SCORPION: Robust Spatial-Temporal Collaborative Perception Model Design on Lossy Network

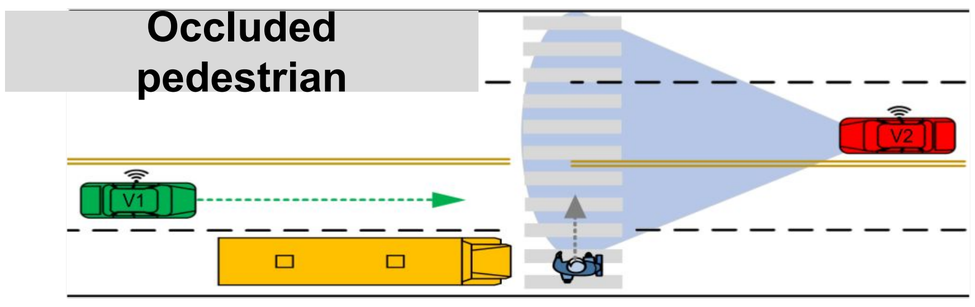
Ruiyang Zhu , Minkyoung Cho , Shuqing Zeng†, Fan Bai†, Z. Morley Mao

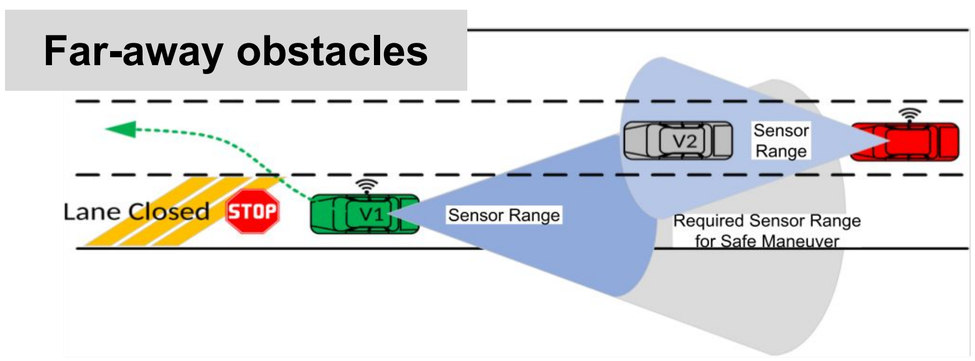
University of Michigan † General Motors Research & Development

Background

Collaborative Perception

- *Effective way to mitigate the limited sensing on single AV perception*

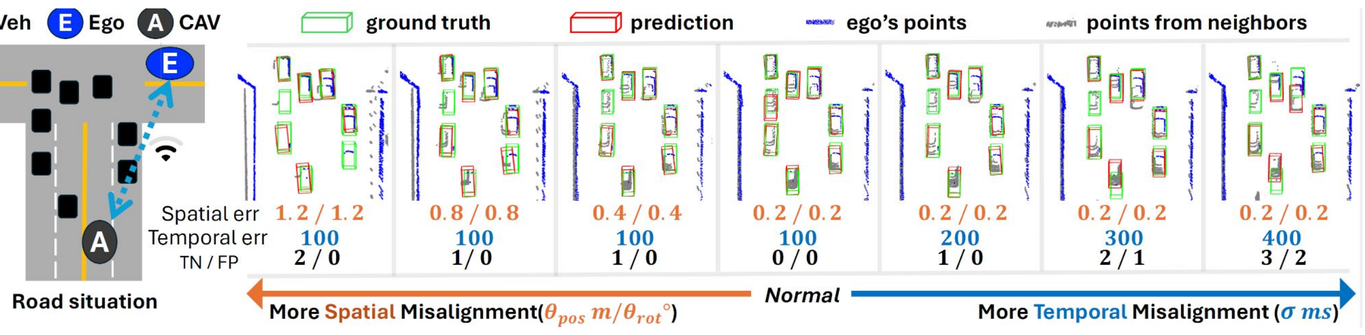




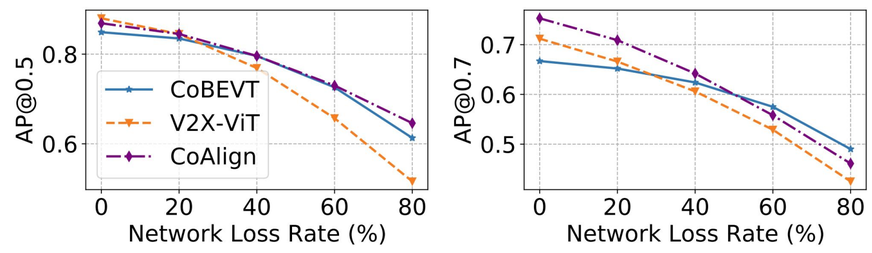
Main Challenges

Imperfections in underlying system layers

- *Spatial misalignments occur due to sensing errors or dropped network packets*
- *Temporal misalignments arise from sensor asynchronization and network delays*

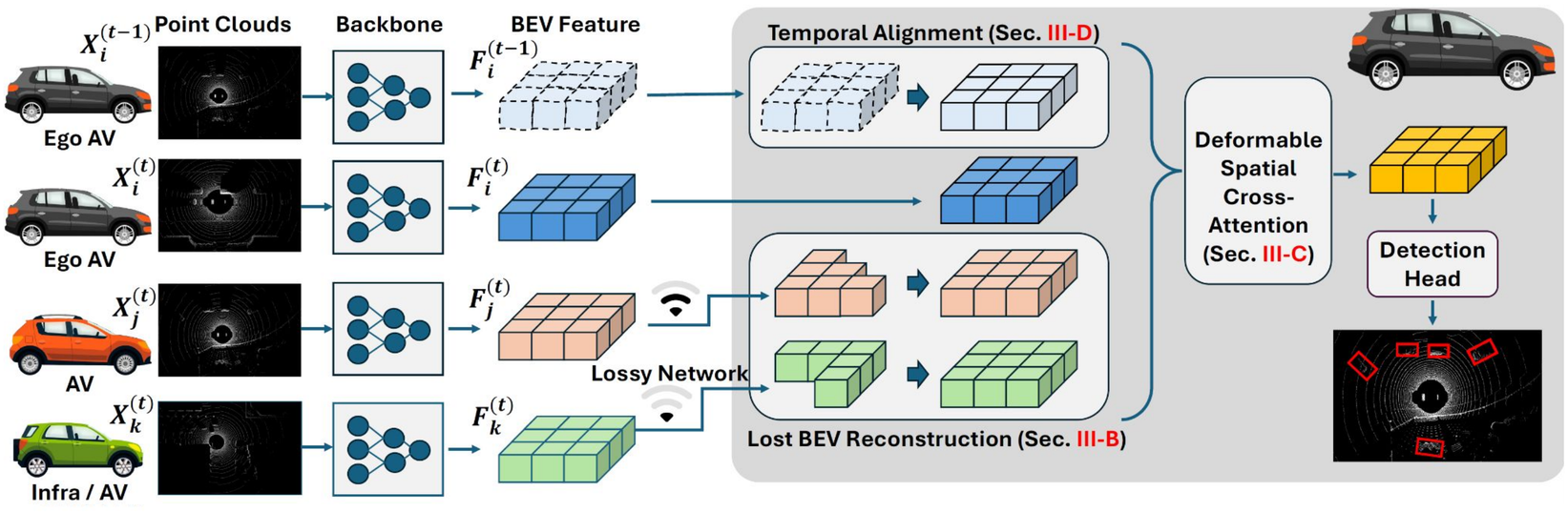


Lossy V2X Network Transmission



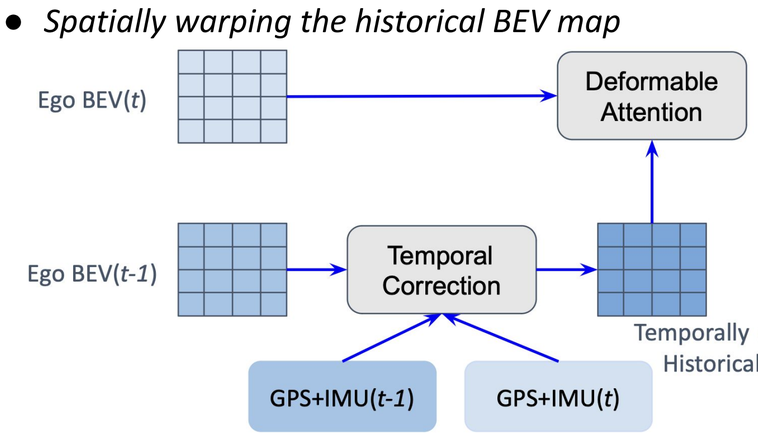
SCORPION: Spatial-temporal Collaborative Perception model on lossy Network

Goal: end-to-end Intermediate-fusion model to address and compensate for the imperfections in system layers



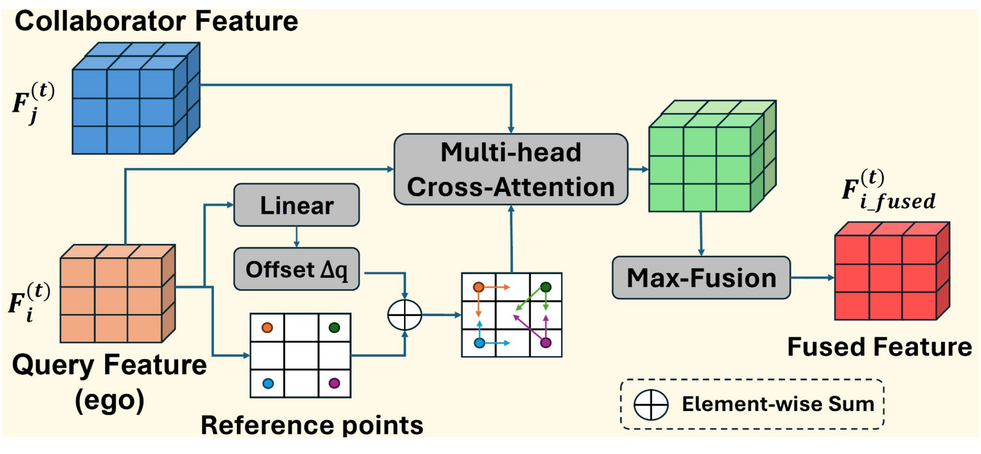
Historical BEV Temporal Alignment (TA)

- *Spatially warping the historical BEV map*



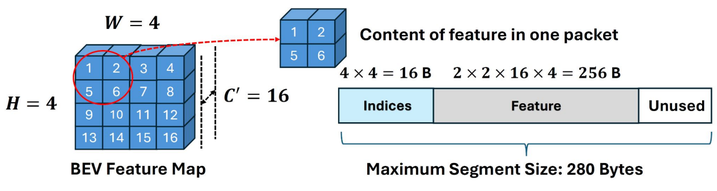
Deformable Spatial Cross Attention (DSCA)

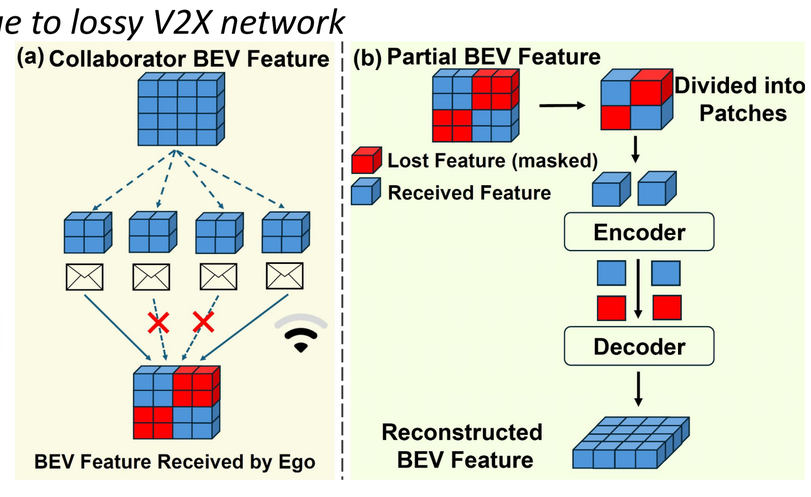
DSCA considers semantic information for localization errors.



Lost BEV Feature Reconstruction (L-BEV-R)

- *The received map has feature indices lost due to lossy V2X network*
- *Masked Autoencoder for reconstruction*
  - *Encoder [1] processed the patches*
  - *Decoder recovers original BEV feature*



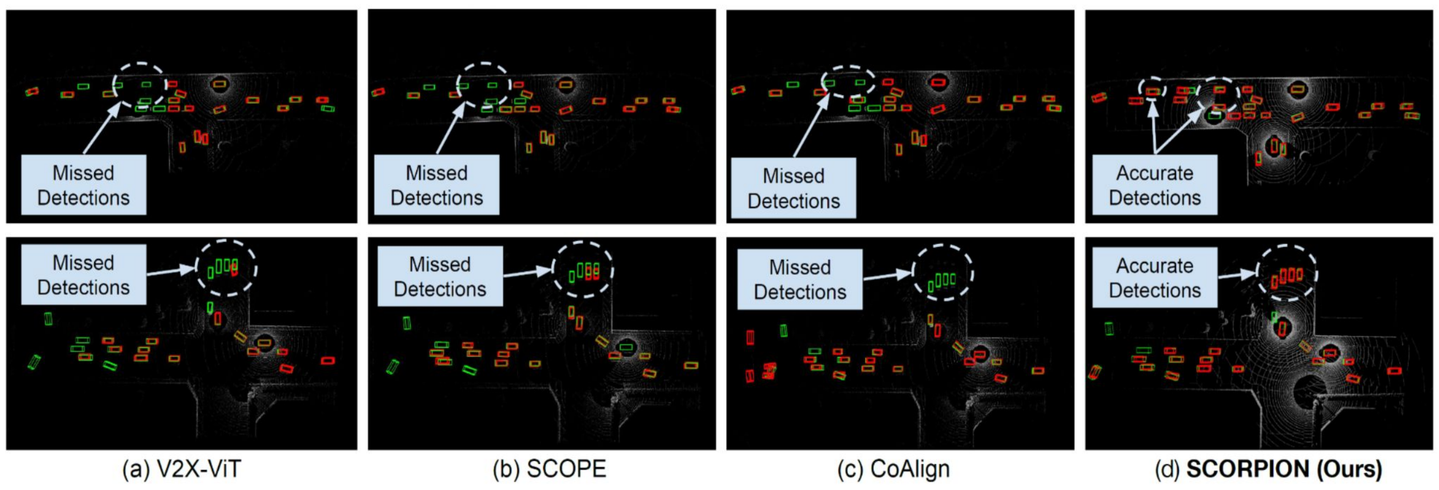


Evaluation Results

SCORPION achieves SOTA performance

Model	V2XSet		OPV2V		DAIR-V2X	
	AP0.5	AP0.7	AP0.5	AP0.7	AP0.5	AP0.7
No Fusion	65.73	52.57	69.38	56.40	63.04	47.39
V2VNet [8]	87.82	74.28	86.76	73.38	65.09	48.18
F-Cooper [10]	82.82	69.38	89.22	79.66	70.54	52.21
AttFuse [7]	81.70	66.24	88.54	72.91	68.02	48.40
CoBEVT [1]	81.00	65.06	88.99	72.80	67.61	55.51
V2X-ViT [2]	82.32	71.21	86.74	75.70	70.87	54.35
CoAlign [5]	86.90	75.31	91.60	82.30	74.02	**56.81**
SCOPE [13]	87.55	75.67	89.60	80.71	74.15	56.52
SCORPION	**88.32**	**77.78**	**93.10**	**85.10**	**74.65**	56.76

Visualization Results Under coexistence of net loss, loc err and sync err



(a) V2X-ViT (b) SCOPE (c) CoAlign (d) **SCORPION (Ours)**

SCORPION outperforms baselines under various levels of network loss & loc/sync errors

References

[1] Masked Autoencoders Are Scalable Vision Learners, CVPR 22

[2] V2X-ViT: Vehicle-to-Everything Cooperative Perception with Vision Transformer, ECCV 22

[3] OPV2V: an open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication, ICRA 21

[4] DAIR-V2X and OpenDAIRV2X: Towards General and Real-World Cooperative Autonomous Driving, CVPR 22