Generative Map Priors for Collaborative BEV Semantic Segmentation

Jiahui Fu[1] , Yue Gong[1] , Luting Wang[1], Shifeng Zhang[2] , Xu Zhou[2] , Si Liu[1]

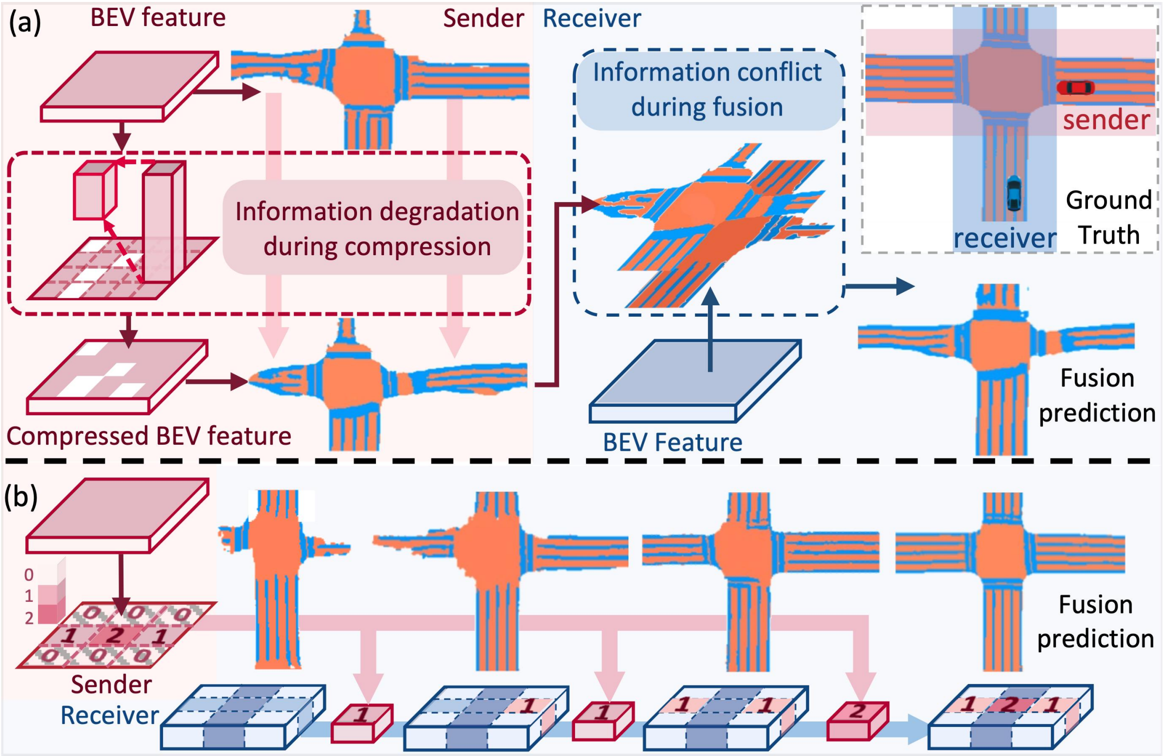
[1]Institute of Artificial Intelligence, Beihang University [2]Sangfor Technologies Inc.

Motivation

Collaborative perception aims for optimal **global perception** among multiple agents with **minimal communication**. To acchieve this target, need to

- ➢ **Information Compression at Sender Side:** Efficiently compress BEV features under limited bandwidth without losing critical semantic information.
- ➢ **Information Aggregation at Receiver Side:** Robustly fuse BEV features from multiple agents while resolving conflicts due to pose and prediction discrepancies.

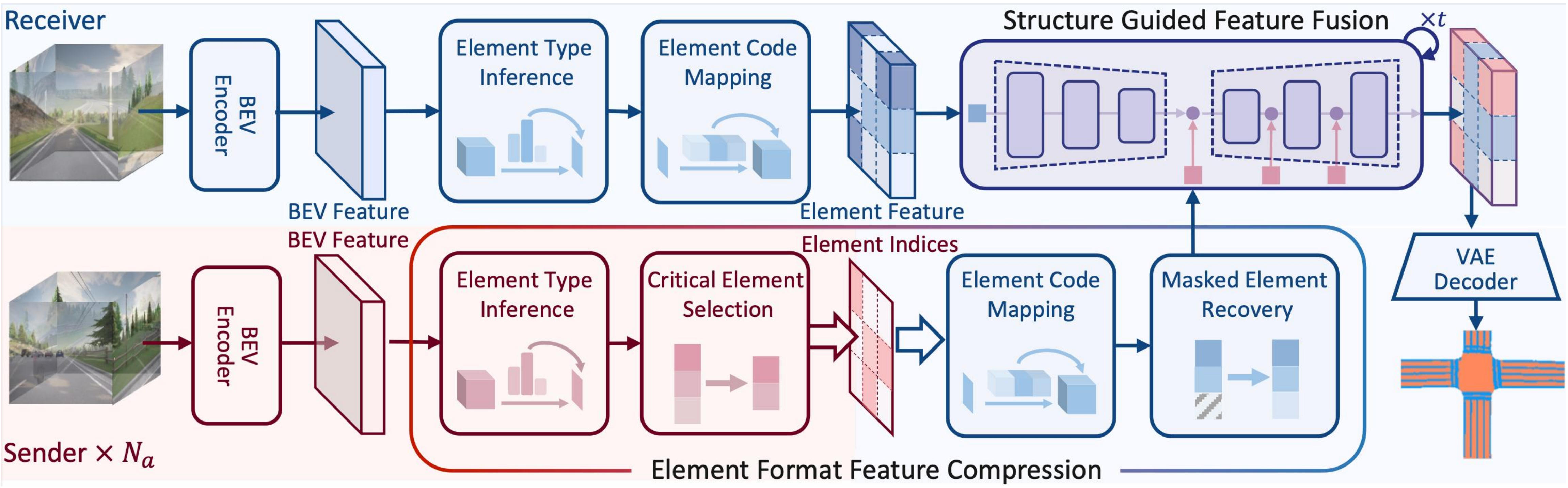
Motivation



- Previous Limitations:
  - **Spatial Mask and Dimensionality Reduction:** Aggressive compression methods cause significant loss of essential spatial context.
  - **Convolution and Attention-based Fusion:** Existing methods overly depend on precise spatial alignment, leading to inaccuracies in dense segmentation scenarios.

We propose a collaborative framework leveraging **generative map priors** such as road layouts and lane continuity to preserve essential information during compression and resolve semantic conflicts

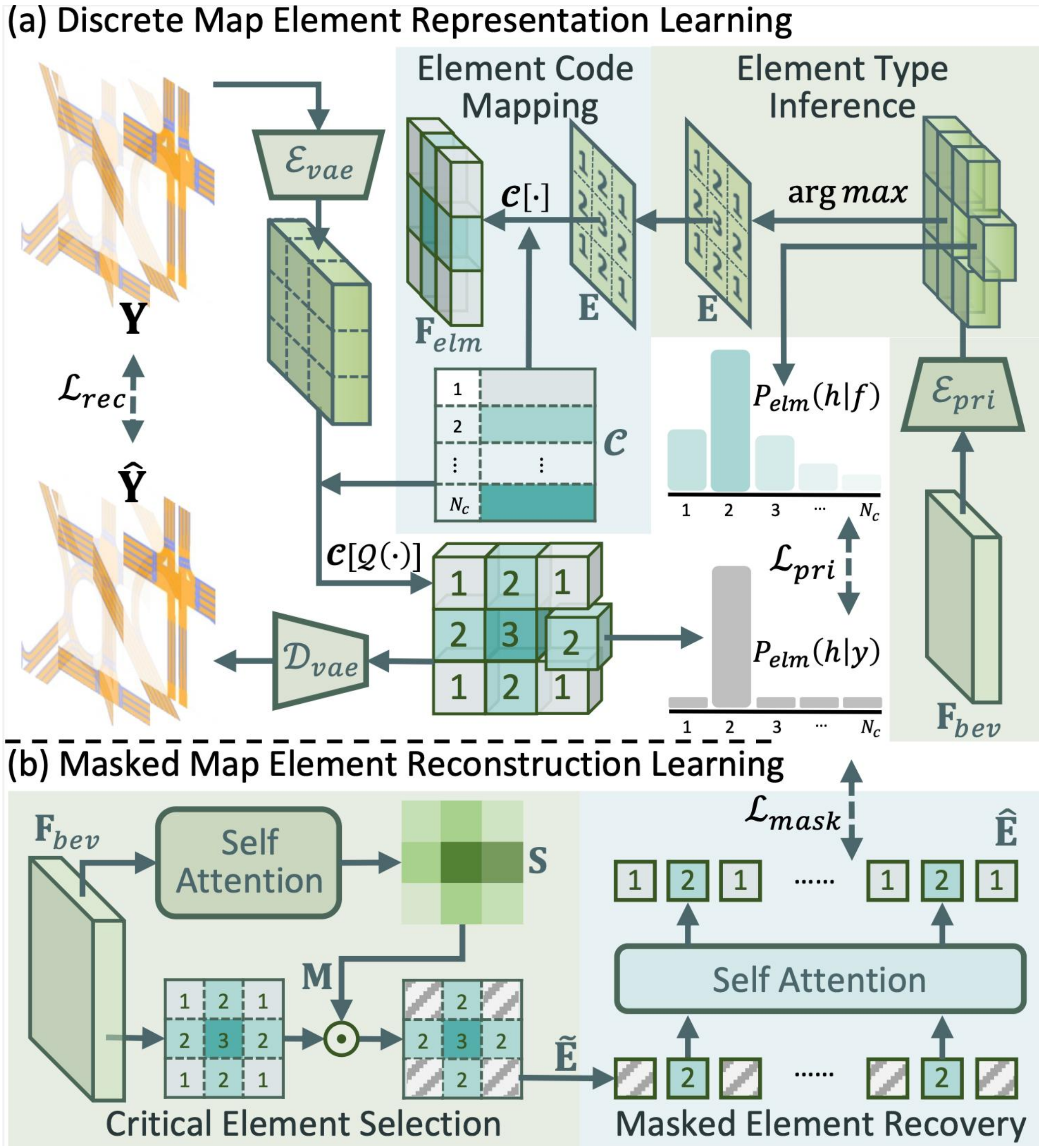
Method Framework



- **In Sender Side:** we apply **Element Format Feature Compression (EFFC)** to use discrete selected key element indices as transmission units, which are obtained from generative VAE training.
- **In Receiver Side:** we integrate the elements from other agents into the ego prediction through **Structure Guided Feature Fusion (SGFF),** which is modeled as an inpainting process based on DDPM.

Method

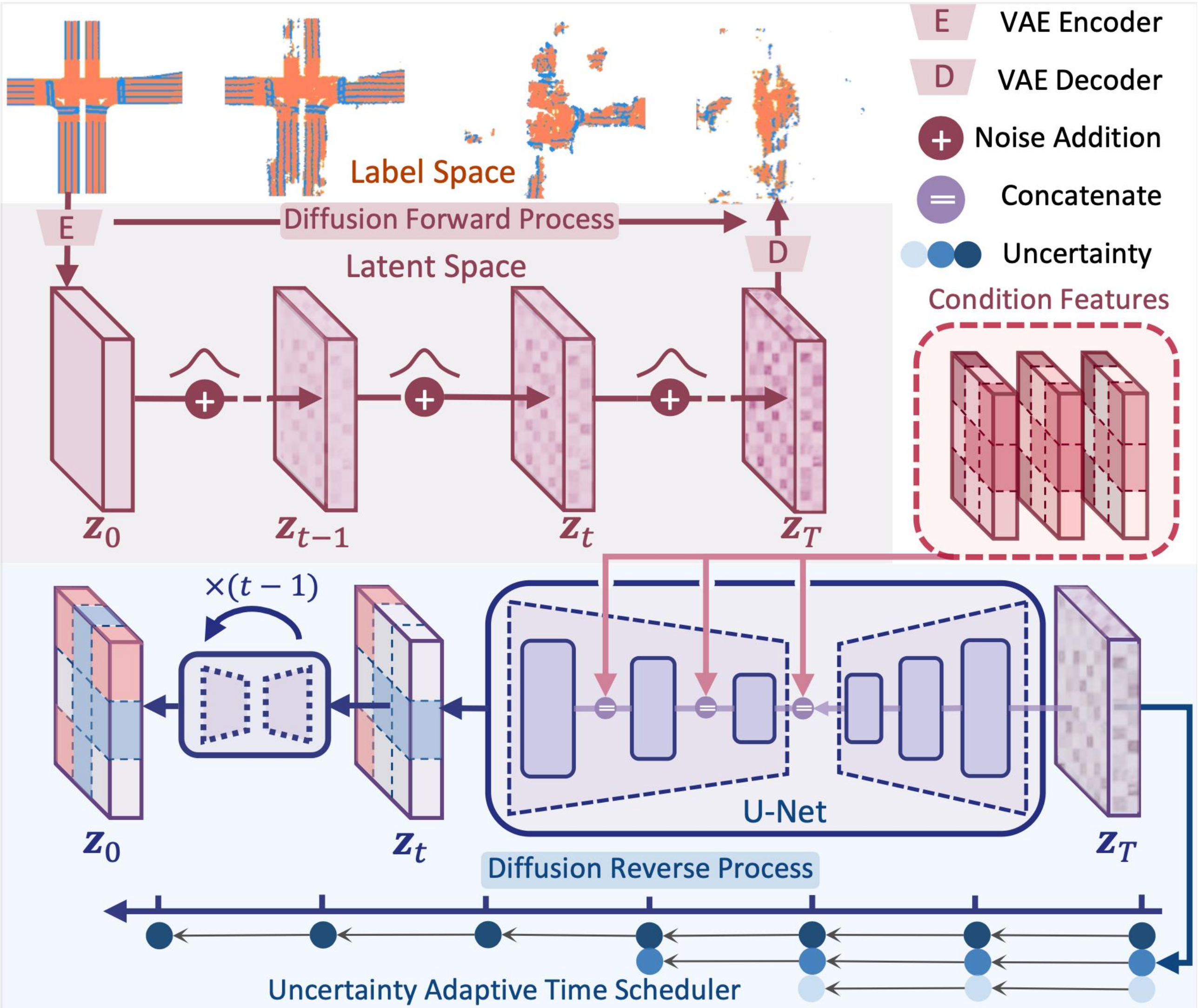
Element Format Feature Compression



- **Discrete map element representation learning:** We first train a VQ-VAE model on the map segmentation data to obtain a **map element codebook**, then train a **convertor to infer the map element type** from the BEV features obtained from the sensor data.
- **Masked map element reconstruction learning:** We optimize a **learnable mask** by the reconstruction training to select the most **critical elements** for transmission, and recover other elements on the receiver side based on the transmitted information.

Method

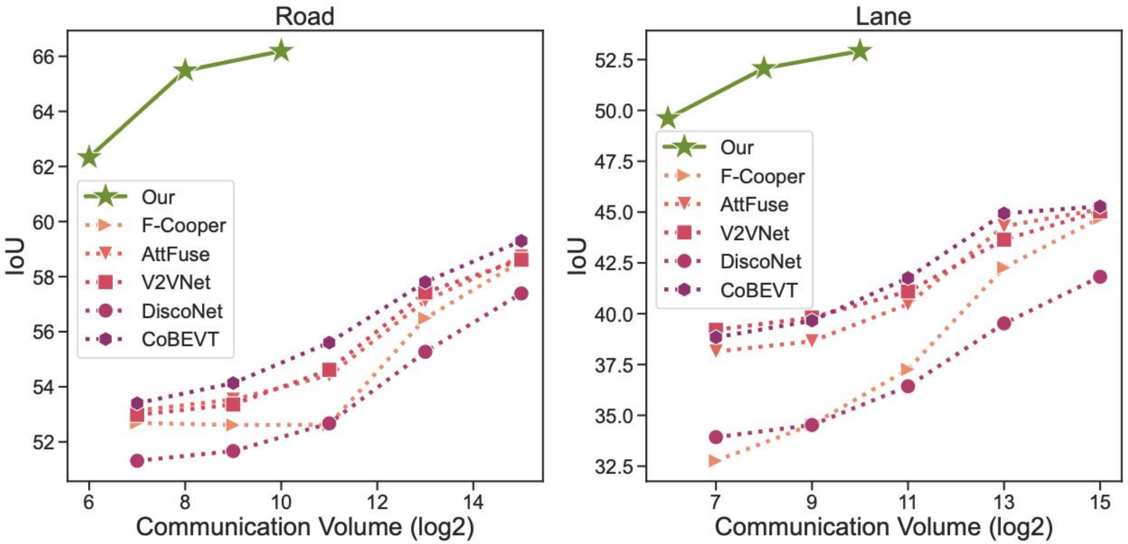
Structure Guided Feature Fusion



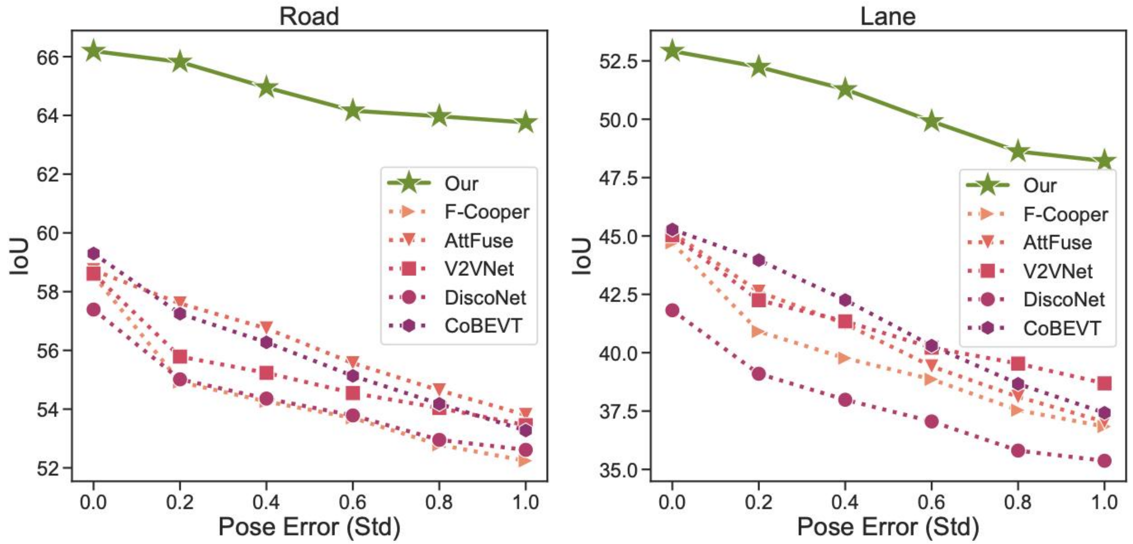
- **Progressive Feature Refinement**: We design a conditional diffusion process that integrates sender features into the ego prediction, guided by global structure priors and conditioned on transformed element features.
- **Efficiency-Optimized Diffusion**: Starting from the ego's element features instead of random noise, we accelerate fusion with an Uncertainty-Adaptive Time Scheduler to dynamically adjust diffusion iterations based on prediction uncertainty.

Experiment

Quantitative Results



(a) Our CoGMP achieves the best performance-bandwidth trade-off.

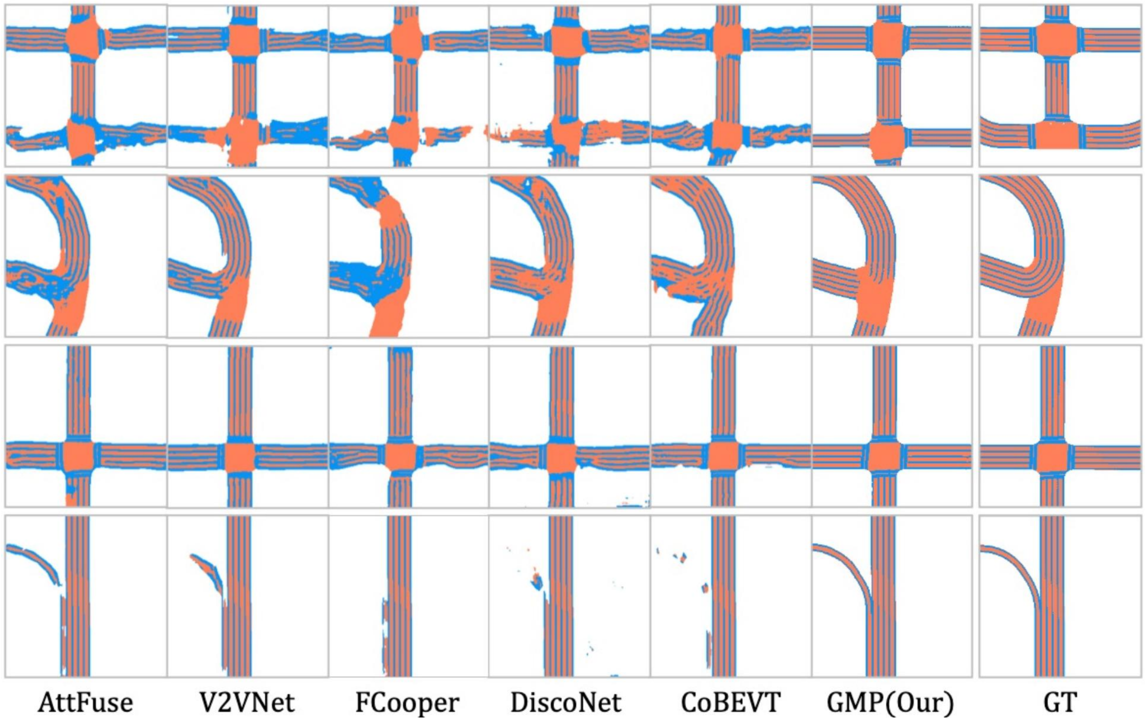


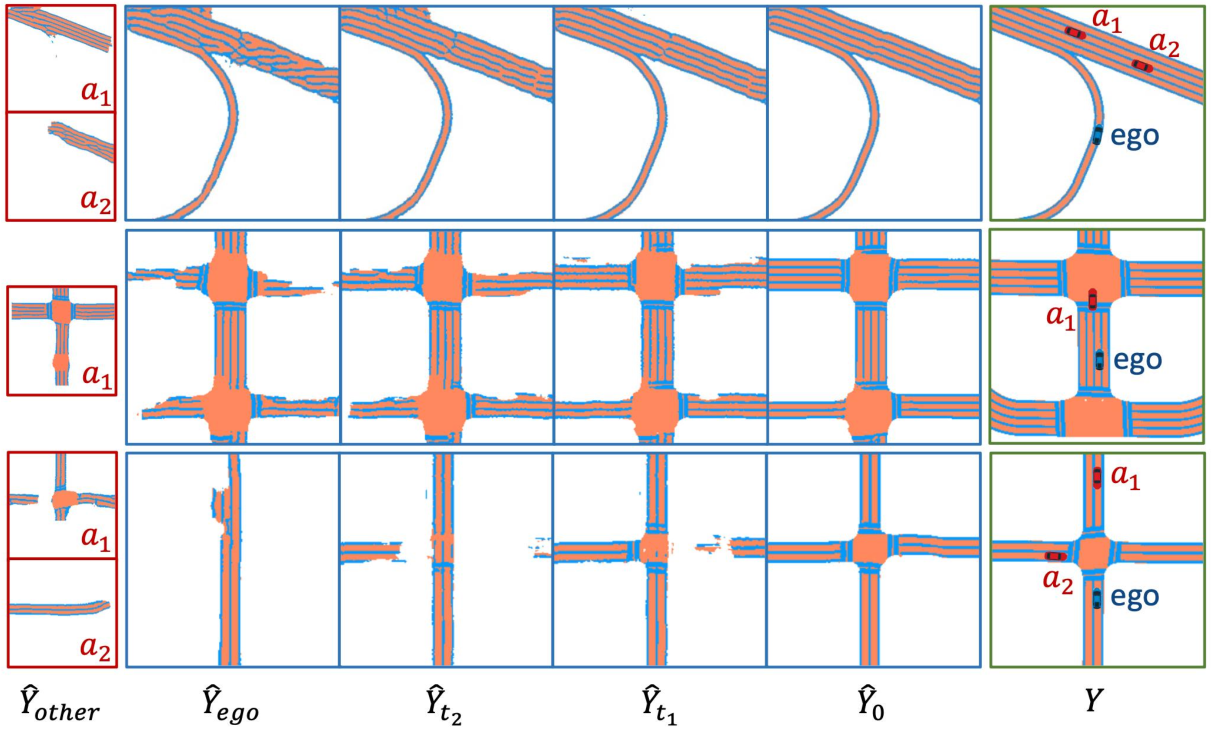
(b) Our CoGMP exhibits the best performance under arbitrary pose error.

- **Efficiency of EFFC Compression** for Performance-Bandwidth Trade-off : we achieve SOTA performance of 66.19/52.92 Road/Lane IoU. With only $\frac{1}{512}$ bandwidth of previous methods, our method still achieves accuracy gains of 3.03/4.33 Road/Lane IoU.
- **Robustness of SGFF Fusion** Under Arbitrary Pose Error : As the pose error noise increases, the performance of our method decreases slightly, with only reduction of 2.43/4.71 Road/Lane IoU, while consistently outperforming previous methods.

Experiment

Qualitative Results





- **Internal priors** provide essential knowledge of road and lane structures, helping to correct incomplete or imprecise ego predictions, such as shaky lane boundaries.
- **External inputs** from other agents help fill in missing details that the ego agent cannot perceive, such as information about distant intersections.

Generative Map Priors for Collaborative BEV Semantic Segmentation

Thanks for your listening