TraF-Align: Trajectory-aware Feature Alignment for Asynchronous Multi-agent Perception

Supplementary Material

In this supplementary material, we present additional experimental results, including PR curves from ablation studies, as well as visualizations of the trajectory field and attention positions.

1. Ablation studies

1.1. Major components

Four combinations of the main components are defined in Table 1, and their PR-curves are shown in Figure 1. Under synchronous conditions, the differences among the ablations are not significant. However, under a 400 ms delay, TraF-Align demonstrates a significant advantage across both entire recall ranges. This highlights the positive impact of the modules in improving robustness against delays.

1.2. Loss

The PR-curves for different loss combinations are shown in Figure 2. As in Figure 1, only the curves under synchronous and 400 ms delay conditions are presented for clarity. The results are consistent with those in Table 3 of the main text.

1.3. Historical frames

Figure 3 presents the Precision-Recall curves for TraF-Align evaluated on the V2V4Real validation dataset under varying frame settings. For a detailed analysis, please refer to Section 4.3 in the main text.

2. Trajectory field

Network. The network for the field predictor (Figure 6) is adapted from UNet. It first reduces the dimensionality of the input feature map, then progressively upsamples it while concatenating the input feature map at each stage. During this process, as the feature map is downsampled, deeper semantic information is progressively extracted. In the upsampling phase, these features are decoded, allowing for a more comprehensive semantic understanding of the feature map. In this paper, the field predictor is used to interpret the feature map output from the backbone and generate the trajectory field, which aligns perfectly with this process.

Table 1. Definition of ablations corresponding to Figure 1.

Item	Field predictor	Offset generator	Attention layers
Ablation 1			
Ablation 2			✓
Ablation 3		✓	✓
TraF-Align	✓	✓	✓

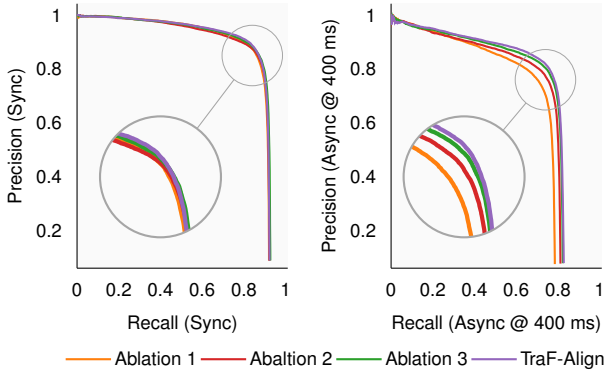

Figure 1. Precision-Recall curve showing the ablation results of major components when the Ego vehicle experiences 0 ms and 400 ms delays on the DAIR-V2X-Seq dataset.

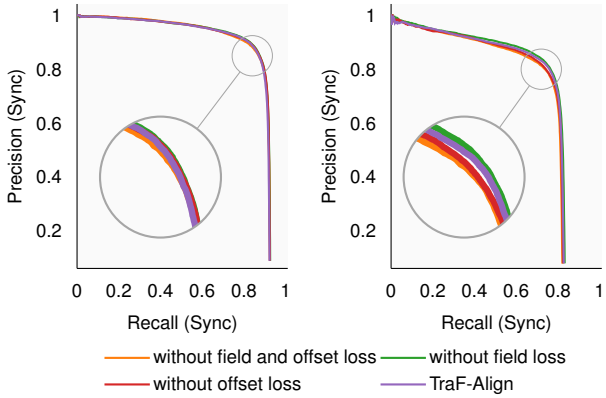

Figure 2. Precision-Recall curve showing the ablation results of loss when the Ego vehicle experiences 0 ms and 400 ms delays on the DAIR-V2X-Seq dataset.

Field visualization. Trajectory field ($\mathbb{R}^{3\times H\times W}$) consists of a position field ($\mathbb{R}^{1\times H\times W}$) and an orientation field ($\mathbb{R}^{2\times H\times W}$). The position field creates heatmap peaks along the object's trajectory, while the orientation field captures the inverse tangent direction of the trajectory. Figure 5 shows several examples of the learned trajectory field.

3. Attention position

Figure 4 visualizes attention positions learned by TraF-Align, and Figure 7 shows Sinkhorn's cost and matching probability matrices.

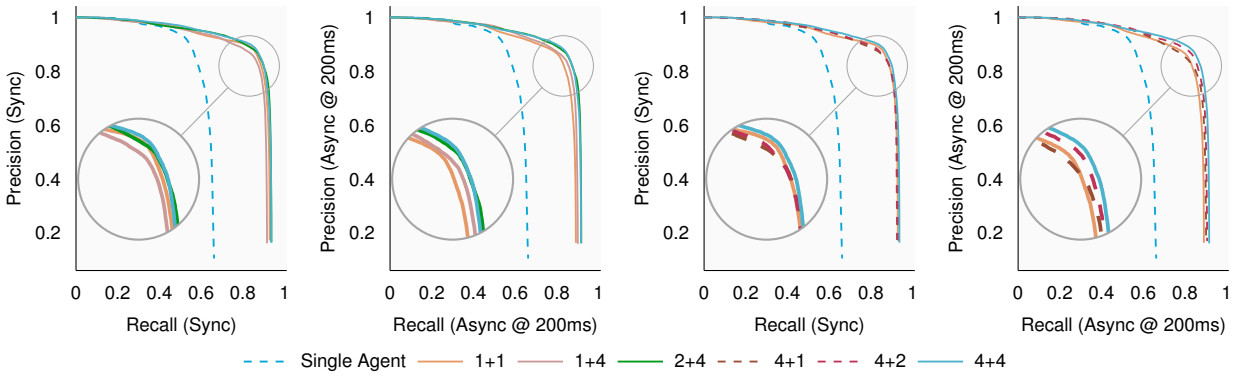

Figure 3. Precision-Recall curves for TraF-Align on the V2V4Real validation dataset under different frame settings. The legend notation $i + j$ represents the use of i frame(s) for the ego vehicle and j frame(s) for the cooperative vehicle.

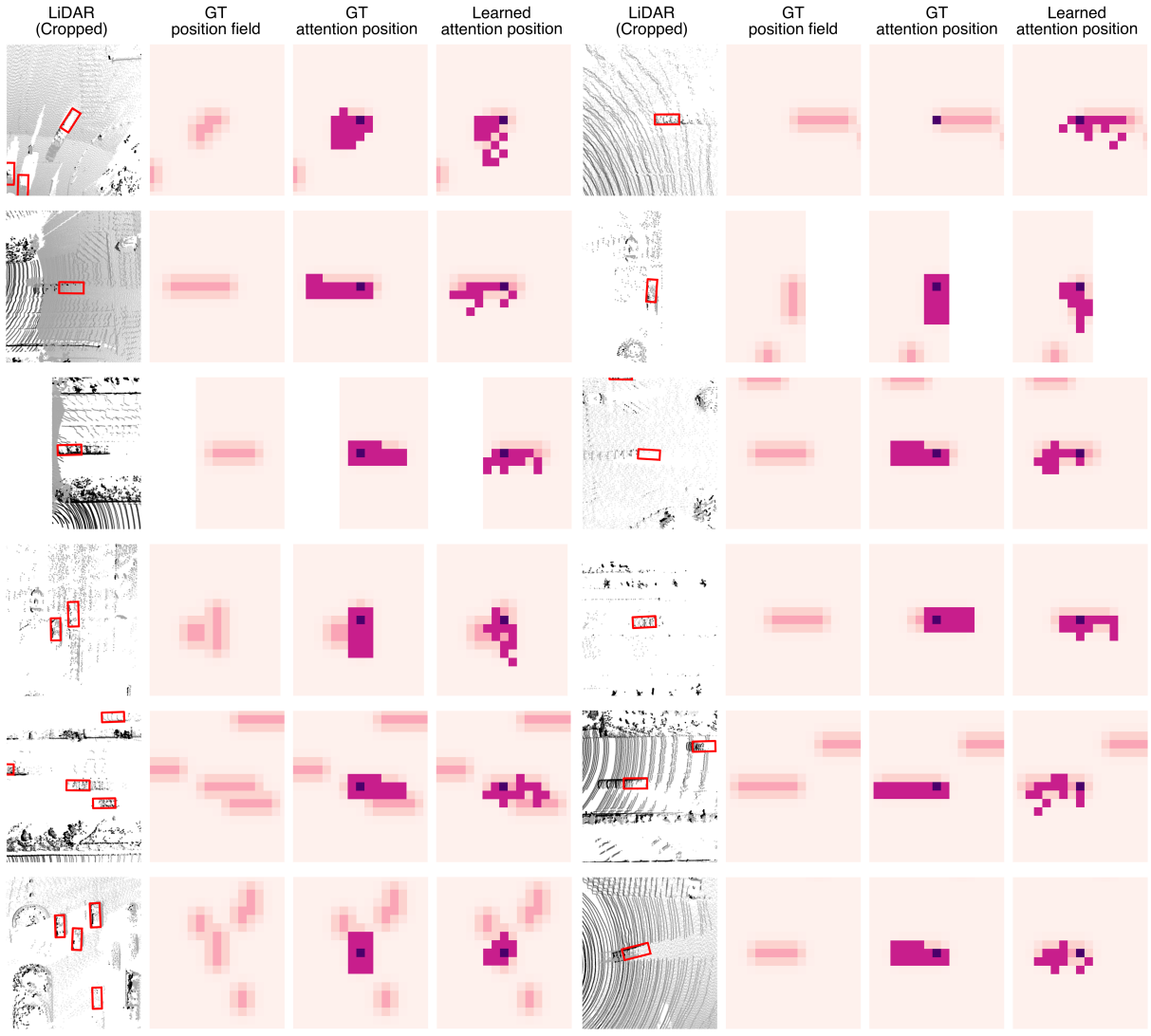

Figure 4. Examples of learned attention positions.

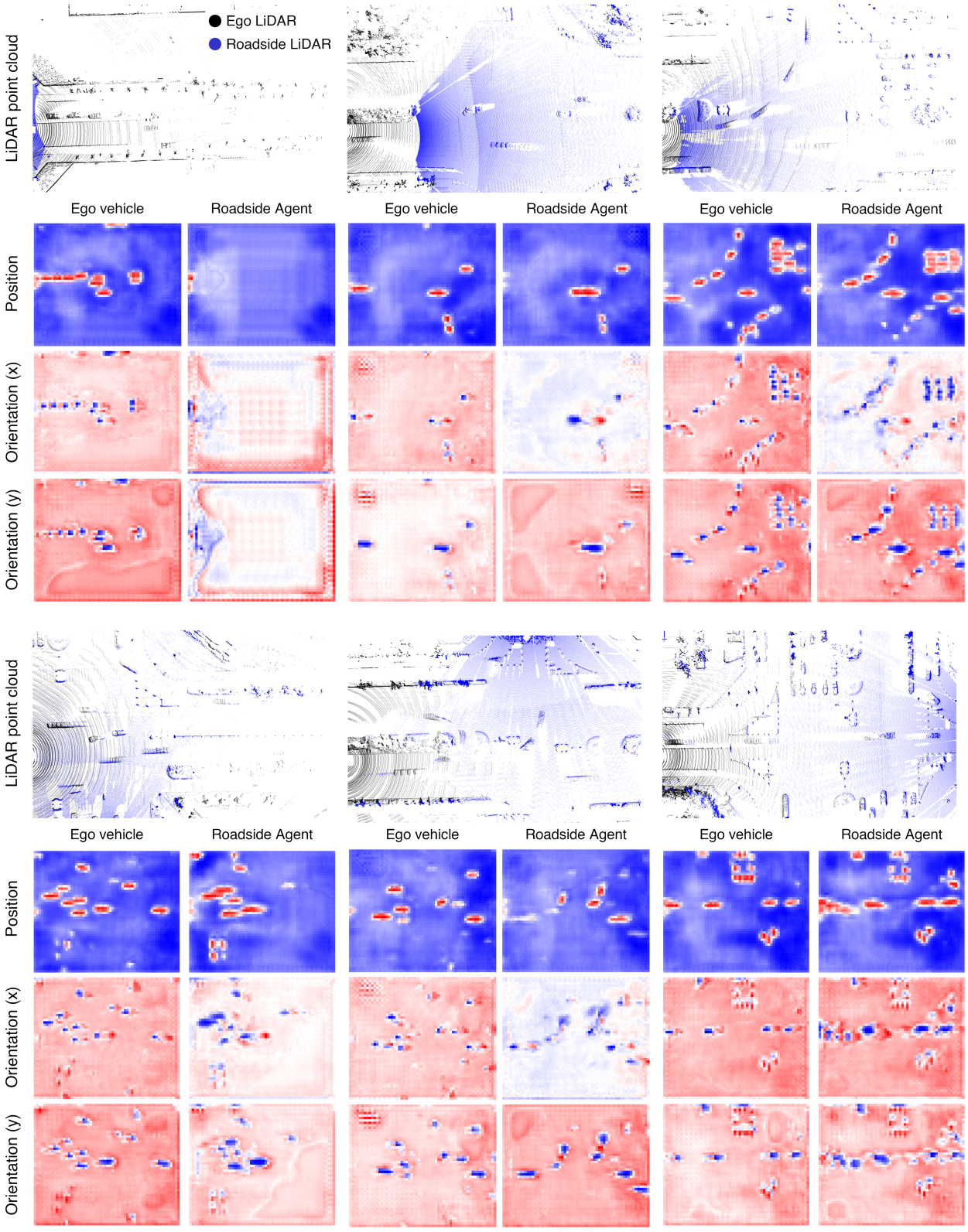

Figure 5. Examples of learned trajectory field by TraF-Align on DAIR-V2X-Seq dataset [43]. The position field creates heatmap peaks along the object's trajectory, while the orientation field captures the inverse tangent direction of the trajectory.

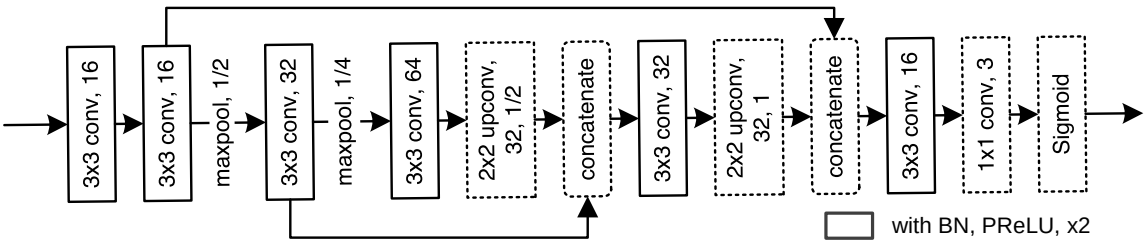

Figure 6. Field predictor.

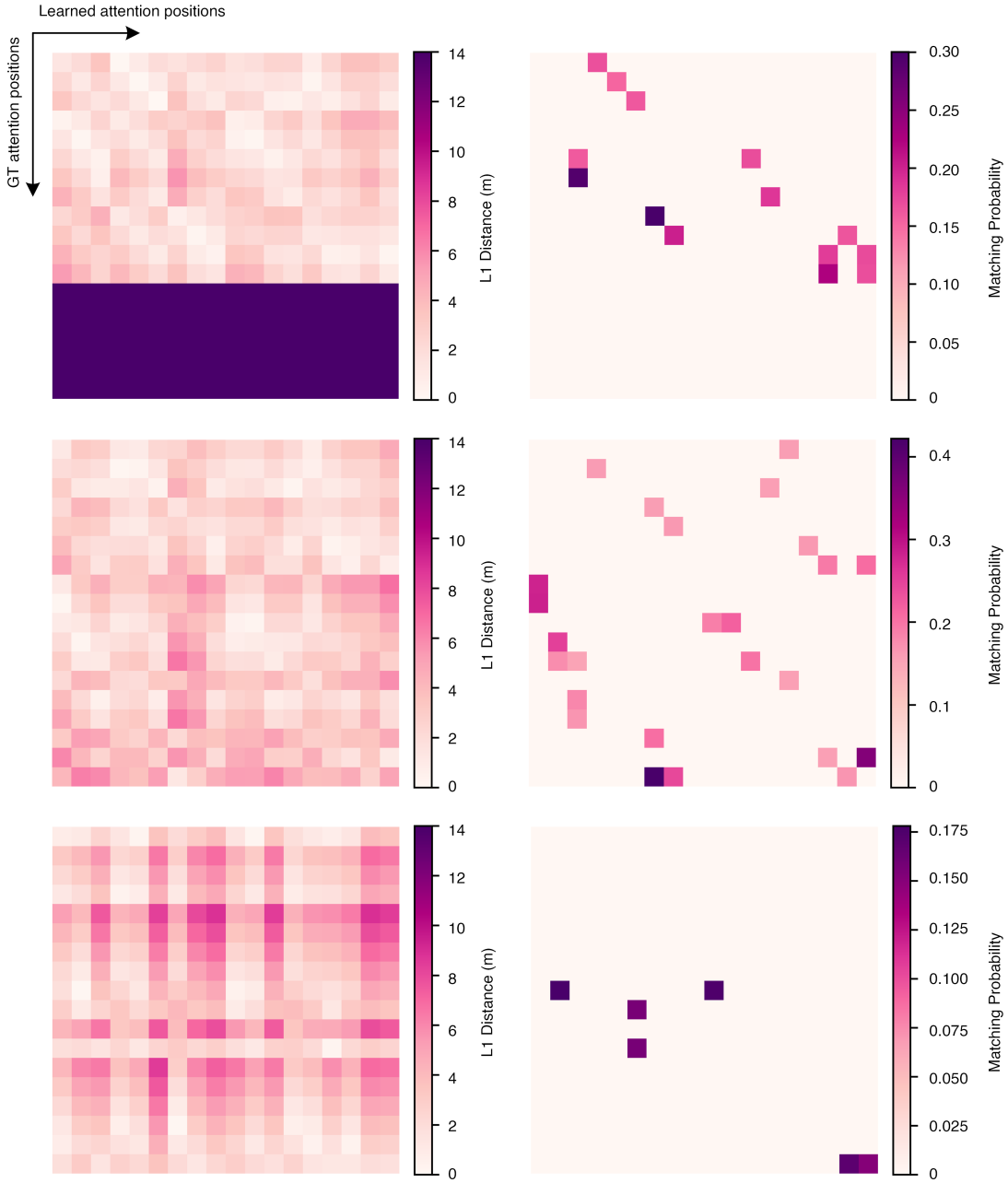

Figure 7. Examples of the cost and assignment probability matrices.