Supplementary Materials for *Towards Capturing the Temporal Dynamics for Trajectory Prediction: a Coarse-to-Fine Approach*

Xiaosong Jia[1,2], **Li Chen**[2], **Penghao Wu**[1,2], **Jia Zeng**[1,2], **Junchi Yan**[1,2*], **Hongyang Li**[1,2], **Yu Qiao**[2]
[*]corresponding author
[1] Shanghai Jiao Tong University [2] Shanghai AI Laboratory
{jiaxiaosong, wupenghaocraig, jia.zeng, yanjunchi}@sjtu.edu.cn
{lichen, lihongyang, qiaoyu}@pjlab.org.cn

A Experiments on Additional Models and Datasets

In order to verify the generality of the conclusions, further experiments are carried out on Argoverse 2 with HDGT. The results on the validation set is in Table 1. We also conducted experiments on LaneGCN on Argoverse 1 in Table 2.

Table 1: Ablation study of the proposed techniques on Argoverse 2 with HDGT.

Temporal Refine	Cumu. Coord. Loss	minADE↓	minFDE↓	MR↓	TRI(%)↓	UR(%)↓
✗	✗	0.8459	1.6962	0.2571	13.03	10.34
✓	✗	0.8375	1.6642	0.2482	**7.97**	**9.48**
✗	✓	0.8328	1.6620	0.2407	14.20	10.44
✓	✓	**0.8298**	**1.6477**	**0.2410**	8.55	9.68

Table 2: Ablation study of the proposed techniques on Argoverse 1 with LaneGCN.

Temporal Refine	Cumu. Coord. Loss	minADE↓	minFDE↓	MR↓	TRI(%)↓	UR(%)↓
✗	✗	0.7208	1.1087	0.1104	0.25	1.08
✓	✗	0.7125	1.0913	0.1066	**0.18**	**0.73**
✗	✓	0.7138	1.0939	0.1072	0.31	1.01
✓	✓	**0.7071**	**1.0856**	**0.1048**	0.2	0.99

One can observe that similar conclusions could be drawn with HDGT on Argoverse 2 and LaneGCN on Argoverse 1, which validates the generality of our method.

B Ablation Study of Refine Loss

To verify the necessity of scratch loss + refine loss, we conducted experiments with HDGT on MLP + 5x1D-CNN with only scratch loss. Specifically, we feed the output feature of the MLP with 1D positional embedding into 1D CNNs and only apply scratch loss on the output of the final CNN. However, we find that the model did not converge well in this case. Thus, we conduct another experiment with residual connections between 1D CNNs. The results on Waymo Motion validation set is in Table 3.

We can observe that With no residual features, the model without the refine loss is even worse than a simpe MLP, which might come from the difficulty of predicting a trajectory based on a sequence of positional embeddings and the shared hidden features. Even with the specific design for the ease of

6th Conference on Robot Learning (CoRL 2022), Auckland, New Zealand.

Table 3: Ablation Study of Refine Loss

Decoder	minADE↓	minFDE↓	MR↓	TRI(%)↓	UR(%)↓
MLP	0.6056	1.2328	0.1723	10.95	2.37
MLP + 5 * 1D CNN w Refine Loss	**0.5871**	**1.1893**	**0.1554**	**5.27**	**0.38**
MLP + 5 * 1D CNN w/o refine loss	0.6376	1.2837	0.1730	7.73	3.52
MLP + 5 * 1D CNN w/o Refine loss + Residual Features	0.6074	1.2197	0.1721	7.81	3.74

optimization, its performance is still similar to a single MLP, which suggests that the performance gain is not merely from 1D CNNs. With refine loss, refine modules could focus on fine-grained details of the trajectory while the initial MLP focuses on predicting a scratch of the future.