IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Long-Term Human Motion Prediction Using Spatio-Temporal Maps of Dynamics

Yufei Zhu[1], Andrey Rudenko[2], Tomasz P. Kucner[3], Achim J. Lilienthal[1,2], Martin Magnusson[1]

***Abstract*—Long-term human motion prediction (LHMP) is important for the safe and efficient operation of autonomous robots and vehicles in environments shared with humans. Accurate predictions are important for applications including motion planning, tracking, human-robot interaction, and safety monitoring. In this paper, we exploit Maps of Dynamics (MoDs), which encode spatial or spatio-temporal motion patterns as environment features, to achieve LHMP for horizons of up to 60 seconds. We propose an MoD-informed LHMP framework that supports various types of MoDs and includes a ranking method to output the most likely predicted trajectory, improving practical utility in robotics. Further, a time-conditioned MoD is introduced to capture motion patterns that vary across different times of day. We evaluate MoD-LHMP instantiated with three types of MoDs. Experiments on two real-world datasets show that MoD-informed method outperforms learning-based ones, with up to 50% improvement in average displacement error, and the time-conditioned variant achieves the highest accuracy overall. Project code is available at https://github.com/test-bai-cpu/LHMP-with-MoDs.git**

***Index Terms*—Human Detection and Tracking, Human and Humanoid Motion Analysis and Synthesis, Probability and Statistical Methods, Human-Aware Motion Planning**

I. Introduction

ENSURING safe and efficient operation of robots in complex and dynamic environments, particularly in the presence of humans, is critical for deploying robotic systems to assist with a wide range of real-world tasks [1, 2]. A key element in achieving this goal is long-term human motion prediction, i.e., anticipating the trajectories of individuals over extended periods. Accurate long-term prediction of future trajectories is a fundamental requirement for various applications, including optimized motion planning, refined tracking, advanced automated driving, improved human-robot interaction, and enhanced intelligent safety monitoring and surveillance.

Human motion is complex, influenced by various factors. These include not only an individual's intrinsic intent and dynamics but also external influences such as social conventions and environmental cues [3]. Predicting trajectories over extended periods, such as 20 seconds or more, requires careful consideration of the impact of large-scale environments on human behavior. While short-term predictions can often rely on current state and immediate interactions, long-term predictions demand more comprehensive modeling to capture how the environment influences and guides human movement.

An effective approach to address long-term human motion prediction (LHMP) is through the use of maps of dynamics (MoDs), which encode spatial or spatio-temporal motion patterns as a feature of the environment. Prior work, CLiFF-LHMP [4], exploits the CLiFF-map representation [5], which is a specific type of MoD that stores a multi-modal, continuous joint distribution of speed and orientation for each discrete map location, to predict human trajectories over long-term horizons. In this work, we extend CLiFF-LHMP to a general MoD-informed LHMP framework, named *MoD-LHMP*, which can be applied with various types of MoDs. By using MoDs, motion prediction can utilize previously observed motion patterns. We also introduce a *ranking* method that enables the prediction of the most likely trajectory output, enhancing its practical utility for robotic applications.

We instantiate MoD-LHMP with multiple types of MoDs, including: (1) the original CLiFF-map; (2) a Time-Conditioned CLiFF-map, which we introduce in this work; and (3) STeF-map [6], a spatio-temporal MoD designed to capture periodic motion patterns. Given that human movement in the same environment varies over time, in this work, we address the problem of capturing spatio-temporal motion patterns and use them for LHMP. We present the Time-Conditioned CLiFF-map, which adapts to varying motion patterns at different times of day. The methods are also compared with Trajectron++ [7], LSTM-based human motion prediction methods [8], a diffusion-based model [9] and a transformer-based model [10] . The evaluation uses two real-world datasets: the ATC [11] and the Edinburgh dataset [12], both capturing open indoor environments. Both datasets span multiple days and exhibit variations in human motion patterns throughout the day. Through this comparative study, we aim to evaluate the performance of spatio-temporal MoDs in the LHMP task.

In summary, we make the following contributions:

- We extend CLiFF-LHMP by introducing a ranking method, enabling most-likely output predicted trajectory, improving its practical applicability in robotics.
- We demonstrate how the framework can be instantiated with a range of different map representations, thus providing a general MoD-LHMP framework.

Manuscript received: May 29, 2025; Revised August 31, 2025; Accepted September 23, 2025. This paper was recommended for publication by Editor Markus Vincze upon evaluation of the Associate Editor and Reviewers' comments. This work received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017274 (DARKO) and 101070596 (euRobin). *(Corresponding author: Yufei Zhu.)*

[1]Yufei Zhu, Achim J. Lilienthal, and Martin Magnusson are with the Centre for Applied Autonomous Sensor Systems (AASS), Örebro University, Sweden {yufei.zhu, achim.lilienthal, martin.magnusson}@oru.se

[2]Andrey Rudenko and Achim J. Lilienthal are with Technical University of Munich, Germany {andrey.rudenko, achim.j.lilienthal}@tum.de

[3]Tomasz P. Kucner is with department of electrical engineering and automation, school of electrical engineering, Aalto University, Finland tomasz.kucner@aalto.fi

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

- We introduce Time-Conditioned CLiFF-map, a temporal variant of CLiFF-map, to improve prediction accuracy.
- We analyze how different instantiations of MoD-LHMP perform on two real-world datasets, compared with learning-based baselines.

II. Related work

Human motion prediction has been studied extensively in recent years. With different prediction horizons, the human motion prediction problem can be short-term (1–2 s), long-term (up to 20 s) [13], and extended long-term (which we define as over 20 s). In this work, we focus on extended LHMP.

Based on the underlying principle for the motion model, motion prediction approaches can be categorized into planning-, pattern- and physics-based approaches [13]. In planning-based methods, prior knowledge of potential goals in the environment can be used. Ikeda et al. [14] use the concept of sub-goals, i.e., way-points that pedestrians tend to move toward before reaching the final destination. After retrieving the position of the sub-goals, this method models the long-term behavior of pedestrians through transition probabilities between sub-goals. Bruckschen et al. [15] propose an approach that uses knowledge about typical human-object interaction sequences. This method learns a transition model of subsequent object interaction and utilizes it to infer the navigation goal using a recursive Bayes' filter. Gorlo et al. [16] predicts extended long-term (up to 60 s) human trajectories using large language models to reason about human interactions with the scene, represented as a 3D dynamic scene graph.

Another main type of approaches is clustering-based methods, which belong to the category of pattern-based approaches. These approaches cluster observed trajectories to create a set of long-term motion patterns. Bennewitz et al. [17] clusters full trajectories into motion patterns and uses hidden Markov models derived from the learned patterns for prediction. Chen et al. [18] propose a dynamic clustering method that can learn the motion pattern which change over time. In the work by Bera et al. [19], global and local motion patterns are learned using Bayesian inference in real-time. One shortcoming of clustering-based methods is their reliance on complete trajectories as input. In practice, especially when using a robot's on-board sensors in cluttered environments, long trajectories are hard to be observed from start to finish. And it is difficult to cluster shorter, incomplete trajectories in a meaningful way.

Clustering-based methods are non-sequential and directly model the distribution over full trajectories. In contrast to that, other pattern-based approaches are sequential and assume that human motion can be described with causally conditional models over time. In the works of [20, 21], models of local motion transition patterns are proposed. There are also approaches for pedestrian crowd prediction. Kiss et al. [22] use a constrained Gaussian process to give a smooth and continuous representation of the crowd dynamics into the future.

In addition to above approaches, physics-based methods predict human motion using kinematic models, typically without modeling underlying forces. One popular example is the constant velocity model (CVM), which is the simplest approach to predict human motion. Schöller et al. [23] showed that CVM can outperform even state-of-the-art neural models at a 4.8 s prediction horizon, but CVM is not reliable for long-term prediction as it ignores environment information.

Our approach to LHMP is connected to the field of maps of dynamics (MoDs). MoDs are maps that explicitly encode changes (e.g., motion). By building spatial and spatio-temporal models, MoDs can capture the patterns followed by dynamic objects (such as humans) in the environment, making them effective for human motion prediction [4, 24].

There are several approaches for creating maps of human motion. Occupancy-based methods focus on mapping human dynamics on occupancy grid maps, modeling motion as shifts in occupancy [25]. Trajectory-based methods extract human trajectories and group them into clusters, with each cluster representing a typical path through the environment [17]. These approaches suffer from noisy or incomplete trajectories. To address this, Chen et al. [26] formulate trajectory modeling as a dictionary learning problem and use augmented semi-nonnegative sparse coding to find local motion patterns characterized by partial trajectory segments.

MoDs can also be based on velocity observations. With velocity mapping, human dynamics can be modeled through flow models. Kucner et al. [5] presented a probabilistic framework for mapping velocity observations, which is named Circular-Linear Flow Field map (CLiFF-map). CLiFF-map can address multimodality in the data and this characteristic can be utilized for LHMP. Also, in contrast to MoDs that map trajectories, CLiFF-map allows building maps of motion patterns from incomplete or spatially sparse data [27]. While originally constructed offline, recent work has proposed online CLiFF-maps that update models with new observations [28].

When building flow models, temporal information can also be incorporated. Molina et al. [6] apply the Frequency Map Enhancement (FreMEn [29]), which is a model describing spatio-temporal dynamics in the frequency domain, to build a time-dependent probabilistic map to model periodic changes in people flow, called STeF-map. The motion orientations in STeF-map are discretized. Another method of incorporating temporal information is proposed by Zhi et al. [30]. Their approach uses a long-short term memory network to provide a multimodal probability distribution of movement directions of a typical object in the environment over time.

III. Method

A. Problem Formulation

Predicting a person's future trajectory is framed as using the past trajectory to estimate a sequence of future states. The observation length is $O_s \in \mathbb{R}^+$ seconds, equivalent to an integer $O_p > 0$ time steps. With the current time-step denoted as the integer $t_0 \geq 0$, the sequence of observed states is $\mathcal{H} = \langle s_{t_0-1}, ..., s_{t_0-O_p} \rangle$, where s_t is the state of a person at time-step t. A state is represented by 2D Cartesian coordinates (x, y), speed ρ and orientation θ: $s = (x, y, \rho, \theta)$.

To predict the future trajectory, we estimate the pedestrian's velocity $(\rho_{\text{obs}}, \theta_{\text{obs}})$ at current state from the observed sequence $\mathcal{H}$. Following the ATLAS benchmark [31], we compute a weighted average of recent velocities using a zero-mean Gaussian kernel with $\sigma = 1.5$, which assign higher weights to more recent observations, such that

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Algorithm 1: MoD-LHMP

Input: $\mathcal{H}$, x_{t_0}, y_{t_0}, Ξ
Output: $\mathcal{T}$, p

1 $\mathcal{T} = \{\}$
2 $\rho_{\text{obs}}, \theta_{\text{obs}} \leftarrow$ getObservedVelocity($\mathcal{H}$)
3 $s_{t_0} = (x_{t_0}, y_{t_0}, \rho_{\text{obs}}, \theta_{\text{obs}})$
4 $p = 1$
5 **for** $t = t_0 + 1, ..., t_0 + T_p$ **do**
6 $\quad x_t, y_t \leftarrow$ getNewPosition(s_{t-1})
7 $\quad \rho_s, \theta_s, p_t \leftarrow$ sampleVelocityFromMoD(x_t, y_t, Ξ)
8 $\quad \rho_t, \theta_t \leftarrow$ predictVelocity($\rho_s, \theta_s, \rho_{t-1}, \theta_{t-1}$)
9 $\quad s_t \leftarrow (x_t, y_t, \rho_t, \theta_t)$
10 $\quad p \leftarrow p * p_t$
11 $\quad \mathcal{T} \leftarrow \mathcal{T} \cup s_t$
12 **return** $\mathcal{T}$, p

Algorithm 2: sampleVelocityFromCLiFF(x, y, Ξ)

Input: x, y, Ξ
Output: ρ_s, θ_s

1 $\Xi_{\text{near}} \leftarrow$ getNearSWGMMs(x, y, Ξ)
2 $\xi \leftarrow$ selectSWGMM(Ξ_{near})
3 $\rho_s, \theta_s, p \leftarrow$ sampleVelociyFromSWGMM(ξ)
4 **return** ρ_s, θ_s, p

$\rho_{\text{obs}} = \sum_{t=1}^{O_p} v_{t_0-t} g(t)$ and $\theta_{\text{obs}} = \sum_{t=1}^{O_p} \theta_{t_0-t} g(t)$, where $g(t) = (\sigma\sqrt{2\pi} e^{\frac{1}{2}(\frac{t}{\sigma})^2})^{-1}$. The current state at t_0 is then represented as $s_{t_0} = (x_{t_0}, y_{t_0}, \rho_{\text{obs}}, \theta_{\text{obs}})$.

From the current state s_{t_0}, the goal is to estimate a sequence of future states up to $T_s \in \mathbb{R}^+$ s. T_s is equivalent to T_p prediction time steps assuming the constant time interval Δt. Thus, the prediction horizon is $T_s = T_p \Delta t$. The future sequence is then denoted as $\mathcal{T} = \langle s_{t_0+1}, s_{t_0+2}, ..., s_{t_0+T_p} \rangle$.

B. MoD-LHMP

CLiFF-LHMP [4] was the first method to exploit an MoD for long-term human motion prediction. Compared with CLiFF-LHMP which uses a specific type of MoD, CLiFF-map, in this work we extend it to *MoD-LHMP* that can be used with all types of MoDs that represent velocities. MoD-LHMP predicts stochastic trajectories by sampling a velocity from MoDs to guide a velocity filtering model.

The algorithm of MoD-LHMP is presented in Alg. 1. The basic algorithm is shown, with an extended version highlighted in red that ranks the predicted trajectories includes additional updates, which will be introduced later in Sec. III-D. In Alg. 1, with the input of an MoD Ξ, past states $\mathcal{H}$ and current location (x_{t_0}, y_{t_0}) of a person, the algorithm predicts a sequence of future states. To estimate $\mathcal{T}$, for each prediction time step, a velocity (ρ_s, θ_s) is sampled from the MoD at the current position (x_t, y_t) to bias the prediction with the learned motion patterns represented by the MoD. The main steps for each prediction iteration are shown in lines 5–9 of Alg. 1.

In each iteration, the new position of prediction time step t (line 6 of Alg. 1) is updated from the previous state:

$$\begin{aligned} x_t &= x_{t-1} + \rho_{t-1} \cos\theta_{t-1} \Delta t, \\ y_t &= y_{t-1} + \rho_{t-1} \sin\theta_{t-1} \Delta t, \end{aligned} \tag{1}$$

Then we estimate the new speed and orientation using a biased version of the CVM. To estimate velocity at t, we

Algorithm 3: sampleVelocityFromSTeF(x, y, Ξ, ρ_{t-1})

Input: x, y, Ξ, ρ_{t-1}
Output: ρ_s, θ_s

1 $\Xi_{\text{near}} \leftarrow$ getNearCell(x, y, Ξ)
2 $\theta_s, p \leftarrow$ sampleDirectionFromCell(ξ)
3 $\rho_s \leftarrow \rho_{t-1}$
4 **return** ρ_s, θ_s, p

sample ρ_s, θ_s from MoD at location (x_t, y_t) in the function `sampleVelocityFromMoD()` (line 7 of Alg. 1). This function is the only part of Alg. 1 that depends on the choice of MoD. In principle, any MoD that allows sampling velocities could be used. The instantiations for CLiFF-map and STeF-maps are shown in Alg. 2 and Alg. 3, respectively.

In line 8, we predict velocity (ρ_t, θ_t) by biasing the last step velocity with the sampled one (ρ_s, θ_s) as:

$$\begin{aligned} \rho_t &= \rho_{t-1} + (\rho_s - \rho_{t-1}) \cdot K(\rho_s - \rho_{t-1}), \\ \theta_t &= \theta_{t-1} + (\theta_s - \theta_{t-1}) \cdot K(\theta_s - \theta_{t-1}), \end{aligned} \tag{2}$$

where $K(\cdot)$ is a kernel function that defines the degree of impact of the MoD. We use a Gaussian kernel with a parameter β that represents the width $K(x) = e^{-\beta \|x\|^2}$.

With kernel K, the MoD term is scaled by the difference between the velocity sampled from the MoD and the current velocity according to the CVM. The sampled velocity is given less weight if it deviates more from the current velocity. This mechanism accounts for outlier trajectories that do not align with the motion patterns encoded in the MoD. A larger β value makes the method behave more like a CVM, and a smaller β makes it more closely follow the MoD.

At the end of each iteration, s_t is added to the predicted trajectory $\mathcal{T}$ and updated t for the next iteration (line 11 of Alg. 1). After iterating for T_p times, there is a sequence $\mathcal{T}$ of future states that represents the predicted trajectory.

C. Examples of MoD-LHMP

In MoD-LHMP, different MoDs can be used to represent velocity. In addition to the original CLiFF-map (Sec. III-C1), we introduce a Time-Conditioned CLiFF-map (Sec. III-C2), which extends the spatially dependent CLiFF-map by incorporating the temporal dimension. This allows human motion patterns that vary across different times of day to be captured. In Sec. III-C3, MoD-LHMP is instantiated with STeF-map, which builds a spatio-temporal model of human motion using the frequency spectrum of human activities.

1) Circular-Linear Flow Field Map (CLiFF-map): CLiFF-map represents motion patterns using multimodal statistics to represent speed and orientation jointly [5]. It associates to an arbitrary set of discrete locations a set of Semi-Wrapped Gaussian Mixture Model (SWGMM) [32] to capture the dependency between the speed and orientation, representing motion patterns based on local observations. The sampling velocity function for CLiFF-map is illustrated in Alg. 2. To sample a direction at location (x, y), from Ξ, we firstly get the SWGMMs Ξ_{near} whose distances to (x, y) are less than the sampling radius r_s. Each SWGMM location in a CLiFF-map is associated with a motion intensity ratio, defined as

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

the ratio of the time during which motion was observed at that location to the total observation time. This ratio provides an estimate of how frequently motion occurs. The SWGMM ξ with highest motion intensity ratio is selected from Ξ_{near}. The speed and orientation are sampled from the selected SWGMM and returned for motion prediction.

2) Time-Conditioned CLiFF-map: In previous work, CLiFF-maps have been trained using cumulative data, integrating data regardless of when they occurred. This approach does not account for variations in motion patterns over time. To address this, we introduce Time-Conditioned CLiFF-maps, which represent motion patterns at different times of day and more accurately capture temporal variations in human flow.

To build Time-Conditioned CLiFF-maps, a day is divided into n time intervals. For each interval, a separate CLiFF-map is trained using the trajectories that occur during that time. Consequently, for a single day, n Time-Conditioned CLiFF-maps are generated, one for each time interval. Fig. 1 shows the CLiFF-map of 10:00, 14:00 and 18:00 on the first day of the ATC dataset [11]. Implementation details are provided later in Sec. IV. These maps visually demonstrate the variations in human motion patterns throughout the day. For a more focused view, as an example, Fig. 3 shows the CLiFF-map at a specific location in the east corridor of the ATC environment, showing how motion patterns change hourly. The Time-Conditioned CLiFF-maps provide a more accurate representation of human motion compared to the general CLiFF-map.

To predict human movement, the current time, t_0, determines the time interval. The corresponding Time-Conditioned CLiFF-map is then used for prediction. The velocity sampling function remains the same as in the standard CLiFF-map and is illustrated in Alg. 2.

3) STeF-map: STeF-map [6] is a spatio-temporal flow map, which models the likelihood of motion directions on a grid-based map by a set of harmonic functions. STeF-map captures long-term changes in crowd movements over time. Each grid cell contains k_{stef} temporal models, corresponding to k_{stef} discrete orientations of motion through the cell over time.

The temporal models are based on the FreMEn framework [29], which uses the Fourier transform to model the probability of a given state as a function of time, represented by harmonic components. In the STeF-map model, the number of people detections in each orientation bin of each cell is counted in a predefined interval t_{stef}. These counts are normalized, and the normalized values update the temporal model spectra. The frequency spectrum is analyzed, and the most m_{stef} prominent spectral components are transferred to the time domain.

To predict motion pattern in a cell at a future time t, based on m_{stef} spectral components, STeF-map model computes a probability $p_{\theta_{\text{stef},i}}(t)$ for each discretized orientation $\theta_{\text{stef},i}$, where $\theta_{\text{stef},i} = i\frac{2\pi}{k_{\text{stef}}}$ and $i \in 0, 1, ..., k_{\text{stef}} - 1$, as described in [6]. The final orientation assigned to the cell in STeF-map is the one with the highest predicted probability. Fig. 2 shows the STeF-map at 10:00, 14:00 and 18:00 on the first day of the ATC dataset. In the Edinburgh dataset [12], the STeF-map of 09:00 (first row) and 14:00 (second row) are shown in Fig. 4, compared with Time-Conditioned CLiFF-maps. Implementation details are provided later in Sec. IV.

Parameter	ATC	Edinburgh
observation horizon O_s	3 s	3 s
max prediction horizon T_s	60 s	20 s
prediction time step Δt	1 s	1 s
CLiFF-map resolution	1 m	1 m
STeF-map resolution	1 m	1 m
sampling radius r_s	1 m	1 m
kernel parameter β	1	1

TABLE I: Parameters used for ATC and Edinburgh datasets

Since STeF-map encodes direction but not speed, STeF-LHMP assumes the speed remains constant, equal to that of the previous time step (Alg. 3, line 3). The direction is sampled from the STeF-map cell at the query location (x_t, y_t). While the standard STeF-map uses the direction with the highest predicted probability, we instead sample from k_{stef} discretized orientations, using their predicted probabilities as weights.

D. Ranking predicted trajectories

When evaluating MoD-LHMP, for an observed sequence, Alg. 1 can be run multiple times to generate a number of predicted trajectories. Based on practical applications for autonomous robots, we rank the predicted trajectories and evaluate with the most likely output. The ranking is based on a fitness value associated with each predicted trajectory, derived from the sample velocities obtained from the MoD.

To calculate the ranking of predicted trajectories, Alg. 1 shows the updated part in red. At each prediction time step, when sampling a velocity from MoD, a corresponding fitness value is returned (line 7 of Alg. 1). For CLiFF-map and Time-Conditioned CLiFF-map, this fitness value corresponds to the likelihood of the sampled velocity under the Semi-Wrapped Gaussian Mixture Model (SWGMM). For STeF-map, the fitness value is given by the weight assigned to the selected discretized orientation. The overall fitness of a predicted sequence $\mathcal{T}$ is computed as the sum of per step fitness values in log space across T_p prediction time steps. An example of ranking is shown in Fig. 5.

IV. Experiments

A. Dataset

To evaluate MoD-LHMP with MoDs that capture changes of human motion patterns over time, it is essential to use datasets that span multiple days and reflect variations in human motion patterns throughout the day. Our experiments were conducted using two real-world datasets, ATC and Edinburgh, both of which provide sufficient multi-day coverage for evaluation. Both datasets represent indoor open area.

1) ATC: The ATC shopping center dataset [11] contains real-world trajectories recorded in a shopping mall in Japan. This dataset covers a large indoor environment, with a total area covered of approximately $900\,\text{m}^2$. Given the extensive duration of the ATC dataset (92 days), we use a subset of 10 days in the experiments. The first day (Oct 24th) is used for training, and the remaining days are used for evaluation.

2) Edinburgh: The Edinburgh dataset [12] consists of pedestrian trajectories collected in the Informatics Forum building at the University of Edinburgh, covering around $180\,\text{m}^2$ over several months of observation. Recordings were

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

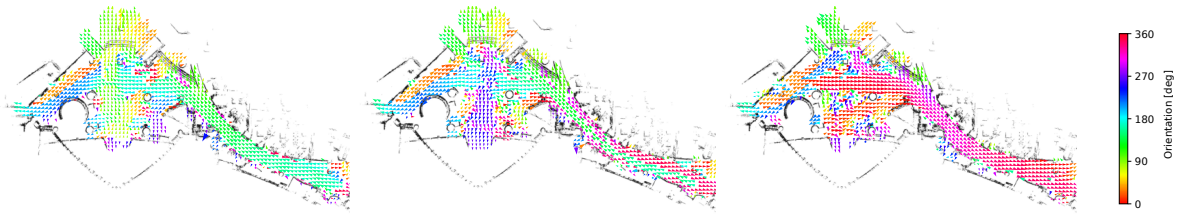

Fig. 1: Time-conditioned CLiFF-map in the ATC dataset, for 10:00 (**left**), 14:00 (**middle**) and 18:00 (**right**), showing changes of motion patterns throughout the day represented by CLiFF-map. At each location, the colored arrow shows the mean of the Gaussian component with maximum weight, where the arrow color encodes orientation and the arrow length encodes speed.

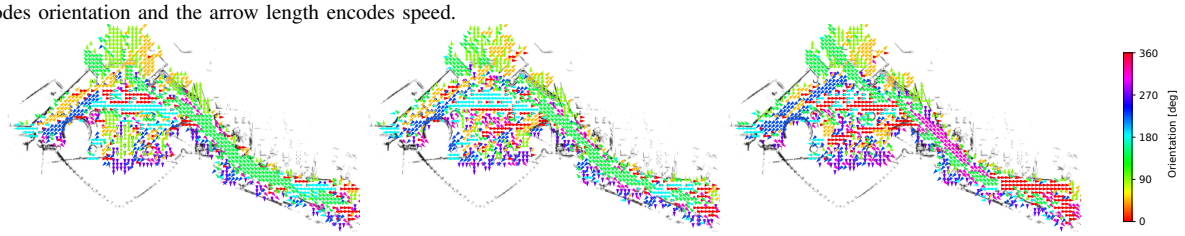

Fig. 2: STeF-map in the ATC dataset, for 10:00 (**left**), 14:00 (**middle**) and 18:00 (**right**), showing changes of motion patterns throughout the day represented by STeF-map. At each location, the colored arrow shows the dominant orientation for each cell in STeF-map, where the arrow color encodes orientation.

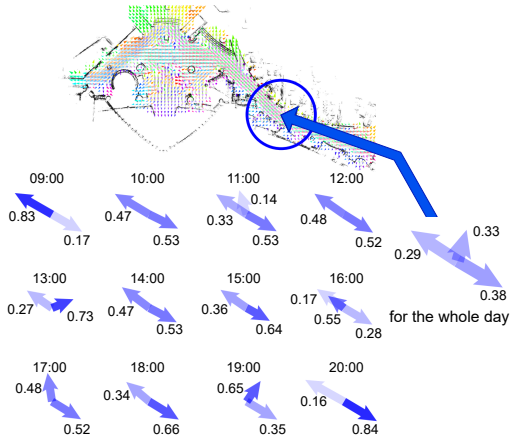

Fig. 3: A focused view of CLiFF-maps at one location in the east corridor of the ATC dataset. For each hour between 9:00 to 21:00, Time-Conditioned CLiFF-maps of the example location are shown, together with the general CLiFF-map of the whole day at the same location. Arrows show the mean of each component in SWGMM, jointly representing speed and orientation. Arrow length encodes speed, while arrow transparency reflects the component weight (lighter arrows correspond to smaller weights).

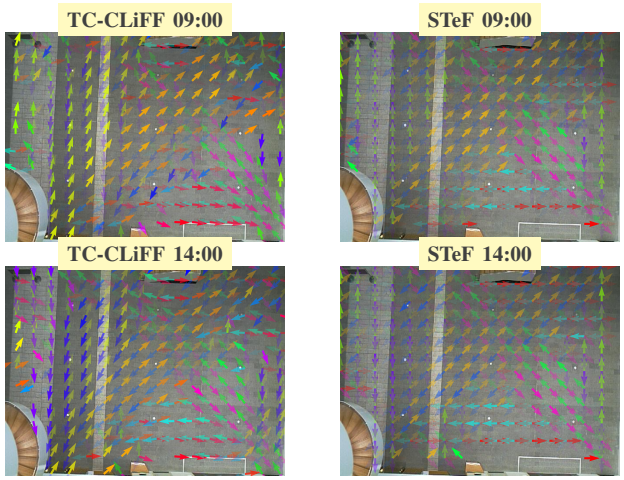

Fig. 4: Edinburgh dataset MoDs, showing motion pattern changes from 09:00 (**first row**) to 14:00 (**second row**). In Time-Conditioned CLiFF-map, colored arrows show the mean of the Gaussian component with transparency indicating the component weights. In STeF-map, colored arrows show each discretized orientation, with transparency reflecting their corresponding probabilities. In both maps, arrow color encodes orientation.

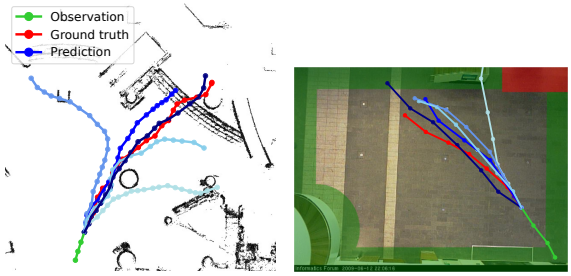

Fig. 5: Examples of predicted trajectory rankings using TC-CLiFF-LHMP in the ATC (**left**) and Edinburgh (**right**) datasets. The **red** line represents the ground truth trajectory, the **green** line represents the observed trajectory and **blue** lines the predicted trajectories, with darker shades of blue indicating higher-ranked predictions. Predictions in darker blue shows higher accuracy, showcasing the effectiveness of the ranking mechanism. **Right** figure also provides a view from the camera in the Edinburgh dataset. The green area shows the marginal region where trajectories start and end. The red area shows the region where people are waiting for a lift.

made on non-consecutive days, sometimes with hours-long gaps. To study the periodic patterns of human motion, we use all the dates when recordings were consecutive during daytime hours, from 08:00 to 16:00. In total, we use 45 days of data, with the first 5 for training and the remaining 40 for testing. For preprocessing, Majecka [12] provides instructions on removing bad trajectories: 1) remove trajectories that start or end outside the marginal area of the scene (shown in green in the right figure of Fig. 5), 2) remove trajectories shorter than 30 points because these could represent spurious detection, and 3) remove trajectories that start and end in the area next to the lifts (shown in red in the right figure of Fig. 5), as these are likely produced by people waiting for a lift.

B. Implementation Details

Both the ATC and Edinburgh datasets in our experiments are downsampled to 1 Hz. For observations, we use 3 s of each trajectory and use the remaining (up to the maximum prediction horizon) as the prediction ground truth.

For the prediction horizon T_s, instead of using a fixed horizon, we explore a wider range of values in our experiments. We evaluate methods using prediction horizons up to a maximum value, which is determined based on the length distribution of each dataset. The 90^{th} percentile values are

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

used as the maximum prediction horizons: 60 s for the ATC dataset and 20 s for the Edinburgh dataset.

To train CLiFF-maps and STeF-maps for both datasets, the map resolution is set to 1 m. For STeF-map, we follow the parameters suggested in [6]. The number of discretized orientations, k_{stef}, is 8. The interval time, t_{stef}, used for creating the STeF-map input histograms, is set to 10 minutes. The number of model components, m_{stef}, is set to 2.

For CLiFF-LHMP and Time-Conditioned CLiFF-LHMP, the sampling radius r_s, which is used when selecting SWGMMs around the given location, is set to 1 m. To bias the current orientation towards the sampled one, we use a default value of β = 1 for both datasets. Prediction stops when no dynamics data (i.e. SWGMMs in CLiFF-map or temporal models in STeF-map) is available within the radius r_s from the sampled location. For each ground truth trajectory, the prediction horizon T_s is either equal to its length or to the maximum planning horizon for longer trajectories. The parameters used for both datasets are detailed in Table I.

Privacy Considerations: The datasets used for training are publicly available and fully anonymized, representing persons only as 2D positions without identifiers or visual data. MoDs further aggregate these trajectories into statistical motion patterns, so no personal information are retained.

C. Baseline

We compare our method against Trajectron++ [7], LSTM-based prediction methods, a transformer-based model (Trajectory Unified TRansformer (TUTR) [10]), and a diffusion-based model (motion indeterminacy diffusion, MID) [9]).

Trajectron++ employs a graph-structured generative neural network based on a conditional-variational autoencoder. To implement Trajectron++ we used publicly available code and trained the model for 100 epochs on the training data of both datasets. Social LSTM [8] and vanilla LSTM are used as comparison baseline representing LSTM-based methods. TrajNet++ framework [33] is used for implementing LSTM and S-LSTM. For training LSTM-based models, we used the parameters proposed in the Social LSTM method, with the pooling size set at 32, sum pooling window size set to be 8 $\times$ 8, the dimension of hidden state for all LSTM models at 128, and the learning rate set at 0.003.

TUTR unifies social interaction modeling and multimodal trajectory prediction components in a transformer encoder-decoder architecture. It predicts multiple trajectories with corresponding probabilities, and we evaluate using the trajectory with the highest probability. MID formulate trajectory prediction as a reverse process of motion indeterminacy diffusion, which gradually discards the indeterminacy to obtain desired trajectory from ambiguous walkable areas. Both MID and TUTR are trained for 100 epochs on both datasets.

D. Evaluation Metrics

For the evaluation of predictive performance we used the following metrics: *Average* and *Final Displacement Errors* (ADE and FDE). ADE describes the mean L^2 distance between predicted trajectories and the ground truth. FDE describes the L^2 distance between the predicted and the ground truth positions at the last prediction time step. For each ground truth trajectory we generated k = 5 prediction trajectories and evaluate with the most likely output with the proposed ranking method. When probability information for the predicted trajectories is not available, we report the mean ADE and FDE values over the predicted trajectories. This applies to the experiments on the no-ranking method (results shown in Fig. 8) and to the baseline method, MID, since MID does not provide probability information for its predictions.

Method	**ADE/FDE** (m)	
	ATC ($T_s = 60$)	Edinburgh ($T_s = 20$)
TC-CLiFF	**5.332** / **11.215**	**3.035** / 5.787
CLiFF	5.557 / 11.736	3.094 / 5.830
STeF	6.026 / 12.633	3.064 / 5.847
T++	11.844 / 27.399	3.810 / 4.897
MID	21.124 / 44.373	5.621 / 7.031
TUTR	12.127 / 26.739	3.375 / **4.432**
S-LSTM	12.704 / 28.086	3.835 / 5.195
LSTM	12.975 / 28.773	3.835 / 5.201
CVM	16.191 / 35.108	7.919 / 17.112

TABLE II: Long-term prediction horizon results in the ATC and Edinburgh datasets. T_s is prediction horizon.

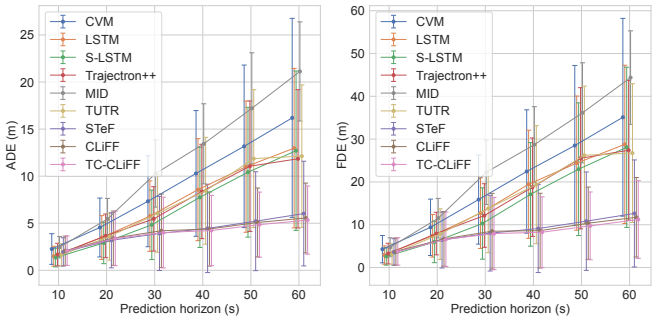

Fig. 6: ADE/FDE (mean $\pm$ one std. dev.) in the ATC dataset with planning horizon 1–60 s.

V. Results

A. Quantitative evaluation

Sec. V presents the evaluation results for the ATC and Edinburgh datasets, featuring MoD-LHMP methods, which includes Time-Conditioned CLiFF-LHMP (TC-CLiFF), CLiFF-LHMP (CLiFF), STeF-LHMP (STeF), as well as deep learning-based methods including Trajectory Unified TRansformer (TUTR), motion indeterminacy diffusion (MID) , Trajectron++ (T++), Social LSTM (S-LSTM) and vanilla LSTM. Constant velocity model (CVM) is also compared as the baseline. The table shows the performance at the maximum prediction horizon for both datasets. Results for other prediction horizon are detailed in Fig. 6 for the ATC dataset and Fig. 7 for the Edinburgh dataset.

In short-term predictions, at 10 s for ATC and 5 s for Edinburgh, deep learning-based methods perform slightly better than the MoD-based ones. However, as the prediction horizon increases, MoD-based methods achieve better accuracy. At the prediction horizon 60 s, TC-CLiFF-LHMP achieves over 50% better ADE accuracy than the deep learning-based methods, with paired t-tests showing $p < 0.001$, under the null hypothesis that our method yields equal or higher mean error than the baseline. In the Edinburgh dataset, whose environment area

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

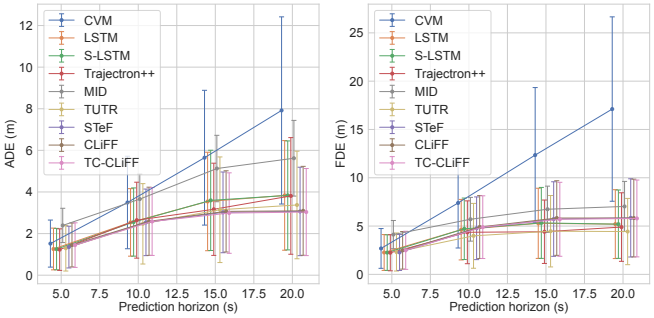

Fig. 7: ADE/FDE (mean ± one std. dev.) in the Edinburgh dataset with planning horizon 1–20 s.

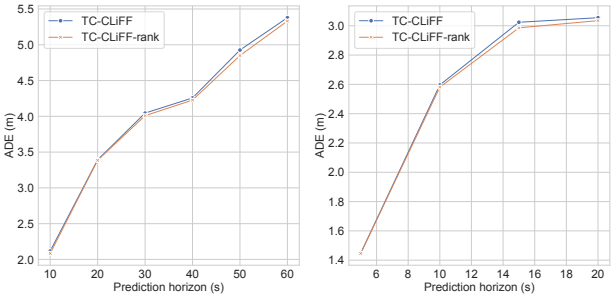

Fig. 8: ADE of Time-Conditioned CLiFF-LHMP in the ATC (**left**) and Edinburgh (**right**) dataset with and without ranking.

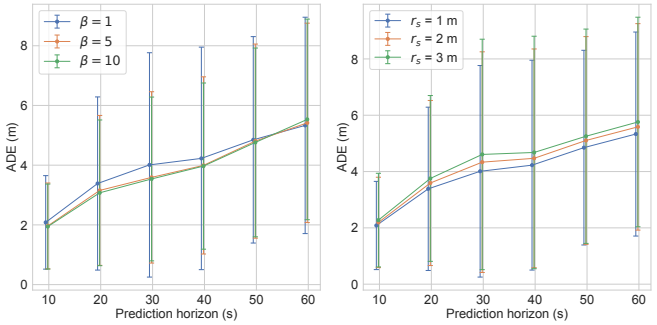

Fig. 9: Parameter analysis on the ATC dataset, for TC-CLiFF-LHMP method, showing the ADE (mean ± one std. dev.) over different prediction horizons vs the kernel parameter β **(left)** and sampling radius r_s **(right)**.

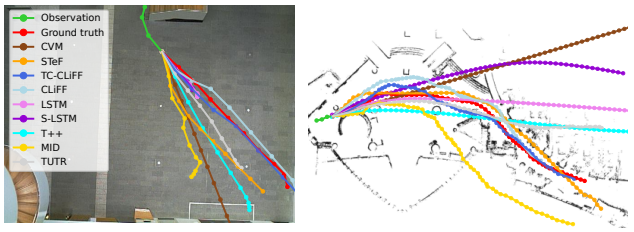

Fig. 10: Prediction examples in Edinburgh dataset (**left**) with a 10 s prediction horizon and the ATC dataset (**right**) with a 60 s horizon. MoD-LHMP methods make more accurate predictions than deep learning methods. In the ATC dataset, when the trajectory predicted by S-LSTM, LSTM, Trajectron++, TUTR and MID are unfeasible by crossing the walls, MoD-LHMP predictions are following the topology of the environment.

is about 20% of ATC, deep learning-based methods achieve lower FDE. This is due to idling behaviors in Edinburgh, where pedestrians stop or slow down at the end of a path. Sequence-based models such as LSTMs and transformers capture this pattern and generate slowed predictions, reducing FDE. MoD-based methods can also capture idling if it frequently occurs at the same location (e.g., a resting point), but when the dominant motion pattern corresponds to normal walking speed, they have a higher probability of sampling continued motion rather than idling, thereby increasing FDE in such trajectories. In contrast, the ATC dataset spans a much larger area with longer, more continuous trajectories, where local motion patterns captured by CLiFF-maps provide a stronger advantage for long-horizon prediction, resulting in a larger performance margin over the baselines.

Within the MoD-LHMP framework, TC-CLiFF-LHMP consistently outperforms the standard CLiFF-LHMP, with paired t-tests showing $p < 0.001$ for both ADE and FDE, under the null hypothesis that our method yields equal or higher mean error than the baseline. Given that motion patterns change over time in both datasets, TC-CLiFF-LHMP captures shifts in them and achieves better accuracy. This improvement remains at the higher prediction horizons. In the long-term, particularly over 30 s in the ATC dataset, CLiFF-based methods outperform STeF-LHMP, with paired t-tests showing $p < 0.001$ for both ADE and FDE, under the same null hypothesis. Unlike STeF-map which uses a discrete 8-bin histogram of orientation distribution and ignores speed, CLiFF-map models a continuous joint distribution of speed and orientation, achieving more accurate long-term predictions.

The evaluation of the ranking method in CLiFF-LHMP, which outputs the most likely predicted trajectory, is shown in Fig. 8. Its benefits increase continuously with the prediction horizon, becoming more pronounced as it extends.

Runtime analysis: At each prediction step, MoD-LHMP queries the MoD, with M stored locations, to find location neighbors and samples a velocity, followed by a constant-time kernel update. The total complexity for N trajectories over a horizon of T_p time steps is $\mathcal{O}(NT_pM)$, which scales linearly with both horizon and dataset size. For inference performance measurements, we used a laptop running Ubuntu 20.04 with an Intel Core i7-1185G7 CPU. When evaluating TC-CLiFF-LHMP, for ATC dataset, the average inference time is 0.11 s per trajectory for prediction horizon up to 60 s. For Edinburgh dataset, the average inference time is 0.04 s per trajectory for prediction horizon up to 20 s. CPU-only inference supports ∼10 Hz trajectory prediction rates, which are sufficient for embedded deployment. The inference process maintains a steady resident memory footprint of approximately 273 MB on the laptop CPU. The storage size of each Time-Conditioned CLiFF-map is about 152 kB per hour on average.

B. Parameter Analysis

Figure 9 presents a sensitivity analysis for the parameters of CLiFF-based methods: kernel parameter β and sample radius r_s. The parameter β scales CLiFF-map term based on the difference between the sampled and current directions. A higher β value makes CLiFF-LHMP to behave more like a CVM, whereas a lower β value results in predictions that more closely follow the sampled velocity. As the prediction horizon increases, a lower β value achieves better performance, showing that relying more on the CLiFF-map enhances accuracy. The sample radius r_s determines the selected nearby SWGMMs within the CLiFF-map. With a sample radius r_s

close to the map resolution ($1\,\mathrm{m}$), more accurate motion patterns can be captured, leading to better prediction accuracy.

C. Qualitative evaluation

Fig. 10 presents qualitative prediction examples. In both datasets, MoD-LHMP outperform deep learning methods. In the ATC environment, as no explicit knowledge of the obstacle layout is provided, deep learning methods predict infeasible trajectories, such as crossing walls. In contrast, MoD-based methods leverage learned motion patterns in MoDs to predict realistic trajectories that follow the complex topology of the environment, implicitly taking obstacles into account.

VI. Conclusion

In this work, we introduce MoD-informed LHMP framework, which is compatible with various types of MoDs. To handle dynamics human flow variations, we present a Time-Conditioned CLiFF-LHMP, which adapts to changing motion patterns throughout the day. Experiments on two real-world datasets show that MoD-informed LHMP approaches outperform state-of-the-art deep learning methods, and that incorporating temporal information further improves prediction accuracy. As the current method discretizes both spatial and temporal dimension into uniform grids, future work will include continuous mapping of human dynamics in spatial and temporal domains, enhancing representation accuracy. To further improve prediction performance, learning the kernel parameter β online from residual errors could help the model adapt to scene-specific dynamics.

References

[1] R. Triebel et al. “Spencer: A socially aware service robot for passenger guidance and help in busy airports”. In: *Field and service robotics*. Springer. 2016, pp. 607–622.

[2] S. Molina et al. “The ILIAD safety stack: human-aware infrastructure-free navigation of industrial mobile robots”. In: *IEEE Robotics & Automation Magazine* (2023).

[3] D. Helbing and P. Molnar. “Social force model for pedestrian dynamics”. In: *Physical review E* 51.5 (1995), p. 4282.

[4] Y. Zhu, A. Rudenko, T. Kucner, L. Palmieri, K. Arras, A. Lilienthal, and M. Magnusson. “CLiFF-LHMP: Using Spatial Dynamics Patterns for Long- Term Human Motion Prediction”. In: *Proc. of the IEEE Int. Conf. on Intell. Robots and Syst. (IROS)*. 2023.

[5] T. Kucner, M. Magnusson, E. Schaffernicht, V. Bennetts, and A. Lilienthal. “Enabling Flow Awareness for Mobile Robots in Partially Observable Environments”. In: *IEEE Robotics and Automation Letters* (2017).

[6] S. Molina, G. Cielniak, and T. Duckett. “Robotic Exploration for Learning Human Motion Patterns”. In: *IEEE Trans. on Robotics and Automation (TRO)* (2022).

[7] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone. “Trajectron++: Dynamically-Feasible Trajectory Forecasting With Heterogeneous Data”. In: *Proc. of the Europ. Conf. on Comp. Vision (ECCV)*. 2020, pp. 683–700.

[8] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. “Social LSTM: Human trajectory prediction in crowded spaces”. In: *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*. 2016, pp. 961–971.

[9] T. Gu, G. Chen, J. Li, C. Lin, Y. Rao, J. Zhou, and J. Lu. “Stochastic Trajectory Prediction via Motion Indeterminacy Diffusion”. In: *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*. 2022.

[10] L. Shi, L. Wang, S. Zhou, and G. Hua. “Trajectory Unified Transformer for Pedestrian Trajectory Prediction”. In: *Proc. of the IEEE Int. Conf. on Computer Vision (ICCV)*. Oct. 2023, pp. 9675–9684.

[11] D. Brščić, T. Kanda, T. Ikeda, and T. Miyashita. “Person tracking in large public spaces using 3-D range sensors”. In: *IEEE Trans. on Human-Machine Systems* 43.6 (2013), pp. 522–534.

[12] B. Majecka. “Statistical models of pedestrian behaviour in the forum”. In: *M.Sc. thesis, School of Informatics, University of Edinburgh* (2009).

[13] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras. “Human motion trajectory prediction: A survey”. In: *Int. J. of Robotics Research* 39.8 (2020), pp. 895–935.

[14] T. Ikeda, Y. Chigodo, D. Rea, F. Zanlungo, M. Shiomi, and T. Kanda. “Modeling and prediction of pedestrian behavior based on the sub-goal concept”. In: *Proc. of the Robotics: Science and Systems (RSS)* (2012).

[15] L. Bruckschen, N. Dengler, and M. Bennewitz. “Human motion prediction based on object interactions”. In: *Proc. of the European Conf. on Mobile Robots (ECMR)*. IEEE. 2019, pp. 1–6.

[16] N. Gorlo, L. Schmid, and L. Carlone. “Long-Term Human Trajectory Prediction Using 3D Dynamic Scene Graphs”. In: *IEEE Robotics and Automation Letters* 9.12 (2024), pp. 10978–10985.

[17] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun. “Learning motion patterns of people for compliant robot motion”. In: *Int. J. of Robotics Research* 24.1 (2005), pp. 31–48.

[18] Z. Chen, D. C. K. Ngai, and N. H. C. Yung. “Pedestrian behavior prediction based on motion patterns for vehicle-to-pedestrian collision avoidance”. In: *Proc. of the IEEE Int. Conf. on Intell. Transp. Syst. (ITSC)*. 2008, pp. 316–321.

[19] A. Bera, S. Kim, T. Randhavane, S. Pratapa, and D. Manocha. “GLMP-realtime pedestrian path prediction using global and local movement patterns”. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2016, pp. 5528–5535.

[20] S. Thompson, T. Horiuchi, and S. Kagami. “A probabilistic model of human motion and navigation intent for mobile robot path planning”. In: *Proc. of the IEEE Int. Conf. on Autonomous Robots and Agents (ICARA)*. 2009, pp. 663–668.

[21] L. Ballan, F. Castaldo, A. Alahi, F. Palmieri, and S. Savarese. “Knowledge transfer for scene-specific motion prediction”. In: *Proc. of the Europ. Conf. on Comp. Vision (ECCV)*. Springer. 2016, pp. 697–713.

[22] S. H. Kiss, K. Katuwandeniya, A. Alempijevic, and T. Vidal-Calleja. “Constrained Gaussian Processes With Integrated Kernels for Long-Horizon Prediction of Dense Pedestrian Crowd Flows”. In: *IEEE Robotics and Automation Letters* 7.3 (2022), pp. 7343–7350.

[23] C. Schöller, V. Aravantinos, F. Lay, and A. Knoll. “What the constant velocity model can teach us about pedestrian motion prediction”. In: *IEEE Robotics and Automation Letters* 5.2 (2020), pp. 1696–1703.

[24] Y. Zhu, H. Fan, A. Rudenko, M. Magnusson, E. Schaffernicht, and A. Lilienthal. “LaCE-LHMP: Airflow Modelling-Inspired Long-Term Human Motion Prediction By Enhancing Laminar Characteristics in Human Flow”. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2024.

[25] Z. Wang, P. Jensfelt, and J. Folkesson. “Building a human behavior map from local observations”. In: *Proc. of the IEEE Int. Symp. on Robot and Human Interactive Comm. (RO-MAN)*. 2016, pp. 64–70.

[26] Y. F. Chen, M. Liu, and J. P. How. “Augmented dictionary learning for motion prediction”. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2016, pp. 2527–2534.

[27] T. R. de Almeida, Y. Zhu, A. Rudenko, T. P. Kucner, J. A. Stork, M. Magnusson, and A. Lilienthal. “Trajectory Prediction for Heterogeneous Agents: A Performance Analysis on Small and Imbalanced Datasets”. In: *IEEE Robotics and Automation Letters* (2024), pp. 1–8.

[28] Y. Zhu, A. Rudenko, L. Palmieri, L. Heuer, A. Lilienthal, and M. Magnusson. “Fast Online Learning of CLiFF-maps in Changing Environments”. In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2025.

[29] T. Krajník, J. Fentanes, J. Santos, and T. Duckett. “FreMEn: Frequency Map Enhancement for Long-Term Mobile Robot Autonomy in Changing Environments”. In: *IEEE Trans. on Robotics (TRO)* (2017).

[30] W. Zhi, R. Senanayake, L. Ott, and F. Ramos. “Spatiotemporal Learning of Directional Uncertainty in Urban Environments With Kernel Recurrent Mixture Density Networks”. In: *IEEE Robotics and Automation Letters* 4.4 (2019), pp. 4306–4313.

[31] A. Rudenko, L. Palmieri, W. Huang, A. Lilienthal, and K. Arras. “The Atlas Benchmark: an Automated Evaluation Framework for Human Motion Prediction”. In: *Proc. of the IEEE Int. Symp. on Robot and Human Interactive Comm. (RO-MAN)*. 2022.

[32] A. Roy, S. K. Parui, and U. Roy. “SWGMM: a semi-wrapped Gaussian mixture model for clustering of circular-linear data”. In: *Pattern Anal. Appl.* 19.3 (2016), pp. 631–645.

[33] P. Kothari, S. Kreiss, and A. Alahi. “Human Trajectory Forecasting in Crowds: A Deep Learning Perspective”. In: *IEEE Trans. on Intell. Transp. Syst. (TITS)* (2021), pp. 1–15.