SRefiner: Soft-Braid Attention for Multi-Agent Trajectory Refinement

Supplementary Material

The overall structure of the supplementary material is listed as follows:

▷ Sec. 6: *Details of evaluation metrics.*
▷ Sec. 7: *Discussion with BeTop [14].*
▷ Sec. 8: *Implementation details.*
▷ Sec. 9: *Ablation study on other baseline and dataset.*

6. Evaluation metrics

We first introduce three evaluation metrics used in the Argoverse v2 dataset [29]: avgMinFDE, avgMinADE and actorMR.

avgMinFDE. The average Final Displacement Error (FDE) associated with the predicted world that has the lowest mean FDE among all (K) predicted worlds. FDE is defined as the L2 distance between the endpoint of the predicted trajectory and the ground truth. The mean FDE is the FDE averaged across all scored actors within a scenario. The index of the world with the lowest mean FDE is determined as follows:

$$idx_{FDE} = \arg\min_{k\in[1,K]} \frac{1}{N}\sum_{i=1}^{N} \|Y_{i,k}(T_+) - Y_{i,\text{gt}}(T_+)\|, \quad (23)$$

where $Y_{i,k}(T_+)$ is the endpoint of the $i-th$ predicted trajectory in the $k-th$ mode, and $Y_{i,\text{gt}}(T_+)$ is the endpoint of the $i-th$ ground truth trajectory. The avgMinFDE is then calculated as:

$$\text{avgMinFDE} = \frac{1}{N}\sum_{i=1}^{N} \|Y_{i,idx_{FDE}}(T_+) - Y_{i,\text{gt}}(T_+)\|. \quad (24)$$

avgMinADE. The mean ADE(Average Displacement Error) is associated with a predicted world that has the lowest mean ADE among all K predicted worlds. ADE is defined as the mean L2 distance between the predicted trajectory and the ground truth. The mean ADE is the ADE averaged across all scored actors within a scenario. The index of the world with the lowest mean ADE is determined as follows:

$$idx_{ADE} = \arg\min_{k\in[1,K]} \frac{1}{N}\sum_{i=1}^{N} \|Y_{i,k} - Y_{i,\text{gt}}\|, \quad (25)$$

where $Y_{i,k}$ is the $i-th$ predicted trajectory in the $k-th$ mode, and $Y_{i,\text{gt}}(T_+)$ is the $i-th$ ground truth trajectory. The avgMinADE is then calculated as:

$$\text{avgMinFDE} = \frac{1}{N}\sum_{i=1}^{N} \|Y_{i,idx_{ADE}} - Y_{i,\text{gt}}\|. \quad (26)$$

actorMR. The actor Miss Rate (actorMR) is defined as the proportion of actor predictions that are considered to have "missed"($>$ 2m FDE) in the the "best" (lowest minFDE) predicted world:

$$\text{actorMR} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_{\{\|Y_{i,idx_{FDE}} - Y_{i,\text{gt}}\|>2\}}. \quad (27)$$

For the INTERACTIONS dataset [34], the evaluation metric minJointFDE is defined identically to avgMinFDE in the Argoverse v2 dataset. The same applies to minJointADE and avgMinADE. However, the definition of minJointMR in the INTERACTIONS dataset differs from actorMR in the choice of the threshold:

$$\text{minJointMR} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_{\{\|Y_{i,idx_{FDE}} - Y_{i,\text{gt}}\|>\tau\}}, \quad (28)$$

where τ denotes the miss rate threshold, defined as follows:

$$\tau = \begin{cases} 1, & v \le 1.4m/s \\ 1 + \frac{v-1.4}{11-1.4}, & 1.4m/s < v \le 11m/s \\ 2, & otherwise, \end{cases} \quad (29)$$

where v is the ground-truth velocity at the final timestep.

7. Discussion with BeTop

BeTop [14] focuses on the integration of trajectory prediction and planning (IPP) in autonomous driving. It predicts the intersection relationships between future trajectories and uses ground truth (GT) to supervise these relationships. It is a direct application of braid topology. Our SRefiner focuses on multi-agent trajectory refinement. We propose a soft-braid topology to capture the spatio-temporal topological relationships among all predicted future trajectories. Additionally, SRefiner uses the soft-braid topology to model the spatio-temporal topological relationships between future trajectories and lanes, whereas BeTop only models the intersections between trajectories. As demonstrated in Table 4, for the task of trajectory refinement, the performance of our proposed soft-braid attention surpasses that of the braid attention used in BeTop.

8. Inplementation details

For both Argoverse v2 [29] and INTERACTIONS [34] dataset, we train the model for 64 epochs with a batch size

Table 6. Ablation study of the effect of each components of SRefiner with FJMP [19] on Argoverse v2 [29].

Method	Traj-Traj Soft-Braid	Traj-Lane Soft-Braid	Topology Update	avgMinFDE ↓	avgMinADE ↓	actorMR ↓
M1				1.920	0.819	0.235
M2		✓	✓	1.780	0.772	0.223
M3	✓		✓	1.756	0.759	0.222
M4	✓	✓		1.787	0.765	0.226
M5	✓	✓	✓	**1.736**	**0.747**	**0.221**

Table 7. Ablation study of the effect of each components of SRefiner with FJMP [19] on INTERACTIONS [34].

Method	Traj-Traj Soft-Braid	Traj-Lane Soft-Braid	Topology Update	avgMinFDE ↓	avgMinADE ↓	actorMR ↓
M1				0.630	0.190	0.122
M2		✓	✓	0.591	0.178	0.112
M3	✓		✓	0.585	0.174	**0.110**
M4	✓	✓		0.593	0.175	0.114
M5	✓	✓	✓	**0.579**	**0.170**	**0.110**

Table 8. Ablation study on the effect of soft-braid topology with FJMP [19] on Argoverse v2 [29]. † indicates only model the topology between trajectories.

Method	avgMinFDE ↓
Baseline	1.920
No Topology	1.794
Braid Topology (BeTop [14])	1.770
Soft-Braid Topology (Ours)†	1.756
Soft-Braid Topology (Ours)	**1.736**

Table 9. Ablation study on the effect of soft-braid topology with FJMP [19] on INTERACTIONS [34]. † indicates only model the topology between trajectories.

Method	minJointFDE ↓
Baseline	0.630
No Topology	0.601
Braid Topology (BeTop [14])	0.594
Soft-Braid Topology (Ours)†	0.585
Soft-Braid Topology (Ours)	**0.579**

Table 10. Ablation study on the Trajectory-Trajectory Soft-Braid Attention local radius and Trajectory-Lane Soft-Braid Attention local radius with FJMP [19] on Argoverse v2 [29].

T-T local radius	avgMinFDE ↓	T-L local radius	avgMinFDE ↓
10	1.752	2	1.746
30	1.739	5	1.738
50	**1.736**	10	**1.736**
100	**1.736**	20	1.738

Table 11. Ablation study on the Trajectory-Trajectory Soft-Braid Attention local radius and Trajectory-Lane Soft-Braid Attention local radius with FJMP [19] on INTERACTIONS [34].

T-T local radius	minJointFDE ↓	T-L local radius	minJointFDE ↓
10	0.585	2	0.584
30	0.582	5	0.580
50	**0.579**	10	**0.579**
100	**0.579**	20	0.580

of 16 on a single RTX 3090 GPU. We utilize the AdamW optimizer with a cosine learning rate schedule and a weight decay of 0.0001. For Argoverse v2, the initial learning rate is set to 1×10^{-4}, while for INTERACTIONS, it is set to 3×10^{-4}. The embedding dimension of the model is 64.

9. More ablation study

In the main paper, we report ablation studies using Forecast-MAE [6] on Argoverse v2 dataset [29] due to the page limit. Here, we present ablation studies using FJMP [19] on both two datasets. Tables 6 and 7 illustrate the ablation study of the impact of each component of SRefiner with FJMP, demonstrating that each component significantly contributes to the overall performance. Tables 8 and 9 show the ablation study on the effect of the soft-braid topology with FJMP on the two datasets. The results indicate that using the soft-braid topology to guide the trajectory refinement yields significant performance improvements and outperform the direct use of the braid topology. Tables 10 and 11 present the ablation study on the choice of trajectory-trajectory soft-braid attention local radius τ_a and trajectory-lane soft-braid attention local radius τ_l, which are set to 50m and 10m by default.