SIM) that handles interactions between humans by incorporating details about their location and speed within a defined local area.

The model presented in [39] proposes a probabilistic method for predicting trajectories by leveraging latent belief energy. A key feature highlighted in this approach is incorporating a cost function within the latent space, ensuring that the model considers both the historical movement patterns and the social context of individuals. This emphasis on latent variables and social considerations contributes to a more comprehensive, context-aware trajectory forecasting model.

The paper [62] focuses on studying how groups of people interact and how these interactions affect the future paths of individuals within the groups. They introduce a model using Long Short-Term Memory (LSTM) with an encoder-decoder architecture to predict future trajectories, considering information about group interactions. They created a dataset using a motion planning algorithm. The results demonstrate that their approach performs better than the Trajnet++ benchmark when predicting group behaviors, even though both use a similar framework.

In [63], the authors introduce a hierarchical network to analyze where pedestrians intend to go and how they'll move in the future. They also create a unique graph structure, the Intention-Interaction Graph (IIG), to represent individual intentions and social interactions. The paper then proposes Interaction Gated Graph Attention Networks (IGGAN) to gather information from the IIG, helping to understand how neighboring pedestrians and planned destinations affect each other.

In [64], the authors introduce a novel concept known as the "scene-gated social feature," designed to enhance trajectory predictions by incorporating both scene and social cues. Through meticulous comparison studies with other feature merging methods, such as concatenation and addition. Secondly, the paper proposes a dynamic star graph data structure to model social graph features around pedestrians. This innovative structure proves memory efficient and remarkably effectively achieves state-of-the-art prediction results during experimental evaluations. Lastly, the paper extensively tests its proposed method, SGS, in various scenarios, showcasing its ability to outperform existing techniques in predicting single and multimodal future trajectories.

In this paper [65], the author deals with the challenge of forgetting important information, called the catastrophic forgetting problem, when learning to predict pedestrians' movements. The author introduces a new method called Social-GR, which uses a pseudo-rehearsal approach based on a unique Generative Replay (GR) model. This model aligns well with how crowds move and doesn't explicitly rely on remembering past experiences. To illustrate the issue of forgetting and to test their solution, they create the Continual Trajectory Prediction Benchmark with four tasks, each representing a real-world dataset of pedestrian paths in different environments. Through various experiments, they demonstrate that their Social-GR approach performs significantly better than other continual learning methods that rely on explicit experience replay, including the state-of-the-art conditional-GR model.

In [66], authors created a model called Neural Social Physics (NSP) to understand how people move in crowds. This model combines different ways of understanding movement and is built like a computer program. It has two main parts: the first is a deterministic model to understand how people move, and the good part is that it learns this formula from accurate data. The second part deals with the uncertainties in how people move and what the authors observe. It's like the model is trying to understand the patterns in how people move and the unpredictability of their actions. Overall, NSP is like an intelligent computer program that learns from real-world examples to determine how people behave in crowds.

In [67], the authors use a generative flow to mimic how humans move, from simple to more complex movements. Unlike older methods, this one can better understand how people move by improving accuracy. Additionally, they introduced a "dual grapheme" to know how people interact in time and space. This new approach is better than previous methods because it considers how people move over time and how they interact.

In [39], authors suggested a new model based on an inverse reinforcement learning (IRL) concept that understands how agents move differently. In this paper, considering social aspects, the method predicts possible paths people or different road agents might take in the future.

Social Interaction Methods can be divided further as well:

- Recursive Social Behavior Graph
- Social Information Encoding
- Social-Pyramid-based Prediction Method

- Recursive Social Behaviour Graph

The Recursive Social Behavior Graph is like drawing a map where each person is a dot, and the lines between them show how they interact [60]. It is unique because it doesn't just create a static map; it changes and updates as people's behaviors change over time [68]. This is important for predicting human trajectory in places where social interactions can affect how they walk [60]. So, in simpler terms, it's a clever way of using a constantly updating map to predict human trajectories based on how they interact [60]. In Figure 11, each color represents a different set of people in observation, and arrows show how these groups interact. We can define these groups and their interactions as a graph, a typical way to show how features spread [60].

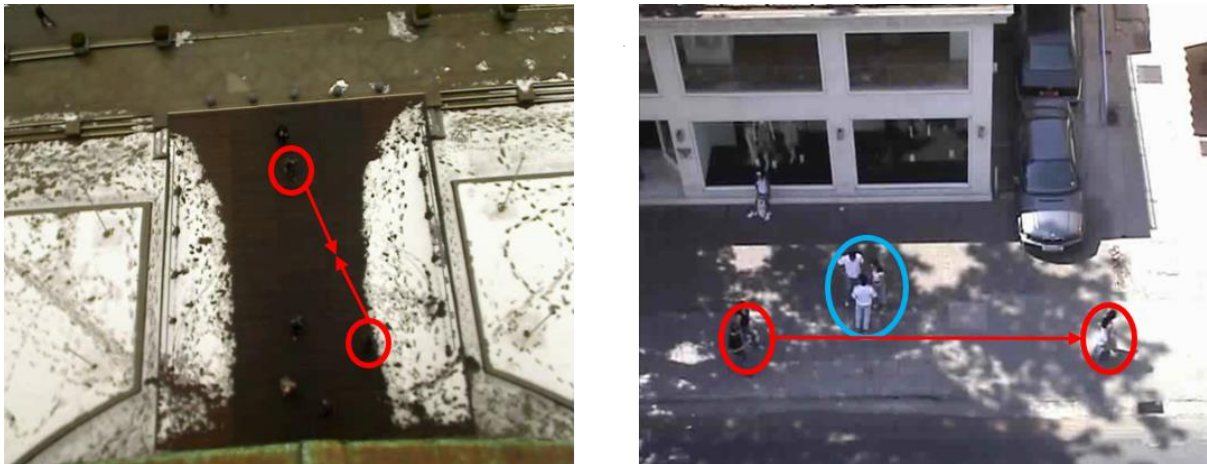


Figure 11, Examples of groups and their interactions: Red and blue circles show different groups. The arrows indicate the direction of influence in these interactions. Thus representing recursive social behavior. This image is taken from [60].

In [60], the authors introduce the Recursive Social Behavior Graph, which shows how people interact socially. They create it using a particular neural network, which helps uncover hidden

connections between people. This network learns from expert annotations about social behavior, a first in prediction tasks. Additionally, they use Graph Convolutional Networks (GCNs) to include human social behaviors in predicting what happens in different situations.

- Social Information Encoding

It's not just about predicting where a person will go; it's also about understanding their role in a group and how they connect with others. So, in simpler terms, Social Information Encoding is like adding a social touch to predictions, making them more accurate by considering how people move in social settings [69].

In [70], the authors recognized the effectiveness of sequence-to-sequence architecture in predicting sequences. They also introduced a unique sequence-to-sequence design to capture the interactions between objects and model the motion patterns of each trajectory. This novel approach aims to improve their understanding of how objects move about the environment.

- Social Pyramid-Based Prediction Method

The Social Pyramid-Based Prediction Method considers people's roles in a social group. It uses a social pyramid framework, like thinking about social hierarchies or who holds different positions in a group [71]. This method focuses on social roles, leadership, and how people interact to predict paths [7]. By understanding that human movement involves social aspects, this method makes predictions more accurate by including social relationships and group dynamics. It gives a deeper understanding of how people move in shared spaces, considering their movements and roles in a social group [7].

The model presented in [33], introduces a social pyramid-based prediction method (SPP), which incorporates two encoders to capture motion and social information. The paper presents a unique social pyramid map structure for the Social encoder, differentiating the influence of other pedestrians in nearby or remote areas based on their spatial locations. This approach provides a deeper understanding of trajectory prediction by considering spatial and social contexts.

4.1.1.2.4 Intention Estimation

The Intention Estimation method in predicting human trajectory is like figuring out what they might do next by watching what they're doing now. This method tries to understand why people move how they do, like if they're trying to get somewhere or avoid obstacles, to make more accurate predictions about where they'll go next [72]. It recognizes that people often have specific reasons for their actions and aims to include this vital information in prediction models [73]. This approach brings a cognitive aspect to predicting paths, helping us better understand how people move in different situations. As depicted in Fig. 12, the driver of the car will be able to choose the appropriate manoeuvres and routes to avoid potential collisions with pedestrians and improve driving comfort at the same time [74].



Figure 12: Here, the person seems to be intending to cross the street. This image shows that people's intentions depend on each other and other objects. It represents the intention estimation method, and it is taken from [74].

Following are the contributions from the research papers falling under this category:

The framework in [75], merges a Dynamic Bayesian network, sequence-to-sequence models, and an adaptive online weighting method. The Dynamic Bayesian network is highlighted for using environmental features and kinematic information to probabilistically understand the intentions behind pedestrian movements.

Unlike other approaches, the model in [76], focuses on generating and guiding goals. It achieves this by creating weighted goals through unsupervised learning of latent intentions. Goal-conditioned transformer networks then handle the prediction of trajectories, providing a probabilistic aspect to the proposals as the final trajectories.

Model in [77], the approach involves using pose estimation from CCTV videos to gather variables related to pedestrians. Four machine learning models are subsequently used to predict the crossing intentions of pedestrians, specifically at red-light intersections. The most accurate model achieved a 0.920 accuracy and an AUC value of 0.849, using data from three intersections. It was tested with different prediction times, up to 4 seconds into the future. As the prediction time increased, the model's accuracy decreased due to a smaller sample size. However, the model still performed well for predictions up to 2 seconds ahead, maintaining an AUC value of 0.841. Key factors contributing to the predictions included the angles between the ankle and knee on the left side and the elbow and shoulder on the right side.

In [78], the paper examines and predicts how pedestrians move when approaching crosswalks, a critical aspect of pedestrian and vehicle interactions in city environments. In one previous research study, the authors evaluated a data-driven architecture employing a Support Vector

Machine to understand pedestrian crossing intentions, mainly focusing on specific features. They extended their work in this paper by using previous results as a baseline to compare with a new neural network architecture for time-series classification. Specifically, they evaluated the effectiveness of dense and LSTM networks. The authors also plan to enhance their feature vectors by incorporating LiDAR images into the classification process.

In [50], the Authors proposed trajectory prediction for on-board camera systems, which utilizes behavioral features, including body orientation, head orientation, and pedestrian pose, to understand intentions. The authors also suggested a model in the image plane that uses the behavioral features of pedestrians.

On the other hand, in [79], the authors highlight the scarcity of public datasets for studying pedestrian behavior in the context of intelligent driving. The authors propose a new dataset, PIE, designed for pedestrian intention estimation to address this gap. They explain that a large-scale human experiment was conducted to collect reference data that captures the diversity of pedestrian intentions in traffic scenarios. Based on the PIE data, they created a model to understand pedestrians' intention to cross the road and thus predict pedestrian trajectory.

Lastly, [63], mentions the hierarchical network architecture, emphasizing its focus on predicting future trajectories. The novel aspect of the model, the Intention-Interaction Graph (IIG), is highlighted, indicating its role in capturing and modeling individual intentions and interactions. A model called Interaction Gated Graph Attention Networks (IGGAN) is proposed to gather information in IIG. This model uses gate and attention mechanisms to understand how neighbouring pedestrians and destinations influence each other.

4.1.1.2.5 Goal-Driven Method

The goal-driven method in predicting human trajectory is a unique way of figuring out where people might go next by understanding and including their intended goals or destinations [80]. This method focuses on the primary reasons driving people to move, which helps better predict where they'll go next [81]. Essentially, the Goal-Driven Method predicts human trajectory based on where they want to go. It recognizes that people walk around with specific goals, like reaching a particular place or avoiding certain areas. Figure 13, The Goal-driven Trajectory Prediction model breaks down a pedestrian's movement into two parts: one focuses on the long-term goal of reaching a destination. In contrast, the other part handles the detailed steps of moving to get there [81]. Following are the contributions from the research papers falling under this category:

The model presented in [82], emphasizes the method's goal-conditioned behavior of a Conditional Variational Autoencoder (CVAE) and incorporates a bidirectional decoder to improve accuracy, particularly in predicting longer-term trajectories.

In [28], the author emphasizes its attention-based recurrent backbone focused on past positions. It notes that while the method initially shows promise, its accuracy improves when combined with a scene-aware goal estimation module. The module utilizes a U-Net architecture to extract semantic information, enhancing the predictions' alignment with the scene's characteristics.

In [83], the authors highlight introducing the pedestrian's destination as a hidden variable, which changes the trajectory prediction problem into a planning problem. The planning is carried out based on the pedestrians' current dynamics, and a particle filter is used to model the distribution over potential destinations.

In [76], the author emphasizes the importance of generating and guiding goals in the context of human trajectory prediction. The Probabilistic Proposal Network (PPNet) is introduced as a method that produces multiple weighted goals from diverse latent intentions obtained through unsupervised learning. The network uses goal-conditioned Transformer networks to predict probabilistic proposals representing the final trajectories.

Presented in [84], the authors described the model as a network that estimates goals at multiple temporal scales. It consists of specific components, such as an encoder, a stepwise goal estimator, and a decoder, each with its role in the trajectory prediction process.

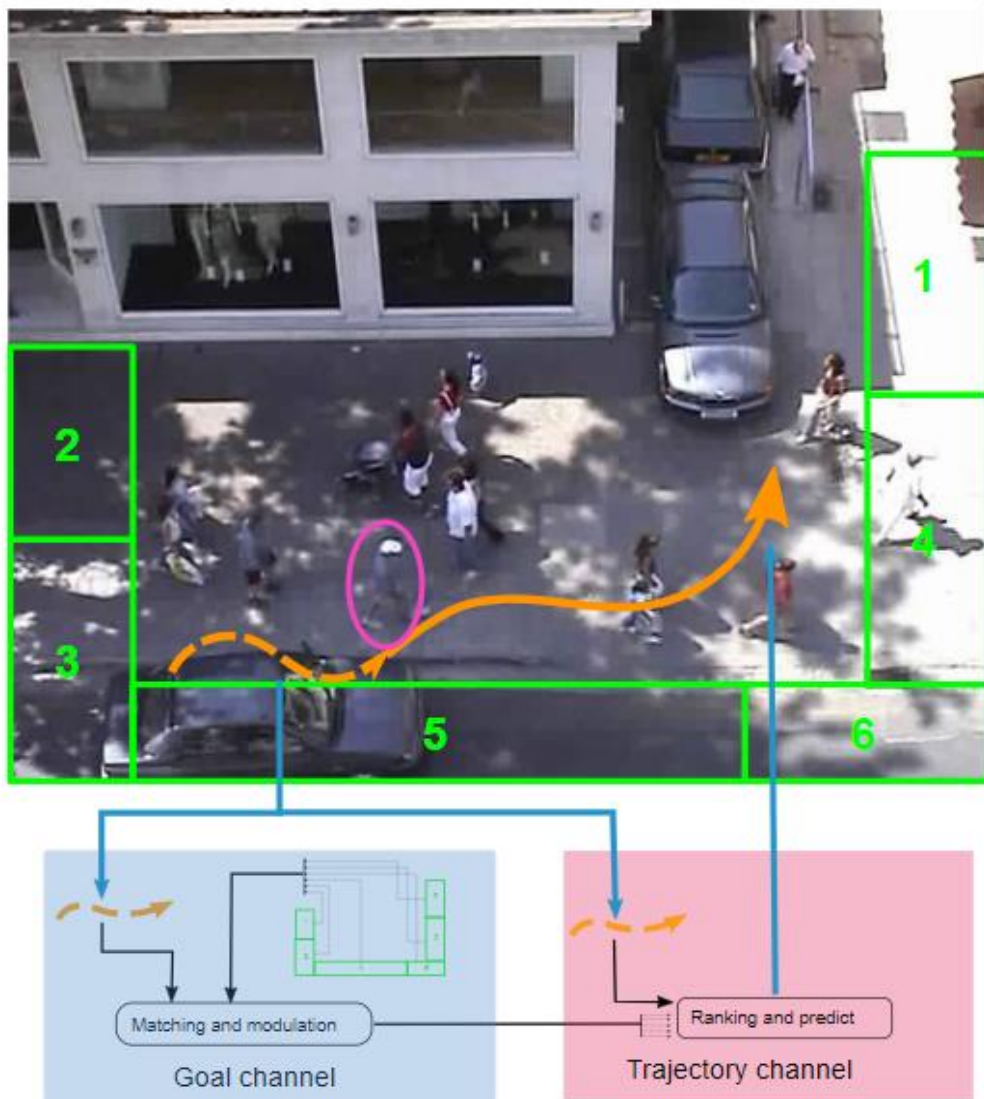


Figure 13 shows the trajectory according to a person's goal. The person encircled by pink wants to move towards the street on the left. Thus, his trajectory is predicted according to that. This image is taken from [81].

The model in [85], addresses the efficiency of robot movements by employing maximum entropy inverse optimal control to model pedestrians' goal-directed trajectories. One notable feature is the generality of the learned cost function, enabling the model to adapt to environmental changes and diverse scenarios.

Lastly, in [86], the presented model is highlighted for its interpretability and end-to-end trainability. The model is inspired by human navigation, employing a two-stage process

involving goal estimation to predict likely target positions and a routing module to estimate feasible trajectories leading toward the expected goal.

4.1.2 Pattern-based Approaches

Pattern-based methods for predicting how people move focus on finding and using repeated patterns in movement data to make better predictions. These methods know that people often follow the same movement patterns [36]. Understanding and using these patterns in predictive models can help us make more accurate predictions.

These methods are beneficial when certain movement sequences or behaviors happen regularly. For example, pedestrians might follow specific paths in busy cities when crossing crowded areas or intersections [87]. Recognizing and including these repeating patterns in models improves the accuracy of predictions in structured situations. It can be split into two categories:

- Spatio-Temporal Graphs
- Sequence-to-Sequence Model

4.1.2.1 Spatio-Temporal Graphs

A Spatiotemporal Graph is a unique method for predicting how people move. It uses a graph to show where and when movements happen [88]. This new approach aims to understand how people move by creating a map of the connections between locations over time, making predictions about movements more accurate.

In this type of graph, each point on the map represents a specific place, and the lines between these points show how they connect in time. Each moment holds details about its location, and the lines represent how different areas are linked over time [89]. By adding time information to the map, this method considers the order of human movement, recognizing that people move through spaces in a connected and time-related way. Figure 14, the solid blue lines represent the paths already taken, while the dashed blue lines show possible future paths. The influence levels between two individuals, described as α , vary based on how they move relative to each other. For instance, the influence of agent i on agent j (α_{ij}) is different from the influence of agent j on agent i (α_{ji}) [88]. The following are the contributions from the research papers that fall under this category:

In [31], the model is presented as a generative neural system specifically designed to predict trajectories in situations involving multiple agents. It emphasizes the importance of interactions by using relational inductive biases and a dynamic graph representation. This approach considers both the trajectories and the scene's context.

The approach in [29], presents a spatiotemporal graph that explicitly represents global interactions among all pedestrians and local interactions with stationary objects. Furthermore, the method proposes a new attention mechanism that operates in spatial and temporal dimensions, focusing on each pedestrian trajectory individually for more accurate predictions.

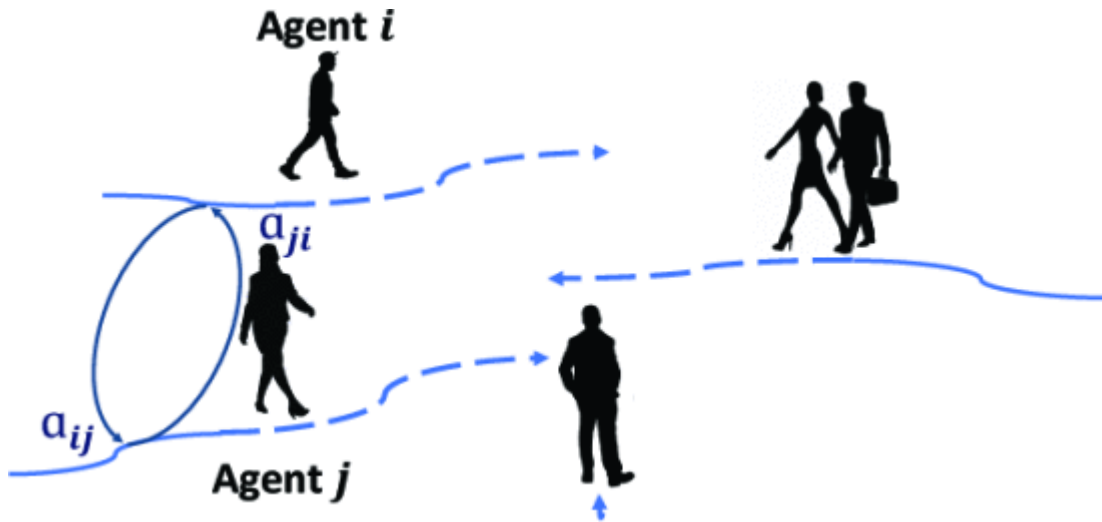


Figure 14, Trajectories of humans w.r.t others at a point in time. Thus, representing spatiotemporal methods. This image is taken from [88].

Following are the contributions from the research papers falling under this category:

In [90], the authors introduce a versatile framework applicable to various space and time tasks. It incorporates meta-path enhancements, which improve the network's understanding of the structural relationships within recurrent neural networks. This enhancement aims to provide better insights into the complex patterns present in spatiotemporal data.

In [75], the author introduces a stochastic trajectory prediction framework with motion indeterminacy diffusion, gradually removing uncertainty to determine the most likely path. They first turn human movement history and social interactions into a particular format called state embedding using a spatial-temporal graph network. Then, this state embedding will be used to help guide the learning of a reverse diffusion process in a Markov chain. To capture how movement changes over time, they created a transformer-based network as the central part of our MID framework.

As described in [91], the authors employ a novel transformer-based graph convolution mechanism to model intra-graph crowd interaction. The proposed model also introduces separate temporal Transformers to capture inter-graph temporal dependencies, effectively capturing complex spatiotemporal interactions through interleaving spatial and temporal Transformers.

4.1.2.2 Sequence-to-Sequence Model

The term "sequence-to-sequence" in the context of human trajectory prediction denotes a model architecture designed to handle sequential data both as input and output [92]. In simpler terms, it means the Model takes in a series of information (sequence) to make predictions and produces another sequence as its output. For trajectory prediction, this involves inputting a sequence of past locations or movements and generating a sequence representing the forecasted future trajectory [93]. This model is well-suited for tasks where understanding and predicting the order of events, such as the movement of pedestrians, is crucial. Figure 1: Given the observed paths of agents in the scene, a predictor (S-LSTM) can reasonably forecast their future positions (blue lines). However, a slight change of less than 5cm in the observed path (in red) can result in an unacceptable collision [94].

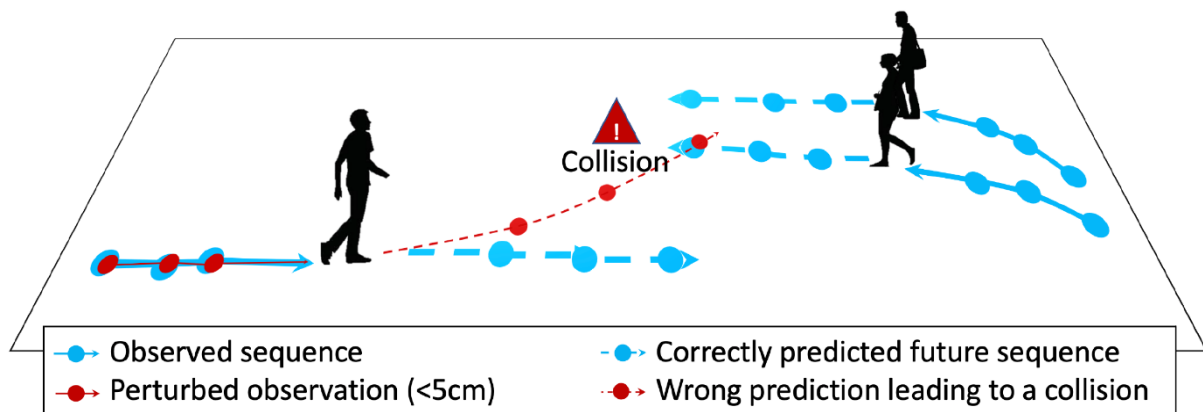


Figure 15 Shows how human trajectory prediction is based on the previous sequence of steps. If the previous sequence of steps differs by a certain angle, the future trajectory will show up as a collision. This image is taken from [94].

Following are the contributions from the research papers falling under this category:

The approach in [95], introduces a model examining how people interact in space and time. Spatial interactions, like how close people are, are handled using Graph Attention (GAT). This helps figure out which people are essential in each scene. Then, they use an LSTM to understand how these interactions change over time. By looking at all the spatial and temporal interactions between people, they predict future paths using a sequence-to-sequence (seq2seq) method. To account for different possible paths, they use variety loss to generate multiple likely outcomes.

In contrast, another approach in [23], employs a sequence-to-sequence model with a soft attention mechanism. Unlike some methods that use predefined navigational modes, this approach directly predicts human trajectories, emphasizing capturing human movement patterns' diverse and variable nature. This strategy enhances the model's adaptability to different navigation scenarios.

Another unique approach is introduced in [24]; the method employs a sequence of cross-attention and transformer blocks designed to capture relationships within high-dimensional data embeddings from historical information. The focus here is on enhancing the modeling of temporal dependencies in trajectory data, explicitly improving the understanding of detailed relationships within historical data.

The model in [96], emphasizes its focus on being interaction-aware in the context of human trajectory prediction. This method considers human interactions and models the joint probability distribution over movement sequences, recognizing individual interactions' crucial role in predicting trajectories. The goal is to provide a holistic perspective on human movement dynamics.

The method introduced in [97], enhances stochastic motion prediction in the context of human trajectory prediction. This improvement is achieved by incorporating randomness through a conditional variational autoencoder, which generates more diverse pose sequences. The outcome is a significant advancement in techniques for predicting a more comprehensive range of possible future trajectories.

The approach presented in [98], integrates a Dynamic Bayesian Network and a Sequence-to-Sequence model. This combined framework is complete, considering environmental features

and kinematic information. The method is versatile, as it can be applied to different spatiotemporal tasks, emphasizing its adaptability and capability to provide a holistic understanding of the environmental context.

4.1.3 Probabilistic Approaches

Probabilistic approaches in human trajectory prediction involve applying the principles of probability theory to model and predict the movement of individuals [128]. In contrast to deterministic methods that offer a single predicted trajectory, probabilistic models present a range of potential trajectories, each associated with a probability. This recognizes the inherent uncertainty in human movement and allows for a more comprehensive understanding of possible outcomes. It can further be divided into the following categories:

- Equilibrium-based Interaction Modeling (Nash Equilibrium)
- Latent Belief Energy-Based Model
- Latent Variational Model

4.1.3.1 Equilibrium-based Interaction Modeling (Nash Equilibrium)

It is an advanced strategy in human trajectory prediction that derives insights from game theory, specifically focusing on the Nash Equilibrium concept. In this context, "equilibrium" refers to a stable state where the predicted trajectories align with the optimal responses of all interacting agents [99].

This methodology operates on the assumption that individuals, as rational decision-makers, seek to optimize their trajectories by anticipating the movements of others. Nash Equilibrium, a fundamental principle in game theory, occurs when no agent can unilaterally deviate from their chosen trajectory to achieve a more favourable outcome [100]. In human trajectory prediction, this equilibrium-based approach models' interactions among individuals, considering their rational responses to others' movements.

The paper [101] deals with how people interact with each other using tools from multi-agent decision theory. It introduces a new way to solve problems using game theory, precisely the Nash equilibrium concept. This modification helps to predict paths for multiple pedestrians without needing supervision. The new method is tested on different scenarios and datasets to ensure it works in various situations.

4.1.3.2 Latent Belief Energy-Based Model

The Latent Belief Energy-Based Model for human trajectory prediction introduces an inventive approach that integrates latent beliefs and energy-based principles to enhance the precision of trajectory forecasts [39]. In this method, "latent beliefs" refer to unobservable factors or internal states that influence human decision-making during movement.

The central concept revolves around representing individuals' latent beliefs as energy functions. These functions encapsulate inherent preferences, intentions, and contextual considerations guiding human trajectories. The Model posits that individuals seek to minimize this energy or maximize alignment with their preferences.

In [39], authors came up with a model that operates as a probabilistic model with a cost function defined in the latent space, capturing things like past movements and social context. This helps make their predictions more accurate and diverse, covering different possibilities.

In [76], the authors introduced a model that uses intelligent networks to figure out other possibilities. It employs unsupervised learning to obtain diverse latent intentions, generating multiple weighted goals automatically. It learns from various intentions and creates different goals, predicting likely paths for people. This method highlights the importance of latent intentions and guiding predictions based on goals in forecasting trajectories.

In [44], the authors devised an excellent way to predict where people walk, especially for intelligent vehicles. They use a Dynamic Bayesian Network, which looks at different factors, such as how aware people are of their surroundings, how critical the situation is, and how the environment is laid out. Considering these hidden factors helps predict where pedestrians might go, especially in changing conditions.

These diverse methods collectively showcase the application of latent variables and belief energy-based models in human trajectory prediction. While [39] emphasizes a probabilistic model with a latent space cost function, [76] focuses on goal generation and guidance using latent intentions. The Dynamic Bayesian Network [44] also incorporates latent states for improved prediction accuracy in dynamic environments. Each approach offers a unique perspective on leveraging latent information to enhance the robustness and diversity of human trajectory forecasts.

4.1.3.3 Latent Variational Model

The Latent Variational Model for human trajectory prediction introduces an innovative strategy incorporating latent variables and variational inference to enhance the precision of trajectory forecasts [43]. In this method, "latent variables" refer to unobservable factors or hidden states influencing human movements.

The fundamental concept revolves around representing latent variables in a probabilistic manner. This probabilistic representation enables the capture of uncertainty and variability inherent in human behavior [102]. The Model states that trajectories originate from a latent distribution, and variational inference is utilized to approximate this distribution based on observed data.

In [43], authors suggest a model that combines one encoder RNN and two decoders RNN. The first decoder RNN learns about the observed trajectory and helps create a compact representation. The second decoder RNN predicts future positions using a latent variable recurrent model. This model learns to predict different possible paths based on the context. It captures uncertainty and considers various factors to forecast multiple plausible trajectories. The model's design is flexible and can be trained all at once. They introduce a way for the model to interact with the environment using visual cues from Convolutional Neural Networks (CNN) without needing semantic segmentation like previous methods.

4.2 Comparative Analysis for the Categories

This section presents a more detailed description of the categories in Fig 1. Given the hierarchical nature of the categories defined in Fig 1, it is logical to initiate the comparison at the highest level of the hierarchy and progressively descend through each subsequent level. This approach thoroughly examines each category, allowing for a detailed assessment of its characteristics and contributions to the study. Thus, offering insights into how categories at different levels interact and influence one another. This comprehensive method of categorical comparison explores the relationships and dependencies between the various categories at each

hierarchical level. This section also includes the benefits and applications of approaches for different categories. For some categories, situations where each category would be the most suitable choice are also described.

4.2.1 Comparative study for the methods at the highest level of the Hierarchy tree presented in Fig1

Cognitive approaches, pattern-based approaches, and Probabilistic approaches share a central focus on understanding and integrating human behavior into trajectory prediction. Their common goal is to improve the precision of trajectory predictions.

Cognitive Approaches:

Benefit:

- Focuses on understanding decision-making, intention estimation, and goal-driven behavior.
- Acknowledges and integrates the complexity of cognitive processes.
- Encompasses diverse methodologies to comprehend and apply cognitive mechanisms.

Application:

- Suited for scenarios where a deep understanding of cognitive intricacies is vital.
- Useful in scenarios where cognitive aspects significantly influence pedestrian behavior.
- Offers flexibility and adaptability in addressing various aspects of human behavior.

Use Cognitive Approaches when:

- Cognitive Understanding is Key: In situations where a deep understanding of cognitive processes is essential to predict human behavior accurately.
- Behavioral details Matter: The goal is to capture delicate aspects of human behavior, such as decision-making and goal-driven behavior.

Pattern-Based Approaches:

Benefit:

- Focuses on identifying and utilizing recurrent patterns in movement data.
- Proves benefits in structured scenarios, such as urban environments with consistent movement sequences.
- Enhances accuracy by recognizing and incorporating specific movement sequences or behaviors.

Application:

- Effective in scenarios where human trajectories exhibit identifiable and repetitive patterns.
- Well-suited for predicting trajectories in specific environments where patterns are prevalent.
- This is particularly effective in scenarios where certain behaviors consistently occur.

Use Pattern-Based Approaches when:

- Identifiable Patterns Exist: In scenarios where human trajectories exhibit identifiable and repetitive patterns.

- **Structured Environments:** In environments where specific movement sequences or behaviors consistently occur.

Probabilistic Approaches:

Benefit:

- Explicitly deals with and models uncertainty through probability theory.
- Presents a range of potential trajectories with associated probabilities.
- Effective in scenarios where a probabilistic outlook is essential.

Application:

- This is particularly advantageous when uncertainty significantly influences human movement prediction.
- Useful when a comprehensive understanding of various possible outcomes is crucial.
- This is particularly applicable in situations where multiple potential outcomes are plausible.

Use Probabilistic Approaches when:

- **Uncertainty is High:** In scenarios where uncertainty plays a crucial role, probabilistic approaches offer a more explicit handling of uncertainties.
- **Comprehensive Outlook is Required:** A comprehensive understanding of potential outcomes and associated probabilities is needed.

4.2.1.1 Comparative analysis for cognitive approaches:

Cognitive approaches are divided into two categories: surrounding-based and behavioral-based methods. Both share a foundational principle: recognizing the critical role of human-centric factors in trajectory prediction. Both approaches aim to enhance the precision of trajectory predictions. Surrounding-based methods leverage environmental context, while behavioral-based methods utilize insights from human behavior to achieve this goal. Each of these categories is divided further into different categories.

Surrounding-Based Methods:

Benefit:

- Primarily concentrates on understanding the external environment's influence on pedestrian trajectories.
- Highlights the importance of external elements (context, surroundings) in determining pedestrian trajectories.

Application:

- Well-suited for scenarios where environmental factors, such as dynamic urban landscapes, significantly shape human movement.
- Provides a detailed approach by categorizing methods based on attention to specific features or considering broader contextual information.
- Particularly effective when the surrounding environment plays a crucial role in influencing movement patterns.

Surrounding-Based Methods suit when:

- **Context is Crucial:** In scenarios where the external environment significantly affects pedestrian trajectories.
- **Dynamic Surroundings:** Attention to specific features or a broader context is essential when surroundings are dynamic.

Behavioral-Based Methods:

Benefit:

- Leverages insights from human behavior to enhance trajectory predictions.
- Encompasses various approaches, each focusing on different aspects of human behavior.
- Acknowledges that diverse behavioral aspects influence pedestrians' movements.

Application:

- Effective in scenarios where understanding and modeling human actions are vital for accurate predictions.
- Offers flexibility to tailor the prediction model based on specific behavioral aspects relevant to the scenario.
- Effective in capturing the complexities of human behavior, considering factors like walking speed, changes in direction, and social interactions.

Behavioral-based methods are effective when:

- **Human Behavior Understanding is Essential:** In situations where insights into human behavior, such as walking patterns and interactions, are critical for accurate predictions.
- **Behavioral Variability Matters:** When diverse aspects of human behavior must be captured and modeled, movement variability is a concern.

Consideration:

- **Hybrid Approach:** A combined or hybrid approach, integrating elements of both surrounding-based and behavioral-based methods, may offer a more holistic and accurate prediction model, especially in scenarios where environmental context and human behavior play pivotal roles.

4.2.1.1.a Comparative analysis for surrounding based methods:

Surrounding-based methods are divided into two categories: Attention-Based Methods and Context-Aware Methods. Both share a common emphasis on considering the environment in human trajectory prediction. They recognize that environmental factors are crucial in shaping and influencing human movement.

Attention-Based Methods:

Focus:

- Concentrate on directing attention to specific environmental features, utilizing techniques like Attention Methods and Location-Velocity Attention.

- Specifically, target critical elements, such as locations or velocities, for effective trajectory prediction.

Enhancement of Prediction:

- Enhance prediction effectiveness by emphasizing crucial elements within the environment.

Insights Offered:

- Tend to provide detailed insights by focusing on elements significantly impacting human movement.

Context-Aware Methods:

Broader Approach:

- Take a broader approach by considering the overall contextual information and recognizing the environment's role in shaping trajectories.

Comprehensive Understanding:

- Encompass a broader range of contextual factors beyond specific features, aiming for a more comprehensive understanding of the environment's impact.

Accuracy Improvement:

- Aim to improve accuracy by considering a broader set of contextual information and recognizing the holistic influence of the environment.

Holistic View:

- Offer a holistic view by considering diverse contextual factors to understand the environment's role better.

Differences between Attention-based methods and Context-aware methods:

- Attention-based methods narrow down their focus to specific features for detailed predictions.
- Context-aware methods adopt a broader perspective, incorporating various contextual factors to understand the environment comprehensively.
- Attention-based methods enhance prediction by emphasizing specific elements.
- Context-aware methods improve accuracy by considering a broader set of contextual information.

Summary:

- While Attention-Based Methods and Context-Aware Methods share a common goal of recognizing the environment's importance, they diverge in approach and scope.
- Attention-based methods narrow their focus to specific features for detailed predictions, offering insights into elements that significantly impact human movement.
- Context-aware methods adopt a broader perspective, considering various contextual factors for a more detailed and comprehensive understanding of the environment's influence on human trajectory.

- The decision to choose between the two methods would depend on the specific requirements of the trajectory prediction task and the desired level of detail in considering environmental factors.

4.2.1.1.a.1 Comparative analysis for attention-based methods:

Attention-based methods are divided into two categories, Attention Methods and Location-Velocity Attention Methods, which involve a spatial component in their predictive models.

Attention Methods:

Focus: Primarily identifies crucial locations or time intervals within a trajectory.

Spatial Emphasis: Offers detailed insights into spatial dynamics, emphasizing the importance of specific locations.

Temporal Consideration: Provides varying attention levels for different trajectory segments, contributing to a detailed analysis of temporal dynamics.

Insights Offered: Offers detailed insights into the complexities of human movement patterns, pivotal moments, and locations.

Location-Velocity Attention Method:

Spatial and Temporal Emphasis:

- Specifically, it incorporates the rate of change in position (velocity) and spatial information.
- Offers a more comprehensive temporal consideration by combining spatial and temporal dimensions.

Attention Mechanisms:

- Prioritizes attention mechanisms for both spatial (location) and temporal (velocity) aspects.

Comprehensive Understanding:

- Aims for a holistic understanding of human trajectories by considering spatial and temporal elements

Insights Offered:

- Provides insights into detailed patterns and dynamics, considering the interplay of spatial and temporal factors.

Summary of Differences:

Focus:

- Attention Methods primarily focus on spatial dynamics, while the Location-Velocity Attention Method considers spatial and temporal dimensions with velocity.

Insight Depth:

- Attention Methods provide detailed insights into specific spatial aspects.

- The Location-Velocity Attention Method offers a more comprehensive insight by considering spatial and temporal dimensions.

In summary, Attention Methods and Location-Velocity Attention Methods contribute to human trajectory prediction by recognizing the importance of location and incorporating dynamic analyses. Attention Methods offer detailed insights into spatial dynamics. At the same time, the Location-Velocity Attention Method aims for a more comprehensive understanding by considering both spatial and temporal dimensions, providing detailed insights into detailed patterns and dynamics over time.

4.2.1.1.b Comparative analysis for behavioral methods:

These trajectory prediction methods share a common extensive goal: understanding human behavior and predicting future movements. However, their distinctions become apparent in how they approach this shared objective.

Behavioral Features Method:

Focus: Keenly examines discernible behavioral patterns.

Approach: This approach enhances trajectory predictions by closely analyzing observable actions like walking speed, direction changes, and environment interactions.

Significance: Prioritizes direct observation of behaviors to uncover patterns that contribute to more accurate trajectory predictions.

Rational Behaviour Method:

Focus: Integrates principles of rational decision-making.

Approach: Goes beyond behavioral observation, considering logical processes such as goals, intentions, and environmental considerations.

Significance: Recognizes that human movements are guided by intentional and logical decision processes, contributing to a deeper understanding of the cognitive aspects influencing trajectories.

Social Interaction Method:

Focus: Explicitly factors in the dynamics of social interactions.

Approach: Recognizes that pedestrian paths are shaped by individual actions and profoundly influenced by engagements with others in the surrounding environment.

Significance: Acknowledges the social context and dynamics, emphasizing the collective influence on pedestrian trajectories.

Intention Estimation Method:

Focus: Introduces a cognitive layer to the prediction process.

Approach: Predicts inherent intentions steering human movements by deducing future actions or objectives based on the analysis of observed behavior.

Significance: Enriches the predictive model with cognitive insights, considering the underlying intentions that drive human actions.

Goal-Driven Method:

Focus: Hones in on forecasting trajectories based on anticipated goals.

Approach: Accentuates the purposeful nature ingrained in human locomotion, considering the objectives individuals aim to achieve in their movements.

Significance: Emphasizes human behavior's intentional and goal-oriented nature, contributing to trajectory predictions based on anticipated objectives.

Summary:

Ultimately, the nuances in these methods lie in the specific aspects of cognition and behavior they prioritize. While all share a common ground in cognitive focus and trajectory prediction, each technique brings a unique perspective by emphasizing distinct facets of human behavior. This diversity enriches the field by offering varied approaches to foreseeing future pedestrian trajectories, making the overall understanding of human movement more comprehensive.

4.2.1.1.b.1 Comparative analysis for social interaction methods:

This section provides a comparison between the methods within the Social interaction category.

Recursive Social Behavior Graph:

Approach:

- Utilizes graphical representation to model the evolving nature of social relationships.
- Predicts future trajectories by analyzing the dynamic interactions within the graph.

Application:

- Effective in scenarios where pedestrian social relationships are dynamic and evolve.
- Captures intricate social dynamics influencing individual movements.

Social Information Encoding:

Approach:

- Directly integrates social information into the prediction model.
- Anticipates trajectory adjustments based on various social contexts and relationships.

Application:

- Well-suited for scenarios where group dynamics and interpersonal relationships significantly impact pedestrian movements.
- Provides a detailed understanding of how social information influences trajectory predictions.

Social Pyramid-Based Prediction Method:

Approach:

- Utilizes a hierarchical structure to predict trajectories based on social roles, leadership, and group dynamics.
- Consider the influence of social hierarchies on individual movements.

Application:

- Effective in scenarios where social roles and hierarchical structures shape pedestrian trajectories.
- Captures the impact of leadership and group dynamics on individual movements.

Comparison: Comparison of Recursive Social Behavior Graph, Social Information Encoding, and Social Pyramid-Based Prediction in the context of:

Graphical Representation:

- Recursive Social Behavior Graph: A dynamic graph represents social interactions.
- Social Information Encoding: Does not rely on evolving graphical representations.
- Social Pyramid-Based Prediction: Utilizes a hierarchical social structure without a dynamic graph.

Social Dynamics:

- Recursive Social Behavior Graph: Captures evolving social dynamics within a graphical framework.
- Social Information Encoding: Considers social cues and information without focusing on graphical evolution.
- Social Pyramid-Based Prediction: Models hierarchical social structures and their influence on trajectories.

Hierarchical Representation:

- Recursive Social Behavior Graph: Represents interactions without a specific hierarchical structure.
- Social Information Encoding: Considers various social aspects without a hierarchical framework.
- Social Pyramid-Based Prediction: Employs a social pyramid hierarchy for prediction.

In summary, these methods differ in their approaches to modeling and utilizing social information. The Recursive Social Behavior Graph emphasizes dynamic graphical representations, Social Information Encoding focuses on directly integrating social cues, and the Social Pyramid-Based Prediction Method employs a hierarchical structure to capture social hierarchies and roles in trajectory prediction. The choice depends on the scenario and the detail level required to model social interactions.

4.2.1.2 Comparative analysis for pattern-based approaches

Both spatiotemporal graphs and sequence-to-sequence models aim to enhance the accuracy of trajectory forecasts. The Spatiotemporal Graph and Sequence-to-Sequence Model align with the broader machine-learning paradigm for sequence prediction tasks. Both methods utilize information from past trajectories to inform predictions about future movements.

Spatio-Temporal Graph:

Benefit:

- Provides a visual and structural representation of the relationships between locations over time.
- Acknowledges the sequential and time-dependent nature of human movement.
- Edges encode temporal connections or transitions, capturing the dependencies between different locations.

Application:

- Well-suited for scenarios where understanding the inherent structure of human mobility and capturing temporal dependencies are crucial.
- Spatiotemporal graphs can predict potential events or gatherings in crowded spaces based on individuals' collective movement patterns.
- Spatio-temporal graphs can predict potential events or gatherings in crowded spaces based on the collective movement patterns of individuals.

Spatio-temporal graphs are effective when understanding the spatial relationships and temporal dependencies between locations is paramount.

Sequence-to-Sequence Model:

Benefits:

- Offers a systematic approach to capturing the order and progression of human movement.
- Characterized by a model architecture specifically crafted for sequential data.
- Takes a sequence of past locations or movements as input and generates a sequence representing the anticipated future trajectory.

Application:

- Particularly effective in tasks where predicting the order of events, such as pedestrian movement, is essential.
- Well-suited for scenarios where understanding the sequence of events is critical for accurate predictions.
- Suitable for various applications where sequence-based prediction is required beyond human trajectory prediction.

Sequence-to-Sequence model is suitable when predicting future trajectories based on the sequential order of past events, which is crucial.

Consideration: A hybrid approach that combines elements of both methods may offer a more comprehensive model, leveraging the strengths of each in scenarios where both spatial relationships and sequential order are significant.

4.2.1.3 Comparative analysis for probabilistic methods

This section provides a comparison between probabilistic methods.

Equilibrium-Based Interaction Modeling (Nash Equilibrium):

Benefits:

- Nash Equilibrium ensures no agent can avoid achieving a more favorable outcome.
- Incorporates principles from game theory, focusing on rational decision-making by individuals.
- Models interactions among individuals, considering rational responses to others' movements.

Applications:

- Suitable for scenarios where predicting trajectories involves understanding how individuals optimize their movements based on anticipating others' actions.
- Valuable in situations where stable and coordinated movements are essential, such as crowded spaces or scenarios with multiple interacting pedestrians.
- Useful in environments with intricate dynamics, such as intersections or areas with diverse pedestrian behaviors.

Use Equilibrium-Based Interaction Modeling when:

- The emphasis is on stable states and optimal responses in predicting trajectories.
- Rational decision-making and game theory principles are crucial for understanding interactions.

Latent Belief Energy-Based Model:

Benefits:

- Captures unobservable factors and internal states influencing human decision-making during movement.
- Represents latent beliefs as energy functions.
- Consider contextual considerations guiding human trajectories.

Applications:

- Well-suited for applications requiring personalized trajectory forecasts.
- This is relevant when considering the influence of latent beliefs on decision-making.
- Effective in scenarios where a detailed understanding of latent beliefs is critical.

Use Latent Belief Energy-Based Model when:

- Emphasis is on capturing inherent preferences and intentions influencing human trajectories.
- Internal states and latent beliefs are critical factors for trajectory prediction.

Latent Variational Model:

Benefits:

- Captures unobservable factors or hidden states influencing human movements.
- Enables the capture of uncertainty in human behavior.
- Utilizes variational inference to approximate latent distribution.

Applications:

- Effective in scenarios where uncertainty in human behavior is prominent.
- Appropriate for applications requiring probabilistic trajectory predictions.
- Relevant when capturing variability in human behavior is crucial.

Use the Latent Variational Model when:

- Capturing uncertainty through a probabilistic representation is crucial for trajectory prediction.
- The focus is on incorporating latent variables and understanding their influence on human movements.

Consideration: A hybrid approach that combines elements from Equilibrium-Based Modeling, Latent Belief Energy-Based Model, and Latent Variational Model may offer a more comprehensive trajectory prediction model, especially in scenarios where stable states, latent beliefs, and uncertainty play significant roles.

4.3 Summary Table

Table 3 represents references for the research papers used to formulate categories. It's essential to know that some papers might not neatly fit into these presented categories because a research paper might use many categories altogether. So, the same research paper reference might be mentioned in front of many categories.

4.4 Dataset

This section describes the dataset used by most of the research papers studied in this survey for their prediction models. Thus, the focus was given to the papers using this dataset so the performances of different models from different papers could be accessed on the same dataset, leaving less chance of dataset-specific bias. The dataset is intended to assist researchers in forecasting pedestrians' future positions by analyzing their previous movements.

The testing set includes 1,804 images across three video clips. These images are captured by a stereo rig mounted on a car, with a resolution of 640 x 480 (layered) and a frame rate of 13-14 frames per second (FPS)¹⁰. The UCY dataset contains accurate pedestrian trajectories with diverse multi-human interaction scenarios. These trajectories are captured at a rate of 2.5 Hz (with a time interval of 0.4 seconds). The dataset comprises three sequences (Zara01, Zara02, and UCY) filmed from a top-view perspective in public spaces.

¹⁰ <https://paperswithcode.com/dataset/eth>

Table 3. This table provides the references for the research papers using the method written before them for Human trajectory prediction.

Cognitive Approaches	Surrounding Based Approaches	Attention Based Methods	Attention Method	[23][24][25][26][27][28][29][30][31][32][33][38][91]	
			Location Velocity Attention Method	[25][30]	
		Context-Aware Methods		[27][31][38][39][40][41][42][43][44][46]	
	Behavioral methods	Behavioral Features Method		[40][49][50][51][52]	
		Rational Behaviour Method		[55]	
		Social Interaction Methods		[26][33][38][39][41][45][58][59][60][61][62][63][64][65][66][67][70][75]	
		Intention Estimation		[50][75][76][77][78][79][63]	
		Goal-Driven Methods		[28][76][82][83][84][85][86]	
	Pattern Based Approaches	Spatio-temporal graphs	[29][31][75][90][91]		
		Sequence-to-sequence models	[23][24][95][96][97][98]		
Probabilistic Approaches	Equilibrium Based Interaction Modeling	[101]			
	Latent Belief Energy-based Model	[39][44][76]			
	Latent Variational Model	[40]			

4.5 Performances

This section discusses the performance of each category presented in Table 3. Since many papers use multiple categories simultaneously, stating the performance for each category is quite tricky. However, this section provides a rough idea about the performance of categories, which gives better results regarding Average Displacement Error (ADE) and Final Displacement Error (FDE). Accuracies were presented in different forms in various papers. However, many papers provided performance in ADE/FDE and tested their model on public datasets – ETH, HOTEL, UNIV, ZARA1, and ZARA2. In this study, ADE/FDE values were

taken out from each of these papers, and then we deduced the ADE/FDE values for each category.

To understand the calculation of category-wise ADE/FDE, let's take an example – A paper uses the attention and context-aware methods together. The paper provides a dataset's ADE/FDE value as 0.4323/0.7810. The ADE/FDE values provided by this paper will be noted for both categories. The same would happen for other papers using the same categories. Thus, each dataset will have multiple ADE/FDE values for each category. The average was taken from the values obtained in the last step to find one value for each category per dataset. Table 4 here provides the ADE/FDE values for each category, calculated as per the above example for each dataset, and references from which these observations are taken.

Table 4. This table provides the averaged ADE/FDE values for each category per dataset (with the ETH/UCY dataset) and the average values for each category for the whole dataset, together with the references for the research papers from which these observations are taken.

Method	ETH ADE/ FDE	HOTEL ADE /FDE	UNIV ADE/ FDE	ZARA1 ADE/ FDE	ZARA2 ADE/ FDE	AVG ADE/ FDE	REFERNCES
Attention method	0.5691/ 0.9609	0.3336/ 0.6227	0.3664/ 0.6309	0.3255/ 0.6355	0.2427/ 0.4027	0.3263/ 0.5272	[23][24][25][26][28] [29][31][32][38][91]
Behavioral Features Method	0.3100/ 0.5700	0.1300/ 0.2100	0.2500/ 0.5300	0.2000/ 0.4100	0.1400/ 0.3100	0.2000/ 0.4100	[40]
Equilibrium Based Interaction Modeling	0.8800/ 1.7800	0.5600/ 1.1200	0.5600/ 1.2100	0.4200/ 0.9100	0.3400/ 0.7300	0.5500/ 1.1500	[101]
Latent Variational Model	0.6100/ 1.0900	0.2800/ 0.5100	0.5900/ 0.1240	0.3700/ 0.7800	0.3000/ 0.6400	0.4300/ 0.8500	[43]
Location Velocity Attention Method	0.5700/ 1.1000	0.4200/ 0.6900	0.5500/ 1.1900	0.4200/ 0.9200	0.3500/ 0.7500	0.4600/ 0.9200	[25]
Context-Aware Methods	0.4040/ 0.7300	0.1800/ 0.2960	0.3440/ 0.4628	0.2320/ 0.4660	0.1920/ 0.3900	0.2700/ 0.5140	[31][38][39][40][43]
Goal-Driven Methods	0.4240/ 0.8040	0.1600/ 0.3000	0.3320/ 0.6640	0.2360/ 0.4640	0.1860/ 0.3680	0.2660/ 0.5180	[28][76][82][84][86]
Intention Estimation	0.4400/ 0.7200	0.1650/ 0.2550	0.3450/ 0.6700	0.2600/ 0.4800	0.2050/ 0.3950	0.2850/ 0.5000	[76]
Latent Belief Energy-based Model	0.2600/ 0.4250	0.1100/ 0.1650	0.2450/ 0.4300	0.1900/ 0.3200	0.1450/ 0.2550	0.1900/ 0.3150	[39][76]
Sequence-to- sequence models	0.5000/ 0.9233	0.2733/ 0.4367	0.4100/ 0.5900	0.3733/ 0.7700	0.3033/ 0.4567	0.3233/ 0.5300	[23][24][96]

Social Interaction Methods	0.5660/ 0.9970	0.2950/ 0.5350	0.4320/ 0.8630	0.2860/ 0.5610	0.2390/ 0.4710	0.3636/ 0.7226	[8][10][21][23][24][36]
Spatio-Temporal graphs	0.5667/ 0.5267	0.2567/ 0.2867	0.4167/ 0.6067	0.3533/ 0.4500	0.3367/ 0.4167	0.3866/ 0.4600	[91][29][31][90]

5 Summary and Emerging Trends

This chapter briefly summarizes the study and discusses emerging trends and future research areas.

5.1 Summary

This study explains a systematic way to read the research papers and conduct a survey study, resulting in a categorization based on the approaches used in the research papers. The formulated categorization can be observed in Fig 1. To understand those categories better, this study describes each category briefly; categories at each level of the tree map were compared, and the benefits and applicability of each category have been mentioned. Then, the research papers falling under each category are presented in Table 3.

To understand the performance for each category, ADE and FDE values for the research papers using those categories were noted. Since many research papers use many categories, the performance measure mentioned in the research papers would not represent the performance for each category in the real sense. So, the average ADE/FDE was calculated for each category by taking ADE and FDE values from all the papers using the respective category. According to the average values for each category presented in Table 4, for the complete data ETH/UCY (Including ETH, UNIV, HOTEL, ZARA1, and ZARA2), the lowest values for ADE and FDE are for the Latent Belief Energy-based Model. Given the limited number of research papers on this method, the result may be biased. So, it would be wiser to consider the category for which the average ADE and FDE values are second least in the table, which is the “Behavioral Feature Method.” According to this study, it is possible to conclude that the behavioral features method is the category to be explored further in future works to achieve better performance.

5.2 Emerging Trends

Traditionally, trajectory prediction focused mainly on historical movement patterns. Still, the current trend emphasizes the importance of considering additional factors, such as the environment, behavioral features, scene semantics, and social interactions. The following are the directions in which the topic can be extended further.

- **Multi-Modal Sensor Fusion:** The increased utilization of various sensor modalities, such as cameras, LiDAR, and radar, is becoming more common. Integrating information from multiple sensors allows for a more comprehensive understanding of the surrounding environment, leading to improved trajectory predictions, especially in complex and dynamic scenarios.
- **Semantic Scene Understanding:** Incorporating semantic information about the scene, such as identifying objects, road structures, and pedestrian intentions, contributes to a more contextually aware prediction model. This trend aligns with advancements in computer vision and scene-understanding techniques.
- **Social Interaction Modeling:** Recognizing and modeling social interactions among pedestrians has gained importance. Predictive models now consider the influence of

group dynamics, social norms, and individual behavior, contributing to more realistic and socially acceptable trajectory predictions.

- **Attention Mechanisms and Transformers:** The adoption of attention mechanisms and transformer architectures in trajectory prediction is rising. These mechanisms enable models to focus on relevant spatial and temporal information, capturing long-range dependencies and intricate patterns in human movement.
- **Uncertainty Estimation:** Acknowledging the inherent uncertainty in predicting human trajectories, recent trends involve developing models that provide point predictions and uncertainty estimates. This is crucial for decision-making systems, especially in safety-critical applications.
- **Meta-Learning and Adaptive Models:** Meta-learning approaches, where models can adapt quickly to new scenarios with limited data, are gaining traction. Adaptive models that continuously learn and update their predictions based on real-time feedback contribute to more responsive and reliable trajectory prediction systems.
- **Real-World Testing and Validation:** As trajectory prediction models advance, there is a growing emphasis on testing and validating these models in real-world scenarios. Evaluation metrics are evolving better to capture models' performance in complex and diverse environments.

In summary, the current trajectory prediction trend involves a holistic approach integrating multi-modal data, scene semantics, and social interactions. This trend aims to create more accurate, context-aware, and adaptable trajectory prediction models suitable for real-world applications.

5.3 Future Research Area

This study explains the categorizing of different methods and compares these categories. Since many research papers use various techniques together to create a model, it would be beneficial for future research to understand and analyze each model separately and then compare them. Existing research papers often compare models based on only ADE/FDE values. However, it would be helpful to explore the advantages and disadvantages of each model compared to others.

Reference

- [1] V. V. Unhelkar, C. Pérez-D'Arpino, L. Stirling, and J. A. Shah, "Human-robot co-navigation using anticipatory indicators of human walking motion," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 6183–6190. doi: [10.1109/ICRA.2015.7140067](https://doi.org/10.1109/ICRA.2015.7140067).
- [2] K. Chen, X. Song, D. Han, J. Sun, Y. Cui, and X. Ren, "Pedestrian behavior prediction model with a convolutional LSTM encoder-decoder," *Physica A: Statistical Mechanics and its Applications*, vol. 560, p. 125132, Dec. 2020, doi: [10.1016/j.physa.2020.125132](https://doi.org/10.1016/j.physa.2020.125132).
- [3] F. Camara *et al.*, "Pedestrian Models for Autonomous Driving Part I: Low-Level Models, From Sensing to Tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, Jun. 2020, doi: [10.1109/TITS.2020.3006768](https://doi.org/10.1109/TITS.2020.3006768).
- [4] M. Gulzar, Y. Muhammad, and N. Muhammad, "A Survey on Motion Prediction of Pedestrians and Vehicles for Autonomous Driving," *IEEE Access*, vol. 9, pp. 137957–137969, 2021, doi: [10.1109/ACCESS.2021.3118224](https://doi.org/10.1109/ACCESS.2021.3118224).
- [5] S. Xu, V. Dutta, X. He, and T. Matsumaru, "A Transformer-Based Model for Super-Resolution of Anime Image," *Sensors*, vol. 22, no. 21, Art. no. 21, Jan. 2022, doi: [10.3390/s22218126](https://doi.org/10.3390/s22218126).
- [6] B. I. Sighencea, "Pedestrian Trajectory Prediction Based on Tree Method using Graph Neural Networks," in *2022 24th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Sep. 2022, pp. 245–249. doi: [10.1109/SYNASC57785.2022.00046](https://doi.org/10.1109/SYNASC57785.2022.00046).
- [7] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: a survey," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, Jul. 2020, doi: [10.1177/0278364920917446](https://doi.org/10.1177/0278364920917446).
- [8] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, and H. Chen, "A Survey on Trajectory-Prediction Methods for Autonomous Driving," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 652–674, Sep. 2022, doi: [10.1109/TIV.2022.3167103](https://doi.org/10.1109/TIV.2022.3167103).
- [9] B. Sighencea, R. Stanciu, and C. Căleanu, "A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction," *Sensors*, vol. 21, p. 7543, Nov. 2021, doi: [10.3390/s21227543](https://doi.org/10.3390/s21227543).
- [10] E. Schuetz and F. B. Flohr, "A Review of Trajectory Prediction Methods for the Vulnerable Road User," *Robotics*, vol. 13, no. 1, Art. no. 1, Jan. 2024, doi: [10.3390/robotics13010001](https://doi.org/10.3390/robotics13010001).
- [11] K. Lyu, H. Chen, Z. Liu, B. Zhang, and R. Wang, "3D human motion prediction: A survey," *Neurocomputing*, vol. 489, pp. 345–365, Jun. 2022, doi: [10.1016/j.neucom.2022.02.045](https://doi.org/10.1016/j.neucom.2022.02.045).
- [12] R. Korbmacher and A. Tordeux, "Review of Pedestrian Trajectory Prediction Methods: Comparing Deep Learning and Knowledge-Based Approaches," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24126–24144, Dec. 2022, doi: [10.1109/TITS.2022.3205676](https://doi.org/10.1109/TITS.2022.3205676).
- [13] H. Handoll, W. Gillespie, L. Gillespie, and R. Madhok, "The Cochrane Collaboration: A leading role in producing reliable evidence to inform healthcare decisions in musculoskeletal trauma and disorders," *Indian journal of orthopaedics*, vol. 42, pp. 247–51, Jul. 2008, doi: [10.4103/0019-5413.41849](https://doi.org/10.4103/0019-5413.41849).

- [14] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and PRISMA Group, “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement,” *PLoS Med*, vol. 6, no. 7, p. e1000097, Jul. 2009, doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097).
- [15] M. Bahari, S. Saadatnejad, A. A. Farsangi, S.-M. Moosavi-Dezfooli, and A. Alahi, “Certified Human Trajectory Prediction,” Mar. 20, 2024, *arXiv*: arXiv:2403.13778. doi: [10.48550/arXiv.2403.13778](https://doi.org/10.48550/arXiv.2403.13778).
- [16] N. Gorlo, L. Schmid, and L. Carlone, “Long-Term Human Trajectory Prediction using 3D Dynamic Scene Graphs,” May 01, 2024, *arXiv*: arXiv:2405.00552. doi: [10.48550/arXiv.2405.00552](https://doi.org/10.48550/arXiv.2405.00552).
- [17] X. Lin, T. Liang, J. Lai, and J.-F. Hu, “Progressive Pretext Task Learning for Human Trajectory Prediction,” Jul. 16, 2024, *arXiv*: arXiv:2407.11588. doi: [10.48550/arXiv.2407.11588](https://doi.org/10.48550/arXiv.2407.11588).
- [18] P. M. Torrens, A. Nara, X. Li, H. Zhu, W. A. Griffin, and S. B. Brown, “An extensible simulation environment and movement metrics for testing walking behavior in agent-based models,” *Computers, Environment and Urban Systems*, vol. 36, no. 1, pp. 1–17, Jan. 2012, doi: [10.1016/j.compenvurbsys.2011.07.005](https://doi.org/10.1016/j.compenvurbsys.2011.07.005).
- [19] L. G. Galvão and M. N. Huda, “Pedestrian and vehicle behaviour prediction in autonomous vehicle system — A review,” *Expert Syst. Appl.*, vol. 238, no. PC, Feb. 2024, doi: [10.1016/j.eswa.2023.121983](https://doi.org/10.1016/j.eswa.2023.121983).
- [20] A. Sadri, F. D. Salim, Y. Ren, W. Shao, J. C. Krumm, and C. Mascolo, “What Will You Do for the Rest of the Day? An Approach to Continuous Trajectory Prediction,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 4, p. 186:1-186:26, Dec. 2018, doi: [10.1145/3287064](https://doi.org/10.1145/3287064).
- [21] C. Shen, X. Xiao, S. Li, and Y. Tong, “Improved Attention Mechanism for Human-like Intelligent Vehicle Trajectory Prediction,” *Electronics*, vol. 12, p. 3993, Sep. 2023, doi: [10.3390/electronics12193993](https://doi.org/10.3390/electronics12193993).
- [22] A. Manafi Soltan Ahmadi and S. Hoseini Semnani, “Human trajectory prediction using LSTM with Attention mechanism,” Sep. 01, 2023. doi: [10.48550/arXiv.2309.00331](https://doi.org/10.48550/arXiv.2309.00331).
- [23] D. Varshneya and G. Srinivasaraghavan, “Human Trajectory Prediction using Spatially aware Deep Attention Models,” May 26, 2017, *arXiv*: arXiv:1705.09436. doi: [10.48550/arXiv.1705.09436](https://doi.org/10.48550/arXiv.1705.09436).
- [24] A. Postnikov, A. Gamayunov, and G. Ferrer, “Conditioned Human Trajectory Prediction using Iterative Attention Blocks,” Jun. 29, 2022, *arXiv*: arXiv:2206.14442. doi: [10.48550/arXiv.2206.14442](https://doi.org/10.48550/arXiv.2206.14442).
- [25] H. Xue, D. Q. Huynh, and M. Reynolds, “A Location-Velocity-Temporal Attention LSTM Model for Pedestrian Trajectory Prediction,” *IEEE Access*, vol. 8, pp. 44576–44589, 2020, doi: [10.1109/ACCESS.2020.2977747](https://doi.org/10.1109/ACCESS.2020.2977747).
- [26] L. Pi, Q. Zhang, L. Yang, and Z. Huang, “Social interaction model enhanced with speculation stage for human trajectory prediction,” *Robotics and Autonomous Systems*, vol. 161, p. 104352, Mar. 2023, doi: [10.1016/j.robot.2022.104352](https://doi.org/10.1016/j.robot.2022.104352).
- [27] L. Li, B. Zhou, J. Lian, X. Wang, and Y. Zhou, “Multi-PPTP: Multiple Probabilistic Pedestrian Trajectory Prediction in the Complex Junction Scene,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 13758–13768, Aug. 2022, doi: [10.1109/TITS.2021.3127539](https://doi.org/10.1109/TITS.2021.3127539).

- [28] L. F. Chiara, P. Coscia, S. Das, S. Calderara, R. Cucchiara, and L. Ballan, “Goal-driven Self-Attentive Recurrent Networks for Trajectory Prediction,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2022, pp. 2517–2526. doi: [10.1109/CVPRW56347.2022.00282](https://doi.org/10.1109/CVPRW56347.2022.00282).
- [29] S. Haddad, M. Wu, W. He, and S. K. Lam, “Situation-Aware Pedestrian Trajectory Prediction with Spatio-Temporal Attention Model,” Jan. 2019. doi: [10.3217/978-3-85125-652-9](https://doi.org/10.3217/978-3-85125-652-9).
- [30] H. Xue, D. Huynh, and M. Reynolds, “Location-Velocity Attention for Pedestrian Trajectory Prediction,” in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2019, pp. 2038–2047. doi: [10.1109/WACV.2019.00221](https://doi.org/10.1109/WACV.2019.00221).
- [31] J. Li, H. Ma, Z. Zhang, J. Li, and M. Tomizuka, “Spatio-Temporal Graph Dual-Attention Network for Multi-Agent Prediction and Tracking,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 10556–10569, Aug. 2022, doi: [10.1109/TITS.2021.3094821](https://doi.org/10.1109/TITS.2021.3094821).
- [32] L. Zhou, Y. Zhao, D. Yang, and J. Liu, “GCHGAT: pedestrian trajectory prediction using group constrained hierarchical graph attention networks,” *Appl Intell*, vol. 52, no. 10, pp. 11434–11447, Aug. 2022, doi: [10.1007/s10489-021-02997-w](https://doi.org/10.1007/s10489-021-02997-w).
- [33] H. Xue, D. Q. Huynh, and M. Reynolds, “Pedestrian Trajectory Prediction Using a Social Pyramid,” in *PRICAI 2019: Trends in Artificial Intelligence*, A. C. Nayak and A. Sharma, Eds., Cham: Springer International Publishing, 2019, pp. 439–453. doi: [10.1007/978-3-030-29911-8_34](https://doi.org/10.1007/978-3-030-29911-8_34).
- [34] S. T. L. Chung and S. S. Patel, “Spatial and temporal proximity of objects for maximal crowding,” *Vision Research*, vol. 194, p. 108012, May 2022, doi: [10.1016/j.visres.2022.108012](https://doi.org/10.1016/j.visres.2022.108012).
- [35] L. Lao, D. Du, and P. Chen, “Predicting Pedestrian Trajectories with Deep Adversarial Networks Considering Motion and Spatial Information,” *Algorithms*, vol. 16, p. 566, Dec. 2023, doi: [10.3390/a16120566](https://doi.org/10.3390/a16120566).
- [36] X. Zou, B. Sun, D. Zhao, Z. Zhu, J. Zhao, and Y. He, “Multi-Modal Pedestrian Trajectory Prediction for Edge Agents Based on Spatial-Temporal Graph,” *IEEE Access*, vol. 8, pp. 83321–83332, 2020, doi: [10.1109/ACCESS.2020.2991435](https://doi.org/10.1109/ACCESS.2020.2991435).
- [37] F. Bartoli, G. Lisanti, L. Ballan, and A. Del Bimbo, “Context-Aware Trajectory Prediction,” May 06, 2017, *arXiv*: arXiv:1705.02503. doi: [10.48550/arXiv.1705.02503](https://doi.org/10.48550/arXiv.1705.02503).
- [38] N. Shafiee, T. Padir, and E. Elhamifar, “Introvert: Human Trajectory Prediction via Conditional 3D Attention,” presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Jun. 2021, pp. 16810–16820. doi: [10.1109/CVPR46437.2021.01654](https://doi.org/10.1109/CVPR46437.2021.01654).
- [39] B. Pang, T. Zhao, X. Xie, and Y. N. Wu, “Trajectory Prediction with Latent Belief Energy-Based Model,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 11809–11819. doi: [10.1109/CVPR46437.2021.01164](https://doi.org/10.1109/CVPR46437.2021.01164).
- [40] J. Sun, Y. Li, L. Chai, H.-S. Fang, Y.-L. Li, and C. Lu, “Human Trajectory Prediction with Momentary Observation,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 6457–6466. doi: [10.1109/CVPR52688.2022.00636](https://doi.org/10.1109/CVPR52688.2022.00636).

- [41] J. Long, J. Mei, and G. Ma, “Egocentric Two-Frame Pedestrian Trajectory Prediction Algorithm Based on a Panoramic Camera,” *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–13, 2023, doi: [10.1109/TIM.2022.3225018](https://doi.org/10.1109/TIM.2022.3225018).
- [42] C. Zhou, G. AlRegib, A. Parchami, and K. Singh, “Learning Trajectory-Conditioned Relations to Predict Pedestrian Crossing Behavior,” in *2022 IEEE International Conference on Image Processing (ICIP)*, Oct. 2022, pp. 4088–4092. doi: [10.1109/ICIP46576.2022.9897655](https://doi.org/10.1109/ICIP46576.2022.9897655).
- [43] A. Díaz Berenguer, M. Alioscha-Perez, M. C. Oveneke, and H. Sahli, “Context-Aware Human Trajectories Prediction via Latent Variational Model,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 5, pp. 1876–1889, May 2021, doi: [10.1109/TCSVT.2020.3014869](https://doi.org/10.1109/TCSVT.2020.3014869).
- [44] J. Kooij, N. Schneider, F. Flohr, and D. Gavrilu, “Context-Based Pedestrian Path Prediction,” Sep. 2014. doi: [10.1007/978-3-319-10599-4_40](https://doi.org/10.1007/978-3-319-10599-4_40).
- [45] Y. Li, M. L. Mekhalfi, M. M. Al Rahhal, E. Othman, and H. Dhahri, “Encoding Motion Cues for Pedestrian Path Prediction in Dense Crowd Scenarios,” *IEEE Access*, vol. 5, pp. 24368–24375, 2017, doi: [10.1109/ACCESS.2017.2772029](https://doi.org/10.1109/ACCESS.2017.2772029).
- [46] A. Kalatian and B. Farooq, “A context-aware pedestrian trajectory prediction framework for automated vehicles,” *Transportation Research Part C: Emerging Technologies*, vol. 134, p. 103453, Jan. 2022, doi: [10.1016/j.trc.2021.103453](https://doi.org/10.1016/j.trc.2021.103453).
- [47] D. Ridel, E. Rehder, M. Lauer, C. Stiller, and D. Wolf, “A Literature Review on the Prediction of Pedestrian Behavior in Urban Scenarios,” in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 3105–3112. doi: [10.1109/ITSC.2018.8569415](https://doi.org/10.1109/ITSC.2018.8569415).
- [48] B. Williams, J. Myerson, and S. Hale, “Individual Differences, Intelligence, and Behavior Analysis,” *Journal of the experimental analysis of behavior*, vol. 90, pp. 219–31, Oct. 2008, doi: [10.1901/jeab.2008.90-219](https://doi.org/10.1901/jeab.2008.90-219).
- [49] A. Yildiz, N. Takemura, Y. Iwai, and K. Sato, “Regression Based Trajectory Learning and Prediction for Human Motion,” in *Image and Video Technology – PSIVT 2013 Workshops*, F. Huang and A. Sugimoto, Eds., Berlin, Heidelberg: Springer, 2014, pp. 193–202. doi: [10.1007/978-3-642-53926-8_18](https://doi.org/10.1007/978-3-642-53926-8_18).
- [50] P. Czech, M. Braun, U. Kreßel, and B. Yang, “On-Board Pedestrian Trajectory Prediction Using Behavioral Features,” Oct. 21, 2022, *arXiv*: arXiv:2210.11999. doi: [10.48550/arXiv.2210.11999](https://doi.org/10.48550/arXiv.2210.11999).
- [51] J. Yang, A. Gui, J. Wang, and J. Ma, “Pedestrian Behavior Interpretation from Pose Estimation,” in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, Sep. 2021, pp. 3110–3115. doi: [10.1109/ITSC48978.2021.9565098](https://doi.org/10.1109/ITSC48978.2021.9565098).
- [52] L. Chen, Y. Zhao, Y. Zhang, S. Fan, and Y. Jia, “An Online Prediction and Trajectory Tracking Method for Human Activity Recognition,” in *Proceedings of the 2020 9th International Conference on Computing and Pattern Recognition*, in ICCPR ’20. New York, NY, USA: Association for Computing Machinery, Jan. 2021, pp. 18–23. doi: [10.1145/3436369.3436458](https://doi.org/10.1145/3436369.3436458).
- [53] Y. Ma, E. W. Lee, Z. Hu, M. Shi, and R. K. Yuen, “An Intelligence-Based Approach for Prediction of Microscopic Pedestrian Walking Behavior,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3964–3980, Oct. 2019, doi: [10.1109/TITS.2019.2931892](https://doi.org/10.1109/TITS.2019.2931892).

- [54] O. C. Yıldırım, A. Sungur, and D. G. Ozer, “Exploring multidimensional aspects of walkability: An innovative analysis approach in Besiktas, Istanbul,” *Frontiers of Architectural Research*, vol. 12, no. 6, pp. 1097–1126, Dec. 2023, doi: [10.1016/j.foar.2023.07.004](https://doi.org/10.1016/j.foar.2023.07.004).
- [55] R. Bailo, J. Carrillo, and P. Degond, “Pedestrian Models Based on Rational Behaviour: Theory, Models, and Safety Problems,” in *Modeling and Simulation in Science, Engineering and Technology*, 2018, pp. 259–292. doi: [10.1007/978-3-030-05129-7_9](https://doi.org/10.1007/978-3-030-05129-7_9).
- [56] K. Chen, X. Song, and X. Ren, “Modeling social interaction and intention for pedestrian trajectory prediction,” *Physica A: Statistical Mechanics and its Applications*, vol. 570, p. 125790, May 2021, doi: [10.1016/j.physa.2021.125790](https://doi.org/10.1016/j.physa.2021.125790).
- [57] Y. Liu, B. Li, X. Wang, C. Sammut, and L. Yao, “Attention-Aware Social Graph Transformer Networks for Stochastic Trajectory Prediction,” *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–14, 2024, doi: [10.1109/TKDE.2024.3390765](https://doi.org/10.1109/TKDE.2024.3390765).
- [58] Q. Liu, X. Shi, R. Jiang, H. Zhang, L. Lu, and R. Shibasaki, “Modeling interpretable social interactions for pedestrian trajectory,” *Transportation Research Part C: Emerging Technologies*, vol. 162, p. 104617, May 2024, doi: [10.1016/j.trc.2024.104617](https://doi.org/10.1016/j.trc.2024.104617).
- [59] C. Anderson, X. Du, R. Vasudevan, and M. Johnson-Roberson, “Stochastic Sampling Simulation for Pedestrian Trajectory Prediction,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2019, pp. 4236–4243. doi: [10.1109/IROS40897.2019.8967857](https://doi.org/10.1109/IROS40897.2019.8967857).
- [60] J. Sun, Q. Jiang, and C. Lu, “Recursive Social Behavior Graph for Trajectory Prediction,” Apr. 22, 2020, *arXiv*: arXiv:2004.10402. doi: [10.48550/arXiv.2004.10402](https://doi.org/10.48550/arXiv.2004.10402).
- [61] S. Zhang, J. Wu, J. Dong, and L. Liu, “Social-Interaction GAN: Pedestrian Trajectory Prediction,” in *Wireless Algorithms, Systems, and Applications*, Z. Liu, F. Wu, and S. K. Das, Eds., Cham: Springer International Publishing, 2021, pp. 429–440. doi: [10.1007/978-3-030-86137-7_46](https://doi.org/10.1007/978-3-030-86137-7_46).
- [62] I. Sethi *et al.*, “Group-Aware Human Trajectory Prediction,” in *2022 IEEE 6th Conference on Information and Communication Technology (CICT)*, Nov. 2022, pp. 1–5. doi: [10.1109/CICT56698.2022.9997980](https://doi.org/10.1109/CICT56698.2022.9997980).
- [63] C. Li, H. Yang, and J. Sun, “Intention-Interaction Graph Based Hierarchical Reasoning Networks for Human Trajectory Prediction,” *IEEE Transactions on Multimedia*, pp. 1–12, 2022, doi: [10.1109/TMM.2022.3182151](https://doi.org/10.1109/TMM.2022.3182151).
- [64] H. Xue, D. Q. Huynh, and M. Reynolds, “Scene Gated Social Graph: Pedestrian Trajectory Prediction Based on Dynamic Social Graphs and Scene Constraints,” Oct. 12, 2020, *arXiv*: arXiv:2010.05507. doi: [10.48550/arXiv.2010.05507](https://doi.org/10.48550/arXiv.2010.05507).
- [65] Y. Wu, A. Bighashdel, G. Chen, G. Dubbelman, and P. Jancura, “Continual Pedestrian Trajectory Learning With Social Generative Replay,” *IEEE Robot. Autom. Lett.*, vol. 8, no. 2, pp. 848–855, Feb. 2023, doi: [10.1109/LRA.2022.3231833](https://doi.org/10.1109/LRA.2022.3231833).
- [66] J. Yue, D. Manocha, and H. Wang, “Human Trajectory Prediction via Neural Social Physics,” Mar. 31, 2023, *arXiv*: arXiv:2207.10435. doi: [10.48550/arXiv.2207.10435](https://doi.org/10.48550/arXiv.2207.10435).
- [67] R. Liang, Y. Li, J. Zhou, and X. Li, “STGlow: A Flow-based Generative Framework with Dual Graphormer for Pedestrian Trajectory Prediction,” *IEEE Trans. Neural Netw. Learning Syst.*, pp. 1–14, 2023, doi: [10.1109/TNNLS.2023.3294998](https://doi.org/10.1109/TNNLS.2023.3294998).
- [68] J. Zhou *et al.*, “Graph neural networks: A review of methods and applications,” *AI Open*, vol. 1, pp. 57–81, Jan. 2020, doi: [10.1016/j.aiopen.2021.01.001](https://doi.org/10.1016/j.aiopen.2021.01.001).

- [69] K. Mangalam *et al.*, “It Is Not the Journey but the Destination: Endpoint Conditioned Trajectory Prediction,” Jul. 18, 2020, *arXiv*: arXiv:2004.02025. doi: [10.48550/arXiv.2004.02025](https://doi.org/10.48550/arXiv.2004.02025).
- [70] Z. Pei, X. Qi, Y. Zhang, M. Ma, and Y.-H. Yang, “Human trajectory prediction in crowded scene using social-affinity Long Short-Term Memory,” *Pattern Recognition*, vol. 93, pp. 273–282, Sep. 2019, doi: [10.1016/j.patcog.2019.04.025](https://doi.org/10.1016/j.patcog.2019.04.025).
- [71] P. Rathore, D. Kumar, S. Rajasegarar, M. Palaniswami, and J. C. Bezdek, “A Scalable Framework for Trajectory Prediction,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3860–3874, Oct. 2019, doi: [10.1109/TITS.2019.2899179](https://doi.org/10.1109/TITS.2019.2899179).
- [72] I. Hasan *et al.*, “Forecasting People Trajectories and Head Poses by Jointly Reasoning on Tracklets and Vislets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 4, pp. 1267–1278, Apr. 2021, doi: [10.1109/TPAMI.2019.2949414](https://doi.org/10.1109/TPAMI.2019.2949414).
- [73] L. Nadel and W. Sinnott-Armstrong, *Conscious Will and Responsibility: A Tribute to Benjamin Libet*. 2010, p. 288. doi: [10.1093/acprof:oso/9780195381641.001.0001](https://doi.org/10.1093/acprof:oso/9780195381641.001.0001).
- [74] J. Lian, F. Yu, L. Li, and Y. Zhou, “Early intention prediction of pedestrians using contextual attention-based LSTM,” *Multimed Tools Appl*, vol. 82, no. 10, pp. 14713–14729, Apr. 2023, doi: [10.1007/s11042-022-13814-y](https://doi.org/10.1007/s11042-022-13814-y).
- [75] T. Gu *et al.*, “Stochastic Trajectory Prediction via Motion Indeterminacy Diffusion,” Mar. 25, 2022, *arXiv*: arXiv:2203.13777. doi: [10.48550/arXiv.2203.13777](https://doi.org/10.48550/arXiv.2203.13777).
- [76] W. Chen, Z. Yang, L. Xue, J. Duan, H. Sun, and N. Zheng, “Multimodal Pedestrian Trajectory Prediction Using Probabilistic Proposal Network,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 6, pp. 2877–2891, Jun. 2023, doi: [10.1109/TCSVT.2022.3229694](https://doi.org/10.1109/TCSVT.2022.3229694).
- [77] 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*, vol. 23, no. 3, pp. 2331–2339, Mar. 2022, doi: [10.1109/TITS.2021.3074829](https://doi.org/10.1109/TITS.2021.3074829).
- [78] 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)*, Nov. 2016, pp. 2607–2612. doi: [10.1109/ITSC.2016.7795975](https://doi.org/10.1109/ITSC.2016.7795975).
- [79] A. Rasouli, I. Kotseruba, T. Kunic, and J. Tsotsos, “PIE: A Large-Scale Dataset and Models for Pedestrian Intention Estimation and Trajectory Prediction,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 6261–6270. doi: [10.1109/ICCV.2019.00636](https://doi.org/10.1109/ICCV.2019.00636).
- [80] J. Zhou, X. Bai, F. Weiping, B. Ning, and R. Li, “Pedestrian intention estimation and trajectory prediction based on data and knowledge-driven method,” *IET Intelligent Transport Systems*, vol. 18, p. n/a-n/a, Nov. 2023, doi: [10.1049/itr2.12453](https://doi.org/10.1049/itr2.12453).
- [81] H. Tran, V. Le, and T. Tran, “Goal-driven Long-Term Trajectory Prediction,” in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2021, pp. 796–805. doi: [10.1109/WACV48630.2021.00084](https://doi.org/10.1109/WACV48630.2021.00084).
- [82] Y. Yao, E. Atkins, M. Johnson-Roberson, R. Vasudevan, and X. Du, “BiTraP: Bi-Directional Pedestrian Trajectory Prediction With Multi-Modal Goal Estimation,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1463–1470, Apr. 2021, doi: [10.1109/LRA.2021.3056339](https://doi.org/10.1109/LRA.2021.3056339).

- [83] E. Rehder and H. Kloeden, “Goal-Directed Pedestrian Prediction,” in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, Dec. 2015, pp. 139–147. doi: [10.1109/ICCVW.2015.28](https://doi.org/10.1109/ICCVW.2015.28).
- [84] C. Wang, Y. Wang, M. Xu, and D. J. Crandall, “Stepwise Goal-Driven Networks for Trajectory Prediction,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2716–2723, Apr. 2022, doi: [10.1109/LRA.2022.3145090](https://doi.org/10.1109/LRA.2022.3145090).
- [85] B. D. Ziebart *et al.*, “Planning-based prediction for pedestrians,” in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct. 2009, pp. 3931–3936. doi: [10.1109/IROS.2009.5354147](https://doi.org/10.1109/IROS.2009.5354147).
- [86] P. Dendorfer, A. Ošep, and L. Leal-Taixé, “Goal-GAN: Multimodal Trajectory Prediction Based on Goal Position Estimation,” arXiv.org. Accessed: Aug. 08, 2024. [Online]. Available: <https://arxiv.org/abs/2010.01114v1>
- [87] N. Molyneaux and M. Bierlaire, “Controlling pedestrian flows with moving walkways,” *Transportation Research Part C: Emerging Technologies*, vol. 141, p. 103672, Aug. 2022, doi: [10.1016/j.trc.2022.103672](https://doi.org/10.1016/j.trc.2022.103672).
- [88] C. Wang, S. Cai, and G. Tan, “GraphTCN: Spatio-Temporal Interaction Modeling for Human Trajectory Prediction,” in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2021, pp. 3449–3458. doi: [10.1109/WACV48630.2021.00349](https://doi.org/10.1109/WACV48630.2021.00349).
- [89] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, “Structural-RNN: Deep Learning on Spatio-Temporal Graphs,” presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Jun. 2016, pp. 5308–5317. doi: [10.1109/CVPR.2016.573](https://doi.org/10.1109/CVPR.2016.573).
- [90] A. Hasan, P. Sriram, and K. Driggs-Campbell, “Meta-path Analysis on Spatio-Temporal Graphs for Pedestrian Trajectory Prediction,” in *2022 International Conference on Robotics and Automation (ICRA)*, May 2022, pp. 617–624. doi: [10.1109/ICRA46639.2022.9811632](https://doi.org/10.1109/ICRA46639.2022.9811632).
- [91] J. Zhu, Z. Lian, and Z. Jiang, “A Spatio-Temporal Graph Transformer Network for Multi-Pedestrian Trajectory Prediction,” in *2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCoM/CyberSciTech)*, Sep. 2022, pp. 1–5. doi: [10.1109/DASC/PiCom/CBDCoM/Cy55231.2022.9927798](https://doi.org/10.1109/DASC/PiCom/CBDCoM/Cy55231.2022.9927798).
- [92] D.-D. Nguyen, C. Van, and M. I. Ali, “Vessel Trajectory Prediction using Sequence-to-Sequence Models over Spatial Grid,” Jun. 2018, pp. 258–261. doi: [10.1145/3210284.3219775](https://doi.org/10.1145/3210284.3219775).
- [93] P. Lv *et al.*, “Trajectory distributions: A new description of movement for trajectory prediction,” *Computational Visual Media*, vol. 8, pp. 213–224, Jun. 2022, doi: [10.1007/s41095-021-0236-6](https://doi.org/10.1007/s41095-021-0236-6).
- [94] S. Saadatnejad, M. Bahari, P. Khorsandi, M. Saneian, S.-M. Moosavi-Dezfooli, and A. Alahi, “Are socially-aware trajectory prediction models really socially-aware?,” *Transportation Research Part C: Emerging Technologies*, vol. 141, p. 103705, Aug. 2022, doi: [10.1016/j.trc.2022.103705](https://doi.org/10.1016/j.trc.2022.103705).
- [95] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, “STGAT: Modeling Spatial-Temporal Interactions for Human Trajectory Prediction,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 6271–6280. doi: [10.1109/ICCV.2019.00637](https://doi.org/10.1109/ICCV.2019.00637).

- [96] B. Brito, H. Zhu, W. Pan, and J. Alonso-Mora, “Social-VRNN: One-Shot Multi-modal Trajectory Prediction for Interacting Pedestrians,” Nov. 14, 2020, *arXiv*: arXiv:2010.09056. doi: [10.48550/arXiv.2010.09056](https://doi.org/10.48550/arXiv.2010.09056).
- [97] S. Aliakbarian, F. Sadat Saleh, M. Salzmann, L. Petersson, and S. Gould, “A Stochastic Conditioning Scheme for Diverse Human Motion Prediction,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 5222–5231. doi: [10.1109/CVPR42600.2020.00527](https://doi.org/10.1109/CVPR42600.2020.00527).
- [98] B. Ivanovic and M. Pavone, “The Trajectron: Probabilistic Multi-Agent Trajectory Modeling With Dynamic Spatiotemporal Graphs,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 2375–2384. doi: [10.1109/ICCV.2019.00246](https://doi.org/10.1109/ICCV.2019.00246).
- [99] A. Colman, “Cooperation, psychological game theory, and limitations of rationality in social interaction,” *The Behavioral and brain sciences*, vol. 26, pp. 139–53; discussion 153, May 2003, doi: [10.1017/S0140525X03000050](https://doi.org/10.1017/S0140525X03000050).
- [100] A. Bighashdel, P. Meletis, and G. Dubbelman, “Towards Equilibrium-based Interaction Modeling for Pedestrian Path Prediction,” in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, Sep. 2020, pp. 1–8. doi: [10.1109/ITSC45102.2020.9294530](https://doi.org/10.1109/ITSC45102.2020.9294530).
- [101] M. Abdar *et al.*, “A review of uncertainty quantification in deep learning: Techniques, applications and challenges,” *Information Fusion*, vol. 76, pp. 243–297, Dec. 2021, doi: [10.1016/j.inffus.2021.05.008](https://doi.org/10.1016/j.inffus.2021.05.008).

License

Non-exclusive licence to reproduce the thesis and make the thesis public

I, **Ankit Garg**,

1. grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

Machine learning solutions for the task of pedestrian trajectory prediction – a systematic literature review,

supervised by **Dr. Naveed Muhammad** and Co-supervised by **Dmytro Zabolotnii**

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Ankit Garg

13/08/2024