Learning Cooperative Trajectory Representations for Motion Forecasting

Hongzhi Ruan, Haibao Yu, Wenxian Yang, Siqi Fan, Zaiqing Nie

NeurIPS 2024

Why Cooperative Motion Forecasting

- Most of the existing works pay more attention to cooperative perception, mainly focusing on single-frame cooperation to enhance perception ability.

- Motion forecasting is a downstream task of perception, which directly influences the actions of the autonomous vehicle and has not been well explored.

Problem Formulation

Input

Cooperative Scenario $\mathcal{S} = \{\mathbf{T}, \mathbf{L}\}$

- Multi-source Trajectory

  $\mathbf{T} = \{\mathbf{T}_{ego}, \mathbf{T}_{other}\}$ $\mathbf{T} \in \mathbb{R}^{N_t \times T \times C_t}$ C_t includes id, location, heading, detection bounding etc

- Vector Map

  $\mathbf{L} \in \mathbb{R}^{N_l \times 2 \times C_l}$ C_l includes location and road type etc

Output

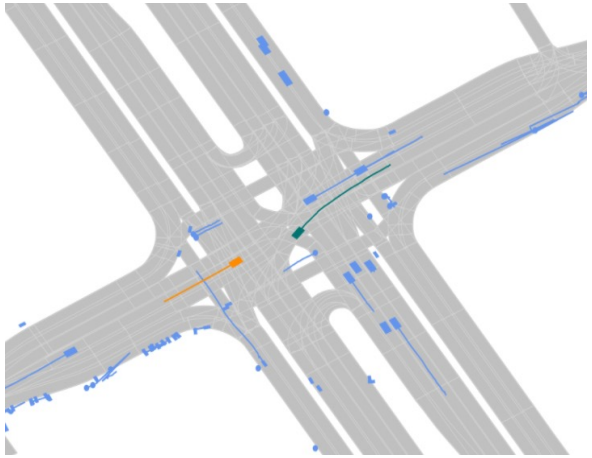
- K future trajectories of the target agent $\mathcal{O} = \{(\mathbf{p}_k^1, \mathbf{p}_k^2, \cdots, \mathbf{p}_k^H)\}_{k=1}^{\mathcal{K}}$, $\mathbf{p}_k^t = (x_k^t, y_k^t)$

Challenges of Cooperative Motion Forecasting

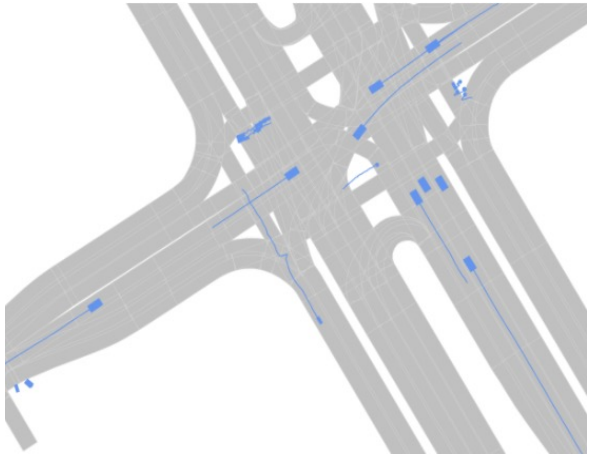
NEURAL INFORMATION PROCESSING SYSTEMS

- Observations of the agents from different views may different due to various sensor perspectives and configurations;

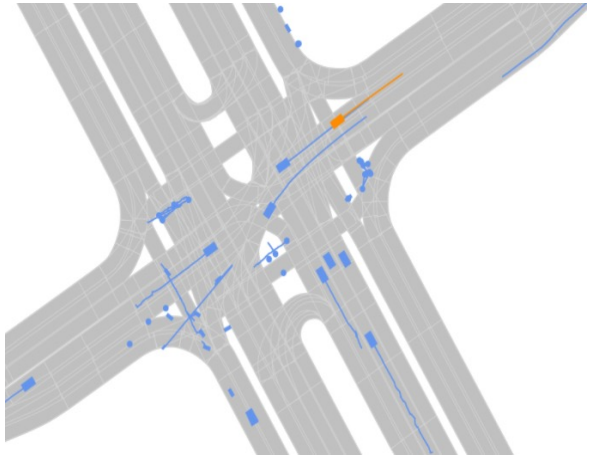
- In the cooperative scenario, there are multi-view observations of multi-agents, and the redundant data need to be leveraged interpretably.



Ego-view

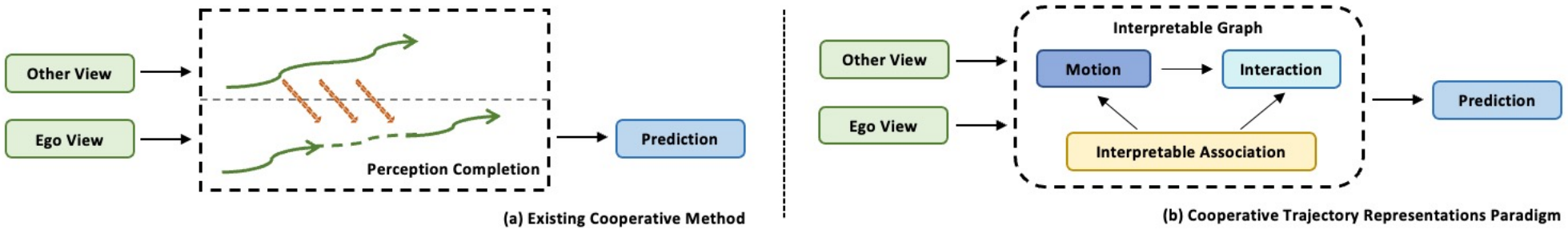


Infra-view



Veh-view

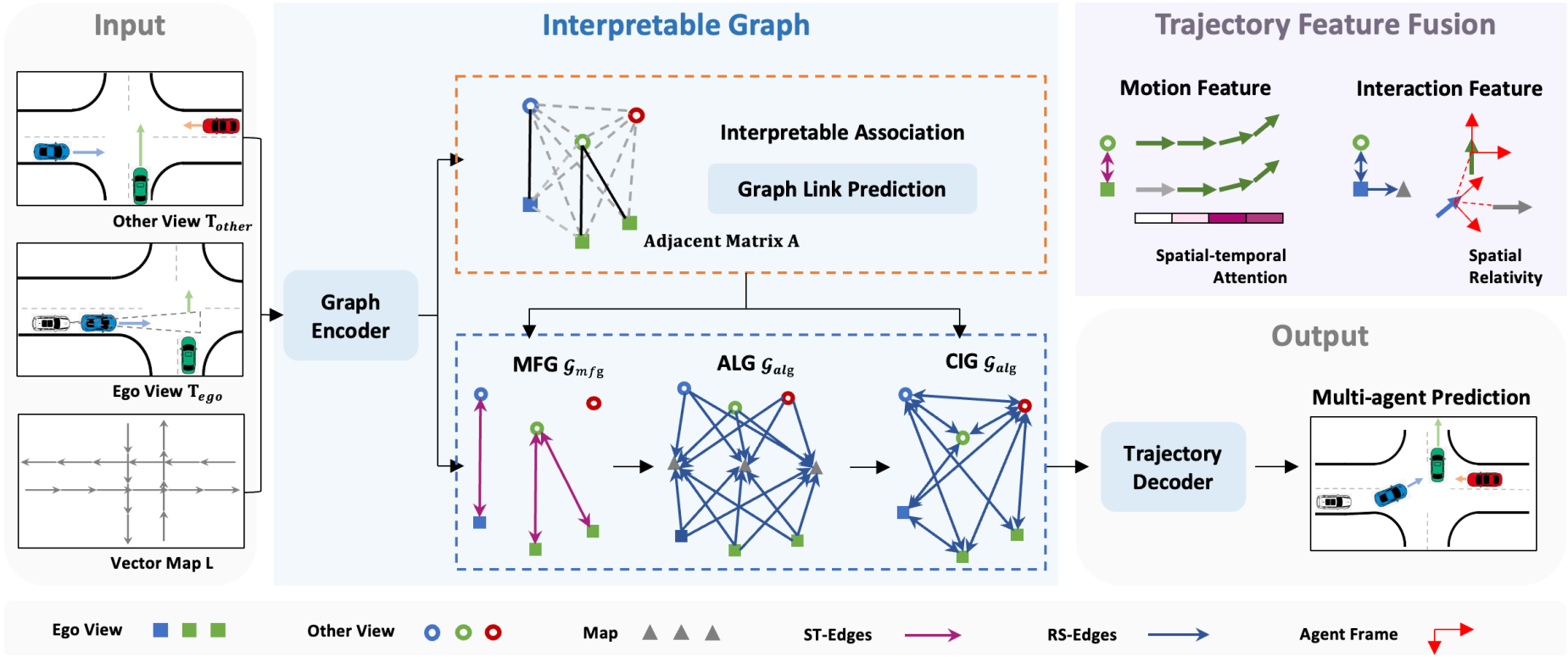
Main Idea



- Existing single-frame methods obtain the agent state at each frame individually, failing to sufficiently model the historical behavior.

- Our research pioneers the exploration of trajectory-based feature fusion for cooperative motion forecasting.

V2X-Graph Overview

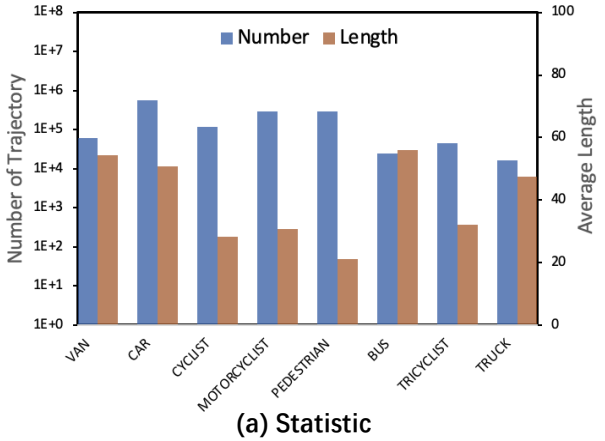


- Forecasting-oriented cooperative trajectory representations of motion and interaction features.
- Graph-guided interpretable cross-view feature fusion.

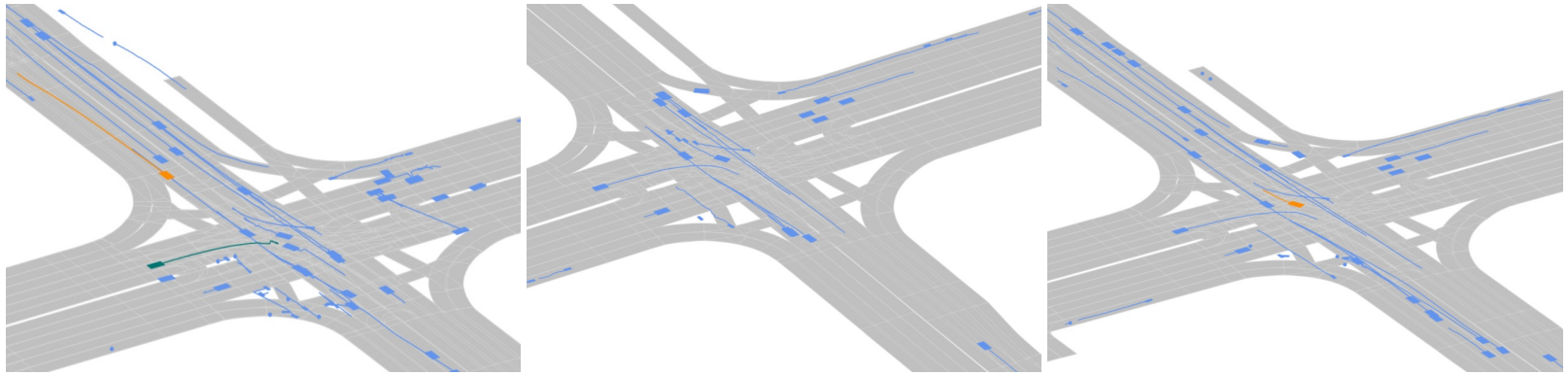
V2X-Traj Dataset

NEURAL INFORMATION PROCESSING SYSTEMS

- The first real-world and public V2X motion forecasting dataset, containing V2V and V2I in every scenario.
- It comprises 10,102 scenarios in challenging intersections, with each scenario lasting for 8 seconds.



(a) Statistic



(b) Visualizations of Ego-Infra-Veh Views

Experimental Results

NEURAL INFORMATION PROCESSING SYSTEMS

Cooperative method comparison on V2X-Seq

Method	DenseTNT[13]		HiVT[58]		V2X-Graph		
	Vehicle-only	PP-VIC[51]	Vehicle-only	PP-VIC[51]	Vehicle-only	PP-VIC[51]	Feature Fusion
minADE	1.71	1.84	1.28	1.12	1.16	1.12	**1.05**
minFDE	2.43	2.56	2.15	1.97	2.04	1.98	**1.79**
MR	0.27	0.28	0.31	0.30	0.30	0.30	**0.25**

- Perception completion leads to error propagation, resulting in performance degradation of motion forecasting.
- V2X-Graph demonstrates its effectiveness with clear performance improvements over single-frame methods.

Experimental Results

NEURAL INFORMATION PROCESSING SYSTEMS

Graph-based methods comparison on V2X-Traj

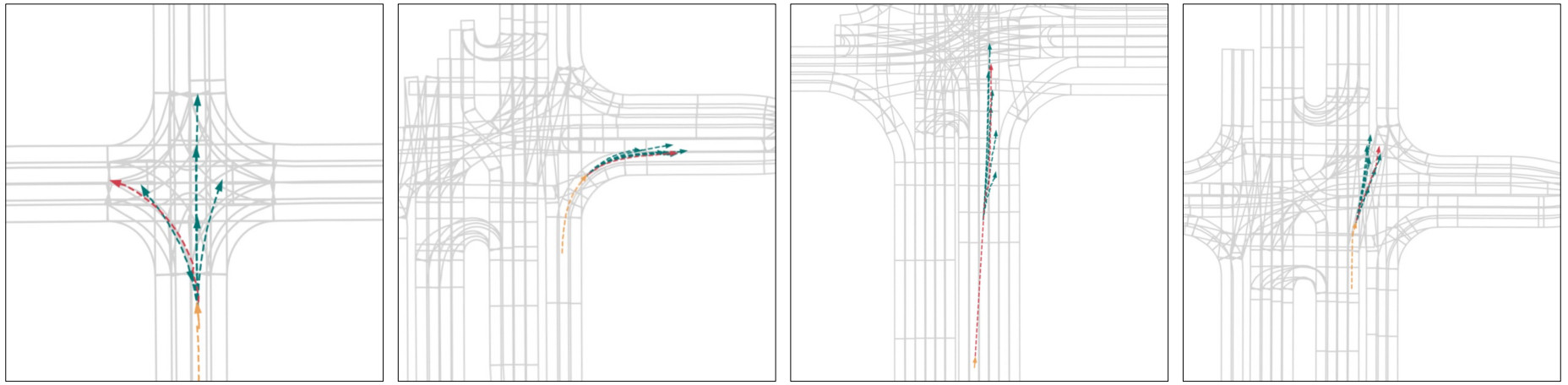
Method	Vehicle-only			V2V			V2I			V2V&I		
	minADE	minFDE	MR	minADE	minFDE	MR	minADE	minFDE	MR	minADE	minFDE	MR
DenseTNT[13]	1.23	2.09	0.25	1.20	2.04	0.25	1.32	2.34	0.29	1.26	2.24	0.28
HDGT[19]	0.91	1.48	0.14	0.94	1.57	0.17	0.94	1.59	0.16	0.94	1.56	0.17
V2X-Graph	0.90	1.56	0.17	0.77	1.26	0.12	0.80	1.30	0.13	**0.72**	**1.13**	**0.11**

- V2X-Graph outperforms the compared methods by large margins in all cooperative settings.
- The method has the potential to achieve further improved performance in more views.

Experimental Results

NEURAL INFORMATION PROCESSING SYSTEMS

Qualitative Results



Thanks for Watching!