KnowMTP: A Knowledge-Guided Framework for Multi-Agent Trajectory Prediction in Autonomous Driving

Rufan Bai[1,*] Tianyi Xue[1,*] Tiantian Zhou[1] Weiwei Wu[1] Changle Li[2] Yuhuan Lu[3,†]
[1]Southeast University, Nanjing, China
[2]Xidian University, Xi'an, China
[3]Digital Future Institute, Khalifa University, Abu Dhabi, United Arab Emirates

Abstract

Accurate multi-agent trajectory prediction (MTP) is essential for autonomous driving, enabling vehicles to anticipate the future behaviors of surrounding agents and plan safe, efficient maneuvers. Although recent data-driven approaches based on the encoder-decoder architecture have achieved notable success, they often rely solely on attention mechanisms to process uniformly structured inputs, overlooking valuable domain knowledge and introducing unnecessary complexity. This paper presents KnowMTP, a knowledge-guided framework for MTP that integrates driving safety and behavioral knowledge to intelligently reorganize input data. KnowMTP achieves this through two complementary filtering mechanisms that combine a Motion-Similar Agent Selection filter identifying relevant neighbors based on motion similarity with a Safety-Critical Agent Selection filter prioritizing agents posing potential collision risks, thereby embedding cognitive awareness of safety constraints. The framework is model-agnostic and integrates seamlessly into existing encoder-decoder architectures. Experiments on benchmark datasets show that KnowMTP consistently improves the performance of state-of-the-art MTP baselines, achieving 18.8% gains on WOMD with 22.1% lower computational cost while enhancing the plausibility and safety awareness of predicted trajectories.

1. Introduction

Multi-agent trajectory prediction (MTP) is critical for autonomous driving systems [13, 35, 41], allowing ego vehicles to predict the future movements of surrounding agents, such as other vehicles and pedestrians. Accurate trajectory prediction for vehicles, pedestrians, and other agents is essential for planning safe and smooth maneuvers in dynamic driving scenarios, such as lane changes [28], overtaking [2], and obstacle avoidance [40]. However, trajectory prediction in complex driving environments is inherently challenging due to the dynamic interactions between multiple agents, the constraints imposed by road topology and physical laws, and the uncertainties inherent in human behaviors. Thus, modern MTP methods typically output multi-modal predictions, accompanied by confidence scores, to comprehensively capture various plausible future trajectories.

In recent years, deep learning approaches have become the dominant paradigm for trajectory prediction. Among them, encoder-decoder architectures based on Transformer [6, 23, 42] or graph neural network [25, 37] backbones are widely adopted. These models encode historical trajectories along with scene context before decoding them into predicted future trajectories. They have achieved good performance in both individual prediction and joint multi-agent settings. GameFormer [14] introduces a game-theoretic Transformer with hierarchical interaction modeling, while MTR++ [32] employs symmetric scene encoding and structured attention to generate consistent joint predictions. These methods reflect ongoing efforts within the research community to refine neural network architectures that better capture complex social interactions and temporal dependencies.

However, existing MTP methods remain purely data-driven. They treat all nearby agents and static map elements as equally relevant, and often rely on learned attention mechanisms to infer interactions. This uniform processing increases model complexity and introduces noise, since many surrounding agents have limited or no influence on the future trajectory of the target. Moreover, by giving equal weight to all neighbors, these models often fail to explicitly prioritize the most critical interactions and can produce unrealistic or unsafe predictions.

To address these limitations, in this paper, we propose KnowMTP, a novel knowledge-guided framework designed explicitly for multi-agent trajectory prediction in autonomous driving. KnowMTP differentiates itself by incorporating domain knowledge about driving behavior and

*Equal contribution.
†Corresponding author.

safety to intelligently reorganize input data under top-down guidance. Our core insight is that a target agent's future behavior is primarily influenced by a small subset of neighboring agents, whose actions are driven by two key factors: motion similarity and safety criticality. KnowMTP implements this insight through a structured input processing module that incorporates two complementary agent filters.

The Motion-Similar Agent Selection filter leverages motion pattern similarity to dynamically identify neighboring agents whose trajectories exhibit comparable driving intentions or styles to the target agent. By computing similarity scores from learned trajectory embeddings, the model selectively retains the most informative agents while suppressing irrelevant ones. This targeted filtering reduces input noise and enhances the focus of the model on meaningful interactions, thereby improving prediction accuracy.

The Safety-Critical Agent Selection filter prioritizes agents that pose a potential safety risk based on proximity and relative motion. This is determined using traffic safety metrics like Time-to-Collision (TTC) and its derivatives (TET, TIT). By explicitly modeling these safety constraints, the prediction framework captures a cognitive understanding of driving safety, resulting in more robust and risk-aware trajectory predictions.

The integration of domain knowledge through these two filters marks a significant departure from traditional purely data-driven approaches. While previous works primarily focused on refining neural architectures or attention mechanisms, KnowMTP systematically organizes input data using insights from driving psychology and safety, incorporating structured prior knowledge into the learning process. The two filters complement each other by focusing both on motion similarity and safety relevance, together improving prediction accuracy and robustness. Our framework is model-agnostic, easily integrating with existing state-of-the-art encoder-decoder architectures to enhance performance without requiring architectural changes.

Extensive experiments demonstrate that augmenting high-performing encoder-decoder MTP baselines with the plug-in modules of KnowMTP consistently improves performance on widely used benchmarks. Specifically, our approach yields substantial gains in prediction accuracy metrics, including lower trajectory errors and higher success rates in anticipating agent behaviors. In addition, the targeted filtering of relevant agents leads to improved computational efficiency, underscoring the practicality of KnowMTP for real-time autonomous driving applications.

In summary, the main contributions can be summarized as follows:

- We propose KnowMTP, a novel knowledge-guided trajectory prediction framework that integrates motion similarity and safety criticality to improve prediction accuracy and robustness.

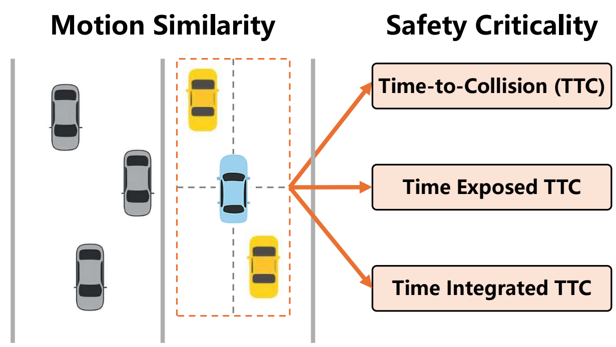

Figure 1. Philosophy of KnowMTP. The blue vehicle denotes the target agent to be predicted. KnowMTP dynamically selects the most relevant neighboring vehicles moving in the same direction (yellow vehicles) based on motion similarity, while ignoring vehicles in the opposite direction (gray vehicles), even if they are spatially closer. KnowMTP also considers the safety criticality of surrounding vehicles with respect to the target agent through metrics such as TTC, Time-Exposed TTC, and Time-Integrated TTC, and injects this safety awareness into encoder-decoder architectures to enhance prediction.

- We introduce a novel agent filtering mechanism based on motion similarity to dynamically select critical neighboring agents, enhancing both accuracy and computational efficiency.
- We integrate the proposed filtering framework into state-of-the-art prediction models and show that it consistently enhances trajectory prediction performance. The framework achieves up to 18.8% improvement on the WOMD dataset while reducing computational cost by 22.1%. Moreover, it enables socially and safety-aware trajectory generation, demonstrating the effectiveness of embedding structured domain knowledge into data-driven multi-agent prediction models..

This work establishes a new paradigm for MTP by demonstrating that the structural integration of domain knowledge through targeted agent filtering is key to achieving superior performance over purely data-driven approaches.

2. Related Work

2.1. Data-Driven Trajectory Prediction

Trajectory prediction in autonomous driving is typically framed as a supervised sequence-to-sequence task aiming to capture complex spatio-temporal dependencies and agent interactions. Early approaches centered on LSTM-based models like Social LSTM [1] and its attention-enhanced variants [29, 34], which excel at sequential dependency modeling but struggle with complex spatial relationships. Raster-based CNN methods subsequently emerged, processing scenes as bird's-eye view images [9] to better capture spatial context. The field then evolved toward vectorized representations where agents and lanes form graph nodes processed by attention or GNN layers, enabling explicit relational modeling [8]. More recently, Transformer architectures have become dominant in trajectory prediction by leveraging self-attention to model complex multi-agent interactions [15, 22, 27, 31]. Models such as S2TNet [5], MTR [31], and DenseTNT [10] effectively capture long-range spatial and temporal dependencies, making them suitable for scenarios involving dense and dynamic agent interactions. Building on this, recent advancements include query-centric paradigms such as QCNet [33] for efficient scene encoding and proposal generation, along with methods that incorporate driving style and enhance prediction accuracy.

However, these data-driven methods often treat all agents equally and assume a static scene context, limiting their effectiveness in dynamic and heterogeneous traffic environments. In contrast, KnowMTP systematically incorporates traffic engineering principles by embedding domain knowledge into the prediction process through interpretable and quantifiable metrics, enabling more robust and adaptive performance in real-world settings.

2.2. Safety Awareness in Trajectory Prediction

Integrating safety awareness is essential for robust trajectory prediction and decision-making in autonomous driving, moving beyond simple geometric collision avoidance. Recent research has emphasized incorporating traffic safety indicators, where metrics such as Time-to-Collision (TTC) and its derivatives (TET, TIT) serve as safety surrogates to filter inputs or constrain prediction outputs [3, 18, 38]. Modern MTP methods adopt safety-aware prediction to refine multi-modal outputs, ensuring generated trajectories adhere to safety standards [16, 30]. Beyond physical constraints, studies on perceived safety and cognitive modeling enhance game-theoretic frameworks with cognitive parameters to capture how agents adjust maneuvers under perceived threats [11, 36]. Other approaches refine GNN-based models to filter interactions based on proximity and potential danger, identifying truly critical interactions [21]. However, most existing methods integrate safety constraints heuristically or apply them only after prediction. KnowMTP advances this field by structurally embedding safety knowledge before the encoder stage. Its Safety-Critical Agent Selection filter proactively leverages established safety metrics (TTC, TET, TIT) to prune inputs, ensuring the model focuses on the most influential and potentially hazardous neighboring agents, thereby achieving more accurate and robust trajectory predictions.

3. Preliminaries

3.1. Problem Formulation

At time $t = 0$, we consider a driving scenario comprising N agents, denoted by $A_{1:N}$, together with the scenario map $\mathbf{M}$ that includes lanes, traffic lights, traffic signs, crosswalks, and so on. Each agent's state history over the past horizon T_h is represented by:

$$\mathbf{X}_n = \left\{\mathbf{x}^{-T_h:0}\right\}_n, \quad n \in [1, N]. \tag{1}$$

Our multi-agent trajectory prediction (MTP) objective is to jointly forecast each agent's trajectory over the future horizon T_f. To better capture the uncertainty, we produce M multimodal outputs per agent, with each candidate trajectory accompanied by a confidence score:

$$\mathbf{Y}_n^{1:T_f} = \left\{\left(\mathbf{y}_{n,j}^{1:T_f}, \delta_{n,j}\right) \mid j = 1, \ldots, M\right\}, \tag{2}$$

where $\mathbf{y}_{n,j}^{1:T_f}$ denotes the j-th predicted coordinate sequence for agent n and $\delta_{n,j}$ is the confidence score (i.e., probability) of the trajectory.

3.2. Encoder-Decoder Architecture for MTP

The encoder-decoder architecture is currently the most prevalent and widely adopted framework for the MTP task, which proves to be effective across different driving scenarios [14, 20, 32, 39]. As shown in the Figure 2, this architecture (the gray part) is composed of an encoder (mostly Transformer) and a (mostly attention-based) decoder. Specifically, for each agent, the encoder captures the sophisticated interactions among its own state, the states of nearby agents, and local map elements, yielding a context embedding that encapsulates its environment-aware behavior. The decoder then extends the capability of this architecture by processing the context embeddings of all agents in the scenario, modeling their interdependencies to generate the final predicted trajectories.

In this study, we inherit the encoder-decoder architecture as the backbone of our framework, and integrate multiple state-of-the-art (SOTA) MTP methods to corroborate the effectiveness of our framework in the experiments.

4. Methodology

4.1. General Scheme

Current MTP methods primarily focus on designing the encoder and/or decoder modules, often overlooking the impact of input data organization on prediction performance. To address this limitation, we propose **KnowMTP**, a knowledge-guided MTP framework that subtly reorganizes the input data under the top-down government of domain knowledge. Despite its simplicity, KnowMTP

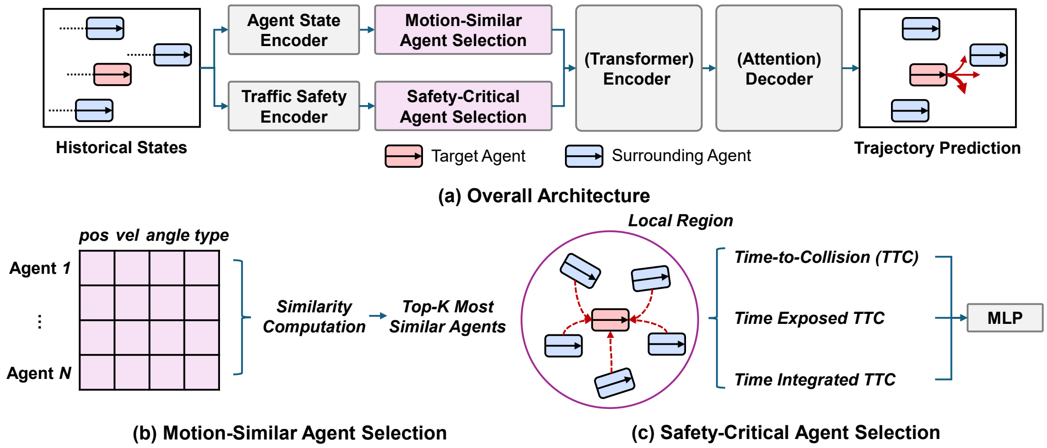

Figure 2. (a) The proposed KnowMTP framework for the multi-agent trajectory prediction (MTP) task. The gray region represents the universal encoder-decoder architecture, while the purple region illustrates our input data reorganization design, which includes two modules. (b) The global agent selection module, responsible for identifying the most relevant surrounding agents for the target agent. (c) The local environment profiling module, characterizing the comprehensive local environmental dynamics surrounding the target agent.

significantly enhances the prediction accuracy of existing encoder-decoder-based MTP models.

The core principle behind KnowMTP is that not all nearby agents' maneuvers are relevant to a given agent's future trajectory. Therefore, only the historical states of the most relevant agents are fed into the model. To operationalize this idea, KnowMTP introduces two agent filters that identify the most relevant neighboring agents for each target agent. The design philosophy of these filters is grounded in two key factors influencing a target agent's behavior: **motion similarity** and **safety criticality**. In the following sections, we detail the design of the two agent filters.

4.2. Motion-Similar Agent Selection

The motion-similar agent filter is designed to identify the most relevant nearby agents for a target agent based on motion similarity, thereby providing valuable contextual information for forecasting the target's future trajectory. Agents exhibiting similar motion patterns to the target are more likely to exert influence, as they tend to share comparable driving intentions and behaviors. Accordingly, we define motion-similar agents as those exhibiting the highest motion similarity to the target agent and use them as inputs to the encoder.

As illustrated in Figure 2(b), the agent filter computes the similarity between the historical states of the target agent and its neighboring agents, selecting the top-K most similar ones. Specifically, the historical state of each agent is first transformed into a state embedding:

$$\mathbf{s}_n = \varphi_s\left(\mathbf{X}_n\right), \tag{3}$$

where $\varphi_s\left(\cdot\right)$ is a state encoding function, implemented using either LSTM networks [14] or MLP networks [32], depending on the underlying MTP baseline. Afterward, the similarity score between agents i and j is computed using the L2 norm:

$$c_{ij} = \left\|\mathbf{s}_i - \mathbf{s}_j\right\|_2. \tag{4}$$

Notably, the similarity score is order-independent, i.e., $c_{ij} = c_{ji}$, ensuring symmetry in the measurement. This score effectively quantifies the consistency in motion patterns between the agents. For each target agent, similarity scores are computed with all neighboring agents, and the top-K most similar ones are retained while the rest are filtered out. This motion-similar filtering mechanism enables the model to focus on the most relevant and influential neighbors from a motion perspective, thereby reducing input noise and enhancing the precision of interaction modeling.

4.3. Safety-Critical Agent Selection

In addition to motion similarity, a target agent's future maneuvers are also constrained by the behaviors of its neighboring agents due to driving safety considerations. Agents that exhibit similar levels of perceived safety are more likely to mutually influence each other's behaviors and should therefore be prioritized by the model. Inspired by the concept of perceived safety among human drivers [12], we incorporate safety-awareness into the agent filtering process. In the traffic safety domain, perceived safety is commonly characterized by three risk metrics: Time-to-Collision (TTC), Time Exposed Time-to-Collision (TET), and Time Integrated Time-to-Collision (TIT) [24]. These metrics describe the one-to-one interactions between any pair of vehicles. We adapt these metrics in our framework as follows:

(1) Time-to-Collision (TTC): TTC represents the time remaining before two agents collide if they continue along their current trajectories. It serves as an indicator of imminent risk and is widely used as an early warning signal. Since conventional TTC is defined for one-dimensional longitudinal motion, we extend it to a two-dimensional setting to better reflect complex driving scenarios. The refined TTC for a target agent i with respect to a surrounding agent j at the time interval t is calculated as:

$$e_{TTC_{i,j}^t} = -\frac{d_{i,j}^t}{\hat{d}_{i,j}^t} \tag{5}$$

$d_{i,j}^t$ denotes the distance between agents i and j and $\hat{d}_{i,j}^t$ is the change rate of $d_{i,j}^t$:

$$\begin{cases} d_{i,j}^t = \sqrt{\left(p_i^t - p_j^t\right)^\top \left(p_i^t - p_j^t\right)} \\ \hat{d}_{i,j}^t = \frac{1}{d_{i,j}^t}\left(p_i^t - p_j^t\right)^\top \left(v_i^t - v_j^t\right) \end{cases} \tag{6}$$

where $\left(p_i^t, v_i^t\right)$ and $\left(p_j^t, v_j^t\right)$ are 2D coordinates and velocities of agents i and j, respectively.

(2) Time Exposed Time-to-Collision (TET): TET measures the cumulative duration during which TTC values remain within safety-critical thresholds over a specified time horizon t_H:

$$e_{TET_i^t} = \sum_{t_k=t-t_H+1}^{t} \delta_i\left(t_k\right) \cdot \tau_{\text{sc}} \tag{7}$$

where τ_{sc} is set to 0.1s, signifying the minimum time step in which the measured TTC values do not change. $\delta_i\left(t_k\right)$ is a switching variable:

$$\delta_i(t_k) = \begin{cases} 1, & \forall 0 \leq e_{TTC_i^{t_k}} \leq e_{TTC^*} \\ 0, & \text{else} \end{cases} \tag{8}$$

where e_{TTC^*} is the safety-critical TTC threshold, set to 2.5 seconds in this study.

(3) Time Integrated Time-to-Collision (TIT): While TET captures the duration of safety-critical exposure, it does not differentiate between varying risk levels. To address this limitation, TIT integrates the TTC profile over time, providing a more nuanced assessment of safety risk:

$$\begin{gathered} e_{TIT_i^t} = \sum_{t_k=t-t_H+1}^{t} \left[e_{TTC^*} - e_{TTC_i^{t_k}}\right] \cdot \tau_{\text{sc}} \\ \forall 0 \leq e_{TTC_i^{t_k}} \leq e_{TTC^*} \end{gathered} \tag{9}$$

Higher TTC, TET, and TIT values indicate greater exposure to potential collision risks and consequently lower perceived safety. We combine these three perceived safety metrics for all neighboring agents into a unified comprehensive safety indicator, denoted as $\mathbf{e}_i$, which characterizes the overall perceived safety of each agent. Following the same embedding and similarity computation process as in Equations (3)-(4), we identify the top-K agents most similar to the target in terms of perceived safety. This safety-critical filtering mechanism enables the model to incorporate cognitive awareness of driving safety, thereby improving the accuracy and robustness of future trajectory predictions.

4.4. Model Training

With the two agent filters described above, the historical states of the selected agents are fed into the prediction model to yield the future trajectories of the target agents. Since KnowMTP modifies only the organization of the input data, it remains fully compatible with the original training losses used in existing MTP baselines. Moreover, by filtering out redundant agent states, KnowMTP reduces input complexity, making the baseline models more efficient when integrated into the framework. This improvement in efficiency is empirically validated in our experiments.

5. Experiments

5.1. Experimental Setup

We conduct experiments on two real-world datasets, WOMD [7] and nuPlan [4], to demonstrate its ability to enhance the trajectory prediction performance of SOTA models. The WOMD dataset contains approximately 487,000 training scenes, 44,000 validation scenes, and 44,000 testing scenes. We use 1 second of historical trajectory data to predict 8 seconds into the future. For the nuPlan dataset, there are 189,377, 23,672, and 23,672 samples for training, validation, and testing, respectively. In this case, 2 seconds of history are used as input, and 8 seconds are predicted ahead. To ensure a fair comparison, we follow the standard evaluation metrics defined for each dataset, including minimum average displacement error (minADE), minimum final displacement error (minFDE), and miss rate (MR). Detailed hyperparameter settings for integrating KnowMTP with different baseline models are provided in Appendix A.

5.2. Main Results

We evaluate the performance of KnowMTP by integrating it with four representative methods: GameFormer [14], MTR [32], BeTop [19], and EDA [17]. These methods share a common Transformer-based encoder but differ in their attention-based decoders, allowing us to validate the effectiveness of KnowMTP across diverse decoder designs.

Prediction results are summarized in Table 1 for the WOMD dataset and Table 2 for the nuPlan dataset. As shown, KnowMTP consistently improves the prediction performance of all baseline models across both datasets,

Table 1. Prediction results on the WOMD dataset. For each baseline and its KnowMTP-enhanced counterpart, the better value for each metric is highlighted in bold.

Method	Evaluation Metrics		
	minADE↓	minFDE↓	MR↓
GameFormer	1.5017	4.2293	0.6703
KnowMTP+GameF	**1.0757**	**3.0788**	**0.3223**
MTR	0.7451	1.5678	0.2017
KnowMTP+MTR	**0.6567**	**1.2712**	**0.1603**
BeTop	0.7181	1.4957	0.1806
KnowMTP+BeTop	**0.6692**	**1.3208**	**0.1664**
EDA	0.7036	1.4646	0.1724
KnowMTP+EDA	**0.6266**	**1.2275**	**0.1492**

Table 2. Prediction results on the nuPlan dataset.

Method	Evaluation Metrics		
	minADE↓	minFDE↓	MR↓
GameFormer	4.2176	10.4640	0.5923
KnowMTP+GameF	**4.1137**	**10.3267**	**0.5201**
MTR	1.4211	2.8260	**0.3111**
KnowMTP+MTR	**1.2120**	**2.4240**	0.3156
BeTop	1.5390	2.8657	**0.3085**
KnowMTP+BeTop	**1.2480**	**2.6590**	0.3122
EDA	1.0131	1.9934	0.2972
KnowMTP+EDA	**0.9353**	**1.8092**	**0.2970**

demonstrating strong generalization and robustness. On the WOMD dataset, KnowMTP achieves an average improvement of **18.8%** over the four baselines, with performance gains of *35.8%*, *17.1%*, *8.8%*, and *13.5%* for GameFormer, MTR, BeTop, and EDA, respectively. On the nuPlan dataset, KnowMTP delivers average improvements ranging from **5.3%** to **9.2%** across the same baselines.

Notably, while KnowMTP introduces a slight increase in MR for MTR and BeTop on nuPlan, this can be attributed to the agent selection mechanism slightly constraining information propagation in topological graph modeling. However, this minor trade-off is negligible, as KnowMTP still enhances overall performance across all key metrics. Furthermore, KnowMTP achieves larger improvements on WOMD than on nuPlan, where all baselines generally perform worse, indicating that the nuPlan dataset presents greater challenges for trajectory prediction. Despite this, KnowMTP consistently boosts the performance of all baselines on nuPlan, underscoring its effectiveness and adaptability in complex multi-agent environments.

Table 3. Ablation results on the WOMD dataset.

Method	Evaluation Metrics		
	minADE↓	minFDE↓	MR↓
KnowMTP+GameF	**1.0757**	**3.0788**	**0.3223**
w/o pos	1.0766	3.0960	0.3626
w/o vel	1.2853	3.6621	0.4908
w/o angle	1.1893	3.4534	0.6245
w/o type	1.1062	3.1633	0.3498
w/o safety	1.2501	3.5922	0.4652
KnowMTP+MTR	**0.6567**	**1.2712**	**0.1603**
w/o pos	0.7194	1.5160	0.1995
w/o vel	0.7486	1.5312	0.2016
w/o angle	0.7617	1.5616	0.2053
w/o type	0.7441	1.5181	0.1987
w/o safety	0.7663	1.5652	0.1965
KnowMTP+BeTop	**0.6692**	**1.3208**	**0.1664**
w/o pos	0.7048	1.4827	0.1754
w/o vel	0.7162	1.5163	0.1895
w/o angle	0.7207	1.5197	0.1946
w/o type	0.7211	1.5111	0.1876
w/o safety	0.7124	1.4733	0.1812
KnowMTP+EDA	**0.6266**	**1.2275**	**0.1492**
w/o pos	0.7586	1.5447	0.1900
w/o vel	0.7055	1.4827	0.1810
w/o angle	0.7299	1.5219	0.1877
w/o type	0.7121	1.4964	0.1771
w/o safety	0.7282	1.5256	0.1834

5.3. Ablation Studies

The KnowMTP framework consists of two key components: motion-similar agent selection and safety-critical agent selection, which together contribute to its strong performance. However, it is important to assess the individual contribution of each component to the overall predictive capability. To this end, we conduct a series of ablation studies and present detailed results on the WOMD dataset in Table 3.

Specifically, to thoroughly examine the effectiveness of the motion-similar agent selection component, we analyze the impact of different agent state factors used in the selection process. The *w/o pos* variant removes the position state from the agent selection mechanism. This results in a noticeable performance drop across all KnowMTP-integrated baselines, underscoring the importance of positional information in identifying and filtering out irrelevant agents. However, the impact of this variant is less pronounced compared to others, suggesting that proximity alone does not fully determine influence. Instead, a more comprehensive assessment of motion similarity is necessary to effectively identify influential agents. The *w/o vel* variant excludes the velocity state during agent selection. The results show

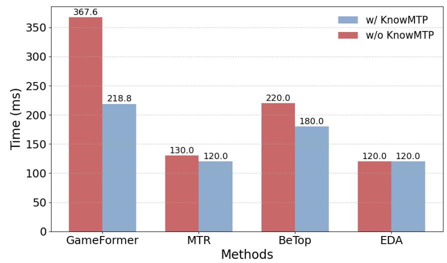

Figure 3. Inference time comparison on the WOMD dataset.

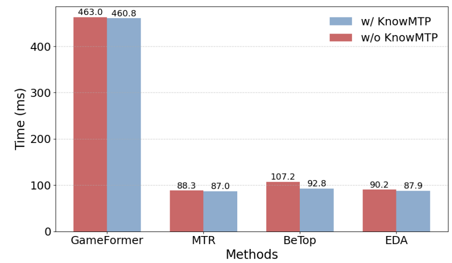

Figure 4. Inference time comparison on the nuPlan dataset.

a consistent decrease in performance across all combined baselines, particularly for KnowMTP+GameFormer, suggesting that velocity is a key factor in accurately identifying influential agents. The *w/o angle* variant removes the orientation (angle) information from the selection process. This modification leads to a consistent drop in performance across different models, particularly for KnowMTP+MTR and KnowMTP+BeTop. Among all motion-specific variants, this yields the most significant degradation, indicating that angle-based selection plays a crucial role in enhancing prediction accuracy. These results highlight the strong influence of orientation in determining motion similarity with the target agent and its substantial contribution to the overall performance of the KnowMTP framework. Lastly, the *w/o type* variant excludes agent type information. This leads to a decline in performance across multiple combined models, demonstrating its added value in identifying contextually relevant agents. Overall, these ablation results confirm that each state factor used in motion-similar agent selection, including position, velocity, angle, and type, makes a meaningful and deliberated contribution to the performance of the KnowMTP framework.

Moreover, we investigate the effectiveness of safety-critical agent selection by constructing a variant denoted as *w/o safety*. It is important to note that we do not further decompose the variant into individual traffic safety factors: TTC, TET, and TIT, since they collectively represent perceived safety and together capture the temporal continuity of driving risk. Removing any one of them would compromise the integrity of the comprehensive and continuous driving safety representation. The results show that the *w/o safety* variant leads to a consistent and substantial degradation in prediction performance across different KnowMTP-integrated baselines, thereby validating that modeling local driving safety is crucial for improving multi-agent interaction understanding and overall trajectory prediction accuracy.

5.4. Inference Time Comparison

To evaluate the efficiency of the proposed KnowMTP framework, we conduct a comparative study of inference times on both the WOMD and nuPlan datasets. Figures 3-4 illustrate the inference times for the four SOTA baseline models and their KnowMTP-integrated counterparts on WOMD and nuPlan datasets, respectively. Following the evaluation protocol of VisionTrap [26], we assess each model's efficiency by computing the average inference time per test scenario. All experiments are conducted on benchmark hardware (4×NVIDIA GeForce RTX 4090 24GB GPUs, Ubuntu 22.04.5) to ensure fair comparison. The results demonstrate that KnowMTP consistently enhances the inference efficiency of the baseline models, achieving average improvements of 4.5-22.1% across both datasets. This improvement is primarily attributed to the proposed agent selection mechanism, which filters out irrelevant surrounding agents by measuring motion similarity and safety criticality to the target agent. This not only reduces noise in the input data but also lowers computational overhead. Although the selection process introduces additional computations, such as metric embedding and similarity evaluation, the associated overhead is negligible compared to the efficiency gained through agent filtering. Overall, these improvements underscore the potential of KnowMTP to enable real-time trajectory prediction in practical autonomous driving scenarios.

5.5. Qualitative Analysis

To further examine the benefits that KnowMTP brings to baseline models, we present visualization results to demonstrate performance improvements across different driving scenarios. Specifically, we qualitatively analyze the predicted trajectories of the baseline models and their KnowMTP-enhanced counterparts on the WOMD dataset in two common traffic scenarios: signalized intersections and main roads. As shown in Figure 5, we select GameFormer and MTR for comparison. The first (a) and second (b) rows illustrate the results in signalized intersection and main road scenarios, respectively. From these visualizations, we observe that KnowMTP consistently produces more plausible and realistic predicted trajectories in both scenarios. In contrast, the baseline models (GameFormer and MTR) often generate predictions that violate traffic regulations, such as running red lights or crossing yellow solid lines, which could lead to potential collisions. These results

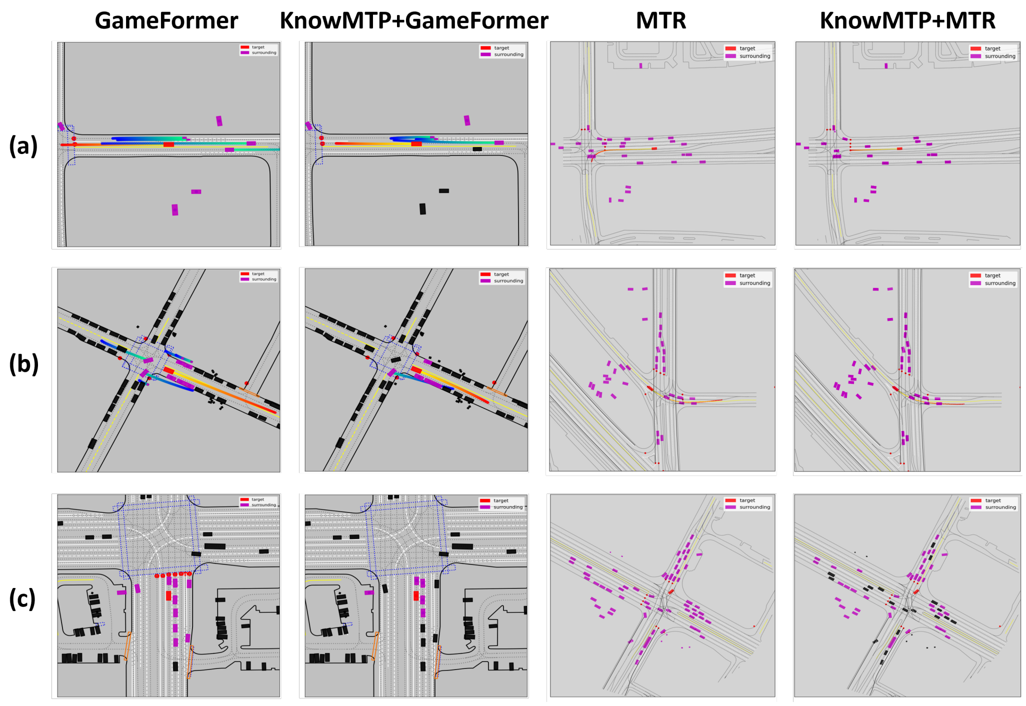

Figure 5. Visualization of the baseline models (GameFormer and MTR) and their KnowMTP-enhanced counterparts on the WOMD dataset. We present prediction results in two common scenarios: (a) signalized intersections (first row) and (b) main roads (second row), both highlighting the socially aware behavior enabled by KnowMTP. Additionally, (c) the third row illustrates the effectiveness of the agent selection mechanism, demonstrating how the agent filter contributes to socially aware trajectory prediction.

demonstrate that KnowMTP enhances the realism of trajectory predictions, making the baseline models more *socially aware*.

To further understand the rationale behind the performance improvements brought by KnowMTP, we visualize the results of the agent selection in the third row (c) of Figure 5. As illustrated, our proposed agent filtering mechanism effectively removes irrelevant surrounding agents (evidenced by the noticeable reduction in purple agents) for the target agent, thereby reducing input noise and allowing the baseline models to make more informed and sensible predictions. For instance, agents driving in the opposite lanes are correctly filtered out, as they have no meaningful interaction with the target agent from both motion and safety perspectives. This selective mechanism contributes to the socially aware property of KnowMTP.

6. Conclusion

In this paper, we propose KnowMTP, a knowledge-guided framework for multi-agent trajectory prediction that explicitly incorporates traffic knowledge to enhance the performance of SOTA prediction models. Experimental results demonstrate that KnowMTP can significantly improve baseline models while reducing computational overhead. Furthermore, qualitative analysis confirms that KnowMTP enables the baselines to generate more socially aware and plausible trajectories, highlighting its potential for advancing knowledge-driven autonomous driving.

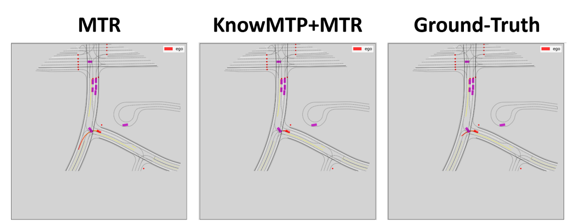

Figure 6. Failure case of KnowMTP, where the enhanced MTP model yields an unrealistic long waiting time at an intersection.

Failure Case: Despite its substantial improvement in prediction accuracy, KnowMTP has a notable limitation. In certain cases, the perceived safety may lead the model to overly conservative behaviors, resulting in unrealistic trajectory predictions. As shown in Figure 6, integrating KnowMTP with MTR causes the agent to make a prolonged stop at a T-intersection, deviating from the ground truth and performing worse than the original MTR. This limitation warrants the future direction in balancing safety awareness and behavioral realism in trajectory prediction.

Acknowledgments

This study is funded in part by the National Natural Science Foundation of China under Grant No. 62402103, in part by the Natural Science Foundation of Jiangsu Province under Grant No. BK20241273, in part by the National Natural Science Foundation of China under Grant No. 62572120, and in part by the Natural Science Foundation of Jiangsu Province under Grant No. BK20230024.

References

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

[2] Nicolas Baumann, Edoardo Ghignone, Cheng Hu, Benedict Hildisch, Tino Hämmerle, Alessandro Bettoni, Andrea Carron, Lei Xie, and Michele Magno. Predictive spliner: Data-driven overtaking in autonomous racing using opponent trajectory prediction. *IEEE Robotics and Automation Letters*, 2024. 1

[3] Charlotte Beaune, Elwan Héry, and Vincent Frémont. Risk-aware nonlinear model predictive control for autonomous navigation: Confidence-based obstacle constraints and time-to-collision safe trajectory planning. In *2025 IEEE Intelligent Vehicles Symposium (IV)*, pages 1633–1639. IEEE, 2025. 3

[4] Holger Caesar, Juraj Kabzan, Kok Seang Tan, Whye Kit Fong, Eric Wolff, Alex Lang, Luke Fletcher, Oscar Beijbom, and Sammy Omari. nuplan: A closed-loop ml-based planning benchmark for autonomous vehicles. *arXiv preprint arXiv:2106.11810*, 2021. 5

[5] Weihuang Chen, Fangfang Wang, and Hongbin Sun. S2tnet: Spatio-temporal transformer networks for trajectory prediction in autonomous driving. In *Asian conference on machine learning*, pages 454–469. PMLR, 2021. 3

[6] 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. 1

[7] Scott Ettinger, Shuyang Cheng, Benjamin Caine, Chenxi Liu, Hang Zhao, Sabeek Pradhan, Yuning Chai, Ben Sapp, Charles R Qi, Yin Zhou, et al. Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9710–9719, 2021. 5

[8] 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 *CVPR*, pages 11522–11530. Computer Vision Foundation / IEEE, 2020. 3

[9] Thomas Gilles, Stefano Sabatini, Dzmitry Tsishkou, Bogdan Stanciulescu, and Fabien Moutarde. HOME: heatmap output for future motion estimation. In *ITSC*, pages 500–507. IEEE, 2021. 2

[10] Junru Gu, Chen Sun, and Hang Zhao. Densetnt: End-to-end trajectory prediction from dense goal sets. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15303–15312, 2021. 3

[11] Peng Hang, Chen Lv, Yang Xing, Chao Huang, and Zhongxu Hu. Human-like decision making for autonomous driving: A noncooperative game theoretic approach. *IEEE Transactions on Intelligent Transportation Systems*, 22(4):2076–2087, 2020. 3

[12] Xiangkun He, Jingda Wu, Zhiyu Huang, Zhongxu Hu, Jun Wang, Alberto Sangiovanni-Vincentelli, and Chen Lv. Fear-neuro-inspired reinforcement learning for safe autonomous driving. *IEEE transactions on pattern analysis and machine intelligence*, 2023. 4

[13] Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, et al. Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17853–17862, 2023. 1

[14] Zhiyu Huang, Haochen Liu, and Chen Lv. Gameformer: Game-theoretic modeling and learning of transformer-based interactive prediction and planning for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3903–3913, 2023. 1, 3, 4, 5

[15] Xiaosong Jia, Penghao Wu, Li Chen, Yu Liu, Hongyang Li, and Junchi Yan. Hdgt: Heterogeneous driving graph transformer for multi-agent trajectory prediction via scene encoding. *IEEE transactions on pattern analysis and machine intelligence*, 45(11):13860–13875, 2023. 3

[16] Boqian Li, Shengbo Wang, Zhenyuan Guo, Song Zhu, Junjian Huang, Junwei Sun, Guanghui Wen, and Shiping Wen. Safe control framework of multi-agent systems from a performance enhancement perspective. *IEEE Transactions on Automation Science and Engineering*, 2024. 3

[17] Longzhong Lin, Xuewu Lin, Tianwei Lin, Lichao Huang, Rong Xiong, and Yue Wang. Eda: Evolving and distinct anchors for multimodal motion prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3432–3440, 2024. 5

[18] Pengfei Lin, Ehsan Javanmardi, Ye Tao, Vishal Chauhan, Jin Nakazato, and Manabu Tsukada. Time-to-collision-aware lane-change strategy based on potential field and cubic polynomial for autonomous vehicles. In *2023 IEEE Intelligent Vehicles Symposium (IV)*, pages 1–6. IEEE, 2023. 3

[19] Haochen Liu, Li Chen, Yu Qiao, Chen Lv, and Hongyang Li. Reasoning multi-agent behavioral topology for interactive autonomous driving. *Advances in Neural Information Processing Systems*, 37:92605–92637, 2024. 5

[20] Haochen Liu, Li Chen, Yu Qiao, Chen Lv, and Hongyang Li. Reasoning multi-agent behavioral topology for interactive autonomous driving. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 3

[21] Qichao Liu, Heye Huang, Shiyue Zhao, Lei Shi, Soyoung Ahn, and Xiaopeng Li. Risknet: interaction-aware risk forecasting for autonomous driving in long-tail scenarios. *Transportation Research Part E: Logistics and Transportation Review*, 205:104478, 2026. 3

[22] Yuhuan Lu, Wei Wang, Rufan Bai, Shengwei Zhou, Lalit Garg, Ali Kashif Bashir, Weiwei Jiang, and Xiping Hu. Hyper-relational interaction modeling in multi-modal trajectory prediction for intelligent connected vehicles in smart cites. *Information Fusion*, 114:102682, 2025. 3

[23] Yuhuan Lu, Pengpeng Xu, Xinyu Jiang, Ali Kashif Bashir, Thippa Reddy Gadekallu, Wei Wang, and Xiping Hu. Lane change prediction for autonomous driving with transferred trajectory interaction. *IEEE Transactions on Intelligent Transportation Systems*, 2025. 1

[24] Michiel M Minderhoud and Piet HL Bovy. Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis & Prevention*, 33(1):89–97, 2001. 4

[25] Xiaoyu Mo, Yang Xing, and Chen Lv. Graph and recurrent neural network-based vehicle trajectory prediction for highway driving. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 1934–1939. IEEE, 2021. 1

[26] Seokha Moon, Hyun Woo, Hongbeen Park, Haeji Jung, Reza Mahjourian, Hyung-gun Chi, Hyerin Lim, Sangpil Kim, and Jinkyu Kim. Visiontrap: Vision-augmented trajectory prediction guided by textual descriptions. In *European Conference on Computer Vision*, pages 361–379. Springer, 2024. 7

[27] Jiquan Ngiam, Benjamin Caine, Vijay Vasudevan, Zhengdong Zhang, Hao-Tien Lewis Chiang, Jeffrey Ling, Rebecca Roelofs, Alex Bewley, Chenxi Liu, Ashish Venugopal, et al. Scene transformer: A unified architecture for predicting multiple agent trajectories. *arXiv preprint arXiv:2106.08417*, 2021. 3

[28] Mingxing Peng, Xusen Guo, Xianda Chen, Kehua Chen, Meixin Zhu, Long Chen, and Fei-Yue Wang. Lc-llm: Explainable lane-change intention and trajectory predictions with large language models. *Communications in Transportation Research*, 5:100170, 2025. 1

[29] Yusheng Peng, Gaofeng Zhang, Jun Shi, Benzhu Xu, and Liping Zheng. Srai-lstm: A social relation attention-based interaction-aware lstm for human trajectory prediction. *Neurocomputing*, 490:258–268, 2022. 2

[30] Xuanchi Ren, Tao Yang, Li Erran Li, Alexandre Alahi, and Qifeng Chen. Safety-aware motion prediction with unseen vehicles for autonomous driving. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15731–15740, 2021. 3

[31] Shaoshuai Shi, Li Jiang, Dengxin Dai, and Bernt Schiele. Motion transformer with global intention localization and local movement refinement. *Advances in Neural Information Processing Systems*, 35:6531–6543, 2022. 3

[32] Shaoshuai Shi, Li Jiang, Dengxin Dai, and Bernt Schiele. Mtr++: Multi-agent motion prediction with symmetric scene modeling and guided intention querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5):3955–3971, 2024. 1, 3, 4, 5

[33] Jie Sun, Senbo Yan, and Xiaowen Song. Qcnet: query context network for salient object detection of automatic surface inspection. *The Visual Computer*, 39(10):4391–4403, 2023. 3

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

[35] Licheng Wen, Daocheng Fu, Xin Li, Xinyu Cai, Tao Ma, Pinlong Cai, Min Dou, Botian Shi, Liang He, and Yu Qiao. Dilu: A knowledge-driven approach to autonomous driving with large language models. In *ICLR*. OpenReview.net, 2024. 1

[36] Yongjun Yan, Lin Peng, Tong Shen, Jinxiang Wang, Dawei Pi, Dongpu Cao, and Guodong Yin. A multi-vehicle game-theoretic framework for decision making and planning of autonomous vehicles in mixed traffic. *IEEE Transactions on Intelligent Vehicles*, 8(11):4572–4587, 2023. 3

[37] Mingxia Yang, Boliang Zhang, Tingting Wang, Jijing Cai, Xiang Weng, Hailin Feng, and Kai Fang. Vehicle interactive dynamic graph neural network-based trajectory prediction for internet of vehicles. *IEEE Internet of Things Journal*, 11(22):35777–35790, 2024. 1

[38] Yuxiang Zhang, Xiaoling Liang, Dongyu Li, Shuzhi Sam Ge, Bingzhao Gao, Hong Chen, and Tong Heng Lee. Adaptive safe reinforcement learning with full-state constraints and constrained adaptation for autonomous vehicles. *IEEE Transactions on Cybernetics*, 54(3):1907–1920, 2023. 3

[39] Zhejun Zhang, Alexander Liniger, Christos Sakaridis, Fisher Yu, and Luc V Gool. Real-time motion prediction via heterogeneous polyline transformer with relative pose encoding. *Advances in Neural Information Processing Systems*, 36:57481–57499, 2023. 3

[40] Ze Zhang, Georg Hess, Emmanuel Dean, Lennart Svensson, and Knut Åkesson. Future-oriented navigation: Dynamic obstacle avoidance with one-shot energy-based multimodal motion prediction. *IEEE Robotics and Automation Letters*, 2025. 1

[41] Yang Zhou, Hao Shao, Letian Wang, Steven L Waslander, Hongsheng Li, and Yu Liu. Smartrefine: A scenario-adaptive refinement framework for efficient motion prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15281–15290, 2024. 1

[42] Zikang Zhou, Luyao Ye, Jianping Wang, Kui Wu, and Kejie Lu. Hivt: Hierarchical vector transformer for multi-agent motion prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8823–8833, 2022. 1