MFTraj: Map-Free, Behavior-Driven Trajectory Prediction for Autonomous Driving

Haicheng Liao[1] , **Zhenning Li**[1] , **Chengyue Wang**[1] , **Huanming Shen**[2] , **Dongping Liao**[1] , **Bonan Wang**[1] , **Guofa Li**[3] , **Chengzhong Xu**[1]
[1]University of Macau
[2]University of Electronic Science and Technology of China
[3]Chongqing University
{yc27979, zhenningli, chengyuewang, yb97428, mc35002, czxu}@um.edu.com,
HuanmingShen@outlook.com.com, hanshan198@gmail.com

Abstract

This paper introduces a trajectory prediction model tailored for autonomous driving, focusing on capturing complex interactions in dynamic traffic scenarios without reliance on high-definition maps. The model, termed MFTraj, harnesses historical trajectory data combined with a novel dynamic geometric graph-based behavior-aware module. At its core, an adaptive structure-aware interactive graph convolutional network captures both positional and behavioral features of road users, preserving spatial-temporal intricacies. Enhanced by a linear attention mechanism, the model achieves computational efficiency and reduced parameter overhead. Evaluations on the Argoverse, NGSIM, HighD, and MoCAD datasets underscore MFTraj's robustness and adaptability, outperforming numerous benchmarks even in data-challenged scenarios without the need for additional information such as HD maps or vectorized maps. Importantly, it maintains competitive performance even in scenarios with substantial missing data, on par with most existing state-of-the-art models. The results and methodology suggest a significant advancement in autonomous driving trajectory prediction, paving the way for safer and more efficient autonomous systems.

1 Introduction

The integration of autonomous vehicles (AVs) with human-driven vehicles and pedestrians necessitates advanced trajectory prediction models. Central to these models is their ability to predict the future trajectories of various road users, leveraging historical data. Despite significant advancements, a pivotal challenge persists: modeling the often unpredictable *driving behaviors* of road users.

These behaviors, shaped by a blend of traffic dynamics, road layouts, and individual cognitive inclinations, offer a unique window into the real-time decision-making processes of humans in complex traffic settings [Schwarting *et al.*, 2019]. Our research has illuminated the pivotal role of understanding human behavioral patterns in trajectory predictions. Recognizing and predicting human driving behavior is not merely about tracing a vehicle's path; it's about understanding the cognitive processes that dictate those paths. By understanding behaviors, AVs can anticipate sudden changes in human-driven vehicles or pedestrian movements, leading to safer co-navigation. Furthermore, behavior-focused predictions can aid in scenarios where traditional data might be ambiguous or incomplete, relying on human behavioral patterns to fill in the gaps. Through the integration of decision-making theories, cognitive psychology, and traffic behavior studies [Yin *et al.*, 2021], trajectory prediction models can be enriched, fostering a harmonious coexistence of AVs and human-driven entities on the road.

High Definition (HD) maps, conventionally considered pivotal for trajectory prediction, pose intrinsic challenges. Their creation is resource-intensive, and in the rapidly changing milieu of urban environments, they can quickly become obsolete [Gao *et al.*, 2020]. This has given rise to *map-free* models, a paradigm shift that operates independently of HD map data. However, while these models adeptly handle dynamic environments, they may lack the granularity provided by comprehensive road network data. This gap is aptly addressed by the advent of deep learning techniques, notably Graph Neural Networks (GNNs) [Liang *et al.*, 2020; Gao *et al.*, 2020]. GNNs, adept at assimilating extensive data from road users, offer nuanced insights into their interactions and the overarching socio-cognitive context, thereby compensating for the lack of detailed HD maps.

Our contributions are as follows:

1. An **advanced map-free architecture** for trajectory prediction that obviates the need for HD maps, resulting in significant computational savings.
2. A novel dynamic geometric graph that captures the **essence of continuous driving behavior**, circumventing the limitations of manual labeling. We have integrated behavioral metrics and criteria, drawing from traffic psychology, cognitive neuroscience, and decision-making frameworks, to craft a model that offers more than mere predictions—it elucidates.
3. Benchmark assessments underscore MFTraj's superior-

ity. Demonstrating a commendable **performance elevation of nearly 5.0%** on the Argoverse, NGSIM, HighD, and MoCAD datasets, its robustness is further accentuated by its consistent performance even with a data shortfall of 25%- 62.5%, underscoring its adaptability and profound understanding of diverse traffic dynamics.

2 Related Work

Recent years have seen an explosion of research in trajectory prediction for autonomous driving (AD), thanks to the burgeoning field of deep learning. These cutting-edge approaches [Liao *et al.*, 2024a; Messaoud *et al.*, 2021; Tian *et al.*, 2024; Liao *et al.*, 2024e] have demonstrated superior performance in complex traffic scenarios. However, they often encounter challenges in adequately representing spatial relationships, such as graphic inputs of the scene. To address this, HD maps, rich in scene and semantic information, have attracted increasing research attention. Considering that Convolutional Neural Networks (CNNs) excel at extracting spatial features from input data, such as spatial features from inputs like vectorized maps or raster images, several studies [Zhao *et al.*, 2021; Gilles *et al.*, 2021; Khandelwal *et al.*, 2020] have merged sequential networks with CNNs. This hybrid approach effectively captures both temporal and spatial features from HD maps, providing enriched contextual information for motion prediction tasks. Recent research has explored Graph Neural Networks (GNNs) [Liang *et al.*, 2020; Mohamed *et al.*, 2020; Liao *et al.*, 2024c], such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs), Transformers [Zhou *et al.*, 2022; Liao *et al.*, 2024d; Chen *et al.*, 2022a; Ngiam *et al.*, 2022], and generative models such as Adversarial Networks (GANs) [Zhou *et al.*, 2023] and Variational Auto Encoders (VAEs) [Walters *et al.*, 2021] for direct encoding of HD maps. For example, VectorNet [Gao *et al.*, 2020] simplifies maps by extracting key points from lane splines and encoding them using GNNs. Moreover, LaneGCN [Liang *et al.*, 2020] and TPCN [Ye *et al.*, 2021] build lane graphs using central line segments, employing GCN to capture dynamic interaction. In addition, HiVT [Zhou *et al.*, 2022] and SSL-Lanes [Bhattacharyya *et al.*, 2023] represent map elements by relative positions, improving the transformer model for time series trajectory data.

Despite their effectiveness, the limited availability of HD maps and the extensive resources needed for their creation and maintenance impede the widespread adoption of autonomous driving, particularly in areas lacking current HD map coverage. In response to these challenges, this study introduces a map-free model that utilizes generative models and a VRNN [Chung *et al.*, 2015] to account for the variable nature of traffic scenarios. We propose a novel adaptive GCN to model the complexity of real-time interactions in traffic scenes. To streamline model complexity, we apply the Linformer framework [Wang *et al.*, 2020] for a balance between computational efficiency and prediction accuracy in AD applications.

3 Methodologies

3.1 Inputs and Outputs

This study focuses on predicting the trajectory of the *target vehicle* in interactive driving scenarios, considering all vehicles within the AV's (termed the *target vehicle*) sensing range. At time t, the ego vehicle anticipates the target vehicle's trajectory for the upcoming t_f steps. Our model, drawing from historical data $\boldsymbol{X}$, considers past trajectories of both the target vehicle (indexed by 0) and its surrounding agents (indexed from 1 to n) over a predefined horizon t_h. Formally,

$$\boldsymbol{X} = \left\{\boldsymbol{X}_0^{t-t_h:t}; \boldsymbol{X}_i^{t-t_h:t}\ \forall i \in [1,n]\right\} \tag{1}$$

where $\boldsymbol{X}_0^{t-t_h:t} = \{x_0^{t-t_h}, x_0^{t-t_h+1}, \ldots, x_0^t\}$ and $\boldsymbol{X}_i^{t-t_h:t} = \{x_i^{t-t_h}, x_i^{t-t_h+1}, \ldots, x_i^t\}$ represent the historical trajectories of the target vehicle and those of the surrounding agents from time $t - t_h$ to t, respectively.

The output of the model is the future trajectory of the target vehicle during the prediction horizon t_f:

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

where y_0^t is the 2D coordinates of the target vehicle at time t.

Our model uniquely operates without relying on maps, using only the historical trajectory data of the target and nearby agents. The model needs an input sequence of length t_h and remains functional even if the historical data is not perfectly sequential. For sporadic missing data points, due to reasons like occlusions or sensor glitches, we employ simple linear imputation or similar methods. In addition, Figure 1 illustrates our proposed model's hierarchical design. Following the encoder-decoder format, it features four key components: behavior-aware, position-aware, interaction-aware modules, and the residual decoder. We delve into each module's specifics below.

3.2 Behavior-aware Module

Moving away from traditional methods that classify driver behaviors into fixed and discrete categories, we offer a more adaptable and flexible solution with our behavior-aware module, which utilizes a continuous portrayal of behavioral attributes. This approach draws inspiration from the multi-policy decision-making framework [Markkula *et al.*, 2020], integrating elements of traffic psychology and dynamic geometric graphs (DGGs) [Boguna *et al.*, 2021] to effectively capture intricate driving behaviors amid ongoing driving maneuvers and evolving traffic conditions.

Dynamic Geometric Graphs. We first model the interactions of different agents with a DGG. At time t, the graph G^t is defined as:

$$G^t = \{V^t, E^t\} \tag{3}$$

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 agent, $E^t = \{e_0^t, e_1^t \ldots, e_n^t\}$ is the set of edges representing potential interactions between agents, and e_i^t is the edge between the node v_i^t and other agents that have potential influences with it. An interaction is assumed to exist only if two agents, e.g., v_i and v_j, are in close proximity to each other, i.e., their shortest distance

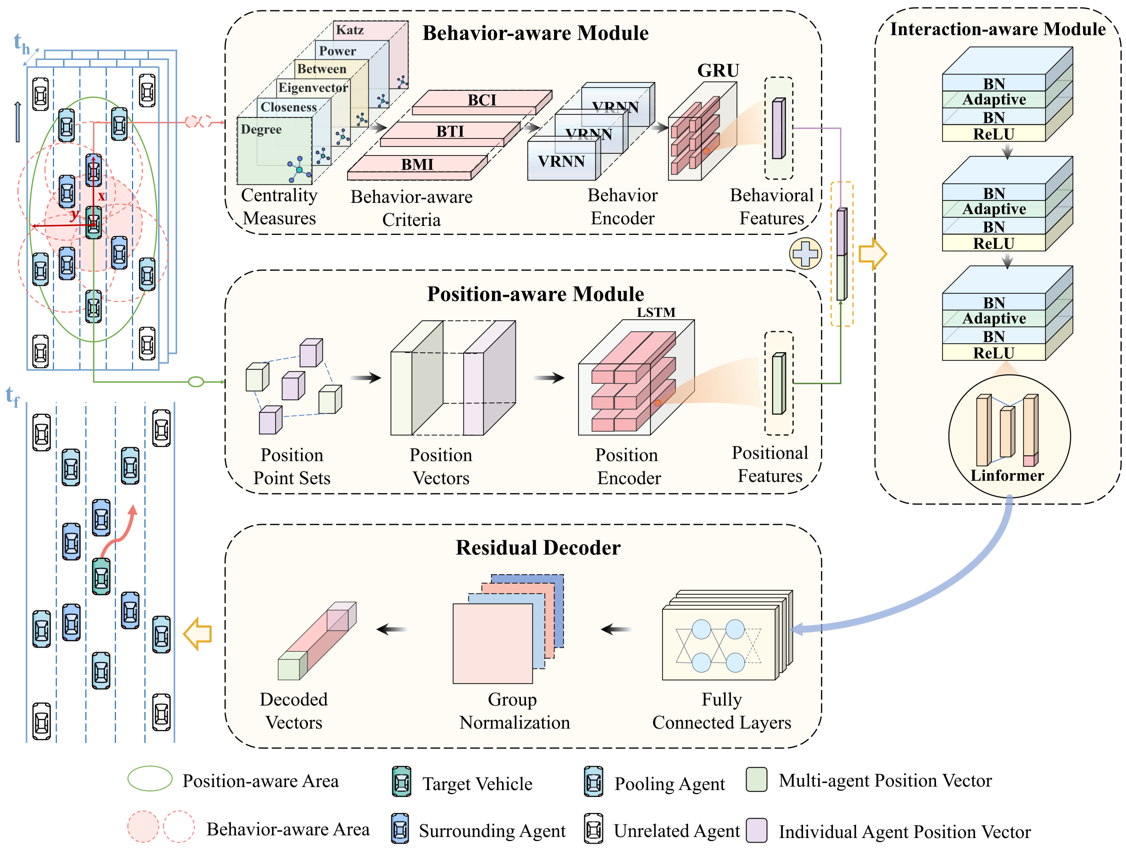

Figure 1: Architecture of the proposed trajectory prediction model.

$d\left(v_i^t, v_j^t\right)$ is less than or equal to a predefined threshold r. Therefore, we define

$$e_i^t = \{v_i^t v_j^t \mid (j \in N_i^t)\} \tag{4}$$

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:

$$A^t(i,j) = \begin{cases} d\left(v_i^t, v_j^t\right) & \text{if } d\left(v_i^t, v_j^t\right) \leq r, i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Centrality Measures. Centrality measures are graph-theoretic metrics that are widely used to understand various aspects of network structures. These measures provide valuable insights into the importance, influence, and connectivity of nodes or vertices within a graph. As shown in Table 1, we use six centralities to characterize driving behavior. These measures allow evaluation of individual agent behavior within the DGG and reveal key agents and overall connectivity of the traffic graph.

(1) Degree Centrality: Reflects the number of connections an agent has, correlating with its influence on and susceptibility to others. It's defined as:

$$J_i^t(D) = \left|N_i^t\right| + J_i^{t-1}(D) \tag{6}$$

where $|\mathcal{N}i^t|$ denotes the total elements in N_i^t.

(2) Closeness Centrality: Indicates an agent's reachability, suggesting its potential influence over others. Defined by:

$$J_i^t(C) = \frac{|N_i^t| - 1}{\sum_{\forall v_j^t \in N_i^t} d\left(v_i^t, v_j^t\right)} \tag{7}$$

(3) Eigenvector Centrality: Measures the importance of an agent by considering both the quantity and quality of connections. Expressed as:

$$J_i^t(E) = \frac{\sum_{\forall v_j^t \in N_i^t} d\left(v_i^t, v_j^t\right)}{\lambda} \tag{8}$$

where λ is the eigenvalue, indicating the collective influence exerted by an agent and its network.

(4) Betweenness Centrality: Highlights agents that act as bridges or bottlenecks in traffic, crucial in congested situations. Formulated as:

$$J_i^t(B) = \sum_{\forall v_s^t, v_k^t \in V^t} \frac{\sigma_{j,k}(v_i^t)}{\sigma_{j,k}} \tag{9}$$

where V^t denotes the set of all agents present in the scene, $\sigma_{j,k}$ signifies the total number of shortest paths between agent v_j^t and agent v_k^t, and $\sigma_{j,k}(v_i)$ represents the number of those paths traversing the agent v_i^t.

(5) Power Centrality: Identifies agents in recurrent interactions, hinting at traffic patterns. Defined by:

$$J_i^t(P) = \sum_k \frac{A_{ii}^k}{k!} \tag{10}$$

Centrality Measures	Magnitude (Original Measure)	Gradient (1st Derivative)	Curvature (2nd Derivative)
Degree ($\mathcal{J}_i^t(D)$) Closeness ($\mathcal{J}_i^t(C)$)	Agent's potential and capability for interaction in the traffic environment	Agent's sensitivity to traffic density variations	Driver's capability to react to fluctuating traffic conditions
Eigenvector ($\mathcal{J}_i^t(E)$) Betweenness ($\mathcal{J}_i^t(B)$)	Agent's significance in dynamic traffic scenarios	Variation in agent's importance in dynamic traffic scenes	Influence of driver behavior alterations on overall traffic conditions
Power ($\mathcal{J}_i^t(P)$) Katz ($\mathcal{J}_i^t(K)$)	Extent of influence an agent exerts on others via direct and indirect interactions at time t	Agent's adaptability to shifts in driving behaviors	Agent's capacity to modify interactions in complex and congested traffic scenarios

Table 1: Centrality measures and their interpretations.

where A_{ii}^k denotes the i-th diagonal element of the adjacency matrix raised to the k-th power, while $k!$ signifies the factorial of k, shedding light on its contribution to the network's structural integrity and dynamism.
(6) Katz Centrality: Emphasizes both direct and distant interactions of an agent, capturing intricate driving patterns. Given as:

$$J_i^t(K) = \sum_k \sum_j \alpha^k A_{ij}^k + \beta^k, \forall i, j \in [0, n], \text{ where } \alpha^k < \frac{1}{\lambda_{\max}} \tag{11}$$

where n represents the number of agents in the real-time traffic scenario, α^k is the decay factor, β^k denotes the weight assigned to the immediate surrounding agents, and A_{ij}^k is the i,j-th element of the k-th power of the adjacency matrix.
Behavior-aware Criteria. Inspired by the relationship between velocity, acceleration, and jerk, we introduce behavioral criteria. These criteria, consisting of Behavior Magnitude Index (BMI) $\mathcal{C}_i^t$, Behavior Tendency Index (BTI) $\mathcal{L}_i^t$, and Behavior Curvature Index (BCI) $\mathcal{I}_i^t$, evaluate different driving behaviors for the target vehicle and its surroundings. They compute thresholds, gradients, and concavities of centrality measures that reflect behaviors such as lane changes, acceleration, deceleration, and driving style. We find that behaviors with significant, fluctuating centrality values in short time frames are likely to affect nearby agents, consistent with human risk perception and time-sensitive decision-making. They are respectively given as follows:

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

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

$$\mathcal{I}_i^t = \left[\left|\frac{\partial^2 \mathcal{J}_i^t(D)}{\partial^2 t}\right|, \left|\frac{\partial^2 \mathcal{J}_i^t(C)}{\partial^2 t}\right|, \ldots, \left|\frac{\partial \mathcal{J}_i^t(K)}{\partial^2 t}\right|\right]^T \tag{14}$$

Behavior Encoder. Incorporating behavior-aware criteria, symbolized as $\mathcal{J} = \{\mathcal{C}_{0:n}^{t-t_h:t}, \mathcal{L}_{0:n}^{t-t_h:t}, \mathcal{I}_{0:n}^{t-t_h:t}\}$, our behavior encoder comprises VRNN and GRU components. This encoder succinctly models relationships between random variables across time, yielding precise sequential behavioral features $\bar{O}_{behavior}^{t-t_h:t}$. Formally:

$$\bar{O}_{behavior}^{t-t_h:t} = \phi_{GRU}\left(\phi_{VRNN}(\mathcal{J})\right) \tag{15}$$

where ϕ_{GRU} and ϕ_{VRNN} denote the GRU and VRNN functions. This encoder captures human driving patterns and their temporal dynamics. Next, behavioral features $\bar{O}_{behavior}^{t-t_h:t}$, fuse with positional features from the position-aware module, subsequently processed by the interaction-aware module for comprehensive feature extraction.

3.3 Position-aware Module

Contrary to traditional methods that emphasize absolute positions [Wang *et al.*, 2022a; Gao *et al.*, 2020] or fixed grids [Deo and Trivedi, 2018], our model emphasizes relative positions. The position-aware module captures individual and group spatial dynamics, interpreting the scene's geometric nuances. These insights are then encoded to produce positional features.
Pooling Mechanism. Our pooling mechanism captures dynamic position data from the traffic environment around the target vehicle, utilizing individual $\mathbf{s}_i^{t_k}$ and multi-agent $\mathbf{s}_{i,j}^{t_k}$ position vectors. This strategy gleans historical trajectories and spatial relationships without depending on fixed positions or grids. The relationships are formulated as:

$$\mathbf{s}_i^{t_k} = \{p_i^{t_k} - p_i^{t_k-1}\}, \ \mathbf{p}_{i,j}^{t_k} = \{p_i^{t_k} - p_j^{t_k}\} \tag{16}$$

Position Encoder. The position encoder employs an LSTM to transform discrete position vectors into continuous spatio-temporal representations, thereby enhancing temporal and spatial interactions between agents and map elements. Given the historical position vectors for the target and surrounding agents, it embeds them temporally:

$$\bar{O}_{position}^{t-t_h:t} = \phi_{LSTM}\left(\bar{\mathbf{h}}_i^{t-t_h:t-1}, \mathbf{s}_i^{t-t_h:t-1}, \mathbf{p}_{i,j}^{t-t_h:t-1}\right) \tag{17}$$

where $\bar{O}_{position}^{t-t_h:t-1}$ is the positional features output by the position encoder, and ϕ_{LSTM} denotes the two-layer LSTM encoder, and $\bar{\mathbf{h}}_i^{t-t_h:t-1}$ represents the hidden position state updated by the encoder on a frame-by-frame basis, with the weights of the LSTM shared among all agents.

3.4 Interaction-aware Module

Effective trajectory prediction in complex traffic scenarios hinges upon a system's ability to comprehend and anticipate interactions among vehicles. Classic GCN-based methods, although proficient at encapsulating geometric inter-agent relationships, often exhibit limitations in fluid traffic conditions due to their fixed adjacency matrix configurations. To tackle this, we introduce a novel adaptive structure-aware GCN, taking cues from advancements in crystal graphs and material design. This novel approach stands out by its capability to craft spatial feature matrices dynamically, adjusting to the number of agents observed in real-time, which ensures a more fluid and adaptable response to traffic changes. A graphical illustration of this concept is provided in Figure 2.

Breaking away from conventional models that predominantly lean on distance-based positional features, our design holistically blends continuous behavioral features into

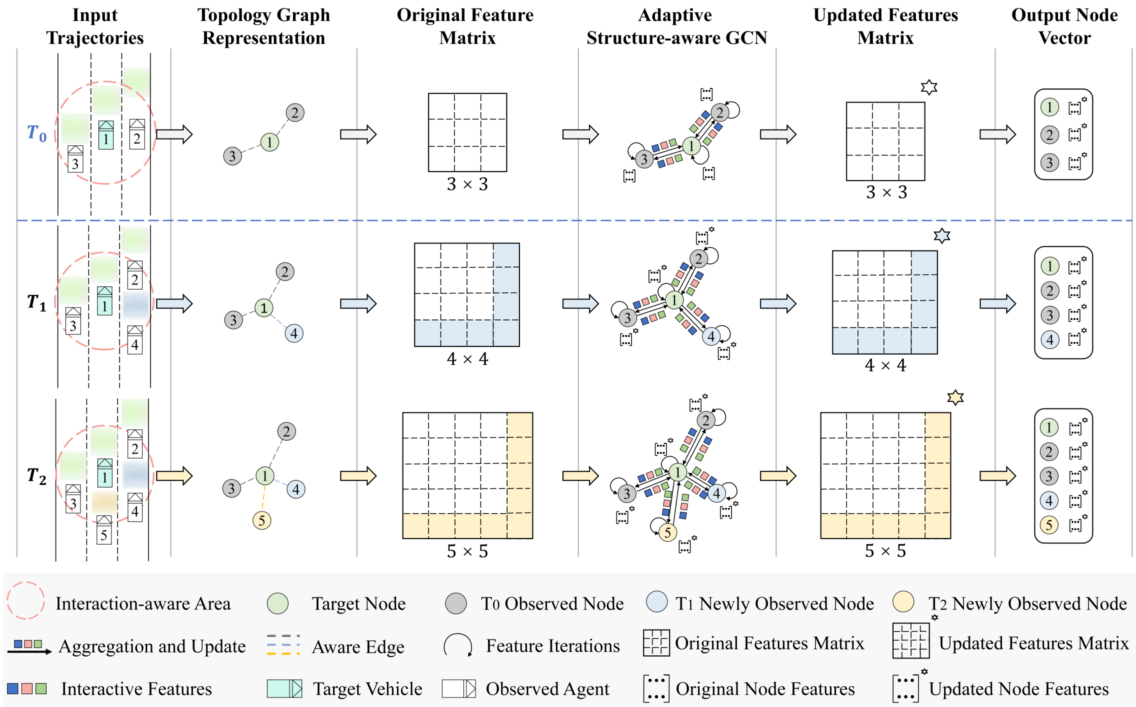

Figure 2: Overview of our adaptive structure-aware GCN. The real-time trajectories of the target and observed agents are captured using a topology graph to form a feature matrix. This matrix undergoes aggregation, updating, and iteration within the GCN. As new agents are observed in real-time, the GCN dynamically adjusts its topology, updating features for the added nodes.

its graph structure. This not only addresses the multifaceted spatio-temporal interactions but also considers the intricate physical interplays between agents, offering a noticeable enhancement in prediction precision. Our design blueprint encompasses an adaptive convolutional neural network rooted in a fully connected interaction multigraph. This structure is adept at simultaneously capturing sequential behavioral and dynamic positional interactions among agents. The multigraph's operational layer is distinguished by nodes, which symbolize sequential behavioral features $\bar{O}_{behavior}^{t-th:t}$ and edges representing positional features $\bar{O}_{position}^{t-th:t}$, as defined below:

$$\begin{aligned}\tilde{\mathbf{z}}_i^k &= F\left(\tilde{\mathbf{z}}_i^{k-1}, \tilde{r}_{i,j}^{k-1}\right)\\ &= \tilde{\mathbf{z}}_i^{k-1} + \phi_{\text{sgm}}\left(\tilde{\mathbf{r}}_{i,j}^{k-1}\mathbf{W}_{\text{g}}^{k-1} + \mathbf{b}_{\text{g}}^{k-1}\right) \odot \phi_{\text{spu}}\left(\tilde{\mathbf{r}}_{i,j}^{k-1}\mathbf{W}_{\text{h}}^{k-1} + \mathbf{b}_{\text{h}}^{k-1}\right)\end{aligned} \tag{18}$$

where the variable k denotes the layer within the GCN, $k \in [1, 3]$, and the symbols $\odot$, ϕ_{sgm}, and ϕ_{spu} represent the element-wise product, sigmoid activation function, and softplus activation function, respectively. Consequently, $\mathbf{W}_{\text{g}}^{k-1}$ and $\mathbf{W}_{\text{h}}^{k-1}$ are learnable matrices, $\mathbf{b}_{\text{g}}^{k-1}$, and $\mathbf{b}_{\text{h}}^{k-1}$ are the bias of the $k-$th layer. $\tilde{r}_{i,j}^{k-1}$ can be represented as follows:

$$\tilde{r}_{i,j}^{k-1} = \left(\tilde{\mathbf{z}}_i^{k-1} \middle\| \tilde{\mathbf{z}}_j^{k-1} \middle\| \mathbf{p}_{i,j}^{t-t_h:t}\right) \tag{19}$$

Additionally, the initial feature vector $\tilde{\mathbf{z}}_i^{(0)}$ is defined as:

$$\tilde{\mathbf{z}}_i^{(0)} = \left(\bar{O}_{behavior}^{t-t_h:t} \| \bar{O}_{position}^{t-t_h:t}\right) \tag{20}$$

Furthermore, the output of the adaptive structure-aware GCN for the target vehicle i is then passed to Linformer, an extension architecture of Transformer,

Furthermore, the output of the adaptive structure-aware GCN for the target vehicle i is subsequently fed into a lightweight transformer-based framework—— Linformer [Wang *et al.*, 2020], to efficiently quantify and compute the dynamic attention weight vectors for the surrounding agents, ultimately output the contextual mapping $\bar{O}$. This allows for a favorable trade-off between accuracy and efficiency.

3.5 Residual Decoder

The residual decoder, comprising a linear residual and projection layer, processes node vectors to forecast the target vehicle's future trajectory, producing the prediction $\mathbf{Y}_0^{t:t+tf}$. This is given by:

$$\boldsymbol{Y} = \mathbf{Y}_{\mathbf{0}}^{\boldsymbol{t:t+t_f}} = F_\theta\left(F_\theta(\bar{O})\right) \tag{21}$$

such that,

$$F_\theta(\cdot) = \phi_{\text{ReLU}}\left[\phi_{\text{GN}}\left(\phi_{\text{Linear}}(\cdot)\right)\right] \tag{22}$$

where ϕ_{ReLU} denotes the ReLU activation function, and ϕ_{GN} denotes the Group Normalization (GN) function [Wu and He, 2018], which is applied to improve the training stability of our model. In addition, the ϕ_{Linear} corresponds to the fully connected layer, while F_θ denotes the residual decoder function.

Model	Input	minADE (m)↓	minFDE (m)↓	MR (%)↓
Argoverse Baseline [Chang *et al.*, 2019]	Map + Traj.	2.96	6.81	81.00
Constant Velocity [Chang *et al.*, 2019]	-	3.55	7.89	75.00
SGAN [Gupta *et al.*, 2018]	-	3.61	5.39	87.11
TPNet [Fang *et al.*, 2020]	Map + Traj.	2.33	5.29	-
PRIME [Song *et al.*, 2022]	Map + Traj.	1.91	3.82	58.67
Uulm-mrm (2rd) [Chang *et al.*, 2019]	Map + Traj.	1.90	4.19	63.47
Jean (1st) [Mercat *et al.*, 2020]	Map + Traj.	1.74	4.24	68.56
WIMP [Khandelwal *et al.*, 2020]	Map + Traj.	1.82	4.03	62.88
Scene-Transformer [Ngiam *et al.*, 2022]	Map + Traj.	1.81	4.06	59.21
TNT [Zhao *et al.*, 2021]	Map + Traj.	1.77	3.91	59.72
mmTransformer [Liu *et al.*, 2021]	Map + Traj.	1.77	4.00	61.78
CtsConv (Aug.) [Walters *et al.*, 2021]	Map + Traj.	1.77	4.05	-
HOME [Gilles *et al.*, 2021]	Map + Traj.	1.72	3.73	58.40
LaneGCN [Liang *et al.*, 2020]	Map + Traj.	1.71	3.78	59.05
GOHOME [Gilles *et al.*, 2022]	Map + Traj.	1.69	3.65	57.21
DenseTNT [Gu *et al.*, 2021]	Map + Traj.	1.68	3.63	58.43
VectorNet [Gao *et al.*, 2020]	Map + Traj.	1.66	3.67	-
TPCN [Ye *et al.*, 2021]	Map + Traj.	1.66	3.69	58.80
SSL-Lanes [Bhattacharyya *et al.*, 2023]	Map + Traj.	1.63	3.56	56.71
LTP [Wang *et al.*, 2022a]	Map + Traj.	1.62	3.55	<u>56.25</u>
HiVT-128 [Zhou *et al.*, 2022]	Map + Traj.	<u>1.60</u>	<u>3.52</u>	-
MFTraj	Traj.	**1.59**	**3.51**	**55.44**
MFTraj (drop 3-frames)	Traj.	1.68	3.59	56.95
MFTraj (drop 5-frames)	Traj.	1.76	3.74	59.08
MFTraj (drop 8-frames)	Traj.	1.86	3.90	61.12
MFTraj (drop 10-frames)	Traj.	1.97	3.96	62.72

Table 2: Performance comparison of various models on *complete* and *missing* datasets for Argoverse. Models use either HD map or vectorized map (Map) and trajectory (Traj.) data or solely Trajectory data, with some not specifying ('-'). Metrics include minADE (k=1), minFDE (k=1), and MR (k=1). Bold and underlined values represent the best and second-best performance in each category.

4 Experiments

4.1 Experimental Setup

Datasets. We tested model's efficacy on Argoverse [Chang *et al.*, 2019], NGSIM [Deo and Trivedi, 2018], HighD [Krajewski *et al.*, 2018], and MoCAD [Liao *et al.*, 2024b] datasets.

Data Segmentation. For Argoverse, we predicted a 3-second trajectory from a 2-second observation, while for NGSIM, HighD, and MoCAD, we use 6-second intervals split into 2 seconds of observation and 4 seconds of prediction. These datasets, referred to as the *complete* dataset, help assess our model in diverse traffic scenarios. Recognizing that real-world conditions often lead to incomplete data, we further assessed our model's resilience using the Argoverse dataset by introducing four subsets with varying levels of missing data: *drop 3-frames*, *drop 5-frames*, *drop 8-frames*, and *drop 10-frames*. These *missing* datasets simulate data loss scenarios. For data gaps, we applied simple linear interpolation.

Metrics. Our experimental protocol was aligned with the Argoverse Motion Forecasting Challenge and prior work [Liao *et al.*, 2024b], we evaluated the performance of our model using standard metrics: minADE, minFDE, MR, and RMSE.

Implementation Details. We implemented our model using PyTorch and PyTorch-lightning on an NVIDIA DGX-2 with eight V100 GPUs. Using the smooth L1 loss as our loss function, the model was trained with the Adam optimizer, a batch size of 32, and learning rates of 10^{-3} and 10^{-4}.

4.2 Experimental Results

Performance Evaluation on the *Complete* Dataset. Tables 2 and Table 4 present a comparative evaluation of our trajectory prediction model against 25 baselines from 2016 to 2023. Unlike most approaches that depend on HD maps or vectorized map data, our model omits map-based inputs. Still, it consistently outperforms the baselines across metrics like minADE, minFDE, MR, and RMSE for both Argoverse and MoCAD datasets. Specifically, for the Argoverse dataset, MFTraj outperforms most of the SOTA models by margins of 2.9% in minADE, 2.4% in minFDE, and 3.8% in MR, while being on par with HiVT. It excels particularly in challenging long-term predictions (4s-5s) on NGSIM, HighD, and MoCAD datasets, with reductions in forecast error surpassing at least 11.5%, 29.6%, and 21.9%, respectively. This emphasizes its potential for accurate long-term predictions in highway and urban settings.

Performance Evaluation on the *Missing* Dataset. Table 2 showcases the resilience of our model when faced with incomplete data sets. Our model consistently outperforms all other baselines on the *drop 3-frames* and *drop 5-frames* datasets. Notably, on the *drop 3-frames* dataset, it surpasses nearly all state-of-the-art (SOTA) models trained on full data, highlighting its remarkable predictive strength even with missing data. While its performance on the *drop 5-frames* dataset excels over most baselines, there are exceptions in specific metrics against models like TNT, WIMP, and mm Transformer. As the number of missing frames increases, as in the *drop 8-frames* and *drop 10-frames* datasets, there's an expected decline in performance. Yet, even with half the input data missing, our model still competes strongly against top baselines, emphasizing its potential in environments with data interruptions.

Comparative Analysis of Model Performance and Complexity. In Table 3, we compare our model's performance and complexity with various SOTA baselines. While our model doesn't have the lowest parameter count, it excels in all performance metrics. Impressively, it achieves this while using 90.42% and 87.18% fewer parameters than WIMP and Scene-Transformer, respectively. Compared to top-10 SOTA models, our model not only surpasses them in accuracy but is also as efficient, if not more so, than HiVT-128, SSL-Lanes, LaneGCN, and HOME+GOHOME. This underlines our model's optimal balance of robustness, efficiency, and trajectory prediction accuracy.

Model	minADE (m)↓	minFDE (m)↓	MR (%)↓	#Param (K)
WIMP [Khandelwal *et al.*, 2020]	1.82	4.03	62.88	>20,000
Scene-Transformer [Ngiam *et al.*, 2022]	1.81	4.06	59.21	15,296
CtsConv (Aug.) [Walters *et al.*, 2021]	1.77	4.05	-	<u>1,078</u>
mmTransformer [Liu *et al.*, 2021]	1.77	4.00	61.78	2,607
LaneGCN [Liang *et al.*, 2020]	1.71	3.78	59.05	3,701
HOME+GOHOME [Gilles *et al.*, 2022]	1.69	3.65	57.21	5,100
DenseTNT [Gu *et al.*, 2021]	1.68	3.63	58.43	1,103
SSL-Lanes [Bhattacharyya *et al.*, 2023]	1.63	3.56	56.71	1,840
HiVT-128 [Zhou *et al.*, 2022]	1.60	3.52	-	2,529
MFTraj	**1.59**	**3.51**	**55.44**	**1,961**

Table 3: Comparative evaluation of MFTraj with SOTA baselines.

4.3 Ablation Studies

We executed an ablation study to assess the impact of individual components within our trajectory prediction model, with the results summarized in Table 5. Model F, i.e., MFTraj, which integrates all components, stands out in all metrics, signifying the synergy of its parts. When the behavior-aware module is excluded in Model A, there are noticeable drops in minADE, minFDE, and MR by 12.6%, 8.8%, and 8.5% respectively, highlighting its pivotal role. Model B, with abso-

Dataset	Model	Prediction Horizon (s)				
		1	2	3	4	5
NGSIM	S-GAN [Gupta *et al.*, 2018]	0.57	1.32	2.22	3.26	4.40
	CS-LSTM [Deo and Trivedi, 2018]	0.61	1.27	2.09	3.10	4.37
	DRBP[Gao *et al.*, 2023]	1.18	2.83	4.22	5.82	-
	WSiP [Wang *et al.*, 2023]	0.56	1.23	2.05	3.08	4.34
	CF-LSTM [Xie *et al.*, 2021]	0.55	1.10	1.78	2.73	3.82
	MHA-LSTM [Messaoud *et al.*, 2021]	0.41	1.01	1.74	2.67	3.83
	TS-GAN [Wang *et al.*, 2022b]	0.60	1.24	1.95	2.78	3.72
	Stdan [Chen *et al.*, 2022b]	0.39	0.96	1.61	2.56	3.67
	iNATran [Chen *et al.*, 2022a]	0.39	0.96	1.61	2.42	3.43
	DACR-AMTP [Cong *et al.*, 2023]	0.57	1.07	1.68	2.53	3.40
	FHIF [Zuo *et al.*, 2023]	0.40	0.98	1.66	2.52	3.63
	MFTraj	**0.38**	**0.87**	**1.52**	**2.23**	**2.95**
HighD	S-GAN [Gupta *et al.*, 2018]	0.30	0.78	1.46	2.34	3.41
	WSiP [Wang *et al.*, 2023]	0.20	0.60	1.21	2.07	3.14
	CS-LSTM [Deo and Trivedi, 2018]	0.22	0.61	1.24	2.10	3.27
	MHA-LSTM [Messaoud *et al.*, 2021]	0.19	0.55	1.10	1.84	2.78
	NLS-LSTM [Messaoud *et al.*, 2019]	0.20	0.57	1.14	1.90	2.91
	DRBP[Gao *et al.*, 2023]	0.41	0.79	1.11	1.40	-
	CF-LSTM [Xie *et al.*, 2021]	0.18	0.42	1.07	1.72	2.44
	Stdan [Chen *et al.*, 2022b]	0.19	0.27	0.48	0.91	1.66
	iNATran [Chen *et al.*, 2022a]	**0.04**	**0.05**	0.21	0.54	1.10
	DACR-AMTP [Cong *et al.*, 2023]	0.10	0.17	0.31	0.54	1.01
	GaVa [Liao *et al.*, 2024d]	0.17	0.24	0.42	0.86	1.31
	MFTraj	0.07	0.10	**0.19**	**0.38**	**0.56**
MoCAD	S-GAN [Gupta *et al.*, 2018]	1.69	2.25	3.30	3.89	4.69
	CS-LSTM [Deo and Trivedi, 2018]	1.45	1.98	2.94	3.56	4.49
	MHA-LSTM [Messaoud *et al.*, 2021]	1.25	1.48	2.57	3.22	4.20
	NLS-LSTM [Messaoud *et al.*, 2019]	0.96	1.27	2.08	2.86	3.93
	WSiP [Wang *et al.*, 2023]	0.70	0.87	1.70	2.56	3.47
	CF-LSTM [Xie *et al.*, 2021]	0.72	0.91	1.73	2.59	3.44
	Stdan [Chen *et al.*, 2022b]	0.62	0.85	1.62	2.51	3.32
	HLTP [Liao *et al.*, 2024a]	0.55	0.76	1.44	2.39	3.21
	BAT [Liao *et al.*, 2024b]	0.35	0.74	1.39	2.19	2.88
	MFTraj	**0.34**	**0.70**	**1.32**	**2.01**	**2.57**

Table 4: Evaluation results for MFTraj and the other SOTA baselines without using HD maps in the NGSIM, HighD and MoCAD datasets over a different horizon. RMSE (m) is the evaluation metric.

lute coordinates, underperforms, emphasizing the relevance of spatial relationships. Model C, without the interaction-aware module and Linformer extension, and Model D, lacking Linformer, both show diminished performance. Similarly, Model E, which uses a standard GCN instead of the adaptive one, also lags, underscoring the latter's efficiency. In essence, this study solidifies the importance of each component in Model F. Every part, from understanding behavioral nuances to updating features effectively, bolsters the model's precision and resilience. In essence, this study solidifies the importance of each component in Model F. Every part, from understanding behavioral nuances to updating features effectively, bolsters the model's precision and resilience.

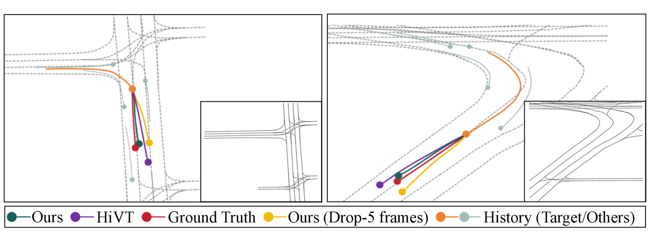

Figure 3: Ablation analysis of individual components in Argoverse.

4.4 Qualitative Results

Figure 3 presents the qualitative results of our model using the Argoverse dataset. We've limited the display to the target vehicle's trajectories for clarity. Interestingly, without the aid of HD maps, our model adeptly discerns road semantics, enabling it to make precise and logical predictions for target vehicles in intricate urban settings. Importantly, Figure 4 illustrates a comparison between the trajectories predicted by MFTraj and the SOTA baselines in the same traffic scenarios. MFTraj outperforms Stdan and WSiP in trajectory prediction, especially in complex scenarios such as lane changes and merging. These results demonstrate the superior adaptability and reliability of MFTraj in complex traffic conditions.

Ablation Models ($\Delta Model\ F$)	minADE (m)↓	minFDE (m)↓	MR (%)↓
Model A	1.82	3.85	60.61
Model B	1.69	3.59	56.14
Model C	1.78	3.71	59.07
Model D	1.71	3.61	57.59
Model E	1.68	3.70	56.94
Model F	**1.59**	**3.51**	**55.44**

Table 5: Qualitative results of MFTraj and HiVT on Agroverse.

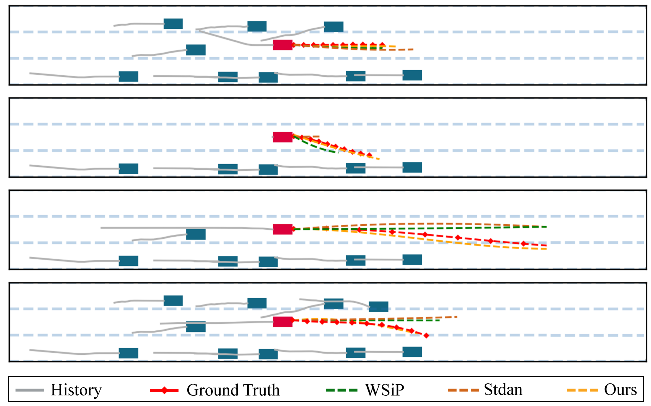

Figure 4: Qualitative results of MFTraj on NGSIM. Target vehicle is depicted in red, while its surrounding agents are shown in blue.

5 Conclusion

This work presents a map-free and behavior-aware trajectory prediction model for AVs, integrating four components: behavior-aware, position-aware, interaction-aware modules, and a residual decoder. These components work in concert to analyze and interpret various inputs, understand human-machine interactions, and account for the inherent uncertainty and variability in the prediction. Evaluated with the Argoverse, NGSIM, HighD, and MoCAD datasets, MFTraj outperformed SOTA baselines in prediction accuracy and efficiency without additional map information. Furthermore, this approach ensures its robustness and adaptability even in the presence of significant missing data; it achieved impressive performance even with a 50% sequential input data deficit. This underscores the resilience and efficiency of MFTraj in predicting future vehicle trajectories and suggests its potential to drastically reduce the data requirements for training AVs, especially in data-missing and limited data scenes.

Acknowledgements

This research is supported by the Science and Technology Development Fund of Macau SAR (File no. 0021/2022/ITP, 0081/2022/A2, 001/2024/SKL), Shenzhen-Hong Kong-Macau Science and Technology Program Category C (SGDX20230821095159012), and University of Macau (SRG2023-00037-IOTSC). Haicheng Liao and Zhenning Li contributed equally to this work. Please ask Dr. Zhenning Li (zhenningli@um.edu.mo) for correspondence.

References

[Bhattacharyya *et al.*, 2023] Prarthana Bhattacharyya, Chengjie Huang, and Krzysztof Czarnecki. Ssl-lanes: Self-supervised learning for motion forecasting in autonomous driving. In *Conference on Robot Learning*, pages 1793–1805. PMLR, 2023.

[Boguna *et al.*, 2021] Marian Boguna, Ivan Bonamassa, Manlio De Domenico, Shlomo Havlin, Dmitri Krioukov, and M Ángeles Serrano. Network geometry. *Nature Reviews Physics*, 3(2):114–135, 2021.

[Chang *et al.*, 2019] 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 CVPR*, 2019.

[Chen *et al.*, 2022a] Xiaobo Chen, Huanjia Zhang, Feng Zhao, Yingfeng Cai, Hai Wang, and Qiaolin Ye. 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, 2022.

[Chen *et al.*, 2022b] Xiaobo Chen, Huanjia Zhang, Feng Zhao, Yu Hu, Chenkai Tan, and Jian Yang. 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, 2022.

[Chung *et al.*, 2015] Junyoung Chung, Kyle Kastner, Laurent Dinh, Kratarth Goel, Courville, and Yoshua Bengio. A recurrent latent variable model for sequential data. *Advances in neural information processing systems*, 2015.

[Cong *et al.*, 2023] Peichao Cong, Yixuan Xiao, Xianquan Wan, Murong Deng, Jiaxing Li, and Xin Zhang. Dacr-amtp: Adaptive multi-modal vehicle trajectory prediction for dynamic drivable areas based on collision risk. *IEEE Transactions on Intelligent Vehicles*, 2023.

[Deo and Trivedi, 2018] Nachiket Deo and Mohan M Trivedi. Convolutional social pooling for vehicle trajectory prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1468–1476, 2018.

[Fang *et al.*, 2020] Liangji Fang, Qinhong Jiang, Jianping Shi, and Bolei Zhou. Tpnet: Trajectory proposal network for motion prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6797–6806, 2020.

[Gao *et al.*, 2020] Jiyang Gao, Chen Sun, Hang Zhao, Yi Shen, Dragomir Anguelov, Congcong Li, and Cordelia Schmid. Vectornet: Encoding hd maps and agent dynamics from vectorized representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11525–11533, 2020.

[Gao *et al.*, 2023] Kai Gao, Xunhao Li, Bin Chen, Lin Hu, Jian Liu, Ronghua Du, and Yongfu Li. Dual transformer based prediction for lane change intentions and trajectories in mixed traffic environment. *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[Gilles *et al.*, 2021] Thomas Gilles, Stefano Sabatini, Dzmitry Tsishkou, Bogdan Stanciulescu, and Fabien Moutarde. Home: Heatmap output for future motion estimation. In *IEEE ITSC*, 2021.

[Gilles *et al.*, 2022] Thomas Gilles, Stefano Sabatini, Dzmitry Tsishkou, Bogdan Stanciulescu, and Fabien Moutarde. Gohome: Graph-oriented heatmap output for future motion estimation. In *Proceedings of ICRA*, 2022.

[Gu *et al.*, 2021] Junru Gu, Chen Sun, and Hang Zhao. Densetnt: End-to-end trajectory prediction from dense goal sets. In *Proceedings of the IEEE/CVF CVPR*, 2021.

[Gupta *et al.*, 2018] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the CVPR*, 2018.

[Khandelwal *et al.*, 2020] Siddhesh Khandelwal, William Qi, Jagjeet Singh, Andrew Hartnett, and Deva Ramanan. What-if motion prediction for autonomous driving. *arXiv preprint arXiv:2008.10587*, 2020.

[Krajewski *et al.*, 2018] Robert Krajewski, Julian Bock, Laurent Kloeker, and Lutz Eckstein. The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In *Proceeding of the ITSC*, 2018.

[Liang *et al.*, 2020] Ming Liang, Bin Yang, Rui Hu, Yun Chen, Renjie Liao, Song Feng, and Raquel Urtasun. Learning lane graph representations for motion forecasting. In *In Proceedings of the ECCV*, 2020.

[Liao *et al.*, 2024a] Haicheng Liao, Yongkang Li, Zhenning Li, Chengyue Wang, Zhiyong Cui, Shengbo Eben Li, and Chengzhong Xu. A cognitive-based trajectory prediction approach for autonomous driving. *IEEE Transactions on Intelligent Vehicles*, pages 1–12, 2024.

[Liao *et al.*, 2024b] Haicheng Liao, Zhenning Li, Huanming Shen, Wenxuan Zeng, Dongping Liao, Guofa Li, and Chengzhong Xu. Bat: Behavior-aware human-like trajectory prediction for autonomous driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 10332–10340, 2024.

[Liao *et al.*, 2024c] Haicheng Liao, Zhenning Li, Chengyue Wang, Bonan Wang, Hanlin Kong, Yanchen Guan, Guofa Li, Zhiyong Cui, and Chengzhong Xu. A cognitive-driven trajectory prediction model for autonomous driving in mixed autonomy environment. *arXiv preprint arXiv:2404.17520*, 2024.

[Liao *et al.*, 2024d] Haicheng Liao, Shangqian Liu, Yongkang Li, Zhenning Li, Chengyue Wang, Bonan Wang, Yanchen Guan, and Chengzhong Xu. Human observation-inspired trajectory prediction for autonomous driving in mixed-autonomy traffic environments. *arXiv preprint arXiv:2402.04318*, 2024.

[Liao *et al.*, 2024e] Haicheng Liao, Chengyue Wang, Zhenning Li, Yongkang Li, Bonan Wang, Guofa Li, and Chengzhong Xu. Physics-informed trajectory prediction for autonomous driving under missing observation. *Available at SSRN 4809575*, 2024.

[Liu *et al.*, 2021] Yicheng Liu, Jinghuai Zhang, Liangji Fang, Qinhong Jiang, and Bolei Zhou. Multimodal motion prediction with stacked transformers. In *Proceedings of the IEEE/CVF CVPR*, 2021.

[Markkula *et al.*, 2020] Gustav Markkula, Ruth Madigan, Dimitris Nathanael, Evangelia Portouli, Yee Mun Lee, André Dietrich, Jac Billington, Anna Schieben, and Natasha Merat. Defining interactions: A conceptual framework for understanding interactive behaviour in human and automated road traffic. *Theoretical Issues in Ergonomics Science*, 21(6):728–752, 2020.

[Mercat *et al.*, 2020] Jean Mercat, Thomas Gilles, Nicole El Zoghby, Guillaume Sandou, Dominique Beauvois, and Guillermo Pita Gil. Multi-head attention for multi-modal joint vehicle motion forecasting. In *IEEE ICRA*, 2020.

[Messaoud *et al.*, 2019] Kaouther Messaoud, Itheri Yahiaoui, Anne Verroust-Blondet, and Fawzi Nashashibi. Non-local social pooling for vehicle trajectory prediction. In *IEEE Vehicles Symposium*, 2019.

[Messaoud *et al.*, 2021] Kaouther Messaoud, Itheri Yahiaoui, Anne Verroust-Blondet, and Fawzi Nashashibi. Attention based vehicle trajectory prediction. *IEEE Transactions on Intelligent Vehicles*, 6(1):175–185, 2021.

[Mohamed *et al.*, 2020] 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 CVPR*, 2020.

[Ngiam *et al.*, 2022] Jiquan Ngiam, Vijay Vasudevan, Benjamin Caine, Zhengdong Zhang, Hao-Tien Lewis Chiang, Jeffrey Ling, Rebecca Roelofs, Alex Bewley, Chenxi Liu, Ashish Venugopal, David J Weiss, Ben Sapp, Zhifeng Chen, and Jonathon Shlens. Scene transformer: A unified architecture for predicting future trajectories of multiple agents. In *Proceedings of the ICLR*, 2022.

[Schwarting *et al.*, 2019] Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences*, 116:24972, 2019.

[Song *et al.*, 2022] Haoran Song, Di Luan, Wenchao Ding, Michael Y Wang, and Qifeng Chen. Learning to predict vehicle trajectories with model-based planning. In *Conference on Robot Learning*, pages 1035–1045. PMLR, 2022.

[Tian *et al.*, 2024] Chunlin Tian, Zhan Shi, Zhijiang Guo, Li Li, and Chengzhong Xu. Hydralora: An asymmetric lora architecture for efficient fine-tuning, 2024.

[Walters *et al.*, 2021] Robin Walters, Jinxi Li, and Rose Yu. Trajectory prediction using equivariant continuous convolution. In *Proceedings of the ICLR*, 2021.

[Wang *et al.*, 2020] Sinong Wang, Belinda Z Li, Madian Khabsa, Han Fang, and Hao Ma. Linformer: Self-attention with linear complexity. *arXiv preprint:2006.04768*, 2020.

[Wang *et al.*, 2022a] Jingke Wang, Tengju Ye, Ziqing Gu, and Junbo Chen. Ltp: Lane-based trajectory prediction for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17134–17142, 2022.

[Wang *et al.*, 2022b] Yu Wang, Shengjie Zhao, Rongqing Zhang, Xiang Cheng, and Liuqing Yang. Multi-vehicle collaborative learning for trajectory prediction with spatio-temporal tensor fusion. *IEEE Transactions on Intelligent Transportation Systems*, 23(1):236–248, 2022.

[Wang *et al.*, 2023] Renzhi Wang, Senzhang Wang, Hao Yan, and Xiang Wang. Wsip: Wave superposition inspired pooling for dynamic interactions-aware trajectory prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4685–4692, 2023.

[Wu and He, 2018] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the ECCV*, 2018.

[Xie *et al.*, 2021] Xu Xie, Chi Zhang, Yixin Zhu, Ying Nian Wu, and Song-Chun Zhu. Congestion-aware multi-agent trajectory prediction for collision avoidance. In *Proceedings of the IEEE ICRA*, 2021.

[Ye *et al.*, 2021] Maosheng Ye, Tongyi Cao, and Qifeng Chen. Tpcn: Temporal point cloud networks for motion forecasting. In *Proceedings of the CVPR*, 2021.

[Yin *et al.*, 2021] Ziyi Yin, Ruijin Liu, Zhiliang Xiong, and Zejian Yuan. Multimodal transformer networks for pedestrian trajectory prediction. In *IJCAI*, 2021.

[Zhao *et al.*, 2021] 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.

[Zhou *et al.*, 2022] 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 CVPR*, 2022.

[Zhou *et al.*, 2023] Zikang Zhou, Jianping Wang, Yung-Hui Li, and Yu-Kai Huang. Query-centric trajectory prediction. In *Proceedings of the IEEE/CVF CVPR*, 2023.

[Zuo *et al.*, 2023] Zhiqiang Zuo, Xinyu Wang, Songlin Guo, Zhengxuan Liu, Zheng Li, and Yijing Wang. Trajectory prediction network of autonomous vehicles with fusion of historical interactive features. *IEEE Transactions on Intelligent Vehicles*, 2023.