BERGISCHE
UNIVERSITÄT
WUPPERTAL

Evaluation of the use of deep learning algorithms for the analysis of high-density pedestrian dynamics

Dissertation
to obtain a doctoral degree

in the
School of Mechanical Engineering and Safety Engineering
at

University of Wuppertal

Submitted by:
Raphael Adrian Korbmacher

from Hanau

Wuppertal, May 2024

Abstract

Data-based approaches, particularly deep learning algorithms, have attracted considerable attention for their performance in various scientific disciplines and practical applications. This dissertation focuses on evaluating these methods in their application to the study of pedestrian dynamics. Through three consecutive journal articles, the research objectives are systematically elaborated based on four research questions, addressing the paradigm shift in this research field, the effectiveness and relevance of various methodologies for trajectory prediction, and the influence of scene density on performance.

Publication I lays the foundation by providing a comprehensive overview of physics-based models and deep learning algorithms, highlighting the technical and applicational divergences between the two approaches.
Publication II further advances this topic by employing datasets characterized by differing densities to conduct pedestrian trajectory prediction experiments. Here, the effectiveness of different approaches is tested, revealing that while the algorithms can achieve high accuracy in terms of distance metrics, they are weak in collision avoidance, a critical aspect of pedestrian dynamics. By integrating the concept of 'time-to-collision' into the deep learning algorithms, a hybrid approach is proposed.
Publication III builds on these findings with a novel dataset generated during the Lyon Festival of Lights. A two-stage prediction approach is presented that first classifies scenes based on density before making predictions. It is shown how this can significantly improve the accuracy of predictions.

In essence, this dissertation bridges the gap between traditional physics-based models and contemporary deep learning algorithms, advocating a hybrid approach for pedestrian trajectory prediction. Through a careful examination of different methods under varying conditions, it finds that data-based approaches offer promising advances, but the integration of concepts from physics-based models can further improve their effectiveness.

Zusammenfassung

Datenbasierte Ansätze, insbesondere Deep Learning Algorithmen, haben aufgrund ihrer Leistung in verschiedenen wissenschaftlichen Disziplinen und praktischen Anwendungen erhebliche Aufmerksamkeit erregt. Diese Dissertation befasst sich mit der Bewertung dieser Methoden in der Anwendung für die Untersuchung von Fußgängerdynamiken. Durch drei aufeinanderfolgende Fachartikel werden die Forschungsziele systematisch anhand von vier Forschungsfragen ausgearbeitet, die sich mit dem Paradigmenwechsel in diesem Forschungsgebiet, der Wirksamkeit und Relevanz verschiedener Methoden zur Prognose von Trajektorien und dem Einfluss der Szenendichte auf die Performance befassen.

Publikation I legt den Grundstein, indem sie einen umfassenden Überblick über physikbasierte Modelle und Deep Learning Algorithmn bietet und die technischen und anwendungsbezogenen Unterschiede zwischen den beiden Ansätzen aufzeigt.
Publikation II vertieft diese Thematik weiter, indem Datensätze verwendet werden, die sich durch unterschiedliche Dichten auszeichnen, um Experimente zur Prognose von Fußgängertrajektorien durchzuführen. Hier wird die Wirksamkeit der verschiedenen Ansätze überprüft, wobei sich zeigt, dass die Algorithmen zwar in Bezug auf Distanzmetriken eine hohe Genauigkeit erzielen können, jedoch bei der Vermeidung von Kollisionen – einem kritischen Aspekt der Fußgängerdynamik – Schwächen aufweisen. Durch die Integration des Konzepts der 'Zeit-zur-Kollision' in die Algorithmen wird ein hybrider Ansatz vorgeschlagen.
Publikation III baut auf diesen Erkenntnissen auf mit einem neuartigen Datensatz, der während des Lichterfestes in Lyon generiert wurde. Ein zweistufiger Prognoseansatz wird präsentiert, der zuerst Szenen auf Basis der Dichte klassifiziert, bevor Prognosen getroffen werden. Es wird aufgezeigt, wie die Prognosegenauigkeit dadurch erheblich verbessert werden kann.

Im Wesentlichen überbrückt diese Dissertation die Lücke zwischen traditionellen physik-basierten Modellen und neuartigen Deep Learning Algorithmen und befürwortet einen hybriden Ansatz zur Prognose von Trajektorien. Durch eine sorgfältige Untersuchung verschiedener Methoden unter variierenden Bedingungen wird festgestellt, dass datenbasierte Ansätze vielversprechende Fortschritte bieten, die Integration von Konzepten aus der Modellierung jedoch ihre Wirksamkeit noch weiter verbessern kann.

Acknowledgement

As I reflect on the journey of my doctoral studies, my heart is full of gratitude for the numerous individuals who have enriched this experience, making it both profound and transformative. I embarked on this path with a passion for learning and a zeal for solving complex problems on my own, imagining the PhD to be an extension of my Master's thesis but on a grander scale. However, this journey taught me that research transcends individual effort; it thrives on collaboration, discussion, and collective problem-solving.

First and foremost, my sincerest thanks to Prof. Antoine Tordeux, my supervisor and closest collaborator. His fantastic guidance, the endless hours he spent discussing new ideas with me, and his patience in tackling problems have been nothing short of inspirational. Prof. Tordeux has been a cornerstone of my PhD journey, and I owe him immensely for his mentorship.

My experience was further enriched by my involvement in the MADRAS Project, which was not just an academic assignment but a journey into the heart of teamwork and collaboration. Special thanks are due to Benoit Gaudou, Nicolas Verstaevel, and Hu Tuu Dang from Toulouse. The three months spent working closely with these esteemed computer scientists were pivotal, with Hu Tuu Dang deserving particular gratitude. Our collaboration blossomed into a friendship that I cherish deeply.

I am also grateful to Jakob Cordes and Mohcine Chraibi. My numerous travels with Jakob—to Lyon, Paris, and Jülich—were filled with enriching discussions that shaped my thinking and research. Mohcine's creative insights often illuminated new paths and perspectives, significantly enhancing my work. A heartfelt thank you to Oskar Defour and Alexandre Nicolas in Lyon, who were not only outstanding hosts but also played a crucial role in the successful data generation during the Festival of Lights. Their hospitality and support were remarkable.

Last but certainly not least, my deepest appreciation goes to my family— Ingrid Korbmacher, Michael Beier-Korbmacher, Melina Korbmacher, Felix Korbmacher, Elisabeth Korbmacher und Karl-Heinz Korbmacher. Their unwavering support and belief in my abilities have been my foundation and comfort throughout this academic pursuit.

Table of Contents

Preamble

Publications

Appendix

List of Publications

Journal Publications included in this Thesis

I **Raphael Korbmacher**, and Antoine Tordeux. "Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches." *IEEE Transactions on Intelligent Transportation Systems* 23.12 (2022): 24126-24144.

II **Raphael Korbmacher**, Huu-Tu Dang, and Antoine Tordeux. "Predicting pedestrian trajectories at different densities: A multi-criteria empirical analysis." *Physica A: Statistical Mechanics and its Applications* 634 (2024): 129440.

III **Raphael Korbmacher**, and Antoine Tordeux. "Towards Better Predictive Models: The Role of Density in Pedestrian Trajectory Predictions." *Sensors* 24.7 (2024): 2356.

Other Publications

iv Basma Khelfa, **Raphael Korbmacher**, Andreas Schadschneider, and Antoine Tordeux "Heterogeneity-induced lane and band formation in self-driven particle systems." *Scientific reports* 12.1 (2022): 4768.

v Basma Khelfa, **Raphael Korbmacher**, Andreas Schadschneider, and Antoine Tordeux "Initiating lane and band formation in heterogeneous pedestrian dynamics." *Collective Dynamics* 6 (2021): 1-13.

vi Huu-Tu Dang, **Raphael Korbmacher**, Antoine Tordeux,Benoit Gaudou, & Nicolas Verstaevel, "TTC-SLSTM: Human trajectory prediction using time-to-collision interaction energy." 15th International Conference on Knowledge and Systems Engineering (KSE). *IEEE*, 2023.

vii **Raphael Korbmacher**, Huu-Tu Dang, Antoine Tordeux, Benoit Gaudou, & Nicolas Verstaevel. "Differences in pedestrian trajectory predictions for high-and low-density situations." 14th International Conference On Traffic And Granular Flow (TGF) *Springer*, 2022.

viii **Raphael Korbmacher**, Alexandre Nicolas, Antoine Tordeux, & Claudia Totzeck. "Time-continuous microscopic pedestrian models: an overview." *Crowd Dynamics, Volume 4: Analytics and Human Factors in Crowd Modeling* (2023): 55-80.

ix **Raphael Korbmacher**, and Antoine Tordeux. "Deep Learning for Predicting Pedestrian Trajectories in Crowds." *Intelligent Systems Conference*. Cham: Springer Nature Switzerland, 2023.

Talks at conferences

x Review of Pedestrian Trajectory Prediction methods: Comparing physics-based and data-based approaches. 17ième Journées de la Matière Condensée (JMC17), 24-27 August 2021, Online, Rennes, France

xi Review of Pedestrian Trajectory Prediction methods: Comparing physics-based and data-based approaches. 10th International Conference on Pedestrian and Evacuation Dynamics (PED2021). 28-30 November 2021, Online, Melbourne, Australia.

xii Differences in pedestrian trajectory predictions for high- and low-density situations. Traffic and Granular Flow 2022 (TGF22), 15–17 October 2022, online, IIT Delhi, India.

xiii Using synthetic data to improve performance of data-driven algorithms in high density pedestrian situations. 2nd GAMA Days Conference, 22-24 June 2022, online, France

xiv Poster: Using time-to-collision in the loss function of deep learning algorithm to improve pedestrian trajectory predictions. Pedestrian and Evacuation Dynamics 2023 Conference (PED23), 27-30 June 2023, Netherlands

Awards Runner-up paper award at 15th International Conference on Knowledge and Systems Engineering 2023 for the work "TTC-SLSTM: Human trajectory prediction using time-to-collision interaction energy."

Author's Contributions

Publication I

- Conceptualization: Raphael Korbmacher, Antoine Tordeux
- Methodology: Raphael Korbmacher, Antoine Tordeux
- Validation: Raphael Korbmacher
- Formal Analysis: Raphael Korbmacher, Antoine Tordeux
- Visualization: Raphael Korbmacher
- Writing – Original Draft Preparation: Raphael Korbmacher
- Writing – Review and Editing: Raphael Korbmacher, Antoine Tordeux

Publication II

- Conceptualization: Raphael Korbmacher, Antoine Tordeux
- Methodology: Raphael Korbmacher, Antoine Tordeux
- Software: Raphael Korbmacher, Tu Dang-Huu
- Validation: Raphael Korbmacher
- Visualization: Raphael Korbmacher
- Formal Analysis: Raphael Korbmacher
- Data Curation: Raphael Korbmacher, Tu Dang-Huu
- Writing – Original Draft Preparation: Raphael Korbmacher
- Writing – Review and Editing: Raphael Korbmacher, Antoine Tordeux, Tu Dang-Huu

Publication III

- Conceptualization: Raphael Korbmacher, Antoine Tordeux
- Methodology: Raphael Korbmacher, Antoine Tordeux

- Software: Raphael Korbmacher
- Validation: Raphael Korbmacher
- Visualization: Raphael Korbmacher
- Formal Analysis: Raphael Korbmacher
- Data Curation: Raphael Korbmacher
- Writing – Original Draft Preparation: Raphael Korbmacher
- Writing – Review and Editing: Raphael Korbmacher, Antoine Tordeux

Preamble

1. Introduction

Within the framework of the Franco-German project MADRAS (Multi-agent modelling of dense crowd dynamics: Predict & Understand), this cumulative doctoral thesis investigates the application of deep learning algorithms for the analysis of high-density pedestrian dynamics. This thesis consists of three consecutive journal publications, which can be found in subsequent chapters. These publications collectively examine the methodologies behind pedestrian trajectory predictions, offering a comprehensive analysis of how deep learning can be leveraged to predict the complex and multifaceted behaviors of pedestrians.

1.1 Motivation

The number of people living in urban areas has been rapidly increasing for decades. Particularly in developing countries, the urban population is increasing so rapidly that local infrastructures are overloaded. This leads to more situations where there are so many people in one place that they can be considered as crowds or high-density situations (2-8 ped/m^2). These scenarios, ranging from everyday urban congestion to large-scale events and emergency evacuations, necessitate a meticulous study of pedestrian behavior to ensure public safety and effective infrastructure planning. Traditionally, scientists have leaned on physics-based (PB) models to simulate and study these behaviors, providing a framework to predict and manage crowd dynamics safely. However, the advent of autonomous technologies, such as vehicles and service robots, has cast pedestrian behavior analysis in a new light. The seamless integration of these technologies into urban life requires predictions of pedestrian movements, necessitating methods that can adapt to the fluidity and unpredictability of human behavior. Herein lies the pivot to data-based algorithms, particularly those grounded in deep learning (DL). Unlike PB models, these algorithms thrive on vast datasets, learning and predicting pedestrian dynamics without explicit theoretical underpinnings. This dissertation seeks to explore this dichotomy. By evaluating the application of DL algorithms in predicting trajectories in high-density situations, this work aims to not only benchmark their performance against PB models but also to investigate how a synergy between DL algorithms and PB models might unlock new potentials in pedestrian trajectory prediction.

1.2 Objectives and Scope

The topic of this dissertation is the evaluation of DL approaches within the realm of pedestrian dynamics. Given the inherent complexity of this subject, it necessitates a multifaceted approach, integrating various perspectives and innovative concepts. A common strategy for dissecting complex issues involves breaking them down into more manageable questions, thereby facilitating a deeper understanding of the individual components. This thesis is structured around four principal research questions (RQ), which are explored through three distinct publications:

- RQ1: Is there a paradigm shift occurring in the field of pedestrian dynamics?
- RQ2: Among the available methodologies, which is most effective for predicting pedestrian trajectories?
- RQ3: To what extent does the density of a scene influence performance? Furthermore, which approach excels in scenarios of high density?
- RQ4: In the field of pedestrian trajectory prediction, will PB models remain relevant, or will DL emerge as the universal answer?

Table 1.1 Scope of Publications I-III.

	RQ1	RQ2	RQ3	RQ4
Publication I	X			X
Publication II		X	X	X
Publication III			X	X

Publication I offers a comprehensive literature review of PB models and DL algorithms, highlighting technical and application differences. This comparative analysis provides insights into RQ1 and RQ4. Publication II employs datasets with varied characteristics to conduct experiments on pedestrian trajectory prediction using both PB models and DL algorithms. This allows making statements about the effectiveness of the methods for pedestrian trajectory prediction, which is the content of RQ2. A significant focus of this publication is on the impact of scene density on prediction accuracy. By assessing performance across a spectrum of densities, we provide insights into RQ3. A key finding is that while DL algorithms excel in minimizing distance-based error metrics, they underperform in collision avoidance, especially at high densities, compared to PB models. Therefore, we integrate the 'time-to-collision' (TTC) concept from the PB models [8] into the loss function of DL algorithms. This innovative hybrid model serves as a preliminary affirmation that PB models could maintain their relevance, offering a nuanced perspective on RQ4. In publication III replicates the methodology from Publication II but leverages a novel dataset collected during the Festival of Lights in Lyon. This dataset is characterized by its rapidly fluctuating densities, making it particularly suitable for exploring RQ3. We introduce a two-stage prediction method, initially classifying scenes before proceeding with trajectory predictions. This approach is further enhanced by incorporating the TTC-based algorithm from Publication II, demonstrating once again how integrating PB concepts can augment DL algorithms. This methodology not only bolsters the prediction accuracy, but also provides additional evidence in response to RQ4, showcasing the potential symbiosis between PB and DL models in the field of pedestrian dynamics.

1.3 Organization of the Dissertation

In the following chapter 2 the necessary theoretical framework is provided. We introduce preliminary concepts and explain the main ideas of the PB models and DL algorithms. In chapter 3 the contributions of the publications are summarised. Furthermore, we illustrating how they collectively help to address RQ1-4. In the last chapter 4 we discuss the RQs and provide a outlook for future works. Following this, the full texts of the three publications are provided.

2. Theoretical Framework

In this chapter, we present a detailed overview of the theoretical framework crucial for comprehending the methodologies and objectives of this dissertation. The three publications included within this work are united by their focus on pedestrian trajectory predictions. Consequently, the theoretical foundation is specifically tailored to address this challenge, providing a comprehensive understanding of the key concepts and methodologies involved in this challenge.

2.1 Preliminary Concepts

Trajectory A trajectory reefers to the time-profile of the pedestrian's position. Mathematically, it is represented by the image coordinates (x_t^i, y_t^i) for each time instant $t = k \cdot dt$, where $k \in \mathbb{N}$. Typically, we predict trajectories over 12 time steps k, equating to 4 seconds in duration at a frame rate of three observations per second.

Scene The trajectories of pedestrians are significantly influenced by the surrounding pedestrians as well as objects in the environment. In prediction tasks, it is essential to anticipate the movements of all i^{th} pedestrians within our area of interest at time t, which we call a scene.

Primary pedestrian In any given scene with n pedestrians, one pedestrian is designated as the primary pedestrian i for analysis. We predict the trajectories for the entire scene and then calculate the prediction error specifically for the primary pedestrian. Subsequently, the process is repeated by selecting a new primary pedestrian, $i + 1$, within the same scene to assess their trajectory prediction error. This iterative method is continued until the trajectory predictions for all pedestrians, up to $i = n$, have been comprehensively evaluated.

ADE/FDE In the realm of supervised deep learning, the evaluation process is crucial. Evaluating pedestrian trajectory prediction is primarily conducted using two Euclidean distance metrics.

The first, Average Displacement Error (ADE) [12], captures the deviation between the predicted and actual trajectories at each time step t

$$\textbf{ADE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\hat{x}_i(t) - x_i(t)\|. \tag{2.1}$$

Where $x_i(t)$ denotes the true position of the i^{th} pedestrian at time t, while $\hat{x}_i(t)$ represents the predicted position. The notation $|\cdot|$ signifies the Euclidean distance. The second metric, the Final Displacement Error (FDE) [11], measures the discrepancy between the final points $t = T_{pred}$ of the predicted and ground truth trajectories:

$$\textbf{FDE} = \frac{1}{N}\sum_{i=1}^{N}\|\hat{x}_i(T) - x_i(T)\|. \tag{2.2}$$

Both distance-based metrics are widely used in pedestrian trajectory predictions for their effectiveness in quantifying the goodness-of-fit.

Collision metric While the aforementioned metrics effectively quantify fit quality, they overlook the repulsive forces critical in pedestrian interactions, potentially missing overlaps or collisions. To address this, the collision metric is employed, enhancing the evaluation process

$$\textbf{COL} = \frac{1}{|S|}\sum_{\hat{X}\in S} COL(\hat{X}), \tag{2.3}$$

with

$$COL(\hat{X}) = \min\left(1, \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j>i}^{N}\big[||\hat{x}_i(t) - \hat{x}_j(t)|| \leq 2R\big]\right). \tag{2.4}$$

In these Equation, S represents all scenes in the test set, and $\hat{X}$ denotes a scene's prediction containing N agents. The Iverson bracket $[\cdot]$ indicates that the expression within it evaluates to 1 if true, and 0 otherwise

$$[P] = \begin{cases} 1 & \text{if } P \text{ is true,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

This metric considers a prediction to collide when two predicted pedestrian trajectories intersect, thereby reflecting the proportion of predictions where collisions are predicted. The pedestrian radius $R = 0.2$ m is a critical parameter in this calculation, influencing the detected collision count.

2.2 Physics-based Approach

The physics-based approach comprises a large number of models ranging from macroscopic, mesoscopic, and microscopic model [4]. Due to the fact, that we predict individual trajectory of one primary pedestrians, the microscopic approach is relevant for our task. In the class of microscopic models we differentiate between force-based, velocity-based, and decision-based models. In the following the concepts of force-based and velocity-based models are introduced.

Force-based models The fundamental premise of a force-based models is that the pedestrian behavior is governed by forced attaching him from his surroundings [9]

$$\frac{d}{dt}x_i = v_i, \qquad i = 1, \ldots, N, \qquad\qquad x_i(0) = x_{0,i}, \tag{2.6a}$$

$$\frac{d}{dt}v_i = D_i(t, x_i, v_i) + I_i(t, x, v) + B(t, x_i, v_i), \qquad v_i(0) = v_{0,i}. \tag{2.6b}$$

In Eq. 2.6b, three types of forces impact the pedestrian's behavior. The force B describes environmental interactions, like with walls or obstacles. The drift term D_i generally models relaxation of the velocity v_i towards a desired velocity u_i [7, 3]

$$D_i(t, x_i, v_i) = \tau(u_i(x_i) - v_i). \tag{2.7}$$

The main modelling component of force-based models is the interaction term I_i which defines social forces derived from proximity concepts and behaviors aimed at avoiding collisions

$$I_i(t, x, v) = -\sum_{j=1}^{N} \nabla U(x_j - x_i)\, \omega(\phi_{ij}) \tag{2.8}$$

where U represents the repulsive potential, and ω the anisotropic factor depending for example on the relative bearing angle with the neighbors ϕ_{ij}. A famous force-based model, which we use in the later follwoing experiments is the social-force model from Helbing and Molnár [7]. In it the functions reads

$$U(x) = AB\exp(-|x - \ell|/B) \qquad \text{and} \qquad \omega(x) = \begin{cases} 1 & \text{if } |\phi_{ij}| < \kappa \\ c & \text{otherwise} \end{cases}$$

where κ is the vision cone angle, and $0 < 1 < c$ is a reduced perception factor. The parameters A and B are the repulsion rate and distance, respectively.

Velocity-based models Velocity-based models are systems of first-order differential equations. They are given by

$$v_i = V_i(t, x, v), \qquad v_i(0) = v_{0,i} \tag{2.9}$$

with the velocity function $V_i \in \mathbb{R}^d$, or by

$$v_i = \omega_i(t, x, v)\, e_i(t, x, v), \qquad v_I(0) = v_{0,i} \tag{2.10}$$

where the scalar speed model $\omega_i(t, x, v) \in \mathbb{R}$ is distinguished from the pedestrian direction model $e_i(t, x, v) \in \mathbb{R}^d$. Pioneer works in the field of velocity-based models date back to the end of the 1990s and the work of Fiorini et al. with velocity obstacle and velocity avoidance sets [5]. The approach consists of linearly extrapolating the trajectories of the pedestrians to determine collision sets. These sets describe the so-called *collision cones*. The pedestrian model is obtained by minimizing the deviation from the desired velocity outside the collision cones

$$v_i = \underset{v \notin \cup_{j \neq i} \mathrm{VO}_{ij}}{\arg\min} \|v - u_i\|^2 \tag{2.11}$$

with u_i the desired velocity and VO_{ij} the collision cone of the i-th pedestrian with the j-th neighboring pedestrian. In the following work we especially use the so called optimal reciprocal collision avoidance model (ORCA) from Van den Berg et al. [13] that extends the collision-free dynamics to a general framework of agents acting independently without communicating with each other.

2.3 Data-based Approach

The foregoing approach heavily relies on a theoretical modeling framework. Pre-identified key mechanisms were formulated in mathematical equations with a few meaningful parameters that have to be calibrated and validated. After that, the model can be used to simulate the scene and predict future trajectories based on current position and velocity of the pedestrians [10].

The DL algorithms do not need any assumptions regarding behavioral mechanisms, obviating the need for prior knowledge and depending solely on data. Due to the fact that the temporal sequence of positions plays a decisive role in trajectories, algorithms are required that can process time series data. Long Short-Term Memory Networks (LSTM) and Generative Adversarial Networks (GAN) stand out for their ability to process time series data and have shown to be capable of predicting realistic pedestrian movement patterns [2, 6].

In the scholarly discourse, a prevalent framework is employed for training algorithms to learn pedestrian behaviors. During the training and evaluation phases, individual trajectories are segmented into input and output sequences. Specifically, the algorithm is provided with $k = 9$ time steps of the input trajectory, encompassing both the primary pedestrian and their immediate neighbors, and is tasked with predicting the subsequent $k = 12$ time steps. Subsequently, this predicted output is compared to the ground truth data to assess the model's accuracy, utilizing error metrics predominantly in the form of ADE and FDE, as detailed in Section 2.1. The primary objective during the training phase is to refine the algorithms such that the Euclidean distance between the predicted trajectory and the ground truth is minimized, thereby enhancing the precision of the predictions.

Long short-term memory networks A recurrent neural network (RNN) is an important class of machine learning methods, that uses feedback connections to store representations of recent input events in form of activations [14]. These feedback connections can exist between different time steps providing a temporal memory to the network [139]. Because of this capability, they are especially suited for sequence modelling tasks such as time series prediction and sequence labelling tasks [9]. Most successful are the LSTM architectures of RNNs, that use purpose-built memory cells to store information. LSTM networks can be trained for sequence generation by processing real data sequences one step at a time and predicting what comes next so that temporal coherence can be ensured. The training of a LSTM network can be lengthy and difficult, but the forward-propagation of new data (testing) can happen quite fast. The Most famous LSTM for pedestrian trajectory prediction is the Social-LSTM (S-LSTM) proposed by Alahi et al. [5]. It is the first major DL application that takes the interactions between pedestrians into account. The main contribution of the S-LSTM is a novel pooling layer called social pooling that gathers the hidden states of nearby pedestrians (see Fig. 2.1). With that technique, the influence of neighboring pedestrians on the movement of the ego pedestrian can be included as an input for the following prediction step.

In Fg. 2.1 the core idea of the S-LSTM is displayed. For each pedestrian in a scene a separate LSTM network with separated input is used. Through the social pooling layer (S-pooling) the LSTM are connected with each other allowing to share information.

Generative adversarial networks GANs are DL algorithms that are characterised by an architecture containing a discriminator and a generator. The generator takes the training data as input and creates different output samples. In turn, the discriminator takes the generated samples and the training data and tries to distinguish whether a given sample belongs to the true data distribution or is generated by the generator. Both components are engaged in a competition similar to a two-

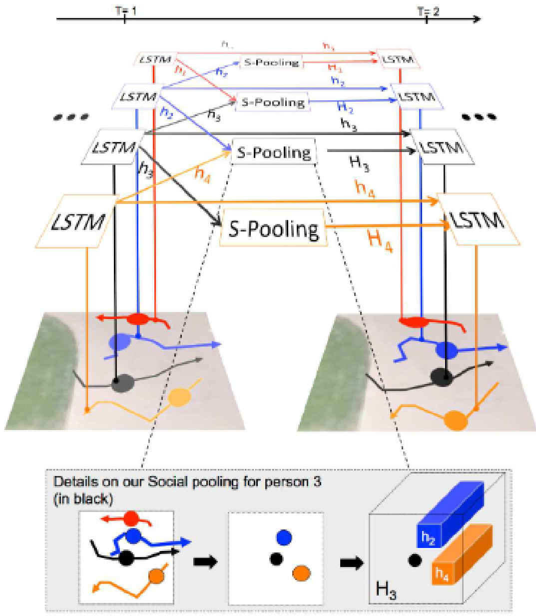


Figure 2.1 Incorpatation of neighbour trajectories by using a social pooling layer in the S-LSTM from Alahi et al. [2].

player min-max game where each one tries to outsmart the other one [76, 136]. From this process, the generator learns to generate data that resemble the true data distribution, which allows GANs to generate multimodal prediction samples.
The most famous GAN for pedestrian trajectory prediction is the Social-GAN (S-GAN) from Gupta et al. [6]. In this algorithm the generator is composed of an LSTM-based encoder, a context pooling module, and an LSTM-based decoder. The discriminator uses LSTMs as well as shown in Fg. 2.2. The pooling module adopted in the S-GAN uses a uniform weight for all surrounding pedestrians.

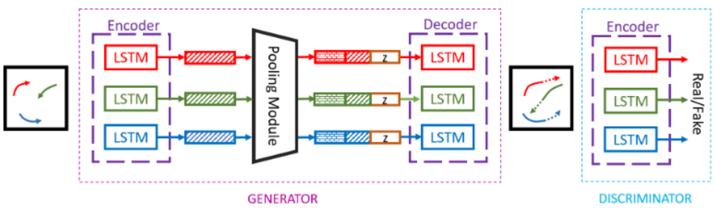


Figure 2.2 Architecture of the S-GAN with its three key components Generator, Pooling Module, and Discriminator [6].

3. Contributions of the Publications

In this section, we will outline the contributions made by the publications that constitute this cumulative dissertation. The sequence of these publications is deliberately structured to build upon each other, creating a coherent thread throughout the dissertation. The initial publication provides a thorough literature review on the DL algorithms and the PB models, exploring both technical and application differences between the two approaches. In the publication II, we extend beyond a qualitative comparison of these methodologies to a quantitative analysis, utilizing both low-density field data and high-density experimental data. This comparison is facilitated by the introduction of three distinct error metrics. Concluding this paper, we introduce an innovative algorithm that surpasses existing DL algorithms in performance. This algorithm is a hybrid model, incorporating the architectural strengths of DL algorithms while integrating TTC a concept derived from PB models as discussed by Karamouzas [8]—to simulate pedestrian behavior more accurately. The last publication continues the exploration begun in the second, albeit with a shift in methodology: instead of relying on experimental data for high-density situations, we utilize field data. Given the scarcity of high-density field data, the research team undertook field experiments during the Festival of Lights in Lyon, generating valuable high-density pedestrian trajectory data. In this publication III, we leverage this unique dataset to validate our experiments, which demonstrate that a prior density-based classification significantly enhances prediction accuracy. We introduce a two-stage TTC-SLSTM algorithm that exhibits exceptional performance by both Euclidean and collision-based error metrics, marking a significant advancement in pedestrian behavior prediction.

3.1 Publication I: Review of Pedestrian Trajectory Prediction Methods

Publication I serves as a comprehensive review paper, primarily aimed at presenting a detailed overview of pedestrian trajectory prediction methods. A pivotal element of this study is the systematic differentiation between DL algorithms and PB models as prediction tools. Beyond a thorough literature review encompassing both approaches, the paper delves into the technical nuances and application-specific distinctions that set these approaches apart. In its concluding sections, the paper addresses unresolved questions and outlines potential avenues for future research, signalling the evolving nature of this field. A notable observation is the divergent perspectives, terminologies, and objectives prevalent among machine learning researchers who employ DL algorithms and the traditional pedestrian dynamics field. This disparity underscores the necessity for a unified framework and lexicon that can bridge the gap between these communities. Consequently, this paper endeavors to serve as a foundational resource, offering standardized terminologies and frameworks to foster mutual understanding and collaboration across the disciplines. Publication I help to answer research question RQ1 and RQ4. RQ1 examining whether a paradigm shift is occurring within pedestrian dynamics studies. Figure 3.1 from publication I indicates that DL algorithms are indeed a hot topic in the last decade.

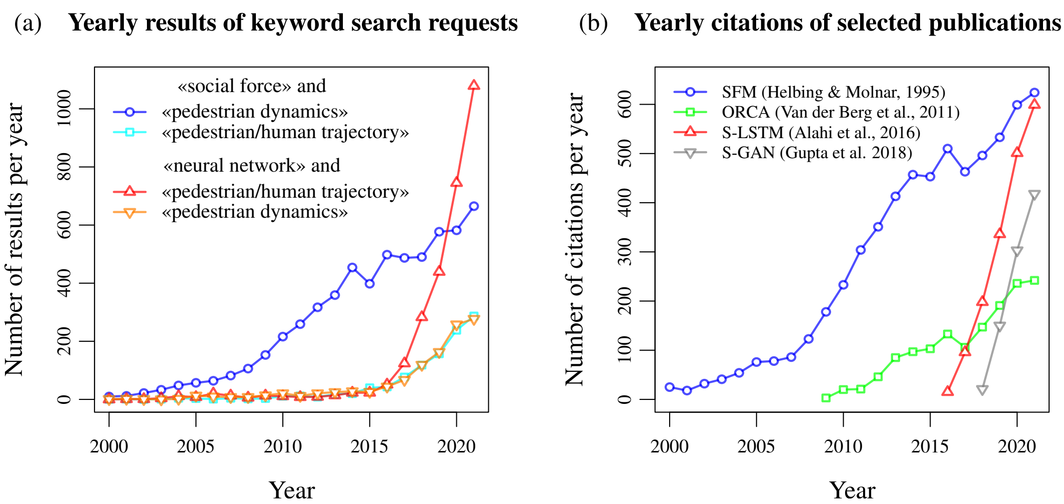


Figure 3.1 Number of annual citations estimated by performing an online search with the engine Google Scholar [1]. Panel (a): Results of the tags «neural network» and «social force» in combination with either «pedestrian/human trajectory» or «pedestrian dynamics». Panel (b): Yearly citations for the articles by Helbing and Molnár [7] (SFM), Van den Berg et al. [13] (ORCA), Alahi et al. [2] (S-LSTM), and Gupta et al. [6] (S-GAN). SFM and ORCA are PB models, whereas S-LSTM and S-GAN are DL approaches.

Notably, the S-LSTM introduced by Alahi et al. [2] and the S-GAN by Gupta et al. [6] stand out as particularly influential contributions. Despite their relatively recent publication, these works have garnered a considerable number of citations annually, underscoring their impact and the keen interest they have sparked in the research community.

Regarding RQ4, our findings in publication I indicate that the current applications of the two approaches are different. While DL algorithms demonstrate superior performance over PB models in predicting individual trajectories at low densities, their efficacy in simulating large-scale events and capturing collective dynamics is yet to be established.

3.2 Publication II: Predicting Pedestrian Trajectories at Different Densities

After finishing the comprehensive literature review on both approaches we wanted to dive into empirical experiments. We employed both low-density field datasets and high-density experimental datasets to assess the performances. At the initial step of our evaluation, we selected an appropriate error metrics. Commonly, ADE and FDE, which are based on Euclidean distance, are utilized for this purpose. Our comparative analysis revealed that, in terms of ADE/FDE accuracy, DL algorithms consistently outperformed PB models across all datasets. To deepen our investigation, we incorporated an additional metric designed to quantify the frequency of collisions or overlaps in the predicted trajectories. Interestingly, despite the superior ADE/FDE performance of DL algorithms, a significant increase in collision instances was observed, particularly in scenarios characterized by high density. To address this issue we developed a novel continuous collision metric based on pedestrians' time-to-collision and added it to the loss function of the DL algorithms. By doing so the algorithms are not just trained on minimizing euclidean distance, but also on reducing the likelihood of collisions with neigbours. This method, which is refer to as the hybrid approach, synthesizes insights from PB models with DL algorithm frameworks to enhance performance. The enhanced loss function is articulated as follows:

$$\mathbf{L}_i = \sum_{t=1}^{T} \|x_i(t) - \hat{x}_i(t)\|^2 + \lambda \sum_{t=1}^{T} f(\min_{j \neq i}\{\tau_{ij}\}), \tag{3.1}$$

where the first term represents the ADE and the second term encapsulates our collision-based metric. The collision metric employs a sigmoid penalty function f, defined as:

$$f(\tau) = \frac{1}{1 + e^{s(\tau - \delta)}}, \tag{3.2}$$

with s and δ denoting the slope and threshold parameters, respectively. The TTC between entities i and j is calculated using:

$$\tau_{ij} = \frac{-x_{ij} \cdot v_{ij} - \sqrt{(x_{ij} \cdot v_{ij})^2 - ||v_{ij}||^2(||x_{ij}||^2 - 4R^2)}}{||v_{ij}||^2}, \tag{3.3}$$

A collision between i-th and j-th pedestrians occurs if there exists a time $\tau > 0$ such that $x_{ij} + v_{ij}\tau$ lies within a circle centred at $(0, 0)$ with radius $2R$. This criterion is mathematically represented as $\|x_{ij} + v_{ij}t\| < 2R$ where $\| \cdot \|$ denotes Euclidean norm.

In Publication II, we delve into RQ2, RQ3, and RQ4, offering substantial insights into each. Our analysis demonstrates that, regarding prediction accuracy measured by distance-based metrics ADE/FDE, DL algorithms consistently surpass PB models (RQ2). However, in high-density situations, the performance gap between DL algorithms and PB models narrows, and DL algorithms exhibit a significant increase in collision occurrences, indicating a potential area of weakness (RQ3). To address that, we integrated the TTC concept from PB models in the loss function of the DL algorithms (RQ4).

3.3 Publication III: Toward Better Pedestrian Trajectory Predictions

In publication III, we aim to build upon the concepts introduced in publication II, further refining these ideas and applying them to our novel dataset. Our dataset, originating from a field study conducted by our team during the Festival of Lights in Lyon, stands out as, to our knowledge, the only high-density pedestrian trajectory field dataset capturing long trajectories in a non-artificial setting. This dataset is distinguished by its dynamic density fluctuations: during the light shows, the tracking area experiences relatively low density, featuring multidirectional flow with long interaction ranges. However, once a show concludes, the density surges dramatically as pedestrians move en masse to the next viewing spot, with densities reaching beyond 2.2 ped/m^2 in a mainly unidirectional flow (see appendix of publication III). Preliminary data analysis revealed that cluster algorithms such as K-Means and Agglomerative Hierarchical Clustering (AHC) effectively identify distinct clusters within the data, as illustrated in Figure 3.2.

Figure 3.2 demonstrates that the density is a characteristic, that explains the behaviour patterns to a certain degree. This knowledge will be used to enhancement of our DL algorithms. We introduce a novel 2-stage approach, where initially, the density of the scene is calculate before proceeding with trajectory predictions. This methodology is depicted in Figure 3.3, illustrating our predictive

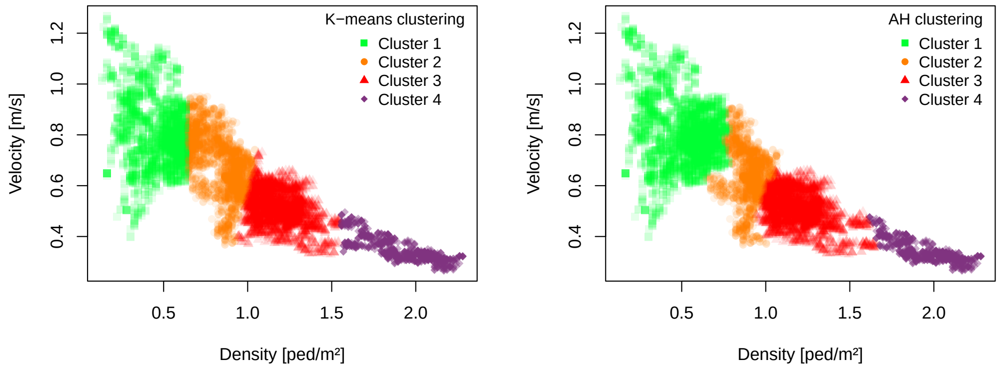


Figure 3.2 Results of the K-Means and the agglomerative hierarchical clustering. Trajectory scenes are clustered as shown by the different colours of the points.

schema.

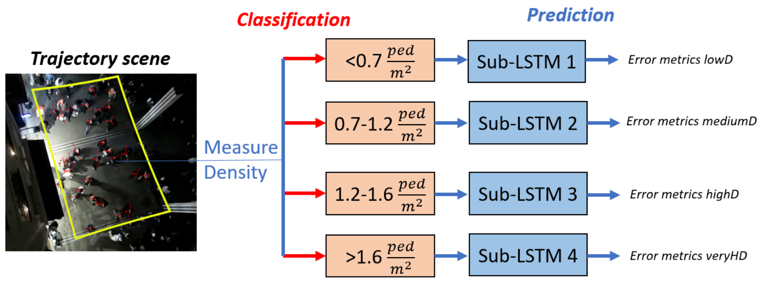


Figure 3.3 Schemata of our two-stage prediction approach

By implementing this schema, we have improved the accuracy of predictions for both S-LSTM and the S-GAN. Moreover, integrating this hybrid approach with the TTC loss function concept from publication II has resulted in an algorithm that exhibits superior performance in terms of ADE/FDE and the collision metric.

This publication contributes significantly to addressing RQ3. Figure 3.2 validates the premise that scene density is a pivotal characteristic, underscoring that differentiating between densities enhances the algorithm's adaptability and prediction accuracy. Additionally, our results reaffirm that integrating insights from PB models with DL algorithms yields improvements. Specifically, in publication III, the incorporation of TTC and density differentiation has proven advantageous, leading to the conclusion that a purely DL approach may not be as effective.

4. Discussion

This dissertation embarked on an explorative journey through the evolving landscape of pedestrian trajectory prediction, with a particular focus on the applicability and efficacy of DL algorithms in comparison to traditional PB models. Through a meticulous examination spread across three journal publications, this work has sought to address four pivotal research questions (RQs), weaving a narrative that not only delves into the potential paradigm shift in pedestrian dynamics but also scrutinizes the capabilities of DL and PB models in predicting pedestrian trajectories across varying densities.

4.1 Contributions to Research Questions

4.1.1 RQ1: The Paradigm Shift

The year 2022 marked a pivotal moment on pedestrian dynamics researchers, as the S-LSTM surpassed the SF Model, the historically most cited work in the field, in annual citations. This shift underscores a growing interest for DL algorithms, signalling a paradigmatic transition. In our publication I we discussed this phenomena and emphasised, that it is caused by a shift in the application of the approaches from simulation of crowds to prediction of single trajectories.

4.1.2 RQ2: Predicting Pedestrian Trajectories

PB models heavily rely on a theoretical modelling framework with pre-identified key mechanisms that were formulated in equations with a few meaningful parameters. DL algorithms do not have interpretable parameters, but extensive arrays of coefficients, enabling them to fit a wide spectrum of data. Given this difference, it is unsurprising that DL algorithm generally outperform PB models in predicting trajectories on a specific dataset they have been trained on. However, choosing the right evaluation system is crucial question for assess predicted pedestrian behaviour. In publication II, we show that reliance on distance metrics alone may result in unrealistic behaviour, such as overlapping, especially in high-density situations. To address that, we introduced a novel TTC based evaluation metric and enhanced the realism and accuracy of the DL algorithm by incorporating this metric into the loss function of the algorithm.

4.1.3 RQ3: Influence of Density on Performance

Our findings in publication II clearly indicate that the density of the scene significantly impacts the effectiveness of both DL and PB approaches. At high densities, the superiority of DL algorithms over PB models is marginal in terms of distance metrics and substantially inferior concerning collision metrics. In publication III we conducted a cluster analysis of our data, that showed that the observed behaviour differences can be mainly explained by the density of the scene. For that reason, we proposed a two-stage framework that initially classifies scenes by density before making predictions, thereby improving the adaptability of our models across various density scenarios.

4.1.4 RQ4: The Future of Pedestrian Dynamics Models

The future of pedestrian dynamics does not envisage the obsolescence of PB models in favour of DL algorithms. Particularly in high-density scenarios and crowd simulations, PB models remain invaluable tools. Our research demonstrates that integrating the strengths of both models can yield superior results for pedestrian trajectory prediction. The synthesis of PB principles with DL innovations, as evidenced in our hybrid models, underscores a synergistic approach that capitalizes on the unique advantages of each, suggesting a balanced and multifaceted future for pedestrian dynamics modelling.

4.2 Outlook

Looking ahead, several avenues for future research emerge from this work. Firstly, the development of more sophisticated hybrid models that seamlessly integrate the predictive strengths of DL algorithms with the nuanced collision avoidance capabilities of PB models could further refine pedestrian trajectory prediction. We used TTC concept in the loss function, but future research should also explore alternative loss functions, particularly for high-density scenarios, where the traditional ADE based approaches may not suffice. Another promising research avenue involves the creation and deployment of specialized algorithms designed for specific scene characteristics. Although this we emphasized density as a critical factor, other scene attributes could also significantly influence model performance. Identifying and incorporating these attributes could lead to more nuanced and effective predictive algorithms. Lastly, our framework focused on short-term predictions at the operational level. Future works needs to overcome this narrow focus and move to the tactical level, where new challenges such as intricate environments, multiple potential pathways, and the intricacies of group dynamics are prevalent. If the DL algorithms are capable to tackle those challenges remains to be demonstrated.

Bibliography

[1] Google Scholar. Accessed: 2021-12-31.

[2] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social LSTM: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 961–971, 2016.

[3] Mohcine Chraibi, Ulrich Kemloh, Andreas Schadschneider, and Armin Seyfried. Force-based models of pedestrian dynamics. *Netw. Heterog. Media*, 6(3):425, 2011.

[4] Mohcine Chraibi, Antoine Tordeux, Andreas Schadschneider, and Armin Seyfried. Modelling of pedestrian and evacuation dynamics. *Encyclopedia of complexity and systems science*, pages 1–22, 2018.

[5] Paolo Fiorini and Zvi Shiller. Motion planning in dynamic environments using velocity obstacles. *The international journal of robotics research*, 17(7):760–772, 1998.

[6] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2255–2264, 2018.

[7] D. Helbing and P. Molnár. Social force model for pedestrian dynamics. *Phys. Rev. E*, 51(5):4282–4286, 1995.

[8] Ioannis Karamouzas, Brian Skinner, and Stephen J Guy. Universal power law governing pedestrian interactions. *Physical review letters*, 113(23):238701, 2014.

[9] Raphael Korbmacher, Alexandre Nicolas, Antoine Tordeux, and Claudia Totzeck. Time-continuous microscopic pedestrian models: an overview. *Crowd Dynamics, Volume 4: Analytics and Human Factors in Crowd Modeling*, pages 55–80, 2023.

[10] Raphael Korbmacher and Antoine Tordeux. Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches. *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[11] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by example. In *Computer graphics forum*, volume 26(3), pages 655–664. Wiley Online Library, 2007.

[12] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th international conference on computer vision*, pages 261–268. IEEE, 2009.

[13] J. van den Berg, S. Guy, M. Lin, and D. Manocha. Reciprocal n-body collision avoidance. In *Robotics Research: The 14th International Symposium ISRR*, pages 3–19, Berlin, Heidelberg, 2011. Springer.

Publications

I. Review of Pedestrian Trajectory Prediction Methods: Comparing Deep Learning and Knowledge-based Approaches

Outline

Bibliographic Information

Korbmacher, R., & Tordeux, A. (2022). Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches. IEEE Transactions on Intelligent Transportation Systems, 23(12), 24126-24144. https://ieeexplore.ieee.org/.

Author's contribution

Conceptualization, R.K. and A.T.; Methodology, R.K. and A.T.; Validation, R.K.; Formal Analysis, R.K.; Visualization, R.K.; Writing – Original Draft Preparation, R.K.; Writing – Review and Editing, R.K. and A.T.

Copyright Notice

Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches.

Raphael Korbmacher[1], and Antoine Tordeux[1]

[1]School of Mechanical Engineering and Safety Engineering, University of Wuppertal, Gaußstr. 20, Germany

Abstract

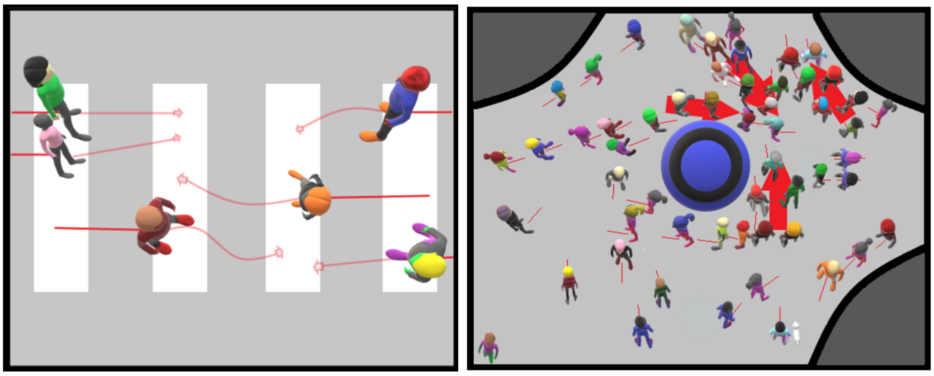
In crowd scenarios, predicting trajectories of pedestrians is a complex and challenging task depending on many external factors. The topology of the scene and the interactions between the pedestrians are just some of them. Due to advancements in data-science and data collection technologies deep learning methods have recently become a research hotspot in numerous domains. Therefore, it is not surprising that more and more researchers apply these methods to predict trajectories of pedestrians. This paper compares these relatively new deep learning algorithms with classical knowledge-based models that are widely used to simulate pedestrian dynamics. It provides a comprehensive literature review of both approaches, explores technical and application oriented differences, and addresses open questions as well as future development directions. Our investigations point out that the pertinence of knowledge-based models to predict local trajectories is nowadays questionable because of the high accuracy of the deep learning algorithms. Nevertheless, the ability of deep-learning algorithms for large-scale simulation and the description of collective dynamics remains to be demonstrated. Furthermore, the comparison shows that the combination of both approaches (the hybrid approach) seems to be promising to overcome disadvantages like the missing explainability of the deep learning approach.

Keywords: Pedestrian trajectory prediction, deep learning, high-density, collision, time-to-collision

I.1 Introduction

The prediction of future trajectories of pedestrians is a valuable but challenging task. Traditionally, researchers try to model and understand human behavior with simple rules and mechanisms that can be used to simulate realistic behavior and predict future trajectories [91, 25]. In the last decade, deep learning methods, that have the interesting property of being able to learn complex features from data alone, have caught a lot of attention in a variety of domains [170, 192]. They have led to a large number of practical applications, especially in machine perception, and have shown to outperform the prediction accuracy of the traditional models in many scientific disciplines. Prominent examples occur in theoretical biology and active matter [202, 40], medicine [53, 74], or materials and chemical science [200, 66, 26, 181].

In the discipline of pedestrian dynamics, there is a high interest in making accurate predictions of pedestrian trajectories due to numerous real-world applications like facility, infrastructure, and building design [20], notably in the case of evacuation [23, 87, 249], autonomous driving cars [176], human-robot interactions [197], assistive technologies in industrial scenarios [135], and entertainment (e.g., augmented and virtual reality) [186].



(a) (b)

Figure I.1 For the prediction of pedestrian trajectories, it is useful to distinguish between scenes with few pedestrians (see (a)) and crowds of pedestrians (see (b)). Deep learning algorithms turn out to be accurate prediction tools for the scene (a), whereas knowledge-based approaches allow to describe collective phenomena at higher scales such as those described in the scene (b).

A trajectory is defined as the time-profile of the pedestrian's position. Changes of the position in a given time step can be interpreted as the velocity. Predicting trajectories means to assess the future motion of pedestrians in a given scene.

Besides the prediction of the simple physical motion of a single human, based on current or past observations, an important issue is to take into account their possible interactions as well as arising collective dynamics. Collective dynamics can be considered by predicting trajectories of many pedestrians in a scene simultaneously. In Fig. I.1(a) few pedestrians are present. The challenge consists in predicting individual trajectories in local interactions with the neighbors over relatively small horizon times. In Fig. I.1(b) a pedestrian crowd is shown. Predicting multiple trajectories and resulting collective crowd behavior at higher scale is necessary.

One approach to tackle the challenge of predicting trajectories is to use the knowledge-based (KB) approach, sometimes referred to by other names, such as physics-based [216], reasoning-based [129], expert-based [33] or traditional approach [19]. The KB approach contains models that are defined by basic rules or generic functions and considers physical as well as social or psychological factors of the pedestrians. They can be specified by a few parameters which generally have physical *interpretations* and allow to adjust the model. Prominent microscopic KB models are the social force model (SFM) [91] or the optimal reciprocal collision avoidance (ORCA) [222]. In general, the goal of these models is to improve the understanding of pedestrian dynamics and to identify microscopic mechanisms and parameters for the emerge of patterns and organisation in given scenes [178, 86, 36, 15, 157, 204, 23]. These models can be successfully applied for trajectory predictions in both scenes shown in Fig. I.1.

The other promising possibility to predict pedestrian trajectory consists of the use of *supervised deep learning* (DL) methods, especially long short-term memory networks (LSTM), convolutional neural networks (CNN), and generative adversarial networks (GAN). In the literature, this approach is often referred to as learning-based [129], pattern-based [188], or data-based approach. Deep learning refers to methods for training neural networks with more than two hidden layers (deep neural networks) [62] and supervised to methods that learn from data. Improving trajectory predictions for autonomous vehicles, service robots, and urban video surveillance are main goals [188]. In opposite to the KB approach, there are no interpretable parameters and rules necessary, but large amounts of data and flexible algorithms.

The KB approach is scientifically relevant since several decades, see for illustration in Fig. I.2 the citation numbers of one of the most famous KB models, the social force model by Helbing and Molnár [91]. On the other hand, the DL approach is a youthful methodology that started to become highly relevant for the prediction of pedestrian trajectories after the publication of the social-LSTM from Alahi et al. [5]. After this date, the social force model started to get associated with pedestrian trajectory as well, see the publication number with the keywords combining social force and pedestrian trajectory in Fig. I.2. Yet, the applications of KB and DL approaches are not identical. KB models are mainly designed to describe and understand collective phenomena implying high numbers of agents in high density scenarios like in Fig. I.1(b). They are based on few parameters that can be interpreted and calibrated. This results in high flexibility of the prediction scenarios, and notably the possibility to change pedestrian motion preferences (e.g., higher desired time-gap during pandemics). DL methods are designed to predict pedestrian trajectories at very local scales in space (few meters) and time (few seconds). If their predictions are accurate, they do not allow to control the motion and to reproduce different kinds of behaviors. In general, they are used for low density scenarios at local scale like in scene (a) presented Fig. I.1. The question of whether DL approaches could also be suitable for large-scale simulation or successfully used to initiate collective dynamics is still open. On the other hand, the pertinence of KB models to predict local trajectories is, regarding the high accuracy of DL algorithms, questionable.

The combination of both KB and DL approaches seems to be especially promising. Recently, some authors try to implement components of the KB models in the DL algorithms to overcome crucial limitations of the DL approach, like the lack of interpretability or generality [7, 100, 125]. Furthermore, another possibility is to use DL methods to improve the accuracy of the KB models. This can be done by estimating the parameters based on the results of an DL algorithm or to implement a neural network in the KB simulations [247, 111]. These combinations are called the hybrid approach and they benefit from the strengths of both approaches and avoid their shortcomings. Other promising algorithms close to hybrid approaches rely on reinforcement learning (RL) and inverse reinforcement learning (IRL). The agents learn from their own experiences in RL methods while IRL partly rely on data.

In this article, we address a comprehensive bibliographical review of KB and DL approaches for the modeling and prediction of pedestrian trajectories. We critically compare the two modelling approaches from their technical aspects as well as their application fields. We highlight similarities and differences between the two methodologies and draw future development perspectives. The manuscript is organised as follows. A thorough literature review of microscopic KB pedestrian models is presented in the next section. In Sec. I.3, we give a literature overview about the DL approach distinguishing between long short-term memory networks, convolutional neural networks, and generative adversarial neural networks. In Sec. I.4 we show a comprehensive comparison of both approaches focusing on their methodologies, phenomena of interest, and application scales. In the last section, we discuss future directions and perspectives of common developments of the KB and DL modelling approaches.

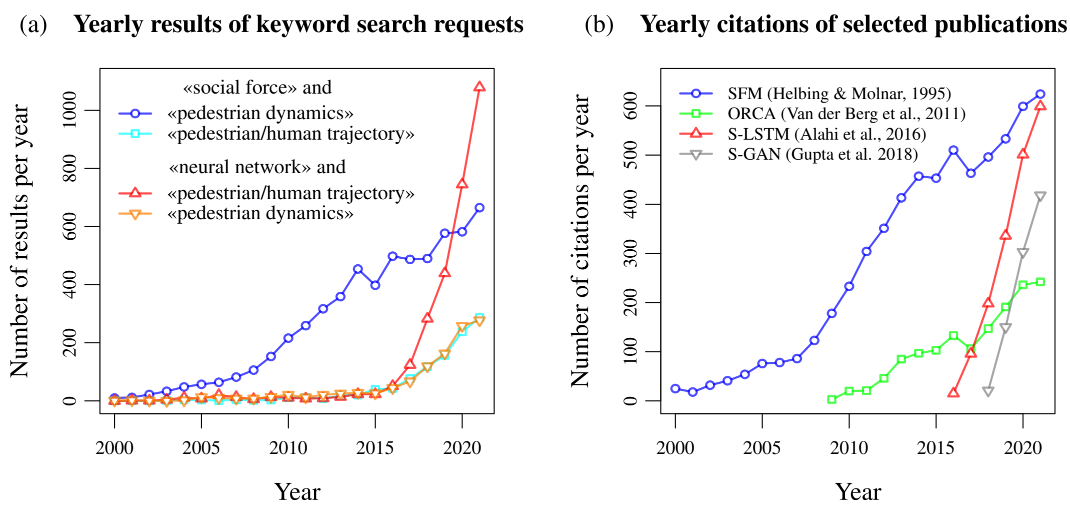


Figure I.2 Number of annual citations estimated by performing an online search with the engine Google Scholar [2]. Panel (a): Results of the tags «neural network» and «social force» in combination with either «pedestrian/human trajectory» or «pedestrian dynamics». Panel (b): Yearly citations for the articles by Helbing and Molnár [91] (SFM), Van den Berg et al. [222] (ORCA), Alahi et al. [5] (S-LSTM), and Gupta et al. [76] (S-GAN). SFM and ORCA are knowledge-based, whereas S-LSTM and S-GAN are deep learning approaches.

I.2 The knowledge-based approach

In the beginnings of the discipline, pedestrian dynamics researchers mostly used direct observations, photographs and time-lapse films to improve knowledge about the behavior of pedestrians [167]. This knowledge was applied to develop level-of-service concepts, design elements of pedestrian facilities and planning guidelines [63, 177, 86, 97]. These concepts and guidelines are useful to understand and control pedestrian dynamics, but are not suited for predictions of pedestrian flows or trajectories. Therefore, in the next step researchers started to create simulation models like force-based microscopic models [95], queuing models [141], the transition matrix model [65] or the models of Henderson [92, 93] which conjectured that the behavior of pedestrian crowds is similar to that of gases or fluids. These last models describe aggregated quantities and not individual pedestrian performances. They are called macroscopic models. Currently, KB pedestrian models range from macroscopic, mesoscopic, and microscopic models among others modeling scale characteristics. Macroscopic and mesosopic approaches are borrowed from continuous fluid dynamics or gas-kinetic models describing the dynamics at an aggregated level, while microscopic approaches model individual pedestrian motions. In the literature, many reviews focus on the modelling scales of pedestrian dynamics and the passages from a modelling scale to another [36, 85, 27, 14, 150, 38]. Some other reviews highlight pedestrian collective dynamics [195, 49, 23, 194] or applications in layout design [48]. In the following, we propose a review of KB pedestrian models focusing on microscopic approaches and their applications for prediction of pedestrian trajectories.

I.2.1 Microscopic pedestrian models

In the three last decades, many researchers focused on individual motion of pedestrians using different microscopic models. One of the advantages of microscopic approaches compared to macroscopic ones is their natural ability to reproduce heterogeneous behaviors. Indeed, the pedestrian being individually considered, it is straightforward to attribute specific characteristics to each agent and to take into account behavioral heterogeneity, among other heterogeneous aspects. On the other hand, microscopic models can be computationally expensive and their use is limited in case of large-scale simulation.

Microscopic pedestrian models consider the individual behaviors and interactions between individuals. The crowd dynamic phenomena result from mutual influences at a macroscopic level [48]. Microscopic models are mainly designed to reproduce characteristics macroscopic features such as fundamental diagrams or collective organisations like band formation [43, 98]. These kinds of models, describing individually pedestrian dynamics, can be used to predict trajectories of pedestrians at any scale. The individual pedestrian behavior is described according to certain KB rules ground on physical, social or psychological factors [38]. These rules are formulated in hand-crafted dynamic equations based on Newton's laws of motion. Given the input information about the initial status of the pedestrians like position, velocity, and acceleration a forward simulation of these rules can be used to predict the future trajectories.

Depending on the inputs and outputs of the model, the motion of the pedestrians to a new position can be determined in different ways. If the output of the model is the new velocity or acceleration, which then allows calculating the new position, the model is classified as a velocity- or acceleration-based model, respectively. If the position is directly determined by certain rules and is not based on differential equations, the class of models is called decision-based.

Acceleration-based models Acceleration-based models, typically *force-based* models, describe a class of microscopic models where the movement of pedestrians is defined by a superposition of exterior forces. One of the first force-based models dates backs to the 1975s and the work by Hirai and Tairu [95]. Nowadays, most of the acceleration-based models are force-based consisting of a relaxation (or anisotropic) term to the desired direction and an interaction term. This last term is generally the sum of repulsion (social force) with the neighbours and obstacles [39, 217]. The interaction force is the gradient of a potential on the distances and eventually the speed differences with the neighbors. This gradient may be exponential as in the social force model [91], algebraic as for the centrifugal and generalised centrifugal force models [244, 37], or partly linear as in the two-dimensional optimal velocity model [161]. The interaction force is weighted by a vision field concept, attributing more importance to the obstacles in front. Acceleration-based models, being of the second order, require relatively fine discretization scheme and may be subjected to numerical pitfalls [123]. Many currently developed acceleration-based pedestrian models are extensions of the social force model (see the review [31] and references therein). Yet, not all acceleration-based models rely on force concepts. See for instance the model by Moussaid et al. [158] based on the concept of desired time gap, the model by Karamouzas et al. [106] relying on time-to-collision variables, or a recent model by Lu et al. [143] based on anticipation mechanisms.

Velocity-based models Acceleration and force-based models allow describing inertial effects, as well as delays in the dynamics. Such mechanisms are questionable for pedestrian dynamics. Indeed, in contrast to a driver in a vehicle, inertial effects are minor for a pedestrian and almost no latency takes place in the motion process. These assertions, upon other, carry the current of *velocity-based* modeling approaches. Velocity-based models have been developed since the 2000s in the literature, later than the acceleration-based models [38]. They are partly inspired by robotic. Technically, velocity-based models rely on first order-differential equations, whereas acceleration-based models are based on second order equations. Velocity-based models are speed functions depending on the position differences with neighbors and obstacles. As for acceleration-based models, the speed of the neighbors (or speed difference) may be taken into account as well (making the system of speed equations implicit). A large class of velocity-based models are based on collision cones and collision avoidance techniques [171, 223, 222, 174, 77, 78, 113, 75]. The models are formulated as optimisation problems on the ensemble of feasible trajectories devoid of collisions [153, 152]. The presence of collisions is generally determined by assuming that the velocity of the neighbors is constant. The velocity ensemble leading to collisions describes then cones in space [171]. To avoid unrealistic oscillation effect (ping-pong effects), the models are extended to reciprocal velocity obstacle model (RVO) [223] or optimal reciprocal collision avoidance (ORCA) [222] for which avoidance techniques are determined in coordination between the pedestrians. ORCA models and their extensions are frequently used in computer graphics to reproduce crowd behaviors (see, e.g., [242, 28]). Other velocity-based models derive from concepts of bearing angle [169], gradient navigation [47], or, inspired from vehicular dynamics, time gap variable [213, 237].

Decision-based models and cellular automata In *decision-based* or *rule-based* models the pedestrian behavior is not modeled based on differential equations, but on rules or decisions determining the new agent positions, velocities, etc. [38]. The time is considered to be discrete for this class of models. In synchronous approaches, the pedestrians make decisions at time $t + \Delta t$ knowing the state of the system at time t. The time step Δt, playing the role of reaction time, has a direct physical meaning and can be used for the calibration of the model. Cellular automata (CA) are typical decision-based models. Not only the time is discrete in CA models, the space and state (i.e., velocity) of the pedestrians are discrete as well. The pedestrians evolve on a lattice, that is generally squared or hexagonal. A cell can be occupied by a single pedestrian only (exclusion rule). The size of a cell corresponds to the size of a pedestrian, generally 40 cm $\times$ 40 cm on a squared lattice, i.e., a maximal density of 6.25 ped/m^2 [232]. The first pedestrian CA models date back to the end of the 1990s [64, 160, 21, 119]. In floor field CA [25, 24, 115, 193], the rules and transition probabilities to the neighboring cells result from static and dynamic floor fields. The static floor field describe the desired velocity of the pedestrian. The dynamic floor field models the interactions with the neighbors. These interactions are inspired from the process of chemotaxis [16] used by some insects, typically pheromones with ants. An important modeling part of the CA approaches consists in solving conflict cases when two pedestrians covet the same cell at the same time. A priority rule may be defined, which can be random [115]. In [116], friction probabilities are introduced for which no pedestrian reaches the desired cell in case of conflict. Such a mechanism allows notably to explain clogging effects at bottlenecks. Recent decision-based models are based on cognitive effect [228, 229] or learning process [247].

I.2.2 Trends during the past decades

The modeling and experimentation of pedestrian dynamics is a quite young research field. First investigations and models date back to the 1960s and the 1970s [167, 95, 65, 92]. However, the topic has mainly been the subject of significant research over the past three decades. Experimental studies on pedestrian dynamics in laboratory conditions have been intensively carried out during the 2010s. Pedestrian experiments rely on uni-directional flow, counter-flow, bottleneck, intersecting flow, etc. An open access data archive can be found in Germany; see [1] and references therein. At the same time, authors developed different types of KB pedestrian models, ranging from microscopic to macroscopic modeling scales; see, e.g., the reviews [36, 85, 27, 195, 38]. Most important in the literature of the KB models is undoubtedly the microscopic social force model by Helbing and Molnár, and more generally the force-based microscopic modeling approach (see Table I.1). Traditional KB approaches by cellular automata, queuing processes, or in analogy to fluid or gas dynamics seen currently to reach a plateau, even if the trends are still light increasing (see Fig. I.3). The microscopic force-based models and approaches based on collision avoidance techniques remain relevant with an expanding number of citations. This is because they commonly serve as benchmark references to evaluate the quality of the predictions with deep learning methods (see, e.g., [5, 82]).

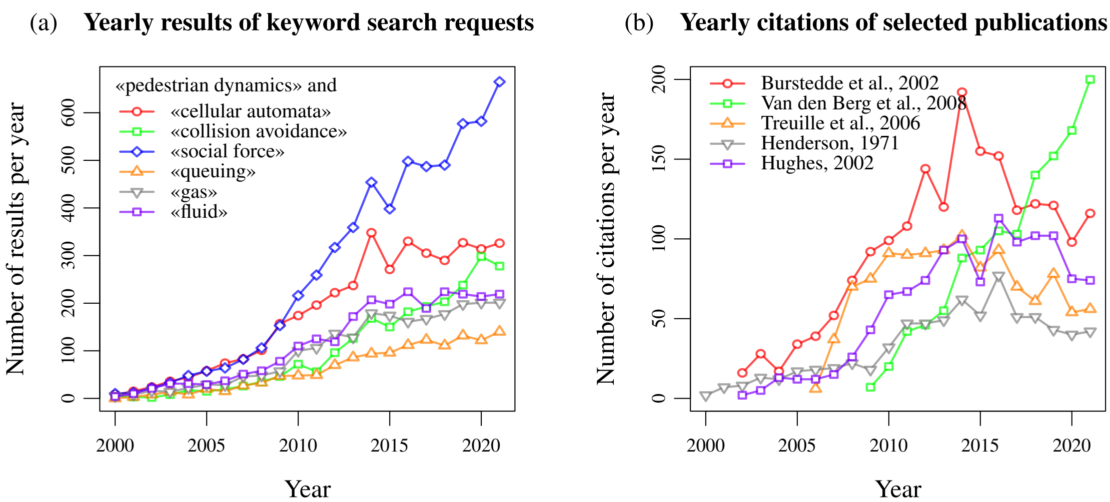


Figure I.3 Number of annual citations estimated by performing an online search with the engine Google Scholar [2]. Panel (a): Results for the tags «pedestrian dynamics» and keywords related to different model classes. Panel (b): Yearly citations for the articles by Burstedde et al. [24], Van den Berg et al. [223], Treuille et al. [221], Henderson [92], and Hughes [103].

I.2.3 Knowledge-based models for understanding and predicting

The aim of knowledge-based models is mainly to identify mechanisms and fundamental parameters operating collectively in the pedestrian dynamics. First of all, body exclusion effects are responsible for jamming and clogging effects, and

Table I.1 Important articles in knowledge-based pedestrian models focusing on microscopic approaches. (*Citations as of 31/12/2021*)

Author, Yr	Title and Reference	Family	Cites
Burstedde, 2001	Simulation of pedestrian dynamics using a two-dimensional cellular automaton [25]	Cellular automata	1931
Kirchner, 2002	Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics [115]	—	1119
Van den Berg, 2008	Reciprocal velocity obstacles for real-time multi-agent navigation [223]	Collision avoidance	1363
Pellegrini, 2009	You'll never walk alone: Modeling social behavior for multi-target tracking [174]	—	1206
Van den Berg, 2011	Reciprocal n-Body Collision Avoidance [222]	—	1440
Helbing, 1995	Social force model for pedestrian dynamics [91]	Force-based	6245
Helbing, 2000	Simulating dynamical features of escape panic [88]	—	5382
Chraibi, 2010	Generalized centrifugal-force model for pedestrian dynamics [37]	—	335
Moussaid, 2011	How simple rules determine pedestrian behavior and crowd disasters [158]	—	1009
Karamouzas, 2014	A universal power law governing pedestrian interactions [106]	—	247
Treuille, 2006	Continuum crowds [221]	Queuing	1176
Henderson, 1971	The statistics of crowd fluids [92]	Gas-kinetic	778
Hughes, 2002	A continuum theory for the flow of pedestrians [103]	Fluid dynamics	1174
Chowdhury, 2000	Statistical physics of vehicular traffic and some related systems [36]	Review	2862
Helbing, 2001	Traffic and related self-driven many-particle systems [85]	—	3979
Castellano, 2009	Statistical physics of social dynamics [27]	—	3963
Schadschneider, 2009	Evacuation dynamics: Empirical results, modeling and applications [195]	—	766
Bellomo, 2011	On the modeling of traffic and crowds: A survey of models, speculations, and perspectives [14]	—	464
Bechinger, 2016	Active particles in complex and crowded environments [13]	—	1507

notions of maximal density. The parameters of KB models rely on the concept of *fundamental diagram*, a phenomenological uni-modal relationship between the flow and the density. The fundamental diagram relationship has been pointed out as early as the 1960s. Yet, investigations on its shape and scattering are still actively ongoing [81, 168, 162, 203, 29, 246, 209]. Identified parameters are, upon others, the desired speed, the agent size, or the reaction time at microscopic scales. They are the maximal density or the capacity at the macroscopic level. The quantitative estimations of the parameters, as well as their numbers and nature, are controversial and subject to diverse influences, type of the flows (e.g., uni-directional, bi-directional), context and motivation, ages, or cultural effects (see [23] and references therein). Simple microscopic rules allows explaining macroscopic shapes of the fundamental diagram [196, 158, 69]. Here, as for traffic flow, temporal parameters such as the reaction time or the time gap with the next pedestrian ahead turn out to be highly relevant [158, 213].

One of the main highlights of KB models is the identification of self-organisation phenomena and the emergence of coordinated dynamics, patterns, structures, and orders at macroscopic scales. Multi-scale approaches allow understanding how microscopic individual behaviors initiate the emergence of macroscopic collective dynamics [43, 98]. Prominent examples of collective dynamics are lane formation [44, 68], stop-and-go waves [11, 61], freezing-by-heating effect [89, 208], herding effect [89, 115] or oscillations, intermittent flow, and pattern formation at bottlenecks and in intersections [86, 90, 41, 164]; see the review [36, 15, 23] and references therein. Comparable self-organisation phenomena arise in social systems and social networks, notably for opinion formation [94, 159, 218]. This includes a large class of non-equilibrium systems of self-driven or active particles often called *active matter* in the literature of statistical physics [201, 27, 179, 226, 147, 51, 13]. Understanding complex non-linear dynamics operating at different modeling scales remains challenging and currently attracts much attention, notably through data-based approaches [204, 163, 40, 170, 50].

The aims of microscopic KB models are mainly to provide a better understanding of large-scale dynamics from individual walking behaviors. Predictions of trajectories for a given dataset is no direct goal of KB models, but an indirect one. In the literature of pedestrian dynamics, some KB models have been proven to be useful for predicting pedestrian trajectories. They are implemented in multi-agent simulation tools [121, 45, 211, 118] to analyse pedestrian dynamics in different types of infrastructures or for specific outdoor events. The KB model, typically the social force model, is the technical kernel of the dynamics in complement to furthers mechanisms and controls describing different types of behaviors or motivation levels, agent characteristics, and further context effects.

I.3 The deep learning approach

The foregoing approach heavily relies on a theoretical modeling framework. Pre-identified key mechanisms were formulated in equations with a few meaningful parameters that have to be calibrated and validated. After that, the model can be used to simulate the scene, which can generate useful information for many applications. The predictive capacity for pedestrian trajectories almost comes as a by-product in the KB approach. In the DL approach the prediction of trajectories is not a by-product, but the main focus. As in many other domains, the DL approach captured a lot of attention over the last decade due to an increase of real and experimental databases, improvements in the computational capacity of computers [38], and the requirement of accurate pedestrian predictions for applications like autonomous vehicles or service robots [187].

There exist different possibilities in the literature to classify the DL methods. Rudenko et al. [187] classify the DL approach into sequential methods and non-sequential methods, based on the type of function approximation they are used for. Bighashdel and Dubbelman [19] classify the methods according to their main focus: The interaction-based approach, where the interactions between the pedestrians are addressed, the path-planning approach, where the trajectories are highly affected by the destinations, and the intention-based approach, where the intention is estimated. In this article, we differentiate between three classes of supervised DL algorithms: LSTM, CNN, and GAN. Currently, these three classes of methods seem to be most relevant in research, although there are also promising publications on transformer networks [67, 4] or variational autoencoder [104].
Before researchers start to use these DL algorithms, statistical models were applied to make predictions based on data. These statistical models learn pedestrian behavior by fitting different function estimates to data. One possibility is to use linear models with Kalman filters or extended Kalman filters [198, 154, 5]. Kalman filters can be used for predictions by propagating the current state with a dynamical model without the inclusion of new measurements. These simple models are not able to account for interactions between humans and thus do not fit for predictions of crowded scenes. The first statistical models that could learn interactions are those based on Gaussian process (GP) like the IGP from Trautman et al. [219]. They proposed an interacting GP where the trajectory of a pedestrian is represented as a Gaussian process. The interaction potential combines multiple trajectories, so that multi-modal distributions can be represented with relatively few parameters. It has been demonstrated, that they perform well with noisy observations and have closed-form predictive uncertainty. Also in [18, 220, 109, 52, 112] GP based models where proposed to model future pedestrian behavior. Other common statistical approaches that are not based on deep learning are approaches

based on Markov property [58]. These approaches include hidden Markov models, in which the hidden state is the pedestrians intent [18, 112]. Whereas the GP uses the entire observed trajectory for the prediction of future trajectories, the predictions with Markov models only depend on the current state [58].
Besides these statistical models, reinforcement learning (RL) is an important method for modelling crowd behavior. Most of all, RL techniques are relevant for robotics to anticipate surrounding pedestrian behavior and to plan a collision-free paths [32, 231]. RL is an unsupervised machine learning method, where an objective is learned via trial and error associated to a *reward function* that rewards or penalizes agent behaviors. Therefore, no data is required. It is assumed that the reward function contains the necessary information [151]. An exception to this is *inverse* RL (IRL) where the design of the reward function is based on data. Kretzschmar et al. [127] introduce an IRL algorithm that uses a maximum entropy probability distribution for a joint set of continuous state-space for mobile robot navigation in crowds. Kitani et al. [117] propose an well-established IRL algorithm for plausible path predictions. To extend the flexibility of the IRL algorithms recent works use deep IRL algorithms that can estimate non-linear, continuous reward functions. These deep IRL show promising results especially in robot path planning and vehicle driving behavior [235]. If it is the aim to predict pedestrian trajectories, these deep IRL algorithms are often combined with LSTM networks, which are supervised learning algorithms that will be discussed in the following. For more literature about RL in pedestrian dynamics refer to, e.g., [131, 149, 148]. For IRL methods used for trajectory predictions see [117, 182], for deep IRL methods see [134, 132, 54], or to get a wider perspective, refer to the surveys [224, 150].

I.3.1 Long short-term memory networks

A recurrent neural network (RNN) is an important class of machine learning methods, that uses feedback connections to store representations of recent input events in form of activations [96]. These feedback connections can exist between different time steps providing a temporal memory to the network [139]. Because of this capability, they are especially suited for sequence modeling tasks such as time series prediction and sequence labeling tasks [9]. Most successful are the LSTM architectures of RNNs, that use purpose-built memory cells to store information. They have achieved impressive results in many sequential prediction tasks like speech recognition [73], machine translation [210], handwriting recognition and generation [72], and image captioning [227]. LSTM networks can be trained for sequence generation by processing real data sequences one step at a time and predicting what comes next so that temporal coherence can be ensured. The training of a LSTM network can be lengthy and difficult, but the forward-propagation of

Table I.2 Selection of important articles of deep learning algorithms predicting pedestrian trajectories.(Number of citation based on search from 31/12/2021)

First author, year	Article's title and reference	Family	Citations
Trautman, 2010	Unfreezing the robot: Navigation in dense, interacting crowds [220]	Gaussian	459
Kitani, 2012	Activity forecasting [117]	IRL	727
Alahi, 2016	Social-LSTM: Human trajectory prediction in crowded spaces [5]	LSTM	1770
Kretzschmar, 2016	Socially compliant mobile robot navigation via IRL[127] [117]	IRL	326
Yi, 2016	Pedestrian behavior understanding and prediction with deep neural networks [241]	CNN	124
Lee, 2017	DESIRE: Distant future prediction in dynamic scenes with interacting agents [134]	LSTM	604
Fernando, 2018	Soft+ hardwired attention: A LSTM framework for human trajectory prediction and abnormal event detection [59]	LSTM	219
Gupta, 2018	Social GAN: Socially acceptable trajectories with generative adversarial networks [76]	GAN	910
Nikhil, 2018	Convolutional neural network for trajectory prediction [166]	CNN	83
Vemula, 2018	Social attention: Modeling attention in human crowds [225]	LSTM	357
Xue, 2018	SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction [239]	LSTM	217
Sadeghian, 2019	SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints [189]	GAN	457
Rudenko, 2020	Human motion trajectory prediction: A survey [187]	Review	282

new data (testing) can happen quite fast. That is why they are a popular choice for real-time tracking and predicting.

Recently, many researchers use these LSTM networks to predict trajectories

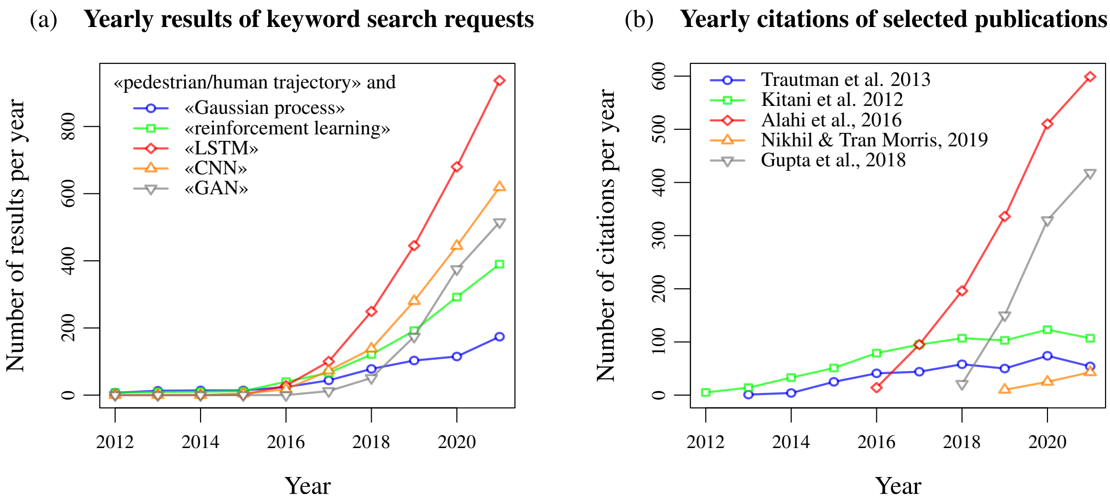


Figure I.4 Number of annual citations estimated by performing an online search with the engine Google Scholar [2]. Panel (a): Results for the tags «pedestrian trajectory» and keywords related to the class of algorithm. Panel (b): Yearly citations for the articles by Trautman et al. [219], Kitani et al. [117], Alahi et al. [5], Yi et al. [241], and Gupta et al. [76].

of pedestrians. Most famous is the social-LSTM proposed by Alahi et al. [5]. It is the first major DL application that takes the interactions between pedestrians in crowded scenarios into account. The main contribution of the social-LSTM is a novel pooling layer called social pooling that gathers the hidden states of nearby pedestrians. With that technique, the influence of neighboring pedestrians on the movement of the ego pedestrian can be included as an input for the following prediction step. Inspired by the success of the social-LSTM many researchers try to use LSTM networks with different settings. The scene-LSTM incorporates additional scene information [146]. The social-scene LSTM [239] considers the scene influence, the social information of the person, and scene scale information in three different LSTM networks. Pfeiffer et al. [175] incorporate static obstacles that have to be avoided. Other approaches that process scene information are [134, 140]. Besides the capturing of scene information, some LSTM algorithms integrate an attention mechanism in the interaction module to capture the relative importance of each person in the scene. Other LSTM networks focus on attention mechanisms in the interaction module to capture the relative importance of each person in the scene. In some of these works, the attention weights are learned by data or by added handcrafted based on domain knowledge [59, 80, 102, 212]. Another relevant work is STGAT by Huang et al. [102] that uses graph neural networks (GNN) instead of a pooling module. To share information across the pedestrians, each agent is treated as a node of a graph. Derived from this work DAG-Net [156] presents a double attentive GNN combined with a RNN and STR-GGRNN [79] introduces an online framework that automatically infers the social interactions by completing the graph edges. An overview of DL pedestrian trajectory prediction algorithms relying on LSTM networks is provided in Table I.3. The works are sorted by the publication year to give the reader an impression about the overall progress in this field.

I.3.2 Convolutional neural network

Convolutional neural network (CNN) are machine learning methods mostly used for computer vision tasks like image classification [128], object detection, object tracking and image segmentation [17]. For studies of pedestrian dynamic these methods are highly important because of their impressive performance in classifying and tracking of objects like pedestrians or vehicles [55, 184, 243]. Another advantage is the reduced computational load outperforming regular neural networks [155, 166]. However, CNN are not widely used to predict pedestrian trajectories, because these are non-sequential methods, which makes it difficult to design the network input and output [241]. They are mostly used for trajectory predictions of road vehicles [136], the prediction of pedestrian behaviors for autonomous vehicles [180, 3], or pose/action recognition [240, 35]. The first CNN designed to model and predict pedestrian trajectories is the "Behavior-CNN" from Yi et al. [241]. It is a 3-stage deep CNN that encodes the pedestrian behavior into sparse displacement volumes which can be directly used as network input at the first stage. The CNN from Nikhil et al. [166] utilizes a highly parallelizable convolutional layer to handle temporal dependencies. Trajectory histories are embedded by means of a fully connected layer. Stacked convolutional layers are used to learn temporal dependencies in a consistent manner. The features from the final convolutional layer are passed through a fully-connected layer which generates all predicted positions simultaneously.

An extension of the CNNs are the graph CNNs which were first introduced by [114]. Mohammed et al. [155] propose the Social-STGCNN that uses graph CNN by modeling the interactions as a spatio-temporal graph, whose edges model the social interactions between the pedestrians. Dan [46] proposed a graph CNN combined with a LSTM network. The graph CNN extract the feature from the pedestrians and the scene for which every pedestrian is regarded as a node and the relationship between each node and its neighbors is obtained by graph embedding. The LSTM encodes the relationship so that the model can predict nodes trajectories. Other approaches combing CNNs with LSTM networks are [238, 105, 183]. An overview of DL pedestrian trajectory prediction algorithms relying on convo-

Table I.3 Overview of DL pedestrian trajectory prediction algorithms relying on LSTM networks.

First author, year	Name	Main characteristics
Alahi, 2016	Social-LSTM	Social pooling layer to handle interactions
Lee, 2017	DESIRE	Ranks and refines the generated trajectories
Fernando, 2018	Soft + Hardwired Attention	Utilises “soft attention” as well as “hardwired” attention
Xue, 2018	SS-LSTM	Three different LSTMs to capture person, social and scene scale information
Manh, 2018	Scene-LSTM	Incorporates scene information
Hasan, 2018	MX-LSTM	Takes head pose and vision range into account
Vemula, 2018	Social-Attention	Based on social-LSTM, but captures the relative importance of each person when navigating in crowds
Bisagno, 2018	Group-LSTM	Coherent filtering algorithm to segment groups
Cheng, 2018	Social-Grid LSTM	Combines social pooling and grid-LSTM methods
Bartoli, 2018	Context-aware-Social-LSTM	Interactions with static elements and dynamic agents
Pfeiffer, 2018	Static-LSTM	Angular pedestrian grid combined with CNN
Ivanovic, 2019	The Trajectron	LSTMs combined with CVAEs and dynamic spatiotemporal graphical structures
Lisotto, 2019	SNS-LSTM	Social, navigation, and semantic pooling mechanism
Huang, 2019	STGAT	Spatial-Temporal Graph neural network combined with LSTM
Haddad, 2019	Situation-Aware-LSTM	Spatio-temporal graph that operates on the local and global contexts
Syed, 2019	SSeg-LSTM	Semantic segmentation to incorporate scene information
Zhao, 2019	MATF	Encodes past trajectory and scene context into a Multi-Agent Tensor
Monti, 2020	DAG-Net	Double attentive GNN that deals with past interactions and future goals

lutional neural networks is given in Table I.4.

Table I.4 Overview of DL pedestrian trajectory prediction algorithms relying on convolutional neural networks.

First author, year	Name	Main characteristics
Yi, 2016	Behavior CNN	3-stage deep CNN that creates sparse displacement volumes
Varshneya, 2017	SSCN	Static spatial context modeled with CNN
Rehder, 2018	RMDN	CNN for inferring destination from images and position. LSTM for prediction
Nikhil, 2018	CNN	Highly parallelizable CNN to handle temporal dependencies
Mohamed, 2020	Social-STGCNN	Models the interactions as a graph using social spatio-temporal graph CNN
Yu, 2020	TGConv	Transformer-based graph convolution mechanism
Dan, 2020	Spatial-Temporal Block	Spatial Temporal Graph CNN combined with LSTM
Ridel, 2020	COVLSTM	2-D grid combined with CNN and LSTM
Jain, 2020	DRF-Net	Discrete residual flow network
Zhang, 2021	Social-IWSTCNN	CNN with spatial and temporal features
Zhao, 2021	STUGCN	CNN with spatio-temporal graph architecture
Zamboni, 2021	Conv2D	2D Convolutional models with different network architectures

I.3.3 Generative adversarial networks

A problem that is inherent with predictions of trajectories is the multimodal nature of future pedestrian trajectories. GANs have a high potential to cope with this problem, because of their capabilities to generate multimodal samples [101]. With the use of GANs, it is possible to predict a distribution of potential future trajectories and not just the best single trajectory [191]. The architecture of a GAN consists of two part: a generator and a discriminator. In the DL approach, a neural network is trained to match the desired data distribution and achieve a low error rate. In a non-intuitive way, the generator of a GAN is trained so that the error rate of the discriminator increases at first. The generator takes the training data as input and creates different output samples. In turn, the discriminator takes the generated samples and the training data and tries to distinguish whether a given sample belongs to the true data distribution or is generated by the generator. Both components are engaged in a competition similar to a two-player min-max game where each one tries to outsmart the other one [76, 136]. From this process, the generator learns to generate data that resemble the true data distribution. Although the results of the GAN-based methods are promising, there are two main difficulties. First, GANs can be hard to train and second, GAN learning often suffers from mode collapse [8, 190]. Applied for trajectory prediction, most GANs are combined with LSTM networks. In the social GAN from Gupta et al. [76] the generator is composed of an LSTM-based encoder, a context pooling module, and an LSTM-based decoder. The discriminator uses LSTMs as well. The pooling module adopted in the social GAN uses a uniform weight for all surrounding pedestrians. Therefore, it can not distinguish the different effects exerted on a target pedestrian by pedestrians at different distances and traveling at different speeds. For that reason, many authors added attention mechanisms. Sadegehian et al. [189] add an attention module to assign different soft attention distribution weights to the surrounding pedestrians and the static environment. Using an attention module with a physical and social component and a feature extractor module composed of a CNN and several LSTM network encoder, the model SoPhie learns the interaction information of different agents and extracts the most important information from the neighbors. Social Ways [6] applies info-GAN [30], which introduces latent code to enhance the multi-modality of prediction and computes the attentive social features to generate a more convincing result. It uses discrimination loss for the discriminator, adversarial loss for the generator and information loss for both. Social-BiGAT [122] relies on BiGAN architecture to help reduce the variance of the predicted trajectory distributions and allow for better generalization. In Lv et al. [144] the authors propose a GAN combined with transformer networks which generates trajectory distributions to capture the uncertainty of the predictions. However, these methods focus on the trajectory prediction in homogeneous environment without considering the types of road users. Lai et al. [130] use an attention module, containing two components, in order to alleviate the issues given by the complexity of a scene with many heterogeneous interacting agents [130]. An overview of DL pedestrian trajectory prediction algorithms relying on generative adversarial neural networks is proposed in Table I.5.

Table I.5 Overview of DL pedestrian trajectory prediction algorithms relying on generative adversarial neural networks.

First author, year	**Name**	**Main characteristics**
Gupta, 2018	Social-GAN	GAN composed of LSTM encoder, context-pooling module and LSTM decoder
Fernando, 2018	GD-GAN	GAN for pedestrian trajectory predicting and group detection
Sadeghian, 2019	SoPhie	GAN that uses path history and scene information
Amirian, 2019	Social Ways	Uses discrimination-, adversarial- and information loss
Kosaraju, 2019	Social-BiGAT	GAN for multimodal trajectory predictions
Lai, 2020	AEE-GAN	Info-GAN architecture with recurrent feedback
Huang, 2021	STI-GAN	Combination of graph attention network and GAN

I.4 Comparing the approaches

In this Section, we present a comparison of the KB and the supervised DL approaches. First, we focus on technical differences (Sec. I.4.1) and second, we describe the differences regarding the applications for predicting pedestrian trajectories (Sec. I.4.2).

I.4.1 Technically oriented comparison

There are considerable differences between the two approaches, which makes it difficult to create a common framework that both can share for a fair comparison. Kothari et al. [124] define trajectory predictions as "given the past trajectories of all humans in a scene, forecast the future trajectories which conform to the social norms". This definition fits for the DL approach, but not the KB approach where the past trajectories are not used to predict (simulate) the future ones. In KB methods, inputs describe the current state of the system through, e.g., instantaneous relative position [91], relative velocity [37], time gap [158], or collision-related indicators such as bearing angle [169], collision cone [223] or time-to-collision [106]. Moreover, the approaches have different criteria in terms of quality of the fit. A "good" KB model has few interpretable parameters, can simulate realistic pedestrian trajectories, and improves understanding of the phenomena. To evaluate the performance of a KB method, the parameters have to be calibrated and the model needs to be validated. The calibration phase is done using knowledge on the values of the parameters (i.e., 1 or 1.5 m/s for the pedestrian desired speed, see [232]) or is formulated using empirical data as an optimisation problem (by least squares or by maximum likelihood, see, e.g., [42, 142, 214, 22]). The validation of KB methods is generally done using three typical ways: by using fundamental diagrams [196, 57], using data [138, 185, 120] and by comparing the resulting phenomena with real-world self-organization phenomena. The goal of the DL approach is to predict trajectories that are as close as possible to the real trajectories. Common key figures for the performance of the DL approaches are the average displacement error (ADE) [174] and the final displacement error (FDE) [5]. The first one averages the Euclidean distance between points of the predicted trajectory and the ground truth that have the same temporal distance and the second one measures the distance between the final predicted position and the ground-truth position. When the two approaches are compared in the literature the ADE and FDE are used as reference (see, e.g., [5, 82, 76]).

The algorithms of the DL approach have a large number of parameters that have no direct physical meaning and therefore can be referred to as coefficients. These coefficients have to be trained by one part of the data set (training phase). In general, this is done by using an training algorithm like the back-propagation algorithm [165]. The aim of this training algorithm is to adjust the network settings in a way, that minimizes the given cost function [71]. Common cost functions are the mean square error or the cross entropy. After the training is complete, the algorithms are feed with the rest of the data set to evaluate predictions (testing phase). This cross-validation method allows detecting overfitting phenomena, when the algorithm presents low training error but poor performance for the prediction of new data. The training is usually computationally expensive and is made offline, whereas the predictions of the trajectories (i.e., the computation), once the training is completed, are quite fast and can be made online (in real-time) [7]. We summarize our findings regarding the technical differences of both approaches in Table I.6.

I.4.2 Application oriented comparison

Besides the technical orientated differences, there are also great differences regarding the applications of both approaches. KB approaches are mostly used for crowded situations. Collective dynamics are described at macroscopic scales. Applications mainly rely on large-scale simulations to analyze, e.g., infrastructure design or evacuation situations. In contrast to that, the DL approach focuses more on single pedestrians and their interactions with other pedestrians or the environment locally and in low density situations. Applications are mainly for automatised mobile systems like autonomous vehicles or industrial robots that have to anticipate the future behavior/trajectory of pedestrians to avoid collisions. The DL approach surpasses the KB approach in complexity (i.e., number of inputs as well as number of parameters/coefficients). This provides a high flexibility and enables the DL algorithms to learn complex interactions and motion patterns when the amount of data is sufficient. This is especially useful in situations where complexity prohibits the explicit programming of a system´s exact physical nature.

The importance of data There are great differences in the role of data for the two KB and DL approaches. The performance of the DL algorithms highly depends on the quality of the data. The majority of researchers in this discipline use datasets containing low density situations with few interacting pedestrians to train the algorithms. A collection of datasets that are widely used to train and evaluate the algorithms are TrajNet++ [124] or OpenTraj [2]. The KB models do not need data to learn pedestrian behavior. Data is just needed for calibrating the parameters. In practice, this is often done by using experimental data including high density situations, see, e.g., [1]. The data used for the KB approach contains the trajectories of each pedestrian as well as the structure of the environment and locations of obstacles. For the DL algorithms, the data can be more varied

Table I.6 Technically oriented comparison of the KB and the DL approach for prediction of pedestrian dynamics.

	Knowledge-based	**Deep learning**
Semantics	Model Parameter Calibration Validation	Algorithm Coefficient Training Testing
Methodology	Differential equation systems or cellular automata	Neural network, mostly RNN, LSTM, CNN, GAN
Inputs	System actual state (e.g., pedestrian relative positions, velocities, etc. at time t)	Past trajectories discretised over the time interval $[t-T, t]$, $T \approx$ 2–4 s
Outputs	Future pedestrian positions at time $t + \delta t$, δt ranging from 0.01 s (discretised differential system) to 0.5–1 s (cellular automata)	Future pedestrian positions discretised over time interval $[t, t + T^\star]$, $T^\star \approx$ 3–5 s
Modeling concepts	Fundamental diagram	Learning human-interaction
	Body exclusion, maximal density	Social pooling
	Desired velocity and social force	Attention mechanisms
	Static and dynamic floor field	Group dynamics
	Collision avoidance techniques, spatial and temporal anticipation mechanisms	Human-space interactions
	Vision angle, bearing angle	
	Temporal interaction indicators (time gap, time-to-collision, time-to-interaction)	
Performance evaluation	Flow-density relationship	Average displacement error (ADE)
	Comparison to experimental data	Final displacement error (FDE)
	Description of collective behaviors and self-organised phenomena	Modified Hausdorff distance (MHD)

because the algorithms are more flexible in terms of inputs. In the graph neural networks, the pedestrians and their interactions are described through a graph [225]. Other authors use spatial information represented as points of interest [12] or as occupancy maps [175]. By using CNNs it is even possible to use images or videos as inputs for the predictions [189, 122].Without sufficient amounts of data, the DL algorithms are not capable to learn pedestrian behavior and predict future trajectories successfully. In addition to the amount of data, pre-processing is an important step to archive good results and efficiently train the algorithm. One pre-processing technique is data normalization where the coordinates of the trajectories are normalized to coordinates with origin in the first observation, coordinates with origin in the last observation, or relative coordinates [245]. Another important pre-processing technique is data augmentation. Using GANs [205] is one form of data augmentation, but there are also more basic forms like rotating the input trajectories, mirror the trajectories, or applying Gaussian filter [245]. Schöller et al. [199] show that data augmentation helps to prevent the DL algorithm from learning environmental priors instead of pedestrian behavior.

Numerical comparison Many studies report on numerical comparisons of KB and DL approaches for the prediction of trajectories. Authors generally used average and final displacement errors (ADE and FDE) to quantify the prediction accuracy. Several articles include the social-LSTM [5] and the social force model [91] as references for, respectively, DL and KB approach. In Table I.7, the ADE and FDE metrics from selected studies using the social-LSTM as a benchmark are shown. All these articles use the same algorithm, the same datasets, the same error metrics, the same length of observed (3.2s) and predicted trajectories (4.8s), and the same cross-validation methods. Nevertheless, the results vary significantly from one analysis to another. This is not just a problem of the social-LSTM, but of the evaluation of the DL approach in general. Reasons for this high variation could be the randomness of the dataset splitting, the randomness of the starting coefficients, differences in the implementation of the algorithms, or different hyperparameter settings for the training, among others. An attempt to solve this problem is the trajectory forecasting challenge Trajnet++ [124]. The challenge provides a uniform sampling and evaluation system allowing to compare rigorously different algorithms in the same framework. Table I.8 reports on ADE and FDE metrics for selected articles using the social force model and social-LSTM algorithm. We can observe that the error estimates also vary significantly with the social force model. Yet, in contrast to Table I.7, the studies do not systematically use the same dataset and prediction length making direct comparisons potentially biased. The objective is to compare the KB and DL approaches using relative errors. Assuming that the setting for KB and DL approaches are identical for a given study, the comparison of relative errors is fair. The variations from one study to another are again very significant. However, except in the work of Cheng et al. [34], the social-LSTM systematically outperforms the social force model in terms of prediction accuracy. The mean ADE and FDE relative errors are, respectively, 105 and 80.4%. In the work from Hasan et al. [83] or from Song et al. [207] the differences are especially huge, the social-LSTM being up to 341% more accurate. Such results are statistically not surprising since the DL approach is based on much more free parameters (coefficients) than the KB approach.

Table I.7 Quantitative comparison of ADE and FDE metrics for articles using social-LSTM as benchmark with different datasets.

First author, year	Average	ETH	HOTEL	ZARA1	ZARA2	UCY
Syed, 2019	0.08 / 0.14	0.15 / 0.295	0.05 / 0.08	0.05 / 0.08	0.07 / 0.1	0.1 / 0.16
Xue, 2018	0.12 / 0.17	0.2 / 0.37	0.08 / 0.13	0.08 / 0.11	0.07 / 0.12	0.2 / 0.24
Manh, 2018	0.25 / 0.22	0.18 / 0.34	0.25 / 0.29	0.37 / 0.33	0.19 / 0.1	0.25 / 0.03
Alahi, 2016	0.27 / 0.61	0.5 / 1.07	0.11 / 0.23	0.22 / 0.48	0.25 / 0.5	0.27 / 0.77
Vemula, 2018	0.37 / 3.32	0.46 / 4.56	0.42 / 3.57	0.21 / 0.65	0.41 / 3.39	0.36 / 4.45
Hossain, 2022	0.44 / 0.98	0.60 / 1.31	0.15 / 0.33	0.43 / 0.93	0.51 / 1.09	0.52 / 1.25
Zhu, 2019	0.45 / 0.91	0.73 / 1.48	0.49 / 1.01	0.27 / 0.56	0.33 / 0.7	0.41 / 0.84
Hasan, 2018	0.64 / 1.45	—	—	0.68 / 1.53	0.63 / 1.43	0.62 / 1.40
Gupta, 2018	0.72 / 1.54	1.09 / 2.35	0.79 / 1.76	0.47 / 1.00	0.56 / 1.17	0.67 / 1.40

Table I.8 Quantitative comparison of the social force model and the social-LSTM (average ADE/average FDE).

First author, Approach	Alahi	Fernando	Cheng	Fernando	Hasan	Song
Social force model	0.39 / 0.60	3.24 / 4.86	0.37 / 1.27	1.5 / 2.46	4.28 / 7.63	0.61 / 0.96
Social-LSTM	0.27 / 0.61	1.76 / 3.51	0.67 / 3.1	1.1 / 1.89	0.97 / 2.08	0.25 / 0.22
Relative error [%]	44 / -1.6	84 / 38	-80 / -145	36 / 24	341 / 267	144 / 336

Advantages and disadvantages of KB and DL approaches The last paragraph shows that the DL algorithms outperform the KB models in terms of prediction accuracy of pedestrian trajectories in low density situations and that the results of the error metrics have a high fluctuation. This fluctuation is mainly a problem of the DL algorithms because the evaluation of these algorithms is based on these error metrics. In addition to this lack of reproducibility, there is also the problem of missing explainability of the DL approach. This means that the coefficients can not be physically understood and interpreted. Therefore it is not clear why the predicted trajectories have the given shapes. In applications where autonomous systems make decisions, it is important to understand and communicate why specific choices have been made [172]. But these in large numbers existing coefficients are also an advantage of the DL approach because they make it possible to learn complex behaviors by fitting a set of values such that the predicted behavior fits the observed behavior. These algorithms do not need a priori knowledge about the system, just observations.

The comparison of the KB models with the DL algorithms is mostly done numerically, disregarding the missing reproducibility and high fluctuation. Furthermore, important advantages of the KB models are not taken to account in the numerical comparison. First of all, the KB models have the advantage of simple forms and interpretable parameters, which makes it easy to reproduce the results and to understand the predictions. The interpretability of the parameters makes the models flexible and easy to cope with environmental and behavioral changes. For example, if the preferences of pedestrians change because of a pandemic, which entails social distance and desired time-gap to rise, the KB methods can change the parameters to model these new situations, without needing new data. In the DL approach, new data is necessary to learn behavioral changes. But the KB models also have disadvantages, like the need of domain knowledge. Furthermore, only average pedestrian behavior is simulated and it is difficult to capture the complete crowd behavior range with a single model [158]. The advantages and disadvantages, as well as other differences regarding the applications of both KB and DL approaches, are summarized in Table I.9.

I.5 Future directions

In the last part of this work we want to look into the future and describe which trends we can identify.

I.5.1 The hybrid approach

In Table I.8 it is shown that many points that can be criticized in the predictions with the DL algorithms are strengths of the KB models. KB models have few parameters with physical interpretations requiring few data for calibration, while the coefficients of the DL algorithms are generally not interpretable and much data is needed in the training phase. For this reason, the combination of both approaches seems to have potential as some pioneer studies point out [173, 100]. This combination is called the hybrid approach and even though the idea of combining both approaches has picked up momentum just in the last few years, there is already a vast amount of work in diverse disciplines. Examples of the hybrid approach are given in applications like the discovering of novel climate patterns [108, 56],

Table I.9 Application oriented comparison of the KB and the DL approach for prediction of pedestrian dynamics.

	Knowledge-based	**Deep learning**
Applications	Simulation tool Infrastructure design Evacuation situations Intelligent transport systems	Socially-aware mobile robots Design of intelligent tracking systems Pedestrian trajectory prediction for autonomous vehicles
Application scales	Large-scale simulation Large infrastructures (train station, stadium, commercial mall), urban centers, outdoor events	Local prediction scale (in time and space) Few interacting pedestrians
Crowd density	Low density (long range interaction, e.g., collision avoidance models) High density (short range interaction, e.g., force-based models)	Low density situations Long range pedestrian interaction
Advantages	Interpretable parameter Explainable predictions Reproducible Few data needed	Accurate predictions Learn complex interaction No domain knowledge necessary No modelling-bias Can process different types of data
Disadvantages	Low use of data Averaged behavior only Not suitable to complex interaction Complete crowd behavior range with a single model difficult to capture	Not interpretable Not reproducible in praxis Lack of generalisation Require large amount of data Necessary complexity of the network unknown

the finding of novel compounds in material science [84, 60], the designing density functionals in quantum chemistry [137], or the improving imaging technologies in bio-medical science [234, 236]. Recently, the first applications for pedestrian trajectory predictions can be found in [247, 7, 125].

In general, there are three ways of combining both KB and DL approaches. First, KB models can be used to improve the DL algorithms. One possibility to do so is *data generation*. Simulations based on KB models are carried out to obtain a synthetic data set, that is used to train and test the neural network. A popular application of data generation can be found in the training of autonomous vehicles for the augmentation of data for scenarios that are not sufficiently represented in the available data set [133, 230]. In [110, 247, 5], the authors apply the social force model to generate a synthetic data-set for the training of neural networks and setting of hyper-parameters. *Knowledge-guided design of architecture* is another possibility to improve the DL algorithms with knowledge. The modular and flexible nature of the networks enables the use of knowledge to specify node connections that capture dependencies among variables. In this way, Antonucci et al. [7] successfully embed the social force model in the architecture of a neural network to generate predictions of human motion. A common technique to improve the output of DL algorithms is *knowledge-guided loss function*. It makes the output consistent with physical laws so that unrealistic prediction can be ruled out [233, 206]. Because the training of DL algorithms is an iterative process, they require an initial choice of coefficients as a first step to commence the learning process. *Knowledge-guided initialization of the network* can be used to guide the network at an early stage to archive generalizable and physically consistent results [107].

The second way of combining the approaches is obtained by using DL algorithms to improve the prediction accuracy of the KB models. One of the oldest and most common way for addressing the imperfections of the KB approach is *residual modeling*. A DL algorithm learns to correct the errors made by the KB model by predicting the model residuals [233]. Bahari et al. [10] proposed a so-called "realistic residual block" to improve vehicle trajectory predictions. Another way to improve the KB predictions consists in calibrating the parameters of the KB models by using DL algorithms. Göttlich et al. [70] use neural networks to estimate the parameters of the social force model. Hossain et al. [99] extend the social force model [91] using group forces, based on neural networks, to consider interactions with static obstacles, other pedestrian, and pedestrian groups. Kreiss [126] uses deep neural networks to estimate different interaction potentials for the social force model.

Given enough data, DL algorithms are capable of predicting pedestrian trajectories of a given scene with relative high accuracy. However, pedestrians experience different interactions in different situations. Whether they are able to make accurate predictions for the whole range of possible scenes and interactions is an open question. The hybrid approach is useful for this problem because it can compensate scarcity of data with available knowledge. The hybrid approaches are promising to improve the predictions of pedestrian trajectories because they benefit from the strengths of both approaches and reduce shortcomings like the missing explainability of the DL approach.

I.5.2 Other directions

As it has been shown in Fig. I.2, the DL approach has gained attention just a few years ago. At the current state-of-the-art, DL is mostly used for learning human behavior and predicting single pedestrian trajectories. In the future, the DL approach could be used in more applications in the discipline of pedestrian dynamics and notably for large scale simulation and the simulation of collective dynamics. Nowadays, the behavior and interactions of agents in simulation platforms currently available rely on KB approaches [121, 45, 211]. Following the success of the DL algorithms for predictions of pedestrians in low density situations, there is a high potential to use these methods for accurate predictions of crowd dynamics like in evacuation situations.

In addition to the KB models, RL algorithms are often used for simulating crowd dynamics, but they have some disadvantages, like the need of a reward function, a priori given goal/destination, or the difficulty to incorporate interactions. Such difficulties are overcome by the supervised DL algorithms. The combination of RL and DL algorithms seems to be promising to solve these disadvantages as Everett et al. [54] have recently shown. Other works have shown successful applications of deep RL for crowd simulations [132] or even evacuation dynamics [248].

There are still open questions to be tackled, before supervised DL algorithms could be successfully used for large-scale crowd simulations, like:

- Which kind of neural networks, which complexity and which training data should be used for large-scale simulation including different types of geometries?

- Should the type and complexity of appropriate networks as well as the data used for the training depend on the scenarios of the simulation?

Besides these considerations, one may expect to develop deep networks trained on large amounts of data that could predict accurate trajectories for any density levels and any type of facility. Such networks could be used in agent-based simulation platforms whose objective is to simulate any type of scenario. Yet, the question whether such universal networks, as supervised approaches, require training on datasets representing the full diversity of pedestrian dynamics remains open.It may be possible, as in unsupervised approaches, that training on few data is sufficient to obtain accurate predictions even for scenarios the networks are not trained for. Such question raises more generally on the robustness of the predictions against new data, i.e., new situations, scenarios, density levels, or types of facilities. Some preliminary results obtained in a corridor and a bottleneck have shown that neural networks are quite robust to new types of facilities and may even overcome KB models in terms of prediction robustness [215, 216].

Another possible development direction of DL approaches for prediction pedestrian trajectory concerns the nature and type of the inputs. The inputs of the algorithms are mostly the trajectories of the pedestrian over finite past horizon. Yet, studies of pedestrian dynamics with KB models identified different relevant variables like, besides the relative position [91], the relative velocity [37], the time gap [158], or indicators related to possibilities of collision such as bearing angle [169], collision cone [223] or time-to-collision [106]. The use of these variables as inputs, even if they can theoretically be deduced from the trajectories, could allow obtaining prediction improvement, especially in the case of low amount of data for the training phase. We can already observe a trend to use more inputs and notably the speed difference [145, 216, 124] and other hidden variables estimated by training [76, 6]. One may expect to include further variables as inputs. The time-to-collision and the time gap, whose roles are fundamental in KB pedestrian dynamic models [106, 158], are promising candidates. Such variables contain information about the relevance of the interactions and may even substitute the interaction or attention modules of DL approaches.

Bibliography

[1] Forschungszentrum Jülich, Jülich Supercomputing Centre. Data archive of experiments on pedestrian dynamic. Accessed: 2021-08-01.

[2] Google Scholar. Accessed: 2021-12-31.

[3] Karam M Abughalieh and Shadi G Alawneh. Predicting pedestrian intention to cross the road. *IEEE Access*, 8:72558–72569, 2020.

[4] Lina Achaji, Julien Moreau, Thibault Fouqueray, François Aioun, and François Charpillet. Is attention to bounding boxes all you need for pedestrian action prediction? *CoRR*, abs/2107.08031, 2021.

[5] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social LSTM: Human trajectory prediction in crowded spaces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 961–971, 2016.

[6] Javad Amirian, Jean-Bernard Hayet, and Julien Pettré. Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 2964–2972, 2019.

[7] Alessandro Antonucci, Gastone Pietro Rosati Papini, Luigi Palopoli, and Daniele Fontanelli. Generating reliable and efficient predictions of human motion: A promising encounter between physics and neural networks. *arXiv preprint arXiv:2006.08429*, 2020.

[8] Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*, 2017.

[9] A. Azzouni and G. Pujolle. A long short-term memory recurrent neural network framework for network traffic matrix prediction. *arXiv preprint arXiv:1705.05690*, 2017.

[10] Mohammadhossein Bahari, Ismail Nejjar, and Alexandre Alahi. Injecting knowledge in data-driven vehicle trajectory predictors. *Transportation Research Part C: Emerging Technologies*, 128:103010, 2021.

[11] Nicolas Bain and Denis Bartolo. Dynamic response and hydrodynamics of polarized crowds. *Science*, 363(6422):46–49, 2019.

[12] Federico Bartoli, Giuseppe Lisanti, Lamberto Ballan, and Alberto Del Bimbo. Context-aware trajectory prediction. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 1941–1946. IEEE, 2018.

[13] Clemens Bechinger, Roberto Di Leonardo, Hartmut Löwen, Charles Reichhardt, Giorgio Volpe, and Giovanni Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88(4):045006, 2016.

[14] Nicola Bellomo and Christian Dogbe. On the modeling of traffic and crowds: A survey of models, speculations, and perspectives. *SIAM review*, 53(3):409–463, 2011.

[15] Nicola Bellomo, Benedetto Piccoli, and Andrea Tosin. Modeling crowd dynamics from a complex system viewpoint. *Math. Models Methods Appl. Sci.*, 22(supp02):1230004, 2012.

[16] E. Ben-Jacob. From snowflake formation to growth of bacterial colonies. Part II. Cooperative formation of complex colonial patterns. *Contemp. Phys.*, 38:205–241, 1997.

[17] Yoshua Bengio. *Learning deep architectures for AI*. Now Publishers Inc, 2009.

[18] Maren Bennewitz, Wolfram Burgard, Grzegorz Cielniak, and Sebastian Thrun. Learning motion patterns of people for compliant robot motion. *The International Journal of Robotics Research*, 24(1):31–48, 2005.

[19] Ariyan Bighashdel and Gijs Dubbelman. A survey on path prediction techniques for vulnerable road users: From traditional to deep-learning approaches. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 1039–1046. IEEE, 2019.

[20] Stephen Bitgood. An analysis of visitor circulation: Movement patterns and the general value principle. *Curator: The Museum Journal*, 49(4):463–475, 2006.

[21] V. J. Blue and J. L. Adler. Cellular automata microsimulation of bidirectional pedestrian flows. *J. Res. Rec.*, 1678:135–141, 2000.

[22] N. W. W. Bode and E. Ronchi. Statistical model fitting and model selection in pedestrian dynamics research. *Collective Dynamics*, 4:1–32, 2019.

[23] M. Boltes, J. Zhang, A. Tordeux, A. Schadschneider, and A. Seyfried. *Empirical Results of Pedestrian and Evacuation Dynamics*, pages 1–29. Springer, Berlin, Heidelberg, 2018.

[24] C. Burstedde, A. Kirchner, K. Klauck, A. Schadschneider, and J. Zittartz. Cellular automaton approach to pedestrian dynamics – applications. In *Pedestrian and Evacuation Dynamics*, pages 87–98, 2002.

[25] C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A*, 295:507–525, 2001.

[26] Keith T Butler, Daniel W Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine learning for molecular and materials science. *Nature*, 559(7715):547–555, 2018.

[27] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591–646, 2009.

[28] John Charlton, Luis Rene Montana Gonzalez, Steve Maddock, and Paul Richmond. Fast simulation of crowd collision avoidance. In *Computer Graphics International Conference*, pages 266–277. Springer, 2019.

[29] U. Chattaraj, A. Seyfried, and P. Chakroborty. Comparison of pedestrian fundamental diagram across cultures. *Adv. Complex Syst.*, 12(03):393–405, 2009.

[30] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 2180–2188, 2016.

[31] Xu Chen, Martin Treiber, Venkatesan Kanagaraj, and Haiying Li. Social force models for pedestrian traffic–state of the art. *Transp. Rev.*, 38(5):625–653, 2018.

[32] Yu Fan Chen, Michael Everett, Miao Liu, and Jonathan P How. Socially aware motion planning with deep reinforcement learning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1343–1350. IEEE, 2017.

[33] Hao Cheng, Fatema T Johora, Monika Sester, and Jörg P Müller. Trajectory modelling in shared spaces: Expert-based vs. deep learning approach? In

International Workshop on Multi-Agent Systems and Agent-Based Simulation, pages 13–27. Springer, 2020.

[34] Hao Cheng, Wentong Liao, Michael Ying Yang, Bodo Rosenhahn, and Monika Sester. AMENet: Attentive maps encoder network for trajectory prediction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 172:253–266, 2021.

[35] Guilhem Chéron, Ivan Laptev, and Cordelia Schmid. P-cnn: Pose-based cnn features for action recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 3218–3226, 2015.

[36] D. Chowdhury, L. Santen, and A. Schadschneider. Statistical physics of vehicular traffic and some related systems. *Phys. Rep.*, 329(4-6):199–329, 2000.

[37] M. Chraibi, A. Seyfried, and A. Schadschneider. Generalized centrifugal-force model for pedestrian dynamics. *Phys. Rev. E*, 82(4):046111, 2010.

[38] M. Chraibi, A. Tordeux, A. Schadschneider, and A. Seyfried. *Modelling of Pedestrian and Evacuation Dynamics*, pages 1–22. Springer Berlin Heidelberg, 2018.

[39] Mohcine Chraibi, Ulrich Kemloh, Andreas Schadschneider, and Armin Seyfried. Force-based models of pedestrian dynamics. *Netw. Heterog. Media*, 6(3):425, 2011.

[40] Frank Cichos, Kristian Gustavsson, Bernhard Mehlig, and G. Volpe. Machine learning for active matter. *Nat. Mach. Intell.*, 2(2):94–103, 2020.

[41] Julien Cividini, Cécile Appert-Rolland, and Hendrik-Jan Hilhorst. Diagonal patterns and chevron effect in intersecting traffic flows. *EPL*, 102(2):20002, 2013.

[42] Alessandro Corbetta, Adrian Muntean, and Kiamars Vafayi. Parameter estimation of social forces in pedestrian dynamics models via a probabilistic method. *Mathematical biosciences and engineering*, 12(2):337–356, 2015.

[43] Emiliano Cristiani, Benedetto Piccoli, and Andrea Tosin. *Multiscale modeling of pedestrian dynamics*, volume 12. Springer, 2014.

[44] J. Cristìn, V. Méndez, and D. Campos. General scaling in bidirectional flows of self-avoiding agents. *Sci. Rep.*, 9:18488, 2019.

[45] Sean Curtis, Andrew Best, and Dinesh Manocha. Menge: A modular framework for simulating crowd movement. *Collective Dynamics*, 1, 2016.

[46] Xiong Dan. Spatial-temporal block and LSTM network for pedestrian trajectories prediction. *arXiv preprint arXiv:2009.10468*, 2020.

[47] F. Dietrich and G. Köster. Gradient navigation model for pedestrian dynamics. *Phys. Rev. E*, 89:062801, 2014.

[48] Hairong Dong, Min Zhou, Qianling Wang, Xiaoxia Yang, and Fei-Yue Wang. State-of-the-art pedestrian and evacuation dynamics. *IEEE Trans. Intell. Transp. Syst.*, 21(5):1849–1866, 2020.

[49] D. C. Duives, W. Daamen, and S. P. Hoogendoorn. State-of-the-art crowd motion simulation models. *Transp. Res. Part C:Emerg. Technol.*, 37:193–209, 2013.

[50] Austin R. Dulaney and John F. Brady. Machine learning for phase behavior in active matter systems. *Soft Matter*, 17(28):6808–6816, 2021.

[51] J Elgeti, R G Winkler, and G Gompper. Physics of microswimmers—single particle motion and collective behavior: a review. *Reports on Progress in Physics*, 78(5):056601, 2015.

[52] David Ellis, Eric Sommerlade, and Ian Reid. Modelling pedestrian trajectory patterns with gaussian processes. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 1229–1234, 2009.

[53] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017.

[54] Michael Everett, Yu Fan Chen, and Jonathan P How. Collision avoidance in pedestrian-rich environments with deep reinforcement learning. *IEEE Access*, 9:10357–10377, 2021.

[55] Matteo Fabbri, Fabio Lanzi, Simone Calderara, Andrea Palazzi, Roberto Vezzani, and Rita Cucchiara. Learning to detect and track visible and occluded body joints in a virtual world. In *Proceedings of the European conference on computer vision (ECCV)*, pages 430–446, Cham, 2018. Springer International Publishing.

[56] James H Faghmous, Ivy Frenger, Yuanshun Yao, Robert Warmka, Aron Lindell, and Vipin Kumar. A daily global mesoscale ocean eddy dataset from satellite altimetry. *Sci. Data*, 2(1):1–16, 2015.

[57] Claudio Feliciani and Katsuhiro Nishinari. An improved cellular automata model to simulate the behavior of high density crowd and validation by experimental data. *Physica A*, 451:135–148, 2016.

[58] Sarah Ferguson, Brandon Luders, Robert C Grande, and Jonathan P How. Real-time predictive modeling and robust avoidance of pedestrians with uncertain, changing intentions. In *Algorithmic Foundations of Robotics XI*, pages 161–177. Springer, 2015.

[59] Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. Soft+ hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection. *Neural Netw.*, 108:466–478, 2018.

[60] Christopher C Fischer, Kevin J Tibbetts, Dane Morgan, and Gerbrand Ceder. Predicting crystal structure by merging data mining with quantum mechanics. *Nat. Mater.*, 5(8):641–646, 2006.

[61] Martin Friesen, Hanno Gottschalk, Barbara Rüdiger, and Antoine Tordeux. Spontaneous wave formation in stochastic self-driven particle systems. *SIAM J. Appl. Math.*, 81(3):853–870, 2021.

[62] Jörg Frochte. *Maschinelles Lernen: Grundlagen und Algorithmen in Python*. Carl Hanser Verlag GmbH Co KG, 2019.

[63] John Joseph Fruin. *Designing for pedestrians a level of service concept*. Polytechnic University, 1970.

[64] M. Fukui and Y. Ishibashi. Self-organized phase transitions in cellular automaton models for pedestrians. *J. Phys. Soc. Jpn.*, 68:2861, 1999.

[65] Dietrich Garbrecht. Describing pedestrian and car trips by transition matrices. *Traffic Quarterly*, 27(1), 1973.

[66] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

[67] Francesco Giuliari, Irtiza Hasan, Marco Cristani, and Fabio Galasso. Transformer networks for trajectory forecasting. In *2020 25th international conference on pattern recognition (ICPR)*, pages 10335–10342. IEEE, 2021.

[68] G. H. Goldsztein. Self-organization when pedestrians move in opposite directions. multi-lane circular track model. *Appl. Sci.*, 10(2), 2020.

[69] Susana N Gomes, Andrew M Stuart, and Marie-Therese Wolfram. Parameter estimation for macroscopic pedestrian dynamics models from microscopic data. *SIAM J. Appl. Math.*, 79(4):1475–1500, 2019.

[70] Simone Göttlich and Stephan Knapp. Artificial neural networks for the estimation of pedestrian interaction forces. In *Crowd Dynamics, Volume 2*, pages 11–32. Springer, 2020.

[71] Simone Göttlich and Claudia Totzeck. Parameter calibration with stochastic gradient descent for interacting particle systems driven by neural networks. *Mathematics of Control, Signals, and Systems*, pages 1–30, 2021.

[72] Alex Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.

[73] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. Ieee, 2013.

[74] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[75] Ke Guo, Dawei Wang, Tingxiang Fan, and Jia Pan. VR-ORCA: Variable responsibility optimal reciprocal collision avoidance. *IEEE Robot. Autom. Lett.*, 6(3):4520–4527, 2021.

[76] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social GAN: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2255–2264, 2018.

[77] S. Guy, J. Chhugani, C. Kim, S. Satish, M. Lin, D. Manocha, and P. Dubey. Clearpath: Highly parallel collision avoidance for multi-agent simulation. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation 2009, New Orleans, LA, USA*, pages 177–187, 2009.

[78] S. Guy, M. Lin, and D. Manocha. Modeling collision avoidance behavior for virtual humans. *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, Toronto, ON, Canada*, 2:575–582, 2010.

[79] Sirin Haddad and Siew-Kei Lam. Self-growing spatial graph network for context-aware pedestrian trajectory prediction. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1029–1033, 2021.

[80] Sirin Haddad, Meiqing Wu, He Wei, and Siew Kei Lam. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model. In *24th Computer Vision Winter Workshop*, pages 4–16, 2019.

[81] BD Hankin and Richard A Wright. Passenger flow in subways. *Journal of the Operational Research Society*, 9(2):81–88, 1958.

[82] Irtiza Hasan, Francesco Setti, Theodore Tsesmelis, Vasileios Belagiannis, Sikandar Amin, Alessio Del Bue, Marco Cristani, and Fabio Galasso. Forecasting people trajectories and head poses by jointly reasoning on tracklets and vislets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(4):1267–1278, 2019.

[83] Irtiza Hasan, Francesco Setti, Theodore Tsesmelis, Alessio Del Bue, Fabio Galasso, and Marco Cristani. MX-LSTM: mixing tracklets and vislets to jointly forecast trajectories and head poses. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6067–6076, 2018.

[84] Geoffroy Hautier, Christopher C Fischer, Anubhav Jain, Tim Mueller, and Gerbrand Ceder. Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chem. Mater.*, 22(12):3762–3767, 2010.

[85] D Helbing. Traffic and related self-driven many-particle systems. *Phys. Mod. Phys.*, 73(4):1067–1141, 2001.

[86] D. Helbing, L. Buzna, A. Johansson, and T. Werner. Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transp. Sci.*, 39(1):1–24, 2005.

[87] D. Helbing, I. Farkas, P. Molnár, and T. Vicsek. Simulation of pedestrian crowds in normal and evacuation situations. In *Pedestrian and Evacuation Dynamics*, pages 21–58, 2002.

[88] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407:487–490, 2000.

[89] D. Helbing, I.J. Farkas, and T. Vicsek. Freezing by heating in a driven mesoscopic system. *Phys. Rev. Lett.*, 84(6):1240–1243, 2000.

[90] D. Helbing and A. Johansson. *Pedestrian, Crowd and Evacuation Dynamics*, pages 6476–6495. Springer New York, 2009.

[91] D. Helbing and P. Molnár. Social force model for pedestrian dynamics. *Phys. Rev. E*, 51(5):4282–4286, 1995.

[92] L. F. Henderson. The statistics of crowd fluids. *Nature*, 229:381–383, 1971.

[93] L. F. Henderson. On the fluid mechanics of human crowd motion. *Transportation Research*, 8:509–515, 1974.

[94] G Hermann and J Touboul. Heterogeneous connections induce oscillations in large-scale networks. *Phys. Rev. Lett.*, 109:018702, 2012.

[95] K Hirai and K Tarui. A simulation of the behavior of a crowd in panic. In *Proceedings of the 1975 International Conference on Cybernetics and Society*, pages 409–411, 1975.

[96] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[97] Stefan Holl, Maik Boltes, and Armin Seyfried. Level-of-Safety-Konzept für den Fußverkehr bei Großveranstaltungen. In *Veranstaltungskommunikation*, pages 253–277. Springer, 2019.

[98] Serge P Hoogendoorn, Femke LM van Wageningen-Kessels, Winnie Daamen, and Dorine C Duives. Continuum modelling of pedestrian flows: From microscopic principles to self-organised macroscopic phenomena. *Physica A*, 416:684–694, 2014.

[99] Sakif Hossain, Fatema T Johora, Jörg P Müller, Sven Hartmann, and Andreas Reinhardt. SFMGNet: A physics-based neural network to predict pedestrian trajectories. *arXiv preprint arXiv:2202.02791*, 2022.

[100] Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, Eduard Hovy, and Eric Xing. Harnessing deep neural networks with logic rules. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2410–2420, Berlin, Germany, 2016. Association for Computational Linguistics.

[101] Lei Huang, Jihui Zhuang, Xiaoming Cheng, Riming Xu, and Hongjie Ma. STI-GAN: Multimodal pedestrian trajectory prediction using spatiotemporal interactions and a generative adversarial network. *IEEE Access*, 9:50846–50856, 2021.

[102] Yingfan Huang, HuiKun Bi, Zhaoxin Li, Tianlu Mao, and Zhaoqi Wang. STGAT: Modeling spatial-temporal interactions for human trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6272–6281, 2019.

[103] Roger L Hughes. A continuum theory for the flow of pedestrians. *Transp. Res. Part B Meth.*, 36(6):507–535, 2002.

[104] Boris Ivanovic, Karen Leung, Edward Schmerling, and Marco Pavone. Multimodal deep generative models for trajectory prediction: A conditional variational autoencoder approach. *IEEE Robotics and Automation Letters*, 6(2):295–302, 2020.

[105] Ajay Jain, Sergio Casas, Renjie Liao, Yuwen Xiong, Song Feng, Sean Segal, and Raquel Urtasun. Discrete residual flow for probabilistic pedestrian behavior prediction. In *Conference on Robot Learning*, pages 407–419. PMLR, 2020.

[106] I. Karamouzas, B. Skinner, and S. J. Guy. A universal power law governing pedestrian interactions. *Phys. Rev. Lett.*, 113(5):238701, 2014.

[107] Anuj Karpatne, Gowtham Atluri, James H Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Trans. Knowl. Data Eng.*, 29(10):2318–2331, 2017.

[108] Jaya Kawale, Stefan Liess, Arjun Kumar, Michael Steinbach, Peter Snyder, Vipin Kumar, Auroop R Ganguly, Nagiza F Samatova, and Fredrick Semazzi. A graph-based approach to find teleconnections in climate data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 6(3):158–179, 2013.

[109] Christoph G Keller and Dariu M Gavrila. Will the pedestrian cross? a study on pedestrian path prediction. *IEEE Transactions on Intelligent Transportation Systems*, 15(2):494–506, 2013.

[110] A. Khadka, P. Remagnino, and V. Argyriou. Synthetic crowd and pedestrian generator for deep learning problems. In *ICASSP 2020 - 2020*

IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 4052–4056, 2020.

[111] Peter Kielar and André Borrmann. An artificial neural network framework for pedestrian walking behavior modeling and simulation. *Collective Dynamics*, 5:290–298, 2020.

[112] Kihwan Kim, Dongryeol Lee, and Irfan Essa. Gaussian process regression flow for analysis of motion trajectories. In *2011 International Conference on Computer Vision*, pages 1164–1171. IEEE, 2011.

[113] S. Kim, G. Stephen, S. Guy, W. Liu, D. Wilkie, R. Lau, M. Lin, and D. Manocha. BRVO: Predicting pedestrian trajectories using velocity-space reasoning. *Int. J. Robot. Res.*, 34:201–217, 2014.

[114] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[115] A. Kirchner and A. Schadschneider. Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics. *Physica A*, 312:260–276, 2002.

[116] Ansgar Kirchner, Katsuhiro Nishinari, and Andreas Schadschneider. Friction effects and clogging in a cellular automaton model for pedestrian dynamics. *Phys. Rev. E*, 67:056122, 2003.

[117] Kris M Kitani, Brian D Ziebart, James Andrew Bagnell, and Martial Hebert. Activity forecasting. In *European conference on computer vision*, pages 201–214. Springer, 2012.

[118] Benedikt Kleinmeier, Benedikt Zönnchen, Marion Gödel, and Gerta Köster. Vadere: An open-source simulation framework to promote interdisciplinary understanding. *Collective Dynamics*, 4, 2019.

[119] H. Klüpfel, T. Meyer-König, J. Wahle, and M. Schreckenberg. Microscopic simulation of evacuation processes on passenger ships. In *Theory and Practical Issues on Cellular Automata*, Berlin Heidelberg, 2000. Springer.

[120] Moonsoo Ko, Taewan Kim, and Keemin Sohn. Calibrating a social-force-based pedestrian walking model based on maximum likelihood estimation. *Transportation*, 40(1):91–107, 2013.

[121] T. Korhonen, S. Hostikka, S. Heliövaara, and H. Ehtamo. FDS+Evac: An agent based fire evacuation model. In *Pedestrian and Evacuation Dynamics 2008*, pages 109–120. 2010.

[122] Vineet Kosaraju, Amir Sadeghian, Roberto Martín-Martín, Ian Reid, S Hamid Rezatofighi, and Silvio Savarese. Social-BiGAT: Multimodal trajectory forecasting using bicycle-gan and graph attention networks. *arXiv preprint arXiv:1907.03395*, 2019.

[123] G. Köster, F. Treml, and M. Gödel. Avoiding numerical pitfalls in social force models. *Phys. Rev. E*, 87:063305, 2013.

[124] Parth Kothari, Sven Kreiss, and Alexandre Alahi. Human trajectory forecasting in crowds: A deep learning perspective. *IEEE Trans. Intell. Transp. Syst.*, pages 1–15, 2021.

[125] Parth Kothari, Brian Sifringer, and Alexandre Alahi. Interpretable social anchors for human trajectory forecasting in crowds. *arXiv preprint arXiv:2105.03136*, 2021.

[126] Sven Kreiss. Deep social force. *arXiv preprint arXiv:2109.12081*, 2021.

[127] Henrik Kretzschmar, Markus Spies, Christoph Sprunk, and Wolfram Burgard. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research*, 35(11):1289–1307, 2016.

[128] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. *Adv. Neural. Inf. Process. Syst.*, 25:1097–1105, 2012.

[129] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. Human-aware robot navigation: A survey. *Rob. Auton. Syst.*, 61(12):1726–1743, 2013.

[130] Wei-Cheng Lai, Zi-Xiang Xia, Hao-Siang Lin, Lien-Feng Hsu, Hong-Han Shuai, I-Hong Jhuo, and Wen-Huang Cheng. Trajectory prediction in heterogeneous environment via attended ecology embedding. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 202–210, 2020.

[131] Truc Viet Le, Siyuan Liu, and Hoong Chuin Lau. A reinforcement learning framework for trajectory prediction under uncertainty and budget constraint. In *ECAI*, pages 347–354, 2016.

[132] Jaedong Lee, Jungdam Won, and Jehee Lee. Crowd simulation by deep reinforcement learning. In *Proceedings of the 11th Annual International Conference on Motion, Interaction, and Games*, pages 1–7, 2018.

[133] Kuan-Hui Lee, German Ros, Jie Li, and Adrien Gaidon. Spigan: Privileged adversarial learning from simulation. *arXiv preprint arXiv:1810.03756*, 2018.

[134] Namhoon Lee, Wongun Choi, Paul Vernaza, Christopher B. Choy, Philip H. S. Torr, and Manmohan Chandraker. DESIRE: Distant future prediction in dynamic scenes with interacting agents. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[135] Marco Leo, G Medioni, M Trivedi, Takeo Kanade, and Giovanni Maria Farinella. Computer vision for assistive technologies. *Comput. Vis. Image Underst.*, 154:1–15, 2017.

[136] Florin Leon and Marius Gavrilescu. A review of tracking and trajectory prediction methods for autonomous driving. *Mathematics*, 9(6), 2021.

[137] Li Li, John C Snyder, Isabelle M Pelaschier, Jessica Huang, Uma-Naresh Niranjan, Paul Duncan, Matthias Rupp, Klaus-Robert Müller, and Kieron Burke. Understanding machine-learned density functionals. *Int. J. Quantum Chem.*, 116(11):819–833, 2016.

[138] Meifang Li, Yongxiang Zhao, Lerong He, Wenxiao Chen, and Xianfeng Xu. The parameter calibration and optimization of social force model for the real-life 2013 Ya'an earthquake evacuation in China. *Saf. Sci.*, 79:243–253, 2015.

[139] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao. Independently recurrent neural network (IndRNN): Building a longer and deeper rnn. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5457–5466, 2019.

[140] Matteo Lisotto, Pasquale Coscia, and Lamberto Ballan. Social and scene-aware trajectory prediction in crowded spaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2019.

[141] Gunnar G Løvås. Modeling and simulation of pedestrian traffic flow. *Transp. Res. Part B Meth.*, 28(6):429–443, 1994.

[142] Ruggiero Lovreglio, Enrico Ronchi, and Daniel Nilsson. Calibrating floor field cellular automaton models for pedestrian dynamics by using likelihood function optimization. *Physica A: Statistical Mechanics and its Applications*, 438:308–320, 2015.

[143] Yi-Xuan Lü, Zhi-Xi Wu, and Jian-Yue Guan. Pedestrian dynamics with mechanisms of anticipation and attraction. *Phys. Rev. Research*, 2:043250, 2020.

[144] Zezheng Lv, Xiaoci Huang, and Wenguan Cao. An improved gan with transformers for pedestrian trajectory prediction models. *International Journal of Intelligent Systems*, 2021.

[145] Y. Ma, E. W. M. Lee, and R. K. K. Yuen. An artificial intelligence-based approach for simulating pedestrian movement. *IEEE Trans. Intell. Transp. Syst.*, 17(11):3159–3170, 2016.

[146] Huynh Manh and Gita Alaghband. Scene-LSTM: A model for human trajectory prediction. *arXiv preprint arXiv:1808.04018*, 2018.

[147] M Cristina Marchetti, Jean-François Joanny, Sriram Ramaswamy, Tanniemola B Liverpool, Jacques Prost, Madan Rao, and R Aditi Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85(3):1143, 2013.

[148] Francisco Martinez-Gil, Miguel Lozano, and Fernando Fernández. Multi-agent reinforcement learning for simulating pedestrian navigation. In *International Workshop on Adaptive and Learning Agents*, pages 54–69. Springer, 2011.

[149] Francisco Martinez-Gil, Miguel Lozano, and Fernando Fernández. Marl-ped: A multi-agent reinforcement learning based framework to simulate pedestrian groups. *Simulation Modelling Practice and Theory*, 47:259–275, 2014.

[150] Francisco Martinez-Gil, Miguel Lozano, Ignacio García-Fernández, and Fernando Fernández. Modeling, evaluation, and scale on artificial pedestrians: a literature review. *ACM Computing Surveys (CSUR)*, 50(5):1–35, 2017.

[151] Francisco Martinez-Gil, Miguel Lozano, Ignacio García-Fernández, Pau Romero, Dolors Serra, and Rafael Sebastián. Using inverse reinforcement learning with real trajectories to get more trustworthy pedestrian simulations. *Mathematics*, 8(9):1479, 2020.

[152] B. Maury and J. Venel. A discrete contact model for crowd motion. *Esaim Math. Model. Numer. Anal.*, 45:145–168, 2011.

[153] B. Maury, J. Venel, A.-H. Olivier, and S. Donikian. A mathematical framework for a crowd motion model. *Comptes Rendus Math.*, 346:1245–1250, 2008.

[154] Mirko Meuter, Uri Iurgel, Su-Birm Park, and Anton Kummert. The unscented kalman filter for pedestrian tracking from a moving host. In *2008 IEEE Intelligent Vehicles Symposium*, pages 37–42. IEEE, 2008.

[155] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14424–14432, 2020.

[156] Alessio Monti, Alessia Bertugli, Simone Calderara, and Rita Cucchiara. Dag-net: Double attentive graph neural network for trajectory forecasting. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2551–2558, 2021.

[157] M. Moussaïd, E.G. Guillot, M. Moreau, J. Fehrenbach, O. Chabiron, S. Lemercier, J. Pettré, C. Appert-Rolland, P. Degond, and G. Theraulaz. Traffic instabilities in self-organized pedestrian crowds. *PLOS Comput. Biol.*, 8(3):1–10, 2012.

[158] Mehdi Moussaïd, Dirk Helbing, and Guy Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *Proc. Natl. Acad. Sci. U.S.A.*, 108(17):6884–6888, 2011.

[159] Mehdi Moussaïd, Juliane E. Kämmer, Pantelis P. Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PLOS ONE*, 8(11):1–8, 2013.

[160] M. Muramatsu, T. Irie, and T. Nagatani. Jamming transition in pedestrian counter flow. *Physica A*, 267:487–498, 1999.

[161] A. Nakayama, K. Hasebe, and Y. Sugiyama. Instability of pedestrian flow and phase structure in a two-dimensional optimal velocity model. *Phys. Rev. E*, 71:036121, 2005.

[162] Francis P Navin and Robert J Wheeler. Pedestrian flow characteristics. *Traffic Engineering, Inst Traffic Engr*, 39, 1969.

[163] Daniel Needleman and Zvonimir Dogic. Active matter at the interface between materials science and cell biology. *Nat. Rev. Mater.*, 2(9):1–14, 2017.

[164] Alexandre Nicolas, Santiago Ibáñez, Marcelo N Kuperman, and Sebastián Bouzat. A counterintuitive way to speed up pedestrian and granular bottleneck flows prone to clogging: Can 'more'escape faster? *Journal of Statistical Mechanics: Theory and Experiment*, 2018(8):083403, 2018.

[165] Michael A Nielsen. *Neural networks and deep learning*, volume 25. Determination press San Francisco, CA, 2015.

[166] Nishant Nikhil and Brendan Tran Morris. Convolutional neural network for trajectory prediction. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 186–196, Cham, 2018. Springer International Publishing.

[167] D. Oeding. Verkehrsbelastung und Dimensionierung von Gehwegen und anderen Anlagen des Fußgängerverkehrs. Forschungsbericht 22, Technische Hochschule Braunschweig, 1963.

[168] SJ Older. Movement of pedestrians on footways in shopping streets. *Traffic Eng. Control.*, 10(4), 1968.

[169] J. Ondřej and J. Pettré. A synthetic-vision based steering approach for crowd simulation. *ACM Trans. Graph.*, 29:123, 2010.

[170] Abbas Ourmazd. Science in the age of machine learning. *Nat. Rev. Phys.*, 2(7):342–343, 2020.

[171] S. Paris, J. Pettré, and S. Donikian. Pedestrian reactive navigation for crowd simulation: a predictive approach. *Computer Graphics Forum*, 26:665–674, 2007.

[172] Frank Pasquale. Toward a fourth law of robotics: Preserving attribution, responsibility, and explainability in an algorithmic society. *Ohio St. LJ*, 78:1243, 2017.

[173] Dino Pedreschi, Fosca Giannotti, Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, and Franco Turini. Meaningful explanations of black box AI decision systems. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33(1), pages 9780–9784, 2019.

[174] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool. You'll Never Walk Alone: Modeling Social Behavior for Multi-target Tracking. In *2009 IEEE 12th International Conference on Computer Vision*, pages 261–268. IEEE, 2009.

[175] Mark Pfeiffer, Giuseppe Paolo, Hannes Sommer, Juan Nieto, Rol Siegwart, and Cesar Cadena. A data-driven model for interaction-aware pedestrian motion prediction in object cluttered environments. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5921–5928. IEEE, 2018.

[176] Atanas Poibrenski, Matthias Klusch, Igor Vozniak, and Christian Müller. M2P3: multimodal multi-pedestrian path prediction by self-driving cars with egocentric vision. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pages 190–197, 2020.

[177] Abishai Polus, Joseph L Schofer, and Ariela Ushpiz. Pedestrian flow and level of service. *J. Transp. Eng.*, 109(1):46–56, 1983.

[178] A. Portz and A. Seyfried. Analyzing stop-and-go waves by experiment and modeling. In *Pedestrian and Evacuation Dynamics*, pages 577–586, 2011.

[179] Sriram Ramaswamy. The mechanics and statistics of active matter. *Annual Review of Condensed Matter Physics*, 1(1):323–345, 2010.

[180] Eike Rehder, Florian Wirth, Martin Lauer, and Christoph Stiller. Pedestrian prediction by planning using deep neural networks. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5903–5908. IEEE, 2018.

[181] Kristofer G Reyes and Benji Maruyama. The machine learning revolution in materials? *MRS Bull.*, 44(7):530–537, 2019.

[182] Nicholas Rhinehart and Kris M Kitani. First-person activity forecasting from video with online inverse reinforcement learning. *IEEE transactions on pattern analysis and machine intelligence*, 42(2):304–317, 2018.

[183] Daniela Ridel, Nachiket Deo, Denis Wolf, and Mohan Trivedi. Scene compliant trajectory forecast with agent-centric spatio-temporal grids. *IEEE Robot. Autom. Lett.*, 5(2):2816–2823, 2020.

[184] Ergys Ristani and Carlo Tomasi. Features for multi-target multi-camera tracking and re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6036–6046, 2018.

[185] Th Robin, Gianluca Antonini, Michel Bierlaire, and Javier Cruz. Specification, estimation and validation of a pedestrian walking behavior model. *Transp. Res. Part B Meth.*, 43(1):36–56, 2009.

[186] Ivan Rodin, Antonino Furnari, Dimitrios Mavroedis, and Giovanni Maria Farinella. Predicting the future from first person (egocentric) vision: A survey. *Comput. Vis. Image Underst.*, page 103252, 2021.

[187] A. Rudenke, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras. Human motion trajectory prediction: a survey. *Int. J. Robot. Res.*, 39(8):895–935, 2020.

[188] Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M Kitani, Dariu M Gavrila, and Kai O Arras. Human motion trajectory prediction: A survey. *Int. J. Robot. Res.*, 39(8):895–935, 2020.

[189] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1349–1358, 2019.

[190] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. *Adv. Neural Inf. Process. Syst.*, 29:2234–2242, 2016.

[191] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *Computer Vision – ECCV 2020*, pages 683–700, Cham, 2020. Springer International Publishing.

[192] Gilbert Saporta. Models for understanding versus models for prediction. In *COMPSTAT 2008*, pages 315–322. Springer, 2008.

[193] A. Schadschneider. Cellular automaton approach to pedestrian dynamics – theory. In *Pedestrian and Evacuation Dynamics*, pages 75–86, 2002.

[194] A. Schadschneider, M. Chraibi, A. Seyfried, A. Tordeux, and J. Zhan. *Pedestrian Dynamics - From Empirical Results to Modeling*. Birkhäuser, Cham, 2018.

[195] A. Schadschneider, W. Klingsch, H. Klüpfel, T. Kretz, C. Rogsch, and A. Seyfried. Evacuation dynamics: Empirical results, modeling and applications. In *Encyclopedia of Complexity and Systems Science*, pages 3142–3176. Springer New York, 2009.

[196] Andreas Schadschneider and Armin Seyfried. Validation of CA models of pedestrian dynamics with fundamental diagrams. *Cybern. Syst.*, 40(5):367–389, 2009.

[197] Stefano Scheggi, Marco Aggravi, Fabio Morbidi, and Domenico Prattichizzo. Cooperative human-robot haptic navigation. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2693–2698. IEEE, 2014.

[198] Nicolas Schneider and Dariu M Gavrila. Pedestrian path prediction with recursive bayesian filters: A comparative study. In *German Conference on Pattern Recognition*, pages 174–183. Springer, 2013.

[199] Christoph Schöller, Vincent Aravantinos, Florian Lay, and Alois Knoll. What the constant velocity model can teach us about pedestrian motion prediction. *IEEE Robotics and Automation Letters*, 5(2):1696–1703, 2020.

[200] Kristof T Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus R Müller, and Alexandre Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature communications*, 8(1):1–8, 2017.

[201] Frank Schweitzer. *Brownian agents and active particles: collective dynamics in the natural and social sciences*. Springer Science & Business Media, 2003.

[202] Andrew W Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander WR Nelson, Alex Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020.

[203] A. Seyfried, B. Steffen, W. Klingsch, and M. Boltes. The fundamental diagram of pedestrian movement revisited. *J. Stat. Mech.: Theory Exp.*, 2005(10):P10002, 2005.

[204] Zahra Shahhoseini and Majid Sarvi. Collective movements of pedestrians: How we can learn from simple experiments with non-human (ant) crowds. *PLOS ONE*, 12:1–20, 2017.

[205] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019.

[206] Mattia Silvestri, Michele Lombardi, and Michela Milano. Injecting domain knowledge in neural networks: a controlled experiment on a constrained problem. In *International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, pages 266–282. Springer, 2021.

[207] Xiao Song, Kai Chen, Xu Li, Jinghan Sun, Baocun Hou, Yong Cui, Baochang Zhang, Gang Xiong, and Zilie Wang. Pedestrian trajectory prediction based on deep convolutional LSTM network. *IEEE Transactions on Intelligent Transportation Systems*, 22(6):3285–3302, 2020.

[208] H.E. Stanley. Non-equilibrium physics: Freezing by heating. *Nature*, 404:718, 2000.

[209] Rudina Subaih, Mohammed Maree, Mohcine Chraibi, Sami Awad, and Tareq Zanoon. Gender-based insights into the fundamental diagram of pedestrian dynamics. In *International Conference on Computational Collective Intelligence*, pages 613–624. Springer, 2019.

[210] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[211] P. Taillandier, B. Gaudou, A. Grignard, Q.-N. Huynh, N. Marilleau, P. Caillou, D. Philippon, and A. Drogoul. Building, composing and experimenting complex spatial models with the gama platform. *Geoinformatica*, 23:299–322, 2019.

[212] Chaofan Tao, Qinhong Jiang, Lixin Duan, and Ping Luo. Dynamic and static context-aware LSTM for multi-agent motion prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 547–563, Cham, 2020. Springer International Publishing.

[213] A. Tordeux, M. Chraibi, and A. Seyfried. Collision-free speed model for pedestrian dynamics. In *Traffic and Granular Flow '15*, pages 225–232, Cham, 2016. Springer.

[214] A. Tordeux and A. Schadschneider. White and relaxed noises in optimal velocity models for pedestrian flow with stop-and-go waves. *J. Phys. A: Math. Theor.*, 49:185101, 2016.

[215] Antoine Tordeux, Mohcine Chraibi, Armin Seyfried, and Andreas Schadschneider. Prediction of pedestrian speed with artificial neural networks. In *International Conference on Traffic and Granular Flow*, pages 327–335. Springer, 2017.

[216] Antoine Tordeux, Mohcine Chraibi, Armin Seyfried, and Andreas Schadschneider. Prediction of pedestrian dynamics in complex architectures with artificial neural networks. *J. Intell. Transp. Syst.*, 24(6):556–568, 2020.

[217] C. Totzeck. An anisotropic interaction model with collision avoidance. *Kinetic & Related Models*, 13(6):1219–1242, 2020.

[218] Jonathan D. Touboul. The hipster effect: When anti-conformists all look the same. *Discrete Continuous Dyn. Syst. Ser. B*, 24:4379, 2019.

[219] Pete Trautman, Jeremy Ma, Richard M Murray, and Andreas Krause. Robot navigation in dense human crowds: Statistical models and experimental studies of human–robot cooperation. *The International Journal of Robotics Research*, 34(3):335–356, 2015.

[220] Peter Trautman and Andreas Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 797–803. IEEE, 2010.

[221] Adrien Treuille, Seth Cooper, and Zoran Popović. Continuum crowds. *ACM Trans. Graph.*, 25(3):1160–1168, 2006.

[222] J. van den Berg, S. Guy, M. Lin, and D. Manocha. Reciprocal n-body collision avoidance. In *Robotics Research: The 14th International Symposium ISRR*, pages 3–19, Berlin, Heidelberg, 2011. Springer.

[223] Jur Van den Berg, Ming Lin, and Dinesh Manocha. Reciprocal velocity obstacles for real-time multi-agent navigation. In *2008 IEEE International Conference on Robotics and Automation*, pages 1928–1935. IEEE, 2008.

[224] Wouter van Toll and Julien Pettré. Algorithms for microscopic crowd simulation: Advancements in the 2010s. In *Computer Graphics Forum*, volume 40, pages 731–754. Wiley Online Library, 2021.

[225] Anirudh Vemula, Katharina Muelling, and Jean Oh. Social attention: Modeling attention in human crowds. In *2018 IEEE international Conference on Robotics and Automation (ICRA)*, pages 4601–4607. IEEE, 2018.

[226] T. Vicsek and A. Zafeiris. Collective motion. *Physics Reports*, 517(3):71–140, 2012.

[227] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164, 2015.

[228] Cornelia von Krüchten and Andreas Schadschneider. A cognitive, decision-based model for pedestrian dynamics. In *Traffic and Granular Flow 2019*, pages 141–147. Springer, 2020.

[229] Cornelia von Krüchten and Andreas Schadschneider. Concept of a decision-based pedestrian model. *Collective Dynamics*, 5:316–323, 2020.

[230] Laura von Rueden, Sebastian Mayer, Rafet Sifa, Christian Bauckhage, and Jochen Garcke. Combining machine learning and simulation to a hybrid modelling approach: Current and future directions. In *International Symposium on Intelligent Data Analysis*, pages 548–560. Springer, Cham, 2020.

[231] Zhiqiang Wan, Chao Jiang, Muhammad Fahad, Zhen Ni, Yi Guo, and Haibo He. Robot-assisted pedestrian regulation based on deep reinforcement learning. *IEEE transactions on cybernetics*, 50(4):1669–1682, 2018.

[232] U. Weidmann. Transporttechnik der Fußgänger - Transporttechnische Eigenschaften des Fußgängerverkehrs (Literaturauswertung) . Schriftenreihe des IVT 90, ETH Zürich, 1993.

[233] Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating scientific knowledge with machine learning for engineering and environmental systems. *arXiv preprint arXiv:2003.04919v5*, 2020.

[234] Ken CL Wong, Linwei Wang, and Pengcheng Shi. Active model with orthotropic hyperelastic material for cardiac image analysis. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 229–238. Springer, 2009.

[235] Markus Wulfmeier, Dominic Zeng Wang, and Ingmar Posner. Watch this: Scalable cost-function learning for path planning in urban environments. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2089–2095. IEEE, 2016.

[236] Jingjia Xu, John L Sapp, Azar Rahimi Dehaghani, Fei Gao, Milan Horacek, and Linwei Wang. Robust transmural electrophysiological imaging: Integrating sparse and dynamic physiological models into ECG-based inference. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 519–527. Springer, 2015.

[237] Qiancheng Xu, Mohcine Chraibi, Antoine Tordeux, and Jun Zhang. Generalized collision-free velocity model for pedestrian dynamics. *Physica A*, 535:122521, 2019.

[238] Yanyu Xu, Zhixin Piao, and Shenghua Gao. Encoding crowd interaction with deep neural network for pedestrian trajectory prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5275–5284, 2018.

[239] Hao Xue, Du Q Huynh, and Mark Reynolds. SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction. In *2018 IEEE Winter Conference on Applications of Computer Vision*, pages 1186–1194. IEEE, 2018.

[240] Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Thirty-second AAAI conference on artificial intelligence*, 2018.

[241] Shuai Yi, Hongsheng Li, and Xiaogang Wang. Pedestrian behavior understanding and prediction with deep neural networks. In *Computer Vision – ECCV 2016*, pages 263–279, Cham, 2016. Springer International Publishing.

[242] Zhipeng Yin, Jinghao Liu, and Lei Wang. Less-effort collision avoidance in virtual pedestrian simulation. In *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*, pages 488–493, 2019.

[243] Young-chul Yoon, Abhijeet Boragule, Young-min Song, Kwangjin Yoon, and Moongu Jeon. Online multi-object tracking with historical appearance matching and scene adaptive detection filtering. In *2018 15th IEEE International conference on advanced video and signal based surveillance (AVSS)*, pages 1–6. IEEE, 2018.

[244] W. Yu, R. Chen, L. Dong, and S. Dai. Centrifugal force model for pedestrian dynamics. *Phys. Rev. E*, 72(2):026112, 2005.

[245] Simone Zamboni, Zekarias Tilahun Kefato, Sarunas Girdzijauskas, Christoffer Norén, and Laura Dal Col. Pedestrian trajectory prediction with convolutional neural networks. *Pattern Recognition*, 121:108252, 2022.

[246] Jun Zhang, Wolfram Klingsch, Andreas Schadschneider, and Armin Seyfried. Ordering in bidirectional pedestrian flows and its influence on the fundamental diagram. *J. Stat. Mech.: Theory Exp.*, 2012(02):P02002, 2012.

[247] Zhe Zhang and Limin Jia. Direction-decision learning based pedestrian flow behavior investigation. *IEEE Access*, 8:15027–15038, 2020.

[248] Zheng Zhang, Dianjie Lu, Jialiuyuan Li, Pingshan Liu, and Guijuan Zhang. Crowd evacuation simulation using hierarchical deep reinforcement learning. In *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 563–568. IEEE, 2021.

[249] Xiaoping Zheng, Tingkuan Zhong, and Mengting Liu. Modeling crowd evacuation of a building based on seven methodological approaches. *Build. Environ.*, 44(3):437–445, 2009.

II. Predicting pedestrian trajectories at different densities: A multi-criteria empirical analysis

Outline

Bibliographic Information

Korbmacher, R., Dang, H. T., & Tordeux, A. (2024). Predicting pedestrian trajectories at different densities: A multi-criteria empirical analysis. Physica A: Statistical Mechanics and its Applications, 634, 129440. https://www.sciencedirect.com.

Author's contribution

Conceptualization, R.K. and A.T.; methodology, R.K. and A.T.; software, R.K.; validation, R.K. and A.T.; formal analysis, R.K.; investigation, R.K.; resources, R.K.; data curation, R.K.; writing—original draft preparation, R.K.; writing—review and editing, R.K. and A.T.; visualization, R.K.; supervision, A.T.; project administration, A.T.; funding acquisition, A.T. All authors have read and agreed to the published version of the manuscript.

Copyright Notice

Predicting pedestrian trajectories at different densities: A multi-criteria empirical analysis.

Raphael Korbmacher[1], Huu Tu-Dang[2], and Antoine Tordeux[1]

[1]School of Mechanical Engineering and Safety Engineering, University of Wuppertal, Gaußstr. 20, Germany
[2]UMR 5505 IRIT, Université Toulouse Capitole, France

Abstract

Predicting human trajectories is a challenging task due to the complexity of pedestrian behavior, which is influenced by external factors such as the scene's topology and interactions with other pedestrians. A special challenge arises from the dependence of the behaviour on the density of the scene. In the literature, deep learning algorithms show the best performance in predicting pedestrian trajectories, but so far just for situations with low densities. In this study, we aim to investigate the suitability of these algorithms for high-density scenarios by evaluating them on different error metrics and comparing their accuracy to that of knowledge-based models that have been used since long time in the literature. The findings indicate that deep learning algorithms provide improved trajectory prediction accuracy in the distance metrics for all tested densities. Nevertheless, we observe a significant number of collisions in the predictions, especially in high-density scenarios. This issue arises partly due to the absence of a collision avoidance mechanism within the algorithms and partly because the distance-based collision metric is inadequate for dense situations. To address these limitations, we propose the introduction of a novel continuous collision metric based on pedestrians' time-to-collision. Subsequently, we outline how this metric can be utilized to enhance the training of the algorithms.

Keywords: Pedestrian trajectory prediction, deep learning, high-density, collision, time-to-collision

---

II.1 Introduction

The task of predicting pedestrian trajectories has emerged as a critical component in a variety of real-world applications, ranging from autonomous vehicles [44] and human-robot interactions [48] to the design of events, infrastructure, and buildings [6], especially in the case of evacuation [8]. This topic has been addressed within academia from the two distinct disciplinary perspectives of pedestrian dynamics and data science [28].

On one hand, the discipline of pedestrian dynamics applies a *knowledge-based* (KB) approach, developing mathematical models that encapsulate the inherent rules governing pedestrian behavior [13]. These models are utilized to conduct simulations in which pedestrian trajectories are computed. The difficulty lies in identifying fundamental mechanisms and parameters that induce realistic pedestrian behavior. On the other hand, computer scientists employ a *data-based* (DB) approach, collecting extensive data and training sophisticated algorithms intended to predict pedestrian trajectories. Their focus predominantly lies in devising efficient algorithm architectures and meticulously fine-tuning the hyperparameters of DB algorithms.

While KB models cater to a broad array of applications and include macroscopic, mesoscopic, and microscopic models [47], DB algorithms are primarily deployed for microscopic trajectory predictions in low-density scenes [28]. Low-density scenes denote situations with a medium pedestrian presence, where individuals possess a high degree of freedom and exhibit long-range interactions. This paper seeks to explore the efficacy of DB algorithms in high-density situations and compare the results with those derived from traditional KB models. One significant challenge inherent in such a comparison lies in devising a fair and comprehensive evaluation. While prior studies have demonstrated the superior performance of DB algorithms in terms of prediction accuracy, these evaluations have exclusively pertained to low-density data [1], focusing solely on distance metrics such as Average Displacement Error (ADE) [43] and Final Displacement Error (FDE) [34]. We propose to extend this evaluation by incorporating two additional metrics: a binary distance-based collision metric as proposed by Kothari et al. [30], and an original continuous time-to-collision-based metric.

Findings indicate that the DB algorithms surpass the KB models across all tested densities in terms of distance metrics. However, the DB algorithm predictions generate a significantly higher number of collisions when compared to the real trajectories and the KB models, which are typically designed with collision avoidance mechanisms.

II.2 Related work

This chapter contains three sections. Initially, we delve into KB models, exploring their intricacies and applications. Next, we examine DB algorithms, highlighting their significance and diferences to the KB models. The final section covers various other methods, selected for their relevance and contribution to this field. Together, these sections provide a thorough groundwork for the research presented in this study.

II.2.1 Knowledge-based models

KB models have a rich history in pedestrian dynamics that dates back to the middle of the 20th century. These models apply principles from physics, such as force fields and particles, to understand and predict the behavior of pedestrians. Currently, KB pedestrian models range from macroscopic, mesoscopic, and microscopic models among others modeling scale characteristics. Macroscopic and mesoscopic approaches are borrowed from continuous fluid dynamics or gas-kinetic models describing the dynamics at an aggregated level, while microscopic approaches model individual pedestrian motions [22]. For pedestrian trajectory predictions macroscopic and mesoscopic models are less relevant, which is why in the following we focus on the microscopic models.

In these models the individual pedestrian behavior is described according to certain rules and mechanisms ground on physical social, or psychological factors [14]. These rules and mechanisms are formulated in hand-crafted dynamic equations based on Newton's laws of motion. Given the input information about the initial status of the pedestrians like position, velocity, and acceleration a forward simulation of KB models can be used to predict the future trajectories. Depending on the modelling order of the model, they can be classified into *decision-based* (zeroth order), *velocity-based* (first order), and *acceleration-based* models (second order) [28]. In acceleration-based models, typically force-based models, the movement of pedestrians is defined by a superposition of exterior forces. Most acceleration-based models are force-based consisting of a relaxation term to the desired direction and an interaction term [28]. This last term is generally the sum of repulsion with the neighbours and obstacles. This is also the case in the most famous force-based model, the *Social Force model* (SF) from Helbing and Molnar [20], where the interaction force is an exponential gradient of a distance-based potential.

Velocity-based models are speed functions, depending on the position differences with neighbors and obstacles. In opposite to the acceleration-based models that are based on second-order differential equations, the velocity-based models rely on first-order equations. Many of these models are based on collision avoidance techniques and are formulated as optimisation problems on some ensemble of feasible trajectories devoid of collisions. The most famous models are the Reciprocal Velocity Obstacle model [52, 56] and the *Optimal Reciprocal Collision Avoidance* (ORCA) [51].

In the last class of models, the decision-based or rule-based models, the pedestrian behavior is not modeled based on differential equations, but on rules or decisions determining the new agent positions, velocities, etc [14]. The time is considered to be discrete for this class of decision-based models, which are typically Cellular automata [9, 35, 7, 27].

KB models in pedestrian dynamics encompass various approaches, including microscopic models, which can be used for trajectory predictions. Most of these models focus on the interactions between pedestrians and the environment and, as a result, are fundamentally based on collision avoidance mechanisms.

II.2.2 Data-based algorithms

The previously outlined approach is fundamentally grounded in a theoretical modeling framework. Essential mechanisms are pre-identified and expressed in the form of equations, equipped with a handful of significant parameters that require calibration and validation. The operational functionality of these models extends to the simulation of pedestrian scenes, enabling the prediction of future trajectories as a consequential byproduct. In the DB approach, the prediction of trajectories is not a secondary outcome, but the main objective. The parameters (or coefficients) of the algorithm have no physical meaning and can not be interpreted. These parameters are determined through a process of training the algorithm with data, with the goal of minimizing a predefined cost function. The common cost function for trajectory predictions is the displacement error metrics ADE or FDE. The trained algorithms are then tested on new data, i.e. data which was previously not used to train the algorithm (cross-validation).

Over the past decade, a multitude of studies employing various data-based methodologies have been published with the aim of predicting pedestrian trajectories. For a comprehensive overview, please refer to the following reviews [28, 30, 46, 5]. The majority of these studies use supervised deep learning techniques with either *Long Short-Term Memory* (LSTM) or *Generative Adversarial Networks* (GAN) architectures. Among the most influential works in this field are *Social-LSTM* by Alahi et al. [1] and *Social-GAN* by Gupta et al. [18]. These groundbreaking papers have served as the inspiration for a number of subsequent studies, which have extended these initial algorithms to incorporate elements such as scene information [57], attention mechanisms [19], graph neural networks [23, 38], and heterogeneity among pedestrians [33]. In addition, it's noteworthy to mention the utilization of convolutional neural networks, which can be trained on video data rather than trajectory data [58, 42, 11].

II.2.3 Other modelling approaches

There are further methods that diverge from the conventional KB models and DB algorithms, yet play a pivotal role in advancing the field. Among these, Reinforcement Learning (RL) stands out as a prominent unsupervised machine learning technique. Unlike supervised methods, RL learns through a trial-and-error approach, guided by a reward function that incentivizes or penalizes agent behaviors. This technique has proven effective in optimizing agent behaviors for objectives such as selecting the fastest path, avoiding collisions, and generating collective dynamics [53, 37, 12]. Expanding upon RL, Inverse Reinforcement Learning (IRL) represents a nuanced approach where the reward function is derived from observed data. IRL, and its more sophisticated counterpart, deep IRL, are instrumental in estimating complex, continuous reward functions, offering enhanced modeling of social behaviors [32, 54].

Another noteworthy direction is the hybridization of KB models with DB algorithms. Whereas the DB algorithms are always based on trajectories, studies on pedestrian dynamics have identified the relevance of variables like the relative velocity [13], the time gap [40], or indicators related to possibilities of collision such as bearing angle, collision cone, or time-to-collision [51, 25]. The use of these variables as inputs or in the loss function of the DB algorithms could allow obtaining prediction improvement, especially in the case of small amount of data for the training phase. Further hybrid approaches are using KB models to generate synthetic training data [1, 26], embedding KB models in the architecture of the DB algorithm [3], or using KB concepts in the loss function to rule out unrealistic predictions [50, 29, 16].

II.3 Method

In this section, we provide a detailed description of the datasets used in this study (Section II.3.1). We then define the models and algorithms employed in the following analysis (Section II.3.2) and introduce distance and collision-based evaluation metrics that will be used to assess the performance of the different models and algorithms (Section II.3.4).

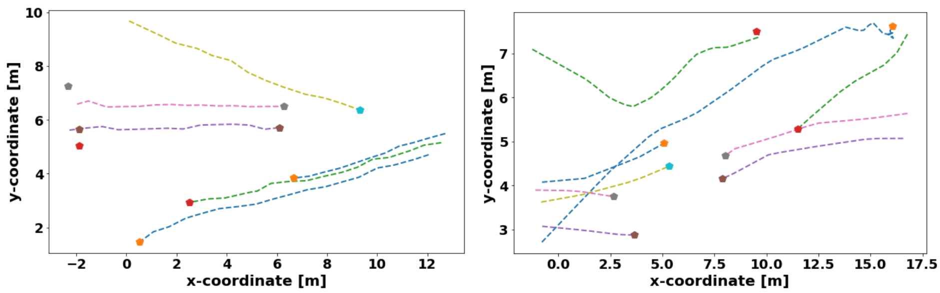
II.3.1 Pedestrian trajectory data

In recent years, a large number of datasets have been collected and made publicly available from an extensive range of studies, which mainly include real-world trajectories of scenarios with low pedestrian densities ranging from 0.1 to 0.4 ped/m^2. For a comprehensive overview, one may refer to [2]. Interestingly, datasets corresponding to higher-density situations are noticeably absent, possibly due to the challenges associated with the data collection (i.e. trajectory extraction).

Low-density dataset

Initially, we assess the performance of the models and algorithms using low-density datasets, which typically feature long-range interactions and scenarios involving less than 0.5 ped/m^2 [28]. Pedestrians in these scenes have high degrees of freedom, and their behaviour is primarily influenced by a few neighbouring people.

Given their emergence as benchmark datasets in pedestrian studies over recent years, we have selected the ETH [43] and UCY [34] datasets for the analysis. The ETH dataset comprises a total of 750 trajectories, divided into two subsets: ETH and Hotel. Fig. II.1 (a) shows an example of a segment from the hotel dataset. The UCY dataset, on the other hand, has been subdivided into three subsets: ZARA01, ZARA02, and UNIV, collectively containing 786 trajectories. An example of ZARA02 is shown in Fig. II.1 (b). Both datasets, collected in outdoor environments, encapsulate a variety of pedestrian traffic patterns, including unidirectional, bidirectional, and multidirectional. These datasets have been recorded at a framerate of 2.5 frames per second.



(a) Example from ETH.

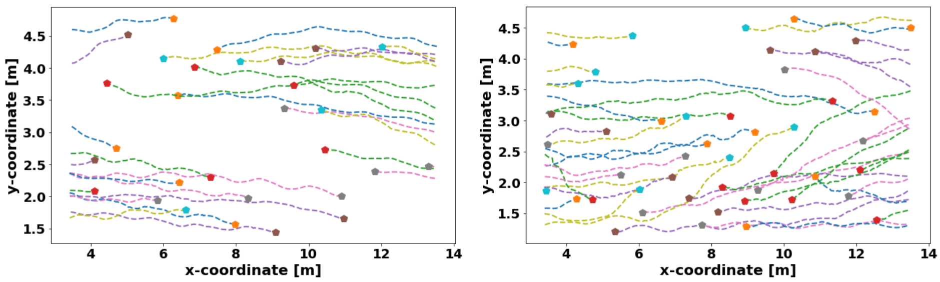
(b) Example from ZARA02.

Figure II.1 Illustrative examples of trajectory samples from low-density datasets (ETH and UCY sets).

High-density datasets

High-density data refers to pedestrian situations characterized by more than 2 ped/m^2, commonly known as crowds. The generation of accurate pedestrian trajectory data in these situations is a challenging task due to the difficulties in the automatic pedestrian identification and tracking. The problems that usually occur are that the pedestrians often gather together, occlude each other, and result in overlapping in pedestrian shapes [55]. These factors contribute to the scarcity of real-world high-density pedestrian trajectory datasets. Nevertheless, there are some rare examples available, which can be found in [24, 45, 39]. In addition to these real-world datasets, Forschungszentrum Jülich has conducted various laboratory experiments, including HERMES, BaSiGo, CroMa, and CrowdDNA, which provide high-density trajectory data [1, 10, 49]. Furthermore, other experimental datasets can be found in [17, 59].

In this study, we primarily focus on the corridor experiments with bidirectional flow from the Forschungszentrum Jülich, as they encompass a diverse array of interactions. The experiment setup incorporates two starting points or entrances, from where pedestrian start their walk. The size of the recording area is a = 10 m, and b_{corr} = 4 m. We utilize data gleaned from six distinct bidirectional corridor settings, as showcased in Table II.1. For illustrative purposes, we depict the trajectories from the experiment exhibiting the third-highest density (bidi3), alongside those from the fifth-highest density (bidi5) in Fig. II.2. In total, the dataset comprises 3096 trajectories, recorded at a framerate of 16 frames per second.



(a) Example from bidi3. (b) Example from bidi5.

Figure II.2 Illustrative examples of trajectory samples from high-density datasets (Juelich experiments).

II.3.2 Models and Algorithms

In the subsequent trajectory predictions, we focus on two crucial elements: predictions across varying pedestrian densities and the utilization of diverse mod-

Table II.1 Overview of pedestrian trajectory datasets.

Dataset	Setting	Number of Pedestrians	Average Density	Maximum Density
ETH				
ETH	Outdoor	361	0.14	0.35
HOTEL	Outdoor	389	0.13	0.32
UCY				
ZARA01	Outdoor	148	0.21	0.51
ZARA02	Outdoor	204	0.27	0.48
UNIV	Outdoor	434	0.38	0.52
JUELICH				
bidi1	Lab	141	0.38	0.55
bidi2	Lab	259	0.58	0.75
bidi3	Lab	480	1.00	1.15
bidi4	Lab	743	2.32	3.03
bidi5	Lab	643	2.64	3.275
bidi6	Lab	830	3.0	3.775

els/algorithms. We systematically use the notation $x_i \in \mathbb{R}^2$ and $v_i \in \mathbb{R}^2$ to represent the position and velocity of the i-th pedestrian. The Euclidean distance is denoted as $|\cdot|$, while x and v refer to the vectors of pedestrian positions and velocities. These vectors have a dimension of $2N$ for N pedestrians. All variables, including $x(t)$ and $x_i(t)$, depend on the time t.

We first select two contemporary knowledge-based models to facilitate this. The first is the Social Force model (SF) by Helbing and Molnar [20], and the second is the Optimal Reciprocal Collision Avoidance (ORCA) approach by van den Berg et al. [4]. The SF model is a widely adopted method that simulates pedestrian movement, treating individuals as particles influenced by various forces. The acceleration within this model is calculated from the summation of three forces as demonstrated in Equation II.1

$$m_i \frac{dv_i}{dt} = m_i \frac{v_i^0 - v_i}{\tau} + \sum_{j \neq i} \nabla U(x_j - x_i) + \sum_{W} \nabla V(x_W - x_i) \tag{II.1}$$

where m_i, v_i and $v_i^0$are the mass, current velocity and preferred velocity of the i-th pedestrian, respectively, while U and V are distance-based interaction potential, e.g.,

$$U(d) = Ae^{-\|d\|/B}, \qquad A, B > 0. \tag{II.2}$$

The first term of Equation II.1 denotes the driving force that the i-th pedestrian experiences to achieve their desired speed and direction within the reaction time $\tau > 0$. The second term represents the summation of the social forces derived from the repulsive effects of pedestrians maintaining distance from each other. The third force denotes the cumulative interaction forces between pedestrian i and obstacles.

In contrast, ORCA is centered on efficiently determining collision-free velocities for multiple agents within a shared environment [4]. It is based on a geometric approach to model agent interaction, identifying the range of velocities that guarantees collision avoidance within a specified time horizon by extrapolating linearly (i.e., assuming constant the velocity) the trajectories. These computations result in collision cones between the pedestrians, that can be empty. Further mechanisms are taken into account to avoid unrealistic oscillation effects and makes the agents acting independently without communicating with each other. Ultimately, each agent subsequently selects the optimal velocity v_i that is closest to its ideal (preferred) velocity v_i^0 within the feasible velocity region [41] excluding collisions, as shown in Equation II.3

$$v_i(t + dt) = \underset{v \in \cap_{j \neq i} \mathrm{ORCA}_{ij}(t)}{\arg\min} \|v - v_i^0\|, \tag{II.3}$$

with ORCA_{ij} the set of feasible (collision-free) velocities of the i-th pedestrian with the j-th neighboring pedestrian, and dt a (small) time step (typically equal to $dt = 0.01\,\mathrm{s}$).

For the DL approach, we adopt the LSTM network, which is widely used in pedestrian trajectory prediction. In this architecture, we leverage two algorithms. The first, referred to as the Vanilla LSTM, considers the historical trajectory over $[t-T_o, t]$, with $T_o > 0$ the observation time, to predict the trajectory over $[t, t+T_p]$ with T_p the prediction time

$$x_i(t + t_p) = \mathrm{LSTM}\big(t + t_p, \big(x_i(t - t_o),\ t_o \in [0, T_o]\big)\big), \qquad \forall t_p \in [0, T_p]. \tag{II.4}$$

This algorithms is grid-based which means that the input is discretised in a local grid constructed around the pedestrian. The second algorithm, known as the Social-LSTM [1], incorporates a social pooling mechanism to aggregate information about neighboring entities within the grid, as illustrated in Equation (II.5)

$$x_i(t+t_p) = \mathrm{SLSTM}\big(i, t+t_p, \big(x(t-t_o),\ t_o \in [0, T_o]\big)\big), \qquad \forall t_p \in [0, T_p]. \tag{II.5}$$

With this mechanism, the model can use the historical trajectories of surrounding pedestrians x over $[t - T_o, t]$, enabling the consideration of interactions in the predictions. Table II.2 lists the various types of information required by each model or algorithm to make predictions. The first approach in the Table II.2 is the constant velocity model. It is the most simple approach making prediction assuming the pedestrian velocities remain constant

$$x_i(t + t_p) = x_i(t) + t_p v_i(t), \qquad \forall t_p \in [0, T_p]. \tag{II.6}$$

II.3.3 Implementation details

The DL algorithms are trained with a learning rate of 0.0015, and a RMS-prop is used as the ADAM optimizer. The batch size is 8, and we train for 12 epochs. As a loss function, the mean squared error is used. For the validation and testing, we use a hold-out validation strategy. 15 % of the data is used for validation, 15 % for testing and the rest for training. The computations are performed using the PyTorch library[1]. Two different observation and prediction times are employed in the study. For 1.2-second predictions, a 1-second observation period is utilized. For 4.8-second predictions, the observation length extends to 3.6 seconds. We employ the Stochastic Gradient Descent (SGD) algorithm to fit the parameters of the KB models to the training data by minimizing the ADE metric. For the SF, we optimize the preferred velocity, the interaction potential, and the reaction time, according to [31]. For the ORCA we optimize the distance to pedestrians that are taken into account and the corresponding reaction time. It is worth noting that the reaction time plays a significant role in the model's behavior. A shorter reaction time prompts a quicker response to the presence of other agents but reduces the pedestrian's freedom in choosing their velocities, as mentioned in [51]. On the other hand, the CV model stands apart as it does not require any calibration or training due to its parameter-free nature. An overview of the most strinking differences between the approaches is presented in Table II.2. The displayed information includes the input of primary pedestrians and their neighbors, as well as the optimization method and the parameters being optimized.

II.3.4 Evaluation

Distance-based metrics

A crucial question that emerges when employing pedestrian trajectory prediction algorithms in high-density scenarios is the appropriate method of evaluation. This query is applicable more broadly to the field of pedestrian dynamics. Without an objective metric for evaluation, it is impossible to definitively determine the best-fitting model or algorithm. In low-density trajectory predictions, two metrics

[1]http://pytorch.org

Table II.2 Overview of important features.

Approach	Primary Input	Neighbor Input	Optim. Parameters	Optim. Method
CV	Current State	None	None	None
SF	Current State	Yes	Preferred Velocity, Interaction Potential, Reaction Time	SGD
ORCA	Current State	Yes	Neighbor Distance, Reaction Time	SGD
Vanilla-LSTM	Past Trajectory	None	Large number of coefficients	ADAM
Social-LSTM	Past Trajectory	Yes	Large number of coefficients	ADAM

based on Euclidean distance are commonly utilized. The first is the Average Displacement Error (ADE) [43], which measures the distance between the predicted trajectory and the ground truth trajectory at a set number of points

$$\textbf{ADE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\hat{x}_i(t) - x_i(t)\|, \tag{II.7}$$

$x_i(t)$ being the actual position of the i-th pedestrian at time t while $\hat{x}_i(t)$ is the predicted position. The second Euclidean distance based metric is Final Displacement Error (FDE). Unlike the ADE, which compares the predicted trajectory with the actual trajectory at every prediction step, FDE evaluates this comparison only at the final step, $x_N(t)$. In the following discussion, we will primarily focus on ADE.

Discrete distance-based collision metric

These two distance metrics are effective in guiding the algorithm to make predictions that match actual trajectories. However, a significant drawback is their focus on distance, leading to an underestimation of the repulsive forces that arise between pedestrians. Consequently, these metrics do not account for potential overlaps or collisions between pedestrians.

To tackle this problem, the following distance-based collision metric has been presented by Kothari et al. [30]

$$\textbf{Col} = \frac{1}{|S|} \sum_{\hat{Y} \in S} Col(\hat{Y}), \tag{II.8}$$

with

$$Col(\hat{Y}) = \min\left(1, \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j>i}^{N} \left[||\hat{x}_i(t) - \hat{x}_j(t)|| \leq 2R\right]\right), \tag{II.9}$$

where S includes all scenes in the test set, $\hat{Y}$ represents a scene prediction containing N agents, and $\hat{y}_i$ is the prediction of agent i over the prediction time of T, while $[\cdot]$ is the Iverson bracket

$$[P] = \begin{cases} 1 & \text{if } P \text{ is true,} \\ 0 & \text{otherwise.} \end{cases} \tag{II.10}$$

This last metric counts a prediction as a collision when a predicted pedestrian trajectory intersects with neighboring trajectories, thus indicating the proportion of predictions where collisions occur. A vital factor in this calculation is the chosen pedestrian size (radius), represented by the variable R in equation (II.9). An increase in R will likewise increase the number of collisions.

New TTC-based error metric

The collision metric Col discussed so far has been designed to mitigate overlapping and collisions between pedestrians. This is based on the principle that pedestrians inherently strive to avoid physical contact with others, an aspect not sufficiently captured by ADE. Nevertheless, this collision metric isn't without its shortcomings, which we aim to address. One drawback is that the metric is based on binary collision identifications. As such, it doesn't distinguish between minor instances of contact, such as a shoulder brush between two pedestrians walking side-by-side, and significant collisions such as a head-on crash. Another limitation is its inability to account for scenarios where a prediction results in multiple collisions. Additionally, the metric models pedestrians as circles, represented by radius R, although an elliptical representation would be more accurate. The proposed solution is to introduce a collision metric based on the concept of Time-to-Collision (TTC) between two pedestrians. In this system, a low TTC implies an impending collision.

The TTC is estimated as the time remaining before two moving pedestrians, donated as i and j, collide based on their current velocities. Suppose that R_i and R_j are the radius of i-th and j-th pedestrians, respectively. Given the relative distance and relative velocity between the two pedestrians

$$x_{ij} = x_i - x_j \qquad \text{and} \qquad v_{ij} = v_i - v_j, \tag{II.11}$$

a collision between i-th and j-th pedestrians occurs if there exists a time $\tau > 0$ such that $x_{ij} + v_{ij}\tau$ lies within a circle centered at $(0, 0)$ with radius $2R$. Mathematically, this condition can be expressed as $\|x_{ij} + v_{ij}t\| < 2R$ where $\| \cdot \|$ denotes Euclidean norm. It turns out to solve the quadratic inequality in t, and τ is the smallest positive root:

$$\tau_{ij} = \frac{-x_{ij} \cdot v_{ij} - \sqrt{(x_{ij} \cdot v_{ij})^2 - ||v_{ij}||^2(||x_{ij}||^2 - 4R^2)}}{||v_{ij}||^2}. \tag{II.12}$$

In this scheme, if no collision is imminent, then τ_{ij} is not real-value or is negative. We set in this case $\tau_{ij} = \infty$. Conversely, we assume by convention that $\tau_{ij} = 0$ if the pedestrians are already in collision. To be able to compare the performances of different prediction approaches, the inverse of average TTC (ITTC) is calculation according to

$$ITTC = \frac{NT}{\sum_{i=1}^{N} \sum_{t=1}^{T} \min_{j \neq i} \{\tau_{ij}(t), \tau_{\text{max}}\}}, \tag{II.13}$$

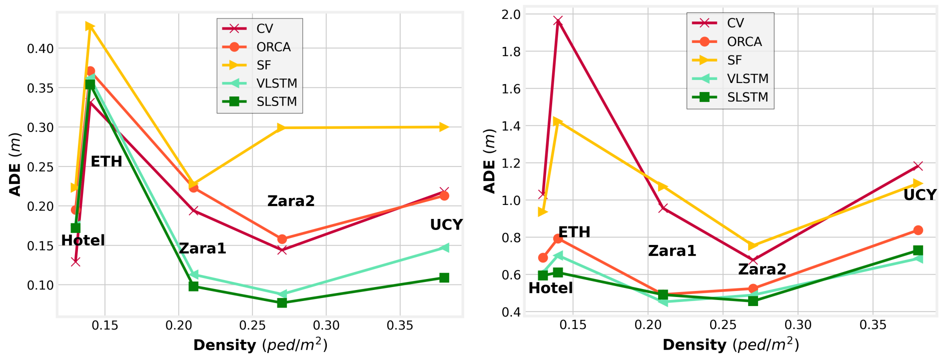
with $\tau_{\text{max}} = 12$ seconds a maximal TTC threshold value.

II.4 Results

In the upcoming chapter, the objective is to showcase the effectiveness of different approaches in making predictions across diverse time intervals and densities. To accomplish this, we will conduct comprehensive evaluations of the predictions using ADE (Section II.4.1), the distance-based collision metric Col (Section II.4.2), and the ITTC (Section II.4.2). Additionally, in Section II.4.3, we will demonstrate how the TTC metric can be leveraged to enhance trajectory predictions.

II.4.1 Distance metric

At first, we will focus on analyzing the distance-error metrics for the low-density datasets. The outcomes of the predictions are presented in Fig. II.3. As previously mentioned, we utilize five datasets for the low-density predictions, with densities ranging between 0.13 and 0.38 ped/m^2. In Fig. II.3, the x-axis displays the average densities of each dataset, while the y-axis represents the ADE metric. On the left side a prediction time T_p of 1.2 seconds is chosen and on the right side a prediction time of 4.6 seconds.



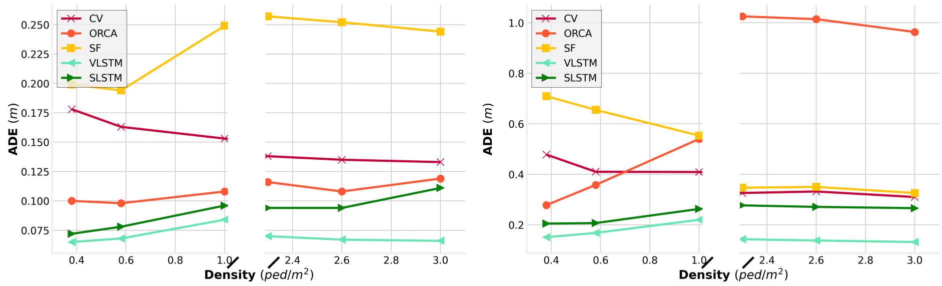
(a) Low-density data with $T_p = 1.2$ sec. (b) Low-density data with $T_p = 4.8$ sec.

Figure II.3 Distance error-metric (ADE) for low-density datasets.

The Fig. II.3 initially demonstrates that the algorithms consistently outperform the models across nearly all low-density datasets, showcasing their superior predictive capabilities. The Social-LSTM exhibits slightly better performance compared to the Vanilla-LSTM.
Furthermore, notable differences can be observed between the two prediction horizons. While the CV approach performs reasonably well at shorter prediction times, the error increases significantly, nearly doubling compared to the other approaches, as the prediction time extends. The substantial error of the CV approach on the ETH dataset highlights its challenging nature, with the highest deviation from keeping speed and direction constant. Interestingly, the most complex approach, namely SLSTM, demonstrates the best performance on this dataset, indicating its effectiveness in handling complexity. Moreover, it is worth noting that an increase in density does not necessarily result in higher prediction errors for the low-density dataset. For instance, Zara02, despite having a higher density than ETH, exhibits a lower average ADE.

In the next step, the same analysis of the different approaches with different prediction horizons is done for the high-density data. The results are presented in Fig. II.4.

As in Fig. II.3, the algorithms consistently outperform the models, irrespective of the dataset density. However, it is noteworthy that, surprisingly, the VLSTM exhibits superior performance compared to the more complex SLSTM. Additionally, it is notable that the error for SLSTM tends to increase with rising density, whereas for the VLSTM, it is lower within the higher density range of 2-3 ped/m^2. This unexpected outcome suggests that higher complexity does not necessarily lead to improved results when dealing with high densities. This assumption gains further support from the observed progression of the ADE for the CV approach,



(a) High-density data with $T_p = 1.2$ sec. (b) High-density data with $T_p = 4.8$ sec.

Figure II.4 Distance error-metric (ADE) for high-density datasets.

where the error appears to decrease with higher densities.

Similar to the results presented in Fig. II.3, we observed that the SF model performs better for longer prediction horizons compared to other approaches. Conversely, the ORCA model struggles to make accurate predictions at densities exceeding 2 ped/m^2. This limitation is attributed to the "freezing problem" highlighted in Luo et al. [36], which becomes prominent at higher densities. For the longer prediction time, the performance of ORCA declines with higher density, while the performance of the SF improves with increasing density.

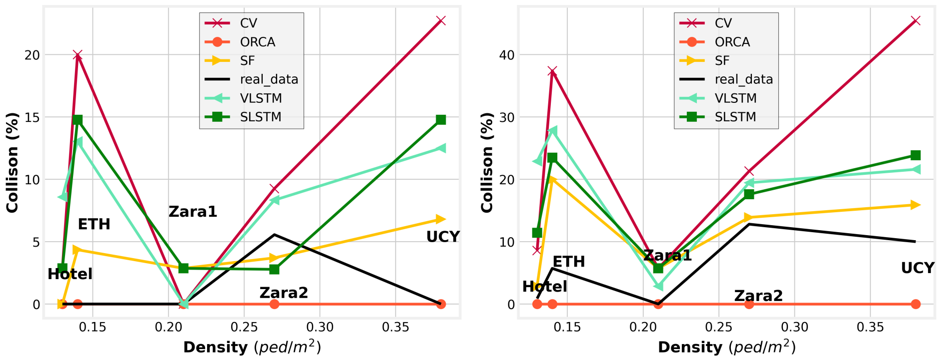
II.4.2 Collision metrics

In this section, we will address an important challenge regarding collisions. In certain instances, pedestrians fail to avoid one another, contrary to what one would expect. This leads to overlapping or collision events, contravening one of the most critical physical criteria that a realistic prediction should satisfy. To assess the magnitude of this phenomenon, we will employ the two collision metrics Col and ITTC described in Section 3.

Distance-based collision metric

An essential aspect of the collision metrics is accurately defining when unrealistic behaviors, such as overlapping, occur. Hence, the shape of the pedestrians plays a crucial role. In the following, we present predictions made for two distinct radii: 0.1 meters and 0.2 meters. The y-axis represents the percentage of predictions demonstrating collisions, as determined by the distance-based collision metric. The black line has not been presented in the figures before and it displays the percentage of collisions in the real datasets.

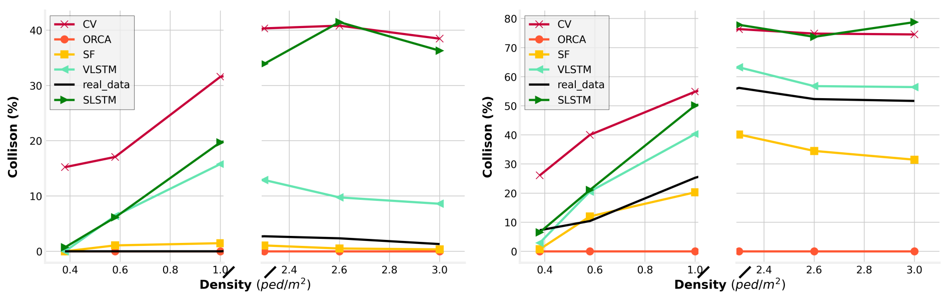
In Fig. II.5, it is evident that, as anticipated, the percentage of collisions is significantly higher for the larger radius. This observation holds true for all ap-



(a) Low-density data, $R = 0.1$. (b) Low-density data, $R = 0.2$.

Figure II.5 Distance-based collision metric for different radius sizes.

proaches and the real data, with the exception of the ORCA model, which demonstrates no collisions for either radius. Regarding the real data, collisions at a radius of 0.1 meters only occur within the Zara2 dataset, whereas collisions occur in the HOTEL, ETH, Zara2, and UCY datasets for a radius of 0.2 meters. This finding is surprising because collisions should not occur in the real trajectories. Collision between pedestrian occur very rarely in the real world. However, upon animating the real trajectories and plotting the colliding trajectories, it becomes apparent that the distance-based collision metric has a drawback: it sometimes considers grouping behavior as collision behavior. When pedestrians walk in groups, they occasionally come so close together that even at a radius of 0.1 meters, they sometimes overlap. In the next Fig. II.6 the distance-based collision metric for the high-density data is presented.



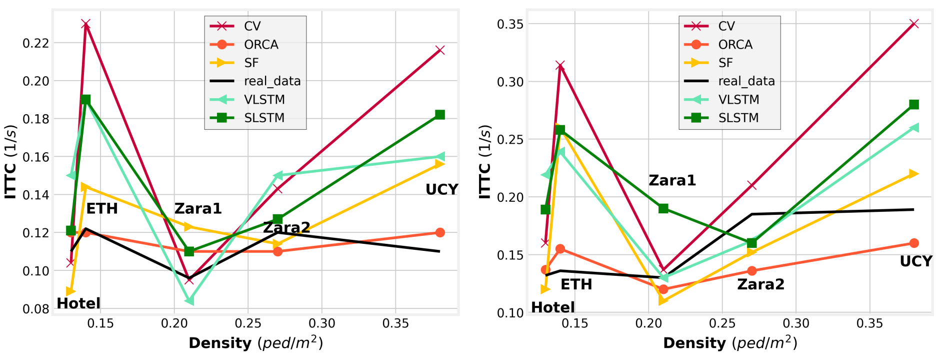
(a) Bidirectional data, $R = 0.1$. (b) Bidirectional data, $R = 0.2$.

Figure II.6 Distance-based collision metric for different radius sizes.

Substantial disparities are evident between the two diagrams representing collision occurrences at varying radii. The predicted collision frequency is nearly double at a radius of 0.2 meters as compared to a smaller radius. The black line, indicative of actual collision data, suggests a negligible number of collisions at high-density datasets for a radius of 0.1 meters. However, when the radius is increased to 0.2 meters, the predictions indicate almost 50 % more collisions in higher densities. These observations suggest that a radius of 0.2 meters is excessively large for such scenarios. Pedestrians at these densities are in such close proximity that they often overlap when represented as circular objects. The ORCA model predicts no collisions for either radius, but this results in extraordinarily high ADE values for large prediction horizons, as depicted in Fig.II.4, right panel. The CV model underperforms and displays the highest collision rate in its predictions. The SLSTM shows comparable performance at lower densities as demonstrated in Fig. II.6, but its effectiveness diminishes at higher densities. The VLSTM model exhibits superior performance to the SLSTM in terms of collision metrics. The percentage of predicted collision of the SF model most closely aligns with actual trajectories.

TTC-based collision metric

In this section, we will discuss the results of the predictions centered around the ITTC. As observed in the preceding figures, the x-axis represents the density of the data. On the other hand, the y-axis denotes ITTC of the predictions (see equation II.13). It is important to note that the maximum TTC value was arbitrarily set to 12 seconds. A higher TTC can be interpreted as an indication of a prediction that is not at risk of collision. Consequently, a lower ITTC value is preferable for collision avoidance compared to a higher one.

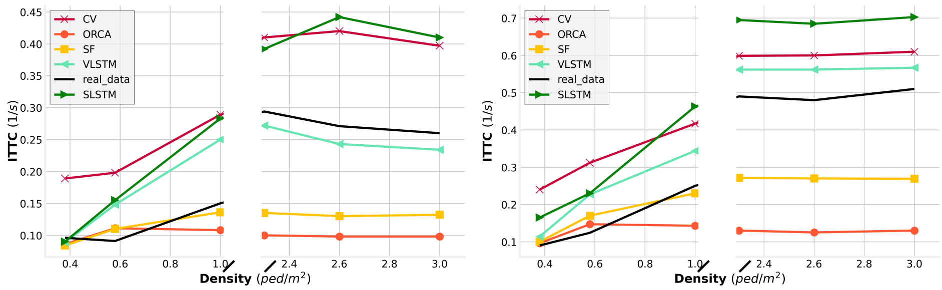


(a) Low-density data, $R = 0.1$. (b) Low-density data, $R = 0.2$.

Figure II.7 TTC collision metric different radius sizes.

Upon initial observation of Fig. II.7, it is evident that the differences between Fig. II.7 (a) and (b) are considerably smaller compared to those seen in the distance-based collision metric. This observation suggests that the TTC-based collision metric is less sensitive to changes in the radius.

Despite this difference, the overall trend of the lines bears a resemblance to that seen in Fig. II.5. However, the information provided by the continuous collision metric is richer and appears to be more accurate. The KB models exhibit lower inverse ATTC values than the algorithms, with the CV model demonstrating the least optimal performance. Notably, the inverse ATTC of ORCA's prediction most closely aligns with the actual trajectories. Proceeding further, we will present the ITTC for high-density data in Fig. II.8.



(a) High-density data, $R = 0.1$.

(b) High-density data, $R = 0.2$.

Figure II.8 TTC collision metric different radius sizes.

Again, it is noteworthy, that the variations between Fig. II.8 (a) and (b) are not as substantial as those seen in Fig. II.6. In the latter virtually no collisions are observable in the actual trajectories (black line) for a radius of 0.1 meters, yet the collision rates for a radius of 0.2 meters are markedly high. In contrast, in Fig. II.8, the difference between (a) and (b) for real trajectories is much less pronounced. Despite this, the trends in Fig. II.6 and Fig. II.8 do not exhibit significant discrepancies, with the KB models consistently surpassing the algorithms and the SLSTM model underperforming in comparison to the VLSTM. In some instances, SLSTM even demonstrates worse performance than the CV model.

II.4.3 TTC Metric for improving performance of the algorithm

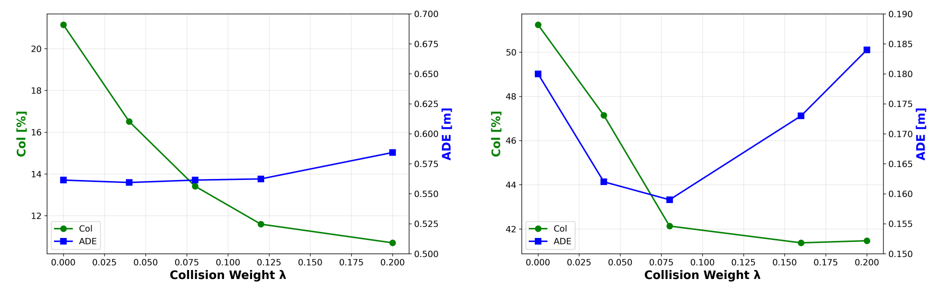
The preceding section demonstrates the promising attributes of the ITTC collision metric. It provides reasonable results that have the advantage of continuity and exhibit less dependence on the pedestrians' shape compared to the distance-based collision metric Col. As such, we have incorporated the TTC into the cost function to see if the prediction can be improved. The cost function quantifies the disparity between the prediction and the real observation. It provides a single indicator L_i that will be minimized during the training. The idea is to penalize predictions with exceptionally low TTC values. Keeping all configurations the same ($T_p = 4.8$ seconds and R=0.2), we have integrated the TTC into the cost function of the SLSTM, as illustrated in equation II.14

$$\mathbf{L}_i = \sum_{t=1}^{T} \|x_i(t) - \hat{x}_i(t)\|^2 + \lambda \sum_{t=1}^{T} f(\min_{j \neq i}\{\tau_{ij}\}). \tag{II.14}$$

The first part of the equation II.14 shows the distance-based metric ADE that is used for training. It compares the actual position of the pedestrian x_i to the predicted one $\hat{x}_i$. The subsequent segment utilizes a sigmoid penalty function f (see Equation II.15), which results in a high penalty for low TTC values

$$f(\tau) = \frac{1}{1 + e^{s(\tau - \delta)}}, \tag{II.15}$$

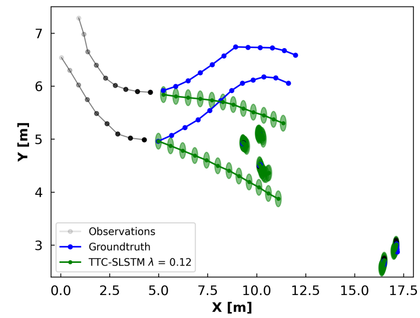
where s and δ are slope and threshold parameters, respectively.
The parameter $\lambda \geq 0$ determines the weight to be given to the second part of the cost function. In Fig. II.9 the results of the predictions in terms of ADE and Col are shown for different settings of λ.



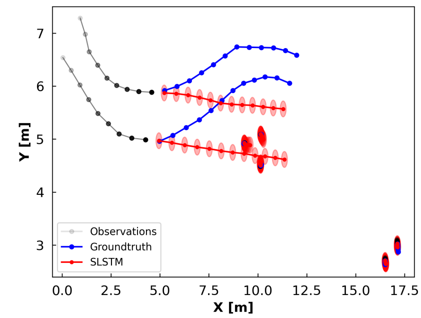
(a) Low-density data. (b) High-density data.

Figure II.9 Illustrating the improvements achieved by incorporating TTC into the training function for both low-and high-density data.

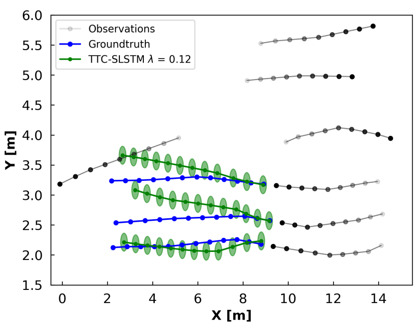
The green line represents the Col metric, while the blue line shows ADE. In the first observation, the value of λ is zero, which means that this model is identical to the SLSTM without a TTC term in the cost function. It can be clearly shown, that the TTC term in the cost function helps to improve avoidance behavior. There is a strong relationship between the value of λ and the number of collisions in the predictions. In the case of low-density data, it is possible to halve the number of collisions without an accompanying increase in ADE. For the high-density data, a reduction in the Col metric by 20 % is achievable, which also leads to a decrease in ADE. The enhancements afforded by the incorporation of TTC into our algorithm are discernible both quantitatively and qualitatively. When visualizing the predicted trajectories, those produced by the TTC-incorporated model appear more realistic, exhibiting superior collision avoidance characteristics. Fig. II.10 presents two scenes from the low density data. On the left side the predictions made by the algorithm trained with TTC are shown and on the right side the predictions with the SLSTM.



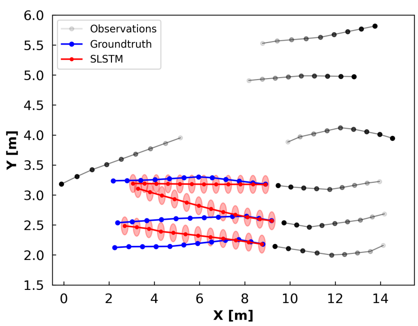
(a) Prediction with TTC in cost function.



(b) Prediction with SLSTM.



(c) Prediction with TTC in cost function.



(d) Prediction with SLSTM.

Figure II.10 Example of trajectory predictions based on the algorithm that was trained with TTC in the cost function and the SLSTM.

In both scenes, the SLSTM's predictions result in collisions. In the upper image, pedestrians do not navigate around the stationary individuals in the middle. Meanwhile, in the lower image, the pedestrians form a group walking so close to each other that it registers as a collision. In the predictions generated using TTC, pedestrians within a group maintain a greater distance from each other. Additionally, in the upper image, pedestrians navigate successfully around those stationary in the center, further demonstrating the benefits of incorporating TTC into trajectory predictions.

II.5 Conclusion

In this study, we have conducted a detailed empirical analysis comparing various pedestrian trajectory prediction approaches. The investigation underscores that the task of predicting pedestrian trajectories is intrinsically complex, with different densities posing additional challenges. While the SLSTM demonstrates excellent performance in low-density scenarios, it struggles to maintain similar accuracy in high-density situations. A particular limitation of the DB algorithms, namely, a significant incidence of collisions in high-density predictions, is addressed by introducing an innovative continuous collision metric that calculates the time-to-collision between pedestrians. This new metric presents a valuable instrument to assess the performance of the approaches, enhancing the overall trajectory prediction accuracy realism feature in terms of hardcore body exclusion.

Although this paper serves as a foundation for refining and expanding the capabilities of DB algorithms for predicting and modelling pedestrian behavior at high densities, it is important to recognize the inherent limitations of this approach. The predictions are predominantly short-term, lacking iterative capabilities, and are confined primarily to the operational level. This narrow focus overlooks the complexities involved in simulating entire pedestrian landscapes, where challenges such as intricate environments, multiple potential pathways, and the intricacies of group dynamics are prevalent. To address these gaps, future research must integrate a tactical level into the modeling process, e.g. for route choice decision making [21, 15]. This addition would enable the simulation of more comprehensive pedestrian scenarios, accounting for the myriad of factors that influence pedestrian behavior in real-world settings.

Bibliography

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social LSTM: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 961–971, 2016.

[2] Javad Amirian, Bingqing Zhang, Francisco Valente Castro, Juan Jose Baldelomar, Jean-Bernard Hayet, and Julien Pettre. OpenTraj: Assessing prediction complexity in human trajectories datasets. In *Asian Conference on Computer Vision (ACCV)*, number CONF, pages 1–17. Springer, 2020.

[3] Alessandro Antonucci, Gastone Pietro Rosati Papini, Paolo Bevilacqua, Luigi Palopoli, and Daniele Fontanelli. Efficient prediction of human motion for real-time robotics applications with physics-inspired neural networks. *IEEE Access*, 10:144–157, 2021.

[4] Jur van den Berg, Stephen J Guy, Ming Lin, and Dinesh Manocha. Reciprocal n-body collision avoidance. In *Robotics research*, pages 3–19. Springer, 2011.

[5] Ariyan Bighashdel and Gijs Dubbelman. A survey on path prediction techniques for vulnerable road users: From traditional to deep-learning approaches. In *2019 IEEE intelligent transportation systems conference (ITSC)*, pages 1039–1046. IEEE, 2019.

[6] Stephen Bitgood. An analysis of visitor circulation: Movement patterns and the general value principle. *Curator: The Museum Journal*, 49(4):463–475, 2006.

[7] Victor J Blue and Jeffrey L Adler. Cellular automata microsimulation of bidirectional pedestrian flows. *Transportation Research Record*, 1678(1):135–141, 1999.

[8] M Boltes, J Zhang, A Tordeux, A Schadschneider, and A Seyfried. Empirical results of pedestrian and evacuation dynamics. *Encyclopedia of complexity and systems science*, 16:1–29, 2018.

[9] Carsten Burstedde, Kai Klauck, Andreas Schadschneider, and Johannes Zittartz. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications*, 295(3-4):507–525, 2001.

[10] Shuchao Cao, Armin Seyfried, Jun Zhang, Stefan Holl, and Weiguo Song. Fundamental diagrams for multidirectional pedestrian flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(3):033404, 2017.

[11] Kai Chen, Xiao Song, Daolin Han, Jinghan Sun, Yong Cui, and Xiaoxiang Ren. Pedestrian behavior prediction model with a convolutional LSTM encoder–decoder. *Physica A: Statistical Mechanics and its Applications*, 560:125132, 2020.

[12] Yu Fan Chen, Michael Everett, Miao Liu, and Jonathan P How. Socially aware motion planning with deep reinforcement learning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1343–1350. IEEE, 2017.

[13] Mohcine Chraibi, Armin Seyfried, and Andreas Schadschneider. Generalized centrifugal-force model for pedestrian dynamics. *Physical Review E*, 82(4):046111, 2010.

[14] Mohcine Chraibi, Antoine Tordeux, Andreas Schadschneider, and Armin Seyfried. Modelling of pedestrian and evacuation dynamics. *Encyclopedia of complexity and systems science*, pages 1–22, 2018.

[15] Huu-Tu Dang, Benoit Gaudou, and Nicolas Verstaevel. A multi-level density-based crowd simulation architecture. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 64–75. Springer, 2023.

[16] Huu-Tu Dang, Raphael Korbmacher, Antoine Tordeux, Benoit Gaudou, and Nicolas Verstaevel. Ttc-slstm: Human trajectory prediction using time-to-collision interaction energy. In *2023 15th International Conference on Knowledge and Systems Engineering (KSE)*, pages 1–6. IEEE, 2023.

[17] Iñaki Echeverría-Huarte, A Garcimartín, RC Hidalgo, César Martín-Gómez, and I Zuriguel. Estimating density limits for walking pedestrians keeping a safe interpersonal distancing. *Scientific reports*, 11(1):1–8, 2021.

[18] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2255–2264, 2018.

[19] Sirin Haddad, Meiqing Wu, Wei He, and Siew Kei Lam. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model. In

Proceedings of the 24th Computer Vision Winter Workshop, pages 4–13, 2019.

[20] Dirk Helbing and Peter Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

[21] Serge P Hoogendoorn and Piet HL Bovy. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological*, 38(2):169–190, 2004.

[22] Serge P Hoogendoorn, Femke LM van Wageningen-Kessels, Winnie Daamen, and Dorine C Duives. Continuum modelling of pedestrian flows: From microscopic principles to self-organised macroscopic phenomena. *Physica A: Statistical Mechanics and its Applications*, 416:684–694, 2014.

[23] Yingfan Huang, Huikun Bi, Zhaoxin Li, Tianlu Mao, and Zhaoqi Wang. STGAT: Modeling spatial-temporal interactions for human trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6272–6281, 2019.

[24] Anders Johansson and Dirk Helbing. Analysis of empirical trajectory data of pedestrians. In *Pedestrian and Evacuation Dynamics 2008*, pages 203–214. Springer, 2009.

[25] Ioannis Karamouzas, Brian Skinner, and Stephen J Guy. Universal power law governing pedestrian interactions. *Physical review letters*, 113(23):238701, 2014.

[26] A Khadka, Paolo Remagnino, and Vasileios Argyriou. Synthetic crowd and pedestrian generator for deep learning problems. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4052–4056. IEEE, 2020.

[27] Ansgar Kirchner and Andreas Schadschneider. Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics. *Physica A: statistical mechanics and its applications*, 312(1-2):260–276, 2002.

[28] Raphael Korbmacher and Antoine Tordeux. Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches. *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[29] Parth Kothari and Alexandre Alahi. Safety-compliant generative adversarial networks for human trajectory forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 24(4):4251–4261, 2023.

[30] Parth Kothari, Sven Kreiss, and Alexandre Alahi. Human trajectory forecasting in crowds: A deep learning perspective. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):7386–7400, 2021.

[31] Sven Kreiss. Deep social force. *arXiv preprint arXiv:2109.12081*, 2021.

[32] Henrik Kretzschmar, Markus Spies, Christoph Sprunk, and Wolfram Burgard. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research*, 35(11):1289–1307, 2016.

[33] Wei-Cheng Lai, Zi-Xiang Xia, Hao-Siang Lin, Lien-Feng Hsu, Hong-Han Shuai, I-Hong Jhuo, and Wen-Huang Cheng. Trajectory prediction in heterogeneous environment via attended ecology embedding. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 202–210, 2020.

[34] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by example. In *Computer graphics forum*, volume 26(3), pages 655–664. Wiley Online Library, 2007.

[35] Ruggiero Lovreglio, Enrico Ronchi, and Daniel Nilsson. Calibrating floor field cellular automaton models for pedestrian dynamics by using likelihood function optimization. *Physica A: Statistical Mechanics and its Applications*, 438:308–320, 2015.

[36] Yuanfu Luo, Panpan Cai, Aniket Bera, David Hsu, Wee Sun Lee, and Dinesh Manocha. Porca: Modeling and planning for autonomous driving among many pedestrians. *IEEE Robotics and Automation Letters*, 3(4):3418–3425, 2018.

[37] Francisco Martinez-Gil, Miguel Lozano, and Fernando Fernández. Emergent behaviors and scalability for multi-agent reinforcement learning-based pedestrian models. *Simulation Modelling Practice and Theory*, 74:117–133, 2017.

[38] Alessio Monti, Alessia Bertugli, Simone Calderara, and Rita Cucchiara. DAG-Net: Double attentive graph neural network for trajectory forecasting. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2551–2558. IEEE, 2021.

[39] Masamitsu Mōri and Hiroshi Tsukaguchi. A new method for evaluation of level of service in pedestrian facilities. *Transportation Research Part A: General*, 21(3):223–234, 1987.

[40] Mehdi Moussaïd, Dirk Helbing, and Guy Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, 108(17):6884–6888, 2011.

[41] Sahil Narang, Andrew Best, Sean Curtis, and Dinesh Manocha. Generating pedestrian trajectories consistent with the fundamental diagram based on physiological and psychological factors. *PLoS one*, 10(4):e0117856, 2015.

[42] Nishant Nikhil and Brendan Tran Morris. Convolutional neural network for trajectory prediction. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.

[43] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th international conference on computer vision*, pages 261–268. IEEE, 2009.

[44] Atanas Poibrenski, Matthias Klusch, Igor Vozniak, and Christian Müller. M2p3: multimodal multi-pedestrian path prediction by self-driving cars with egocentric vision. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pages 190–197, 2020.

[45] Abishai Polus, Joseph L Schofer, and Ariela Ushpiz. Pedestrian flow and level of service. *Journal of transportation engineering*, 109(1):46–56, 1983.

[46] Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M Kitani, Dariu M Gavrila, and Kai O Arras. Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, 39(8):895–935, 2020.

[47] Andreas Schadschneider, Mohcine Chraibi, Armin Seyfried, Antoine Tordeux, and Jun Zhang. Pedestrian dynamics: From empirical results to modeling. *Crowd Dynamics, Volume 1: Theory, Models, and Safety Problems*, pages 63–102, 2018.

[48] Stefano Scheggi, Marco Aggravi, Fabio Morbidi, and Domenico Prattichizzo. Cooperative human-robot haptic navigation. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2693–2698. IEEE, 2014.

[49] Armin Seyfried, Bernhard Steffen, Wolfram Klingsch, and Maik Boltes. The fundamental diagram of pedestrian movement revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(10):P10002, 2005.

[50] Mattia Silvestri, Michele Lombardi, and Michela Milano. Injecting domain knowledge in neural networks: a controlled experiment on a constrained problem. In *Integration of Constraint Programming, Artificial Intelligence, and Operations Research: 18th International Conference, CPAIOR 2021, Vienna, Austria, July 5–8, 2021, Proceedings 18*, pages 266–282. Springer, 2021.

[51] Jur Van Den Berg, Stephen J Guy, Ming Lin, and Dinesh Manocha. Reciprocal n-body collision avoidance. In *Robotics Research: The 14th International Symposium ISRR*, pages 3–19. Springer, 2011.

[52] Jur Van den Berg, Ming Lin, and Dinesh Manocha. Reciprocal velocity obstacles for real-time multi-agent navigation. In *2008 IEEE international conference on robotics and automation*, pages 1928–1935. Ieee, 2008.

[53] Giuseppe Vizzari and Thomas Cecconello. Pedestrian simulation with reinforcement learning: A curriculum-based approach. *Future Internet*, 15(1):12, 2022.

[54] Zhiqiang Wan, Chao Jiang, Muhammad Fahad, Zhen Ni, Yi Guo, and Haibo He. Robot-assisted pedestrian regulation based on deep reinforcement learning. *IEEE Transactions on Cybernetics*, 50(4):1669–1682, 2018.

[55] Xinlong Wang, Tete Xiao, Yuning Jiang, Shuai Shao, Jian Sun, and Chunhua Shen. Repulsion loss: Detecting pedestrians in a crowd. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7774–7783, 2018.

[56] Qiancheng Xu, Mohcine Chraibi, Antoine Tordeux, and Jun Zhang. Generalized collision-free velocity model for pedestrian dynamics. *Physica A: Statistical Mechanics and its Applications*, 535:122521, 2019.

[57] Hao Xue, Du Q Huynh, and Mark Reynolds. SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1186–1194. IEEE, 2018.

[58] Shuai Yi, Hongsheng Li, and Xiaogang Wang. Pedestrian behavior understanding and prediction with deep neural networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 263–279. Springer, 2016.

[59] Francesco Zanlungo, Claudio Feliciani, Zeynep Yücel, Katsuhiro Nishinari, and Takayuki Kanda. Macroscopic and microscopic dynamics of a pedestrian cross-flow: Part ii, modelling. *Safety Science*, 158:105969, 2023.

III. Toward Better Pedestrian Trajectory Predictions: The Role of Density and Time-to-Collision in Hybrid Deep-Learning Algorithms

Outline

Bibliographic Information

Korbmacher, R., & Tordeux, A. (2024). Towards Better Predictive Models: The Role of Density in Pedestrian Trajectory Predictions. https://www.sciencedirect.com.

Author's contribution

The author contributed to . . . . Furthermore, wrote a significant part of . . . .

Copyright Notice

Toward Better Pedestrian Trajectory Predictions: The Role of Density and Time-to-Collision in Hybrid Deep-Learning Algorithms

Raphael Korbmacher[1], and Antoine Tordeux[1]

[1]School of Mechanical Engineering and Safety Engineering, University of Wuppertal, Gaußstr. 20, Germany

---

Abstract

Predicting human trajectories poses a significant challenge due to the complex interplay of pedestrian behavior, which is influenced by environmental layout and interpersonal dynamics. This complexity is further compounded by variations in scene density. To address this, we introduce a novel dataset from the Festival of Lights in Lyon 2022, characterized by a wide range of densities (0.2-2.2 ped/m^2). Our analysis demonstrates that density-based classification of data can significantly enhance the accuracy of predictive algorithms. We propose an innovative two-stage processing approach, surpassing current state-of-the-art methods in performance. Additionally, we utilize a collision-based error metric to better account for collisions in trajectory predictions. Our findings indicate that the effectiveness of this error metric is density-dependent, offering prediction insights. This study not only advances our understanding of human trajectory prediction in dense environments but also presents a methodological framework for integrating density considerations into predictive modeling, thereby improving algorithmic performance and collision avoidance.

Keywords: Pedestrian trajectory prediction, deep learning, pedestrian trajectory dataset, density-based classification, collision avoidance

---

III.1 Introduction

The challenge of predicting pedestrian trajectories has emerged as a pivotal challenge in recent years. This surge in interest is largely attributed to the profound implications it holds for autonomous vehicle navigation [29], service robot deployment [31], and the strategic planning of infrastructure and mass gatherings [4]. In addressing these intricate challenges, researchers have traditionally employed physics-based (PB) models to simulate and understand pedestrian behavior. These models have been instrumental in dissecting collective phenomena and enhancing our understanding of pedestrian dynamics, particularly in high-density contexts relevant to crowd management and evacuation strategies [4]. However, the landscape of pedestrian trajectory prediction has witnessed a paradigm shift over the last decade with the advent and integration of deep learning (DL) algorithms [20]. Despite the opaqueness of these models in terms of interpretability, their superiority in mirroring observed trajectories has been markedly pronounced, especially when juxtaposed with their PB counterparts [1]. Nonetheless, it is important to acknowledge that the domains of applicability for PB and DL models do not entirely overlap. While PB models excel in the realm of high-density simulations, providing insights into collective behavior, DL algorithms predominantly thrive in low-density environments where individual pedestrian movements are characterized by a greater degree of freedom and intricate long-range interactions [20]. This paper introduces a novel, real-world pedestrian trajectory dataset, gathered during the Festival of Lights in Lyon. Field pedestrian trajectory datasets are typically gathered from low-density situations. In contrast, this dataset captures the nuanced dynamics of pedestrian movements across a large spectrum of density levels. It ranges from sparse crowds observed during show moments to the densely packed throngs seen after the event. Utilizing this dataset, we train DL algorithms, including Long Short-Term Memory (LSTM) networks and Generative Adversarial Networks (GAN). Our methodology is underscored by a novel approach: we harness situational classification predicated on crowd density to refine our models' learning process. This two-stage process, which initially classifies the scene based on density and then predicts the trajectories, not only bolsters the efficiency of our models but also substantially elevates the precision of trajectory predictions. This improvement is demonstrated by comparative analyses with traditional DL algorithms and PB models.

Another challenge of trajectory prediction that we face in this paper is overlapping and colliding of predicted trajectories [8]. To tackle that problem, we integrate a time-to-collision (TTC) [16] term into the loss function of the algorithms. A parameter, λ, is utilized to modulate the TTC's influence on the training. Our empirical research uncovers a significant relationship between the optimal λ

values and the density levels, highlighting the intricacies of pedestrian behavior across different densities.

The remainder of this paper is organized as follows: first, we review related studies in Section III.2. Then, our novel dataset is presented in Section III.3 and the methodology for the empirical work is proposed in Section III.4. The results are shown in Section III.5. The last Section III.6 includes a discussion of the results and an outlook on future works.

III.2 Related work

The domain of pedestrian trajectory prediction is multifaceted, drawing insights from various disciplines and methodologies. The two main stream are the physics-based models and data-based algorithms [20]. PB models have been the cornerstone of understanding pedestrian dynamics, especially in high-density scenarios. The Social Force model (SF), introduced by Helbing and Molnar [13], exemplifies this approach, simulating pedestrian movement by balancing attractive and repulsive forces. Other famous PB models are the Optimal Reciprocal Collision Avoidance (ORCA) from Van den Berg et al. [34] or the cellular automata model from Burstedde et al. [6]. However, these models are not without their challenges, particularly when it comes to encapsulating the full range of crowd behavior [20]. For more PB models see reviews like [3, 7, 9].

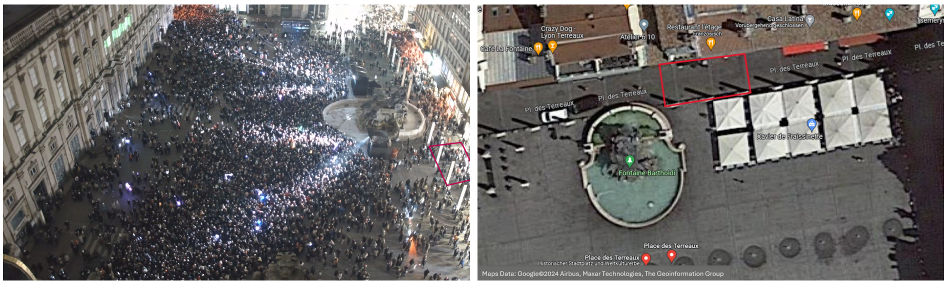
In pursuit of addressing these limitations, the research frontier has gradually shifted towards data-driven methodologies. Notably, the past decade has witnessed a burgeoning interest in DL approaches. Pioneering works like the *Social LSTM* from Alahi et al. [1] introduce the use of Recurrent Neural Networks (RNN), specifically LSTM networks, in conjunction with a novel concept known as *Social Pooling*. This innovative approach incorporates neighbouring information, thereby enriching the model's contextual understanding. This social concept was further enhanced by Gupta et al. [12] through *Social GAN*, where the generative adversarial framework allowed for the generation of multiple plausible future paths, addressing the inherent uncertainty in human movement. Alternative methods employed for social predictions include attention mechanisms [35], graph-based approaches [26], and the utilization of relative coordinates [32]. Additionally, deep learning architectures such as Convolutional Neural Networks [40, 24] and Transformers [39] have been applied in trajectory prediction tasks.

A pivotal aspect of this paper is the innovative classification of trajectory scenes based on crowd density prior to prediction. To the best of our knowledge, this approach is a novel paradigm, potentially owing to the scarcity of high-density, real-world pedestrian trajectory datasets. Xue et al. [37] predict pedestrian destinations using bidirectional LSTM classification. This involves an additional classification stage to distinguish between possible destinations of pedestrians. They classify the route manually into four distinct categories. In another paper from the authors [38] the classification is done based on a clustering algorithm. Kothari et al. [22] categorize pedestrian trajectories based on the nature of interactions observed, identifying behaviors such as collision avoidance, leader-follower dynamics, and grouping behavior. An alternative methodology involves classifying trajectories based on individual pedestrian characteristics. Papathanasopoulou et al. [27] concentrate on attributes such as age, gender, height, and speed to inform their classification. A second cornerstone of this work is the seamless integration of a PB concept, TTC, into the loss function of DL algorithms. This synthesis of PB principles and DL models is not an isolated endeavor. Alahi et al. [1] and Khadka et al. [17] utilized simulated data from PB models for training DL algorithms. Antonucci et al. [2] embedded a PB model directly into the DL architecture. Furthermore, the works of Silvestri et al. [33] and Kothari et al. [21] stand out for their use of PB principles within the loss function to eliminate unrealistic predictions.

III.3 The Dataset

With a growing interest in data-based methods the significance of pedestrian trajectory data has been elevated in recent research. This area has seen a proliferation of datasets published by researchers, which can be categorized into field data and experimental data obtained in laboratory conditions. In the field studies, real-world settings are employed where individuals, unaware of their participation in a study, navigate through various scenarios. Famous field datasets are the ETH [28] and UCY [23] datasets, which are widely used in the machine learning community. Originating from surveillance videos, these datasets capture pedestrians scene of low density (0.1-0.5 ped/m^2). The GLOW dataset from Eindhoven [10], a dataset used for route choice analysis, contains trajectory scenes of higher densities, but only for short length trajectories. Other field datasets are the Stanford Drone Dataset [30], the Grand Central Station Dataset [41], and the Edinburgh Informatics Forum Dataset [25]. None of these have densities above 0.2 ped/m^2. In the following we will present a field dataset with pedestrian densities between 0.2-2.2 ped/m^2.

The data was collected at Lyon´s Festival of Lights. The event running for four days from 7 pm to 11 pm attracts millions (2 million in 2022) of visitors each year. Key attractions are light shows at Place des Terreaux and Place Saint-Jean. We have installed cameras at the Place des Terreaux to film the area which is represented by the red rectangle in Figure III.1.



(a) View over Place des Terreaux (Lyon). (b) View over the tracking area (red box).

Figure III.1 For the prediction of pedestrian trajectories, it is useful to distinguish between scenes with few pedestrians (see (a)) and crowds of pedestrians (see (b)). Deep learning algorithms turn out to be accurate prediction tools for the scene (a), whereas knowledge-based approaches allow to describe collective phenomena at higher scales such as those described in the scene (b).

In Figure III.1(a), the entirety of Place des Terreaux is depicted. The red box on the right-hand side delineates our designated tracking region. Figure III.1(b) offers an aerial perspective of this same area. This designated zone measures 9 meters in length and 6.5 meters in width. On average, we concurrently tracked 55 pedestrians, resulting in a mean density of 0.95 ped/m^2. The distribution of pedestrian density exhibited significant variability. During the light show, the majority of pedestrians congregated in the central area of the square, remaining largely stationary. Consequently, the pedestrian density within our tracking zone was relatively low. However, when the show concluded—which consistently lasts approximately 9 minutes—the crowd dynamics shifted dramatically as most individuals sought to exit towards another event. In this transition phase, the density within our tracking corridor surged, often exceeding 120 pedestrians moving simultaneously. For video calibration, we meticulously established nine calibration points, ensuring the precise tracking of pedestrian trajectories using the PeTrack software [5]. Throughout our study, we recorded 5195 individual trajectories, which averaged a duration of 12.38 seconds and a mean velocity of 0.62 m/s. The pedestrian flow changes between mostly unidirectional flow, after the light show and bidirectional flow during the show. More informations about the data can be found in the appendix. For the training and testing of the algorithms we need trajectories of a minimal length of 7 seconds (see III.4.1). Because many trajectories are more than 14 seconds long, they can be used more than once. All in all we get 7450 trajectories for training and testing.

III.4 Methodology

III.4.1 Overview

For predicting pedestrian trajectories we have i^{th} pedestrian in a scene represented by image coordinates(x_t^i, y_t^i) for each time instant $t = k \cdot dt$, with $k \in \mathbb{N}$ und $dt = 1/3$ s the time step. The observed positions from $t = T_1$ to $t = T_{obs}$ is taken as input and the aim is to predict future trajectories from $t = T_{obs} + dt$ to $t = T_{pred}$. Every scene involves a primary pedestrian and his neigbours over the timespan T_1 to T_{pred}. A neigbours is a pedestrian whose position at T_1 is closer to the position of the primary pedestrian than a radius $r = 5$ m. Our dataset has a framerate of three observation for each second. We choose input trajectories of 9 observations (3 sec.) and want to predict 12 timesteps (4 sec.).
The predicted trajectory of all primary pedestrians are evaluated on two commonly utilized Euclidean distance metric and a collision metric. In the first distance-based metrics, called average displacement error (ADE) [28], the distance between the predicted trajectory and the ground truth trajectory is measured at any time step t

$$\textbf{ADE} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \|\hat{x}_i(t) - x_i(t)\|. \tag{III.1}$$

$x_i(t)$ is the actual position of the i^{th} pedestrian at time t while $\hat{x}_i(t)$ is the predicted position. The Euclidean distance is denoted as $\| \cdot \|$. The second distance-based metric, called final displacement error (FDE) [23] displays the distance between the final point $t = T_{pred}$ of the predicted trajectory and the ground truth trajectory

$$\textbf{FDE} = \frac{1}{N} \sum_{i=1}^{N} \|\hat{x}_i(T) - x_i(T)\|. \tag{III.2}$$

These distance-based metrics are widely used in pedestrian trajectory predictions for their effectiveness in quantifying the goodness-of-fit. However, repulsive forces, which are pivotal in shaping interactions between pedestrians, are not taken into account [18]. Consequently, these metrics do not account for potential overlaps or collisions between pedestrians. Therefore the collision metric is used to enhance the evaluating process

$$\textbf{COL} = \frac{1}{|S|} \sum_{\hat{X} \in S} COL(\hat{X}), \tag{III.3}$$

with

$$COL(\hat{X}) = \min\left(1, \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j>i}^{N}\left[||\hat{x}_i(t) - \hat{x}_j(t)|| \leq 2R\right]\right). \tag{III.4}$$

S includes all scenes in the test set, $\hat{X}$ represents a scene prediction containing N agents, and $\hat{x}_i$ is the prediction of the position of the agent i over the prediction time of T, while $[\cdot]$ is the Iverson bracket

$$[P] = \begin{cases} 1 & \text{if } P \text{ is true,} \\ 0 & \text{otherwise.} \end{cases} \tag{III.5}$$

This metric counts a prediction as a collision when a predicted pedestrian trajectory intersects with neighboring trajectories, thus indicating the proportion of predictions where collisions occur. A vital factor in this calculation is the chosen pedestrian size R. An increase in R will likewise increase the number of collisions. For calculating the collision metrics we use a radius $R = 0.2$ m.

III.4.2 Prediction approaches

In the subsequent trajectory predictions, various trajectory prediction approaches ranging from traditional PB models to modern DL algorithms are chooses as benchmarks for comparison with our two-stage approach. We present the results of the Constant Velocity model (CV) and SF model [13] as well as the results of a Vanilla LSTM, the Social LSTM (SLSTM) [1] and the Social GAN [12]. These approaches, characterized by their diverse features, are commonly selected for comparison and serve as benchmarks that must be surpassed. The position and velocity of the i^{th} pedestrian are denoted as $x_i \in \mathbb{R}^2$ and $v_i \in \mathbb{R}^2$, respectively. For a system of N pedestrians, the position and velocity vectors, $x = (x_1, \ldots, x_N)$ and $v = (v_1, \ldots, v_N)$, have dimensions of $2N$. All variables, including $x(t)$ and $x_i(t)$, are functions of time t.

Constant Velocity Model

The CV model assumes pedestrian velocities remain unchanged over time. It serves as a baseline for more complex models. The future position of a pedestrian is predicted as:

$$x_i(t + t_p) = x_i(t) + t_p v_i(t), \qquad \forall t_p \in [0, T_p]. \tag{III.6}$$

Social Force Model

Introduced by Helbing and Molnar [13], the SF model treats pedestrians as particles influenced by forces. Within this framework, the model calculates acceleration based on the cumulative effect of three distinct forces, as delineated in Eq. III.7

$$m_i \frac{dv_i}{dt} = m_i \frac{v_i^0 - v_i}{\tau} + \sum_{j \neq i} \nabla U(x_j - x_i) + \sum_{W} \nabla V(x_W - x_i) \tag{III.7}$$

Here, m_i, v_i, and v_i^0 signify the mass, current velocity, and desired velocity of pedestrian i. The term $\nabla U(x_j - x_i)$ represents the repulsive force from other pedestrians, while $\nabla V(x_W - x_i)$ indicates the repulsive force from obstacles. The potential functions $U(d)$ and $V(d)$ are given by:

$$U(d) = ABe^{-|d|/B}, \quad A, B > 0 \qquad \text{and} \qquad V(d) = A'B'e^{-|d|/B'}, \quad A', B' > 0 \tag{III.8}$$

where $A, A', B, B' > 0$ are interaction parameters of the social force model.

The first term of Eq. III.7 signifies the driving force experienced by the i^{th} pedestrian. This force propels the individual towards their desired speed and direction within a relaxation time $\tau > 0$. The second term encapsulates the summation of social forces, originating from the repulsive effects as pedestrians endeavour to maintain a comfortable distance from one another. The third term accounts for the aggregate interaction forces between pedestrian i and various obstacles.

Whereas the CV model has no parameter, the SF model has three parameters, preferred velocity, interaction potential, and reaction time, that can be optimized to get accurate predictions.

Vanilla LSTM

LSTM networks, a class of RNN designed to learn long-term dependencies, have proven effective in handling sequential data, particularly for time series prediction tasks. Introduced by Hochreiter and Schmidhuber [14], LSTMs address the vanishing and exploding gradient problems common in traditional RNNs, making them suitable for complex sequence modeling tasks such as trajectory prediction. The vanilla LSTM model considers historical trajectories to predict future positions.

$$x_i(t + t_p) = x_i(t) + \text{LSTM}\big(t_p, \big(x_i(t - t_o),\ t_o \in [0, T_o]\big)\big), \qquad \forall t_p \in [0, T_p]. \tag{III.9}$$

Here, $x_i(t+t_p)$ predicts the future trajectory of a pedestrian i at time $t+t_p$, based on its past positions $x_i(t-t_o)$, over an observation window $t_o \in [0, T_o]$.

Social LSTM

LSTM networks have demonstrated effective performance in sequence learning tasks. One such task, the prediction of pedestrian trajectories, presents the challenges, that the trajectory of a pedestrian can be significantly influenced by the trajectories of surrounding pedestrians. The number of these neighboring influences can fluctuate widely, especially in densely crowded environments [19].

Enhancing the LSTM framework, the SLSTM by Alahi et al. [1] incorporates a social pooling layer, enabling the model to consider the influence of neighboring pedestrians explicitly. This is a key distinction from the Vanilla LSTM, reflecting the model's capacity to capture social interactions:

$$x_i(t+t_p) = x_i(t) + \mathrm{SLSTM}\big(i, t_p, \big(x(t-t_o),\ t_o \in [0, T_o]\big)\big), \qquad \forall t_p \in [0, T_p]. \tag{III.10}$$

where x is the vector of positions of the neighboring pedestrians. In this formulation, the inclusion of the index i and the collective pedestrian state x emphasizes the model's attention to the surrounding pedestrians' trajectories, making it adept at handling complex social behaviors in dense scenarios.

Social GAN

Another approach we take into account is the Social GAN (SGAN) introduced by Gupta et al. [12]. This model extends traditional approaches by incorporating GANs to predict future trajectories. GANs, conceptualized by Goodfellow et al. [11], consist of two competing networks: a Generator, which generates data samples, and a Discriminator, which evaluates the authenticity of the samples against real data. SGAN leverages this architecture to generate plausible future trajectories of pedestrians, addressing the complex dynamics of pedestrian movement in crowded spaces. A key feature of the SGAN model is its pooling mechanism, which processes the relative positions of pedestrians to each other. This mechanism is crucial for understanding the social interactions and dependencies among individuals in crowded environments

$$x_i(t+t_p) = x_i(t) + \mathrm{SGAN}\big(i, t_p, \big(x(t-t_o),\ t_o \in [0, T_o]\big)\big), \qquad \forall t_p \in [0, T_p]. \tag{III.11}$$

III.4.3 Two-stage process

The foundation of our innovative classification framework lies in its capacity to predict trajectories across varying density levels, marking a departure from traditional models that typically utilize a single algorithm to process a wide array of scenarios within a dataset. Our strategy entails segmenting the dataset according to the density of each scene, thereby generating distinct subsets. At the inception of our methodology, we establish well-defined criteria for classification. This process is underpinned by two distinct methodologies: a statistical analysis and a review of existing literature. The results of this clustering process are depicted in Figures III.2. These figures visually represent each dataset item as a point measurement each second, with the left side of Figures III.2 illustrating points based on their average density

$$\rho(t) = \frac{N(t)}{A}, \tag{III.12}$$

and average velocity

$$\bar{v}(t) = \frac{1}{N(t)} \sum_{i \in \mathcal{S}(t)} v_i(t), \tag{III.13}$$

for the K-Means clustering. On the right side the clustering is carried out by using the Agglomerative Hierarchical Clustering algorithm (AHC).

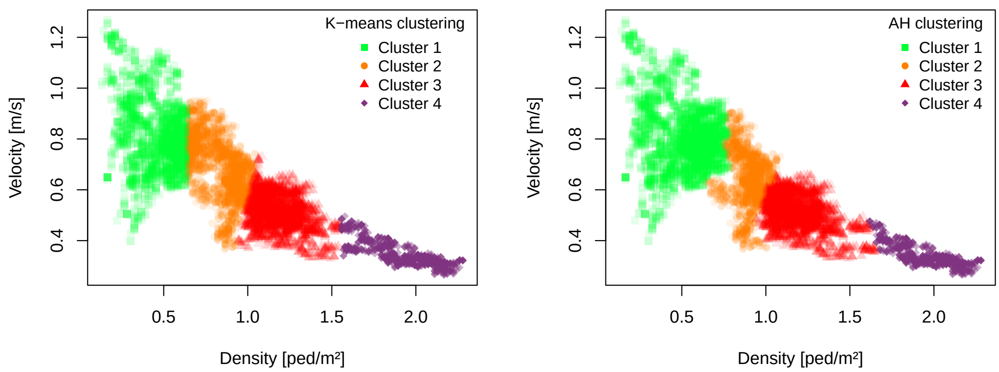


Figure III.2 Results of the K-Means and the agglomerative hierarchical clustering. Trajectory scenes are clustered as shown by the different colors of the points.

We can see clear vertical colour switch's of the points at densities around 0.7, 1.1, and 1.6 ped/m^2 for both cluster algorithms. Remarkably, without presetting the number of clusters or explicitly focusing on density levels, the K-Means and

AHC algorithms autonomously reveal density-dependent clustering. The delineation of clusters and their boundary values align closely with those identified in the literature. Stefan Holl [15] delineates critical density thresholds for various infrastructures, signifying points at which pedestrian behavior undergoes significant changes. According to Holl, densities below 0.7 ped/m^2 indicate a free flow state, densities below 1.3 ped/m^2 represent a bound flow, and values above 1.3 ped/m^2 are indicative of congested flow. In our model, we refine these categories by slightly narrowing the bound flow range and subdividing the congested flow category into two distinct segments.

Figure III.3 illustrates the procedural steps undertaken to evaluate our proposed methodology and within the sizes of the clusters, which are taken from Figure III.2.

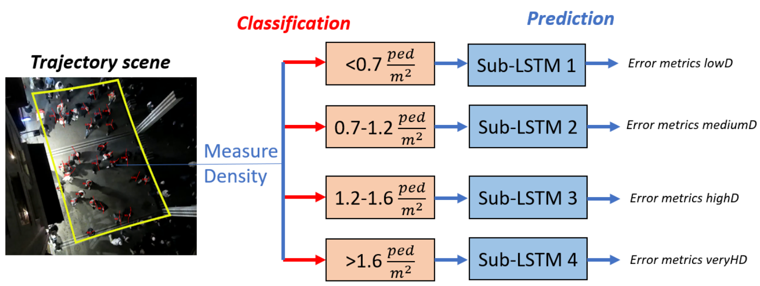
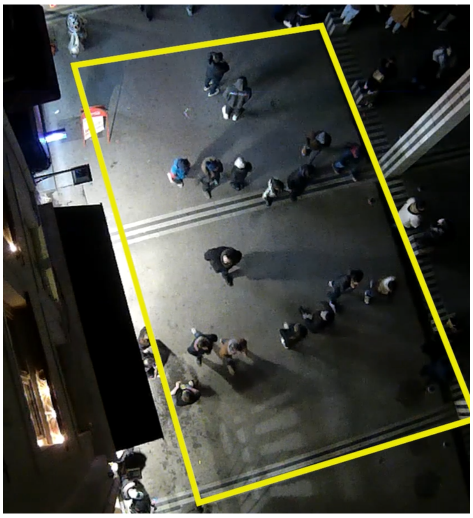


Figure III.3 Schemata of our two-stage prediction approach

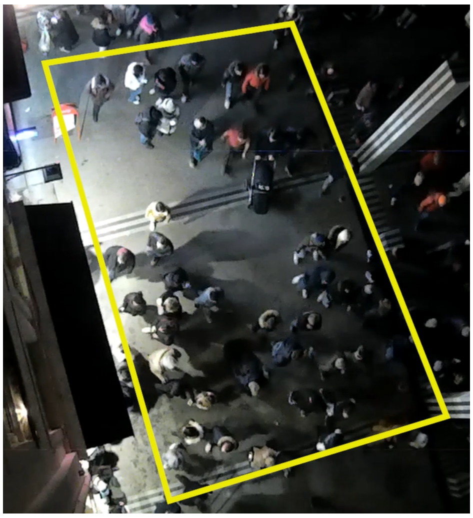
New trajectory scene are given to our framework, where the initial step involves calculating the scene's density using Equation III.12 in individuals per square meter (ped/m^2), N represents the total number of pedestrians observed within the scene, and A is the scene's total area in square meters.

The density categorization is as follows: scenes with a density below 0.7 ped/m^2 are labelled as lowD; densities ranging from 0.7 to 1.2 ped/m^2 are classified as mediumD; densities between 1.2 and 1.6 ped/m^2 are designated as highD; and densities exceeding 1.6 ped/m^2 are identified as veryHD. Following classification, the scene is bifurcated into two segments: the initial segment spans 9 timesteps and serves as input for one of the four specialized Sub-LSTMs, while the subsequent segment, encompassing 12 timesteps, is utilized to appraise the LSTMs' performance through the computation of error metrics ADE, FDE and COL. Figures III.4(a) to III.4(d) provide illustrative examples of each density level encountered in our dataset.

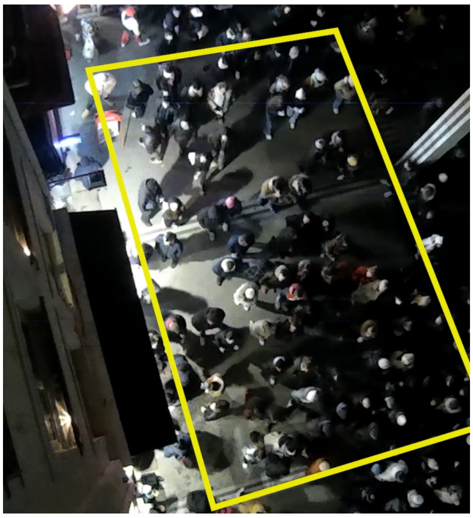
In Figure III.4(a), the scene exhibits very low density, with pedestrian movement primarily from two directions, leading to numerous interactions and avoidance behaviors. This is characteristic of our lowD data. Conversely, Figure



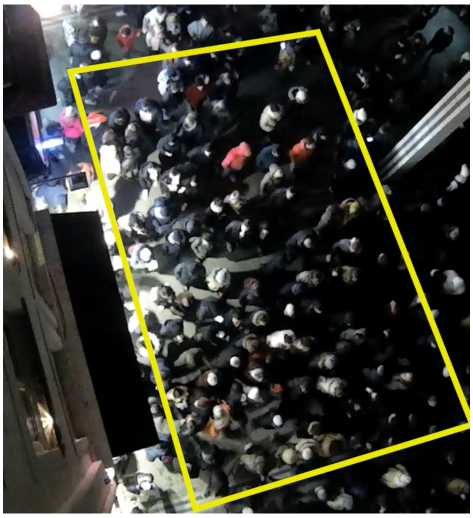
(a) Example of lowD scene.



(b) Example of mediumD scene.



(c) Example of highD scene.



(d) Example of veryHD scene.

Figure III.4 Examples for each of the four density levels.

III.4(b) showcases a moderately higher density, yet still affords space for interactions, avoidance, and bidirectional pedestrian flow. In Figure III.4(c), representing highD data, the dynamics of pedestrian movement markedly differ from those observed in lowD and mediumD scenes, with movement predominantly unidirectional from the top, indicating a tendency to follow the pedestrian ahead. This pattern is even more pronounced in Figure III.4(d), where the flow from the top is so dense that passage from the bottom becomes challenging, leading pedestrians to follow the leader with limited freedom of movement and space. These observed behavioral differences underpin our classification rationale.

III.4.4 Collision weight

Predictions of pedestrian trajectories presents the challenge to predict trajectory paths that do not collide with neighbours. Accurately measuring these collisions is challenging due to the shapes of pedestrians, which can vary from person to person. Traditionally, collisions are defined by the overlap of the radii of two pedestrians, as delineated in Equations III.3 to III.5. However, this method proves sub optimal for inclusion as a penalty function within the loss function of DL algorithms. Analysis of pedestrian trajectory data frequently reveals instances where collisions are not genuine but rather instances of grouping behavior, with individuals walking closely, sometimes shoulder-to-shoulder. It is not these interactions we aim to deter, but rather scenarios in which individuals move directly towards one another without any attempt to avoid collision—behaviors that are unrealistic and undesirable.
To address this, we adopt the TTC concept, a widely recognized principle in the study of pedestrian dynamics [16]. Implementing this variable in the loss function of an DL algorithm would reduce predicted situations, where pedestrians walk straight towards each other without avoidance mechanism. Integrating TTC into a DL algorithm's loss function significantly mitigates predictions where pedestrians are on a direct collision course without any avoidance mechanisms. The TTC term calculates the time until two pedestrians would collide if they continue moving at their current velocities, a concept validated by Karamouzas et al. [16]. The relative position and velocity between the pedestrian i and j can be denoted by $x_{ij} = x_i - x_j \in \mathbb{R}^2$ and $v_{ij} = v_i - v_j \in \mathbb{R}^2$, respectively. A collision between pedestrian i and pedestrian j occurs if a ray, originating from (x_i, y_i) and extending in the direction of v_{ij}, intersects the circle centred at (x_j, y_j) with a radius of $R_i + R_j$ at some time τ_{ij} in the future. This condition can be mathematically represented as $||x_{ij} + v_{ij}.t||^2 < (R_i + R_j)^2$ where $||.||$ denotes Euclidean norm. Solving this quadratic inequality for t yields τ_{ij} as the smallest positive root:

$$\tau_{ij} = \frac{-x_{ij} \cdot v_{ij} - \sqrt{(x_{ij} \cdot v_{ij})^2 - ||v_{ij}||^2(||x_{ij}||^2 - (R_i + R_j)^2)}}{||v_{ij}||^2} \tag{III.14}$$

A collision is imminent when $\tau_{ij} = 0$, whereas a large positive value for τ_{ij} indicates no collision risk. To implement τ_{ij} into the loss function we have to use an sigmoid function f that has high values, if τ_{ij} is low and vice versa:

$$f(\tau) = \frac{1}{1 + e^{s(\tau - \delta)}}, \tag{III.15}$$

where $s = 10$ and $\delta = 0.4$ are slope and threshold parameters, respectively. This function is then integrated into the loss function, traditionally focused solely on minimizing the ADE. The revised loss function combines ADE with TTC loss, optimized through minimization:

$$\mathbf{L}_i = \sum_{t=1}^{T} \|x_i(t) - \hat{x}_i(t)\| + \lambda \sum_{t=1}^{T} f(\min_{j \neq i}\{\tau_{ij}\}), \tag{III.16}$$

where $\lambda > 0$ modulates the influence of the TTC component in their loss function. The calculation of τ_{ij} considers all nearby pedestrians to the primary pedestrian, employing the minimum τ_{ij} to identify and mitigate the most critical potential collision scenario in the model.

III.4.5 Implementation details

The algorithms are implemented in the commonly accepted configurations of related contributions [22]. All computations are performed using the PyTorch framework. The learning rate is set to 0.001 and an ADAM optimizer is utilizied. The batch size is set to 8 and training is carried out for 15 epochs, if not the early stop mechanism interrupt. This is the case, when the validations error starts to rise for three epochs. For validation and testing, a hold-out validation strategy is adopted by allocating 15% of the dataset for each validation and testing, while the remaining data serves as the training set. For capturing pedestrian interactions, we choose a circles with a radius of $r = 5$ m surrounding the primary pedestrian.

III.5 Results

We will unveil the predictive outcomes of our dataset using two distinct yet synergistic methods. Initially, we will showcase the performance of our two-stage prediction framework, comparing it with contemporary state-of-the-art methodologies. Subsequently, we will demonstrate the seamless integration of our two-stage process with the incorporation of the TTC term into the loss function, illustrating its efficacy in mitigating collision instances.

III.5.1 Two-Stage Predictions

The results of our predictions will be presented in Table III.1. As described in Sec. III.4.3 we evaluated the predictions on the different density levels lowD, mediumD, highD, and veryHD. For every approach we measure ADE, FDE and COL metrics.

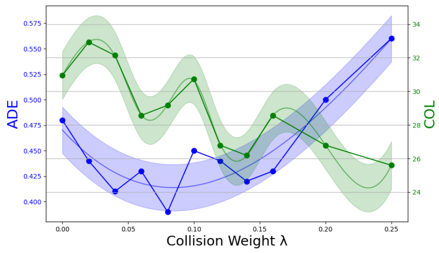
Table III.1 Quantitative comparison of ADE and FDE metrics for articles using Social LSTM as benchmark with different datasets.

Model	LowD		MediumD		HighD		VeryHD	
	ADE/FDE	**COL**	**ADE/FDE**	**COL**	**ADE/FDE**	**COL**	**ADE/FDE**	**COL**
CV	0.71/0.97	54.76	0.85/0.98	45.73	0.53/0.8	62.35	0.44/0.67	81.74
Social Force [13]	0.78/1.33	**24.4**	0.55/0.89	**31.16**	0.5/0.82	36.43	0.36/0.63	54.78
Vanilla LSTM	0.5/0.99	31.55	0.33/0.63	37.69	0.29/0.52	36.43	**0.24/0.41**	63.8
Social LSTM [1]	0.53/1.02	57.74	0.37/0.73	59.3	0.41/0.78	64.26	0.35/0.66	75.37
Social GAN [12]	0.53/0.99	31.36	0.39/0.72	32.16	0.36/0.61	**32.33**	0.25/0.41	55.94
Our 2stg. SLSTM	0.48/0.93	30.95	**0.3/0.63**	36.18	**0.26/0.4**	42.02	**0.24/0.41**	**52.23**
Our 2stg. SGAN	**0.44/0.83**	32.74	**0.27/0.52**	40.2	0.28/0.5	**35.33**	0.26/0.43	58.6
Our 2stg. TTC-SLSTM	**0.39/0.73**	**29.17**	**0.3/0.62**	**22.61**	**0.23/0.36**	36.29	**0.24/0.41**	**52.23**

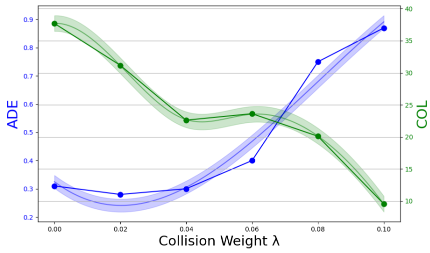
The initial insight gleaned from Table III.1 reveals a clear trend: as density increases, the COL metric rises while the ADE/FDE diminish. This pattern emerges because higher densities naturally lead to reduced distances between individuals, consequently resulting in increased overlaps among agents. Additionally, it's observed that velocities decrease as density intensifies, leading to trajectories that are shorter in spatial extent. This reduction in travel distance directly contributes to the observed decrease in both ADE and FDE metrics at higher densities. Furthermore, it is clear that the DB algorithm outperform the traditional models CV and SF in terms of ADE/FDE. In terms of COL metric SF performs very well.
In the last three rows of Table III.1, we present the effectiveness of our two-stage approach and its combination with the TTC term. The results clearly show a significant improvement in the algorithm's precision, attributed to the strategy of classification before prediction. Our enhanced two-stage SLSTM model consistently outperforms the traditional SLSTM across all evaluated datasets, demonstrating superior performance in terms of ADE, FDE, and COL metrics. Similarly, our adapted SGAN model shows marked improvements over the standard SGAN in three out of four datasets with respect to ADE. Integrating the TTC term further enhances the SLSTM results, notably in reducing collisions. A more detailed discussion on this enhancement is provided in the subsequent section III.5.2.

III.5.2 Collision weight

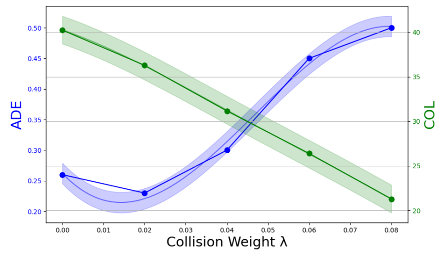
In this study, we propose to integrate TTC in the loss function with the two-stage approach outlined in Sec. III.4.3. As described in Eq. III.16 the collision part in the loss function can be adjusted by a parameter λ [18]. If λ is high, the impact of the TTC term is high compared to the impact of ADE and vice versa. In the following diagram, the impact of different values of λ on the prediction accuracy of the SLSTM algorithm is displayed. In Figure III.5(a) for lowD data in Figure III.5(b) for mediumD data, in Figure III.5(c) for highD data, and in Figure III.5(d) for veryHD data. The first observation in each Figures is for $\lambda = 0$, which means, that it is equivalent to the value of our two-stage SLSTM in Table III.1.



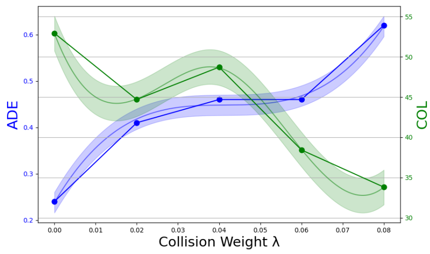
(a) Low density (lowD)



(b) Medium density (mediumD)



(c) High density (highD)



(d) Very high density (veryHD)

Figure III.5 ADE and COL metrics for the two-stage SLSTM algorithm according to the collision weight λ for the four density levels.

Across all figures, a consistent pattern emerges: increasing the collision weight λ generally results in fewer collisions. Optimal predictions occur at a specific λ value, where ADE and FDE are equivalent to or lower than those at $\lambda = 0$. Each dataset exhibits a maximum effective λ value beyond which ADE sharply increases. In the lowD dataset (Figure III.5(a)), the ideal λ is 0.08, reducing ADE by 19% and collisions by 6%. Values slightly higher than 0.08 are still beneficial, yielding fewer collisions and enhanced avoidance behavior, but λ values exceeding 0.16 lead to a significant increase in ADE. In the mediumD dataset

(Figure III.5(b)), the optimal λ is 0.04, reducing ADE and collisions by 3% and 40%, respectively. Here, a λ value above 0.06 results in increased ADE, although a λ of 0.1 reduces the collision metric by 75%. In the highD dataset, improvements in ADE are marginal too, only notable at a λ of 0.02. However, the collision metric significantly decreases, by up to 48%, at a λ of 0.08. Conversely, in the very high-density (veryHD) dataset, increasing collision weight initially results in a rise in ADE, with no subsequent decrease. While the collision metric decreases by 37% at $\lambda = 0.08$ values, there's no improvements for ADE. These empirical observations lead to the insight, that pedestrian behavior at different densities is very different and need different parameter configurations. At lower densities our TTC term can improve overall accuracy (ADE and COL) in higher densities we can only reduce COL, by taking higher ADE into account.

III.6 Conclusions

Pedestrian behavior is inherently complex, exhibiting a wide variety of patterns across different contexts. This paper introduced a novel pedestrian trajectory dataset, characterized by its diversity in situational contexts, including varying densities and motivations. Our analysis of the data reveals variations in pedestrian behaviors correlating with the density of the scene. To address these variations, we propose a novel two-stage classification and prediction process. This approach first classifies scenes based on density and then applies the suitable model for predicting behavior within that specific density context. Implementing this framework enhanced the prediction accuracy of two famous DL algorithms, Social LSTM and Social GAN.
Further, we integrated a TTC based term into the loss function of the SLSTM to improve avoidance behaviors, consequently reducing potential collisions. Our empirical studies indicate that the effectiveness of the TTC-based term varies with density; it significantly benefits scenarios of low density by correlating higher TTC values with reduced collision incidents. However, the outcomes in high-density situations were more ambiguous, suggesting a nuanced impact of density on the efficacy of this approach. This observation could be attributed to the nuanced dynamics of pedestrian behavior across different densities. Specifically, in environments with lower densities, pedestrians tend to navigate more through avoidance and interactions, making TTC particularly relevant. Conversely, in higher density settings, pedestrian movement is more characterized by forced leader-follower dynamics, diminishing the prominence of TTC in explaining behavior. This study underscores the complexity of pedestrian behavior, which varies significantly under different environmental conditions. It highlights the necessity of adopting a flexible modeling approach to accurately predict pedestrian

trajectories in diverse settings.

This research opens several avenues for future investigation in the field of pedestrian trajectory prediction, especially concerning heterogeneous datasets characterized by variable densities. Current methodologies typically rely on a one-size-fits-all model for behavior prediction across all conditions. We advocate for the development and application of multiple specialized models, each tailored to different scene characteristics, with scene density being a pivotal factor. While focusing on density has proven to be a successful strategy, exploring additional factors and individual pedestrian behavior characteristics could yield further improvements. For instance, our methodology utilized an estimate of overall scene density. However, pedestrians do not take global densities for there decision-making into account, but rather local densities. Wirth et al. [36] demonstrate that pedestrian decisions are primarily influenced by their visual neighborhood. Future studies should investigate the impact of assessing local density variations within a scene, which could be particularly beneficial in environments exhibiting a wide range of density levels. This direction could unlock new dimensions of accuracy and reliability in trajectory prediction models.

Moreover, incorporating the TTC concept into the loss function has shown promise in enhancing prediction accuracy at lower density levels. Future research should explore alternative loss functions, particularly for high-density scenarios, where the traditional ADE based approaches may not suffice. Investigating other metrics that could more effectively capture the complexities of high-density pedestrian behavior is crucial for advancing the field.

Bibliography

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 961–971, 2016.

[2] Alessandro Antonucci, Gastone Pietro Rosati Papini, Paolo Bevilacqua, Luigi Palopoli, and Daniele Fontanelli. Efficient prediction of human motion for real-time robotics applications with physics-inspired neural networks. *IEEE Access*, 10:144–157, 2021.

[3] Nicola Bellomo and Christian Dogbe. On the modeling of traffic and crowds: A survey of models, speculations, and perspectives. *SIAM review*, 53(3):409–463, 2011.

[4] M Boltes, J Zhang, A Tordeux, A Schadschneider, and A Seyfried. Empirical results of pedestrian and evacuation dynamics. *Encyclopedia of complexity and systems science*, 16:1–29, 2018.

[5] Maik Boltes and Armin Seyfried. Collecting pedestrian trajectories. *Neurocomputing*, 100:127–133, 2013. Special issue: Behaviours in video.

[6] Carsten Burstedde, Kai Klauck, Andreas Schadschneider, and Johannes Zittartz. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications*, 295(3-4):507–525, 2001.

[7] Mohcine Chraibi, Antoine Tordeux, Andreas Schadschneider, and Armin Seyfried. Modelling of pedestrian and evacuation dynamics. *Encyclopedia of complexity and systems science*, pages 1–22, 2018.

[8] Huu-Tu Dang, Raphael Korbmacher, Antoine Tordeux, Benoit Gaudou, and Nicolas Verstaevel. Ttc-slstm: Human trajectory prediction using time-to-collision interaction energy. In *2023 15th International Conference on Knowledge and Systems Engineering (KSE)*, pages 1–6. IEEE, 2023.

[9] Dorine C Duives, Winnie Daamen, and Serge P Hoogendoorn. State-of-the-art crowd motion simulation models. *Transportation research part C: emerging technologies*, 37:193–209, 2013.

[10] Alessandro Gabbana, Federico Toschi, Philip Ross, Antal Haans, and Alessandro Corbetta. Fluctuations in pedestrian dynamics routing choices. *PNAS nexus*, 1(4):pgac169, 2022.

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[12] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2255–2264, 2018.

[13] Dirk Helbing and Peter Molnar. Social force model for pedestrian dynamics. *Physical Review E*, 51(5):4282, 1995.

[14] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[15] Stefan Holl. *Methoden für die Bemessung der Leistungsfähigkeit multidirektional genutzter Fußverkehrsanlagen*. Number FZJ-2017-00069. Jülich Supercomputing Center, 2016.

[16] Ioannis Karamouzas, Brian Skinner, and Stephen J Guy. Universal power law governing pedestrian interactions. *Physical review letters*, 113(23):238701, 2014.

[17] A Khadka, Paolo Remagnino, and Vasileios Argyriou. Synthetic crowd and pedestrian generator for deep learning problems. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4052–4056. IEEE, 2020.

[18] Raphael Korbmacher, Huu-Tu Dang, and Antoine Tordeux. Predicting pedestrian trajectories at different densities: A multi-criteria empirical analysis. *Physica A: Statistical Mechanics and its Applications*, 634:129440, 2024.

[19] Raphael Korbmacher, Tu Dang-Huu, Antoine Tordeux, Nicolas Verstaevel, and Benoit Gaudou. Differences in pedestrian trajectory predictions for high-and low-density situations. In *14th International Conference on Traffic and Granular Flow (TGF) 2022*, pages à–paraître. Springer, 2022.

[20] Raphael Korbmacher and Antoine Tordeux. Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches. *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[21] Parth Kothari and Alexandre Alahi. Safety-compliant generative adversarial networks for human trajectory forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 24(4):4251–4261, 2023.

[22] Parth Kothari, Sven Kreiss, and Alexandre Alahi. Human trajectory forecasting in crowds: A deep learning perspective. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):7386–7400, 2021.

[23] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by example. In *Computer graphics forum*, volume 26, pages 655–664. Wiley Online Library, 2007.

[24] Yang Li, Huahu Xu, Minjie Bian, and Junsheng Xiao. Attention based cnn-convlstm for pedestrian attribute recognition. *Sensors*, 20(3):811, 2020.

[25] Barbara Majecka. Statistical models of pedestrian behaviour in the forum. *Master's thesis, School of Informatics, University of Edinburgh*, 2009.

[26] Alessio Monti, Alessia Bertugli, Simone Calderara, and Rita Cucchiara. Dag-net: Double attentive graph neural network for trajectory forecasting. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2551–2558. IEEE, 2021.

[27] Vasileia Papathanasopoulou, Ioanna Spyropoulou, Harris Perakis, Vassilis Gikas, and Eleni Andrikopoulou. A data-driven model for pedestrian behavior classification and trajectory prediction. *IEEE Open Journal of Intelligent Transportation Systems*, 3:328–339, 2022.

[28] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th international conference on computer vision*, pages 261–268. IEEE, 2009.

[29] Atanas Poibrenski, Matthias Klusch, Igor Vozniak, and Christian Müller. M2p3: multimodal multi-pedestrian path prediction by self-driving cars with egocentric vision. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pages 190–197, 2020.

[30] Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. Learning social etiquette: Human trajectory understanding in crowded scenes. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*, pages 549–565. Springer, 2016.

[31] Stefano Scheggi, Marco Aggravi, Fabio Morbidi, and Domenico Prattichizzo. Cooperative human-robot haptic navigation. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2693–2698. IEEE, 2014.

[32] Xiaodan Shi, Xiaowei Shao, Zhiling Guo, Guangming Wu, Haoran Zhang, and Ryosuke Shibasaki. Pedestrian trajectory prediction in extremely crowded scenarios. *Sensors*, 19(5):1223, 2019.

[33] Mattia Silvestri, Michele Lombardi, and Michela Milano. Injecting domain knowledge in neural networks: a controlled experiment on a constrained problem. In *Integration of Constraint Programming, Artificial Intelligence, and Operations Research: 18th International Conference, CPAIOR 2021, Vienna, Austria, July 5–8, 2021, Proceedings 18*, pages 266–282. Springer, 2021.

[34] Jur Van Den Berg, Stephen J Guy, Ming Lin, and Dinesh Manocha. Reciprocal n-body collision avoidance. In *Robotics Research: The 14th International Symposium ISRR*, pages 3–19. Springer, 2011.

[35] Anirudh Vemula, Katharina Muelling, and Jean Oh. Social attention: Modeling attention in human crowds. In *2018 IEEE international Conference on Robotics and Automation (ICRA)*, pages 4601–4607. IEEE, 2018.

[36] Trenton D Wirth, Gregory C Dachner, Kevin W Rio, and William H Warren. Is the neighborhood of interaction in human crowds metric, topological, or visual? *PNAS nexus*, 2(5):pgad118, 2023.

[37] Hao Xue, Du Q Huynh, and Mark Reynolds. Bi-prediction: Pedestrian trajectory prediction based on bidirectional lstm classification. In *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8. IEEE, 2017.

[38] Hao Xue, Du Q Huynh, and Mark Reynolds. Poppl: Pedestrian trajectory prediction by lstm with automatic route class clustering. *IEEE transactions on neural networks and learning systems*, 32(1):77–90, 2020.

[39] Hai-Yan Yao, Wang-Gen Wan, and Xiang Li. End-to-end pedestrian trajectory forecasting with transformer network. *ISPRS International Journal of Geo-Information*, 11(1):44, 2022.

[40] Shuai Yi, Hongsheng Li, and Xiaogang Wang. Pedestrian behavior understanding and prediction with deep neural networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 263–279. Springer, 2016.

[41] Bolei Zhou, Xiaogang Wang, and Xiaoou Tang. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents.

In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2871–2878. IEEE, 2012.

Appendix

Curriculum Vitæ

Raphael Korbmacher

Mainzer Str. 35, 42119 Wuppertal | korbmacher@uni-wuppertal.de

Education

2021 –	**PhD Candidate, Universität Wuppertal** Deep Learning for trajectory prediction and autonomous vehicles. Data-based approaches for reliability engineering.
2020 – 2024	**M.Sc. Economics, FernUniversität in Hagen** Thesis topic: *Sentiment Data in Risk Management: A Machine Learning Approach.*
2018 – 2021	**M.Sc. Quality Engineering, University of Wuppertal** Thesis topic: *Development of Heuristic and Machine Learning Models to Predict Lane Changes for Autonomous Vehicles.*
2015 – 2020	**B.Sc. Economics, FernUniversität in Hagen** Thesis topic: *Neural Networks for the Prediction of Macroeconomic Indicators.*
2015 – 2018	**B.Sc. Safety Engineering, University of Wuppertal** Thesis topic: *Multivariate Methods for Reliability Analysis using R and KNIME.*
2007 – 2015	**Abitur, Gymnasium Haan**

Major publications

1. **Raphael Korbmacher**, and Antoine Tordeux. "Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches." *IEEE Transactions on Intelligent Transportation Systems* 23.12 (2022): 24126-24144.

2. **Raphael Korbmacher**, Huu-Tu Dang, and Antoine Tordeux. "Predicting pedestrian trajectories at different densities: A multi-criteria empirical analysis." *Physica A: Statistical Mechanics and its Applications* 634 (2024): 129440.

3. **Raphael Korbmacher**, and Antoine Tordeux. "Towards Better Predictive Models: The Role of Density in Pedestrian Trajectory Predictions." *Sensors* 24.7 (2024): 2356.

Further publications

4. Basma Khelfa, **Raphael Korbmacher**, Andreas Schadschneider, and Antoine Tordeux "Heterogeneity-induced lane and band formation in self-driven particle systems." *Scientific reports* 12.1 (2022): 4768.

5. Basma Khelfa, **Raphael Korbmacher**, Andreas Schadschneider, and Antoine Tordeux "Initiating lane and band formation in heterogeneous pedestrian dynamics." *Collective Dynamics* 6 (2021): 1-13.

6. Huu-Tu Dang, **Raphael Korbmacher**, Antoine Tordeux,Benoit Gaudou, & Nicolas Verstaevel, "TTC-SLSTM: Human trajectory prediction using time-to-collision interaction energy." 15th International Conference on Knowledge and Systems Engineering (KSE). *IEEE*, 2023.

7. **Raphael Korbmacher**, Huu-Tu Dang, Antoine Tordeux, Benoit Gaudou, & Nicolas Verstaevel. "Differences in pedestrian trajectory predictions for high-and low-density situations." 14th International Conference On Traffic And Granular Flow (TGF) *Springer*, 2022.

8. **Raphael Korbmacher**, Alexandre Nicolas, Antoine Tordeux, & Claudia Totzeck. "Time-continuous microscopic pedestrian models: an overview." *Crowd Dynamics, Volume 4: Analytics and Human Factors in Crowd Modeling* (2023): 55-80.

9. **Raphael Korbmacher**, and Antoine Tordeux. "Deep Learning for Predicting Pedestrian Trajectories in Crowds." *Intelligent Systems Conference*. Cham: Springer Nature Switzerland, 2023.

10. Oscar Dufour, Huu-Tu Dang, Jakob Cordes, **Raphael Korbmacher**, Gaudou Benoit, Mohcine Chraibi, Alexandre Nicolas, Antoine Tordeux. "Dense Crowd Dynamics and Pedestrian Trajectories: A Multiscale Field Study at the Fête des Lumières in Lyon"

Project participation
Part of DFG and ANR project Madras (Multi-agent modelling of dense crowd dynamics: Predict & Understand)

Teaching
Reliability engineering with Python lab

Conferences
17ième Journées de la Matière Condensée 2021, 10th International Conference on Pedestrian and Evacuation Dynamics 2021, Traffic and Granular Flow 2022, 2nd GAMA Days Conference 2022, Pedestrian and Evacuation Dynamics 2023

Awards
Runner-up paper award at 15th International Conference on Knowledge and Systems Engineering 2023 for the work "TTC-SLSTM: Human trajectory prediction using time-to-collision interaction energy."