ProphNet: Efficient Agent-Centric Motion Forecasting with Anchor-informed Proposals

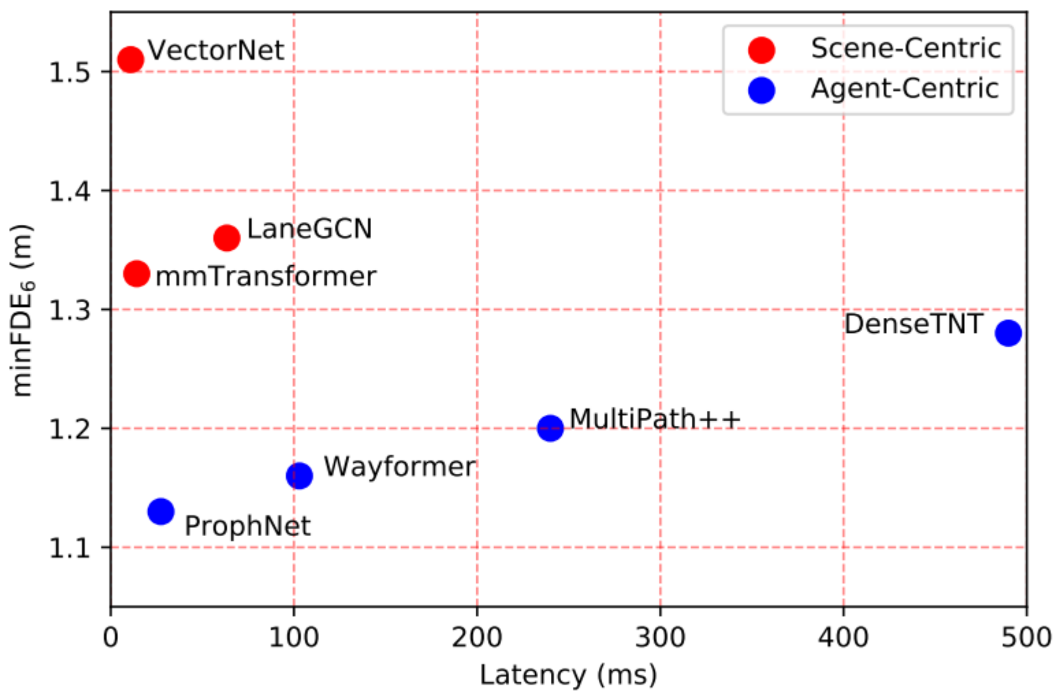
Xishun Wang, Tong Su, Fang Da, Xiaodong Yang

QCraft

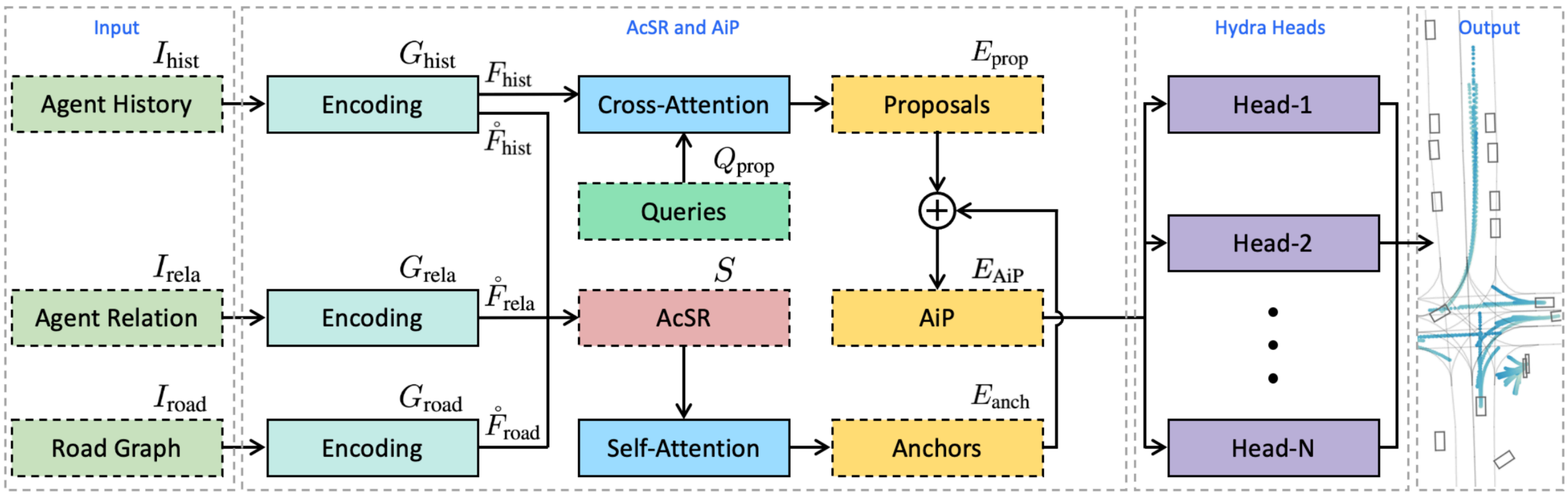
Introduction

- Motion forecasting in autonomous driving is challenging due to heterogenous nature of multi-sourced input, multimodality in agent behavior, and low inference latency required onboard deployment to predict dozens of agents.
- ProphNet involves (i) uniform encoding of heterogeneous input to construct a unified feature representation space agent-centric scene representation (AcSR), (ii) generating proposals and anchors to form anchor-informed proposals (AiP), and (iii) feeding AiP into hydra prediction heads to produce the multimodal output trajectories
- We hope this work would encourage more research toward practical model designs considered for the real-world driving deployment, with not only high prediction accuracy but also succinct architecture and efficient inference.

Introduction



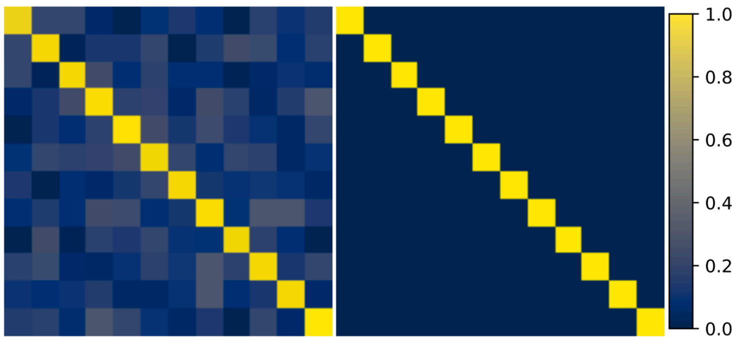
Method



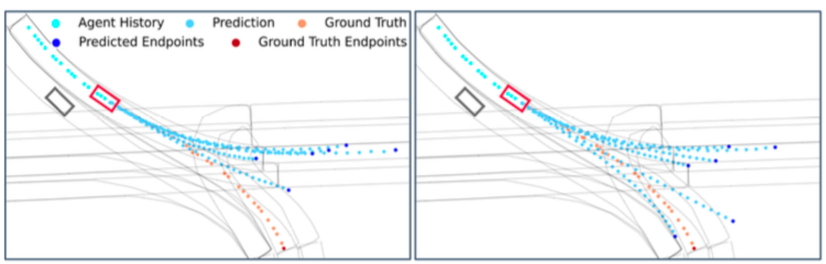
A schematic overview of ProphNet. We uniformly encode the heterogeneous input by gMLP, and combine the compact encoding features to form a unified representation space AcSR. We generate proposals through cross-attention between learnable queries and full history encoding. Anchors are learned based on self-attention of AcSR. We introduce AiP by integrating proposals and anchors, and randomly select subsets from AiP to feed into hydra prediction heads to output the final prediction. We use the solid and dashed boxes to indicate operators and operands in the network, respectively.

Method

Initialize queries orthogonally

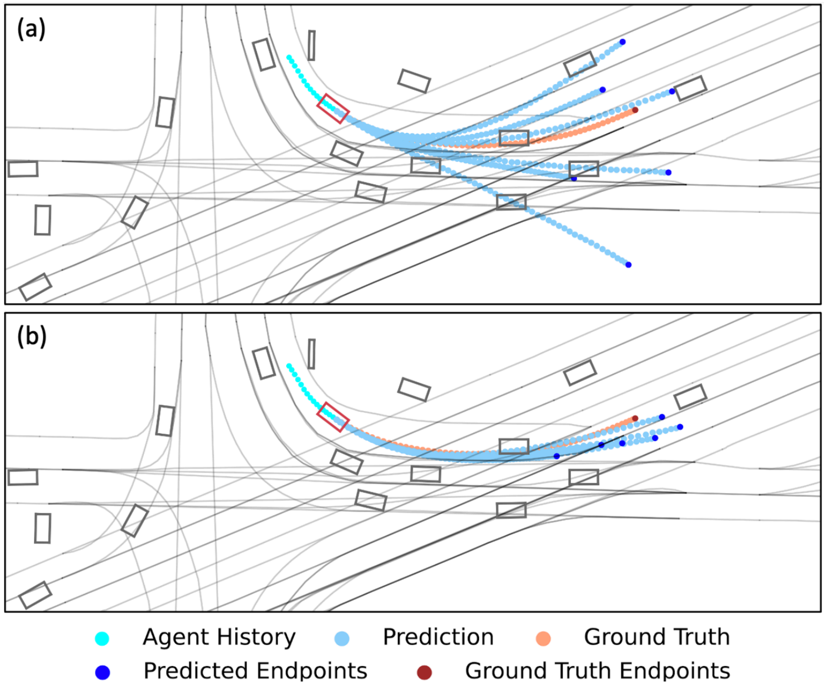


Comparison between random and orthogonal initializations



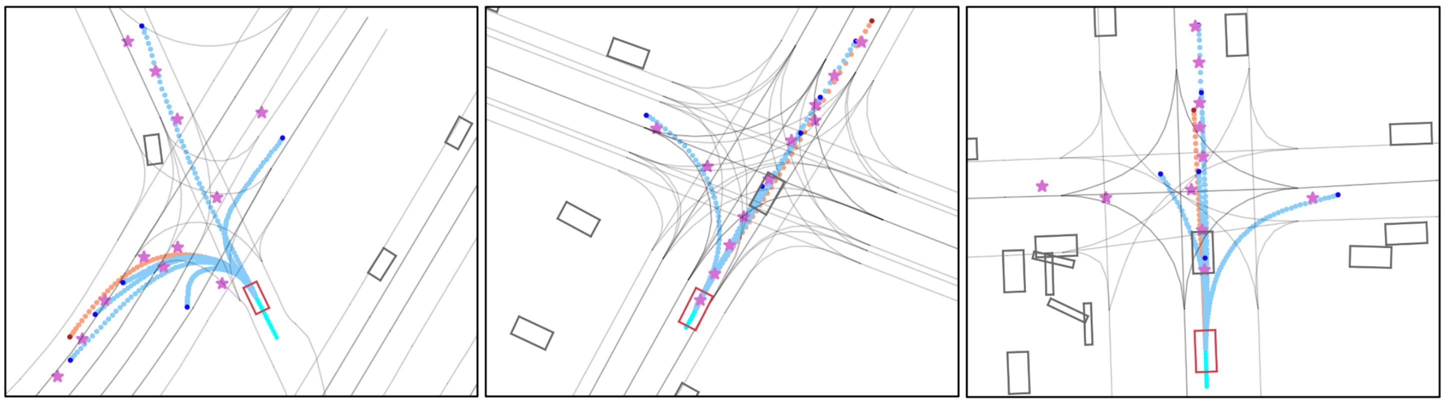
Method

Trajectories predicted by proposals only vs. trajectories predicted by AiP



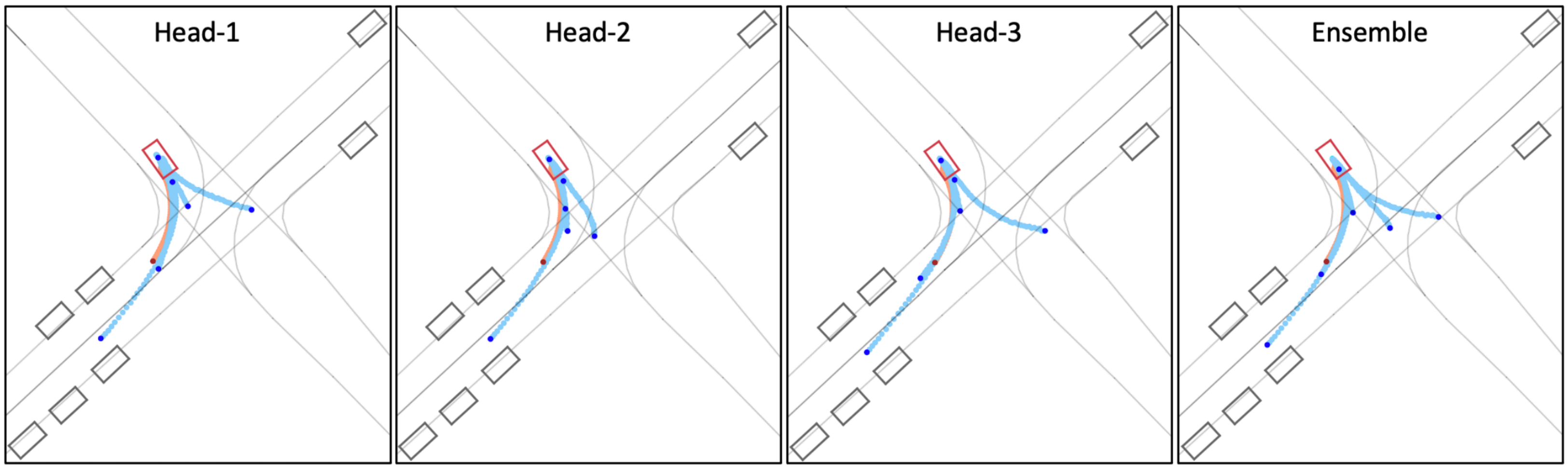
Method

Visualization of learned anchors



Method

Visualization of hydra-head prediction and ensemble effect



Results

Argoverse-1

Method	minADE$_6$	minFDE$_6$	minADE$_1$	minFDE$_1$	MR	brier-minFDE
LaneRCNN [28]	0.9038	1.4526	1.6852	3.6916	0.1232	2.1470
LaneGCN [9]	0.8703	1.3622	1.7019	3.7624	0.1620	2.0539
mmTransformer [12]	0.8436	1.3383	1.7737	4.0033	0.1540	2.0328
TPCN [26]	0.8153	1.2442	1.5752	3.4872	0.1333	1.9286
SceneTransformer [16]	0.8026	1.2321	1.8108	4.0551	0.1255	1.8868
TNT [30]	0.9097	1.4457	2.1740	4.9593	0.1656	2.1401
DenseTNT [7]	0.8817	1.2815	1.6791	3.6321	0.1258	1.9759
MultiPath++ [21]	0.7897	1.2144	1.6235	3.6141	0.1324	1.7932
Wayformer [15]	**0.7676**	1.1616	1.6360	3.6559	0.1186	1.7408
ProphNet	0.7726	**1.1442**	**1.5240**	**3.3341**	**0.1121**	**1.7323**

Argoverse-2

Method	minADE$_6$	minFDE$_6$	minADE$_1$	minFDE$_1$	MR	brier-minFDE
GOHOME [6]	0.88	1.51	1.95	4.71	0.20	2.16
GoRela [3]	0.76	1.48	1.82	4.62	0.22	2.01
TENET [24]	0.70	1.38	1.84	**4.69**	0.19	1.90
ProphNet	**0.68**	**1.33**	**1.80**	4.74	**0.18**	**1.88**

Results

Comparison on inference latency

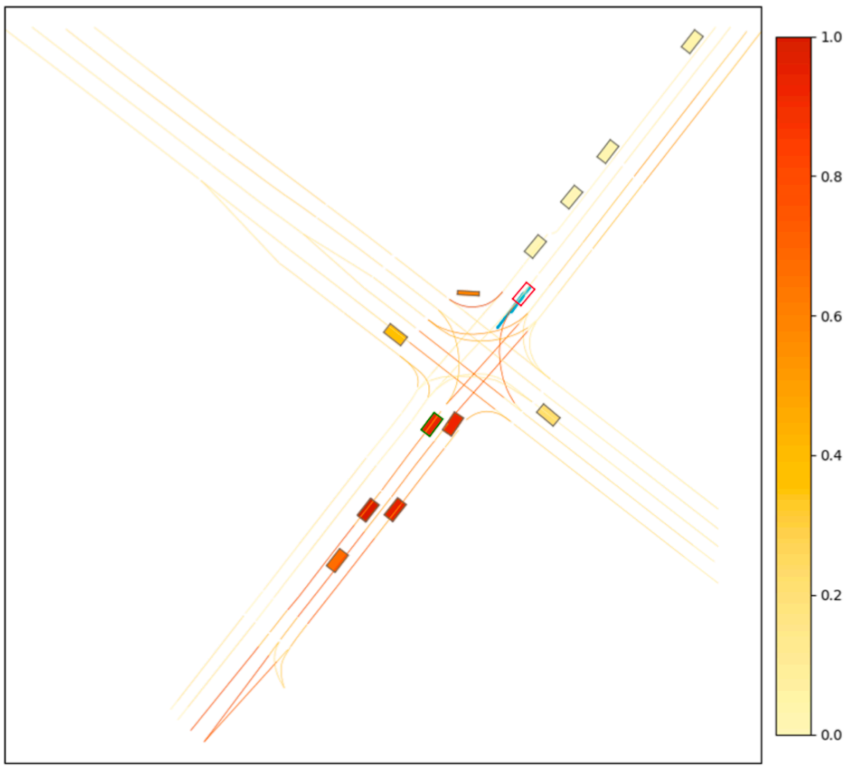
Method	Latency (ms)	GFLOPs
VectorNet [5]	10.9	0.01
LaneGCN [9]	63.4	0.13
mmTransformer [12]	14.2	0.01
DenseTNT [7]	490.1	0.62
MultiPath++ [21]	240.4	2.19
Wayformer [15]	102.3	2.13
ProphNet-S	27.4	0.39
ProphNet	28.0	0.40

Ablation

Model	minADE_1	minFDE_1
(a) Compact History Encoding	3.31	9.15
(b) Proposals	3.27	9.06
(c) Hierarchical Encoding	2.04	5.81
(d) AcSR	2.02	5.62
(e) AcSR with Anchors	2.01	5.47
(f) ProphNet-S	1.98	5.37
(g) ProphNet	1.97	5.33

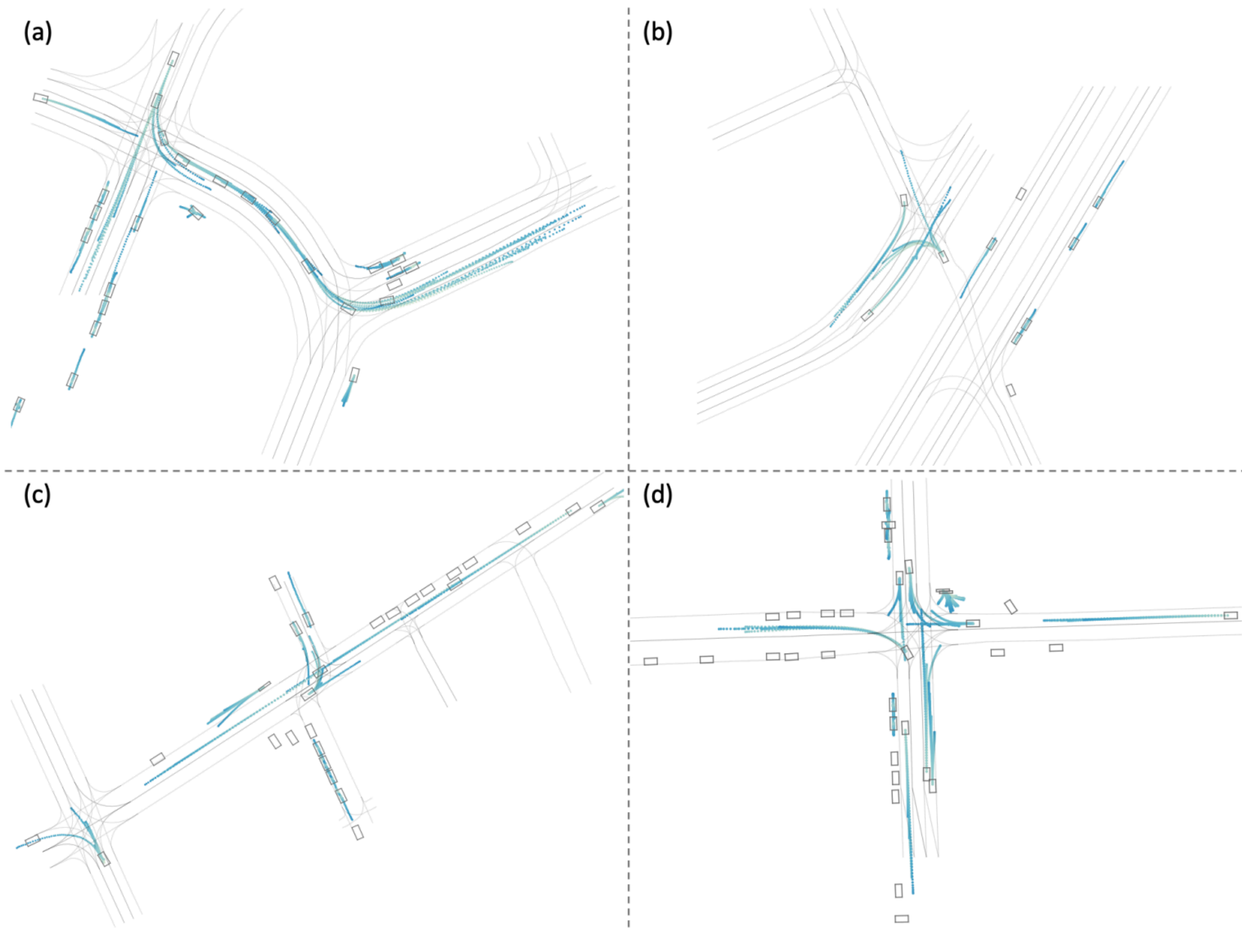
Results

Visualization of learned attention distribution



Results

Visualization in challenging scenarios



Thanks for watching!