Motion Forecasting in Continuous Driving

Nan Song[1] **Bozhou Zhang**[1] **Xiatian Zhu**[2] **Li Zhang**[1*]

[1]School of Data Science, Fudan University [2]University of Surrey

https://github.com/fudan-zvg/RealMotion

The primary distinction between previous methods and ours

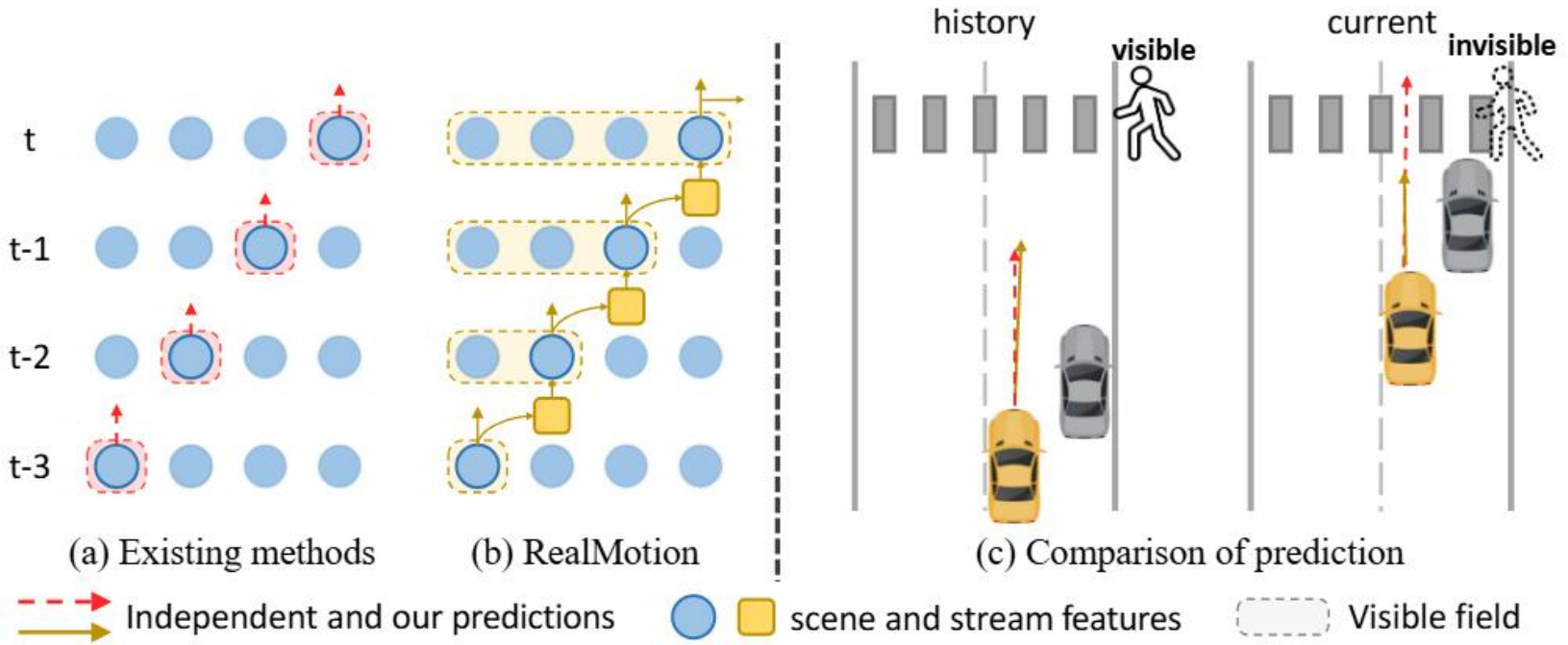

Figure 1: Comparison of **(a) existing methods** independently processing each scene and **(b) our RealMotion** recurrently collecting historical information. **(c)** For example, RealMotion can perceive the currently invisible pedestrian and predict the giving way for the interested agent.

Overview of our Data Reorganization

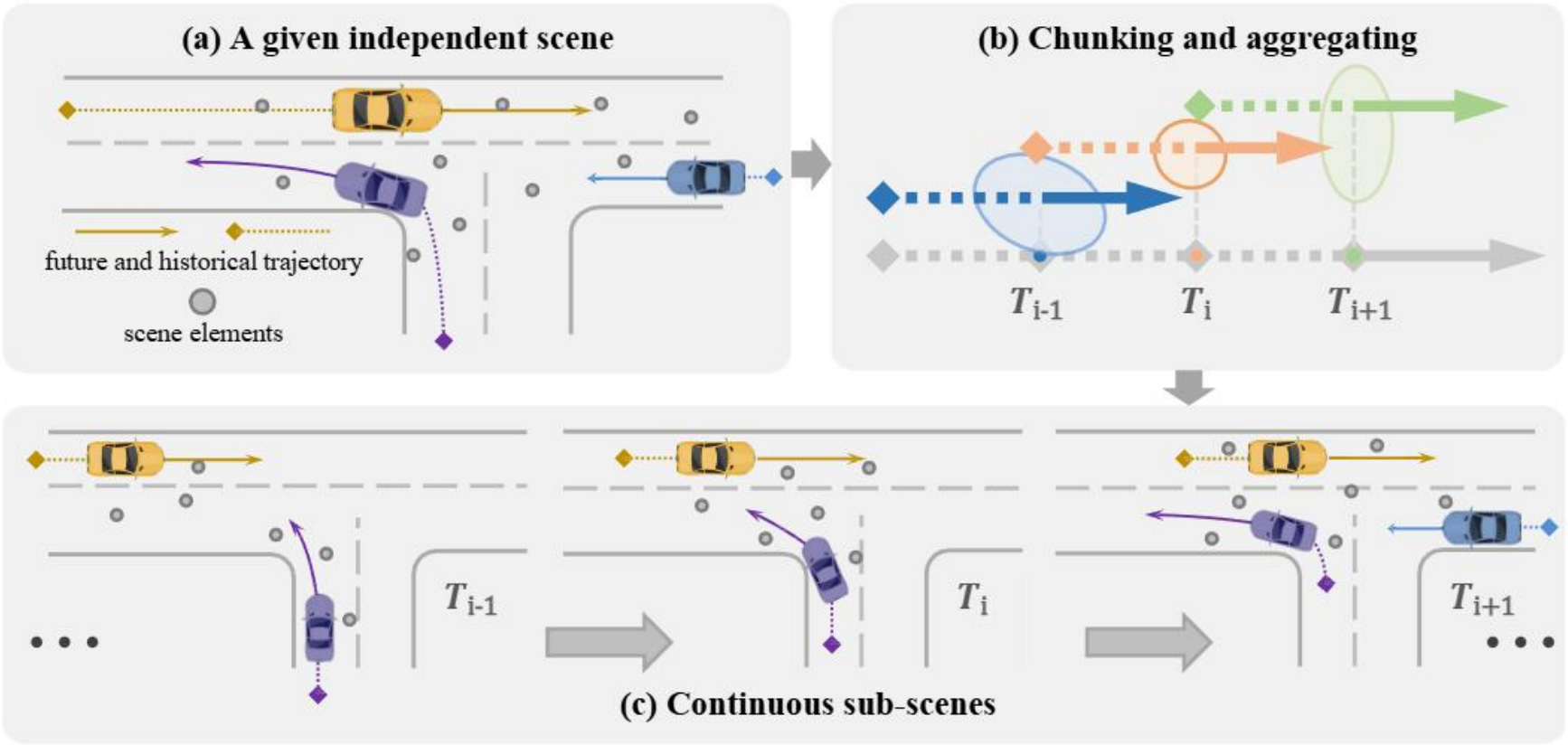

Figure 2: Illustration of our data reorganization strategy, processing (a) a given independent scene by (b) chunking the trajectories into segments and aggregating surrounding elements, generating the (c) continuous sub-scenes.

Overview of our RealMotion framework.

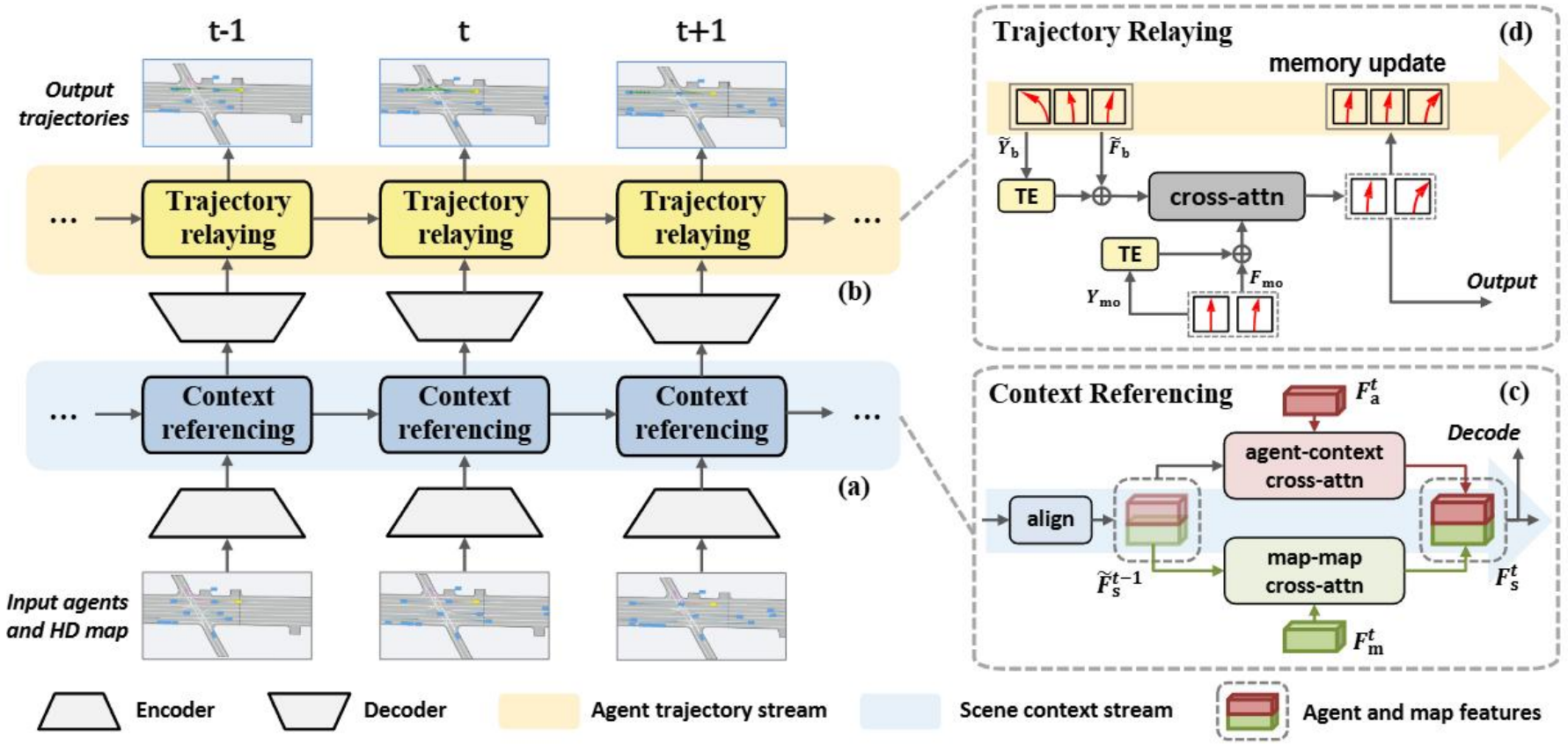

Figure 3: Overview of our RealMotion architecture. RealMotion adopts an encoder-decoder structure with two intermediate streams designed to capture interactive relationships within each scene and across the continuous scenes. The (a) **Scene context stream** and (b) **Agent trajectory stream** iteratively accumulate information for the scene context and rectify the prediction, respectively. The (c) **context referencing** and (d) **trajectory relaying** modules are specially-designed cross-attention mechanism for each stream.

Experiment on Argoverse 2.

Table 1: Performance comparison on *Argoverse 2 test set* in the official leaderboard. For each metric, the best result is in **bold** and the second best result is underlined. "-": Unreported results; "†": Methods that use model ensemble trick. RealMotion-I refers to the independent variant of our model without data reorganization and stream modules, simply taking the original trajectory as input to forecast the motion like previous methods.

Method	$minFDE_1$	$minADE_1$	$minFDE_6$	$minADE_6$	MR_6	$b\text{-}minFDE_6$
HDGT [19]	5.37	2.08	1.60	0.84	0.21	2.24
THOMAS [14]	4.71	1.95	1.51	0.88	0.20	2.16
GoRela [7]	4.62	1.82	1.48	0.76	0.22	2.01
HPTR [48]	4.61	1.84	1.43	0.73	0.19	2.03
QML† [34]	4.98	1.84	1.39	0.69	0.19	1.95
Forecast-MAE [5]	4.36	1.74	1.39	0.71	0.17	2.03
TENET† [42]	4.69	1.84	1.38	0.70	0.19	1.90
BANet† [45]	4.61	1.79	1.36	0.71	0.19	1.92
GANet [39]	4.48	1.78	1.35	0.73	0.17	1.97
SIMPL [47]	-	-	1.43	0.72	0.19	2.05
Gnet† [11]	4.40	1.72	1.34	0.69	0.18	1.90
ProphNet [41]	4.74	1.80	1.33	0.68	0.18	**1.88**
QCNet [51]	<u>4.30</u>	<u>1.69</u>	<u>1.29</u>	**0.65**	<u>0.16</u>	1.91
RealMotion-I	4.42	1.73	1.38	0.70	0.18	2.01
RealMotion	**3.93**	**1.59**	**1.24**	<u>0.66</u>	**0.15**	<u>1.89</u>

Ablation study.

Table 4: Ablation study on the core components of RealMotion on the *Argoverse 2 validation set*. "Con. Data" indicates the processed continuous scenes. "SC Strm" and "AT Strm" indicate our proposed scene context stream and agent trajectory stream, respectively.

ID	Con. Data	SC Strm	AT Strm	$minFDE_1$	$minADE_1$	$minFDE_6$	$minADE_6$	MR_6	$b\text{-}minFDE_6$
1				4.499	1.793	1.423	0.721	0.185	2.054
2	✓			4.397	1.722	1.357	0.687	0.169	2.001
3	✓	✓		4.129	1.648	1.344	0.678	0.160	1.987
4	✓		✓	4.194	1.667	1.331	0.673	0.164	1.976
5	✓	✓	✓	4.091	1.620	1.312	0.664	0.156	1.961

Table 5: Ablation study on (a) (left) the Feature Alignment and Trajectory Embedding and (b) (right) the gradient steps and split points. For (a), "C.A." and "T.A." represent the feature alignment modules used in the Context Referencing and the Trajectory Relaying blocks. "T.E." represents the Trajectory Embedding. For (b), "Grad Steps" indicates the number of steps we take to compute the gradient. "Split Pts" indicates the split points used to divide the trajectory.

C.A.	T.A.	T.E.	$minFDE_6$	$minADE_6$	MR_6
			1.334	0.681	0.163
✓			1.328	0.674	0.160
	✓		1.326	0.673	0.158
✓	✓		1.324	0.670	0.158
✓	✓	✓	1.312	0.664	0.156

Steps	Split Pts	$minFDE_6$	$minADE_6$	MR_6
1	(30, 40, 50)	1.420	0.716	0.175
2	(30, 40, 50)	1.341	0.681	0.162
	(30, 40, 50)	1.312	0.664	0.156
3	(20, 35, 50)	1.331	0.674	0.158
	(40, 45, 50)	1.365	0.692	0.163
5	(30, 35, 40, 45, 50)	1.323	0.668	0.158

Qualitative results.

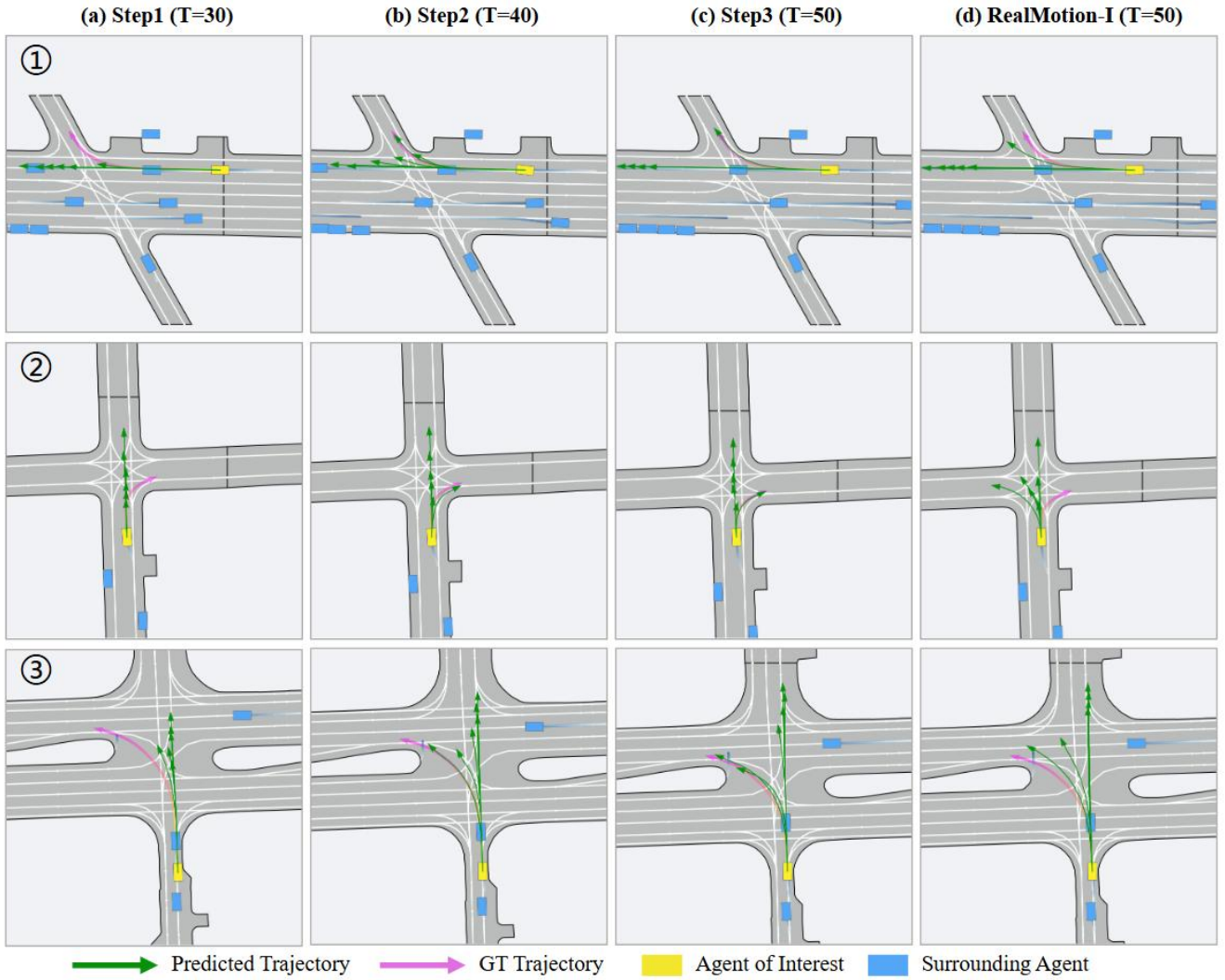

Figure 4: Qualitative results on the *Argoverse 2 validation set*. The panel (a)-(c) demonstrate the progressive forecasting results of our RealMotion, where the panel (c) is the final predictions for evaluation. The panel (d) shows the one-shot forecasting of RealMotion-I.

Thank you for listening!