NATRA: Noise-Agnostic Framework for Trajectory Prediction with Noisy Observations

Rongqing Li[1] Changsheng Li[1‡] Ruilin Lv[1] Yuhang Li[1] Yang Gao[2]
Xiaolu Zhang[3] JUN ZHOU[3]
[1]Beijing Institute of Technology [2]Meituan [3]Ant Group
lirongqing99@gmail.com lcs@bit.edu.cn 3220231454@bit.edu.cn 596983629@qq.com
gaoyang94@meituan.com yueyin.zxl@antfin.com jun.zhoujun@antfin.com

Abstract

*Trajectory prediction aims to forecast an agent's future trajectories based on its historical observed trajectories, which is a critical task for various applications such as autonomous driving, robotics, and surveillance systems. Most existing trajectory prediction methods assume that the observed trajectories collected for forecasting are clean. However, in real-world scenarios, noise is inevitably introduced into the observations, resulting in the collapse of the existing approaches. Therefore, it is essential to perform robust trajectory prediction based on noisy observations, which is a more practical scenario. In this paper, we propose **NATRA**, a **N**oise-**A**gnostic framework capable of tackling the problem of **TRA**jectory prediction with arbitrary types of noisy observations. Specifically, we put forward a mutual information-based mechanism to denoise the original noisy observations. It optimizes the produced trajectories to exhibit a pattern that closely resembles the clean trajectory pattern while deviating from the noisy one. Considering that the trajectory structure may be destroyed through the only optimization of mutual information, we introduce an additional reconstruction loss to preserve the structure information of the produced observed trajectories. Moreover, we further propose a ranking loss to further enhance the performance. Because NATRA does not rely on any specific module tailored to particular noise distributions, it can handle arbitrary types of noise in principle. Additionally, our proposed NATRA can be easily integrated into existing trajectory prediction models. Extensive experiments on both synthetic and real-world noisy datasets demonstrate the effectiveness of our method.*

‡ Corresponding Author

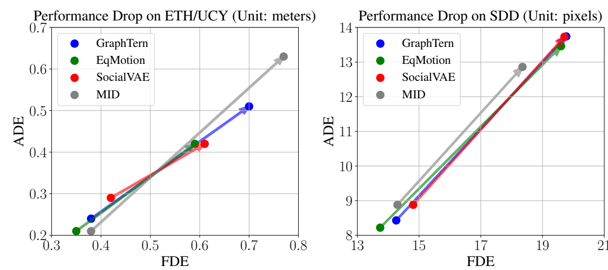

Figure 1. Performance drop of various trajectory prediction methods, including GraphTern [3], EqMotion [87], SocialVAE [88] and MID [21], on the ETH/UCY [33, 53] and Stanford Drone datasets (SDD) [57]. The start of each arrow indicates the performance under clean observations, while the end represents the degraded performance under noisy observations (We add Gaussian noise $\mathcal{N}(0, \sigma = 0.4)$ to the clean observations.). Best viewed in color.

1. Introduction

The objective of trajectory prediction is to anticipate the future trajectories for agents given their past observed trajectories, which is an essential and emerging task in numerous applications, such as autonomous driving [22, 54, 55, 79, 105, 106], drones [15], surveillance systems [76], and robotics [25, 58]. In recent years, trajectory prediction has garnered significant attention in the computer vision and machine learning communities, with numerous methods proposed [4, 5, 11, 12, 14, 81]. Among these methods, they typically assume the observed historical trajectories are clean, and leverage them to predict future trajectories. Recent advances have demonstrated promising performance in trajectory prediction by learning from such clean observed trajectory data.

However, in real-world scenarios, the observed trajectories are inevitably noisy, often caused by adverse weather conditions or occlusions [72, 73, 99], which significantly hinder the performance of existing trajectory prediction methods. To substantiate this point, we introduce noise into the historical trajectories of both the training and testing sets of the ETH/UCY [33, 53] and Stanford Drone (SDD) [57] datasets. We then conduct experiments on several recently proposed trajectory prediction approaches. We compare the performance before and after introducing noise to the observations. As shown in Figure 1, the presence of noise in the observed trajectories leads to a significant performance drop for various trajectory prediction methods on both two datasets. This significant performance degradation highlights the detrimental impact of noise on trajectory prediction accuracy, even for state-of-the-art models. Therefore, it is crucial to devise a robust method for predicting future trajectories based on noisy observations.

In this paper, we propose **NATRA**, a noise-agnostic method designed to address the challenge of trajectory prediction with arbitrary types of noisy observations. Specifically, we first propose a mutual information-based mechanism to filter noise from the original observations. This mechanism ensures the produced trajectories closely resemble the patterns of noise-free trajectories while deviating from the noisy patterns. To this end, we maximize the mutual information between the produced trajectories and the clean future trajectories (i.e., ground-truth), while simultaneously minimizing the mutual information between the produced trajectories and the original noisy observations. In this way, the produced trajectories are forced to collate information from both the noisy trajectories and clean future trajectories, thereby preserving the necessary information while filtering out noise. However, solely relying on optimizing mutual information for denoising may disrupt the structure of the trajectory. Therefore, we propose to randomly mask several observations and attempt to reconstruct the masked locations. By jointly optimizing the mutual information and reconstruction losses, the trajectory denoise model can effectively eliminate noise while preserving the structure information of the trajectory. In the meantime, we design a ranking loss to facilitate the ability of the trajectory prediction module based on an intuitive thought: predictions using the produced denoised observations will be superior to those using noisy observations. It is noteworthy that the ranking loss optimizes not only the trajectory prediction module but also the denoising module, which can further assist in filtering noise to some extent. Since NATRA does not rely on any specific module tailored to a particular noise distribution, it can handle arbitrary noise in principle. Essentially, our proposed NATRA is a plug-and-play approach that is compatible with existing trajectory prediction models, enabling them to gracefully handle cases with noisy observations.

Our main contributions are summarized as follows: 1) We investigate a new problem setting for trajectory prediction with noisy observations, addressing a more practical scenario. To tackle this, we propose a noise-agnostic, plug-and-play approach called NATRA. 2) We design a denoising module that incorporates a mutual information-based loss along with a reconstruction loss, effectively denoising observed trajectories while preserving their structural information. 3) We propose a ranking loss to ensure that denoised observations yield superior future predictions compared to their noisy counterparts, thereby enhancing the accuracy of trajectory predictions. 4) We conduct extensive experiments on the ETH/UCY and SDD datasets, demonstrating that our method significantly outperforms the baselines in predicting trajectory with noisy observations.

2. Related Works

2.1. Trajectory Prediction with Clean Observations

Trajectory prediction has been an active area of research in the computer vision and machine learning communities. Early works employ physics-based methods to model the trajectories of agents [44, 53]. Subsequently, learning-based approaches are proposed, which significantly enhance the performance of trajectory prediction [59, 78, 85, 91, 107]. They model trajectory temporal information and the interaction between agents [1, 2, 50, 65, 93]. One representative approach is social pooling, which aggregates hidden state information of neighbors within a spatial grid [23, 61]. Additionally, attention mechanisms [18, 77], graph neural networks [29, 31, 39, 69] and transformers [66, 96, 97, 108] have also been exploited to model interactions among agents. To further enhance prediction performance, researchers delve into incorporating the map information. Works such as [16, 46, 48, 63, 68] encode RGB scene information, while [20, 28, 74, 94, 103] incorporate lane and road traffic information. Moreover, due to the inherent uncertainty associated with agents, researchers have proposed a series of models to predict multiple plausible future trajectories, including GANs [40, 43, 104], VAEs [34, 35, 70], and diffusion models [6, 26, 38, 56]. Recently, several new task settings have been introduced to address more practical trajectory prediction problems, including trajectory prediction with varying observations [37, 38, 52, 71, 89], long-tailed distribution in trajectory prediction [49, 80, 98], and distribution shift in trajectory prediction [30, 42, 67, 90].

Despite these methods having shown promising performance, they rely on sufficiently clean observed trajectories. As aforementioned, when the observed trajectories are corrupted by noise, the model performance severely deteriorates. In contrast to these approaches, we attempt to tackle the problem of predicting future trajectories with noisy observed trajectories.

2.2. Robust Trajectory Prediction

Robust trajectory prediction aims to achieve accurate future trajectory forecasting with noisy input. Existing approaches primarily take detection results as inputs, and attribute noise to imperfections in tracking modules [82, 83, 92, 95, 101]. For instance, MTP [83] proposes a multi-hypothesis framework to generate multiple potential tracklets and predict diverse future trajectory predictions, thereby mitigating error propagation. Similarly, the work in [82] mitigates noise sensitivity by directly utilizing raw detections and their inter-frame affinity matrices as inputs to trajectory prediction models, bypassing the error-prone data association process. Additionally, recent efforts have also addressed adversarial noise in historical trajectories induced by adversarial attacks [8, 9, 27, 60, 100, 102].

While these methods have proven effective in enhancing robustness and prediction accuracy, they are typically limited to handling specific noise types, which restricts their practical applicability. In contrast, our method is designed to directly eliminate noise in trajectory locations, thereby improving both the robustness and accuracy of trajectory prediction models.

3. Methods

3.1. Problem Formulation

Let $X_{obs} = \{x_{obs}^1, x_{obs}^2, \ldots, x_{obs}^{T_{obs}}\}$ denote the observed trajectories, where T_{obs} is the observation length, and $x_{obs}^i \in \mathbb{R}^2$ is the i^{th} location. We assume the observations are given by $X_{obs} = f_N(S_{obs})$, where $S_{obs} = \{s_{obs}^1, s_{obs}^2, \ldots, s_{obs}^{T_{obs}}\}$ are clean observed trajectories, and f_N is a noise function that introduces arbitrary types of noise, such as Gaussian and Poisson noise, to the clean samples. Moreover, we denote the ground-truth future trajectories as $Y_{fut} = \{y_{fut}^1, y_{fut}^2, \ldots, y_{fut}^{T_{fut}}\}$, where $y_{fut}^i \in \mathbb{R}^2$ represents the i^{th} locations, and T_{fut} is the length of the ground-truth future trajectories. In this work, we define the problem by assuming Y_{fut} are clean, and only the observed historical trajectory is noisy[1] . Different from previous works that typically utilize clean observations S_{obs} for future trajectory prediction, our goal is to develop a robust trajectory prediction method using noisy observations, which is a more practical scenario. Specifically, we aim to use noisy observations X_{obs} to forecast K plausible future trajectories $\{\hat{Y}_{fut}\}_{k=1}^{K}$ under the supervision of Y_{fut}.

3.2. Overall Framework

The overall framework of the proposed NATRA is shown in Figure 2. Our framework consists of two parts: a Trajectory Denoise Model (TDM) and a Trajectory Prediction Backbone (TPB). To eliminate noise from the noisy observations X_{obs}, we first propose a mutual information-based mechanism, which encourages the produced trajectories $\hat{X}_{obs}$ exhibit patterns similar to the noise-free ground-truth future trajectories Y_{fut}, while deviating from the patterns of the noisy observed trajectories X_{obs}. This is achieved through a loss function $\mathcal{L}_{MI}$ that simultaneously maximizes the mutual information between $\hat{X}_{obs}$ and Y_{fut} while minimizing the mutual information between $\hat{X}_{obs}$ and X_{obs}. In this way, the produced trajectories $\hat{X}_{obs}$ are forced to collate information from both the noisy trajectories X_{obs} and clean future trajectories Y_{fut}, thereby filtering out noise. Given that only optimizing mutual information-based loss potentially disrupts the structure of the trajectories, we propose a reconstruction strategy to mitigate this issue. Specifically, we randomly mask M locations of the noisy observations X_{obs} to obtain X_{obs}^{mask}. The X_{obs}^{mask} is then fed into the TDM to reconstruct the masked portion of the original noisy input X_{obs} using $\mathcal{L}_{rec}$. By jointly optimizing $\mathcal{L}_{MI}$ and $\mathcal{L}_{rec}$, the TDM is able to learn to denoise while preserving the structure information of the trajectories. To facilitate more accurate trajectory prediction, we devise a ranking loss. We first input both the denoised observations $\hat{X}_{obs}$ and the original noisy observations X_{obs} into the TPB to forecast future trajectories. Then the ranking loss is applied to encourage the future trajectories predicted from the denoised observations to be more precise than those predicted from the noisy observations, thereby enhancing the trajectory prediction performance. The TDM and TPB modules can benefit from each other: the ranking loss in TPB helps TDM filter noise more effectively, while the denoised trajectories generated by TDM enable TPB to predict future trajectories more accurately. As NATRA is not dependent on any module specialized for a particular noise distribution, it is capable of handling arbitrary noise in principle. In addition, NATRA is essentially a plug-and-play approach and can be readily integrated into existing trajectory prediction models, enabling them to effectively handle scenarios with noisy observations.

3.3. Trajectory Prediction with Noisy Observations

In this section, we introduce the details of our NATRA. We first present the mutual information-based denoising mechanism, followed by the designed ranking loss.

3.3.1. Mutual Information-Based Denoising Mechanism.

Given noisy observations X_{obs}, we expect the noise can be eliminated through a trajectory denoising model $\Phi_{\mathbf{TDM}}$. Inspired by Information Bottleneck [75], we encourage the produced trajectories to exhibit patterns closely resembling noise-free trajectory patterns while deviating from noisy patterns. This is achieved by maximizing the mutual information between the produced trajectories and noise-free ground-truth future trajectories Y_{fut} while minimizing the mutual information between the produced trajectories and

[1]Most datasets, such as Argoverse [10], undergo post-processing, so we can consider the trajectories in the dataset, including historical and future trajectories, as clean. Nevertheless, we can manually introduce noise into the historical trajectories.

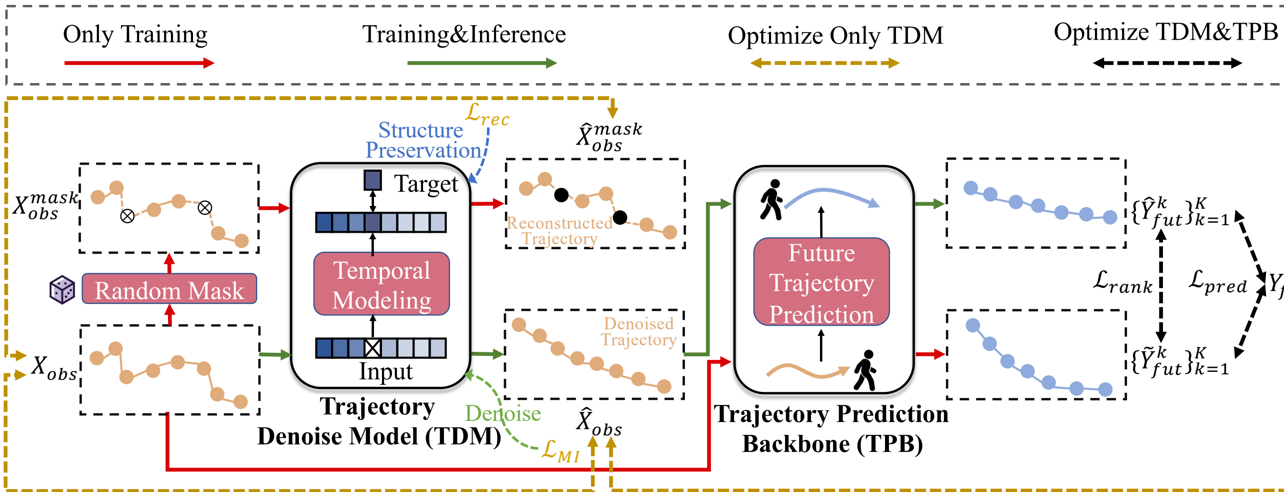

Figure 2. Overview of NATRA framework. It is composed of two modules: a Trajectory Denoise Model (TDM) and a Trajectory Prediction Backbone (TPB). The $\mathcal{L}_{MI}$ denoises the produced trajectories $\hat{X}_{obs}$ by maximizing mutual information between $\hat{X}_{obs}$ and the clean ground-truth Y_{fut}, while minimizing the mutual information between $\hat{X}_{obs}$ and noisy observations X_{obs}. The $\mathcal{L}_{rec}$ reconstructs the masked location of the original trajectories. By jointly optimizing $\mathcal{L}_{rec}$ and $\mathcal{L}_{MI}$, the trajectory denoise model can learn to denoise trajectories while preserving the structure information. The ranking loss $\mathcal{L}_{rank}$ constrains the predictions based on the denoised observations $\hat{X}_{obs}$ to be superior to those based on the noisy observations X_{obs}, thereby further filtering out noise and enhance the ability of trajectory prediction.

the original noisy observations X_{obs}. We define the objective function as:

$$J_{MI} = \min_{\hat{X}_{obs}} \ \alpha I(X_{obs}; \hat{X}_{obs}) - I(\hat{X}_{obs}; Y_{fut}), \quad (1)$$

where $I(\cdot\ ;\ \cdot)$ calculates the mutual information of two variables, and α is a trade-off parameter.

However, directly calculating J_{MI} is intractable. Therefore, we estimate the upper bound of $I(X_{obs}; \hat{X}_{obs})$ by utilizing CLUB [13], and the lower bound of $I(\hat{X}_{obs}; Y_{fut})$ by leveraging the method described in MINE [7]. We first calculate the upper bound of $I(X_{obs}; \hat{X}_{obs})$.

Theorem 3.1. *Given two random variables x and y, the mutual information $I(x; y)$ has the following upper bound*

$$I(x; y) \ \leq \ \mathbb{E}_{p(x,y)}[\log p(y|x)] \ - \ \mathbb{E}_{p(x)}\mathbb{E}_{p(y)}[\log p(y|x)]. \quad (2)$$

Proof. See proof in the Appendix 6.7.

By substituting X_{obs} and $\hat{X}_{obs}$ to the Equation (2), we can obtain the upper bound of $I(X_{obs}; \hat{X}_{obs})$:

$$I(X_{obs}; \hat{X}_{obs}) \leq \mathbb{E}_{p(X_{obs}, \hat{X}_{obs})}[\log p(\hat{X}_{obs}|X_{obs})] \\ - \mathbb{E}_{p(X_{obs})}\mathbb{E}_{p(\hat{X}_{obs})}[\log p(\hat{X}_{obs}|X_{obs})]. \quad (3)$$

Since $p(\hat{X}_{obs}|X_{obs})$ is unknown, we introduce a variational approximation distribution $q_\phi(\hat{X}_{obs}|X_{obs})$ to approximate $p(\hat{X}_{obs}|X_{obs})$ with parameter ϕ, following [13]. Thus, the upper bound can be written as:

$$I(X_{obs}; \hat{X}_{obs}) \leq I_\mu(X_{obs}; \hat{X}_{obs}) \\ = \mathbb{E}_{p(X_{obs}, \hat{X}_{obs})}[\log q_\phi(\hat{X}_{obs}|X_{obs})] \\ - \mathbb{E}_{p(X_{obs})}\mathbb{E}_{p(\hat{X}_{obs})}[\log q_\phi(\hat{X}_{obs}|X_{obs})]. \quad (4)$$

Theorem 3.2 (Donsker-Varadhan representation [17]). *Given two probability distributions $\mathbb{P}$, $\mathbb{Q}$. The Kullback Liebler Divergence admits the following dual representation:*

$$D_{KL}(\mathbb{P}||\mathbb{Q}) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log \mathbb{E}_{\mathbb{Q}}[e^T], \quad (5)$$

Proof. See the proof in the Appendix 6.8.

Next, we calculate the lower bound of the mutual information $I(\hat{X}_{obs}; Y_{fut})$.

Based on the Theorem 3.2, we can obtain the $I(\hat{X}_{obs}, Y_{fut})$ by:

$$I(\hat{X}_{obs}, Y_{fut}) = D_{KL}(p(\hat{X}_{obs}, Y_{fut})||p(\hat{X}_{obs})p(Y_{fut})) \\ = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{p(\hat{X}_{obs}, Y_{fut})}[T] - \log \mathbb{E}_{p(\hat{X}_{obs})p(Y_{fut})}[e^T], \quad (6)$$

where $\Omega = \hat{X}_{obs} \times Y_{fut}$ is the input space. Let $\mathcal{F}$ be any class of functions $T : \Omega \to \mathbb{R}$, and the lower bound of $I(\hat{X}_{obs}, Y_{fut})$ can be expressed as:

$$I(\hat{X}_{obs}, Y_{fut}) \geq I_{\mathcal{F}}(\hat{X}_{obs}, Y_{fut}) \\ = \sup_{T\in\mathcal{F}} \mathbb{E}_{p(\hat{X}_{obs}, Y_{fut})}[T] - \log \mathbb{E}_{p(\hat{X}_{obs})p(Y_{fut})}[e^T]. \quad (7)$$

We choose $\mathcal{F}$ to be the family of functions $T_\psi : \hat{X}_{obs} \times Y_{fut} \to \mathbb{R}$, parameterized by a neural network ψ. Thus, the lower bound can be written as:

$$I(\hat{X}_{obs}, Y_{fut}) \geq I_\psi(\hat{X}_{obs}, Y_{fut}) \\ = \sup_\psi \mathbb{E}_{p(\hat{X}_{obs}, Y_{fut})}[T_\psi] - \log \mathbb{E}_{p(\hat{X}_{obs})p(Y_{fut})}[e^{T_\psi}]. \quad (8)$$

Based on Equation (4) and (8), we derive the upper bound $\mathcal{L}_{MI}$ of the J_{MI} as:

$$J_{MI} \leq \mathcal{L}_{MI} = \alpha I_\mu(X_{obs}; \hat{X}_{obs}) - I_\psi(\hat{X}_{obs}, Y_{fut}) \\ = \alpha \mathbb{E}_{p(X_{obs}, \hat{X}_{obs})}[\log q_\phi(\hat{X}_{obs}|X_{obs})] \\ - \mathbb{E}_{p(X_{obs})}\mathbb{E}_{p(\hat{X}_{obs})}[\log q_\phi(\hat{X}_{obs}|X_{obs})] \\ - \sup_\psi \mathbb{E}_{p(\hat{X}_{obs}, Y_{fut})}[T_\psi] + \log \mathbb{E}_{p(\hat{X}_{obs})p(Y_{fut})}[e^{T_\psi}] \quad (9)$$

By minimizing the upper bound $\mathcal{L}_{MI}$, we can obtain an approximation solution to Equation (1), enabling the trajectory denoise model to learn how to denoise trajectories.

However, only optimizing the mutual information may destroy the structure of the produced trajectories. Therefore, we propose a reconstruction strategy to preserve the structure information. As shown in the left part of Figure 2, we mask locations within the noisy observed trajectories X_{obs} to generate X_{obs}^{mask}, which can be formulated as:

$$X_{obs}^{mask} = X_{obs} \odot \mathcal{M}_{obs}, \quad (10)$$

where $\odot$ represents the element-wise multiplication. $\mathcal{M}_{obs}$ is a 0-1 mask vector, where the value **0** represents the corresponding locations are masked. Subsequently, the X_{obs}^{mask} is fed into the trajectory denoise model $\Phi_{\mathbf{TDM}}$ to produce observed trajectories $\hat{X}_{obs}^{mask}$. To enable TDM to preserve the structural information of the trajectories, we reconstruct the masked locations in the produced trajectories. We define the reconstruction loss as follows:

$$\mathcal{L}_{rec} = \mathcal{J}(\hat{X}_{obs}^{mask} \odot (1 - \mathcal{M}_{obs}), X_{obs} \odot (1 - \mathcal{M}_{obs})), \quad (11)$$

where $\mathcal{J}$ denotes the distance metric, and we empirically adopt L_2 distance in this work. Through optimizing $\mathcal{L}_{rec}$, the trajectory denoise model can learn to preserve the structure information of the observations. By jointly optimizing the reconstruction loss $\mathcal{L}_{rec}$ with $\mathcal{L}_{MI}$, the mutual information-based mechanism effectively denoises the trajectories while preserving their structural information.

3.3.2. Trajectory Prediction Based on Ranking Loss.

After obtaining the denoised observed trajectories $\hat{X}_{obs}$, we design a ranking loss to enhance future trajectory prediction performance. The ranking loss is based on an intuitive thought: leveraging denoised observed trajectories should yield more accurate future predictions compared to using noisy observations. To accomplish this, we first input the denoised observations into the trajectory prediction backbone $\Phi_{\mathbf{TPB}}$. Then, we can predict K plausible future trajectories:

$$\{\hat{Y}_{fut}^k\}_{k=1}^K = \Phi_{\mathbf{TPB}}(\Phi_{\mathbf{TDM}}(X_{obs})), \quad (12)$$

Similarly, we can also predict K possible trajectories based on the noisy observed trajectories:

$$\{\tilde{Y}_{fut}^k\}_{k=1}^K = \Phi_{\mathbf{TPB}}(X_{obs}). \quad (13)$$

After obtaining K possible trajectories based on the denoised and noisy observed trajectories, respectively, we then select the minimal distances $d_{denoise}$ and d_{noise} by calculating the distances between each predicted trajectory and ground-truth trajectory, respectively. Formally,

$$d_{denoise} = \min_{1 \leq k \leq K} ||\hat{Y}_{fut}^k - Y_{fut}||_2, \quad (14)$$

$$d_{noise} = \min_{1 \leq k \leq K} ||\tilde{Y}_{fut}^k - Y_{fut}||_2. \quad (15)$$

Then, we employ the ground-truth future trajectories as supervision for the best-predicted trajectory:

$$\mathcal{L}_{pred} = ||\hat{Y}_{fut}^{best} - Y_{fut}||_2 + ||\tilde{Y}_{fut}^{best} - Y_{fut}||_2, \quad (16)$$

where $best$ represents the trajectory with a minimal distance to the ground-truth. Subsequently, we design a ranking loss to constrain the best prediction $\hat{Y}_{fut}^{best}$ using the denoised observations to be more accurate than that $\tilde{Y}_{fut}^{best}$ using the noisy observations:

$$\mathcal{L}_{rank} = \max(0, d_{denoise} - d_{noise} + \Delta), \quad (17)$$

where Δ is a margin. Since the ranking loss optimizes both the trajectory prediction backbone and the trajectory denoise model, it not only aids in better trajectory prediction but also facilitates the denoising ability of the trajectory denoise model.

3.4. Optimization and Inference

Optimization. We define the total loss function as:

$$\mathcal{L} = \mathcal{L}_{pred} + \beta \mathcal{L}_{rank} + \delta \mathcal{L}_{rec} + \gamma \mathcal{L}_{MI}, \quad (18)$$

where β, δ, and γ are trade-off hyperparameters. The training details are shown in Appendix 6.6.

Inference. After training, the model can be utilized for trajectory prediction based on noisy observations. As shown in the green arrow in Figure 2, we first feed the noisy trajectories X_{obs} into the trajectory denoise model $\Phi_{\mathbf{TDM}}$ to obtain denoised trajectories $\hat{X}_{obs}$. Subsequently, the trajectory prediction backbone $\Phi_{\mathbf{TPB}}$ takes $\hat{X}_{obs}$ as input to obtain the predicted future trajectories $\{\hat{Y}_{fut}^k\}_{k=1}^K$.

Table 1. Comparison of different methods on the ETH/UCY dataset. All methods are trained and tested with noisy observations input. The evaluation metrics are ADE and FDE (Unit: meters). The best results are highlighted in **bold**.

Noise	Method	ETH		HOTEL		UNIV		ZARA1		ZARA2		AVG	
		ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE
$\sigma = 0.2$	GraphTern	0.56	0.81	0.27	0.39	0.39	0.59	0.37	0.58	0.34	0.50	0.39	0.57
	EqMotion	0.51	0.70	0.17	0.24	0.36	0.56	0.33	0.54	0.25	0.37	0.32	0.48
	MID	0.74	0.86	0.37	0.36	0.45	0.58	0.43	0.52	0.40	0.47	0.48	0.56
	SocialImplict	0.67	1.28	0.31	0.49	0.46	0.77	0.39	0.71	0.35	0.63	0.44	0.78
	SocialVAE	0.56	0.89	0.18	0.25	0.40	0.63	0.32	0.49	0.25	0.38	0.34	0.53
	Wavelet+GraphTern	0.50	0.72	0.23	0.35	0.37	0.56	0.32	0.50	0.30	0.42	0.34	0.51
	EMA+GraphTern	0.53	0.76	0.26	0.37	0.38	0.57	0.34	0.54	0.31	0.48	0.36	0.53
	NATRA+GraphTern	**0.48**	**0.68**	**0.19**	**0.26**	**0.35**	**0.53**	**0.28**	**0.44**	**0.27**	**0.36**	**0.31**	**0.45**
	Wavelet+EqMotion	0.48	0.64	**0.16**	0.22	0.32	0.52	0.31	0.48	0.23	0.32	0.30	0.44
	EMA+EqMotion	0.49	0.66	**0.16**	0.22	0.34	0.53	0.31	0.48	0.23	0.33	0.31	0.44
	NATRA+EqMotion	**0.47**	**0.61**	**0.16**	**0.21**	**0.29**	**0.47**	**0.28**	**0.44**	**0.21**	**0.29**	**0.28**	**0.40**
$\sigma = 0.4$	GraphTern	0.67	0.99	0.40	0.51	0.49	0.69	0.51	0.71	0.46	0.61	0.51	0.70
	EqMotion	0.65	0.87	0.25	0.34	0.43	0.64	0.42	0.62	0.34	0.48	0.42	0.59
	MID	0.88	1.08	0.51	0.49	0.60	0.73	0.59	0.84	0.57	0.72	0.63	0.77
	SocialImplict	0.74	1.39	0.40	0.64	0.54	0.86	0.59	0.88	0.53	0.76	0.56	0.61
	SocialVAE	0.67	1.01	0.24	0.32	0.48	0.73	0.41	0.58	0.31	0.43	0.42	0.61
	Wavelet+GraphTern	0.60	0.83	0.33	0.43	0.46	0.64	0.42	0.63	0.37	0.49	0.44	0.60
	EMA+GraphTern	0.64	0.92	0.36	0.46	0.46	0.66	0.47	0.67	0.41	0.58	0.47	0.66
	NATRA+GraphTern	**0.55**	**0.77**	**0.29**	**0.41**	**0.42**	**0.61**	**0.38**	**0.56**	**0.34**	**0.42**	**0.40**	**0.55**
	Wavelet+EqMotion	0.62	0.74	0.23	0.29	0.41	0.58	0.39	0.56	0.32	0.43	0.39	0.52
	EMA+EqMotion	0.63	0.75	0.23	0.29	0.42	0.63	0.40	0.58	0.31	0.42	0.43	0.53
	NATRA+EqMotion	**0.57**	**0.71**	**0.20**	**0.25**	**0.35**	**0.51**	**0.35**	**0.51**	**0.29**	**0.41**	**0.35**	**0.48**

Table 2. Comparison of different methods on the SDD dataset. All methods are trained and tested noisy observed trajectory inputs. The evaluation metrics are ADE and FDE (Unit: pixels). The best results are highlighted in **bold**.

Noise	Method	SDD	
		ADE	FDE
$\sigma = 0.2$	GraphTern	11.67	18.37
	EqMotion	10.62	15.68
	MID	10.26	15.38
	SocialImplicit	15.92	26.82
	SocialVAE	11.67	17.62
	Wavelet+GraphTern	10.52	16.86
	EMA+GraphTern	11.03	17.42
	NATRA+GraphTern	**10.08**	**15.64**
	Wavelet+EqMotion	10.36	15.24
	EMA+EqMotion	10.32	15.22
	NATRA+EqMotion	**10.06**	**14.67**

Noise	Method	SDD	
		ADE	FDE
$\sigma = 0.4$	GraphTern	13.74	19.77
	EqMotion	13.46	19.60
	MID	12.86	18.35
	SocialImplicit	18.07	29.98
	SocialVAE	13.73	19.71
	Wavelet+GraphTern	12.98	18.50
	EMA+GraphTern	13.21	19.11
	NATRA+GraphTern	**12.35**	**17.28**
	Wavelet+EqMotion	12.38	18.25
	EMA+EqMotion	12.79	18.64
	NATRA+EqMotion	**11.92**	**17.65**

4. Experiments

4.1. Experiment Settings

Backbones and Compared Baselines. To validate the efficacy of NATRA, we integrate it into three popular trajectory prediction backbones GraphTern [3], EqMotion [87] and HiVT [105]. We compare our method against five state-of-the-art trajectory prediction models, including **GraphTern**, **EqMotion**, HiVT, **MID** [21], **SocialImplicit** [51], and **SocialVAE** [88]. These methods take original noisy observations as input to predict future trajectories. Considering there are few works focusing on trajectory prediction with noisy observed trajectory inputs, we establish two trajectory denoising baselines by integrating Wavelet and EMA with the trajectory prediction backbones, respectively, for a more comprehensive comparison. The **Wavelet** utilizes the wavelet transform to decompose the signal into multiple scales, obtaining wavelet coefficients of different fre-

Table 3. Comparison of different methods on the Argoverse dataset. All methods are trained and tested noisy observed trajectory inputs.

Noise	Method	Argoverse			
		K=1		K=6	
		ADE	FDE	ADE	FDE
$\sigma = 4$	HiVT	3.16	6.22	0.93	1.48
	EMA+HiVT	3.07	6.29	0.91	1.47
	Wavelet+HiVT	3.09	6.06	0.90	1.42
	NATRA+HiVT	**2.78**	**5.69**	**0.87**	**1.37**

quencies. Then the noise in high-frequency coefficients is removed by the thresholding method. The **EMA** smooths the current trajectory location by taking an exponentially weighted average of the current location and past locations.

Dataset. We evaluate our proposed NATRA on three widely used datasets: the ETH/UCY [33, 53], SDD [57], and Argoverse dataset [10]. **ETH/UCY** dataset is composed of 5 scenes, including ETH, HOTEL, UNIV, ZARA1, and ZARA2, with 1,536 pedestrians recorded in total. Following [4, 24, 45, 88], we adopt the "leave-one-out" strategy, where the models are trained on 4 scenes and tested on the remaining scene. **SDD** consists of 20 scenes captured using a drone in a top-down view around the university campus containing several types of moving agents such as humans, bicyclists, skateboarders, and vehicles, which contains 5,232 trajectories in total. Following works in [21, 84], we use 8 frames of trajectories are used as observations to predict the next 12 frames. **Argoverse** is a large-scale trajectory prediction dataset for autonomous driving, comprising of 324,557 scenes. The historical trajectories contains 20 frames with 10Hz as the sample rate, which is used to predict future 30 frames of future trajectories.

To verify the robustness of NATRA against noise, we add noise into original trajectory prediction datasets, ETH/UCY, SDD and Argoverse. We employ two settings: **1)** We first add Gaussian noise $\mathcal{N}(0, \sigma)$ for noise simulation.[2] **2)** To verify the noise-agnostic property of NATRA, we additionally incorporate Poisson noise $\mathcal{P}(\lambda)$, mixed noise, and multiplicative noise. **3)** To evaluate NATRA's capability in handling real-world noisy data, we extract high-noise samples from the original dataset to construct a noisy dataset that reflects real-world noise conditions.

Evaluation Metrics. Following previous works [21, 47, 61, 64], we employ ADE_k and FDE_k to evaluate the predicted trajectories. ADE is the average L2 error between all future timesteps, and FDE is the error at the final timestamp. We take the best out of $K = 20$ predictions to account for the multi-modality for trajectory prediction on ETH/UCY and SDD dataset, as in [62, 86], and we set $K = 1$ and $K = 6$ on Argoverse Dataset, as in [41, 105]

[2]The Central Limit Theorem [32] suggests that combination of various noise sources tend to approximate a Gaussian distribution.

Implementation Details. Implementation details of NATRA are elaborated in Appendix 6.9

4.2. Results and Analysis

Performance on Trajectory Predictions with Gaussian Noisy Observations. We evaluate the performance of our proposed NATRA and compare it with various baselines on the ETH/UCY and SDD datasets, with σ setting to 0.2 and 0.4.[3] The results are listed in Table 1 and Table 2. Based on the two tables, NATRA+GraphTern and NATRA+EqMotion significantly outperforms GraphTern and EqMotion on the two datasets under the setting of $\sigma = 0.2$ and $\sigma = 0.4$. This illustrates current state-of-the-art methods cannot well tackle the case of noisy observations. In addition, results in Table 3 show that NATRA also outperforms baselines on Argoverse Dataset under the setting of $\sigma = 4$, which further demonstrates its superiority.

However, when integrating our proposed NATRA into these two models, the performance can be significantly improved. This demonstrates the effectiveness of our method for trajectory prediction with noisy observations, and also highlights its compatibility with different trajectory prediction models. Furthermore, NATRA outperforms the Wavelet and EMA denoising methods, further underscoring the superiority of our proposed approach.

Table 4. Comparison of different methods on the noisy subset of ETH/UCY and SDD. All methods are trained and tested noisy observed trajectory inputs.

Method	ETH/UCY		SDD(Noisy)	
	ADE	FDE	ADE	FDE
GraphTern	0.48	0.66	17.92	25.87
EMA+GraphTern	0.45	0.61	16.80	23.33
Wavelet+GraphTern	0.45	0.59	16.11	22.47
NATRA+GraphTern	0.42	0.54	15.29	21.39

Performance under Different Noise Setting. We conduct experiments to verify the effectiveness of NATRA for various noise settings. we added (1) Poisson noise, (2) Mixed noise (Gaussian + Poisson), (3) Noise randomly multiplied by $\delta \in [0.95, 1]$ and (4) Gaussian noise randomly sampled from $\sigma \in \{0.2, 0.4\}$. The results are listed in Table 5. We observe our method consistently outperforms the baselines across various settings, which demonstrates the effectiveness of NATRA and it is agnostic to noise distributions in principle.

Performace on Real-world Noisy Dataset. To evaluate NATRA on real-world noise, we extract a noisy subset from ETH/UCY and SDD dataset by performing cubic regression

[3]We employ these values because they present a challenge. With a pedestrian speed of approximately 1.0 m/s [19], these noise levels correspond to 20% and 40% of the speed, which significantly affects the accuracy of trajectory predictions.

Table 5. Comparison of different methods under different noise setting on the SDD dataset. The evaluation metrics are ADE and FDE (Unit: pixels). The best results are highlighted in **bold**.

(a) Possion Noise ($\lambda = 0.4$).

Noise	Method	SDD	
		ADE	FDE
$\lambda = 0.4$	EqMotion	14.05	19.46
	Wavelet+EqMotion	12.95	17.97
	EMA+EqMotion	13.15	17.58
	NATRA+EqMotion	**12.22**	**16.23**

(b) Mixed noise composed of Gaussian noise ($\sigma = 0.2$) and Poisson noise ($\lambda = 0.2$).

Noise	Method	SDD	
		ADE	FDE
$\sigma = 0.2$ $\lambda = 0.2$	EqMotion	14.66	20.27
	Wavelet+EqMotion	13.48	18.30
	EMA+EqMotion	13.82	18.97
	NATRA+EqMotion	**12.96**	**17.09**

(c) Noise randomly multiplied by $\delta \in [0.95, 1.0]$.

Noise	Method	SDD	
		ADE	FDE
$\delta \in [0.95, 1.0]$	EqMotion	17.48	19.10
	Wavelet+EqMotion	16.17	18.18
	EMA+EqMotion	16.25	18.38
	NATRA+EqMotion	**15.66**	**17.39**

(d) Gaussian Noise sampled from $\sigma \in \{0.2, 0.4\}$.

Noise	Method	SDD	
		ADE	FDE
$\sigma \in \{0.2, 0.4\}$	EqMotion	13.33	18.92
	Wavelet+EqMotion	12.82	15.66
	EMA+EqMotion	12.76	15.50
	NATRA+EqMotion	**12.28**	**15.15**

on observed trajectories and calculating MSE between the regressed and original trajectories. The top 10% of MSE trajectories are selected as the noisy dataset. We conduct experiments for comparison. Results listed in Table 4 show that NATRA still outperforms baselines.

Ablation Studies. We conduct ablation studies on the components of our proposed method. We utilize GraphTern as the backbone and set σ to 0.4. The results are listed in Table 6. When max term of $\mathcal{L}_{MI}$ between denoised and clean trajectory is employed, the performance increases, as it makes denoised trajectories close to clean pattern. If min term of $\mathcal{L}_{MI}$ between denoised and noisy trajectories is added, the performance further improves, as it makes the denoised trajectory not contain noisy information. Nevertheless, since the useful information in observations may also discarded via min term, reconstruction is needed. After adding $\mathcal{L}_{rec}$ term, the useful information in observations are preserved, further boost performance. Finally, we add the ranking loss $\mathcal{L}_{rank}$, which enables our method to achieve the best performance. This indicates that $\mathcal{L}_{rank}$ enhances the capability of the trajectory prediction model.

Table 6. Ablation Studies on each component of NATRA. The best results are highlighted in **bold.**

Component				ETH/UCY		SDD	
max of $\mathcal{L}_{MI}$	min of $\mathcal{L}_{MI}$	$\mathcal{L}_{rec}$	$\mathcal{L}_{rank}$	ADE	FDE	ADE	FDE
				0.51	0.70	13.74	19.77
✓				0.48	0.67	13.36	18.94
✓	✓			0.47	0.65	13.21	18.79
✓	✓	✓		0.42	0.59	12.74	17.96
✓	✓	✓	✓	**0.40**	**0.55**	**12.35**	**17.28**

Qualitative Results. We visualize the denoised observations and predicted future trajectories generated by EWA, Wavelet, and NATRA, using GraphTern as the backbone. The results are shown in Figure 3. We observe that NATRA can generate less noisy observed trajectories and more accurate future trajectories compared to other methods. This demonstrates the proposed mutual information-based mechanism effectively denoises the observations, and the ranking loss aids in forecasting more precise future trajectories.

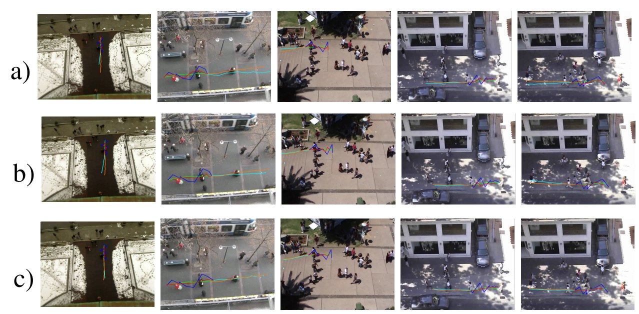

Figure 3. Visualizations of (a) EMA+GraphTern, (b) Wavelet+GraphTern, (c) NATRA+GraphTern on the ETH/UCY Dataset. The clean, noisy, and denoised observations are shown in green, blue, and red, respectively. The noisy observations are obtained through adding Gaussian noise $\mathcal{N}(0, \sigma = 0.4)$ into clean observations. The ground-truth and predicted future trajectories are shown in orange and cyan, respectively.

5. Conclusion

In this paper, we investigated a challenging task of trajectory prediction with noisy observations. We proposed NATRA, a noise-agnostic framework that simultaneously filters out noise and predicts future trajectories. To remove the noise from the observations, we designed a denoising mechanism by jointly optimizing a mutual information-based and a reconstruction loss. Moreover, we devised a ranking loss that requires the prediction performance using denoised observed trajectories to be superior to that using the original noisy observations, thereby further improving the performance of the model. Extensive experiments demonstrated the effectiveness of NATRA and its compatibility with various trajectory prediction models.

Acknowledgements

This work was supported by the NSFC under Grants U2441242.

References

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

[2] Florent Altché and Arnaud de La Fortelle. An lstm network for highway trajectory prediction. In *2017 IEEE 20th international conference on intelligent transportation systems (ITSC)*, pages 353–359. IEEE, 2017. 2

[3] Inhwan Bae and Hae-Gon Jeon. A set of control points conditioned pedestrian trajectory prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 6155–6165, 2023. 1, 6

[4] Inhwan Bae, Jean Oh, and Hae-Gon Jeon. Eigentrajectory: Low-rank descriptors for multi-modal trajectory forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10017–10029, 2023. 1, 7

[5] Inhwan Bae, Junoh Lee, and Hae-Gon Jeon. Can language beat numerical regression? language-based multimodal trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 753–766, 2024. 1

[6] Inhwan Bae, Young-Jae Park, and Hae-Gon Jeon. Singulartrajectory: Universal trajectory predictor using diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17890–17901, 2024. 2

[7] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. Mutual information neural estimation. In *International conference on machine learning*, pages 531–540. PMLR, 2018. 4

[8] Yulong Cao, Chaowei Xiao, Anima Anandkumar, Danfei Xu, and Marco Pavone. Advdo: Realistic adversarial attacks for trajectory prediction. In *European Conference on Computer Vision*, pages 36–52. Springer, 2022. 3

[9] Yulong Cao, Danfei Xu, Xinshuo Weng, Zhuoqing Mao, Anima Anandkumar, Chaowei Xiao, and Marco Pavone. Robust trajectory prediction against adversarial attacks. In *Conference on robot learning*, pages 128–137. PMLR, 2023. 3

[10] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8748–8757, 2019. 3, 7

[11] Guangyi Chen, Zhenhao Chen, Shunxing Fan, and Kun Zhang. Unsupervised sampling promoting for stochastic human trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17874–17884, 2023. 1

[12] Hao Chen, Jiaze Wang, Kun Shao, Furui Liu, Jianye Hao, Chenyong Guan, Guangyong Chen, and Pheng-Ann Heng. Traj-mae: Masked autoencoders for trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8351–8362, 2023. 1

[13] Pengyu Cheng, Weituo Hao, Shuyang Dai, Jiachang Liu, Zhe Gan, and Lawrence Carin. Club: A contrastive log-ratio upper bound of mutual information. In *International conference on machine learning*, pages 1779–1788. PMLR, 2020. 4

[14] Sehwan Choi, Jungho Kim, Junyong Yun, and Jun Won Choi. R-pred: Two-stage motion prediction via tube-query attention-based trajectory refinement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8525–8535, 2023. 1

[15] Matteo Corbetta, Portia Banerjee, Wendy Okolo, George Gorospe, and Dmitry G Luchinsky. Real-time uav trajectory prediction for safety monitoring in low-altitude airspace. In *Aiaa aviation 2019 forum*, page 3514, 2019. 1

[16] Patrick Dendorfer, Sven Elflein, and Laura Leal-Taixé. Mg-gan: A multi-generator model preventing out-of-distribution samples in pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13158–13167, 2021. 2

[17] Monroe D Donsker and SR Srinivasa Varadhan. Asymptotic evaluation of certain markov process expectations for large time. iv. *Communications on pure and applied mathematics*, 36(2):183–212, 1983. 4

[18] Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection. *Neural networks*, 108:466–478, 2018. 2

[19] Albert Forde and Janice Daniel. Pedestrian walking speed at un-signalized midblock crosswalk and its impact on urban street segment performance. *Journal of Traffic and Transportation Engineering (English Edition)*, 8(1):57–69, 2021. 7

[20] Junru Gu, Chen Sun, and Hang Zhao. Densetnt: End-to-end trajectory prediction from dense goal sets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15303–15312, 2021. 2

[21] Tianpei Gu, Guangyi Chen, Junlong Li, Chunze Lin, Yongming Rao, Jie Zhou, and Jiwen Lu. Stochastic trajectory prediction via motion indeterminacy diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17113–17122, 2022. 1, 6, 7

[22] Xunjiang Gu, Guanyu Song, Igor Gilitschenski, Marco Pavone, and Boris Ivanovic. Producing and leveraging online map uncertainty in trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14521–14530, 2024. 1

[23] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceed-*

ings of the IEEE conference on computer vision and pattern recognition, pages 2255–2264, 2018. 2

[24] Yingfan Huang, Huikun Bi, Zhaoxin Li, Tianlu Mao, and Zhaoqi Wang. Stgat: Modeling spatial-temporal interactions for human trajectory prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6272–6281, 2019. 7

[25] Nikolay Jetchev and Marc Toussaint. Trajectory prediction: learning to map situations to robot trajectories. In *Proceedings of the 26th annual international conference on machine learning*, pages 449–456, 2009. 1

[26] Chiyu Jiang, Andre Cornman, Cheolho Park, Benjamin Sapp, Yin Zhou, Dragomir Anguelov, et al. Motiondiffuser: Controllable multi-agent motion prediction using diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9644–9653, 2023. 2

[27] Ruochen Jiao, Xiangguo Liu, Takami Sato, Qi Alfred Chen, and Qi Zhu. Semi-supervised semantics-guided adversarial training for robust trajectory prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8207–8217, 2023. 3

[28] Miao Kang, Shengqi Wang, Sanping Zhou, Ke Ye, Jingjing Jiang, and Nanning Zheng. Ffinet: Future feedback interaction network for motion forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 2024. 2

[29] Sungjune Kim, Hyung-gun Chi, Hyerin Lim, Karthik Ramani, Jinkyu Kim, and Sangpil Kim. Higher-order relational reasoning for pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15251–15260, 2024. 2

[30] Hao Kong, Jie Xu, Shenjian Gong, Jian Yang, and Shanshan Zhang. Adaptive pedestrian trajectory prediction via target-directed angle augmentation. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4065–4069. IEEE, 2024. 2

[31] Vineet Kosaraju, Amir Sadeghian, Roberto Martín-Martín, Ian Reid, Hamid Rezatofighi, and Silvio Savarese. Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks. *Advances in neural information processing systems*, 32, 2019. 2

[32] Sang Gyu Kwak and Jong Hae Kim. Central limit theorem: the cornerstone of modern statistics. *Korean journal of anesthesiology*, 70(2):144–156, 2017. 7

[33] Laura Leal-Taixé, Michele Fenzi, Alina Kuznetsova, Bodo Rosenhahn, and Silvio Savarese. Learning an image-based motion context for multiple people tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3542–3549, 2014. 1, 2, 7

[34] Mihee Lee, Samuel S Sohn, Seonghyeon Moon, Sejong Yoon, Mubbasir Kapadia, and Vladimir Pavlovic. Muse-vae: Multi-scale vae for environment-aware long term trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2221–2230, 2022. 2

[35] Namhoon Lee, Wongun Choi, Paul Vernaza, Christopher B Choy, Philip HS Torr, and Manmohan Chandraker. Desire: Distant future prediction in dynamic scenes with interacting agents. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 336–345, 2017. 2

[36] Changsheng Li, Rongqing Li, Ye Yuan, Guoren Wang, and Dong Xu. Deep unsupervised active learning via matrix sketching. *IEEE Transactions on Image Processing*, 30: 9280–9293, 2021.

[37] Rongqing Li, Changsheng Li, Yuhang Li, Hanjie Li, Yi Chen, Ye Yuan, and Guoren Wang. Itpnet: Towards instantaneous trajectory prediction for autonomous driving. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1643–1654, 2024. 2

[38] Rongqing Li, Changsheng Li, Dongchun Ren, Guangyi Chen, Ye Yuan, and Guoren Wang. Bcdiff: Bidirectional consistent diffusion for instantaneous trajectory prediction. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[39] Xin Li, Xiaowen Ying, and Mooi Choo Chuah. Grip: Graph-based interaction-aware trajectory prediction. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 3960–3966. IEEE, 2019. 2

[40] Yuke Li. Which way are you going? imitative decision learning for path forecasting in dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 294–303, 2019. 2

[41] Yuhang Li, Changsheng Li, Baoyu Fan, Rongqing Li, Ziyue Zhang, Dongchun Ren, Ye Yuan, and Guoren Wang. Fdnet: Feature decoupling framework for trajectory prediction. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9997–10004. IEEE, 2024. 7

[42] Yuhang Li, Changsheng Li, Ruilin Lv, Rongqing Li, Ye Yuan, and Guoren Wang. Lakd: Length-agnostic knowledge distillation for trajectory prediction with any length observations. *Advances in Neural Information Processing Systems*, 37:28720–28744, 2025. 2

[43] Rongqin Liang, Yuanman Li, Xia Li, Yi Tang, Jiantao Zhou, and Wenbin Zou. Temporal pyramid network for pedestrian trajectory prediction with multi-supervision. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2029–2037, 2021. 2

[44] Matthias Luber, Johannes A Stork, Gian Diego Tipaldi, and Kai O Arras. People tracking with human motion predictions from social forces. In *2010 IEEE international conference on robotics and automation*, pages 464–469. IEEE, 2010. 2

[45] Karttikeya Mangalam, Harshayu Girase, Shreyas Agarwal, Kuan-Hui Lee, Ehsan Adeli, Jitendra Malik, and Adrien Gaidon. It is not the journey but the destination: Endpoint conditioned trajectory prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 759–776. Springer, 2020. 7

[46] Karttikeya Mangalam, Yang An, Harshayu Girase, and Jitendra Malik. From goals, waypoints & paths to long term human trajectory forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15233–15242, 2021. 2

[47] Weibo Mao, Chenxin Xu, Qi Zhu, Siheng Chen, and Yanfeng Wang. Leapfrog diffusion model for stochastic trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5517–5526, 2023. 7

[48] Mancheng Meng, Ziyan Wu, Terrence Chen, Xiran Cai, Xiang Zhou, Fan Yang, and Dinggang Shen. Forecasting human trajectory from scene history. *Advances in Neural Information Processing Systems*, 35:24920–24933, 2022. 2

[49] Ray Coden Mercurius, Ehsan Ahmadi, Soheil Mohamad Alizadeh Shabestary, and Amir Rasouli. Amend: A mixture of experts framework for long-tailed trajectory prediction. *arXiv preprint arXiv:2402.08698*, 2024. 2

[50] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14424–14432, 2020. 2

[51] Abduallah Mohamed, Deyao Zhu, Warren Vu, Mohamed Elhoseiny, and Christian Claudel. Social-implicit: Rethinking trajectory prediction evaluation and the effectiveness of implicit maximum likelihood estimation. In *European Conference on Computer Vision*, pages 463–479. Springer, 2022. 6

[52] Alessio Monti, Angelo Porrello, Simone Calderara, Pasquale Coscia, Lamberto Ballan, and Rita Cucchiara. How many observations are enough? knowledge distillation for trajectory forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6553–6562, 2022. 2

[53] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th international conference on computer vision*, pages 261–268. IEEE, 2009. 1, 2, 7

[54] Tran Phong, Haoran Wu, Cunjun Yu, Panpan Cai, Sifa Zheng, and David Hsu. What truly matters in trajectory prediction for autonomous driving? *Advances in Neural Information Processing Systems*, 36, 2024. 1

[55] Mozhgan Pourkeshavarz, Junrui Zhang, and Amir Rasouli. Cadet: a causal disentanglement approach for robust trajectory prediction in autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14874–14884, 2024. 1

[56] Davis Rempe, Zhengyi Luo, Xue Bin Peng, Ye Yuan, Kris Kitani, Karsten Kreis, Sanja Fidler, and Or Litany. Trace and pace: Controllable pedestrian animation via guided trajectory diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13756–13766, 2023. 2

[57] Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. Learning social etiquette: Human trajectory understanding in crowded scenes. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*, pages 549–565. Springer, 2016. 1, 2, 7

[58] Christoph Rösmann, Malte Oeljeklaus, Frank Hoffmann, and Torsten Bertram. Online trajectory prediction and planning for social robot navigation. In *2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 1255–1260. IEEE, 2017. 1

[59] Luke Rowe, Martin Ethier, Eli-Henry Dykhne, and Krzysztof Czarnecki. Fjmp: Factorized joint multi-agent motion prediction over learned directed acyclic interaction graphs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13745–13755, 2023. 2

[60] Saeed Saadatnejad, Mohammadhossein Bahari, Pedram Khorsandi, Mohammad Saneian, Seyed-Mohsen Moosavi-Dezfooli, and Alexandre Alahi. Are socially-aware trajectory prediction models really socially-aware? *Transportation research part C: emerging technologies*, 141:103705, 2022. 3

[61] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. Sophie: An attentive gan for predicting paths compliant to social and physical constraints. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1349–1358, 2019. 2, 7

[62] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*, pages 683–700. Springer, 2020. 7

[63] Nasim Shafiee, Taskin Padir, and Ehsan Elhamifar. Introvert: Human trajectory prediction via conditional 3d attention. In *Proceedings of the IEEE/cvf Conference on Computer Vision and Pattern recognition*, pages 16815–16825, 2021. 2

[64] Liushuai Shi, Le Wang, Chengjiang Long, Sanping Zhou, Mo Zhou, Zhenxing Niu, and Gang Hua. Sgcn: Sparse graph convolution network for pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8994–9003, 2021. 7

[65] Liushuai Shi, Le Wang, Chengjiang Long, Sanping Zhou, Fang Zheng, Nanning Zheng, and Gang Hua. Social interpretable tree for pedestrian trajectory prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2235–2243, 2022. 2

[66] Liushuai Shi, Le Wang, Sanping Zhou, and Gang Hua. Trajectory unified transformer for pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9675–9684, 2023. 2

[67] Benjamin Stoler, Ingrid Navarro, Meghdeep Jana, Soonmin Hwang, Jonathan Francis, and Jean Oh. Safeshift: Safety-informed distribution shifts for robust trajectory prediction in autonomous driving. *arXiv preprint arXiv:2309.08889*, 2023. 2

[68] Hao Sun, Zhiqun Zhao, and Zhihai He. Reciprocal learning networks for human trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7416–7425, 2020. 2

[69] Jianhua Sun, Qinhong Jiang, and Cewu Lu. Recursive social behavior graph for trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 660–669, 2020. 2

[70] Jianhua Sun, Yuxuan Li, Hao-Shu Fang, and Cewu Lu. Three steps to multimodal trajectory prediction: Modality clustering, classification and synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13250–13259, 2021. 2

[71] Jianhua Sun, Yuxuan Li, Liang Chai, Hao-Shu Fang, Yong-Lu Li, and Cewu Lu. Human trajectory prediction with momentary observation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6467–6476, 2022. 2

[72] Jianqiao Sun, Ziheng Cheng, Bo Chen, Xin Yuan, Chunhui Qu, and Hongwei Liu. Mepat: Meta-prior aided transformer for adverse weather condition restoration. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025. 1

[73] Shangquan Sun, Wenqi Ren, Xinwei Gao, Rui Wang, and Xiaochun Cao. Restoring images in adverse weather conditions via histogram transformer. In *European Conference on Computer Vision*, pages 111–129. Springer, 2024. 1

[74] Xiaolong Tang, Meina Kan, Shiguang Shan, Zhilong Ji, Jinfeng Bai, and Xilin Chen. Hpnet: Dynamic trajectory forecasting with historical prediction attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15261–15270, 2024. 2

[75] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. *arXiv preprint physics/0004057*, 2000. 3

[76] Maria Valera and Sergio A Velastin. Intelligent distributed surveillance systems: a review. *IEE Proceedings-Vision, Image and Signal Processing*, 152(2):192–204, 2005. 1

[77] Anirudh Vemula, Katharina Muelling, and Jean Oh. Social attention: Modeling attention in human crowds. In *2018 IEEE international Conference on Robotics and Automation (ICRA)*, pages 4601–4607. IEEE, 2018. 2

[78] Mingkun Wang, Xinge Zhu, Changqian Yu, Wei Li, Yuexin Ma, Ruochun Jin, Xiaoguang Ren, Dongchun Ren, Mingxu Wang, and Wenjing Yang. Ganet: Goal area network for motion forecasting. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1609–1615. IEEE, 2023. 2

[79] Xishun Wang, Tong Su, Fang Da, and Xiaodong Yang. Prophnet: Efficient agent-centric motion forecasting with anchor-informed proposals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21995–22003, 2023. 1

[80] Yuning Wang, Pu Zhang, Lei Bai, and Jianru Xue. Fend: A future enhanced distribution-aware contrastive learning framework for long-tail trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1400–1409, 2023. 2

[81] Di Wen, Haoran Xu, Zhaocheng He, Zhe Wu, Guang Tan, and Peixi Peng. Density-adaptive model based on motif matrix for multi-agent trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14822–14832, 2024. 1

[82] Xinshuo Weng, Boris Ivanovic, Kris Kitani, and Marco Pavone. Whose track is it anyway? improving robustness to tracking errors with affinity-based trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6573–6582, 2022. 3

[83] Xinshuo Weng, Boris Ivanovic, and Marco Pavone. Mtp: Multi-hypothesis tracking and prediction for reduced error propagation. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, pages 1218–1225. IEEE, 2022. 3

[84] Conghao Wong, Beihao Xia, Ziming Hong, Qinmu Peng, Wei Yuan, Qiong Cao, Yibo Yang, and Xinge You. View vertically: A hierarchical network for trajectory prediction via fourier spectrums. In *European Conference on Computer Vision*, pages 682–700. Springer, 2022. 7

[85] Conghao Wong, Beihao Xia, Ziqian Zou, Yulong Wang, and Xinge You. Socialcircle: Learning the angle-based social interaction representation for pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19005–19015, 2024. 2

[86] Chenxin Xu, Maosen Li, Zhenyang Ni, Ya Zhang, and Siheng Chen. Groupnet: Multiscale hypergraph neural networks for trajectory prediction with relational reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6498–6507, 2022. 7

[87] Chenxin Xu, Robby T Tan, Yuhong Tan, Siheng Chen, Yu Guang Wang, Xinchao Wang, and Yanfeng Wang. Eqmotion: Equivariant multi-agent motion prediction with invariant interaction reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1410–1420, 2023. 1, 6

[88] Pei Xu, Jean-Bernard Hayet, and Ioannis Karamouzas. Socialvae: Human trajectory prediction using timewise latents. In *European Conference on Computer Vision*, pages 511–528. Springer, 2022. 1, 6, 7

[89] Yi Xu and Yun Fu. Adapting to length shift: Flexilength network for trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15226–15237, 2024. 2

[90] Yi Xu, Lichen Wang, Yizhou Wang, and Yun Fu. Adaptive trajectory prediction via transferable gnn. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6520–6531, 2022. 2

[91] Yi Xu, Armin Bazarjani, Hyung-gun Chi, Chiho Choi, and Yun Fu. Uncovering the missing pattern: Unified framework towards trajectory imputation and prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9632–9643, 2023. 2

[92] Yihong Xu, Loïck Chambon, Éloi Zablocki, Mickaël Chen, Alexandre Alahi, Matthieu Cord, and Patrick Pérez. Towards motion forecasting with real-world perception inputs: Are end-to-end approaches competitive? In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 18428–18435. IEEE, 2024. 3

[93] Hao Xue, Du Q Huynh, and Mark Reynolds. Ss-lstm: A hierarchical lstm model for pedestrian trajectory prediction.

In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1186–1194. IEEE, 2018. 2

[94] Maosheng Ye, Tongyi Cao, and Qifeng Chen. Tpcn: Temporal point cloud networks for motion forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11318–11327, 2021. 2

[95] Rui Yu and Zihan Zhou. Towards robust human trajectory prediction in raw videos. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8059–8066. IEEE, 2021. 3

[96] Ye Yuan, Xinshuo Weng, Yanglan Ou, and Kris M Kitani. Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9813–9823, 2021. 2

[97] Haichao Zhang, Yi Xu, Hongsheng Lu, Takayuki Shimizu, and Yun Fu. Oostraj: Out-of-sight trajectory prediction with vision-positioning denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14802–14811, 2024. 2

[98] Junrui Zhang, Mozhgan Pourkeshavarz, and Amir Rasouli. Tract: A training dynamics aware contrastive learning framework for long-tail trajectory prediction. *arXiv preprint arXiv:2404.12538*, 2024. 2

[99] Kaihao Zhang, Rongqing Li, Yanjiang Yu, Wenhan Luo, and Changsheng Li. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Transactions on Image Processing*, 30:7419–7431, 2021. 1

[100] Liang Zhang, Nathaniel Xu, Pengfei Yang, Gaojie Jin, Cheng-Chao Huang, and Lijun Zhang. Trajpac: towards robustness verification of pedestrian trajectory prediction models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8327–8339, 2023. 3

[101] Pu Zhang, Lei Bai, Yuning Wang, Jianwu Fang, Jianru Xue, Nanning Zheng, and Wanli Ouyang. Towards trajectory forecasting from detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):12550–12561, 2023. 3

[102] Qingzhao Zhang, Shengtuo Hu, Jiachen Sun, Qi Alfred Chen, and Z Morley Mao. On adversarial robustness of trajectory prediction for autonomous vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15159–15168, 2022. 3

[103] Hang Zhao, Jiyang Gao, Tian Lan, Chen Sun, Ben Sapp, Balakrishnan Varadarajan, Yue Shen, Yi Shen, Yuning Chai, Cordelia Schmid, et al. Tnt: Target-driven trajectory prediction. In *Conference on Robot Learning*, pages 895–904. PMLR, 2021. 2

[104] 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. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12126–12134, 2019. 2

[105] Zikang Zhou, Luyao Ye, Jianping Wang, Kui Wu, and Kejie Lu. Hivt: Hierarchical vector transformer for multi-agent motion prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8823–8833, 2022. 1, 6, 7

[106] Zikang Zhou, Jianping Wang, Yung-Hui Li, and Yu-Kai Huang. Query-centric trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17863–17873, 2023. 1

[107] Dekai Zhu, Guangyao Zhai, Yan Di, Fabian Manhardt, Hendrik Berkemeyer, Tuan Tran, Nassir Navab, Federico Tombari, and Benjamin Busam. Ipcc-tp: Utilizing incremental pearson correlation coefficient for joint multi-agent trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5507–5516, 2023. 2

[108] Yiyao Zhu, Di Luan, and Shaojie Shen. Biff: Bi-level future fusion with polyline-based coordinate for interactive trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8260–8271, 2023. 2