

# Are We Pursuing the Right Objective? A Survey in Human Trajectory Prediction

Manuel Hetzel, Kerim Turacan, Konrad Doll  
University of Applied Sciences Aschaffenburg  
Germany

{firstname.lastname}@th-ab.de

Bernhard Sick  
University of Kassel  
Germany

bsick@uni-kassel.de

## Abstract

*Human Trajectory Prediction (HTP) is a challenging task in computer vision and robotics, with numerous applications, including Autonomous Driving, Smart City Surveillance, Human-Machine Interaction, and Autonomous Robots. Despite significant progress in recent years, existing methods have focused on accuracy, social interaction modeling, and deterministic diversity. In contrast, little attention has been paid to uncertainty modeling, calibration, and forecasts from short observation periods, which are crucial for downstream tasks such as path planning and collision avoidance. Moreover, existing methods often rely on datasets and evaluation metrics that may not align well with the prediction goals of real-world applications, such as integrating HTP into autonomous vehicles or robot path-planning tasks. This survey provides a comprehensive review of HTP methods, datasets, and evaluation metrics, and discusses the need for more application-specific evaluation and the use of diverse datasets in relation to established rules and conditions. By providing a thorough understanding of the current State of the Art in HTP and its use cases, this survey aims to facilitate the development of more accurate and usable predictive models for real-world applications.*

## 1. Introduction

HTP is a profoundly significant and challenging problem in computer vision and robotics. Its relevance is underscored by its numerous applications, such as Autonomous Driving (AD), Smart City Surveillance (SCS), Human-Machine Interaction (HMI), and Autonomous Robots (AR). This task aims to forecast the future motion of humans, such as pedestrians, cyclists, and others, based on their past movements, considering factors such as Environmental Context (EC), Social Interactions (SI), and personal intentions. Human movements are inherently indeterministic, and future predictions can be multimodal, leading to errors. A wide range of methods has been proposed for HTP, rang-

ing from analytical approaches like the Constant Velocity (CV) or Social Forces Model (SFM) to deep learning-based techniques (Long-Short-Term-Memory (LSTM), Convolutional Neural Networks (CNN), Graph Neural Networks (GNN), Generative Adversarial Networks (GAN), Conditional Variable Auto Encoders (CVAE), Normalizing Flow (NF), Transformers (TF), and Diffusion Models (DM). Each method has unique strengths and limitations and has shown varying degrees of success. However, their performance is often evaluated using datasets and metrics that may not accurately reflect the prediction requirements of real-world applications. The evaluation of HTP models is typically performed using metrics such as Average Displacement Error (ADE) and Final Displacement Error (FDE) [100]. While these metrics provide a quantitative assessment of prediction accuracy, they do not necessarily capture the requirements of AD, SCS, HMI, and AR applications. For instance, predicting a single or  $K$  discrete future trajectory hypotheses, without additional uncertainty and reliability assessment, in such applications is of limited use for downstream tasks, such as path planning or collision avoidance, which often internally rely on probability estimates. Therefore, it is essential to model appropriate, calibrated uncertainty estimates for each discrete forecast. Numerous studies support the added value of probabilistic output modeling with uncertainty-related forecasts compared to discrete forecasts without any uncertainty-related output for path planning and situational assessment [30, 32, 63, 95, 114]. If a prediction states that a pedestrian has a 90% probability of being at a specified location, further evaluation is necessary to verify the correctness. Otherwise, downstream tasks cannot trust these predictions. Furthermore, the most commonly used datasets, such as ETH/UCY [71, 100], and SDD [105], provide a good starting point. Still, they may not capture the full range of scenarios and complexities encountered in many real-world applications, especially in traffic-related ones.

This survey reviews existing methods (Sec. 2), datasets (Sec. 3), and evaluation metrics (Sec. 4) for HTP, and discusses the need for more rigorous, application-specific eval-

uation and dataset selection. We discuss rules and conditions and elaborate on gaps to encourage rethinking and refining established principles (Sec. 5).

## 2. HTP Methods

This work presents a comprehensive HTP summary of the relationships and dependencies among methods, data sets, evaluation metrics, research, and application targets. Therefore, we review more than 100 publications on the research topic. Existing surveys build their taxonomy from a feature extraction level [6] [56] [102] or focus on the methods themselves [104] [156], without a coherent relation between the aspects mentioned above. Existing methods can differ in the representation of input data, architecture, and the scope of the predicted output. On the input side, methods can utilize the following data: past trajectories of the target and multiple surrounding agents, as well as additional context information such as semantic maps or behavioral rules. From an architectural standpoint, multiple deep learning-based techniques have been established, including LSTMs, GNNs, CNNs, GANs, CVAEs, NF, TFs, and DMs. These architectures target different aspects [156]. RNNs specialize in the temporal modeling of human trajectories, CNNs focus on spatial relationships and map data integration, GNNs concentrate on spatial relationships with implicit learning of interactions, GANs and CVAEs emphasize the representation of higher diversity in trajectory hypotheses, and TFs and DMs prioritize global attention allocation, learning the temporal and interactive features of movement patterns. Recently, several techniques have been proposed to address social interactions and environmental conditions. On the output side, methods can provide single/multiple Discrete Trajectory Predictions (DTP) as  $K$  discrete hypotheses with [58, 79, 84, 115] or without [35, 55, 64, 70, 88] probability scores/ranking or Probabilistic Trajectory Predictions (PTP) as unimodal/multimodal distributions representing residence areas from which discrete hypotheses with calibrated probabilities can be sampled [44, 46, 151].

**LSTMs** are often used for sequence problems due to their ability to store long-term information. Social-LSTM [2] is a classic framework for HTP that introduces social pooling to capture interactions. Social pooling aggregates hidden states of nearby agents within a spatial grid, enabling the model to encode local interaction patterns such as collision avoidance and group movement. This formulation was one of the first to move beyond independent trajectory forecasting by explicitly modeling social context within a recurrent architecture. Many frameworks adapt the basic idea and further optimize it, like Group-LSTM [12], Scene-LSTM [86], and SS-LSTM [139].

**CNNs** are superior in terms of spatial features and relationships, Y-net [85], MATF [155], and Next [75] are commonly cited examples. Y-net introduces a dual-encoder ar-

chitecture that jointly processes rasterized scene context and observed agent trajectories, fusing semantic map information with motion history. By predicting goal distributions and conditioning future trajectories on sampled goals, it explicitly decouples high-level intention prediction from low-level motion generation, improving long-horizon forecasting in structured environments.

**GNNs** are well-suited for modeling interaction relationships and spatial information. Nodes are commonly used to represent agents, and edges represent interactions between agents. Trajectron++ [108], STGAT [48], and Social-BiGAT [66] are the first frameworks using GNNs for HTP, followed by Social-STGCNN [92], SGCN [80], RSGB [124], GroupNet [136], and GP-Graph [52]. Trajectron++ extends Trajectron by modeling dynamic spatiotemporal graphs, allowing agent interactions to evolve over time. It combines graph neural networks with a CVAE to capture multimodal future distributions while supporting variable numbers of agents and online inference, making it suitable for real-world deployment.

Social-GAN [40] was the first to present a **GAN** architecture for HTP, where features from the past encoder are concatenated with Gaussian noise vectors and sent together into the decoder, where a discriminator is learned to classify whether the future trajectory is the ground truth or the prediction. Other frameworks adapt the concept like Social-Ways [5], Sophie [107], MG-GAN [98], SEEM [132] or Goal-GAN [28].

**CVAEs** differ from GANs; they can explicitly learn the distribution of target trajectories conditioned on historical trajectories by learning the latent distribution from which it samples. Trajectron [14], Trajectron++ [108], and PECNet [64] are some of the most influential representatives. GAN or CVAE models are difficult to train due to the implicit distribution modeling. The **NF** architecture aims to mitigate this by explicitly learning the data distribution via an invertible network that maps a complex distribution to a tractable form via invertible transformations. HBAFlow [11], FloMo [111], STGlow [106], FlowChain [83], and MoFlow [35] are well-known examples. MoFlow combines multiple normalizing flows to capture multimodal trajectory distributions while maintaining exact likelihood estimation. By decomposing complex motion patterns into multiple flow components, it improves sample diversity and training stability compared to GAN- or CVAE-based methods.

**TF-** and **DM-**based architectures have been increasingly used over the past two years, demonstrating strong accuracy performance. They provide self-attention-based memory mechanisms and Denoising Diffusion. TransF [34], STAR [27], AgentFormer [147], MID [37], TDOR [65], LED [88], EqMotion [21], SingularTrajectory [55], MART [70], MNRF [33], and DD-MDN [46]

Model	Year	Arch	Metrics	Results	S	M	P	Code available	Pretrained Models	Results Reproducible
<i>CV Model</i> [110]	-	Physical	ADE,FDE	0.39/0.83	×	×	×	×	×	✓
<i>CA Model</i> [110]	-	Physical	ADE,FDE	0.84/2.11	×	×	×	×	×	✓
<i>SFM Model</i> [140]	2011	Physical	ADE,FDE	0.39/0.60	✓	×	×	×	×	✓
<i>Basic LSTM</i> [2]	2016	LSTM	ADE,FDE	0.44/0.98	×	×	×	×	×	✓
<i>Social-LSTM</i> [2]	2016	LSTM	ADE,FDE	0.27/0.61	✓	×	×	Link	×	✓
<i>Desire</i> [69]	2017	CVAE	ADE,FDE	0.46/1.00	✓	✓	×	Link	×	×
<i>CNN-18</i> [96]	2018	CNN	ADE,FDE	0.60/1.22	×	×	×	×	×	×
<i>Social-GAN</i> [40]	2018	GAN	ADE,FDE	0.78/1.18	✓	×	×	Link	✓	×
<i>Trajectron</i> [14]	2018	CVEA	ADE,FDE	0.34/0.67	✓	×	×	Link	×	✓
<i>S-ATTN</i> [130]	2018	GNN	ADE,FDE	0.30/2.59	✓	×	×	Link	×	✓
<i>Sophie</i> [107]	2019	GAN	ADE,FDE	0.54/1.15	✓	✓	×	Link	×	✓
<i>Social-BiGAT</i> [66]	2019	GNN	ADE,FDE	0.48/1.00	✓	✓	×	×	×	×
<i>STGAT</i> [48]	2019	GNN	ADE,FDE	0.62/0.83	✓	×	×	Link	×	✓
<i>MATF</i> [155]	2019	CNN	ADE,FDE	0.48/0.90	✓	✓	×	Link	×	✓
<i>NEXT</i> [75]	2019	CNN	ADE,FDE	0.46/1.00	✓	✓	×	Link	✓	✓
<i>SR-LSTM</i> [153]	2019	LSTM	ADE,FDE	0.45/0.94	✓	×	×	Link	✓	✓
<i>CVM</i> [110]	2020	Physical	ADE,FDE	0.28/0.56	×	×	×	Link	×	✓
<i>BiTraP</i> [34]	2020	CVAE	ADE,FDE	0.18/0.35	✓	×	×	Link	✓	×
<i>TransF</i> [34]	2020	TF	ADE,FDE	0.54/1.17	✓	×	×	×	×	×
<i>STAR</i> [27]	2020	TF	ADE,FDE	0.26/0.53	✓	×	×	Link	×	✓
<i>PECNet</i> [64]	2020	CVAE	ADE,FDE	0.29/0.48	✓	×	×	Link	✓	✓
<i>Trajectron++</i> [108]	2020	CVAE	ADE,FDE,NLL	0.19/0.41	✓	✓	×	Link	×	×
<i>Social-STGCNN</i> [92]	2020	GNN	ADE,FDE	0.44/0.75	✓	✓	×	Link	✓	✓
<i>Reci</i> [123]	2020	GAN	ADE,FDE	0.44/0.90	✓	✓	×	×	×	×
<i>TPNet</i> [77]	2020	CNN	ADE,FDE	0.27/0.42	✓	✓	×	×	×	×
<i>GTPPO</i> [142]	2020	CVAE	ADE,FDE	0.31/0.50	✓	×	×	×	×	×
<i>NMMP</i> [47]	2020	GAN	ADE,FDE	0.41/0.82	✓	✓	×	Link	✓	✓
<i>Goal-GAN</i> [28]	2020	GAN	ADE,FDE	0.43/0.85	✓	✓	✓	Link	×	✓
<i>RSBG</i> [94]	2020	GNN	ADE,FDE	0.44/0.98	✓	×	×	×	×	×
<i>Mantra</i> [89]	2020	CVAE	ADE,FDE	0.32/0.65	✓	×	×	Link	✓	✓
<i>BNSP-SFM</i> [58]	2021	CVAE	ADE,FDE	0.16/0.24	✓	×	✓	×	×	×
<i>FloMo</i> [111]	2021	NF	ADE,FDE	0.22/0.37	✓	×	×	Link	×	×
<i>SGCN</i> [80]	2021	GNN	ADE,FDE	0.37/0.65	✓	×	×	×	×	×
<i>DMRGCN</i> [49]	2021	GNN	ADE,FDE	0.34/0.58	✓	×	×	Link	✓	✓
<i>S-CSR</i> [41]	2021	CVAE	ADE,FDE	0.10/0.16	✓	×	×	×	×	×
<i>Social-STAGE</i> [84]	2021	GNN	ADE,FDE, $M_1,M_2$	0.32/0.59	✓	×	✓	×	×	×
<i>Expert-Traj</i> [154]	2021	CVAE	ADE,FDE	0.19/0.36	✓	×	×	Link	✓	×
<i>PCCSNet</i> [125]	2021	CVAE	ADE,FDE	0.21/0.42	✓	×	×	Link	✓	×
<i>Agentformer</i> [147]	2021	TF	ADE,FDE	0.23/0.39	✓	×	×	Link	✓	×
<i>MG-GAN</i> [98]	2021	GAN	ADE,FDE	0.36/0.71	✓	✓	×	Link	✓	×
<i>YNet</i> [85]	2021	CNN	ADE,FDE	0.18/0.27	✓	✓	×	Link	✓	×
<i>DisDis</i> [19]	2021	CVAE	ADE,FDE,PCMD	0.17/0.37	✓	×	×	Link	✓	×
<i>LB-EBM</i> [97]	2021	Energy	ADE,FDE	0.21/0.38	✓	×	×	Link	✓	✓
<i>Social-NCE</i> [78]	2021	CVAE	ADE,FDE	0.40/0.98	✓	×	×	Link	×	✓
<i>S-DPF</i> [115]	2021	CVAE	ADE,FDE	0.43/0.63	✓	×	✓	×	×	×
<i>STC-Net</i> [73]	2021	GNN	ADE,FDE	0.38/0.68	✓	×	×	×	×	×
<i>Introvert</i> [113]	2021	CVAE	ADE,FDE	0.21/0.34	✓	✓	×	Link	×	×
<i>TPNMS</i> [76]	2021	GAN	ADE,FDE	0.38/0.73	✓	×	×	Link	✓	✓

Table 1. Overview of HTP-focused methods sorted by years 2011-2021. **Results** present the best of  $K = 20 \min ADE/FDE$  results regarding the **ETH/UCY dataset** in meters; lower is better. **Dataset acronyms**: E: ETH, S: SDD, N: NuScenes, G: GCS, D: inD, I: IMPTC, W: Waymo, A: ApolloScape, F: FPD, T: TrajNet, B: NBA, L: NFL. **S** is Social Interaction integration, **M** is Map/Context Integration, and **P** is With Probability Scores and/or uncertainty calibration. **Additional symbols**: ×: no, ✓: yes, - not available.

are currently some of the most popular SOTA frameworks. MID formulates trajectory prediction as a denoising diffusion process that progressively transforms noise into realistic future trajectories conditioned on observed motion. SingularTrajectory models future trajectories on a low-rank trajectory manifold, decomposing motion into a small set of dominant modes to reduce redundancy while preserving multimodality. MNRF learns a continuous probabilistic

trajectory field over space and time, enabling fine-grained trajectory querying and improved spatial consistency compared to discrete sampling-based methods.

Tab. 1, Tab. 2, Tab. 3, Tab. 4, and Tab. 5 present detailed summaries of more than 100 peer-reviewed methods and frameworks targeting HTP. Methods in Tab. 1 and Tab. 2 use the common and de facto standard ETH/UCY dataset as the primary or only evaluation dataset. Methods

Model	Year	Arch	Metrics	Results	S	M	P	Code available	Pretrained Models	Results Reproducible
<i>TPNSTA</i> [74]	2022	GAN	ADE,FDE	0.37/0.71	✓	×	×	×	×	×
<i>Social-SSL</i> [128]	2022	TF	ADE,FDE	0.44/0.85	✓	×	×	<a href="#">Link</a>	×	✓
<i>SHENet</i> [90]	2022	CVAE	ADE,FDE	0.23/0.36	✓	✓	×	<a href="#">Link</a>	✓	✓
<i>CSCNet</i> [134]	2022	CVAE	ADE,FDE	0.37/0.79	✓	✓	×	×	×	×
<i>STT</i> [94]	2022	CVAE	ADE,FDE	0.43/0.88	✓	×	×	×	×	×
<i>SGNet-ED</i> [131]	2022	CVAE	ADE,FDE	0.18/0.35	✓	×	×	<a href="#">Link</a>	✓	×
<i>Social-Ways</i> [5]	2022	GAN	ADE,FDE	0.46/0.83	✓	×	×	<a href="#">Link</a>	×	✓
<i>CSR</i> [42]	2022	CVAE	ADE,FDE	0.14/0.23	✓	×	×	×	×	×
<i>CNN-22</i> [117]	2022	CNN	ADE,FDE	0.44/0.91	✓	✓	×	×	×	×
<i>CAGN</i> [61]	2022	CNN	ADE,FDE	0.25/0.43	✓	×	×	<a href="#">Link</a>	×	×
<i>SIT</i> [81]	2022	CTT	ADE,FDE	0.23/0.38	✓	×	×	<a href="#">Link</a>	✓	✓
<i>Social-VAE</i> [99]	2022	CVAE	ADE,FDE,NLL	0.21/0.33	✓	×	×	<a href="#">Link</a>	✓	×
<i>Social-Implicit</i> [93]	2022	GAN	ADE,FDE,AMD,AMV	0.33/0.67	✓	×	×	<a href="#">Link</a>	✓	✓
<i>V<sup>2</sup>-Net</i> [25]	2022	MLP	ADE,FDE	0.18/0.28	✓	✓	×	<a href="#">Link</a>	✓	×
<i>NSP-SFM</i> [57]	2022	CVAE	ADE,FDE	0.17/0.24	✓	✓	×	<a href="#">Link</a>	✓	×
<i>MID</i> [37]	2022	DM	ADE,FDE	0.38/0.54	✓	×	×	<a href="#">Link</a>	×	×
<i>MemoNet</i> [137]	2022	CVAE	ADE,FDE	0.21/0.35	✓	×	×	<a href="#">Link</a>	✓	✓
<i>ScePT</i> [148]	2022	CVAE	ADE,FDE	0.12/0.73	✓	✓	×	<a href="#">Link</a>	×	✓
<i>SEEM</i> [132]	2023	GAN	ADE,FDE	0.48/0.95	✓	×	×	×	×	×
<i>GATraj</i> [20]	2023	GNN	ADE,FDE	0.17/0.29	✓	×	×	<a href="#">Link</a>	✓	✓
<i>Graph-TERN</i> [50]	2023	GNN	ADE,FDE	0.24/0.38	✓	×	×	<a href="#">Link</a>	✓	✓
<i>MSLR</i> [149]	2023	GNN	ADE,FDE	0.19/0.33	✓	×	×	<a href="#">Link</a>	✓	✓
<i>E-V<sup>2</sup>-Net</i> [23]	2023	MLP	ADE,FDE	0.17/0.28	✓	×	×	<a href="#">Link</a>	×	×
<i>FlowChain</i> [83]	2023	NF	ADE,FDE,EMD	0.29/0.52	✓	×	×	<a href="#">Link</a>	✓	✓
<i>TUTR</i> [79]	2023	TF	(B)ADE,(B)FDE	0.21/0.36	✓	×	✓	<a href="#">Link</a>	×	✓
<i>LED</i> [88]	2023	DM	ADE,FDE	0.21/0.33	✓	×	×	<a href="#">Link</a>	✓	✓
<i>EqMotion</i> [21]	2023	TF	ADE,FDE	0.21/0.35	✓	×	×	<a href="#">Link</a>	✓	✓
<i>STGlow</i> [106]	2023	NF	ADE,FDE	0.15/0.28	✓	×	×	×	×	×
<i>DySeT</i> [101]	2024	TF	ADE,FDE	0.20/0.33	✓	✓	×	×	×	×
<i>Social-CVAE</i> [133]	2024	CVAE	ADE,FDE	0.03/0.02	✓	✓	×	<a href="#">Link</a>	✓	×
<i>LMTraj-SUP</i> [54]	2024	TF	ADE,FDE	0.22/0.32	✓	×	×	<a href="#">Link</a>	✓	✓
<i>SingularTrajectory</i> [55]	2024	TF	ADE,FDE	0.21/0.32	✓	✓	×	<a href="#">Link</a>	✓	✓
<i>MDN</i> [44]	2024	LSTM	ADE,FDE,RLS,SS	0.26/0.50	×	×	✓	<a href="#">Link</a>	✓	✓
<i>PPT</i> [103]	2024	TF	ADE,FDE	0.20/0.31	✓	×	×	<a href="#">Link</a>	✓	✓
<i>MART</i> [70]	2024	TF	ADE,FDE	0.21/0.33	✓	×	×	<a href="#">Link</a>	✓	✓
<i>CLLS</i> [103]	2025	TF	ADE,FDE	0.20/0.32	✓	×	×	×	×	×
<i>MoFlow</i> [35]	2025	NF	ADE,FDE	0.21/0.33	✓	×	×	<a href="#">Link</a>	✓	✓
<i>MNRF</i> [33]	2025	TF	ADE,FDE	0.19/0.32	✓	×	×	<a href="#">Link</a>	✓	✓
<i>DD-MDN</i> [46]	2026	DM	ADE,FDE,RLS,SS	0.19/0.31	✓	✓	✓	<a href="#">Link</a>	✓	✓

Table 2. Overview of HTP-focused methods, focused on **ETH/UCY** dataset, sorted by years 2022-2025/26. Notation and description are equal to Table 1.

in Tab. 3 utilize the SDD [105] and in Tab. 4 utilize the inD [13] datasets for evaluation.

The tables compare distinctive features, including primary architecture, social interaction awareness (S), context map data usage (M), used datasets, evaluation metrics, probabilistic outputs and/or uncertainty handling (P), and ADE/FDE performance on the most commonly used evaluation datasets, as well as transparency to the research community regarding code/pre-trained model availability and reproducibility. We have aggregated the following conclusions from these reviewed frameworks: **First:** All frameworks use ADE/FDE as standard evaluation metrics, and only 9% apply advanced metrics. **Second:** The ETH/UCY dataset is used in 97% followed by SDD with 31% and inD with 7%. Additional datasets are below 5%. 41% exclusively used the ETH/UCY dataset for evaluation. **Third:** Social interaction is covered by 93%, additional context and map usage is taken into account by 28%, and probability-

related forecasts representing uncertainty handling by 5%. **Fourth:** 71% share their framework/code with the research community, and in most cases, others or ourselves can reproduce published results. The provision of pre-trained models clearly deviates from this with a provisioning rate of only 44%. **Fifth:** Regarding the ETH/UCY dataset, the performance has reached a saturation level, with an ADE/FDE border of 0.15-0.20m ADE and 0.30-0.35m FDE. A few results, such as Social-CVAE [133], CRS [42], or S-CSR [41], have surpassed this limit by employing a recursive prediction scheme called ultra-sampling, which is criticized for selecting the best of  $K$  predictions at every forecast horizon, without considering time correlation or providing probabilities or ranking information that may justify this evaluation procedure.

In addition, an HTP subtask has emerged in recent years, focusing on optimizing previously proposed frameworks. Tab. 5 provides more details. These works try

SDD Dataset											
Model	Year	Arch	Metrics	Results	S	M	P	Code available	Pretrained Models	Results Reproducible	
<i>Sophie</i> [107]	2019	GAN	ADE,FDE	16.27/29.38	✓	✓	×	<a href="#">Link</a>	×	✓	
<i>PECNet</i> [64]	2020	CVAE	ADE,FDE	9.96/15.88	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>BNSP-SFM</i> [58]	2021	CVAE	ADE,FDE	6.46/10.49	✓	×	✓	×	×	×	
<i>FloMo</i> [111]	2021	NF	ADE,FDE	2.60/4.43	✓	×	×	<a href="#">Link</a>	×	×	
<i>Expert-Traj</i> [154]	2021	CVAE	ADE,FDE	7.69/14.38	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>PCCSNet</i> [125]	2021	CVAE	ADE,FDE	8.62/16.16	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>MG-GAN</i> [98]	2021	GAN	ADE,FDE	13.65/25.85	✓	✓	×	<a href="#">Link</a>	✓	×	
<i>YNet</i> [85]	2021	CNN	ADE,FDE	7.85/11.85	✓	✓	×	<a href="#">Link</a>	✓	×	
<i>LB-EBM</i> [97]	2021	Energy	ADE,FDE	8.87/15.61	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>SIT</i> [81]	2022	CTT	ADE,FDE	9.13/15.42	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>Social-VAE</i> [99]	2022	CVAE	ADE,FDE,NLL	8.10/11.72	✓	×	×	<a href="#">Link</a>	✓	×	
<i>V<sup>2</sup>-Net</i> [25]	2022	MLP	ADE,FDE	7.12/11.39	✓	✓	×	<a href="#">Link</a>	✓	×	
<i>NSP-SFM</i> [57]	2022	CVAE	ADE,FDE	8.56/11.85	✓	✓	×	<a href="#">Link</a>	✓	×	
<i>MID</i> [37]	2022	TF	ADE,FDE	9.73/15.32	✓	×	×	<a href="#">Link</a>	×	×	
<i>MUSE-VAE</i> [91]	2022	CVAE	ADE,FDE,NLL,ECFL	6.36/11.10	✓	✓	×	<a href="#">Link</a>	✓	×	
<i>TDOR</i> [65]	2022	TF	ADE,FDE	8.60/13.90	✓	✓	×	<a href="#">Link</a>	✓	✓	
<i>MemoNet</i> [137]	2022	CVAE	ADE,FDE	8.56/12.66	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>Graph-TERN</i> [50]	2023	GNN	ADE,FDE	8.43/14.26	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>MSLR</i> [149]	2023	GNN	ADE,FDE	8.22/13.39	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>E-V<sup>2</sup>-Net</i> [23]	2023	MLP	ADE,FDE	6.57/10.49	✓	×	×	<a href="#">Link</a>	×	×	
<i>FlowChain</i> [83]	2023	NF	ADE,FDE,EMD	9.93/17.17	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>TUTR</i> [79]	2023	TF	(B)ADE,(B)FDE	7.76/12.69	✓	×	✓	<a href="#">Link</a>	×	✓	
<i>LED</i> [88]	2023	TF	ADE,FDE	8.48/11.66	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>Social-CVAE</i> [133]	2024	CVAE	ADE,FDE	1.48/1.06	✓	✓	×	<a href="#">Link</a>	✓	×	
<i>LMTraj-SUP</i> [54]	2024	TF	ADE,FDE	7.85/10.15	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>PPT</i> [103]	2024	TF	ADE,FDE	7.03/10.65	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>MART</i> [70]	2024	TF	ADE,FDE	7.43/11.82	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>CLLS</i> [103]	2025	TF	ADE,FDE	7.25/11.05	✓	×	×	×	×	×	
<i>MoFlow</i> [35]	2025	TF	ADE,FDE	7.50/11.96	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>MNRF</i> [33]	2025	TF	ADE,FDE	7.20/11.29	✓	×	×	<a href="#">Link</a>	✓	✓	
<i>DD-MDN</i> [46]	2026	DM	ADE,FDE,RLS,SS	7.19/11.82	✓	✓	✓	<a href="#">Link</a>	✓	✓	

Table 3. Overview of HTP-focused methods sorted by year. Results present the best  $K = 20 \min ADE/FDE$  results for the **SDD** dataset, in pixels; lower is better. Notation is equal to Table 1 and Table 2.

inD Dataset										
<i>Trajectron++</i> [108]	2020	CVAE	ADE,FDE	0.62/0.98	✓	✓	×	<a href="#">Link</a>	✓	✓
<i>Y-Net</i> [85]	2021	CNN	ADE,FDE	0.55/0.93	✓	✓	×	<a href="#">Link</a>	✓	×
<i>AC-VRNN</i> [10]	2021	RNN	ADE,FDE	0.42/0.80	✓	×	✓	×	×	×
<i>Agentformer</i> [147]	2021	TF	ADE,FDE	0.57/0.87	✓	×	×	<a href="#">Link</a>	✓	×
<i>Goar-SAR</i> [22]	2021	CVAE	ADE,FDE	0.44/0.70	✓	✓	×	<a href="#">Link</a>	✓	✓
<i>Di-Long</i> [118]	2024	TF	ADE,FDE	0.37/0.59	✓	✓	×	×	×	×
<i>DD-MDN</i> [46]	2026	DM	ADE,FDE,RLS,SS	0.20/0.36	✓	✓	✓	<a href="#">Link</a>	✓	✓

Table 4. Overview of HTP-focused methods sorted by year. Results present the best  $K = 20 \min ADE/FDE$  results for the **inD** dataset, in meters; lower is better. Notation is equal to Table 1 and Table 2.

to improve trajectory sampling (DLow [144], LDS [145], NSPM [53], Stimulus [60], BOSampler [38]), group interaction (GroupNet [136], GP-Graphs [52]), or feature representation (MOE [59], EigenTrajectory [51], SocialCircle [24]) to create a more diverse set of discrete prediction samples and improve the performance in complex crowd scenarios. SocialCircle encodes surrounding agents using a circular spatial partition centered on the target agent, capturing relative distance and angular information in a compact representation. This design emphasizes rotational invari-

ance and local interaction structure while remaining computationally efficient. GroupNet explicitly models group-level behaviors by identifying and tracking pedestrian groups over time, enabling the predictor to reason about shared goals and coordinated motion patterns. By incorporating group-aware representations, it improves forecasting in dense crowds where individual interactions alone are insufficient. BOSampler improves trajectory diversity by applying Bayesian optimization to guide sampling of latent variables toward high-quality, diverse futures. Rather than rely-



Model	Year	Improvements	Code	Pre-trained	Reproducible	Idea
<i>DLow</i> [144]	2020	5.5% up to 17.1%	<a href="#">Link</a>	✓	✓	Learn a diversity sampling function generating a diverse yet likely set of future trajectories.
<i>LDS</i> [145]	2021	3.6% up to 31.3%	<a href="#">Link</a>	×	✓	Use a pre-trained flow or VAE model to improve quality and expand sampling diversity.
<i>GroupNet</i> [136]	2022	3.1% up to 17.6%	<a href="#">Link</a>	×	✓	Use a trainable multiscale hypergraph to capture pair-wise and group-wise interactions at multiple group sizes.
<i>GP-Graph</i> [52]	2022	4.8% up to 48.4%	<a href="#">Link</a>	✓	✓	Include collective group representations for effective prediction in crowded environments.
<i>MOE</i> [59]	2022	2.8% up to 16.7%	<a href="#">Link</a>	×	✓	Momentary extractor for a more effective information capturing to reduce input size.
<i>NSPM</i> [53]	2022	-0.9% up to 60.2%	<a href="#">Link</a>	✓	✓	Concept of discrepancy as a measure for a learnable sampling strategy.
<i>EigenTrajectory</i> [51]	2023	-31.1% up to 74.3%	<a href="#">Link</a>	✓	✓	Trajectory descriptor to form an Eigen-Trajectory space, in place of Euclidean space, to represent pedestrian movements.
<i>Stimulus-Verification</i> [60]	2023	4.4% up to 29.3%	<a href="#">Link</a>	✓	✓	Uses factors in the observation that may affect future movements, such as interaction and scene context, to improve sampling.
<i>BOSampler</i> [38]	2023	-3.0% up to 44.0%	<a href="#">Link</a>	×	✓	Encouraging exploration of low-probability choices to improve the diversity of samples.
<i>SocialCircle</i> [24]	2024	0.0% up to 8.0%	<a href="#">Link</a>	✓	✓	A new angle-based trainable social interaction representation.

Table 5. Overview of techniques to improve existing methods sorted by year. Notations, acronyms, and symbols are the same as in Table 1 and Table 2.

ing on random sampling, it actively searches for informative samples that balance likelihood and diversity, often leading to significant performance gains. Performance can improve dramatically, but degradation is also observed.

Uncertainty handling and generating calibrated probability-related forecasts are treated significantly less frequently in HTP; only Social-STAGE [84], TUTR [79], BNSP-SFM [58], MDN [44], and DD-MDN [46] handle these topics from all reviewed frameworks. Social-STAGE [84] ranks multimodal Gaussian distributions by computing probabilities for each mode, where future trajectories are clustered into discrete motion patterns and assigned likelihood scores. This ranking mechanism enables relative comparison between predicted futures but relies on fixed Gaussian assumptions and does not explicitly model calibration. TUTR [79] employs a Transformer architecture to forecast  $K$  future trajectories in parallel, with associated probabilities implicitly derived from the learned attention-based data distribution. While effective at capturing long-range temporal dependencies and interaction patterns, the predicted probabilities are not explicitly calibrated or evaluated for distributional correctness. BNSP-SFM [58] formulates trajectory prediction as a Bayesian nonparametric stochastic flow model, treating trajectory sampling as a Gaussian process and constructing an acquisition function to quantify the expected sampling value. This design encourages exploration of the long-tail region of the trajectory distribution but focuses on sample diversity rather than probabilistic calibration or uncertainty consistency. However, these three methods do not cover uncertainty calibration or distributional correctness in any aspect. Currently, MDN [44] and DD-MDN [46] are the

only frameworks in HTP that address the calibration and evaluation of the underlying probability distributions to ensure they reflect genuine aleatoric uncertainty. DD-MDN [46] addresses the reliability gap in HTP by coupling a few-shot denoising diffusion backbone with a dual Mixture Density Network to generate uncertainty-calibrated residence areas and probability-ranked trajectories. The framework uses a dual-GM representation to ensure time-consistent anchor paths while maintaining calibrated aleatoric uncertainty at each future time step, thereby enabling the derivation of trustworthy confidence regions. Furthermore, it can rank the discrete future hypotheses by their calibrated probabilities. It achieves SOTA accuracy and exhibits unique robustness during momentary observations when the input horizon is limited to just two frames.

Overall, it must be noted that evaluations, results, and comparisons with other methods are sometimes presented in a non-transparent manner. Over the years, various methods have used the data preprocessing of Trajectron++ and propagated new SOTA ADE and FDE results, even though errors in the implementation were already proven in 2021 [120]. The error affected ADE/FDE results by 40 % on average (from 0.19/0.41m to 0.31/0.51m for Trajectron++ itself), and in some cases, methods from 2023 still had been affected or ignored, resulting in incorrect and too good results.

Furthermore, some methods use datasets or data splits that differ from those of others, thereby precluding direct comparability. Several methods achieve outstanding ADE/FDE results but do not publish/provide any code or other evidence. BNSP-SFM with (0.16/0.24m) or S-CSR

with (0.10/0.16m), both from 2021, report accuracy values that even the best reproducible methods from 2025/26 do not achieve by a wide margin. In summary, 47 % of all methods that report results for ETH/UCY have some kind of potential credibility issues.

### 3. Datasets

Evaluating HTP involves testing algorithms on public datasets. However, collecting and annotating large datasets is time-consuming and labor-intensive, particularly for supervised learning methods that rely heavily on high-quality annotations. Existing pedestrian trajectory prediction datasets can be categorized into surveillance-, sport-, and traffic-based application scenarios. Tab. 6 provides a detailed overview of publicly available HTP datasets. Surveillance-based datasets monitor a fixed scene to analyze crowded human scenarios, primarily involving inter-person interactions. Traffic-based datasets are used for pedestrian trajectory prediction in AD and AR applications, including data collected in urban areas by vehicles, drones, or intelligent infrastructure. These datasets provide fine-grained information about multiple traffic participants and mapping data.

The ETH/UCY dataset has been established as the standard benchmark, followed by the SDD dataset. Therefore, as shown in Tab. 1, Tab. 2, and Tab. 3, HTP evaluation is primarily focused on surveillance, and traffic-related assessments are rarely conducted. Moreover, positional accuracy measurements and evaluations on the ETH/UCY dataset have plateaued since 2023/2024. Multiple methods achieved overall similar accuracy, comparable to that of LED or EqMotion (2023), SingularTrajectory or MART (2024), MoFlow (2025), or DD-MDN (2026). A similar pattern is observed in the SDD dataset. TUTR (2023), MART or PPT (2024), and MNRF (2025) or DD-MDN (2026). Progress is still visible, but the step size is decreasing significantly, indicating that, in terms of pure accuracy on the current surveillance datasets, a natural barrier has been reached.

Waymo [126] and Argoverse 2 [9] are two large-scale AD datasets widely used for Vehicle Trajectory Prediction (VTP), which include multiple road-user classes, diverse scenarios, and more than 100k human trajectories. In HTP, only MDN uses the Waymo dataset [126]. In addition, inD [13] and IMPTC [45] are two high-quality, urban-area-focused, context-enriched trajectory datasets recorded by drones and intelligent infrastructure at critical intersections. inD is yet rarely used in HTP, and IMPTC is relatively new. JAAD [3] and PIE [4] include additional scene-related meta-information like gender, age, or intention to cross the street besides classical trajectories. The Forking-Path [62] dataset is currently the only HTP-suitable multi-modal dataset that includes multiple correct future paths for

each input. It is based on simulations. This demonstrates an apparent yet unresolved weakness, leaving room for the release of new datasets. In recent years, several works, such as LED [88], MemoNet [137], and Social-VAE [99], have used the NBA [119] and NFL [68] datasets to evaluate their models. These represent human motion and interactions in sports such as North American football and basketball.

### 4. Evaluation Metrics

The Best-of-K (BoK) ADE/FDE metric has been established as the standard performance evaluation in most publications, with  $K = 20$  (stochastic) samples followed by  $K = 1$  (deterministic). A minority uses additional metrics. Tab. 7 and Tab. 8 present an overview of all introduced or, for HTP evaluation, derived metrics. The tables are split into deterministic and probabilistic metrics.

Regarding deterministic metrics, ADE and FDE [100] are the standard metrics for HTP evaluation, measuring the average and final L2-Distance between the discrete predicted positions and Ground Truth (GT) at every forecast horizon  $H_{fc} = \{1, 2, \dots, n\}$ . Average/Final Self Distance ASD/FSD [143] is introduced to measure the diversity of  $K$  predictions, measuring the distance between two samples. The Miss Rate (MR) [126] counts the number of discrete predicted positions that fall within a defined range around the GT position, an environmental accuracy measure compared with the strict positional accuracy used by ADE/FDE. Mean Average Precision (mAP) [126] is used for trajectory evaluation, measuring whether a prediction is correctly classified into the defined motion classes. As defined by the Waymo dataset [126], eight motion classes exist. The Collision Rate (CR) [31] metric determines whether two objects' predictions will collide, as defined by a collision radius. It is a valuable extension of ADE/FDE, particularly when social interaction handling and evaluation are of interest. Environment Collision-Free Likelihood (ECFL) [109] measures how well an agent can interact with its surrounding environment. Collision Rate [31] measures the occurrence of predicted collisions. Joint-ADE/FDE (JADE/JFDE) [3] extends classical metrics to account for surrounding agents and potential collision scenarios, thereby reducing unnatural predictions, such as colliding or diverging trajectories within groups. Brier-ADE/FDE (BADE/BFDE) [82] incorporates a prediction probability to facilitate accurate predictions with a high confidence level.

Regarding probabilistic metrics, the Negative Log-Likelihood (NLL) [14] quantifies how well the predicted mean and spread align with the GT. Still, NLL does not evaluate the predicted spread itself: if the model predicts a trajectory with high uncertainty (a wide spread of possible positions), but the actual position falls within that range, the NLL will be lower. If the model predicts a trajectory with low uncertainty (a narrow spread of possible posi-

Datasets	Year	View	Sensors	Framerate	Area	Objects	Target
<i>UCY</i> [71]	2007	Bird	Camera	2.5 Hz	Public Areas	P	Surveillance
<i>ETH</i> [100]	2009	Bird	Camera	2.5 Hz	Public Areas	P	Surveillance
<i>GCS</i> [116]	2012	Bird	Camera	2.5 Hz	Public Areas	P	Surveillance
<i>SDD</i> [105]	2016	Bird	Camera	30.0 Hz	Public Areas	P	Surveillance
<i>NBA</i> [119]	2017	Bird	Camera	10.0 Hz	Sport Areas	P	Sport
<i>NFL</i> [68]	2019	Bird	Camera	10.0 Hz	Sport Areas	P	Sport
<i>KITTI</i> [36]	2012	Vehicle	Camera, Lidar	10.0 Hz	Urban Areas	V, P, C	Traffic
<i>Cityscapes</i> [26]	2016	Vehicle	Camera	16.0 Hz	Urban Areas	V, P, C	Traffic
<i>JAAD</i> [3]	2017	Vehicle	Camera	30.0 Hz	Urban Areas	P	Traffic
<i>ApolloScape</i> [135]	2018	Vehicle	Camera, Lidar	10.0 Hz	Urban Areas	V, P, C, B	Traffic
<i>PIE</i> [4]	2019	Vehicle	Camera	30.0 Hz	Urban Areas	P	Traffic
<i>nuScenes</i> [16]	2019	Vehicle	Camera, Lidar, Radar	2.0 Hz	Urban Areas	V, P	Traffic
<i>Waymo</i> [126]	2019	Vehicle	Camera, Lidar	2.0 Hz	Urban Areas	V, P, C	Traffic
<i>Interaction</i> [152]	2019	Bird	Camera	10.0 Hz	Urban Areas	V, P	Traffic
<i>Forking-Paths</i> [62]	2019	Bird	Camera	30.0 Hz	Public Areas	P	Traffic
<i>Argoverse 1</i> [18]	2020	Vehicle	Camera, Lidar	10.0 Hz	Urban Areas	V, P, C	Traffic
<i>Argoverse 2</i> [9]	2021	Vehicle	Camera, Lidar	10.0 Hz	Urban Areas	V, P, C	Traffic
<i>AP</i> [67]	2021	Vehicle	Camera	25.0 Hz	Urban Areas	P, C	Traffic
<i>ONCE</i> [87]	2021	Vehicle	Camera, Lidar	10.0 Hz	Urban Areas	P, C	Traffic
<i>nuPlan</i> [17]	2022	Vehicle	Camera, Lidar	10.0 Hz	Urban Areas	P, C	Traffic
<i>Cyclist-Action</i> [150]	2020	Infrastructure	Camera	50.0 Hz	Urban Intersection	P, C	Traffic
<i>inD</i> [13]	2020	Bird	Camera	25.0 Hz	Urban Intersection	V, P, C	Traffic
<i>SIND</i> [138]	2022	Infrastructure	Camera	10.0 Hz	Urban Intersection	V, P, C	Traffic
<i>IMPTC</i> [45] [43]	2023	Infrastructure	Camera, Lidar	25.0 Hz	Urban Intersection	V, P, C, B, S, E	Traffic
<i>V2X-Seq</i> [146]	2023	Infrastructure	Camera, Radar	10.0 Hz	Urban Intersection	V, P, C	Traffic
<i>OnSiteVRU</i> [141]	2025	Infrastructure	Camera, Radar	25.0 Hz	Urban Intersection	V, P, C	Traffic

Table 6. Overview of public datasets for HTP research and applications. Object classes acronyms: V: Vehicle, P: Pedestrian, C: Cyclist, B: Motorbike, S: Stroller, E: E-scooter.

tions) but the actual position is far from the expected mean, the NLL will be higher. KDE-NLL [108, 112] extends the NLL for methods that cannot provide distributional predictions, utilizing a Kernel Density Estimation (KDE) to create a distributional representation for deterministic predictions, thereby enabling the application of the NLL measurement. Reliability (RLS) and Sharpness (SS) [44, 151] measure the observed relative frequency of occurrences by calculating the estimated Confidence Levels (CL) for every corresponding pair of predicted distribution and ground truth point, and the volumetric measure of the confidence levels. The metrics provide a statistical guarantee of probabilistic and positional correctness and measure the spread size expansion over time. Precision and Recall [98] measure how well a predicted distribution can cover all ground truth trajectories. In this case, the metric depends on the availability of multimodal GT data, a feature currently available only in the FPD dataset.  $M_1$  [84] aims to compute the diversity of predicted trajectories sampled from the best mode of a multimodal distribution in relation to the ground truth, and  $M_2$  [84] evaluates the error contribution of other modes concerning their confidence. PCMD [39] considers the predictions with corresponding probabilities and evaluates the prediction model under the whole latent distribution. Delta Empirical Sigma Value (DESV) [15] targets the identification of over- or underconfidence, using the difference in the fraction of GT positions that fall within the

$i$ - $\sigma$  level set (e.g.,  $1\sigma, 2\sigma, 3\sigma$ ) of the predicted distribution and the fraction from an ideal Gaussian. The Average Mahalanobis Distance (AMD) [93] attempts to quantify the closeness of all generated samples to the ground truth. In contrast, the Average Maximum Eigenvalue (AMV) [93] describes the overall spread of the predictions. Percentage of Trajectory Usage (PTU) [72] evaluates the comprehensiveness of the performance of multi-future prediction. It computes the generated-sample-to-ground-truth ratio and penalizes out-of-distribution predictions. As Precision and Recall, the metric depends on the availability of multimodal GT data. Maximum Mean Discrepancy with Gaussian kernel (MMD) [1] measures the difference between the mean embeddings of the two distributions, and Brier Score (BS) [1] measures the accuracy of predicted outcome probabilities using the mean squared error between the predicted and the actual outcome. The Average and Final-Negative Log-Likelihood (ANLL, FNLL) [127] estimates how well the predicted distribution matches the observed data by verifying if the weight of the best fitting mode to GT is the highest. The metric evaluates the correct weighting of multi-modal mixtures but provides no information about the single components. Percentage of Trajectory Usage (PTU) [72] evaluates the comprehensiveness of the performance of multi-future prediction. It computes the generated-sample-to-ground-truth ratio and penalizes out-of-distribution predictions. PTU demonstrates how many



Metric	Year	Definition	Description
<i>ADE</i> [100]	2009	$ADE = \frac{1}{N} \sum_{t=0}^{H_{fc}} \ y_t - \hat{y}_t\ _2$	Average L2 distance between the predicted trajectory $\hat{y}_t$ and GT trajectory $y_t$ over time $t \in H_{fc}$
<i>FDE</i> [100]	2009	$FDE = \ y_T - \hat{y}_T\ _2$	L2 distance between the final predicted position $\hat{y}_T$ and the final ground truth position $y_T$
<i>ASD</i> [143]	2020	$ASD = \frac{1}{N} \sum_{t=0}^{H_{fc}} \ \hat{y}_t^{(i)} - \hat{y}_t^{(j)}\ _2$	Average L2 distance over all time steps between a forecasted sample $\hat{y}_t^{(i)}$ and its closest neighbor $\hat{y}_t^{(j)}$ with $i, j \in K$ .
<i>FSD</i> [143]	2020	$FSD = \ \hat{y}_T^{(i)} - \hat{y}_T^{(j)}\ _2$	L2 distance between the final position of a forecasted sample $\hat{y}_T^{(i)}$ and its closest neighbors' sample final position $\hat{y}_T^{(j)}$ with $i, j \in K$ .
<i>MR</i> [126]	2021	$\text{IsMatch}(\hat{y}_t, y_t) = \mathbb{1}[\hat{y}_t^k < \lambda^{lon}] \cdot \mathbb{1}[\hat{y}_t^k < \lambda^{lat}]$	A binary match/miss indicator function $\text{ISMATCH}(\hat{y}_t, y_t)$ is assigned to each sample $k$ waypoint at a time $t$ , $\lambda^{lon}$ and $\lambda^{lat}$ are longitudinal and lateral thresholds that vary with time and velocity.
<i>mAP</i> [126]	2021	$mAP = \frac{N}{M}$	With $N$ denoting the total number of correctly classified data samples and $M$ denoting the total number of data samples. The final mAP metric averages over eight GT trajectory shapes: straight, straight-left, straight-right, left, right, left u-turn, right u-turn, and stationery.
<i>ECFL</i> [109]	2022	$ECFL(\hat{y}_a, \mathbf{E}) = \left[ \frac{1}{K} \sum_{k=1}^K \Pi_{t=0}^{H_{fc}} \mathbf{E}[\hat{y}_{a,k,t,1}, \hat{y}_{a,k,t,0}] \right]$	With denoting $K$ possible futures, an agent $a$ can interact with the environment, represented by a binary matrix $\mathbf{E}$ . $\hat{y}_{a,k,t}$ is the agents $a$ current position taking path $k$ at time step $t$ .
<i>CR</i> [31]	2023	$CR_{JADE}(\mathbf{y}) = \frac{1}{A} \sum_{a=1}^A \mathbb{1}[\text{collision}(\mathbf{y}_a^k, \mathbf{y}_{b \neq a}^k)]$	With denoting the predictions over all agent-timestep-samples as $\mathbf{y}$ , and $\text{collision}$ is a function that returns True if two line segments come within a defined radius $r$ of each other.
<i>JADE</i> [31]	2023	$JADE = \frac{1}{H_{fc}A} \min_{k=1}^K \sum_{a=1}^A \sum_{t=1}^{H_{fc}} \ \hat{y}_{t,a}^k - y_{t,a}^k\ _2^2$	With denoting $A$ as the number of agents, and $K$ as the number of discrete predicted samples.
<i>JFDE</i> [31]	2023	$JFDE = \frac{1}{A} \min_{k=1}^K \sum_{a=1}^A \ \hat{y}_{T,a}^k - y_{T,a}^k\ _2^2$	With denoting $A$ as the number of agents, and $K$ as the number of discrete predicted samples.
<i>BADE</i> [82]	2023	$BADE = ADE + (1-p)^2$	With $p$ denoting the probability of the predicted trajectory.
<i>BFDE</i> [82]	2023	$BFDE = FDE + (1-p)^2$	With $p$ denoting the probability of the predicted trajectory.

Table 7. Overview of deterministic evaluation metrics for HTP applications. Unless otherwise noted, the following designations will be used: A predicted trajectory is defined as  $\hat{y}_t$ , the GT as  $y_t$ , with discrete moments in time  $t \in H_{fc}$  as forecast horizon  $H_{fc} = \{1, 2, \dots, n\}$  and the maximum forecast length  $N$ . Multiple methods predict a set of  $K$  discrete future trajectories with several  $A$  agents, and  $p$  represents the probability of a single prediction.

K-sampled discrete trajectories can match N-possible GTs. It is only applicable when multimodal GT data are available, e.g., the Forking Path Dataset (FPD) [62]. Prediction Interval Coverage Probability (PICP) and Mean Prediction Interval Width (MPIW) [8] evaluate if the corresponding GT position lies within the predicted distribution or not, and quantifies the average size of all distributions for each time step.

## 5. Summary and Objectives

The following chapter assesses the current state of HTP, including the disclosure of limitations and the potential for improvement. Our central claim is that the field’s prevailing objective, minimizing average geometric error on a narrow set of convenience benchmarks, does not adequately reflect the demands of autonomous driving, robot navigation, and smart-city surveillance. Two recent surveys sharpen this perspective from complementary angles: Schoeller et al. show that a constant-velocity formulation poses a surprisingly strong baseline on the ETH/UCY dataset, which indicates that many learned models exploit scene- or dataset-specific regularities rather than robust interaction reasoning [110]. Dietl et al. synthesize evidence that widely used datasets underrepresent high-complexity behaviors and of-

fer uneven coverage of social and infrastructural regimes, implying that global averages often reward performance on abundant easy slices rather than on safety-critical tails [29]. Together, these insights motivate a reframing of the common objective from point accuracy on fixed splits to calibrated, risk-aware, and transferable forecasts, whose value is demonstrated under distribution shifts and whose uncertainty is meaningful to downstream decision-making.

### 5.1. Dataset Diversity

As described in Sec. 2, the ETH/UCY dataset is the de facto standard for HTP evaluation, with usage of 97%, followed by SDD at 31%. Notably, 41% of all publications evaluate exclusively on ETH. The dataset is small, comprising 1536 pedestrian trajectories sampled at 2.5 Hz. While having a standard dataset for easy comparison with each other is fine, Schoeller et al. demonstrate that the included motion patterns contain relatively few genuine interaction turning points and are dominated by rectilinear paths [110]. While standardization accelerates comparison, a singular emphasis on ETH/UCY narrows the behavioral envelope against which methods are optimized.

Dietl et al. consolidate dataset-level evidence across ETH/UCY, SDD, and inD, highlighting systematic dif-

Metric	Year	Definition	Description
<i>NLL</i> [14]	2018	$NLL = -\sum_{i=1}^I p_i \log(\hat{p}_i)$	With $p_i$ denoting a binary indicator of the correctness of predicting the data sample in class $i$ , $\hat{p}_i$ is the predicted probability of the data sample belonging to class $i$ , and $I$ is the total number of classes.
<i>KDE-NLL</i> [112]	2018	$KDENLL = -\frac{\sum_{i=1}^n NLL(p_{\hat{y}}^i, p_y^i)}{n}$	With $n$ denoting the total number of trajectory points, $p_{\hat{y}}^i$ denoting the probability distribution of the predicted trajectories estimated by kernel density at step $i$ , $p_y^i$ denoting the probability distribution of the ground truth trajectory, and $NLL$ denoting the negative log-likelihood function.
<i>RLS</i> [121] [122]	2019	$RLS = 1 - \frac{1}{ H_K  \cdot  A } \sum_{h \in H_K} \sum_{\alpha \in A}  1 - \alpha - f_{o;t+h}(1 - \alpha) $	With denoting $f_{o;t+h}$ as the observed relative frequency of occurrences, $\alpha$ represents certain confidence levels from the range $\alpha \in [0.0, 1.0]$ resulting in $\alpha \in A \{0.01, 0.02, \dots, 0.99\}$ for every corresponding pair of predicted distribution and GT point.
<i>SS</i> [121] [122]	2019	$\Omega(1 - \alpha) = \{\hat{y}   \hat{y} \in \mathbb{R}^2 \wedge \overbrace{1 - \alpha(\hat{y})}^{\text{confidence level of } \hat{y}} \geq 1 - \alpha\}$	With denoting $\Omega(1 - \alpha)$ as a certain confidence level $\alpha \in [0.0, 1.0]$ .
<i>Precision</i> [98]	2021	$Precision = \frac{M_i}{M}$	With $M_i$ denoting the total number of samples that are correctly classified in that class and with $M$ denoting the total number of samples classified as the given class, whereas classes refer to given ground truth modes. It measures the ratio of generated samples in the support of the ground truth distribution.
<i>Recall</i> [98]	2021	$Recall = \frac{M_i}{N_i}$	With $M_i$ denoting the total number of data samples which are correctly classified in that class and $N_i$ denoting the total number of samples in the given classes, whereas classes refer to given ground truth modes.
$M_1$ [84]	2021	$M_1 = \frac{1}{M} ((\sum_i e_i) - \hat{e})$	With $M$ denoting the number of modes, error $\hat{e}$ contributed by probability $\hat{p}$ of mode ( $\hat{m}$ ) from expectation of all modes errors $E(e_i)$ .
$M_2$ [84]	2021	$M_2 = (\sum_i p_i * e_i) - p_{max} * e_{p_{max}}$	With $\hat{e} = e_{p_{max}}$ as subtracted weighted error of the best mode, maximum probability mode error with $\hat{p} = p_{max}$ from the weighted expectation of errors of all modes using the probabilities $p$ predicted.
<i>PCMD</i> [39]	2021	$PCMD(k) = \mathbb{E}_{\mathbf{z}} \min\{\mathcal{D}(\mathbf{z})   \mathbf{z} \in \{z_1^*, z_2^*, \dots, z_m^*\}\}$	With denoting $\mathcal{D}(\mathbf{z})$ as the ADE or FDE distance between the GT and the sampled prediction based on the distribution $\mathbf{z}$ , $k = \frac{m}{M}$ as the ranking rate in the numerical calculation, e.g. calculating the minimum ADE/FDE of top 20 probability trajectories, when $k = \frac{20}{M}$ , and sampling $M$ variables $Z = \{\mathbf{z}_i \in \Omega   i = 1, 2, \dots, M\}$ and sort these variables with the probability $p_{\theta}(\mathbf{z}_i   \mathbf{x})$ from large to small to obtain $Z_{sort} = \{\mathbf{z}_i^*\}$ .
<i>DESV</i> [15]	2022	$\Delta ESV_i := \sigma_{p,i} - \sigma_{ideal,i}$	With denoting $\sigma_{p,i}$ is the empirical fraction of GT positions that lie within the $i$ -sigma level set of the prediction distribution and $\sigma_{ideal,i}$ is the expected fraction from a perfectly-calibrated bivariate Gaussian, where $\sigma_{ideal,1} \approx 0.39$ , $\sigma_{ideal,2} \approx 0.86$ , and $\sigma_{ideal,3} \approx 0.99$ .
<i>PTU</i> [72]	2022	$PTU = \frac{\sum_{i=1}^N  \hat{y}_i  /  y_i }{N}$	With $ \hat{y}_i $ denoting the number of predictions used while evaluating with $minADE_K$ and $minFDE_K$ of $K$ samples, $ y_i $ denotes the number of ground truth trajectories in a data sample, and $N$ representing the number of data samples.
<i>AMD</i> [93]	2022	$AMD = \frac{1}{A \times H_{fc}} \sum_{a \in A} \sum_{t \in H_{fc}} M_D(\hat{\mu}_{GMM,t}^a, \hat{G}_t^a, y_t^a)$	With denoting $M_D$ as Mahalanobis Distance, $\hat{\mu}$ as mean of the prediction, $\hat{G}$ the inverse covariances of each mixture components averaged and weighted probabilistically, and $\hat{\mu}_{GMM} = \sum_{k=1}^K \hat{\pi}_k \hat{\mu}_k$ is the mean of the GMM.
<i>AMV</i> [93]	2022	$AMV = \frac{1}{A \times H_{fc}} \sum_{a \in A} \sum_{t \in H_{fc}} \lambda_1^{\downarrow}(\hat{\Sigma}_{GMM,t}^a)$	With denoting $\lambda_1^{\downarrow}$ is the eigenvalue with the largest magnitude from the matrix eigenvalues, and $\hat{\Sigma}_{GMM}$ is the covariance matrix of the predicted GMM distribution.
<i>MMD</i> [1]	2023	$MDD(\mathbb{F}, X, Y) = \sup_{f \in \mathbb{F}} (\frac{1}{m} \sum_{i=1}^m f(x_i) - \frac{1}{n} \sum_{j=1}^n f(y_j))$	With denoting $X$ and $Y$ as two distributions over set $F$ , and $m, n$ denoting the numbers of discrete samples drawn from $X, Y$ .
<i>BS</i> [1]	2023	$BS = \frac{1}{k} \sum_{i=1}^k (y_i - \hat{y}_i)^2$	With denoting $\hat{y}_i$ as prediction, $y_i$ as ground truth, and $k$ as the number of forecasts.
<i>ANLL</i> [127]	2023	$ANLL = \frac{1}{H_{fc}} \sum_{k=1}^{H_{fc}} -\log(\sum_j \pi^j \mathcal{N}(\mathbf{x}_k   \hat{\mathbf{x}}_k^j, \mathbf{P}_k^j))$	With denoting $\mathbf{P}_k^j$ the estimated state covariance of all mixtures $j \in \{1, \dots, H_{fc}\}$ , $\hat{\mathbf{x}}_k^j$ the predicted mean of the states $\mathbf{x}_k$ , $\pi^j$ GMM mixing coefficients, and $\mathcal{N}$ is the normal distribution.
<i>FNLL</i> [127]	2023	$FNLL = -\log(\sum_j \pi^j \mathcal{N}(\mathbf{x}_T   \hat{\mathbf{x}}_T^j, \mathbf{P}_T^j))$	With denoting $\mathbf{P}_T^j$ the estimated state covariance of the final mixture of $H_{fc}$ , $\hat{\mathbf{x}}_T^j$ the predicted mean of the state $\mathbf{x}_T$ , $\pi^j$ GMM mixing coefficients, and $\mathcal{N}$ is the normal distribution.
<i>PICP</i> [8]	2024	$PICP \approx \frac{1}{ \mathcal{D}^* } \sum_{(x,y \in \mathcal{D}^*)} \mathbb{1}(y_k \in \Gamma(X_k))$	Coverage probability for a single state shows whether the ground truth state, $\mathbf{y}_k$ lies within the predicted covariance ellipse $\Gamma(X_k)$ for the state $\mathbf{X}_k$ , $\mathbb{1}$ denotes an indicator function representing Boolean values.
<i>MPIW</i> [8]	2024	$MPIW \approx \frac{1}{ \mathcal{D}^* } \sum_{(x,y \in \mathcal{D}^*)} ( u(x) - l(x) )$	With denoting $u(x)$ and $l(x)$ as lower and upper bounds for the prediction interval, others see <i>PICP</i> .

Table 8. Overview of probabilistic evaluation metrics for HTP applications. Unless otherwise noted, the designations are equal to those in Tab. 7.

ferences in how much social density, infrastructure constraints, and heterogeneous road users are represented [29]. Pedestrian–pedestrian interactions are most prevalent in ETH/UCY-style campus scenes. In contrast, SDD and inD expose richer traffic layouts and multi-class interactions, but often with fewer social-pressure configurations. High-complexity behaviors such as abrupt stops, restarts, sharp

turns, merges, or cross-class negotiations are comparatively rare across the board. Consequently, models can achieve competitive headline numbers without demonstrating competence on the behaviors that carry disproportionate operational risk. Beyond recommending just more datasets, these observations argue for curating evaluation suites that deliberately span social and infrastructural regimes and for

reporting stratified performance by behavioral complexity, so that gains are not merely re-statements of dataset priors but evidence of transferable capability. As already outlined in Sec. 3, pairing ETH/UCY with broader sets such as SDD, Waymo, and inD provides a more faithful proxy for typical AD and SCS requirements. It is more likely to surface weaknesses and generalization problems that single-domain protocols might conceal.

## 5.2. Observation Duration Scaling

The common practice of using long input sequences, typically 3.2 s, is misaligned with several deployment realities. In AD and HMI, prediction must often begin immediately after initial detection, with short, noisy observations and imperfect association in crowded scenes. Under these conditions, longer observation windows can blur predictive skill by compromising tracking stability and accumulating fragmentation errors. Recent studies explicitly probe the scaling of prediction quality with observation horizon. STT [94], MOE [59], and SingularTrajectory [55] report up to 50% degradation in accuracy when the input is reduced from 3.2 s to 0.8 s on ETH/UCY, and a comparative evaluation by Uhlemann et al. [129] finds that Trajectron++ and SGAN degrade less than Y-Net, Social-Implicit, and AgentFormer. Schoeller et al. further show that most of the predictive signal used by typical neural architectures is concentrated in the most recent 0.8 s of motion, with earlier history contributing comparatively little [110]. These converging results indicate that objective design and reporting should treat observation duration as a first-class variable. In particular, short input-horizon calibration, e.g., the ability to represent widening uncertainty as information diminishes, deserves emphasis in robot navigation and autonomous driving, where reaction time is intrinsically limited, and where the benefit of longer lookbacks may broadly reflect tracking convenience rather than robust inference.

## 5.3. Evaluation Metric Expansion

A shared metric vocabulary is indispensable; however, relying on ADE/FDE as primary objectives underspecifies reliability, usability, and trustworthiness. As surveyed in Sec. 4, JADE/JFDE extend geometric assessment to social interactions, while PICP, AMD, PCMD, and RLS target positional correctness, and MPIW, AMV, and SS capture distributional spread, e.g., uncertainty expansion over time. Crucially, PCMD and RLS incorporate prediction probabilities, with RLS additionally incorporating a statistical reliability criterion, making it a natural anchor for probabilistic evaluation. Without probability information and calibration, multi-hypothesis predictions become difficult to rank at decision time, since ground truth is absent during operation [7]. Empirical evidence underscores this gap: evaluations in [44] show that none of several popular methods

deliver well-calibrated predictions on ETH/UCY; MDN attains the strongest reliability despite not achieving the best ADE/FDE accuracy. Most other methods tend to produce overconfident hypotheses, a predictable result of their focus on and optimization of positional accuracy.

The objective should therefore shift from optimizing point-wise displacement to optimizing proper probabilistic scores, along with calibration and sharpness. Negative log-likelihood can prevent illusory gains from overconfident underdispersion. At the same time, calibration diagnostics ensure that nominal coverage holds across social-density strata rather than only on straight, low-risk segments. JADE/JFDE deserve greater weight to reflect interaction handling, and risk-sensitive geometric summaries that up-weight small time-to-collision and close-proximity frames align better with autonomous driving and surveillance priorities than unweighted averages. Taken together, an evaluation built around RLS (for trustworthiness), complemented by AMV or SS (for sharpness) and JADE/JFDE (for interaction quality), offers a principled balance between accuracy and operational value, reducing the incentive to overfit to the narrow conditions under which simple kinematics already excels.

## 5.4. Evaluation Objectives

BoK ADE/FDE are the standard evaluation metrics in HTP; therefore, optimizing for these values has become crucial, leaving less room for other essential evaluation signals. Especially on ETH/UCY, a performance ceiling is visible, and some architectures have moved away from explicitly representing distributions, instead sampling a fixed number  $K$  of discrete predictions (e.g., TUTR, SingularTrajectory), which makes probabilistic assessment more difficult. In parallel, ultra-sampling strategies, such as Social-CVAE, NSP-SFM, and CSR, can achieve impressive accuracy (Social-CVAE reports ADE/FDE of 0.02/0.03 m), yet these scores are below typical sensor and annotation noise levels. This evaluation method lacks a sound foundation or reasoning and is therefore rejected by most researchers. None of the methods that use ultra-sampling for evaluation provides any valid arguments for their usefulness and relevance, nor do they give any probabilities or uncertainty estimates that could support such a measure. These trends suggest the need to re-anchor objectives.

A coherent objective for the next phase of HTP should be application-aligned, complexity-aware, and robust by design. Application alignment refers to the extent to which metrics correlate with downstream planner costs. For autonomous driving, this involves reduced surrogate collision risk and fewer high-jerk maneuvers under tight time-to-collision conditions. For robot navigation, it entails improved short-horizon calibration in dense local neighborhoods. For smart city surveillance, it results in greater sen-

sitivity to deviations from nominal flow. Complexity awareness means that both training and reporting allocate attention to high-social-pressure, non-linear, and start/stop behaviors that are rare but consequential, using stratified supervision and stratified metrics rather than single-number aggregates. Robustness by design requires cross-scene and cross-dataset validation so that gains are portable beyond the scene priors of a single campus or intersection. Within this framework, simple kinematic baselines remain essential reference points; improvements should be considered substantive only when they persist across scenes, horizons, and complexity strata and when they coincide with improved calibration and risk profiles [29, 110].

In short, probability-aware, risk-sensitive, and transfer-conscious objectives offer a more faithful path toward deployable prediction than chasing marginal ADE/FDE gains on narrow protocols.

## 6. Conclusion

This survey repositions the objective of human trajectory prediction around the requirements of real-world autonomy rather than leaderboard convenience and highlights the need for a rethink. It presents multiple high-quality HTP-related datasets and an extensive summary of application-oriented evaluation metrics that account for probabilistic aspects of reliability and robustness. The literature reviewed here shows, on the one hand, that short-horizon kinematics already explain a surprising share of benchmark performance [110], and on the other, that today’s datasets offer uneven behavioral coverage and sparse exposure to the very regimes that matter most in deployment [29].

Recognizing these facts, progress should be judged by whether predictions are calibrated, risk-aware, and reliable under distribution shift, and by whether they translate into tangible gains for downstream decision-making. Concretely, this entails diversified and stratified evaluations across datasets and behaviors; horizon-aware reporting that exposes the information content of short observations; and probabilistic metrics, led by reliability-focused criteria such as RLS and complemented by sharpness and interaction measures that make uncertainty actionable. If future work adopts these objectives, apparent improvements will be more likely to reflect genuine advances in reasoning about people and environments, and less likely to arise from overfitting to dataset idiosyncrasies. The outcome will be prediction systems that transfer across scenes and tasks, communicate uncertainty in a way that planners can trust, and measurably contribute to the safety and efficiency of numerous safety-critical applications and deployments.

## References

- [1] A. Aastha, L. Rebecca, and A. Nisar. Learning to forecast aleatoric and epistemic uncertainties over long horizon trajectories. *IEEE International Conference on Robotics and Automation*, 2023. 8, 10
- [2] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. *IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 2016. 2, 3
- [3] R. Amir and K. Iuliia. Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior. *IEEE International Conference on Computer Vision Workshops*, 2017. 7, 8
- [4] R. Amir and K. Iuliia. Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. *IEEE/CVF International Conference on Computer Vision*, 2019. 7, 8
- [5] Javad Amirian, Jean-Bernard Hayet, and Julien Pettr . Social ways: Learning multi-modal distributions of pedestrian trajectories with gans. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 2, 4
- [6] R. Andrey, P. Luigi, and H. Michael. Human motion trajectory prediction: a survey. *The International Journal of Robotics Research*, 39, 2019. 2
- [7] N Anshul, E Azim, and R. Zachary. Uncertainty estimation of pedestrian future trajectory using bayesian approximation. *IEEE Open Journal of Intelligent Transportation Systems*, 3, 2022. 11
- [8] N Anshul, E. Azim, R. Zachary, and G. Prasenjit. Pedestrian trajectory forecasting using deep ensembles under sensing uncertainty. *IEEE Transactions on Intelligent Transportation Systems*, 2023. 9, 10
- [9] W. Benjamin and Q. William. Argoverse 2: Next generation datasets for self-driving perception and forecasting. *International Conference on Neural Information Processing Systems*, 2023. 7, 8
- [10] Alessia Bertugli and Rita Cucchiara. Ac-vrnn: Attentive conditional-vrnn for multi-future trajectory prediction. *Comput. Vis. Image Underst.*, 2020. 5
- [11] Apratim Bhattacharyya, Christoph Nikolas Straehle, Mario Fritz, and Bernt Schiele. Haar wavelet based block autoregressive flows for trajectories. *Pattern Recognition*, 2020. 2
- [12] Niccol  Bisagno, Bo Zhang, and Nicola Conci. Group lstm: Group trajectory prediction in crowded scenarios. *ECCV Workshops*, 2018. 2
- [13] J. Bock and R. Krajewski. The ind dataset: A drone dataset of naturalistic road user trajectories at German intersections. *IEEE Intelligent Vehicles Symposium*, 2020. 4, 7, 8
- [14] I. Boris and P. Marco. The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs. *2019 IEEE/CVF International Conference on Computer Vision*, 2018. 2, 3, 7, 10
- [15] I. Boris, L. Yifeng, and S. Shubham. Propagating state uncertainty through trajectory forecasting. *IEEE International Conference on Robotics and Automation*, 2022. 8, 10



- [16] H. Caesar and V. Bankiti. nuscenes: A multimodal dataset for autonomous driving. *IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 2020. 8
- [17] Holger Caesar, Juraj Kabzan, Kok Seang Tan, Whye Kit Fong, Eric Wolff, Alex H. Lang, Luke Fletcher, Oscar Beijbom, and Sammy Omari. nuplan: A closed-loop ml-based planning benchmark for autonomous vehicles. In *arXiv preprint arXiv:2106.11810*, 2021. 8
- [18] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 8
- [19] G. Chen, J. Li, and N. Zhou. Personalized trajectory prediction via distribution discrimination. *2021 IEEE/CVF International Conference on Computer Vision*, 2021. 3
- [20] H. Cheng, M. Liu, L. Chen, H. Broszio, M. Sester, and M. Yang. Gatraj: A graph- and attention-based multi-agent trajectory prediction model. *ISPRS Journal Of Photogrammetry And Remote Sensing*, 2023. 4
- [21] X. Chenxin, T. Robby, and T. Yuhong. Eqmotion: Equivariant multi-agent motion prediction with invariant interaction reasoning. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2, 4
- [22] Luigi Filippo Chiara and Lamberto Ballan. Goal-driven self-attentive recurrent networks for trajectory prediction. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2022. 5
- [23] W. Conghao, X. Beihao, and P. Qinmu. Another vertical view: A hierarchical network for heterogeneous trajectory prediction via spectrums. *ArXiv*, 2023. 4, 5
- [24] W. Conghao, X. Beihao, and Y. Xinge. Socialcircle: Learning the angle-based social interaction representation for pedestrian trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 5, 6
- [25] W. Conghao, X. Beihao, and H. Ziming. View vertically: A hierarchical network for trajectory prediction via fourier spectrums. *IEEE/CVF European Conference On Computer Vision*, 2022. 4, 5
- [26] M. Cordts, M. Omran, S. Ramos, and S. Roth. The cityscapes dataset for semantic urban scene understanding. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 8
- [27] Y. Cunjun, M. Xiao, and R. Jiawei. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. *IEEE/CVF European Conference On Computer Vision*, 2020. 2, 3
- [28] Patrick Dendorfer, Aljosa Osep, and Laura Leal-Taixé. Goal-gan: Multimodal trajectory prediction based on goal position estimation. *ArXiv*, 2020. 2, 3
- [29] Laura Dietl and Christian Facchi. Really, pedestrian trajectories: How realistic are the datasets? *IEEE Intelligent Vehicles Symposium*, 2025. 9, 10, 12
- [30] J. Eilbrecht, M. Bieshaar, S. Zernetsch, K. Doll, B. Sick, and O. Stursberg. Model-predictive planning for autonomous vehicles anticipating intentions of vulnerable road users by artificial neural networks. *2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, November 27–Dec. 1, 2017*, pages 1–8, 2017. 1
- [31] W. Erica, H. Hanako, and R. Deva. Joint metrics matter: A better standard for trajectory forecasting. *IEEE/CVF International Conference on Computer Vision*, 2023. 7, 9
- [32] A. Eskandarian and C. Wu. Research advances and challenges of autonomous and connected ground vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 22:683–711, 2019. 1
- [33] Zilin Fang, David Hsu, and Gim Hee Lee. Neuralized markov random field for interaction-aware stochastic human trajectory prediction. *International Conference on Learning Representations*, 2025. 2, 4, 5
- [34] G. Francesco, H. Irtiza, and C. Marco. Transformer networks for trajectory forecasting. *International Conference on Pattern Recognition (ICPR)*, 2020. 2, 3
- [35] Yuxiang Fu, Qi Yan, Lele Wang, Ke Li, and Renjie Liao. Moflow: One-step flow matching for human trajectory forecasting via implicit maximum likelihood estimation based distillation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 2, 4, 5
- [36] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 8
- [37] T. Gu, G. Chen, J. Li, C. Lin, Y. Rao, J. Zhou, and J. Lu. Stochastic trajectory prediction via motion indeterminacy diffusion. *IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 2022. 2, 4, 5
- [38] C. Guan-Hong and C. Zhenhao. Unsupervised sampling promoting for stochastic human trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 5, 6
- [39] C. Guangyi, L. Junlong, and Z. Nuoxing. Personalized trajectory prediction via distribution discrimination. *IEEE/CVF International Conference on Computer Vision*, 2021. 8, 10
- [40] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. *IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 2018. 2, 3
- [41] Z. Hao, R. Dongchun, Y. Xu, and F. Mingyu. Sliding sequential cvae with time variant socially-aware rethinking for trajectory prediction. *ArXiv*, 2021. 3, 4
- [42] Z. Hao, R. Dongchun, Y. Xu, and F. Mingyu. Csr: Cascade conditional variational auto encoder with socially-aware regression for pedestrian trajectory prediction. *Pattern Recognition*, 133, 2022. 4
- [43] M. Hetzel, H. Reichert, K. Doll, and B. Sick. Smart infrastructure: A research junction. In *IEEE International Smart Cities Conference*, 2021. 8
- [44] M. Hetzel, H. Reichert, K. Doll, and B. Sick. Reliable probabilistic human trajectory prediction for autonomous applications. *IEEE/CVF European Conference On Computer Vision (ECCV)*, 2024. 2, 4, 6, 8, 11



- [45] M. Hetzel, H. Reichert, G. Reitberger, E. Fuchs, K. Doll, and B. Sick. The IMPTC dataset: An infrastructural multi-person trajectory and context dataset. *IEEE Intelligent Vehicles Symposium*, 2023. 7, 8
- [46] M. Hetzel, K. Turacan, H. Reichert, K. Doll, and B. Sick. DD-MDN: Human trajectory forecasting with diffusion-based dual mixture density networks and uncertainty self-calibration. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2026. 2, 4, 5, 6
- [47] Yue Hu, Siheng Chen, Ya Zhang, and Xiao Gu. Collaborative motion prediction via neural motion message passing. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 3
- [48] Y. Huang, H. Bi, and Z. Li. Stgat: Modeling spatial-temporal interactions for human trajectory prediction. *2019 IEEE/CVF International Conference on Computer Vision*, 2019. 2, 3
- [49] B. Inhwan and J. Hae-Gon. Disentangled multi-relational graph convolutional network for pedestrian trajectory prediction. *AAAI Conference on Artificial Intelligence*, 2021. 3
- [50] B. Inhwan and J. Hae-Gon. A set of control points conditioned pedestrian trajectory prediction. *AAAI Conference on Artificial Intelligence*, 2023. 4, 5
- [51] B. Inhwan, O. Jean, and J. Hae-Gon. Eigentrajecory: Low-rank descriptors for multi-modal trajectory forecasting. *IEEE/CVF International Conference on Computer Vision*, 2023. 5, 6
- [52] B. Inhwan, P. Jin-Hwi, and J. Hae-Gon. Learning pedestrian group representations for multi-modal trajectory prediction. *IEEE/CVF European Conference On Computer Vision*, 2022. 2, 5, 6
- [53] B. Inhwan, P. Jin-Hwi, and J. Hae-Gon. Non-probability sampling network for stochastic human trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 5, 6
- [54] B. Inhwan, L. Junoh, and J. Hae-Gon. Can language beat numerical regression? language-based multimodal trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 4, 5
- [55] B. Inhwan, P. Young-Jae, and J. Hae-Gon. Singulartrajecory: Universal trajectory predictor using diffusion model. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2, 4, 11
- [56] T. Izzeddin, K. Salman, and S. Ajmal. Vision-based intention and trajectory prediction in autonomous vehicles: A survey. In *International Joint Conference on Artificial Intelligence*, 2022. 2
- [57] Y. Jiangbei, M. Dinesh, and W. He. Human trajectory prediction via neural social physics. *IEEE/CVF European Conference On Computer Vision*, 2022. 4, 5
- [58] Y. Jiangbei, M. Dinesh, and W. He. Human trajectory forecasting with explainable behavioral uncertainty. *ArXiv*, 2023. 2, 3, 5, 6
- [59] S. Jianhua and L. Yuxuan. Human trajectory prediction with momentary observation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 5, 6, 11
- [60] S. Jianhua and L. Yuxuan. Stimulus verification is a universal and effective sampler in multi-modal human trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 5, 6
- [61] D. Jinghai, W. Le, and L. Chengjiang. Complementary attention gated network for pedestrian trajectory prediction. *AAAI Conference on Artificial Intelligence*, 2022. 4
- [62] L. Junwei, J. Lu, P. Kevin, Y. Ting, and H. Alexander. The garden of forking paths: Towards multi-future trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 7, 8, 9
- [63] G. Kahn, A. Villafior, V. Pong, P. Abbeel, and S. Levine. Uncertainty-aware reinforcement learning for collision avoidance. *ArXiv*, abs/1702.01182, 2017. 1
- [64] M. Karttikeya, G. Harshayu, and A. Shreya. It is not the journey but the destination: Endpoint conditioned trajectory prediction. *IEEE/CVF European Conference On Computer Vision*, 2020. 2, 3, 5
- [65] G. Ke, L. Wenxi, and P. Jia-Yu. End-to-end trajectory distribution prediction based on occupancy grid maps. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2, 5
- [66] V. Kosaraju and A. Sadeghian. Social-bigat: multimodal trajectory forecasting using bicycle-gan and graph attention networks. *International Conference on Neural Information Processing Systems*, 2019. 2, 3
- [67] V. Kress, S. Zernetsch, K. Doll, and B. Sick. Aschaffenburg pose dataset. *Zenodo*, 2021. 8
- [68] NFL: North American Football League. Nfl big data bowl dataset. 2019. 7, 8
- [69] Namhoon Lee, Wongun Choi, Paul Vernaza, Christopher Bongsoo Choy, Philip H. S. Torr, and Manmohan Chandraker. Desire: Distant future prediction in dynamic scenes with interacting agents. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 3
- [70] Seongju Lee, Junseok Lee, Yeonguk Yu, Taeri Kim, and Kyoobin Lee. Mart: Multiscale relational transformer networks for multi-agent trajectory prediction. *IEEE/CVF European Conference On Computer Vision*, 2024. 2, 4, 5
- [71] A. Lerner, Y. Chrysanthou, and D. Lischinski. Crowds by example. *Computer Graphics Forum*, 2007. 1, 8
- [72] Lihuan Li, Maurice Pagnucco, and Yang Song. Graph-based spatial transformer with memory replay for multi-future pedestrian trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 8, 10
- [73] Shijie Li, Yanying Zhou, Jinhui Yi, and Juergen Gall. Spatial-temporal consistency network for low-latency trajectory forecasting. *IEEE/CVF International Conference on Computer Vision*, 2021. 3
- [74] Yuanman Li, Rongqin Liang, Wei Wei, Wei Wang, Jiantao Zhou, and Xia Li. Temporal pyramid network with spatial-temporal attention for pedestrian trajectory prediction. *IEEE Transactions on Network Science and Engineering*, 2022. 4
- [75] Junwei Liang, Lu Jiang, Juan Carlos Niebles, Alexander G. Hauptmann, and Li Fei-Fei. Peeking into the future: Predicting future person activities and locations in videos.

- IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 2, 3
- [76] Rongqin Liang, Yuanman Li, Xia Li, Yi Tang, Jiantao Zhou, and Wenbin Zou. Temporal pyramid network for pedestrian trajectory prediction with multi-supervision. In *AAAI Conference on Artificial Intelligence*, 2020. 3
- [77] F. Liangji, J. Qin hong, and S. Jianping. Tpnnet: Trajectory proposal network for motion prediction. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 3
- [78] Yuejiang Liu, Qi Yan, and Alexandre Alahi. Social nce: Contrastive learning of socially-aware motion representations. *IEEE/CVF International Conference on Computer Vision*, 2020. 3
- [79] S. Liushuai and W. Le. Trajectory unified transformer for pedestrian trajectory prediction. *IEEE/CVF International Conference on Computer Vision*, 2023. 2, 4, 5, 6
- [80] S. Liushuai, W. Le, and L. Chengjiang. Sgcnn: Sparse graph convolution network for pedestrian trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 2, 3
- [81] S. Liushuai, W. Le, and L. Chengjiang. Social interpretable tree for pedestrian trajectory prediction. *AAAI Conference on Artificial Intelligence*, 2022. 4, 5
- [82] S. Liushuai, W. Le, and Z. Sanpin. Trajectory unified transformer for pedestrian trajectory prediction. *IEEE/CVF International Conference on Computer Vision*, 2023. 7, 9
- [83] T. Maeda and N. Ukita. Fast inference and update of probabilistic density estimation on trajectory prediction. *IEEE/CVF International Conference On Computer Vision*, 2023. 2, 4, 5
- [84] S. Malla and C. Choi. Social-stage: Spatio-temporal multi-modal future trajectory forecast. *IEEE International Conference on Robotics and Automation (ICRA)*, 2021. 2, 3, 6, 8, 10
- [85] K. Mangalam, Y. An, H. Girase, and J. Malik. From goals, waypoints & paths to long term human trajectory forecasting. *IEEE/CVF International Conference On Computer Vision*, 2021. 2, 3, 5
- [86] Huynh Trung Manh and Gita Alaghband. Scene- lstm: A model for human trajectory prediction. *ArXiv*, 2018. 2
- [87] Jiageng Mao, Minzhe Niu, Chenhan Jiang, Hanxue Liang, and Jingheng Chen. One million scenes for autonomous driving: Once dataset. In *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Datasets and Benchmarks Track*, 2021. 8
- [88] W. Mao, C. Xu, Q. Zhu, S. Chen, and Y. Wang. Leapfrog diffusion model for stochastic trajectory prediction. *IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 2023. 2, 4, 5, 7
- [89] Francesco Marchetti, Federico Becattini, Lorenzo Seidenari, and A. Bimbo. Mantra: Memory augmented networks for multiple trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 3
- [90] Mancheng Meng, Ziyang Wu, Terrence Chen, Xiran Cai, Xiang Sean Zhou, Fan Yang, and Dinggang Shen. Forecasting human trajectory from scene history. *International Conference on Neural Information Processing Systems*, 2022. 4
- [91] L. Mihee, S. Samuel, and M. Seonghyeon. Muse-vae: Multi-scale vae for environment-aware long term trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 5
- [92] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. *IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 2020. 2, 3
- [93] A. Mohamed, D. Zhu, W. Vu, M. Elhoseiny, and C. Claudel. Social-implicit: Rethinking trajectory prediction evaluation and the effectiveness of implicit maximum likelihood estimation. *IEEE/CVF European Conference On Computer Vision*, 2022. 4, 8, 10
- [94] Alessio Monti, Angelo Porrello, Simone Calderara, Pasquale Coscia, Lamberto Ballan, and Rita Cucchiara. How many observations are enough? knowledge distillation for trajectory forecasting. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 3, 4, 11
- [95] A. Nayak and A. Eskandarian. Uncertainty estimation of pedestrian future trajectory using bayesian approximation. *IEEE Open Journal of Intelligent Transportation Systems*, 3:617–630, 2022. 1
- [96] N. Nikhil and T. Morris. Convolutional neural network for trajectory prediction. *IEEE/CVF European Conference On Computer Vision Workshops*, 2019. 3
- [97] B. Pang, T. Zhao, and X. Xie. Trajectory prediction with latent belief energy-based model. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 3, 5
- [98] D. Patrick, E. Sven, and L. Laura. Mg-gan: A multi-generator model preventing out-of-distribution samples in pedestrian trajectory prediction. *2021 IEEE/CVF International Conference on Computer Vision*, 2021. 2, 3, 5, 8, 10
- [99] X. Pei and K. Jean-Bernard, H. and Ioannis. Social-vae: Human trajectory prediction using timewise latents. *IEEE/CVF European Conference On Computer Vision*, 2022. 4, 5, 7
- [100] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool. You’ll never walk alone: Modeling social behavior for multi-target tracking. *IEEE/CVF International Conference on Computer Vision*, 2009. 1, 7, 8, 9
- [101] Mozhgan Pourkeshavarz, Junrui Zhang, and Amir Rasouli. Dyset: a dynamic masked self-distillation approach for robust trajectory prediction. *IEEE/CVF European Conference On Computer Vision*, 2024. 4
- [102] G. Prasenjit, E. Azim, K. Young-Keun, and M. Goodarz. State estimation and motion prediction of vehicles and vulnerable road users for cooperative autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 2022. 2
- [103] Ruiqi Qiu, Jun Gong, Xinyu Zhang, Siqi Luo, Bowen Zhang, and Yi Cen. Adapting to observation length of trajectory prediction via contrastive learning. In *IEEE/CVF*

- Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 4, 5
- [104] H. Renhao, X. Hao, and P. Maurice. Multimodal trajectory prediction: A survey. *ArXiv*, abs/2302.10463, 2023. 2
  - [105] A. Robicquet and A. Sadeghian. Learning social etiquette: Human trajectory understanding in crowded scenes. *IEEE/CVF European Conference On Computer Vision*, 2016. 1, 4, 8
  - [106] L. Rongqin, L. Yuanman, Z. Jiantao, and L. Xia. Stglow: A flow-based generative framework with dual graphormer for pedestrian trajectory prediction. *IEEE Transactions on neural networks and learning systems*, 2022. 2, 4
  - [107] A. Sadeghian, V. Kosaraju, and A. Sadeghian. Sophie: An attentive gan for predicting paths compliant to social and physical constraints. *IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 2018. 2, 3, 5
  - [108] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. *IEEE/CVF European Conference On Computer Vision*, 2020. 2, 3, 5, 8
  - [109] S. Samuel, L. Mihee, and M. Seonghyeon. A2x: An agent and environment interaction benchmark for multi-modal human trajectory prediction. *Proceedings of the 14th ACM SIGGRAPH Conference on Motion, Interaction and Games*, 2021. 7, 9
  - [110] C. Schoeller and V. Aravatinos. What the constant velocity model can teach us about pedestrian motion prediction. *IEEE Robotics and Automation Letters Vol. 5, No. 2*, 2020. 3, 9, 11, 12
  - [111] C. Schöller and A. Knoll. Flomo: Tractable motion prediction with normalizing flows. *IEEE/RSJ International Conference On Intelligent Robots And Systems*, 2021. 2, 3, 5
  - [112] David W. Scott. Multivariate density estimation and visualization. Wiley, 1992. 8, 10
  - [113] Nasim Shafiee, Taşkın Padır, and Ehsan Elhamifar. Introvert: Human trajectory prediction via conditional 3d attention. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 3
  - [114] W. Shao, J. Xu, Z. Cao, H. Wang, and J. Li. Uncertainty-aware prediction and application in planning for autonomous driving: Definitions, methods, and comparison. *ArXiv*, abs/2403.02297, 2024. 1
  - [115] Xiaodan Shi, Xiaowei Shao, Guangming Wu, Haoran Zhang, Zhiling Guo, Renhe Jiang, and Ryosuke Shibasaki. Social-dpf: Socially acceptable distribution prediction of futures. *AAAI Conference on Artificial Intelligence*, 2021. 2, 3
  - [116] Y. Shuai, L. Hongsheng, and W. Xiaogang. Understanding pedestrian behaviors from stationary crowd groups. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2015. 8
  - [117] Z. Simone, T. Zekarias, and G. Sarunas. Pedestrian trajectory prediction with convolutional neural networks. *Pattern Recognition*, 2022. 4
  - [118] Das Sourav and Ballan Lamberto. Distilling knowledge for short-to-long term trajectory prediction. *IEEE International Conference on Intelligent Robots and Systems*, 2023. 5
  - [119] Stats Perform SportVU. Nba sportvu dataset. 2017. 7, 8
  - [120] ssf019. Fixed results for eth/ucy, 2021. Accessed: 2025-05-02. 6
  - [121] Z. Stefan and R. Hannes. Trajectory forecasts with uncertainties of vulnerable road users by means of neural networks. *IEEE Intelligent Vehicles Symposium*, 2019. 10
  - [122] Z. Stefan and R. Hannes. A holistic view on probabilistic trajectory forecasting – case study. cyclist intention detection. *IEEE Intelligent Vehicles Symposium*, 2022. 10
  - [123] Hao Sun, Zhiquan Zhao, and Zhihai He. Reciprocal learning networks for human trajectory prediction. *IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 2020. 3
  - [124] Jianhua Sun, Qinhong Jiang, and Cewu Lu. Recursive social behavior graph for trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 2
  - [125] J. Sun, Y. Li, H. Fang, and C. Lu. Three steps to multi-modal trajectory prediction: Modality clustering, classification and synthesis. *IEEE/CVF International Conference On Computer Vision*, 2021. 3, 5
  - [126] P. Sun and H. Kretzschmar. Scalability in perception for autonomous driving: Waymo open dataset. *IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 2020. 7, 8, 9
  - [127] W. Theodor and O. Joel. Mtp-go: Graph-based probabilistic multi-agent trajectory prediction with neural odes. *IEEE Transactions on Intelligent Vehicles*, 2023. 8, 10
  - [128] Li-Wu Tsao, Yan-Kai Wang, Hao-Siang Lin, Hong-Han Shuai, Lai-Kuan Wong, and Wen-Huang Cheng. Socialssl: Self-supervised cross-sequence representation learning based on transformers for multi-agent trajectory prediction. *European Conference on Computer Vision*, 2022. 4
  - [129] N. Uhlemann, F. Fent, and M. Lienkamp. Evaluating pedestrian trajectory prediction methods with respect to autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 2023. 11
  - [130] Anirudh Vemula, Katharina Muelling, and Jean Oh. Social attention: Modeling attention in human crowds. *IEEE International Conference on Robotics and Automation*, 2018. 3
  - [131] Chuhua Wang, Yuchen Wang, Mingze Xu, and David J. Crandall. Stepwise goal-driven networks for trajectory prediction. *IEEE Robotics and Automation Letters*, 2021. 4
  - [132] Dafeng Wang, Hongbo Liu, Naiyao Wang, Yiyang Wang, Hua Wang, and Seán F. McLoone. Seem: A sequence entropy energy-based model for pedestrian trajectory all-then-one prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 2, 4
  - [133] X. Wei, Y. Haoteng, and W. He. Socialvae: Predicting pedestrian trajectory via interaction conditioned latents. *AAAI Conference on Artificial Intelligence*, 2024. 4, 5
  - [134] Beihao Xia, Conghao Wong, Qinmu Peng, Wei Yuan, and Xinge You. Cscnet: Contextual semantic consistency network for trajectory prediction in crowded spaces. *Pattern Recognit.*, 2022. 4

- [135] H. Xinyu and C. Xinjing. The apolloscape dataset for autonomous driving. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2018. 8
- [136] C. Xu, M. Li, Z. Ni, Y. Zhang, and S. Chen. Groupnet: Multiscale hypergraph neural networks for trajectory prediction with relational reasoning. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2, 5, 6
- [137] C. Xu, W. Mao, W. Zhang, and S. Chen. Remember intentions: Retrospective-memory-based trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 4, 5, 7
- [138] Yanchao Xu, Wenbo Shao, Jun Li, Kai Yang, Weida Wang, Hua Huang, Chen Lv, and Hong Wang. Sind: A drone dataset at signalized intersection in china. In *Proceedings of the 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pages 2471–2478, 2022. 8
- [139] Hao Xue, Du Q. Huynh, and Mark Reynolds. Ss-lstm: A hierarchical lstm model for pedestrian trajectory prediction. *IEEE Winter Conference on Applications of Computer Vision*, 2018. 2
- [140] Kota Yamaguchi, Alexander C. Berg, Luis E. Ortiz, and Tamara L. Berg. Who are you with and where are you going? *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2011. 3
- [141] Zhongcun Yan, Jianqing Li, Peng Hang, and Jian Sun. Onsitevru: A high-resolution trajectory dataset for high-density vulnerable road users. In *arXiv preprint arXiv:2503.23365*, 2025. 8
- [142] Biao Yang, Guocheng Yan, Pin Wang, Ching yao Chan, Xiaofeng Liu, and Yang Chen. A novel graph-based trajectory predictor with pseudo-oracle. *IEEE Transactions on Neural Networks and Learning Systems*, 2020. 3
- [143] Y. Ye and K. Kris. Diverse trajectory forecasting with determinantal point processes. *IEEE International Conference on Learning Representations*, 2019. 7, 9
- [144] Y. Ye and K. Kris. Diverse trajectory forecasting with determinantal point processes. *IEEE International Conference on Learning Representations*, 2020. 5, 6
- [145] M. Yecheng and I. Jeevana. Likelihood-based diverse sampling for trajectory forecasting. *IEEE/CVF International Conference on Computer Vision*, 2021. 5, 6
- [146] Haibao Yu, Wenxian Yang, Hongzhi Ruan, Zhenwei Yang, Yingjuan Tang, Xu Gao, Xin Hao, Yifeng Shi, Yifeng Pan, Ning Sun, Juan Song, Jirui Yuan, Ping Luo, and Zaiqing Nie. V2x-seq: A large-scale sequential dataset for vehicle–infrastructure cooperative perception and forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 8
- [147] Y. Yuan, X. Weng, Y. Ou, and K. Kitani. Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting. *IEEE/CVF International Conference on Computer Vision*, 2021. 2, 3, 5
- [148] C. Yuxiao, I. Boris, and P. Marco. Scept: Scene-consistent, policy-based trajectory predictions for planning. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 4
- [149] W. Yuxuan, W. Le, and Z. Sanping. Multi-stream representation learning for pedestrian trajectory prediction. *AAAI Conference on Artificial Intelligence*, 2023. 4, 5
- [150] S. Zernetsch, V. Kress, K. Doll, and B. Sick. Cyclist actions: Optical flow sequences and trajectories. *Zenodo*, 2020. 8
- [151] S. Zernetsch, H. Reichert, V. Kress, K. Doll, and B. Sick. A holistic view on probabilistic trajectory forecasting - case study. cyclist intention detection. *IEEE Intelligent Vehicles Symposium*, 2022. 2, 8
- [152] W. Zhan, L. Sun, and D. Wang. INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps. *arXiv:1910.03088 [cs, eess]*, 2019. 8
- [153] Pu Zhang, Wanli Ouyang, Pengfei Zhang, Jianru Xue, and Nanning Zheng. Sr-lstm: State refinement for lstm towards pedestrian trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 3
- [154] H. Zhao and R. Wildes. Where are you heading? dynamic trajectory prediction with expert goal examples. *IEEE/CVF International Conference on Computer Vision*, 2021. 3, 5
- [155] Tianyang Zhao, Yifei Xu, Mathew Monfort, Wongun Choi, Chris Baker, Yibiao Zhao, Yizhou Wang, and Ying Nian Wu. Multi-agent tensor fusion for contextual trajectory prediction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 2, 3
- [156] F. Zheng, K. Kun, and X. Chuchu. Summary and reflections on pedestrian trajectory prediction in the field of autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 2024. 2