Collaborative Motion Prediction via Neural Motion Message Passing

Yue Hu[1], Siheng Chen[2 ✉], Ya Zhang[1 ✉], and Xiao Gu[1]

[1] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University
[2] Mitsubishi Electric Research Laboratories
{18671129361, ya_zhang, gugu97} @sjtu.edu.cn, schen@merl.com

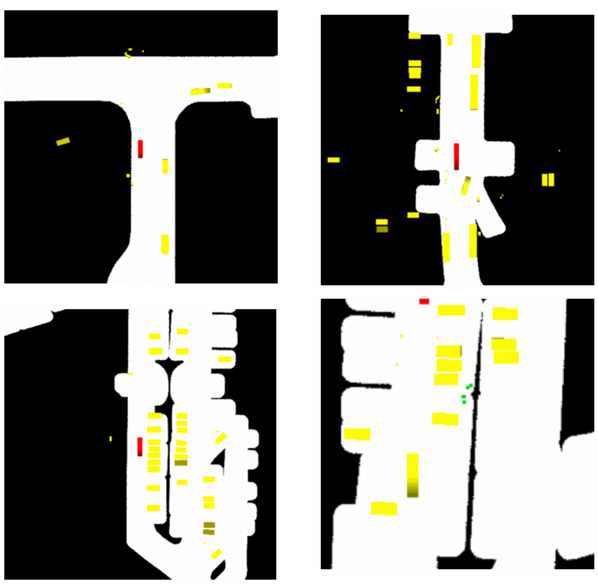

Figure 1. Rasterized scene images. The red box is the SDV, the yellow boxes are the vehicles, and the green dots are the pedestrians. As time goes back, the brightness of the boxes darkens, resulting in the fading tails.

Rasterization. Given the global coordinates and the headings of each traffic actor involved in the scene and the map marked the drivable areas. We encode each actor type into the corresponding vector layer. White represents drivable areas, and vice versa. Vehicles are in yellow polygons while the self-driving vehicle (SDV) is in red and pedestrians in green. Other actors are colored black. Besides, the actors' histories are represented with reduced level of brightness in the same color. The vector layers of the same actor are rasterized one by one in time order, resulting in the fading effect. Then the vector layers are rasterized one by one on top of each other, in the order from drivable areas to traffic actors such as vehicles and pedestrians, resulting in a RGB scene image.

Visualization of interaction embeddings in JMP-NMMP. Figure. 2 shows the visualization of interactions in the embedding domain and the corresponding actor pairs in the spatial domain in the joint motion prediction scenarios. Similar to interaction embedding visualization in Figure 1, the interaction embeddings $\mathbf{e}_{ij}$ are mapped to 2D coordinates via t-SNE and shown in the first column. We randomly pick three pairs of close interaction samples in the embedding space, which are colored red and blue, and plot the corresponding trajectory pairs in the followed two columns with dots. We get the similar conclusions as the interaction embedding analysis in pedestrian motion prediction setting. We see that (i) close interaction embeddings in the embedding domain lead to similar trajectory relations between associated actors in the spatial domain. For example, in the first row, pedestrians are following the front one

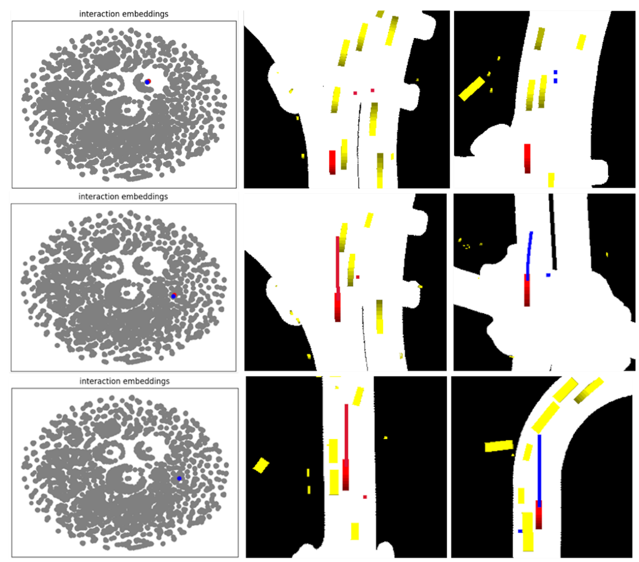

Figure 2. The first column is the t-SNE visualization of interaction embeddings, where each dot represents an interaction and correspond to spatial-domain plots. The close red and blue dots in the first column represent close interaction embeddings in the embedding domain. The right two columns are the corresponding traffic scenes. And the middle column is the corresponding spatial plot of the red dot. The trajectories of the interacted traffic actors are colored in red. Similarly the right column presents spatial plot of the blue dot.

and moving slowly to pass the road; in the second and third rows, pedestrians are obstructed by the moving vehicles on the road; (ii) different interaction types between actors' trajectory in the spatial domain are encoded to different positions in the embedding domain. The bottom two rows are close; both are far from the top row in embedding space. The trajectories show the interaction types of the top two rows are similar and quite different from the last row.