SHARP: Short-Window Streaming for Accurate and Robust Prediction in Motion Forecasting

Alexander Prutsch Christian Fruhwirth-Reisinger David Schinagl Horst Possegger
alexander.prutsch@tugraz.at
Institute of Visual Computing, Graz University of Technology

Project Page | Code | LinkedIn

CVPR JUNE 3-7, 2026 DENVER COLORADO

Motivation and Contributions

- Trajectory prediction is a core component of the classical autonomous vehicle control stack, critical for safe and efficient motion planning.
- Real-world traffic scenes are **constantly changing**, resulting in heterogeneous agent observation lengths ranging from a few frames to several seconds.
- Most state-of-the-art models rely on **fixed-size observation windows without information streaming** – efficiently propagating motion features for as long as agents remain visible is an open challenge, as existing approaches fall short.
  - Prior streaming methods [1, 2] relay information but remain constrained to fixed-observation-length training inputs and use only implicit context streaming.
  - Variable-history models [3, 4, 5] train using knowledge distillation and partial masking but lack streaming capabilities for continuous, real-time evolution.

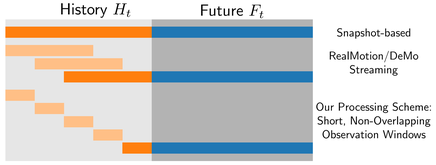
→ We propose **SHARP** (*short-window streaming for accurate and robust predictions*), a novel framework that adaptively handles **heterogeneous observation lengths in continuously evolving traffic scenes**.

→ We introduce an **instance-aware context streaming** module that efficiently propagates temporal context across inference steps.

→ Extensive evaluations on **AV2, nuScenes, and AV1** demonstrate strong robustness across variable observation lengths, a critical requirement for real-world self-driving.

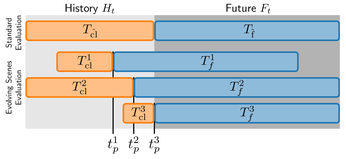
→ Our lightweight approach establishes a new state-of-the-art in the practically relevant setting of **streaming multi-agent trajectory prediction** on the AV2 dataset.

→ We use **non-overlapping short windows**, enabling continuous context accumulation and efficient real-time inference by incrementally integrating new observations.



Evolving Scenes Evaluation Protocol

While standard benchmarks evaluate models only using a fixed history H_t to predict a fixed future horizon F_t, real-world driving demands dynamic adaptability .
To provide a more comprehensive and realistic evaluation, we test at different prediction timesteps t_p, utilizing varying context lengths T_{cl}, and evaluating across multiple future horizons T_f.



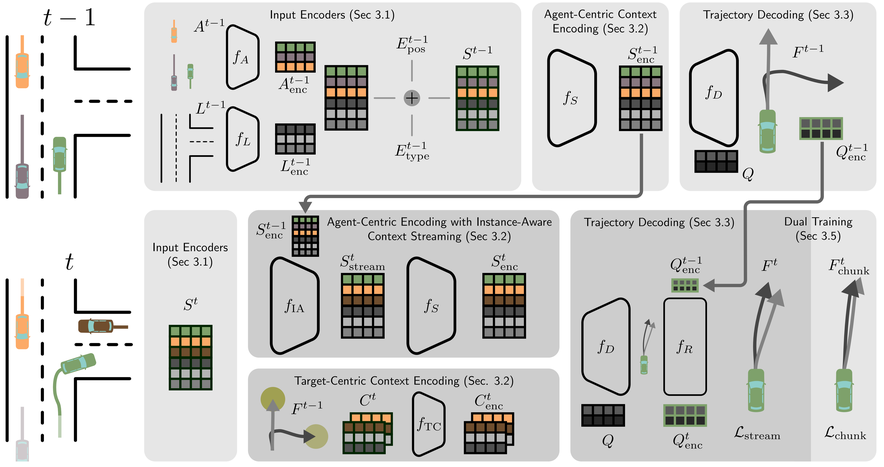
Latency Analysis (NVIDIA RTX 3090)

SHARP achieves an average online latency of just **30 ms per scenario in the AV2 multi-agent setting.**
Single-Agent Setting: SHARP requires only 33 ms, 35 ms, and 57 ms for batches of 16, 32, and 64 agents, respectively. This outpaces DeMo [2] (35/49/88 ms) and RealMotion [1] (44/79/140 ms).

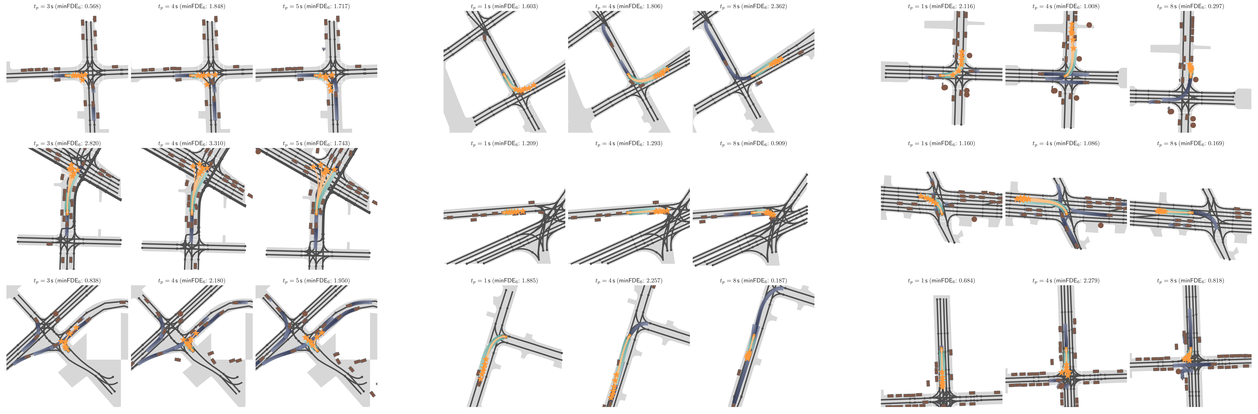
Approach

Lightweight Backbone Architecture: An efficient model processing short observation windows incrementally. It integrates a self-attention-based agent encoder f_A, a PointNet-based [6] vector map encoder f_L, a scene encoder f_S (learning relationships between agent and map elements), and a DETR-like [7] decoder f_D using mode queries Q to provide K prediction hypotheses for an agent.

- **Instance-Aware Context Streaming** f_{IA}: Explicitly models temporal agent correspondences by modulating cross-attention weights with a learned mask.
- **Target-Centric Context** f_{TC}: Leverages prior prediction endpoints as spatial anchors to efficiently extract highly relevant local scene features.
- **Trajectory Relay** [1] f_{TR}: Refines current predictions using past trajectory forecasts via a dedicated cross-attention mechanism.
- **Dual Training Scheme:** Jointly optimizes for predictions based on the accumulated streaming history ($\mathcal{L}_{\text{stream}}$) and single-observation-window predictions ($\mathcal{L}_{\text{chunk}}$) to maximize robustness across varying context lengths.
- **Scene-Wide Joint Predictions:** Combines marginal (single-agent) outputs into globally consistent, multi-agent forecasts by modeling interactions across agents and modes.



Qualitative Results (AV2 Single-Agent Validation Set)



Predictions of SHARP, ground truth, historical observations, and surrounding agents

Evolving Scenes Results (AV2 Single-Agent Validation Set)

Evaluation Setup	Method	mADE$_1$	mFDE$_1$	MR$_6$	mADE$_6$	mFDE$_6$	**b-mFDE$_6$**
$t_p = 4$ s, $T_{\text{cl}} = 3$ s, $T_f = 6$ s	RealMotion [1]	1.70	4.44	0.17	0.68	1.35	2.00
	DeMo [2]	1.58	4.21	0.16	**0.62**	1.26	1.95
	SHARP	**1.54**	**3.84**	**0.13**	0.63	**1.18**	**1.80**
$t_p = 5$ s, $T_{\text{cl}} = 3$ s, $T_f = 6$ s	RealMotion [1]	2.04	5.33	0.23	0.79	1.65	2.32
	DeMo [2]	1.92	5.16	0.25	0.79	1.78	2.48
	SHARP	**1.84**	**4.66**	**0.19**	**0.73**	**1.44**	**2.09**
$t_p = 6$ s, $T_{\text{cl}} = 3$ s, $T_f = 5$ s	RealMotion [1]	1.81	4.81	0.22	0.72	1.56	2.25
	DeMo [2]	1.70	4.66	0.25	0.74	1.75	2.45
	SHARP	**1.58**	**4.07**	**0.18**	**0.66**	**1.36**	**2.02**
$t_p = 6$ s, $T_{\text{cl}} = 6$ s, $T_f = 5$ s	RealMotion [1]	1.37	3.46	0.10	0.56	1.05	1.70
	DeMo [2]	1.23	3.15	0.10	0.52	0.98	1.66
	SHARP	**1.21**	**2.93**	**0.09**	**0.51**	**0.94**	**1.56**
$t_p = 7$ s, $T_{\text{cl}} = 7$ s, $T_f = 4$ s	RealMotion [1]	1.19	3.05	0.09	0.48	0.91	1.58
	DeMo [2]	1.07	2.77	0.11	0.48	0.93	1.62
	SHARP	**0.89**	**2.10**	**0.05**	**0.39**	**0.70**	**1.31**
$t_p = 8$ s, $T_{\text{cl}} = 8$ s, $T_f = 3$ s	RealMotion [1]	0.94	2.36	0.07	0.39	0.70	1.39
	DeMo [2]	0.84	2.15	0.10	0.41	0.78	1.48
	SHARP	**0.60**	**1.36**	**0.02**	**0.28**	**0.48**	**1.08**
$t_p = 4.5$ s, $T_{\text{cl}} = 4$ s, $T_f = 6$ s	RealMotion [1]	1.65	4.31	0.17	0.67	1.34	1.99
	DeMo [2]	**1.54**	4.11	0.15	**0.61**	1.27	1.95
	SHARP	1.59	**4.08**	**0.14**	0.64	**1.23**	**1.87**
$t_p = 5$ s, $T_{\text{cl}} = 4$ s, $T_f = 6$ s	RealMotion [1]	1.77	4.51	0.18	0.71	1.39	2.04
	DeMo [2]	**1.68**	4.33	0.18	0.68	1.38	2.06
	SHARP	1.69	**4.16**	**0.15**	**0.67**	**1.25**	**1.89**
$t_p = 5.5$ s, $T_{\text{cl}} = 4$ s, $T_f = 5.5$ s	RealMotion [1]	1.70	4.37	0.17	0.67	1.34	2.00
	DeMo [2]	1.65	4.29	0.19	0.68	1.44	2.13
	SHARP	**1.60**	**3.98**	**0.15**	**0.60**	**1.21**	**1.86**
$t_p = 5$ s, $T_{\text{cl}} = 5$ s, $T_f = 6$ s	RealMotion [1]	1.65	4.10	0.16	0.67	1.30	1.94
	DeMo [2]	**1.48**	**3.73**	**0.13**	**0.61**	**1.19**	1.86
	SHARP	1.57	3.85	0.14	0.64	1.20	**1.82**

AV2 Multi-Agent Test Set

Method	Streaming	actorMR$_6$	avgMinADE$_6$	avgMinFDE$_6$	**avgBrierMinFDE$_6$**
Forecast-MAE [8]	✗	0.19	0.69	1.55	2.24
SRefiner [9]	✗	0.19	0.68	1.52	2.21
RealMotion [1]	✓	0.18	0.62	1.32	2.01
DeMo [2]	✓	0.16	0.58	1.24	1.93
SHARP (Ours)	✓	**0.15**	**0.56**	**1.15**	**1.80**

Ablation Study (Single-Agent)

Dataset	T_{cl}	TCF	IA	DT	mADE$_6$	mFDE$_6$	**b-mFDE$_6$**
AV2 Val	5 s	✗	✗	✗	0.74	1.28	1.91
		✓	✗	✗	0.64	1.22	1.84
		✓	✓	✗	**0.63**	**1.19**	**1.81**
		✓	✓	✓	0.64	1.20	1.82
	1 s	✗	✗	✗	1.18	2.57	3.28
		✓	✗	✗	1.09	2.49	3.18
		✓	✓	✗	1.11	2.55	3.25
		✓	✓	✓	**0.76**	**1.48**	**2.13**
AV1 Val	2s	✓	✗	✗	0.62	0.96	1.58
		✓	✓	✗	**0.59**	**0.91**	**1.52**
		✓	✓	✓	**0.59**	**0.91**	1.53
	0.5s	✓	✗	✗	0.75	1.22	1.86
		✓	✓	✗	0.71	1.13	1.78
		✓	✓	✓	**0.64**	**1.00**	**1.61**

References

[1] Nan Song et al. "Motion Forecasting in Continuous Driving". In: *NeurIPS*. 2024.
[2] Bozhou Zhang et al. "DeMo: Decoupling Motion Forecasting into Directional Intentions and Dynamic States". In: *NeurIPS*. 2024.
[3] Yi Xu et al. "Adapting to Length Shift: FlexiLength Network for Trajectory Prediction". In: *CVPR*. 2024.
[4] Sheng Wang et al. "Improving Autonomous Driving Safety with POP: A Framework for Accurate Partially Observed Trajectory Predictions". In: *ICRA*. 2024.
[5] Pengfei Zhu et al. "Target-driven Self-Distillation for Partial Observed Trajectories Forecasting". In: *arXiv* (2025).
[6] Charles R Qi et al. "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation". In: *CVPR*. 2017.
[7] Nicolas Carion et al. "End-to-End Object Detection with Transformers". In: *ECCV*. 2020.
[8] Jie Cheng et al. "Forecast-MAE: Self-supervised Pre-training for Motion Forecasting with Masked Autoencoders". In: *ICCV*. 2023.
[9] Liwen Xiao et al. "SRefiner: Soft-Braid Attention for Multi-Agent Trajectory Refinement". In: *CVPR*. 2025.