LTP: Lane-based Trajectory Prediction for Autonomous Driving

Supplementary Material

A. Detailed Architecture

We use a three-layer GNN as shape encoder, and the width of each layer is 32, 64, and 128, as shown in Fig.7. Multi-head attention mechanism is used in the global interaction encoder as shown in Fig.8 and the number of heads is set to 8. We use a four-layer attention network as the global interaction encoder, and a residual module is added at the end of each layer. In order to generate multi-modal trajectories for the agents we are concerned about, we combine the embedding of all A agents that require to be predicted and the embedding of M lane segments together and output multi-mode trajectories with probability weights, as shown in Fig.9.

B. Implementation Details

For each scene, we use vehicles and lanes of 80 meters near the current location of the target agent as input. Since for the Argoverse dataset, only one agent is needed to predict, we set the origin of the coordinate system to the position of the target agent at $t = 0$, and use random rotation for data augmentation during training. While for our inner dataset, the origin of the coordinate system is the current position of the data acquisition vehicle, so other agents in the same scene have equal status. The learning rate is 0.001 at the beginning and multiplied by 0.95 for each epoch until the learning rate reaches 3e-5. We train the model on 8 V100 GPUs for 70 epochs using Adam optimizer with a batch size of 32. The D for the ground truth threshold is set to 2.5. The γ_1, γ_2 and k for variance-based NMS is set to 3.5, 1.0 and 15.0.

C. Detail of further application study

C.1. Vanilla Closed-loop Inference.

In the closed-loop tests, we randomly specify the initial 2s trajectory and use the last two seconds of the trajectory generated by the model as the input for the next inference. We always choose the trajectory with the highest probability, and find that the model prefers to go straight and turn right, which makes it easy to form a circular path in the global map. It is speculated that this is because the ratio of straight going and right turning is higher in the training dataset. If instead of choosing the trajectory with the greatest probability but randomly sampling the trajectory, a variety of paths will be generated, which is more suitable for simulation systems with diverse needs as shown in Fig.10.

C.2. Cost Volume Optimization.

Let $\tau = (\tau^0, \tau^1, ..., \tau^{T-1})$ be a trajectory spanning over T timesteps into the future, where τ^t is the path point at time t. With our LTP, we can predict multi trajectories for all target vehicles parallel, including our ego-vehicle

$$\Gamma_{ego} = \{\tau_{e_1}, \tau_{e_2}, ..., \tau_{e_n}\} \tag{12}$$

and all other dynamic agents

$$\begin{aligned}\Gamma_{oth} = \{&\{\tau_{o_1^1}, \tau_{o_1^2}, ..., \tau_{o_1^n}\},\\ &\{\tau_{o_2^1}, \tau_{o_2^2}, ..., \tau_{o_2^n}\},\\ &\{\tau_{o_k^1}, \tau_{o_k^2}, ..., \tau_{o_k^n}\}\}.\end{aligned} \tag{13}$$

Our objective is to find the best trajectory τ^* that satisfies:

$$\tau^* = \underset{\tau \in \Gamma_{ego}}{\arg\max} \sum_t c^t(\tau^t, \Gamma_{oth}), \tag{14}$$

in where

$$c^t(\tau^t, \Gamma_{oth}) = \sum_{i=0}^{k} \sum_{j=0}^{n} P_{i,j} \mathbf{I}(\tau^t, Occ(\tau^t_{o_i^j})). \tag{15}$$

$P_{i,j}$ is the probability corresponding to the j-th trajectory of agent i. $\mathbf{I}$ is a discriminant function to determine whether the path point τ^t is within the occupied range of $Occ(\tau^t_{o_i^j})$, which represents the occupied space of $\tau^t_{o_i^j}$ and can be quickly generated by convolution. The qualitative results is shown in Fig.11. It can be seen that with cost volume, LTP can safely cope with challenging scenarios such as congestion, cut in, and pedestrian crossing.

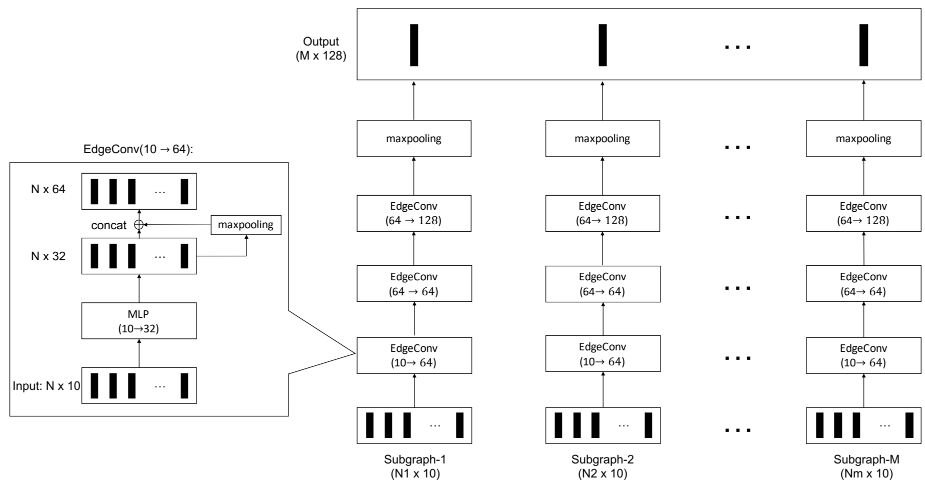

Figure 7. Shape encoder.

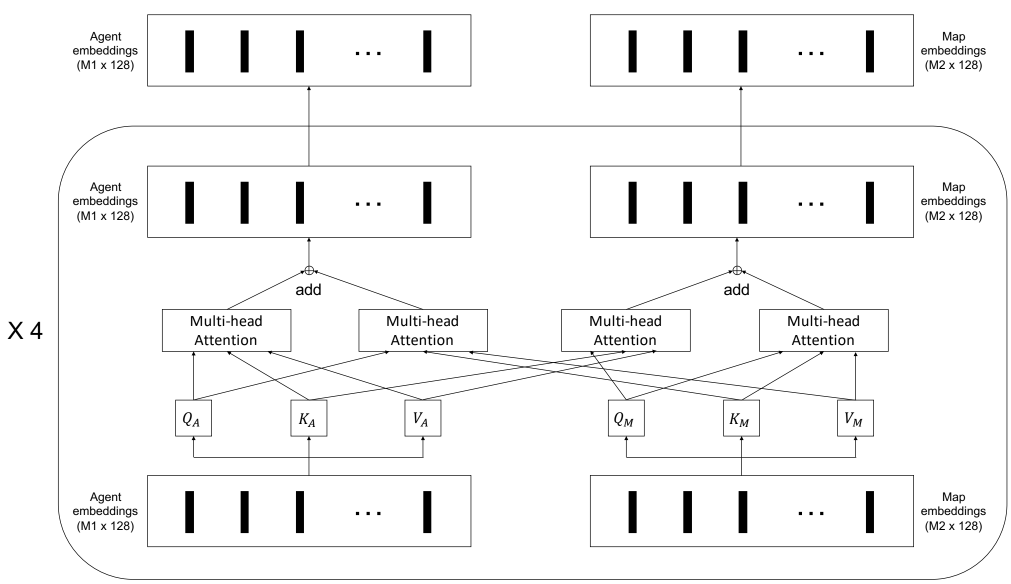

Figure 8. Interaction encoder.

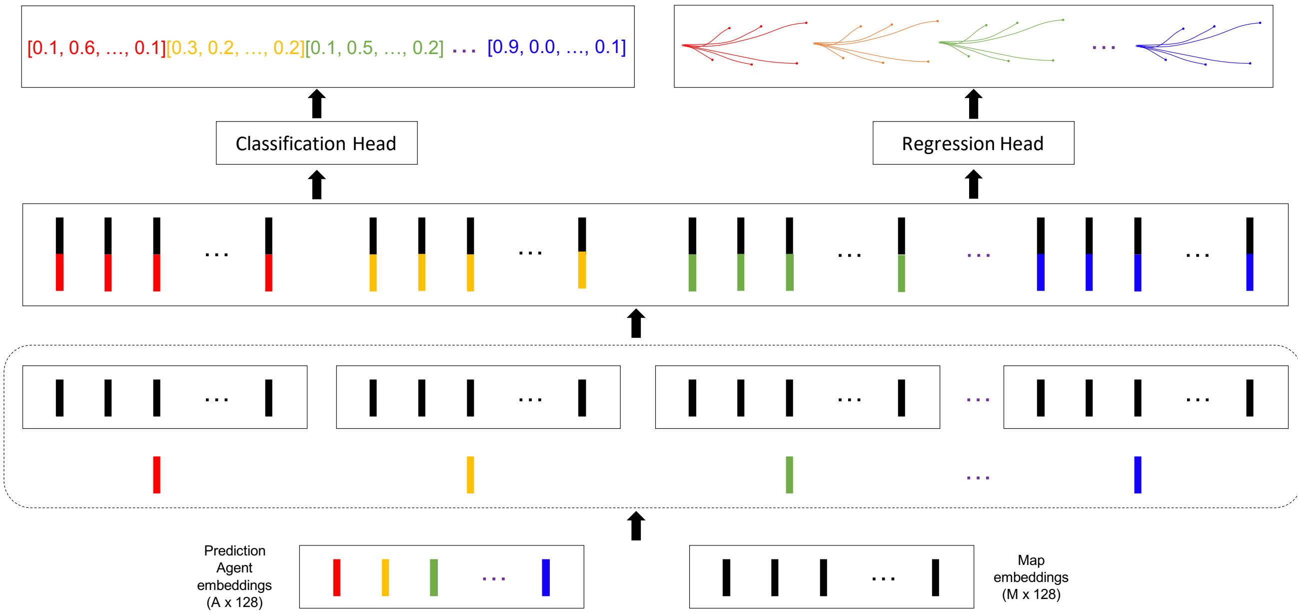

Figure 9. output layer.

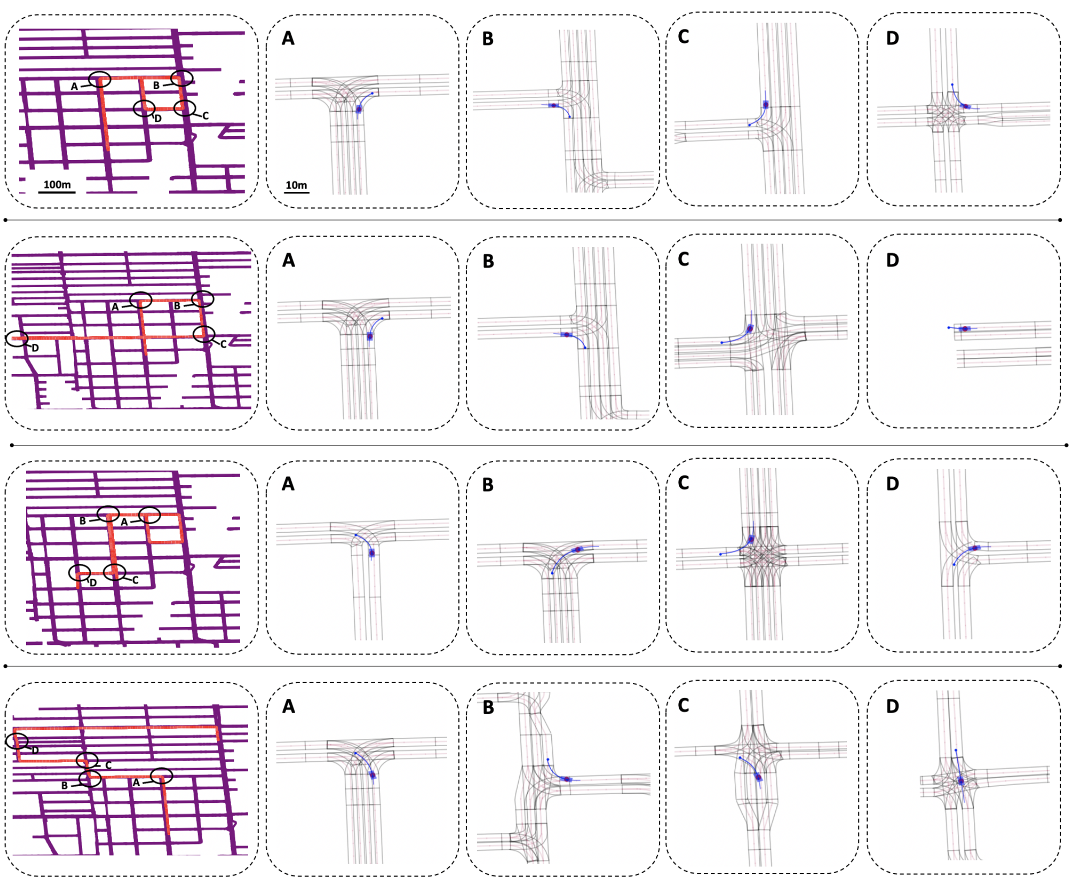

Figure 10. Closed-loop inference can generate diverse paths for simulation.

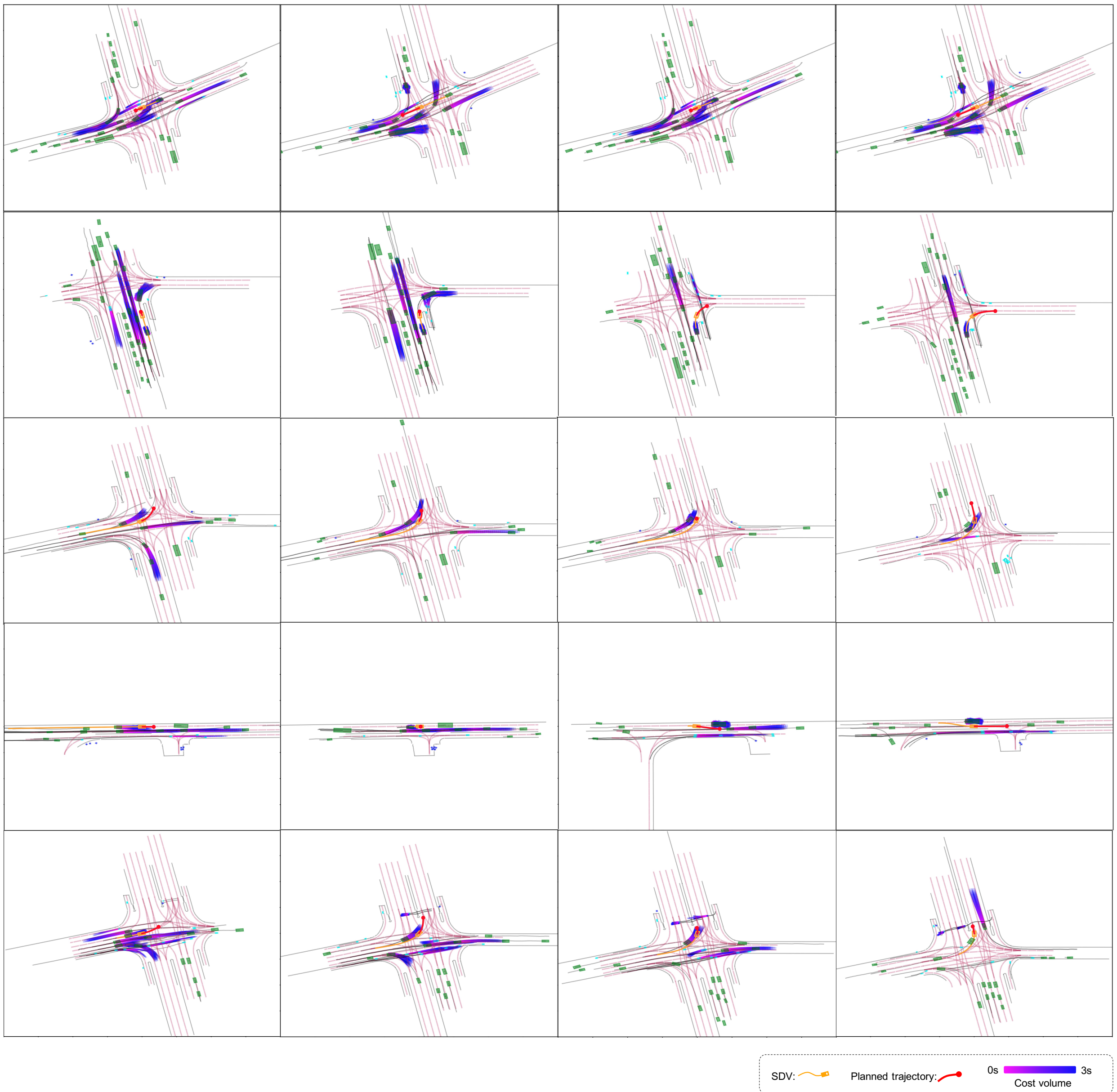

Figure 11. Cost volume result.