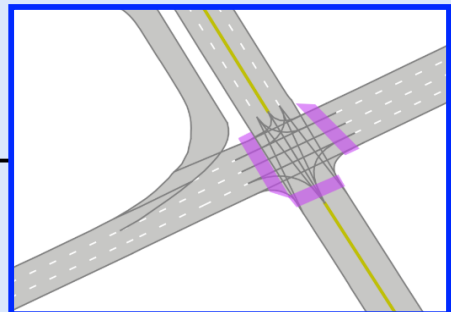
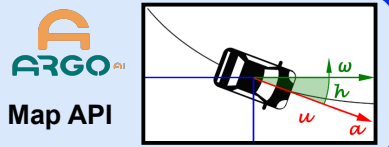


Input

HD Map



Physics-based Heuristics



Agents Past Trajectories



T_0
 T_{-1}
 T_{-2}
 T_{-3}
...
 T_{-n}

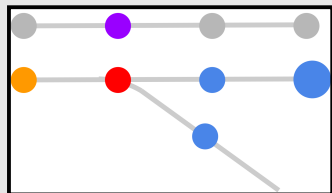
$\Delta v +$
Metadata

Heuristic Lane Proposals

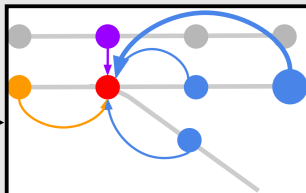
Concat

Map Encoder

Build Graph



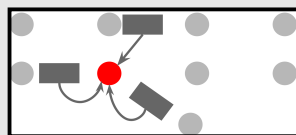
Graph encoding



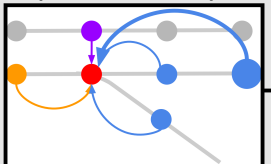
Latent graph

Fusion Cycle

Cross-Attention (Agent to Lane)

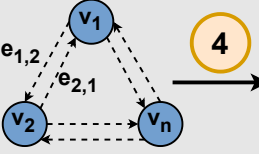


Graph encoding (Lane to Lane)



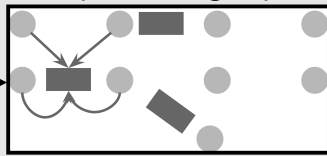
2

Crystal-GCN (Agent to Agent)



3

Cross-Attention (Lane to Agent)



4

Social Encoder

Positional Encoding

Linear Encoder (FC)

Latent actors

Transformer Encoder

$\times L_h$

MHSA

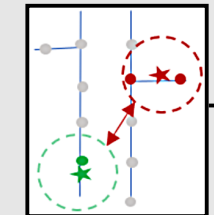
Add & Norm

Feed Forward

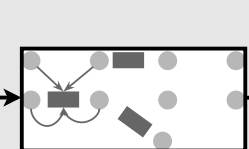
Add & Norm

Goal Areas estimation

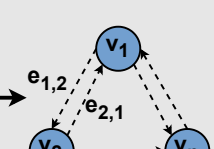
Estimate goals and context



Add deep context to agents



Agent to Agent update



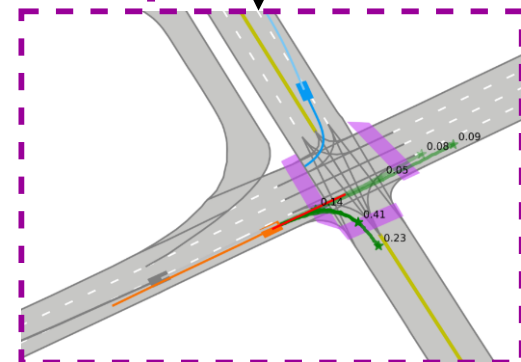
Latent actors with deep area context

Multimodal Decoder (Regression & Confidences)

Preliminary multimodal predictions

Motion Refinement

Output



Refined Multimodal trajectories