BAT: Behavior-Aware Human-Like Trajectory Prediction for Autonomous Driving

Haicheng Liao[1*], Zhenning Li[1*†], Huanming Shen[2], Wenxuan Zeng[3], Dongping Liao[1], Guofa Li[4], Shengbo Eben Li[5], Chengzhong Xu[1]

[1] University of Macau
[2] University of Electronic Science and Technology of China
[3] Peking University
[4] Chongqing University
[5] Tsinghua University
{yc27979, zhenningli, yb97428, czxu}@um.edu.mo,
huanmingshen@outlook.com, zwx.andy@stu.pku.edu.cn, hanshan198@gmail.com, lishbo@tsinghua.edu.cn

Abstract

The ability to accurately predict the trajectory of surrounding vehicles is a critical hurdle to overcome on the journey to fully autonomous vehicles. To address this challenge, we pioneer a novel behavior-aware trajectory prediction model (BAT) that incorporates insights and findings from traffic psychology, human behavior, and decision-making. Our model consists of behavior-aware, interaction-aware, priority-aware, and position-aware modules that perceive and understand the underlying interactions and account for uncertainty and variability in prediction, enabling higher-level learning and flexibility without rigid categorization of driving behavior. Importantly, this approach eliminates the need for manual labeling in the training process and addresses the challenges of non-continuous behavior labeling and the selection of appropriate time windows. We evaluate BAT's performance across the Next Generation Simulation (NGSIM), Highway Drone (HighD), Roundabout Drone (RounD), and Macao Connected Autonomous Driving (MoCAD) datasets, showcasing its superiority over prevailing state-of-the-art (SOTA) benchmarks in terms of prediction accuracy and efficiency. Remarkably, even when trained on reduced portions of the training data (25%), our model outperforms most of the baselines, demonstrating its robustness and efficiency in predicting vehicle trajectories and the potential to reduce the amount of data required to train autonomous vehicles, especially in corner cases. In conclusion, the behavior-aware model represents a significant advancement in the development of autonomous vehicles capable of predicting trajectories with the same level of proficiency as human drivers. The project page is available on our GitHub.

Introduction

Recent advancements in autonomous driving (AD) have been remarkable. Nonetheless, as we move towards the commercialization of high-level AD technology, challenges abound. One of the most significant barriers is equipping autonomous vehicles (AVs) with the ability to anticipate the trajectory of nearby vehicles in intricate situations as skillfully as humans.

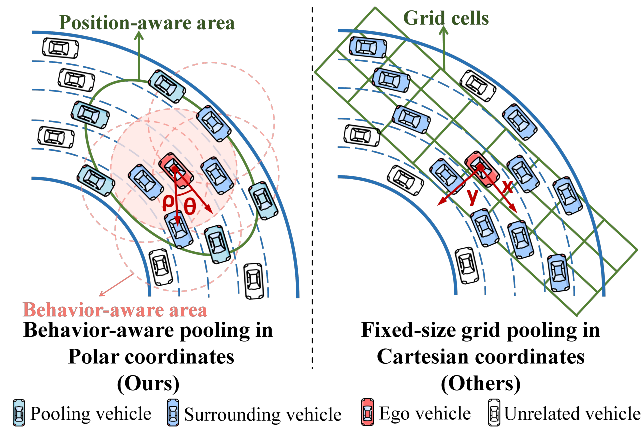

Figure 1: An overview of our proposed behavior-aware pooling mechanism and the classical pooling mechanism. Left: Modeling the vehicle using polar coordinates. Right: Modeling the vehicle using fixed-size grids and representing the position in Cartesian coordinates.

Driving, for humans, necessitates continuous monitoring of the current states of surrounding vehicles and forecasting their future states before actions like acceleration or overtaking. These states, predominantly determined by trajectories, form the bedrock of safe driving and collision prevention. This demands a keen assessment of the *interaction* among vehicles and an unbiased grasp of their *behavior*, in line with traffic regulations and accumulated driving experience (Müller, Risto, and Emmenegger 2016).

In our quest to enhance the trajectory prediction capabilities of AVs, mimicking human-like comprehension and response to surrounding scenarios might be a breakthrough. As highlighted in (Schwarting et al. 2019; Wang et al. 2022a), accounting for the behaviors of other drivers in the decision-making processes of AVs can potentially result in enhanced driving performance. With this understanding, we advocate that a deeper dive into driver behavior can significantly uplift trajectory prediction for AVs.

Previous investigations have posited that there exists a certain relationship between different drivers' behaviors and

*Authors contributed equally; †Corresponding author.

their driving performance (Toledo, Musicant, and Lotan 2008; Chandra et al. 2020; Xie et al. 2020). When confronted with the prospect of another vehicle attempting to overtake, aggressive drivers may accelerate to impede the overtaking vehicle, while cautious drivers may reduce their speed slightly to facilitate safe passing. In addition, driver behavior on the road tends to exhibit a degree of predictability, persistence, and consistency (Hang, Lv, and Chen 2022; Schwarting et al. 2019). For example, an individual who has recently violated the speed limit is likely to continue driving at high speeds as long as circumstances allow, while cautious drivers maintain their conservative driving strategy. These stability and repetition characteristics make it possible to predict and anticipate the behavior of other drivers.

In addition, humans naturally perceive their surroundings in relative terms, especially when it involves spatial understanding. This intrinsic way of processing spatial data based on relative positioning and orientation often does not align with the fixed Cartesian coordinates commonly used in many predictive models. However, polar coordinates, which detail a point's position based on its distance from a reference and the angle from a reference direction, echo this human-centric perception. When driving, humans think in terms like ”slightly ahead and to the right” rather than specific Cartesian coordinates. Adopting this perspective, our pioneering pooling mechanism, as illustrated in Fig.1, captures vehicle positions using polar coordinates, offering a more intuitive representation especially pertinent for trajectory prediction in AVs.

Despite extensive research in AD trajectory prediction, significant gaps remain. To bridge these, we've combined insights from human behavior and decision-making to design an innovative behavior-aware trajectory prediction model. In summary, our work's principal contributions are:

- We present a novel dynamic geometric graph approach that eliminates the need for manual labeling during training. This method addresses the challenges of labeling non-continuous behaviors and selecting appropriate time windows, while effectively capturing continuous driving behavior. Inspired by traffic psychology, decision theory, and driving dynamics, our model incorporates centrality metrics and behavior-aware criteria to provide enhanced flexibility and accuracy in representing driving behavior. To the best of our knowledge, this is the first attempt to incorporate **continuous representation of behavioral knowledge** in trajectory prediction for AVs.
- We propose a novel pooling mechanism, aligned with human observational instincts, that extracts vehicle positions in **polar coordinates**. It simplifies the representation of direction and distance in Cartesian coordinates, accounts for road curvature, and allows modeling in complex scenarios such as roundabouts and intersections.
- We introduce a new Macao Connected Autonomous Driving (MoCAD) dataset, sourced from a L5 autonomous bus with over 300 hours across campus and busy urban routes. Characterized by its unique **right-hand-drive system**, MoCAD, set to be publicly available, is pivotal for research in right-hand-drive dynamics and enhancing trajectory prediction models.
- Our model significantly outperforms the SOTA baseline models when tested on the NGSIM, HighD, RounD, and MoCAD datasets, respectively. Remarkably, it maintains impressive performance even when trained on only 25.0% of the dataset, demonstrating exceptional robustness and adaptability in various traffic scenarios, including **highways**, **roundabouts**, and **busy urban locales**.

Related Work

A plethora of research has been conducted in the realm of trajectory prediction, with a diverse array of approaches being proposed. These approaches can be broadly classified into three categories: physics-based, statistics-based, and deep learning-based approaches.

Physics-based Approaches. These approaches are primarily divided into kinetic and kinematic models (Lin, Ulsoy, and LeBlanc 2000). They use principles from physics and mechanics, taking into account the current state of the vehicle, such as speed and steering angle, to make predictions (Wong et al. 2022). Despite their interpretability and computational efficiency, these methods often exhibit lower prediction accuracy compared to SOTA techniques (Huang et al. 2022).

Statistics-based Approaches. In contrast, statistical-based approaches, both parametric and non-parametric, describe predicted trajectories using predefined maneuver distributions, such as Gaussian processes, hidden Markov models, dynamic Bayesian networks, and support vector machines (Wang et al. 2021; Li et al. 2023b). These methods tend to offer more refined and sophisticated model structures, resulting in better prediction performance than physics-based approaches. Their experiments on real-world data showed significant improvements over baselines.

Deep Learning-based Approaches. The surge in popularity of deep learning has led to extensive research in trajectory prediction for AVs. Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Transformers (Vaswani et al. 2017) are among the most widely used approaches, each offering unique modeling considerations and focuses (Ye, Cao, and Chen 2021; Liang et al. 2020). RNNs, such as Long Short-Term Memory (LSTM), are often used to process time-series trajectory data, while CNNs excel at extracting spatial features from inputs such as bird's-eye or raster images. Some researchers combine RNNs and CNNs to integrate both temporal and spatial features into their models (Liao et al. 2023; Huang, Mo, and Lv 2022; Bhattacharyya, Huang, and Czarnecki 2023; Zhang and Li 2022). Transformers, with their renowned success in many domains, have also demonstrated superior performance in trajectory prediction (Li et al. 2022; Zeng et al. 2023; Li et al. 2023a). Compared to physics-based and statistics-based methods, these data-driven approaches have generally demonstrated superior prediction performance, especially for tasks requiring long-term predictions (beyond 3 seconds).

Problem Formulation

The trajectory prediction task can be formulated as follows. At each time t, we predict the multimodal trajectories of the ego vehicle, based on historical observations of both the ego vehicle and its surrounding vehicles (*agents*). Given the inputs of historical observations $\boldsymbol{X}$, the model aims to predict a multi-modal distribution over future trajectories of the ego vehicle $P(\boldsymbol{Y}|\boldsymbol{X})$.

Inputs and Outputs

The inputs $\boldsymbol{X}$ to our model are the historical trajectories over a fixed time horizon t_h of both the ego vehicle (subscript 0) and all the surrounding vehicles (subscripts 1 to n):

$$\boldsymbol{X}_i^{t-t_h:t} = \left\{p_i^{t-t_h:t}\right\}, \forall i \in [0, n] \tag{1}$$

where $p_{0:n}^{t-t_h:t}$ denotes the 2D position coordinates.

The output of the model is a probability distribution over the future trajectory of the ego vehicle during the prediction horizon t_f :

$$\boldsymbol{Y} = \boldsymbol{Y}_0^{t+1:t+t_f} = \{y_0^{t+1}, y_0^{t+2}, \ldots, y_0^{t+t_f-1}, y_0^{t+t_f}\} \tag{2}$$

As aforementioned, we define the motion of the vehicles in Polar coordinates (shown in Fig.1) rather than the Cartesian coordinates. In the Polar coordinate, we assume the origin O of the stationary frame of reference is fixed at the center of the ego vehicle at time t. Our inputs and outputs can be further written as (take the ego vehicle as an example, for convenience, assume input at instant t and output at instant $t+1$):

$$x_0^t = \{\rho_0^t, \theta_0^t\} \tag{3}$$

and

$$y_0^{t+1} = \{\rho_0^{t+1}, \theta_0^{t+1}\} \tag{4}$$

where ρ and θ are the distance and angle of the vehicle.

The transformation relationship between Cartesian and Polar coordinate systems is illustrated below. Given a vehicle's position history in lateral coordinate $x_i^{t_k}$ and longitudinal coordinate $y_i^{t_k}$ at time t_k, the distance $\rho_i^{t_k}$ and vehicle orientation $\theta_i^{t_k}$ for Polar representation can be computed as the following formula:

$$\begin{cases} \rho_i^{t_k} = \sqrt{\left(x_i^{t_k} - x_0^t\right)^2 + \left(y_i^{t_k} - y_0^t\right)^2} \\ \theta_i^{t_k} = \arctan\left(\frac{y_i^{t_k} - y_0^t}{x_i^{t_k} - x_0^t}\right) \end{cases} \tag{5}$$

where x_0^t and y_0^t are the lateral and longitudinal coordinates of the ego vehicle (defined as the origin $\mathcal{O}$) at time t, respectively. $\rho_i^{t_k}$ is the Polar diameter relative to the origin $\mathcal{O}$, and $\theta_i^{t_k}$ is the orientation of the ith vehicle at time t_k.

Multi-modal Probabilistic Maneuver Prediction

To account for the uncertainty and variability in the prediction, the multimodal prediction framework considers multiple potential maneuvers that the ego vehicle could perform and estimates the probability of each maneuver based on previous observations. This not only provides multiple predictions but also quantifies the confidence level associated with each prediction. This is particularly beneficial for informed decision-making in response to anticipated maneuvers, as it allows AVs to account for the uncertainty inherent in the predictions.

Proposed Model

Fig. 2 shows the architecture of BAT, which is upon the encoder-decoder framework with four modules to capture different aspects of behaviors and interactions between different agents, including behavior-aware, interaction-aware, priority-aware, and position-aware modules. The project page is available on our Github[1].

Behavior-aware Module

The complex and dynamic nature of traffic scenarios presents significant challenges in interpreting and categorizing driver behavior. Unlike previous studies that categorize driver behavior into finite and human-defined classifications, we present a more flexible and adaptable solution, namely the behavior-aware module, by avoiding discrete behavior categories in favor of a continuous representation of behavioral information. Our behavior-aware module is motivated by the multi-policy decision-making (MPDM) framework for human drivers (Markkula et al. 2020) and integrates traffic psychology (Toghi et al. 2022) using dynamic geometric graphs (DGGs) (Dall and Christensen 2002) to model and evaluate human driving behavior.

Dynamic Geometric Graphs At time t, the graph G^t can be given as follow:

$$G^t = \{V^t, E^t\} \tag{6}$$

where $V^t = \{v_0^t, v_1^t \ldots, v_n^t\}$ is the set of nodes, v_i^t is the i-th node representing the i-th vehicle, $E^t = \{e_0^t, e_1^t \ldots, e_n^t\}$ is the set of undirected edges, and e_i^t is the edge between the node v_i^t and other vehicles that have potential influences on it. It is assumed that the interaction only exists when the nodes v_i and v_j are in close proximity to one another, or formally, the shortest distance between them, $d\left(v_i^t, v_j^t\right)$, is less than or equal to the predetermined distance threshold r. Therefore, we define

$$e_i^t = \{v_i^t v_j^t \mid (j \in N_i^t)\} \tag{7}$$

where $N_i^t = \left\{v_j^t \in V^t \backslash \{v_i^t\} \mid d\left(v_i^t, v_j^t\right) \leq r, i \neq j\right\}$.

Correspondingly, the symmetrical adjacency matrix A^t of G^t can be given as:

$$A^t(i,j) = \begin{cases} d\left(v_i^t, v_j^t\right) & \text{if } d\left(v_i^t, v_j^t\right) \leq r, i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Centrality Measures To more accurately capture the potential interactions between the observed traffic agents, we use centrality measures (degree, closeness, and eigenvector centrality measures) (Freeman 1978) as prior knowledge to portray driving behavior in DGGs.

[1]https://github.com/Petrichor625/BATraj-Behavior-aware-Model

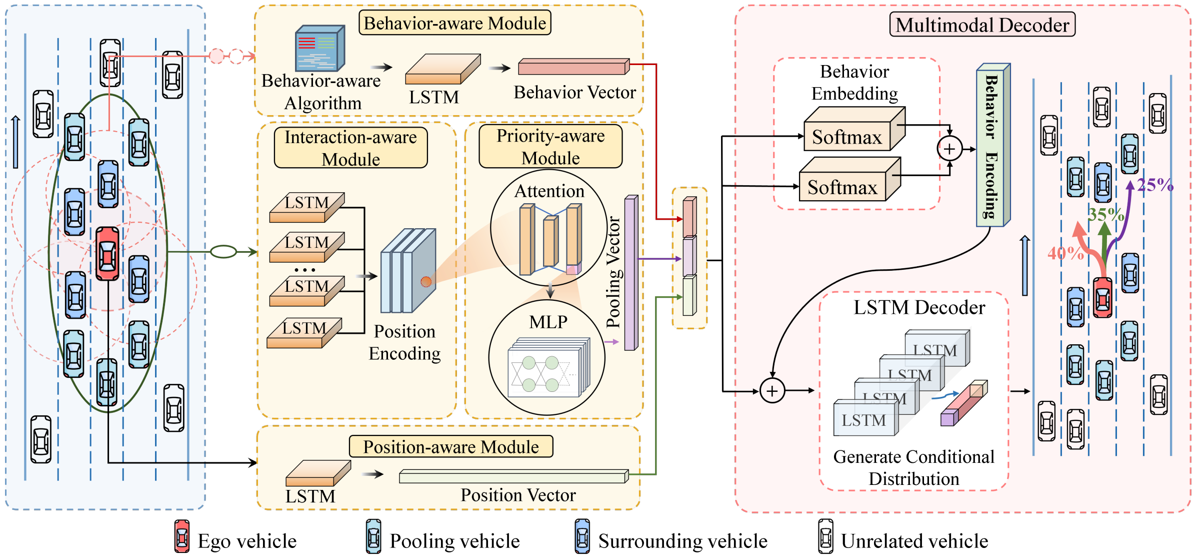

Figure 2: Architecture of behavior-aware trajectory prediction model.

Degree Centrality. Degree centrality is characterized by the count of immediate connections a node has with other nodes within the graph. This concept intuitively suggests that a traffic agent with more connections is both more susceptible to the influences of other agents and more influential in shaping their actions. Formally,

$$\mathcal{J}_i^t(D) = \left|\mathcal{N}_i^t\right| + \mathcal{J}_i^{t-1}(D) \tag{9}$$

where $|\mathcal{N}_i^t|$ is the total number of elements in $\mathcal{N}_i^t$ at time t.

Closeness Centrality. We propose that the closer a vehicle is to its surroundings, the higher its likelihood of interacting with adjacent vehicles. This idea is encapsulated by the closeness centrality metric, which gauges the ease of interaction and accessibility between a vehicle and its neighboring vehicles. Closeness centrality is determined using the shortest paths between the vehicle (node) and other vehicles in the traffic graph. This is achieved by summing the inverse of their distances. Formally,

$$\mathcal{J}_i^t(C) = \frac{|\mathcal{N}_i^t| - 1}{\sum_{\forall v_j^t \in \mathcal{N}_i^t} d\left(v_i^t, v_j^t\right)} \tag{10}$$

Eigenvector Centrality. In the context of understanding driver behavior, a vehicle's eigenvector centrality takes into account both its interactions with nearby vehicles and the influence of those interactions. Specifically, this metric integrates the vehicle's number of connections and the weight of the influence of connected vehicles. This helps to identify influential vehicles in a traffic context and their potential impact on other drivers. Formally,

$$\mathcal{J}_i^t(E) = \frac{\sum_{\forall v_j^t \in \mathcal{N}_i^t} d\left(v_i^t, v_j^t\right)}{\lambda} \tag{11}$$

where λ is the eigenvalue. In addition, the Perron-Frobenius theorem states that for a non-negative matrix (such as the adjacency matrix in our case), there exists a positive eigenvector solution for the greatest eigenvalue of the matrix (Pillai, Suel, and Cha 2005). This means that the eigenvector corresponding to the greatest eigenvalue of the adjacency matrix can be used to compute the eigenvector centrality measure of the nodes in the graph.

Behavior-aware Criterion The behavior-aware criterion is devised to mirror human-like trajectory predictions by leveraging the analytical properties of centrality measures. This aids in detecting and comprehending human driving behavior. By doing so, it removes the necessity for manual labeling, addressing issues like labeling non-continuous behaviors and choosing optimal time frames. Furthermore, this criterion effectively encapsulates continuous driving behaviors. Incorporating this with the Behavior Likelihood Estimate (BLE) and Behavior Intensity Estimate (BIE) refines prediction accuracy and dependability in fluctuating and intricate traffic conditions.

Behavior Likelihood Estimate. The BLE criterion quantifies behavior probabilities using time-based row derivatives, even without explicit behavior classifications. A higher probability of a behavior is indicated by prominent row derivatives and local extrema. For v_i^t at time t, the BLE considering all three centrality measures is as follows:

$$\mathcal{I}_i^t = \left[\left|\frac{\partial \mathcal{J}_i^t(D)}{\partial t}\right|, \left|\frac{\partial \mathcal{J}_i^t(C)}{\partial t}\right|, \left|\frac{\partial \mathcal{J}_i^t(E)}{\partial t}\right|\right]^T \tag{12}$$

where $|\cdot|$ denotes the absolute value operator.

Behavior Intensity Estimate. The BIE quantifies the potential impact intensity of a driving behavior on surrounding vehicles. It takes into account the duration of the behavior, with longer-lasting behaviors assumed to have a greater impact than those that are brief. The BIE for the node v_i^t at time t is built on top of BLE, and is defined as:

$$\mathcal{L}_i^t = \left|\frac{\partial \mathcal{I}_i^t}{\partial t}\right| = \left[\left|\frac{\partial^2 \mathcal{J}_i^t(D)}{\partial^2 t}\right|, \left|\frac{\partial^2 \mathcal{J}_i^t(C)}{\partial^2 t}\right|, \left|\frac{\partial^2 \mathcal{J}_i^t(E)}{\partial^2 t}\right|\right]^T \tag{13}$$

In summary, the BIE in conjunction with the BLE as prior knowledge provides a comprehensive understanding of individual driver behavior. This is achieved by segmenting each traffic scene into behavior-aware regions centered around observed agents of the ego vehicle. For these regions, behavioral features are extracted from the traffic agents, including contextual information (as shown in Fig. 1). These features are then embedded and encoded frame by frame by an LSTM network to generate high-dimensional behavior vectors. By combining insights into the probability and intensity of the behavior, the overall impact on the surrounding traffic is determined. This infusion of human-like reasoning aligns with human perception and cognition, improving the accuracy and efficiency of trajectory prediction for AVs.

Interaction-aware Module

To capture and assemble social interactions between the ego vehicle and its surrounding agents, we introduce an innovative interaction-aware pooling mechanism. This module includes a hierarchical LSTM encoder and a position encoding layer. The LSTM encoder processes recent historical trajectories for both ego and observed agents, updating hidden features with shared LSTM weights frame by frame. The features are then represented in polar coordinates and mapped through the position encoding layer to capture higher-order interactive information.

Priority-aware Module

The priority-aware module uses an attention mechanism layer to compute dynamic attention weight vectors for surrounding agents based on their higher-order interactive information. This attention mechanism (Vaswani et al. 2017) assigns weights that indicate their importance in predicting the ego vehicle's trajectory. These weight vectors express the relative importance of the agents and are used to weight higher-order interaction data in later stages. They are then fed into a multi-layer perceptron (MLP) to produce high-dimensional aggregate pooling vectors through a max-pooling layer.

Position-aware Module

To further enhance the modeling of temporal dependencies and spatial relationships, this module employs a dedicated LSTM network to encode and learn the dynamic position of the ego vehicle. The historical trajectory of the ego vehicle is also represented in polar coordinates and subsequently embedded using an LSTM. This refinement enhances the model's ability to detail the agent's trajectory.

Decoder

The position vector of the ego vehicle is integrated with additional information about the hidden pooling vectors and the high-dimensional behavior vector. This composite undergoes embedding by a softmax activation function (behavior embedding), followed by processing by an MLP (behavior encoding). Finally, the processed input is analyzed by the LSTM decoder, which generates a probability distribution over the possible future trajectories of the ego vehicle.

Experiments

We evaluate the effectiveness of our model using four datasets: NGSIM (Deo and Trivedi 2018), HighD (Krajewski et al. 2018), RounD (Krajewski et al. 2020), and MoCAD. These datasets, sourced from varied and intricate real-world traffic situations like highways, roundabouts, and urban locales, serve as a comprehensive testing ground. To gauge our model's precision, we employed the Root Mean Square Error (RMSE) metric.

Experimental Setup

These data sets were partitioned into training, validation, and test sets using standard sampling. We refer to the complete test set as the *overall* test set. The trajectories for the NGSIM, HighD, and MoCAD datasets were divided into 8-second intervals. The first 3 seconds served as the trajectory history ($t_h = 3$) for input, and the following 5 seconds represented the ground truth ($t_f = 5$) for output. For the RounD dataset, the trajectories were divided into 6-second chunks with $t_h = 2$ and $t_f = 4$. To delve deeper into our model's performance, the NGISM dataset was further split based on distinct vehicular maneuvers, including no lane-change (*keep*), on-ramp lane merging (*merge*), right lane-change (*right*), and left lane-change (*left*). This subset, termed the *maneuver-based* test set, allows for a granular examination of our model's capabilities across different traffic actions.

Training and Implementation Details

Our model is trained to converge using an NVIDIA A100 40GB GPU. We introduce the Negative Log-Likelihood criterion as a complement to the RMSE in the loss function.

Experimental Results

We evaluate BAT against various SOTA trajectory prediction methods from 2016 to 2023. These include well-known benchmarks such as S-LSTM (Alahi et al. 2016), S-GAN (Gupta et al. 2018), CS-LSTM (Deo and Trivedi 2018), S-GAN (Gupta et al. 2018), MATF-GAN (Zhao et al. 2019), NLS (Messaoud et al. 2019), DN-IRL (Fernando et al. 2019), DRBP(Gao et al. 2023), WSiP (Wang et al. 2023), CF-LSTM (Xie et al. 2021), MHA (Messaoud et al. 2021), MATH (Hasan et al. 2021), HMNet (Xue et al. 2021), EA-Net (Cai et al. 2021), TS-GAN (Wang et al. 2022b), STDAN (Chen et al. 2022b), and iNATran (Chen et al. 2022a). The results, displayed in Table 1, highlight our model's significant advancements in trajectory prediction over current SOTA baselines. Using RMSE as the evaluation metric, our model surpasses recent baselines (2021-2023) by 2.6% for short-term predictions (1s-3s) and reduces prediction error by 56.7% for long-term predictions (4s-5s) on the NGSIM dataset. On the HighD dataset, known for its superior data volume and precision, our model significantly outperforms most baselines, showing improvements of 62.7% and 43.6% compared to STDAN and iNATran, respectively, over a 5-second horizon.

The strengths of BAT become more evident in complex scenarios, like urban streets and unstructured roads (RounD

Dataset	Model	Prediction Horizon (s)				
		1	2	3	4	5
NGSIM	S-LSTM	0.65	1.31	2.16	3.25	4.55
	S-GAN	0.57	1.32	2.22	3.26	4.40
	CS-LSTM	0.61	1.27	2.09	3.10	4.37
	MATF-GAN	0.66	1.34	2.08	2.97	4.13
	NLS	0.56	1.22	2.02	3.03	4.30
	DRBP	1.18	2.83	4.22	5.82	-
	DN-IRL	0.54	1.02	1.91	2.43	3.76
	WSiP	0.56	1.23	2.05	3.08	4.34
	CF-LSTM	0.55	1.10	1.78	2.73	3.82
	MHA	0.41	1.01	1.74	2.67	3.83
	HMNet	0.50	1.13	1.89	2.85	4.04
	TS-GAN	0.60	1.24	1.95	2.78	3.72
	STDAN	0.39	0.96	1.61	2.56	3.67
	BAT (25%)	0.31	0.85	1.65	2.69	3.87
	BAT	**0.23**	**0.81**	**1.54**	**2.52**	**3.62**
HighD	S-LSTM	0.22	0.62	1.27	2.15	3.41
	S-GAN	0.30	0.78	1.46	2.34	3.41
	WSiP	0.20	0.60	1.21	2.07	3.14
	CS-LSTM	0.22	0.61	1.24	2.10	3.27
	MHA	0.19	0.55	1.10	1.84	2.78
	NLS	0.20	0.57	1.14	1.90	2.91
	DRBP	0.41	0.79	1.11	1.40	-
	EA-Net	0.15	0.26	0.43	0.78	1.32
	CF-LSTM	0.18	0.42	1.07	1.72	2.44
	STDAN	0.19	0.27	0.48	0.91	1.66
	iNATran	**0.04**	**0.05**	0.21	0.54	1.10
	BAT (25%)	0.14	0.34	0.65	0.89	1.27
	BAT	0.08	0.14	**0.20**	**0.44**	**0.62**
RounD	S-LSTM	0.94	1.82	3.43	5.21	-
	S-GAN	0.72	1.57	3.01	4.78	-
	CS-LSTM	0.71	1.21	2.09	3.92	-
	MATH	0.38	0.80	1.76	3.08	-
	MHA	0.62	0.98	1.88	3.65	-
	NLS	0.62	0.96	1.91	3.48	-
	WSiP	0.52	0.99	1.88	3.07	-
	CF-LSTM	0.51	0.87	1.79	3.14	-
	STDAN	0.35	0.77	1.74	2.92	-
	BAT (25%)	0.32	0.72	1.99	3.12	-
	BAT	**0.23**	**0.55**	**1.43**	**2.46**	-
MoCAD	S-LSTM	1.73	2.46	3.39	4.01	4.93
	S-GAN	1.69	2.25	3.30	3.89	4.69
	CS-LSTM	1.45	1.98	2.94	3.56	4.49
	MHA	1.25	1.48	2.57	3.22	4.20
	NLS	0.96	1.27	2.08	2.86	3.93
	WSiP	0.70	0.87	1.70	2.56	3.47
	CF-LSTM	0.72	0.91	1.73	2.59	3.44
	STDAN	0.62	0.85	1.62	2.51	3.32
	BAT (25%)	0.65	0.99	1.89	2.81	3.58
	BAT	**0.35**	**0.74**	**1.39**	**2.19**	**2.88**

Table 1: Evaluation results for BAT and the baselines in the *overall* test set over a different horizon. Note: RMSE (m) is the evaluation metric, where lower values indicate better performance, with some not specifying ('-'). Values in bold represent the best performance in each category.

and MoCAD datasets). Here, our model consistently surpasses current SOTA baselines, with accuracy gains between 17.8%-75.5% on RounD and 12.7%-79.8% on MoCAD. Such improvements underscore the significance of factoring in driving behavior and our relative distance pooling mechanism, especially in dense traffic scenarios. For scalability testing, even when our model was trained on just 25% of the training data, it still managed to outperform most baselines, indicating a potential reduction in data needs for training AVs in challenging contexts.

Dataset	Model	Prediction Horizon (s)				
		1	2	3	4	5
keep	S-LSTM	0.35	1.01	1.81	2.82	4.15
	S-GAN	0.36	1.01	1.81	2.83	4.15
	CS-LSTM	0.34	0.98	1.75	2.77	4.06
	MATF-GAN	0.37	1.11	1.74	2.66	3.91
	WSiP	0.32	0.89	1.58	2.51	3.59
	HMNet	0.31	0.83	1.56	2.51	3.68
	STDAN	0.28	0.85	1.52	2.53	3.49
	BAT (25%)	0.28	0.86	1.54	2.52	3.73
	BAT	**0.23**	**0.81**	**1.49**	**2.44**	**3.56**
merge	S-LSTM	0.81	1.31	2.51	4.01	5.78
	S-GAN	0.71	1.32	2.53	4.11	5.97
	CS-LSTM	0.61	1.34	2.58	4.12	5.94
	MATF-GAN	0.53	1.41	2.56	3.97	5.52
	WSiP	0.40	1.18	2.41	3.72	5.16
	HMNet	0.34	1.17	2.32	3.63	5.20
	STDAN	0.28	1.19	2.21	3.67	4.95
	BAT (25%)	0.31	0.95	1.95	3.31	4.98
	BAT	**0.25**	**0.89**	**1.83**	**3.04**	**4.45**
left	S-LSTM	0.77	1.68	3.04	4.67	6.59
	S-GAN	0.66	1.68	3.11	4.85	6.8
	CS-LSTM	0.54	1.63	3.01	4.71	6.63
	MATF-GAN	0.61	1.72	3.02	4.62	6.34
	WSiP	0.41	1.46	2.82	4.42	6.22
	HMNet	0.41	1.31	2.87	4.47	6.33
	STDAN	0.35	1.33	2.84	4.51	5.97
	BAT (25%)	0.43	1.24	2.43	4.01	5.91
	BAT	**0.33**	**1.07**	**2.24**	**3.73**	**5.51**
right	S-LSTM	0.69	1.97	3.81	6.17	9.09
	S-GAN	0.72	1.97	3.91	6.32	9.23
	CS-LSTM	0.61	2.01	3.97	6.48	9.48
	MATF-GAN	0.56	1.88	3.90	6.07	9.01
	WSiP	0.52	1.61	3.60	5.78	8.45
	HMNet	0.49	1.62	3.47	5.87	8.59
	STDAN	0.38	1.49	3.46	5.87	7.93
	BAT (25%)	0.47	1.41	3.09	5.19	7.87
	BAT	**0.31**	**1.36**	**2.96**	**5.15**	**6.78**

Table 2: Evaluation results for the proposed model and the baselines in the maneuver-based test set for NGSIM dataset.

We also conducted tests on the *maneuver-based* test set, as detailed in Table 2. Specifically, in the *merge* and *right* test subsets, our model achieves significantly lower RMSE values than the SOTA baselines, demonstrating an improvement of at least 10.1% for a prediction horizon of 5 seconds, which could significantly mitigate the risk of traffic accidents. Moreover, our model shows remarkable improvement in the *keep* and *left* test subsets, highlighting its robustness and effectiveness in accurately predicting future vehicle trajectories in various driving scenarios and maneuvers.

Overall, our findings affirm our model's capability and efficiency in predicting vehicle trajectories for AVs.

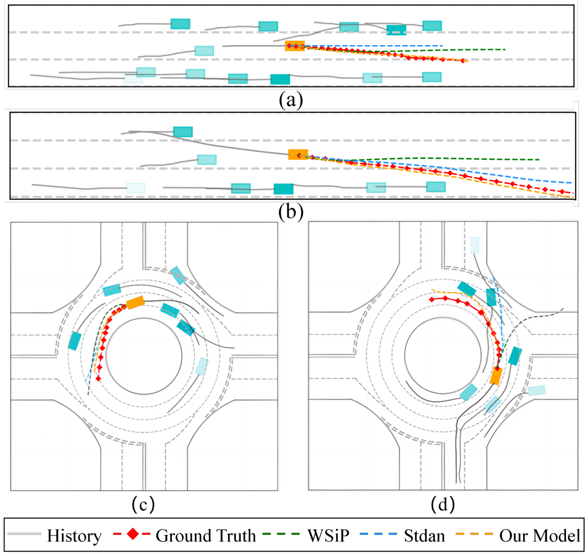

Figure 3: Visualizations and heat maps selected from the NGSIM (a-b) and RounD (c-d) datasets. The target vehicle is depicted in orange, while its surrounding vehicles are shown in blue. The darkness of the blue color indicates the higher importance weight of the surrounding vehicle.

Ablation Studies

Table 3 presents an analysis of four critical components: polar coordinates, behavior-aware, interaction-aware, and priority-aware modules. We tested five models: Model A (using Cartesian coordinates), Model B (excluding the behavior-aware module), Model C (excluding the interaction-aware module), Model D (excluding the priority-aware module), and Model E (with all components).

On evaluating against the NGSIM and RounD datasets, all stripped-down versions (A-D) underperformed compared to the comprehensive Model E. Notably, the integration of interaction-aware and priority-aware modules significantly boosted performance, underlining their importance in enhancing prediction accuracy.

Dataset	Time (s)	Model (ΔMethod E)				
		A	B	C	D	E
NGSIM	1	0.27	0.30	0.27	0.28	**0.23**
	2	0.86	0.89	0.85	0.87	**0.81**
	3	1.63	1.68	1.60	1.63	**1.54**
	4	2.65	2.68	2.62	2.64	**2.52**
	5	4.02	4.08	3.97	3.97	**3.62**
RounD	1	0.76	0.55	0.44	0.35	**0.23**
	2	0.94	0.83	0.76	0.72	**0.55**
	3	1.87	1.72	1.63	1.54	**1.43**
	4	3.02	2.82	2.76	2.68	**2.46**

Table 3: Ablation results for different models on NGSIM and RounD datasets (Evaluation metric: RMSE (m)).

The behavior-aware module's inclusion significantly enhanced performance by capturing dynamic vehicular interactions, vital for accurate trajectory prediction. By factoring in surrounding vehicles' behavior, BAT predicts the ego vehicle's trajectory more insightfully. This mirrors human decision-making, where actions and intentions of other agents, including vehicles, shape trajectory predictions (Baron 2000). Furthermore, adopting the polar coordinate system in Model C outperformed the Cartesian approach, especially in roundabout environments like RounD. This aligns with studies on human perception, suggesting people process goal-relevant information distinctively (Todd and Gigerenzer 2000). The polar system better reflects human cognition of spatial vehicular relationships, emphasizing the significance of both behavioral and spatial considerations in trajectory prediction.

Intuition and Interpretability Analysis

To further underscore the prowess of BAT, we visually dissect its prediction outcomes across diverse scenarios in Fig. 3. For the sake of clarity, we spotlight solely the trajectories deemed most probable for the ego vehicle in each context. We meticulously chose two demanding driving situations: transitioning into the right lane (Fig. 3 (a-b)) and maneuvering through a roundabout (Fig.3 (c-d)). Intriguingly, the heat maps vividly unveil a direct relationship between the proximity of the ego to neighboring vehicles and their respective significance. This exposes pronounced social interplay among the nearby agents. In Fig. 3 (a), the bottom-most vehicle with the red circle exhibits friendly driving behavior, intuitively creating ample space for the ego vehicle's lane transition. On the flip side, the vehicle with the red circle in Fig. 3 (b) manifests aggressive driving tendencies, potentially accelerating to impede the ego vehicle's lane merge. Herein lies the genius of BAT's behavior-aware module: it discerns driver personas, predicting the ego vehicle's inability to seamlessly merge, aligning impeccably with the ground truth. Conversely, models bereft of this driving behavior consideration falter, such as Stdan and WSiP, deviating significantly from the actual trajectory.

In addition, BAT captures the influence of agents even from non-adjacent lanes, attributing this to their distinct driving behavior—a facet frequently sidestepped in conventional studies. To sum up, BAT doesn't just predict; it observes, interprets, and decides like a human. By mirroring human decision-making, BAT offers a promising leap toward autonomous driving that's both accurate and reliable.

Conclusion

Predicting the trajectories of surrounding vehicles with a high degree of accuracy is a fundamental challenge that must be addressed in the quest for full AVs. To address this challenge, we propose a behavior-aware modular model with four components: behavior-aware, interaction-aware, priority-aware, and position-aware modules. Our model outperforms current SOTA baselines in terms of prediction accuracy and efficiency on the NGSIM, HighD, RounD, and MoCAD datasets, even when trained on 25% training set, demonstrating its robustness, applicability, and potential to reduce training data requirements for AVs in challenging or unusual situations such as corner cases, roundabouts.

Acknowledgements

This research is supported by the Science and Technology Development Fund of the Macau SAR (File No. 0021/2022/ITP, 0081/2022/A2, 0015/2019/AKP, SKL-IoTSC(UM)-2021-2023/ORP/GA08/2022, SKL-IoTSC(UM)-2024-2026/ORP/GA06/2023) and the University of Macau (SRG2023-00037-IOTSC).

References

Alahi, A.; Goel, K.; Ramanathan, V.; Robicquet, A.; Fei-Fei, L.; and Savarese, S. 2016. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 961–971.

Baron, J. 2000. *Thinking and deciding*. Cambridge University Press.

Bhattacharyya, P.; Huang, C.; and Czarnecki, K. 2023. Ssl-lanes: Self-supervised learning for motion forecasting in autonomous driving. In *Conference on Robot Learning*, 1793–1805. PMLR.

Cai, Y.; Wang, Z.; Wang, H.; Chen, L.; Li, Y.; Sotelo, M. A.; and Li, Z. 2021. Environment-attention network for vehicle trajectory prediction. *IEEE Transactions on Vehicular Technology*, 70(11): 11216–11227.

Chandra, R.; Bhattacharya, U.; Mittal, T.; Bera, A.; and Manocha, D. 2020. CMetric: A Driving Behavior Measure Using Centrality Functions. *arXiv preprint arXiv:2003.04424*.

Chen, X.; Zhang, H.; Zhao, F.; Cai, Y.; Wang, H.; and Ye, Q. 2022a. Vehicle trajectory prediction based on intention-aware non-autoregressive transformer with multi-attention learning for Internet of Vehicles. *IEEE Transactions on Instrumentation and Measurement*, 71: 1–12.

Chen, X.; Zhang, H.; Zhao, F.; Hu, Y.; Tan, C.; and Yang, J. 2022b. Intention-aware vehicle trajectory prediction based on spatial-temporal dynamic attention network for Internet of Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(10): 19471–19483.

Dall, J.; and Christensen, M. 2002. Random geometric graphs. *Physical review E*, 66(1): 016121.

Deo, N.; and Trivedi, M. M. 2018. Convolutional social pooling for vehicle trajectory prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 1468–1476.

Fernando, T.; Denman, S.; Sridharan, S.; and Fookes, C. 2019. Neighbourhood context embeddings in deep inverse reinforcement learning for predicting pedestrian motion over long time horizons. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 0–0.

Freeman, L. C. 1978. Centrality in social networks conceptual clarification. *Social networks*, 1(3): 215–239.

Gao, K.; Li, X.; Chen, B.; Hu, L.; Liu, J.; Du, R.; and Li, Y. 2023. Dual Transformer Based Prediction for Lane Change Intentions and Trajectories in Mixed Traffic Environment. *IEEE Transactions on Intelligent Transportation Systems*.

Gupta, A.; Johnson, J.; Fei-Fei, L.; Savarese, S.; and Alahi, A. 2018. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2255–2264.

Hang, P.; Lv, C.; and Chen, X. 2022. *Human-Like Decision Making and Control for Autonomous Driving*. CRC Press.

Hasan, M.; Solernou, A.; Paschalidis, E.; Wang, H.; Markkula, G.; and Romano, R. 2021. Maneuver-aware pooling for vehicle trajectory prediction. *arXiv preprint arXiv:2104.14079*.

Huang, Y.; Du, J.; Yang, Z.; Zhou, Z.; Zhang, L.; and Chen, H. 2022. A Survey on Trajectory-Prediction Methods for Autonomous Driving. *IEEE Transactions on Intelligent Vehicles*.

Huang, Z.; Mo, X.; and Lv, C. 2022. Multi-modal Motion Prediction with Transformer-based Neural Network for Autonomous Driving. In *2022 International Conference on Robotics and Automation (ICRA)*, 2605–2611.

Krajewski, R.; Bock, J.; Kloeker, L.; and Eckstein, L. 2018. The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2118–2125.

Krajewski, R.; Moers, T.; Bock, J.; Vater, L.; and Eckstein, L. 2020. The rounD Dataset: A Drone Dataset of Road User Trajectories at Roundabouts in Germany. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 1–6.

Li, G.; Qiu, Y.; Yang, Y.; Li, Z.; Li, S.; Chu, W.; Green, P.; and Li, S. E. 2022. Lane change strategies for autonomous vehicles: a deep reinforcement learning approach based on transformer. *IEEE Transactions on Intelligent Vehicles*.

Li, Z.; Chen, Z.; Li, Y.; and Xu, C. 2023a. Context-aware trajectory prediction for autonomous driving in heterogeneous environments. *Computer-Aided Civil and Infrastructure Engineering*.

Li, Z.; Liao, H.; Tang, R.; Li, G.; Li, Y.; and Xu, C. 2023b. Mitigating the impact of outliers in traffic crash analysis: A robust Bayesian regression approach with application to tunnel crash data. *Accident Analysis & Prevention*, 185: 107019.

Liang, M.; Yang, B.; Hu, R.; Chen, Y.; Liao, R.; Feng, S.; and Urtasun, R. 2020. Learning lane graph representations for motion forecasting. In *European Conference on Computer Vision*, 541–556. Springer.

Liao, H.; Shen, H.; Li, Z.; Wang, C.; Li, G.; Bie, Y.; and Xu, C. 2023. GPT-4 Enhanced Multimodal Grounding for Autonomous Driving: Leveraging Cross-Modal Attention with Large Language Models. *arXiv preprint arXiv:2312.03543*.

Lin, C.-F.; Ulsoy, A. G.; and LeBlanc, D. J. 2000. Vehicle dynamics and external disturbance estimation for vehicle path prediction. *IEEE Transactions on Control Systems Technology*, 8(3): 508–518.

Markkula, G.; Madigan, R.; Nathanael, D.; Portouli, E.; Lee, Y. M.; Dietrich, A.; Billington, J.; Schieben, A.; and Merat, N. 2020. Defining interactions: A conceptual framework for understanding interactive behaviour in human and automated road traffic. *Theoretical Issues in Ergonomics Science*, 21(6): 728–752.

Messaoud, K.; Yahiaoui, I.; Verroust-Blondet, A.; and Nashashibi, F. 2019. Non-local social pooling for vehicle trajectory prediction. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, 975–980. IEEE.

Messaoud, K.; Yahiaoui, I.; Verroust-Blondet, A.; and Nashashibi, F. 2021. Attention Based Vehicle Trajectory Prediction. *IEEE Transactions on Intelligent Vehicles*, 6(1): 175–185.

Müller, L.; Risto, M.; and Emmenegger, C. 2016. The social behavior of autonomous vehicles. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 686–689.

Pillai, S. U.; Suel, T.; and Cha, S. 2005. The Perron-Frobenius theorem: some of its applications. *IEEE Signal Processing Magazine*, 22(2): 62–75.

Schwarting, W.; Pierson, A.; Alonso-Mora, J.; Karaman, S.; and Rus, D. 2019. Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences*, 116(50): 24972–24978.

Todd, P. M.; and Gigerenzer, G. 2000. Précis of simple heuristics that make us smart. *Behavioral and brain sciences*, 23(5): 727–741.

Toghi, B.; Valiente, R.; Sadigh, D.; Pedarsani, R.; and Fallah, Y. P. 2022. Social coordination and altruism in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(12): 24791–24804.

Toledo, T.; Musicant, O.; and Lotan, T. 2008. In-vehicle data recorders for monitoring and feedback on drivers' behavior. *Transportation Research Part C: Emerging Technologies*, 16(3): 320–331.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, R.; Wang, S.; Yan, H.; and Wang, X. 2023. WSiP: Wave Superposition Inspired Pooling for Dynamic Interactions-Aware Trajectory Prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 4685–4692.

Wang, W.; Wang, L.; Zhang, C.; Liu, C.; Sun, L.; et al. 2022a. Social interactions for autonomous driving: A review and perspectives. *Foundations and Trends® in Robotics*, 10(3-4): 198–376.

Wang, Y.; Wang, C.; Zhao, W.; and Xu, C. 2021. Decision-Making and planning method for autonomous vehicles based on motivation and risk assessment. *IEEE Transactions on Vehicular Technology*, 70(1): 107–120.

Wang, Y.; Zhao, S.; Zhang, R.; Cheng, X.; and Yang, L. 2022b. Multi-Vehicle Collaborative Learning for Trajectory Prediction With Spatio-Temporal Tensor Fusion. *IEEE Transactions on Intelligent Transportation Systems*, 23(1): 236–248.

Wong, C.; Xia, B.; Hong, Z.; Peng, Q.; Yuan, W.; Cao, Q.; Yang, Y.; and You, X. 2022. View Vertically: A hierarchical network for trajectory prediction via fourier spectrums. In *European Conference on Computer Vision*, 682–700. Springer.

Xie, G.; Shangguan, A.; Fei, R.; Ji, W.; Ma, W.; and Hei, X. 2020. Motion trajectory prediction based on a CNN-LSTM sequential model. *Science China Information Sciences*, 63(11): 1–21.

Xie, X.; Zhang, C.; Zhu, Y.; Wu, Y. N.; and Zhu, S.-C. 2021. Congestion-aware multi-agent trajectory prediction for collision avoidance. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 13693–13700. IEEE.

Xue, Q.; Li, S.; Li, X.; Zhao, J.; and Zhang, W. 2021. Hierarchical Motion Encoder-Decoder Network for Trajectory Forecasting. *arXiv preprint arXiv:2111.13324*.

Ye, M.; Cao, T.; and Chen, Q. 2021. Tpcn: Temporal point cloud networks for motion forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11318–11327.

Zeng, W.; Li, M.; Xiong, W.; Tong, T.; Lu, W.-j.; Tan, J.; Wang, R.; and Huang, R. 2023. Mpcvit: Searching for accurate and efficient mpc-friendly vision transformer with heterogeneous attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5052–5063.

Zhang, K.; and Li, L. 2022. Explainable multimodal trajectory prediction using attention models. *Transportation Research Part C: Emerging Technologies*, 143: 103829.

Zhao, T.; Xu, Y.; Monfort, M.; Choi, W.; Baker, C.; Zhao, Y.; Wang, Y.; and Wu, Y. N. 2019. Multi-agent tensor fusion for contextual trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12126–12134.