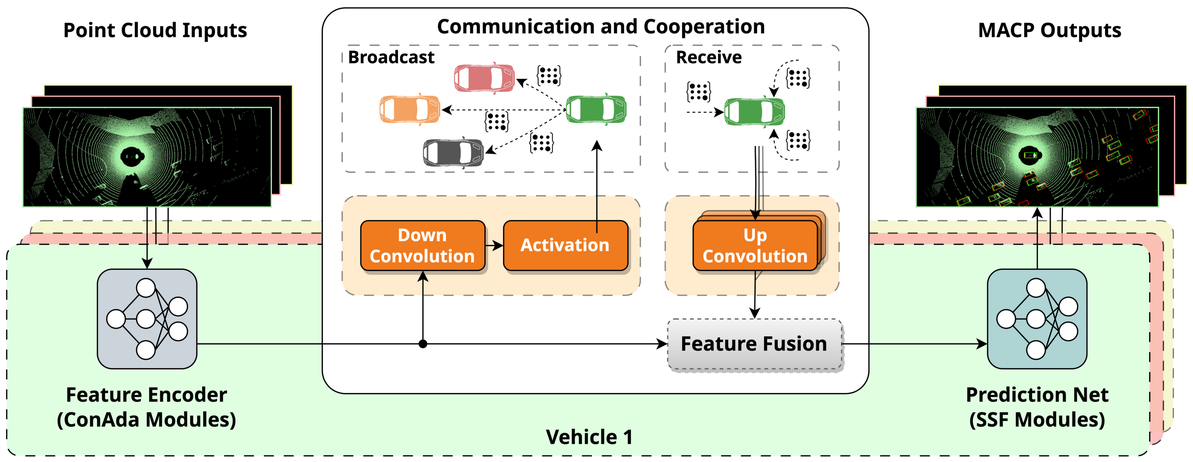
MACP: Efficient Model Adaptation for Cooperative Perception

Yunsheng Ma*, Juanwu Lu*, Can Cui, Sicheng Zhao, Xu Cao, Wenqian Ye, Ziran Wang

JAN 4-8 WACV 2024 WAIKOLOA HAWAII

Overview

- Identify the gaps between the single-agent perception and cooperative perception
- Propose a general framework enabling easy adaptation from single-agent to cooperative perception
- Evaluate our method on simulation and real-world data, demonstrating **SOTA** performance



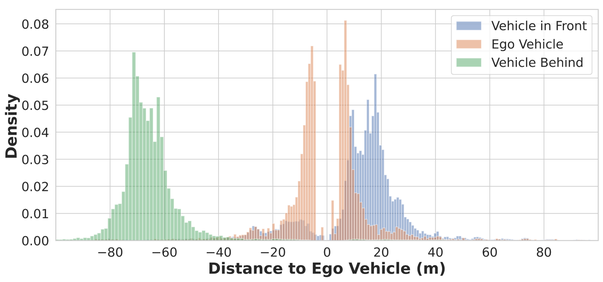
Methodology

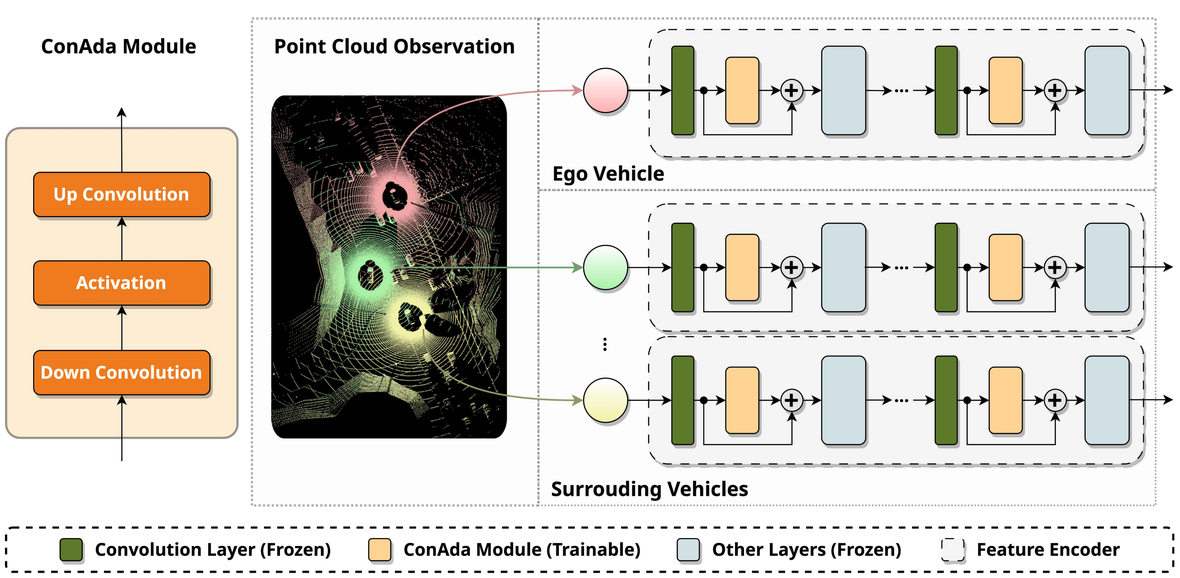
Challenges

- **Spatial Distribution Shifts**
  - Point clouds around different vehicles differ in scale and location
  - Simple fusion, like concatenation, can lead to multi-modal distribution
- **Feature Space Shifts**
  - $p(x, y) = p(y|x)p(x)$ where we assert $p_S(x) \neq p_T(x); p_S(y|x) = p_T(y|x)$
  - Sources of feature space shifts:
    - Difference between pre-train and application cases
    - Enriched information through sharing compressed on the same latent space
- **Communication Bottleneck**
  - Communication channels are limited resources.
  - Information-sharing on ad-hoc vehicular networks can be tricky.

Framework and Modules

- MACP is a **distributed framework**.
  - Each vehicle encodes point-cloud features locally.
  - Information-sharing after point-cloud encoding.
  - Predictions are derived locally using fused features.
- Module: **Convolution Adapter (ConAda)**
  - Projects and filters on low-dimensional space
    $O = Conv[Activation(Conv(I, K_{down})), K_{up})$
  - Compresses and filters shared information
- Module: **Scale and Shift Features** (*Lian et al. 2022*)
  - Rescale and shift feature to align latent spaces
    $X_{i,j}^{output} = \gamma \odot X_{i,j}^{input} + \beta$





Results

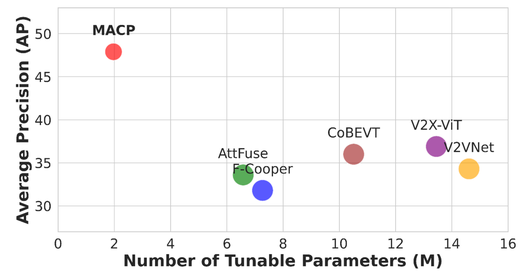
On the V2V4Real Dataset, the proposed MACP model outperforms the leading SOTA model by achieving a 30% improvement in Average Precision (AP) at Intersection over Union (IoU) = 70 while requiring only 15% of the number of tunable parameters and 65% of the volume of data transmission.

Method	Param (M)		AP@IoU=50/70 (↑)				AM (↓)
	Total	Trainable	Overall	0-30m	30-50m	50-100m	(MB)
No Fusion	6.58	6.58	39.8/22.0	69.2/42.6	29.3/14.4	4.8/1.6	0
Late Fusion	6.58	6.58	55.0/26.7	73.5/36.8	43.7/22.2	36.2/17.3	0.003
Early Fusion	6.58	6.58	59.7/32.1	76.1/46.3	42.5/20.8	47.6/21.1	0.96
F-Cooper [7]	7.27	7.27	60.7/31.8	80.8/46.9	45.6/23.6	32.8/13.4	0.20
V2VNet [38]	14.61	14.61	64.5/34.3	80.6/51.4	52.6/26.6	42.6/14.6	0.20
AttFuse [45]	6.58	6.58	64.7/33.6	79.8/44.1	53.1/29.3	43.6/19.3	0.20
V2X-ViT [44]	13.45	13.45	64.9/36.9	82.0/55.3	51.7/26.6	43.2/16.2	0.20
CoBEVT [41]	10.51	10.51	66.5/36.0	82.3/51.1	52.1/28.2	**49.1**/19.5	0.20
MACP (**Ours**)	8.94	1.97	**67.6/47.9**	**83.7/62.1**	**58.4/38.5**	34.6/**23.1**	0.13

1) Performance Comparison on the V2V4Real Dataset

Method	Param (M)		AP@IoU=50/70 (↑)	
	Total	Trainable	Default Towns	Culver City
No Fusion	6.58	6.58	67.9/60.2	55.7/47.1
Late Fusion	6.58	6.58	85.8/78.1	79.9/66.8
Early Fusion	6.58	6.58	89.1/80.0	82.9/69.6
F-Cooper [7]	7.27	7.27	88.7/79.0	84.6/72.8
V2VNet [38]	14.61	14.61	89.7/82.2	86.0/73.4
AttFuse [45]	6.58	6.58	90.8/81.5	85.4/73.5
V2X-ViT [44]	13.45	13.45	89.1/82.6	87.3/73.7
CoBEVT [41]	10.51	10.51	91.4/86.1	85.9/77.2
AdaFusion [30]	7.27	7.27	91.6/85.6	88.1/79.0
MACP (**Ours**)	8.98	2.00	**93.7/90.3**	**91.4/80.7**

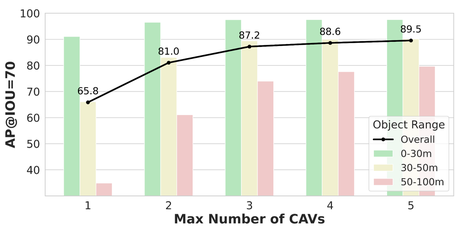
2) Performance Comparison on the OPV2V Dataset



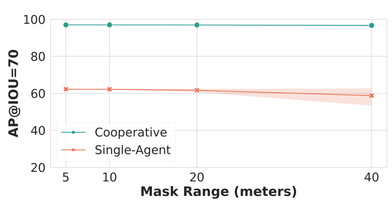
3) Visualization of Results on the V2V4Real Dataset

Method	Param (M)		AP@IoU=50/70 (↑)			
	Total	Trainable	Overall	0-30m	30-50m	50-100m
Full Fine-Tune	8.92	8.92	68.3/**51.9**	**85.0/65.7**	**59.1/41.2**	33.0/**26.9**
Train Fusion & Head Only	8.92	1.94	65.5/42.1	83.1/57.4	51.9/30.0	33.5/19.9
Adapter Only [19]	9.17	2.19	63.8/37.5	81.0/50.8	49.1/26.5	33.7/17.6
SSF Only [25]	8.92	1.95	64.7/44.2	82.7/60.2	50.8/30.9	32.6/21.0
ConAda Only	8.97	2.00	67.5/49.0	84.1/62.2	57.0/38.6	34.1/25.6
MACP	8.98	2.00	**69.4**/49.6	83.2/63.1	58.6/40.0	**38.5**/26.5

4) Effectiveness of Proposed Components



5) Effect of Compression Factors

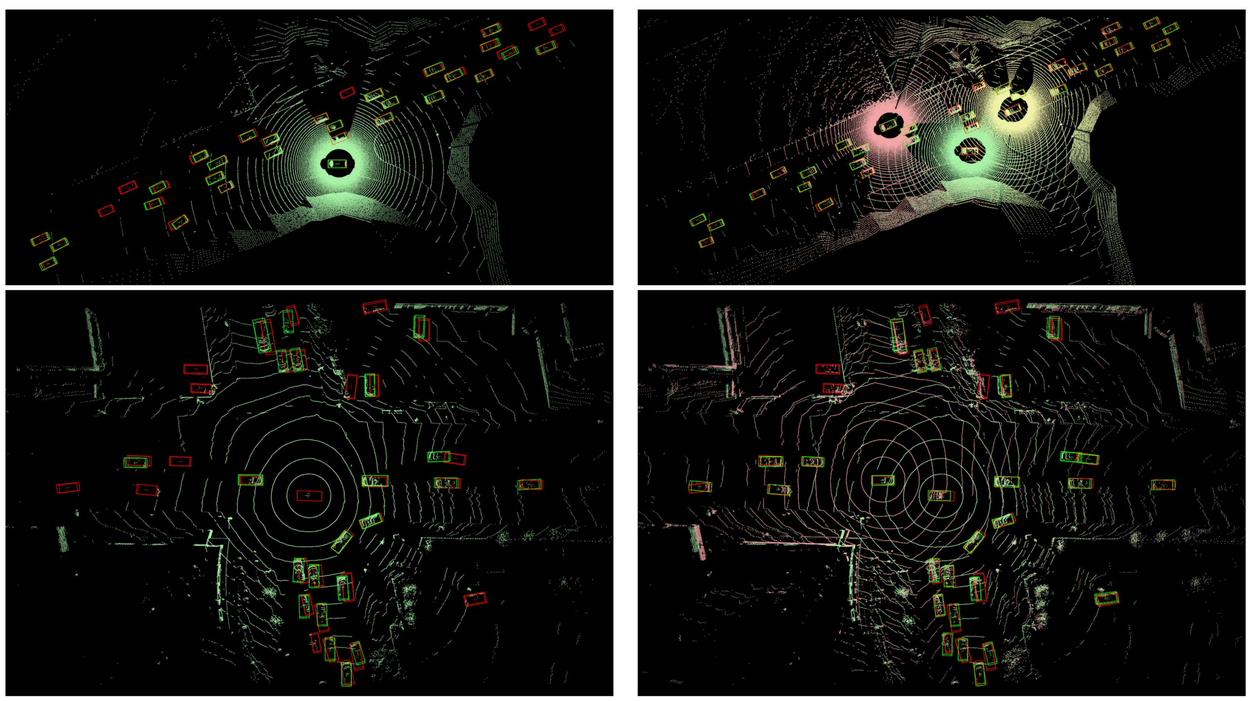


6) Results of the robustness analysis

7) Effect of Max Number of CAVs

Visualization

The 3D bounding boxes in red and green represent the ground-truth and predicted objects, respectively.



1) BEVFusion (Single-Agent Perception)

2) MACP (Cooperative Perception)