

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

SUPPLEMENTARY MATERIAL

Xishun Wang Tong Su Fang Da Xiaodong Yang*
QCraft

In Section A, we report additional comparisons with the leaderboard results on the two benchmarks. Section B presents the visualization of learned attentions across multi-sourced input. Section C describes the details regarding the latency evaluation of various methods. Section D shows more qualitative results to illustrate the prediction of the proposed approach. Section E provides more ablation study.

A. ProphNet on Leaderboards

In the main paper, we have compared ProphNet with various published methods in Tables 2 and 3. Here we present more comparisons with the leading results (at the time of submission) including both published and unpublished methods that are reported on the leaderboards of Arogoverse-1¹ and Argoverse-2².

To compare with the leaderboard results on Arogoverse-1, we follow the common practice of model ensembling and train six models with different random seeds. We then simply use non-maximum suppression to merge all predicted trajectories to produce the final prediction output. As demonstrated in Table 6, ProphNet ranks the 1st among all 278 submissions.

As for the leaderboard comparison on Argoverse-2, we train three different models for ensembling, similar as that in Argoverse-1. As shown in Table 7, ProphNet ranks 2nd among all 32 submissions.

We then collect and compare the leading methods from the two leaderboards to see their generalizability across different benchmarks. As listed in Table 8, one can find that only ProphNet and QCNet (unpublished) yield stable and superior performance on both leaderboards.

To sum up, ProphNet achieves the 1st rank on the leaderboard of Argoverse-1 and the 2nd rank on the leaderboard of Argoverse-2. We have discussed the comparative study with published methods in the main paper. However, as the details of unpublished methods are not disclosed, we are

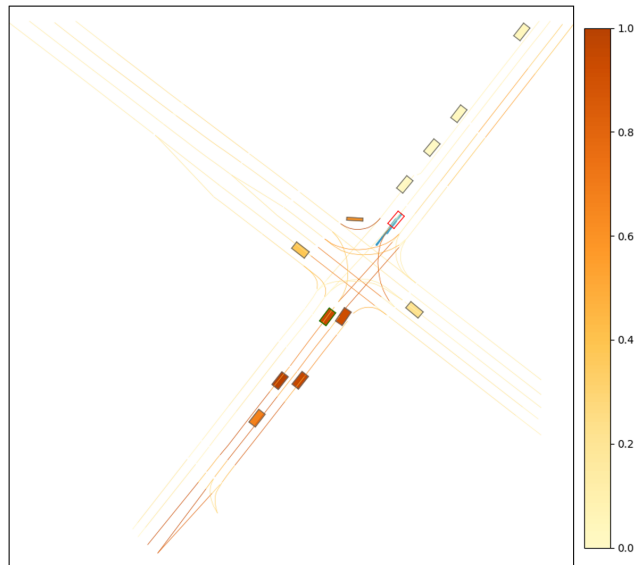


Figure 7. Illustration of the attention distribution of a target agent (red box) over the neighboring agents (ego-vehicle in green box) and road polylines learned by ProphNet.

unable to analyze the comparison with them qualitatively.

B. Visualization of Learned Attentions

Next we visualize the learned attentions to multi-sourced input for in-depth understanding of the internal attention mechanism of ProphNet. Figure 7 illustrates the attention distribution of a target agent to other neighboring agents and road polylines. As shown in this figure, we observe that the target agent pays more attention to its frontal and lateral agents that would be potentially interacted with, while less attention to the agents behind. For road polylines, the target agent attends more on the frontal ones including the straight, left or right polylines that the future trajectories may lie on. This attention distribution learned in ProphNet is reasonably alike how humans attend to different traffic elements at an intersection.

*Corresponding author xiaodong@qcraft.ai

¹<https://eval.ai/web/challenges/challenge-page/454/leaderboard/1279>

²<https://eval.ai/web/challenges/challenge-page/1719/leaderboard/4098>

Method	minADE ₆	minFDE ₆	minADE ₁	minFDE ₁	MR	brier-minFDE
Wayformer [12]	0.77	1.16	1.64	3.66	0.12	1.74
VI LaneIter*	0.77	1.11	1.52	3.28	0.11	1.73
FFINet*	0.76	1.12	1.53	3.36	0.11	1.73
QCNet*	0.74	1.07	1.54	3.37	0.11	1.70
ProphNet	0.76	1.13	1.49	3.26	0.11	1.69

Table 6. Comparison of the top 5 results of both published and unpublished methods on the leaderboard of Argoverse-1 by the date November 18, 2022. Note brier-minFDE is the primary ranking metric. * denotes unpublished methods.

Method	minADE ₆	minFDE ₆	minADE ₁	minFDE ₁	MR	brier-minFDE
OPPred*	0.71	1.36	1.79	4.61	0.19	1.92
TENET [20]	0.70	1.38	1.84	4.69	0.19	1.90
GNet*	0.69	1.34	1.72	4.40	0.18	1.90
ProphNet	0.66	1.31	1.78	4.80	0.17	1.89
QCNet*	0.62	1.19	1.56	3.96	0.14	1.78

Table 7. Comparison of the top 5 results of both published and unpublished methods on the leaderboard of Argoverse-2 by the date November 18, 2022. Note brier-minFDE is the primary ranking metric. * denotes unpublished methods.

Method	Rank (Argo-1)	Rank (Argo-2)	brier-minFDE (Argo-1)	brier-minFDE (Argo-2)
ProIn*	6	7	1.75	1.93
TENET [20]	11	4	1.77	1.90
GNet*	7	3	1.75	1.90
VI LaneIter*	4	10	1.73	2.00
QCNet*	2	1	1.70	1.78
ProphNet	1	2	1.69	1.89

Table 8. Comparison of the top ranking results of both published and unpublished methods across the leaderboards of Argoverse-1 and Argoverse-2 by the date November 18, 2022. * denotes unpublished methods.

C. Details on Latency Evaluation

In this section, we describe the details of evaluating inference latency shown in Figure 1 and Table 4. As pointed out in the main paper, in addition to prediction accuracy, inference latency is an equally important measurement of a motion forecasting model for real-world driving deployment. Table 9 lists the sources that are used to conduct the inference latency evaluation. We employ the open-sourced code and models, and evaluate them under the same setting (tested with a single NVIDIA V100 GPU and the number of agents set to 64). As a special case, since there is no open-sourced material of Wayformer [12], we implement the model following the specifications detailed in the appendix (Table 5) of related paper [12].

D. More Qualitative Results

We further provide more visualization of the predicted trajectories in challenging scenarios. Figure 8 demonstrates four representative scenes with the holistic prediction out-

put by ProphNet. As shown in this figure, (a-b) depict the complex road topologies, where ProphNet is able to produce multiple accurate trajectories that align to the lane centerlines reasonably well; and (c-d) illustrate the crowded intersections, where ProphNet predicts rich and rational multimodal future trajectories.

E. More Ablation Study

Table 10 provides more ablation studies for deeper understanding of our approach. We first compare different ways for initializing proposal queries in (a) and (f), where the random initialization is observed to be inferior to the orthogonal initialization. Figure 9 further shows the qualitative comparison. We then compare different ways of fusing proposals and anchors in (b) and (f). It is found that the simple summation performs better than the attention that incurs a higher computation cost. We next compare the direct use of anchor points with the proposed anchor embeddings in (c) and (f), and find that the latter is superior, validat-

Method	Source
VectorNet [5]	https://github.com/Henry1iu/TNT-Trajectory-Prediction
LaneGCN [9]	https://github.com/uber-research/LaneGCN
mmTransformer [11]	https://github.com/decisionforce/mmTransformer
DenseTNT [7]	https://github.com/Tsinghua-MARS-Lab/DenseTNT
MultiPath++ [18]	https://github.com/stepankonev/waymo-motion-prediction-challenge-2022-multipath-plus-plus
Wayformer [12]	Our implementation following the model and training details in the appendix (Table 5) of [12]

Table 9. Sources of different methods used for inference latency evaluation.

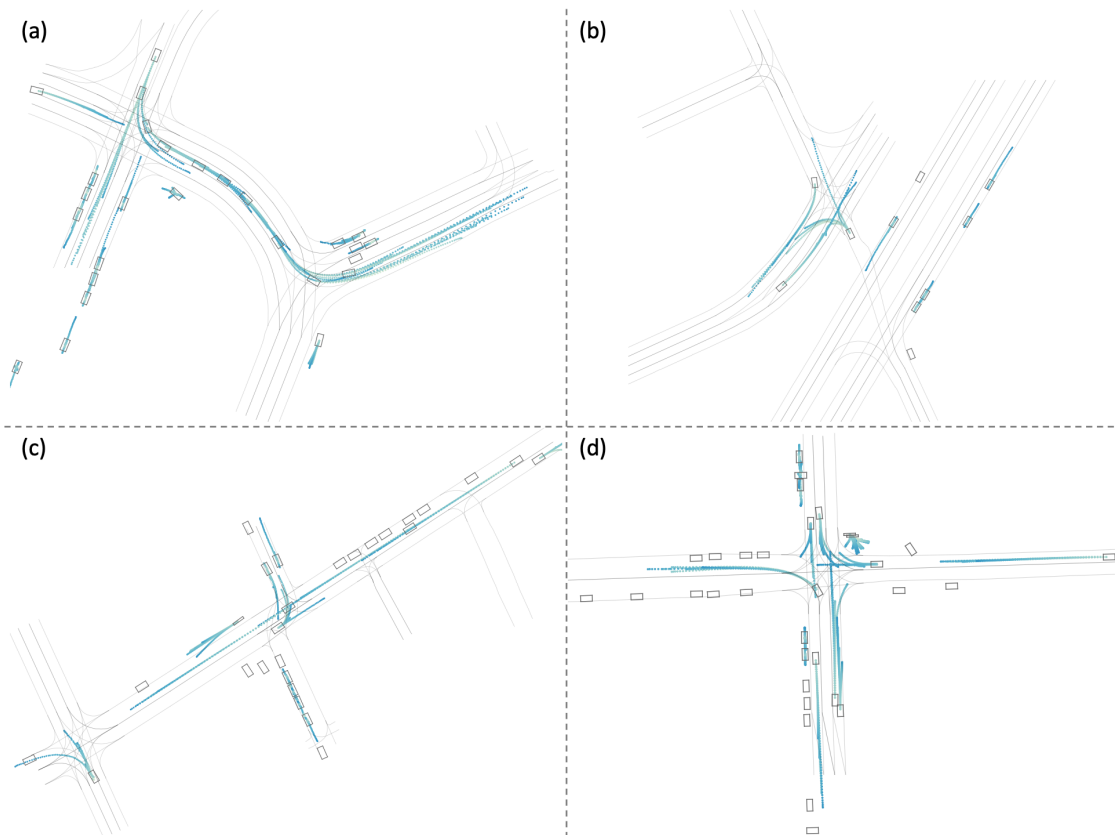


Figure 8. Illustration of the multimodal future trajectories predicted by ProphNet at various challenging scenarios with complex road topologies (a-b) and crowded intersections (c-d).

Model	minADE_1	minFDE_1
(a) Random Initialization	1.30	2.81
(b) Fusion by Attention	1.29	2.81
(c) Use of Anchor Points	1.29	2.80
(d) MLP Encoding	1.40	3.02
(e) 1D-CNN Encoding	1.31	2.83
(f) ProphNet-S	1.28	2.77

Table 10. A variety of additional ablation studies on Argoverse-1.

ing the rich goal-oriented contexts encoded in the anchor embeddings. Finally, we compare different ways of input sequence encoding in (d-f), where gMLP outperforms both MLP and 1D-CNN. As for the failure cases of our approach,

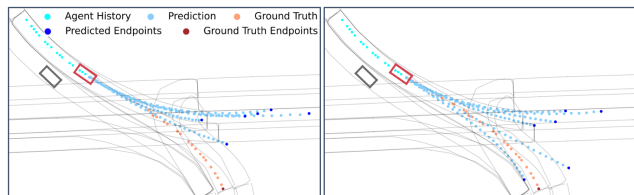


Figure 9. Illustration of the predicted trajectories by ProphNet-S with randomly (left) and orthogonally (right) initialized queries.

we find some long-horizon trajectories predicted in challenging scenes (e.g., roundabout) are occasionally off the road. One solution to mitigating this issue is to enforce an off-road penalty in training.