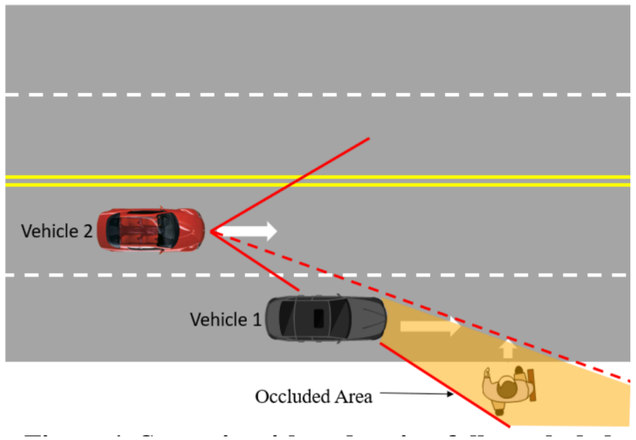
STAMP: Scalable Task- And Model-agnostic Collaborative Perception

ICLR 2025

Xiangbo Gao[1], Runsheng Xu[2], Jiachen Li[3], Ziran Wang[4], Zhiwen Fan[5], Zhengzhong Tu[1,*]

[1]Texas A&M University [2]UCLA [3]UC Riverside
[4]Purdue University [5]UT Austin

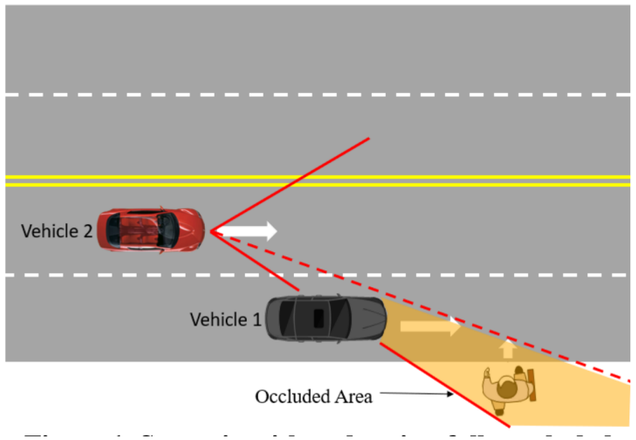
Limitation of Single Agent Perception



Object Occlusion [1]

[1] Zou, Fengjiao, et al. "V2V-based Collision-avoidance Decision Strategy for Autonomous Vehicles Interacting with Fully Occluded Pedestrians at Midblock on Multilane Roadways." ICTD, *2023*.

Limitation of Single Agent Perception

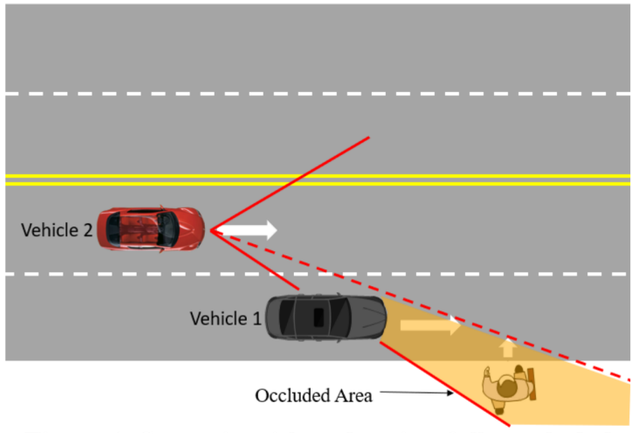


Object Occlusion [1]

Sensor limitation

[1] Zou, Fengjiao, et al. "V2V-based Collision-avoidance Decision Strategy for Autonomous Vehicles Interacting with Fully Occluded Pedestrians at Midblock on Multilane Roadways." ICTD, *2023*.

Limitation of Single Agent Perception



Object Occlusion [1]

Sensor limitation

Severe Weather

[1] Zou, Fengjiao, et al. "V2V-based Collision-avoidance Decision Strategy for Autonomous Vehicles Interacting with Fully Occluded Pedestrians at Midblock on Multilane Roadways." ICTD, *2023*.

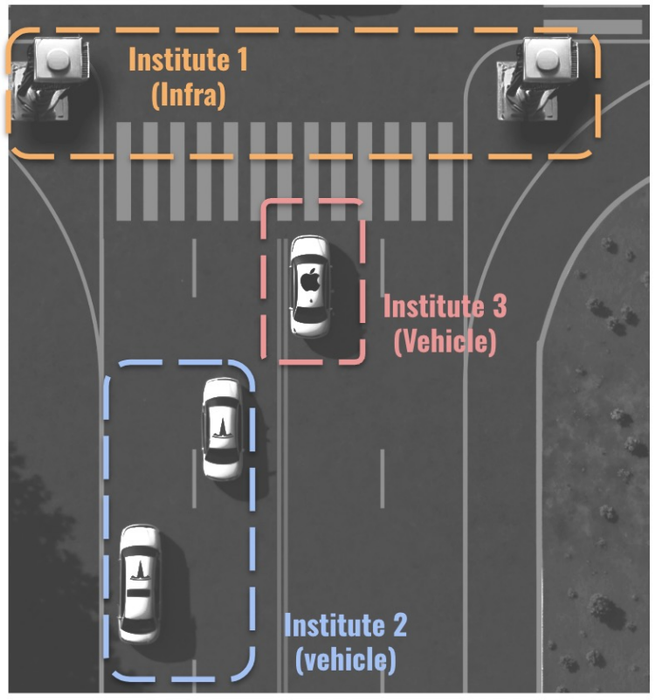
Multi-agent Collaborative Perception

Multi-agent Collaborative Perception [2]

[2] Huang, Tao, et al. "V2X cooperative perception for autonomous driving: Recent advances and challenges." arXiv, 2023.

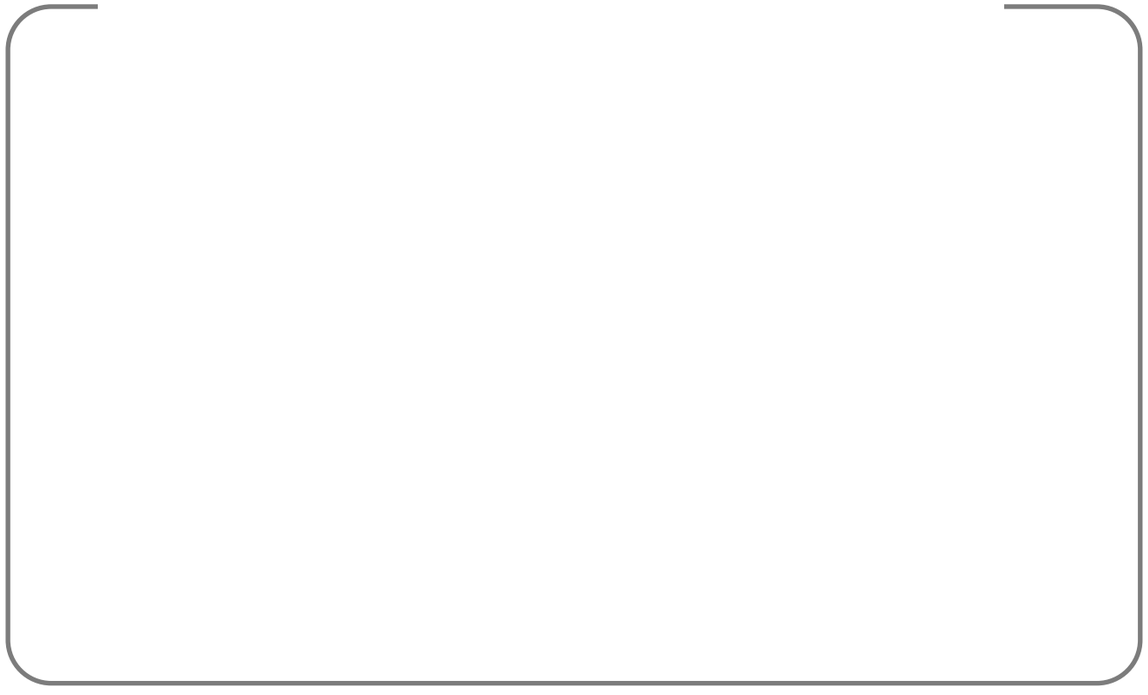
Multi-agent Collaborative Perception

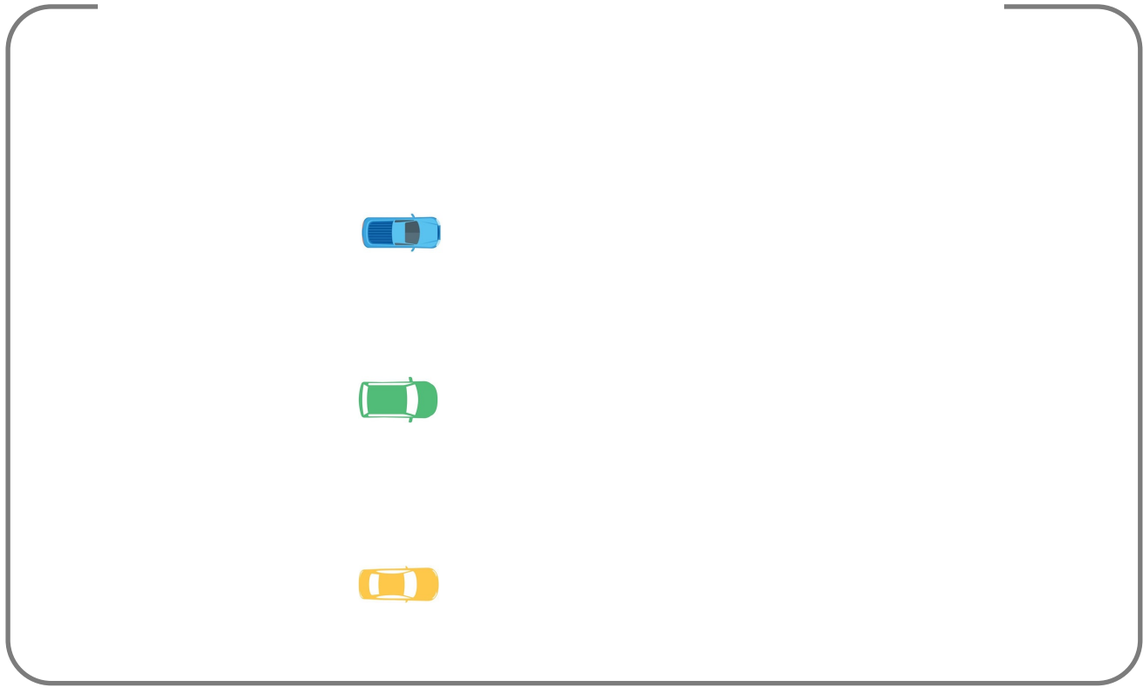
Multi-agent Collaborative Perception [2]

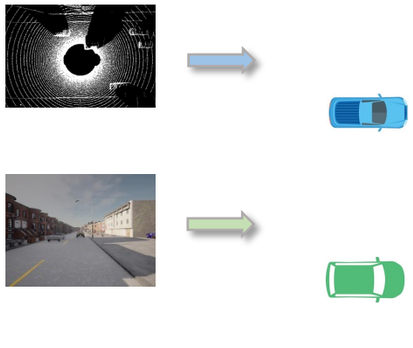


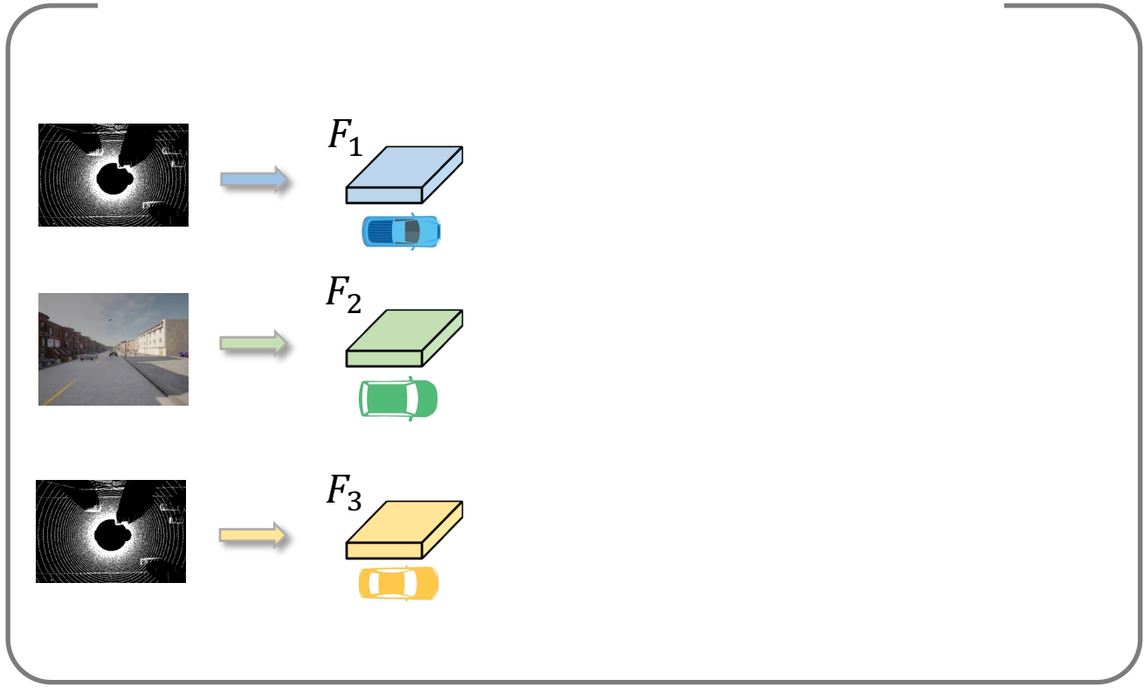
Agent Heterogeneity

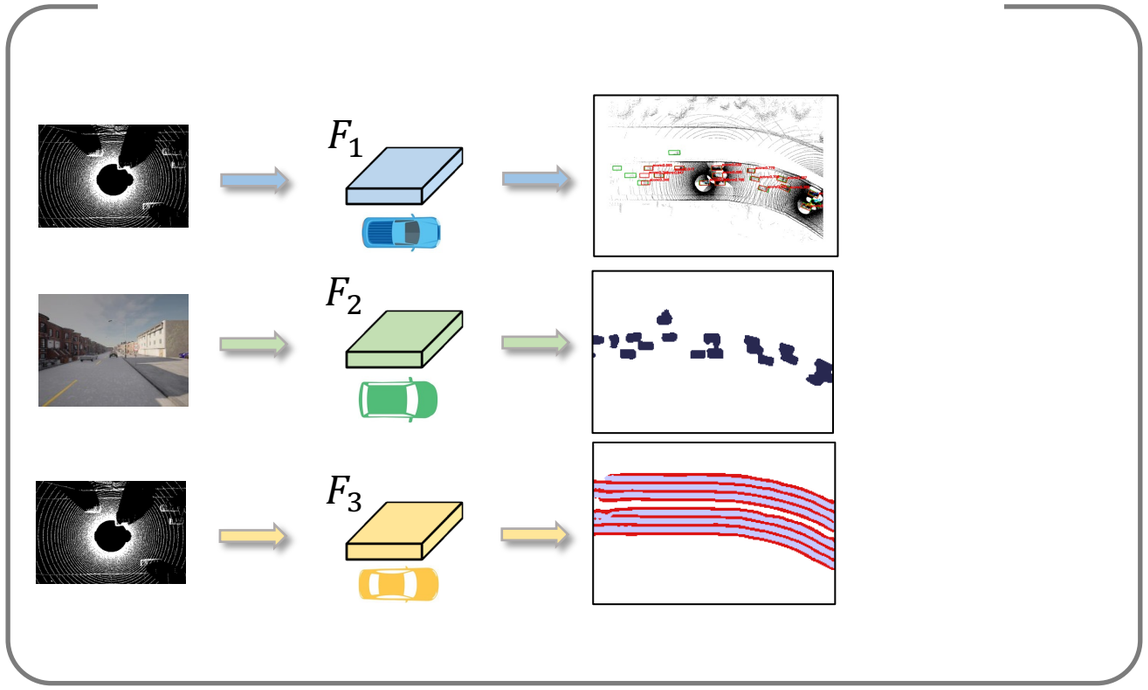
[2] Huang, Tao, et al. "V2X cooperative perception for autonomous driving: Recent advances and challenges." arXiv, 2023.

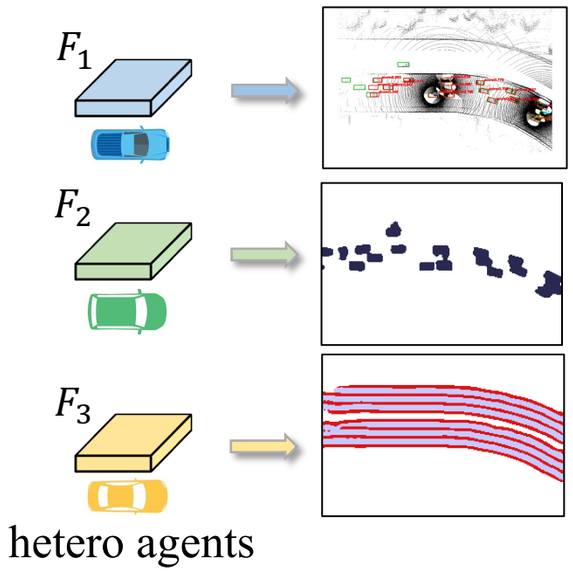




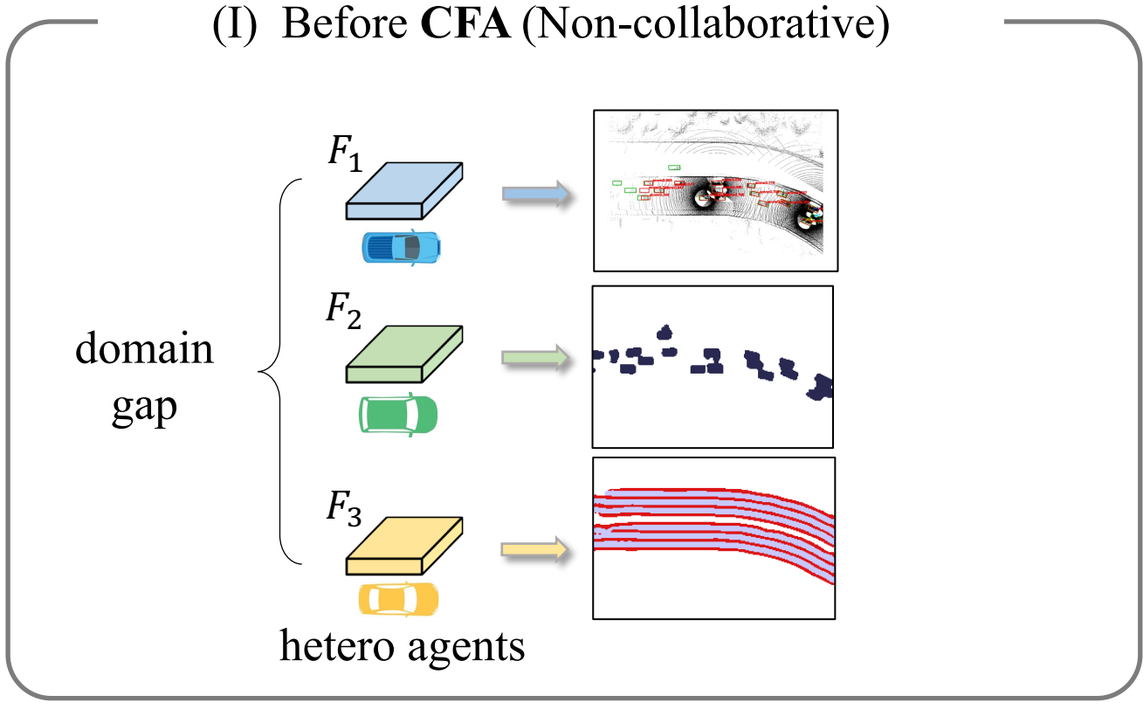



F_1
F_2
F_3

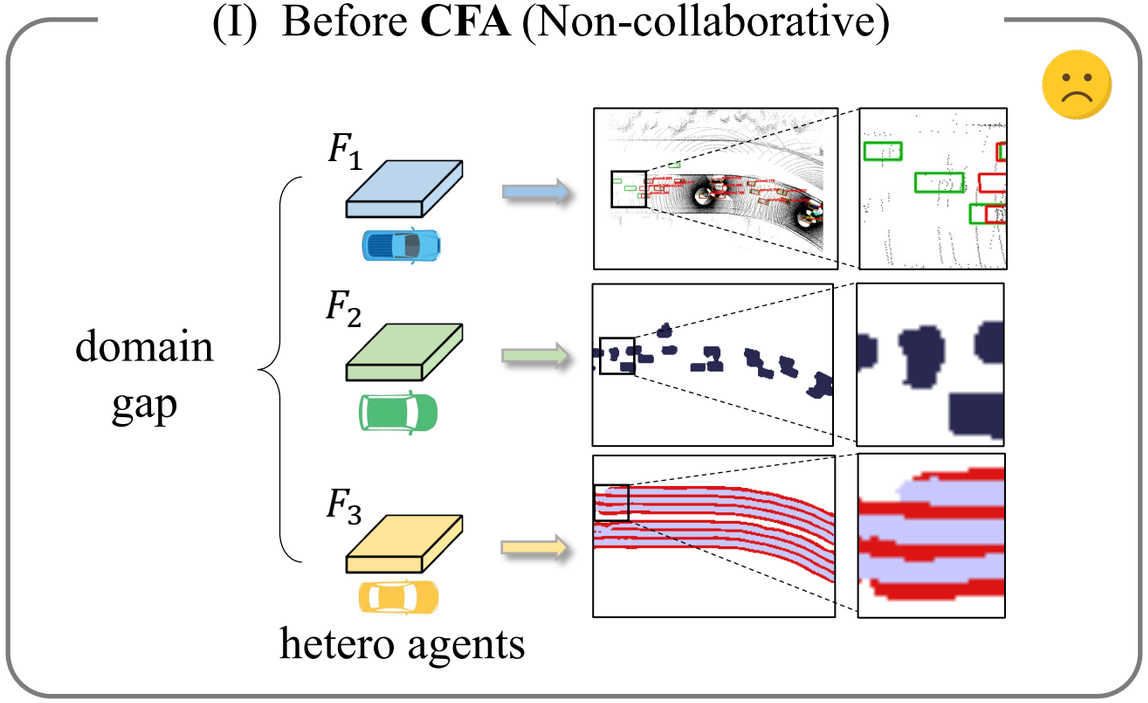

F_1
F_2
F_3



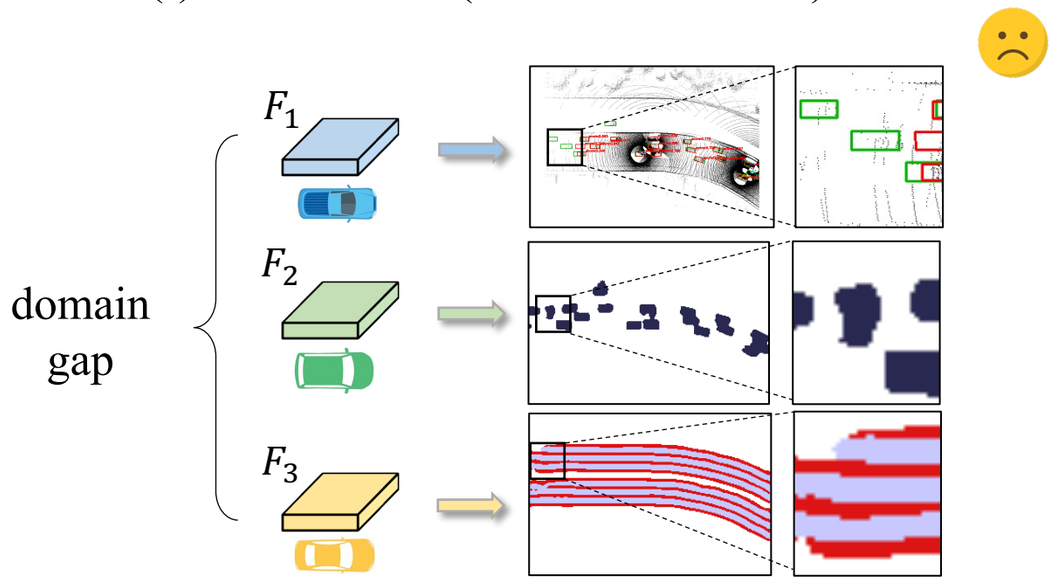
Heterogeneous in input modalities, model architectures, downstream tasks, etc.



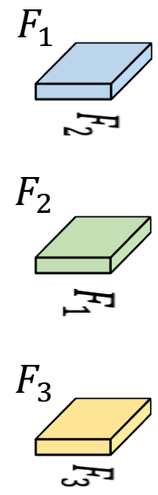
Heterogeneous in input modalities, model architectures, downstream tasks, etc.


(I) Before **CFA** (Non-collaborative)
domain gap
F_1
F_2
F_3
hetero agents

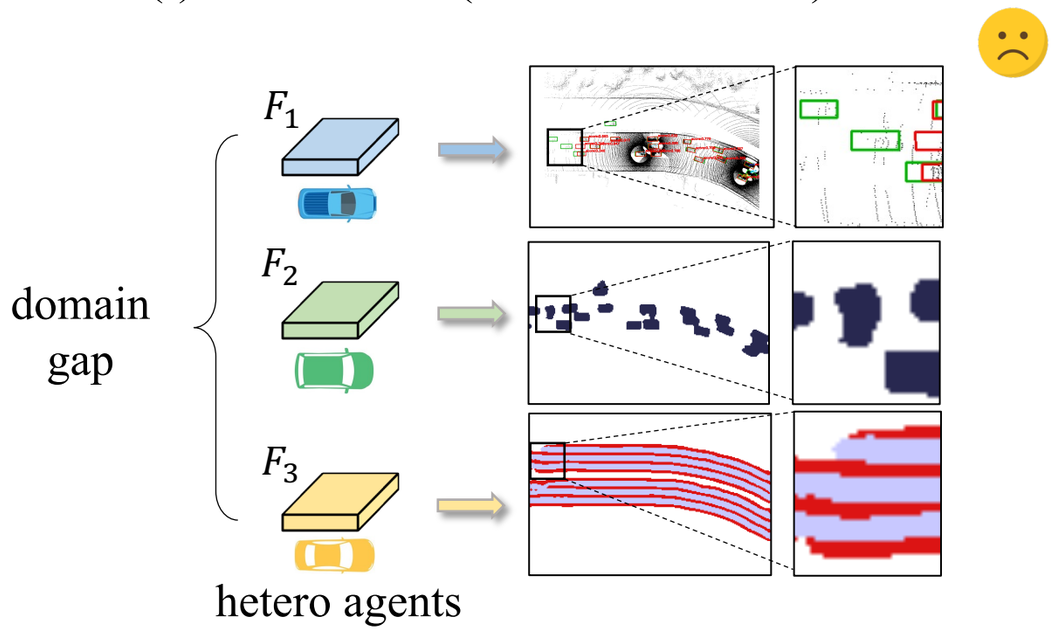
(I) Before **CFA** (Non-collaborative)



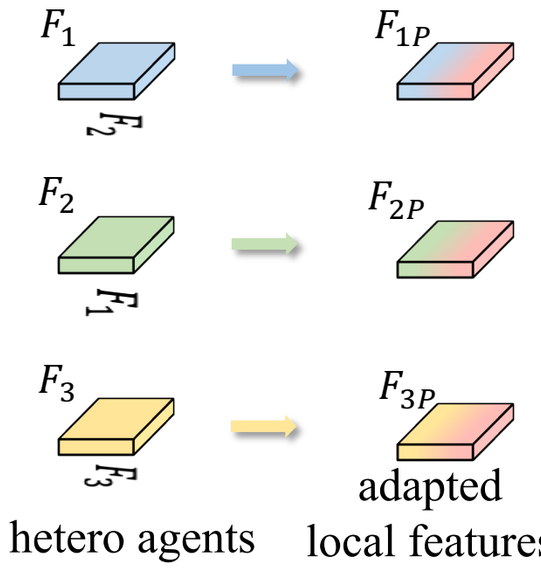
(III) After **CFA** (heterogeneously collaborative)



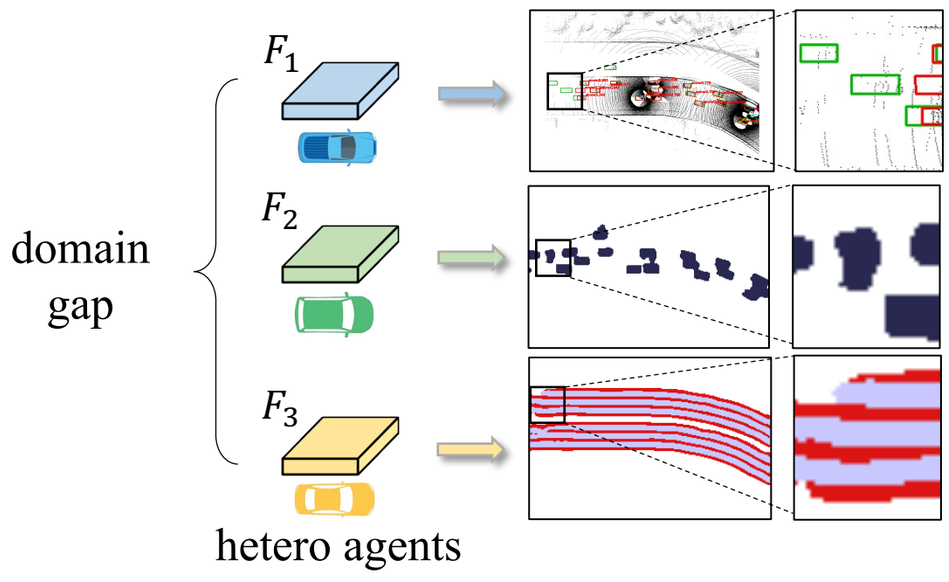
(I) Before **CFA** (Non-collaborative)



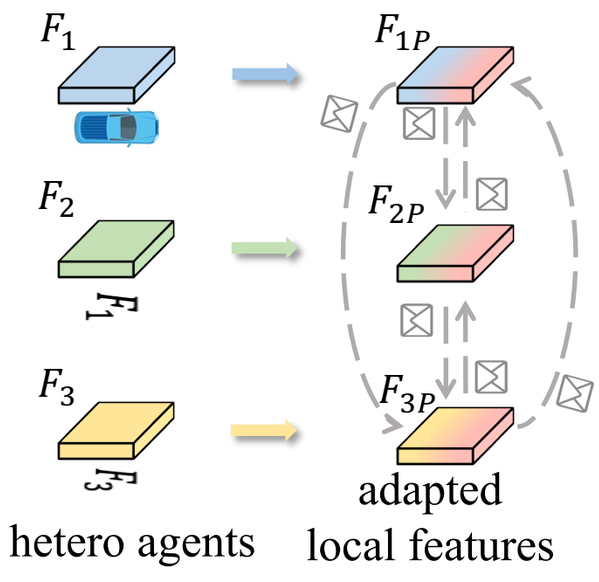
(III) After **CFA** (heterogeneously collaborative)



(I) Before **CFA** (Non-collaborative)



(III) After **CFA** (heterogeneously collaborative)



(I) Before **CFA** (Non-collaborative)

domain gap

F_1

F_2

F_3

hetero agents

(III) After **CFA** (heterogeneously collaborative)

F_1 F_2 F_3

hetero agents

F_{1P} F_{2P} F_{3P}

adapted local features

score:0.525

(I) Before **CFA** (Non-collaborative)

domain gap

F_1

F_2

F_3

hetero agents

(III) After **CFA** (heterogeneously collaborative)

F_1

F_2

F_3

F_{1P}

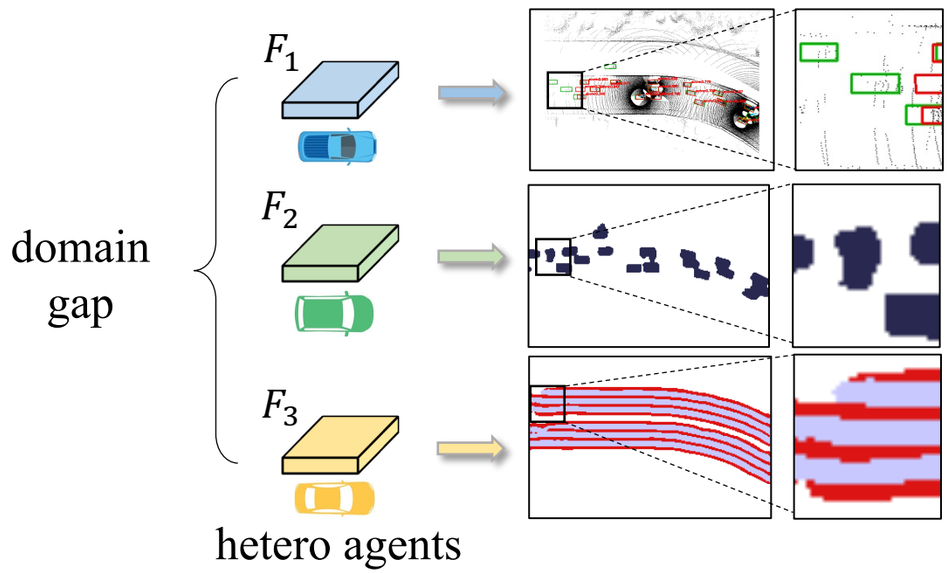
F_{2P}

F_{3P}

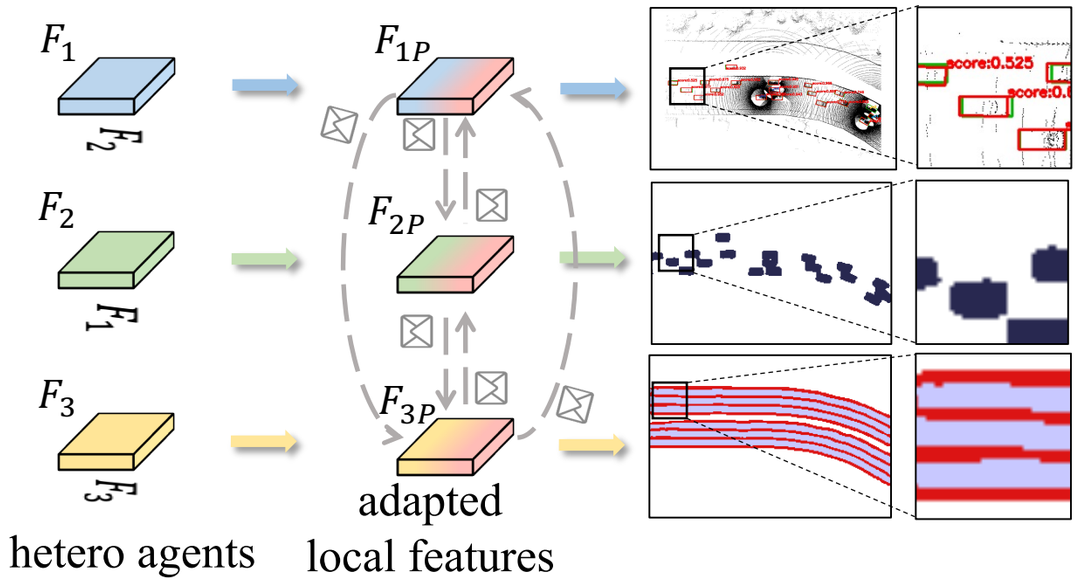
adapted local features

hetero agents

(I) Before **CFA** (Non-collaborative)

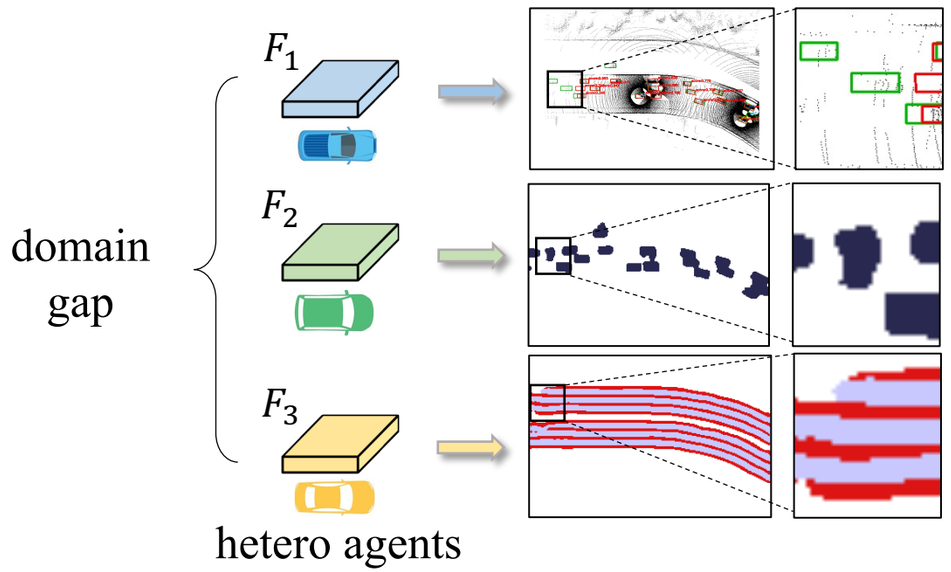


(III) After **CFA** (heterogeneously collaborative)

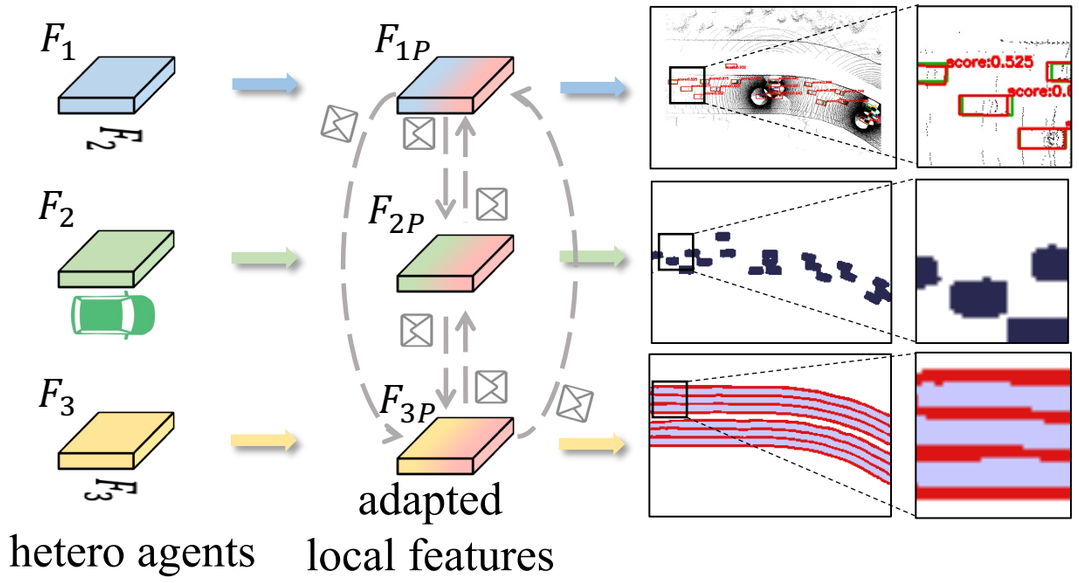


(II) **Collaborative Feature Alignment (CFA)**

(I) Before **CFA** (Non-collaborative)

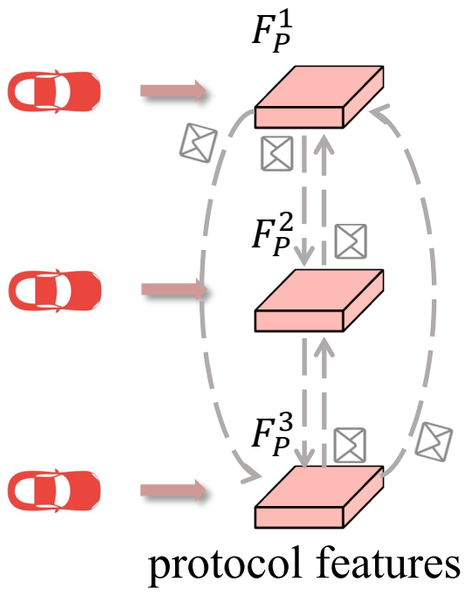


(III) After **CFA** (heterogeneously collaborative)

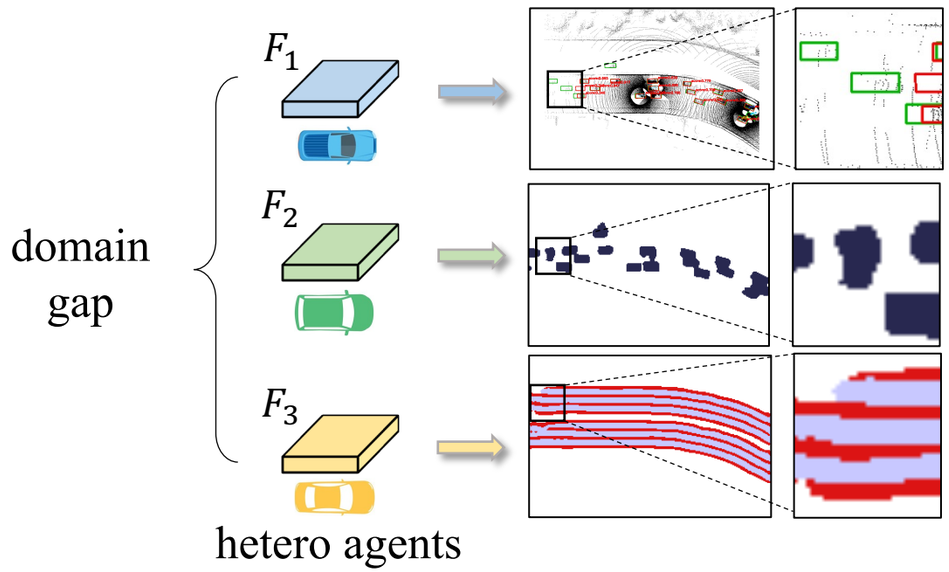


(II) **Collaborative Feature Alignment (CFA)**

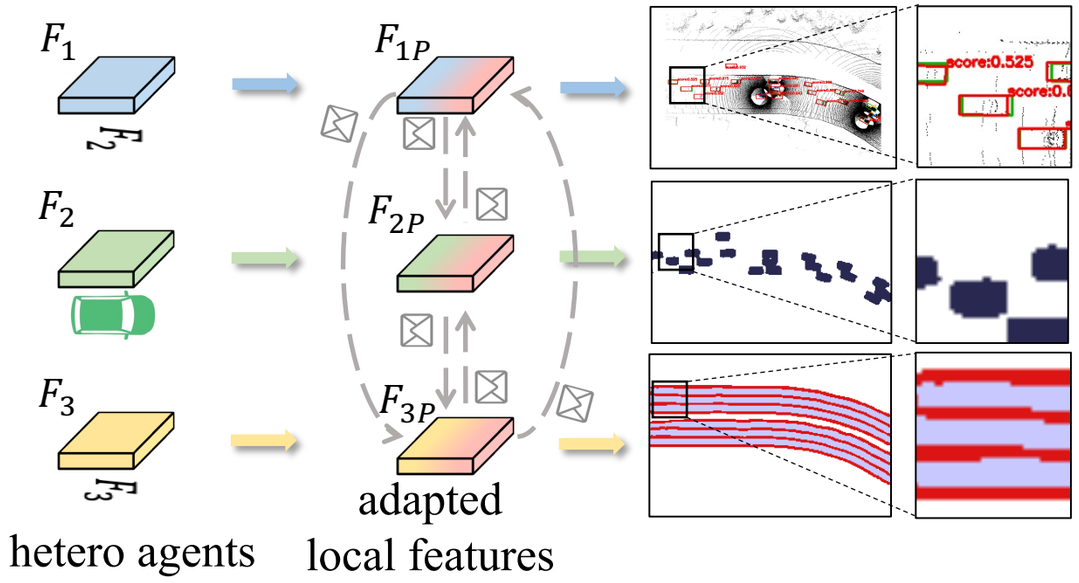
Step 1: Training Protocol Network



(I) Before **CFA** (Non-collaborative)

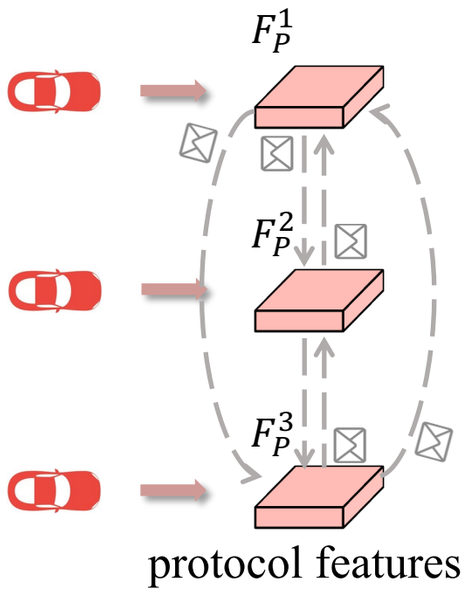


(III) After **CFA** (heterogeneously collaborative)



(II) Collaborative Feature Alignment (CFA)

Step 1: Training Protocol Network



Step 2: Training Local Adapters and Reverters

(I) Before **CFA** (Non-collaborative)

domain gap

F_1

F_2

F_3

hetero agents

(III) After **CFA** (heterogeneously collaborative)

F_1 F_2 F_3

F_{1P} F_{2P} F_{3P}

hetero agents

adapted local features

(II) Collaborative Feature Alignment (CFA)

Step 1: Training Protocol Network

F_P^1 F_P^2 F_P^3

protocol features

Step 2: Training Local Adapters and Reverters

F_2

(I) Before **CFA** (Non-collaborative)

domain gap

F_1

F_2

F_3

hetero agents

(III) After **CFA** (heterogeneously collaborative)

F_1 F_2 F_3

hetero agents

F_{1P} F_{2P} F_{3P}

adapted local features

(II) **Collaborative Feature Alignment (CFA)**

Step 1: Training Protocol Network

F_P^1 F_P^2 F_P^3

protocol features

Step 2: Training Local Adapters and Reverters

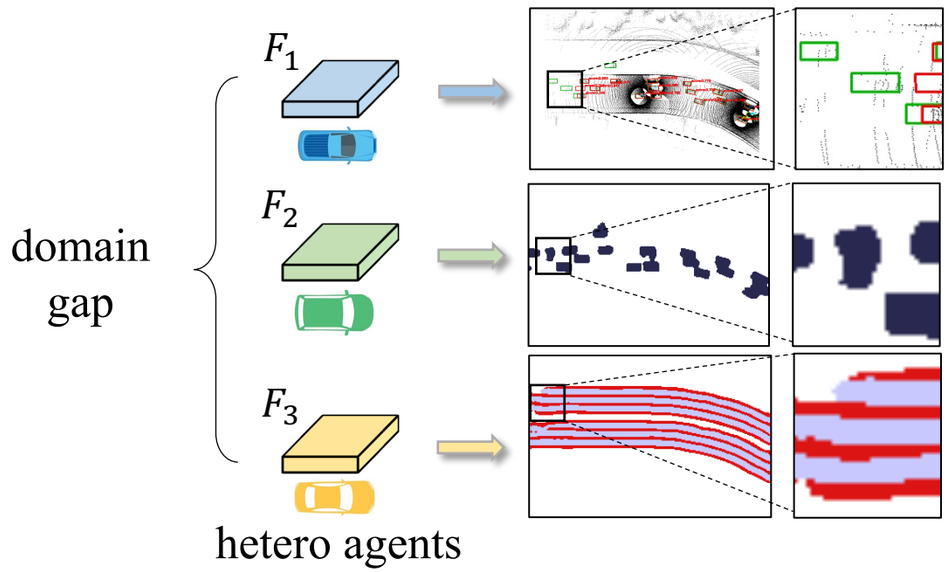
F_2

local feature

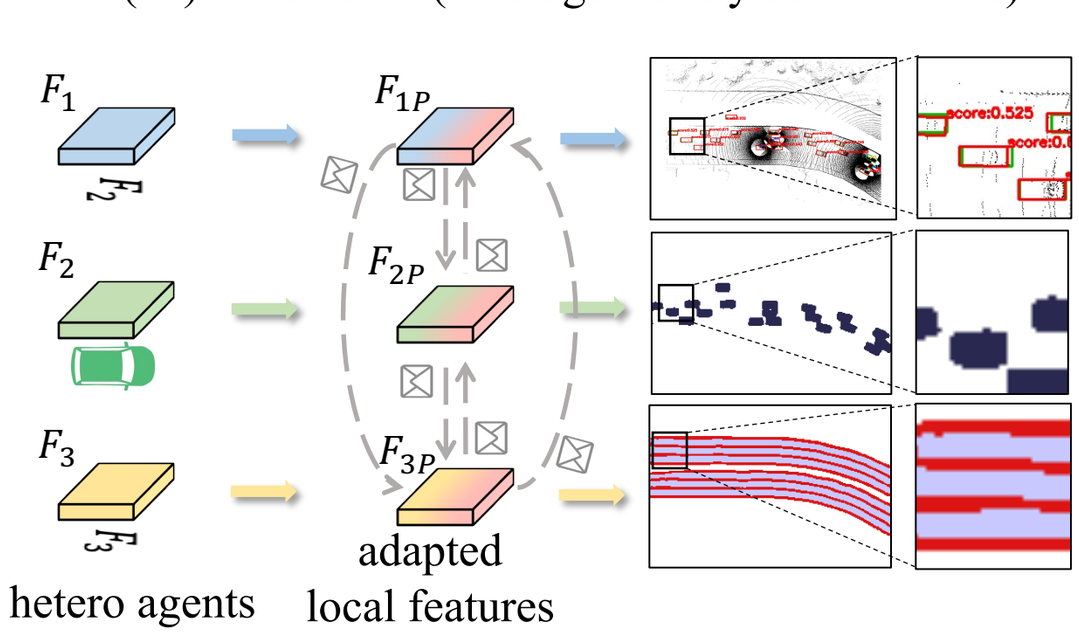
World state

protocol feature

(I) Before **CFA** (Non-collaborative)

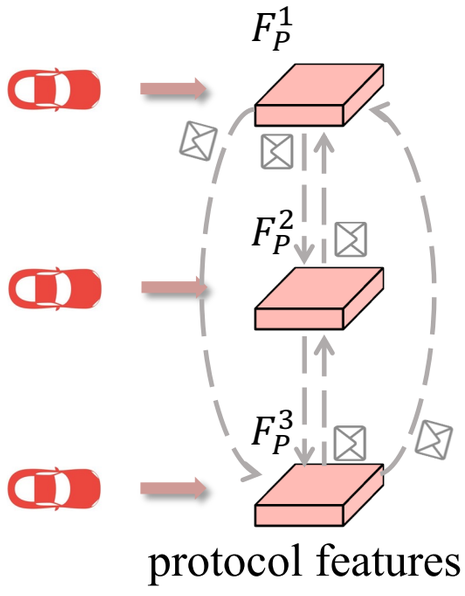


(III) After **CFA** (heterogeneously collaborative)

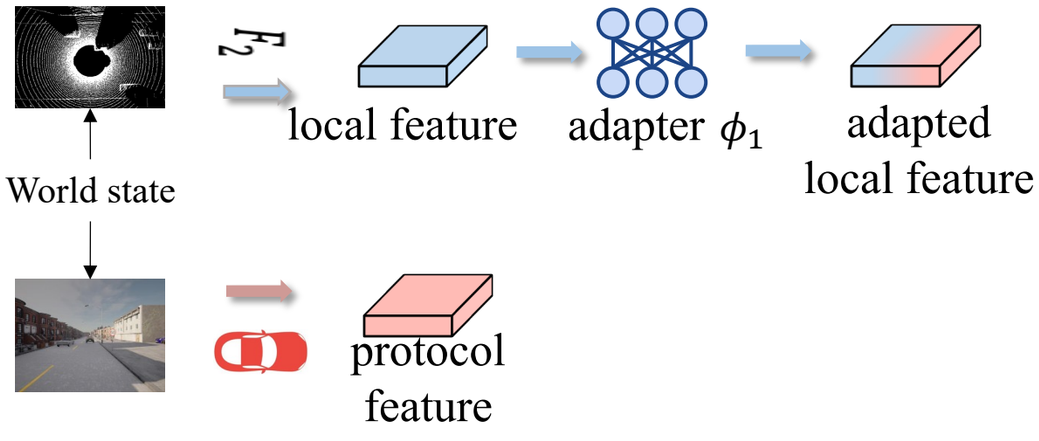


(II) **Collaborative Feature Alignment (CFA)**

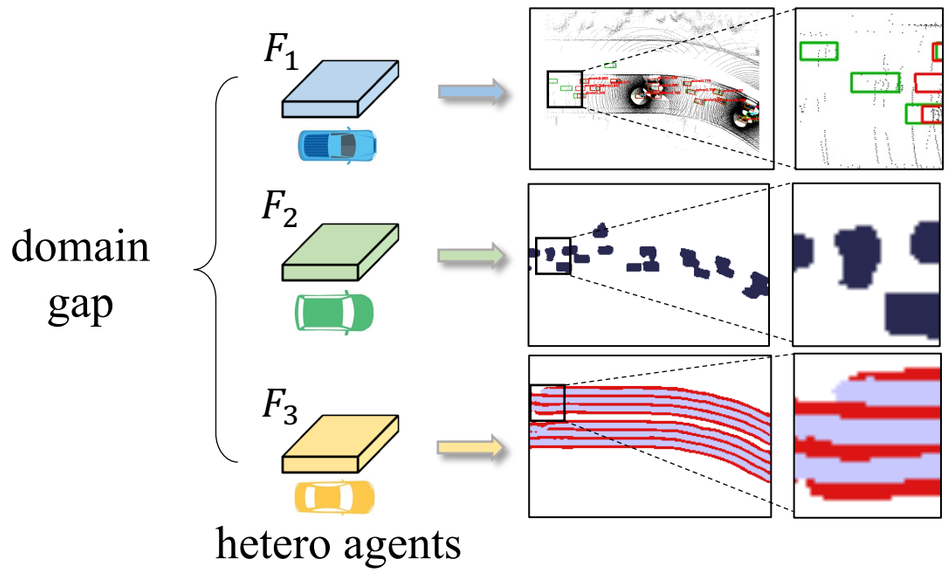
Step 1: Training Protocol Network



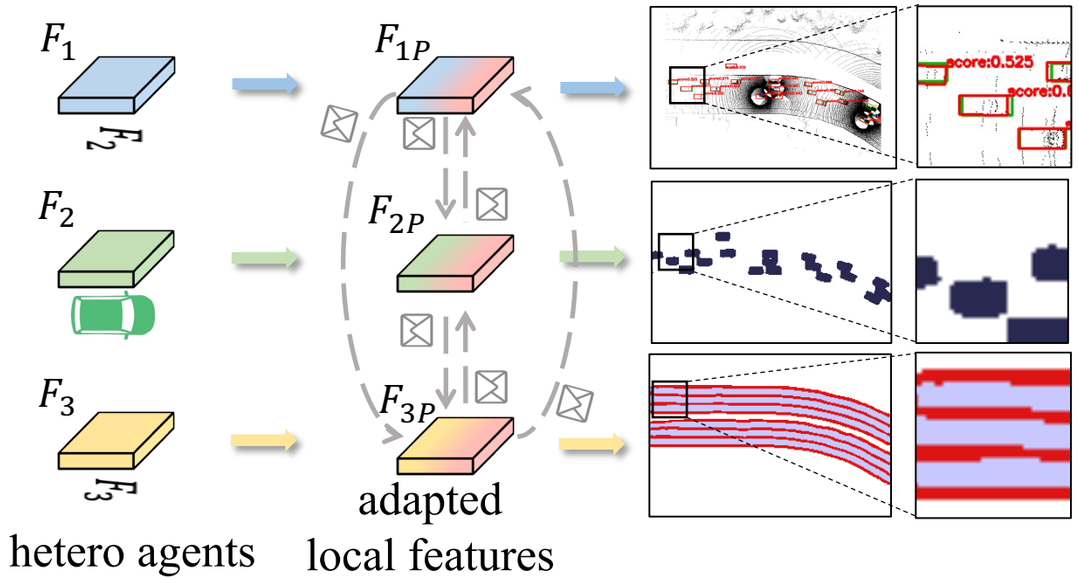
Step 2: Training Local Adapters and Reverters



(I) Before **CFA** (Non-collaborative)

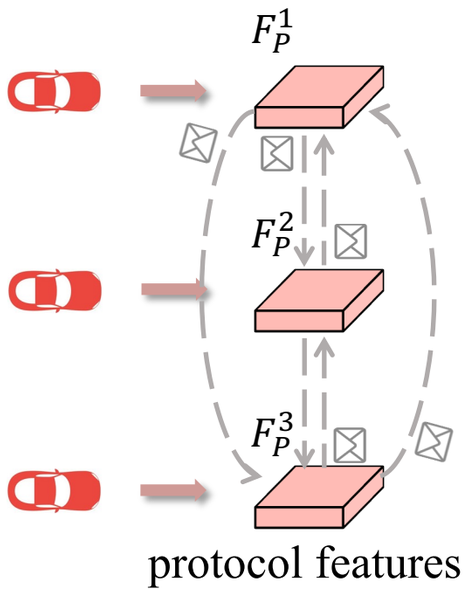


(III) After **CFA** (heterogeneously collaborative)

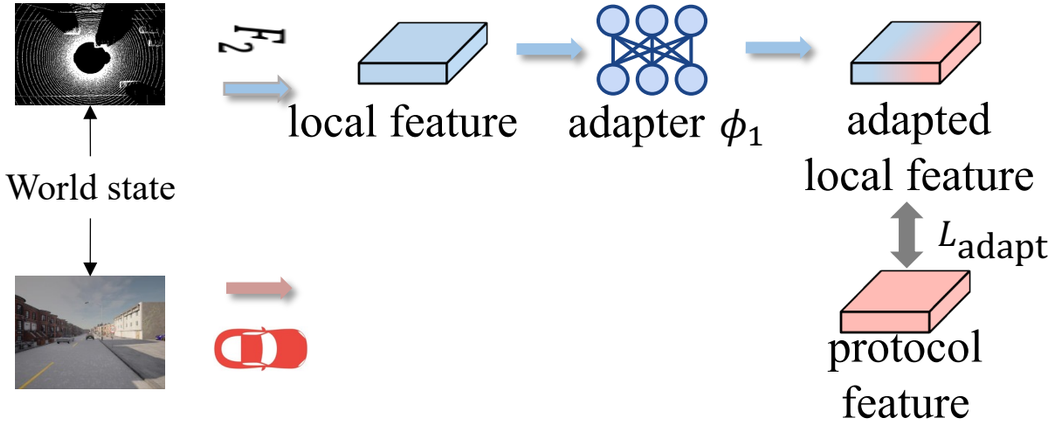


(II) **Collaborative Feature Alignment (CFA)**

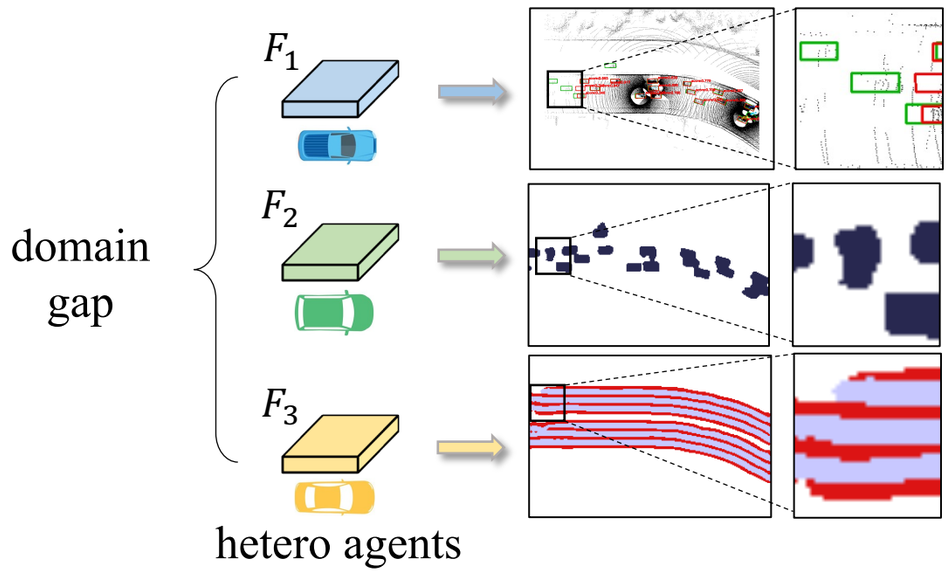
Step 1: Training Protocol Network



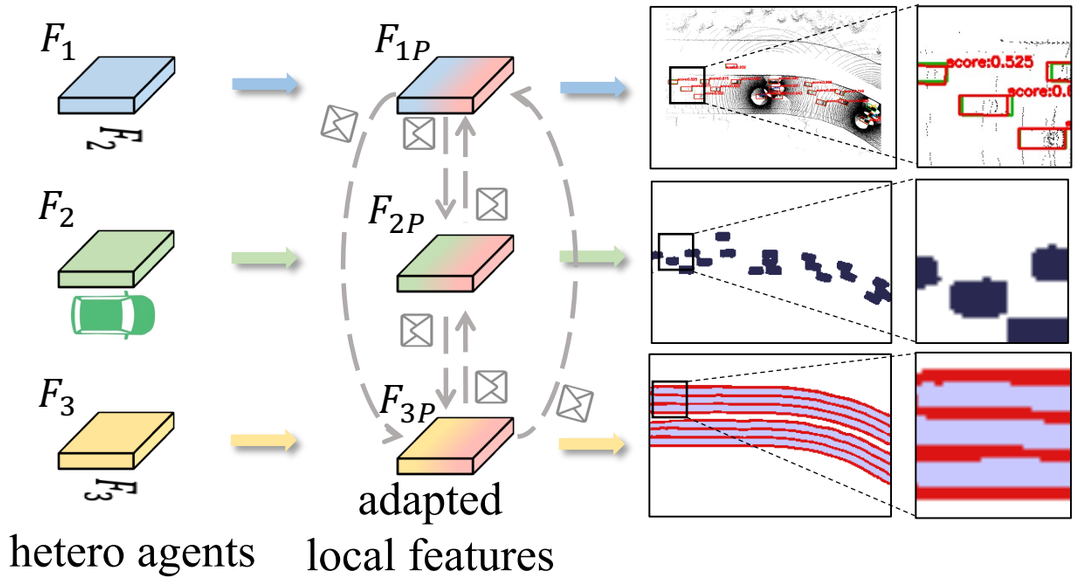
Step 2: Training Local Adapters and Reverters



(I) Before **CFA** (Non-collaborative)

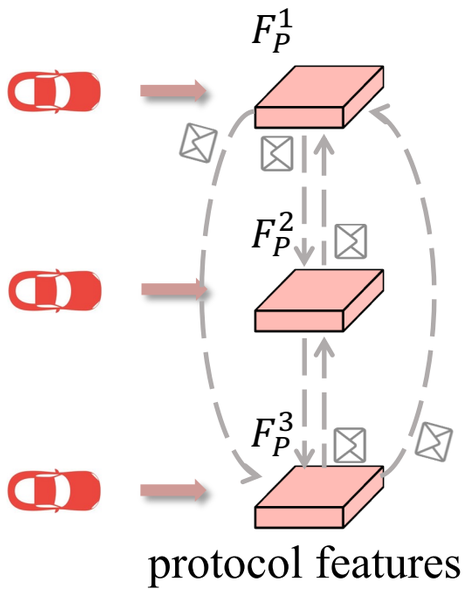


(III) After **CFA** (heterogeneously collaborative)

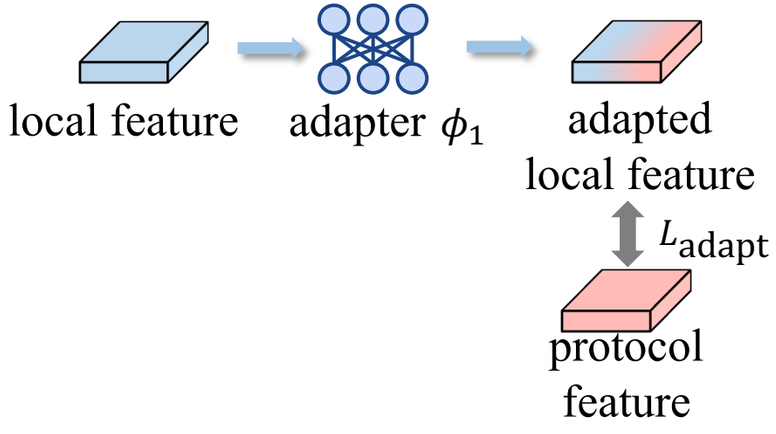


(II) **Collaborative Feature Alignment (CFA)**

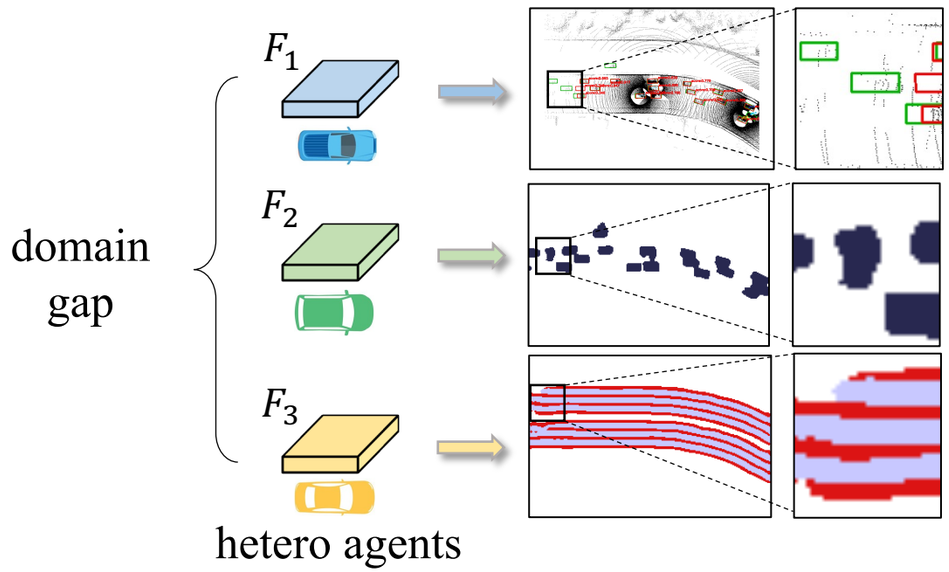
Step 1: Training Protocol Network



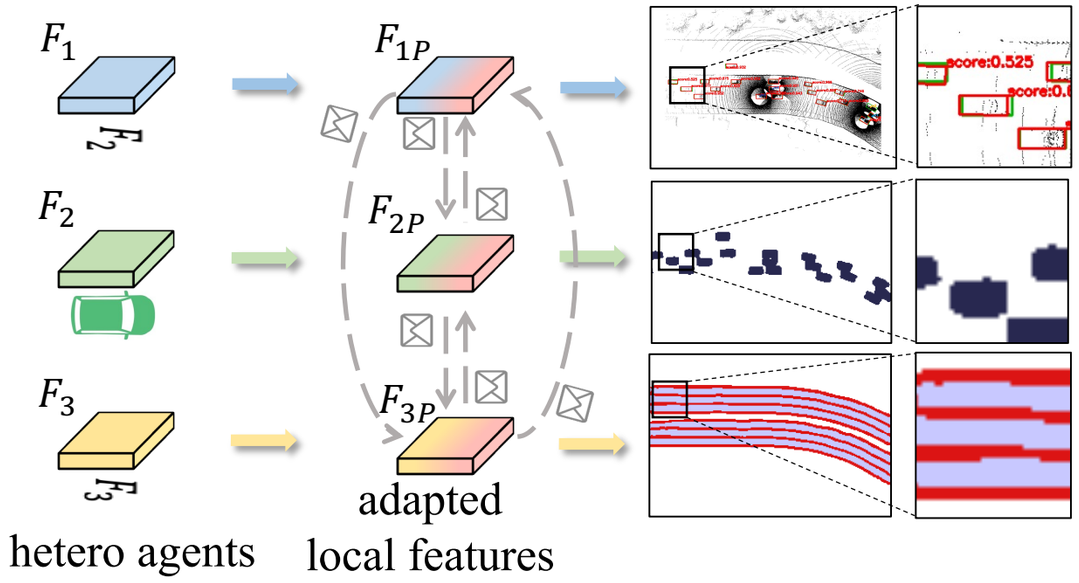
Step 2: Training Local Adapters and Reverters



(I) Before **CFA** (Non-collaborative)

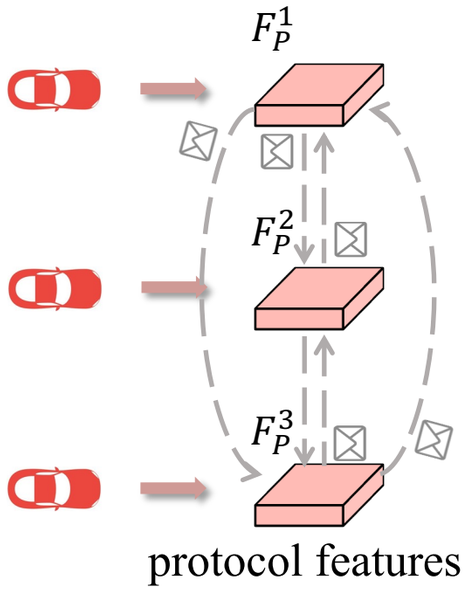


(III) After **CFA** (heterogeneously collaborative)

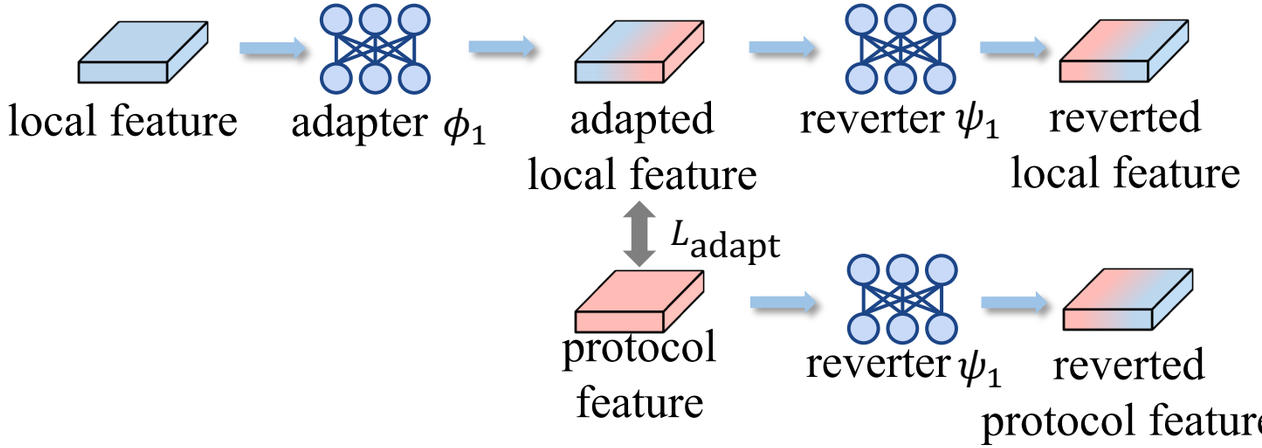


(II) **Collaborative Feature Alignment (CFA)**

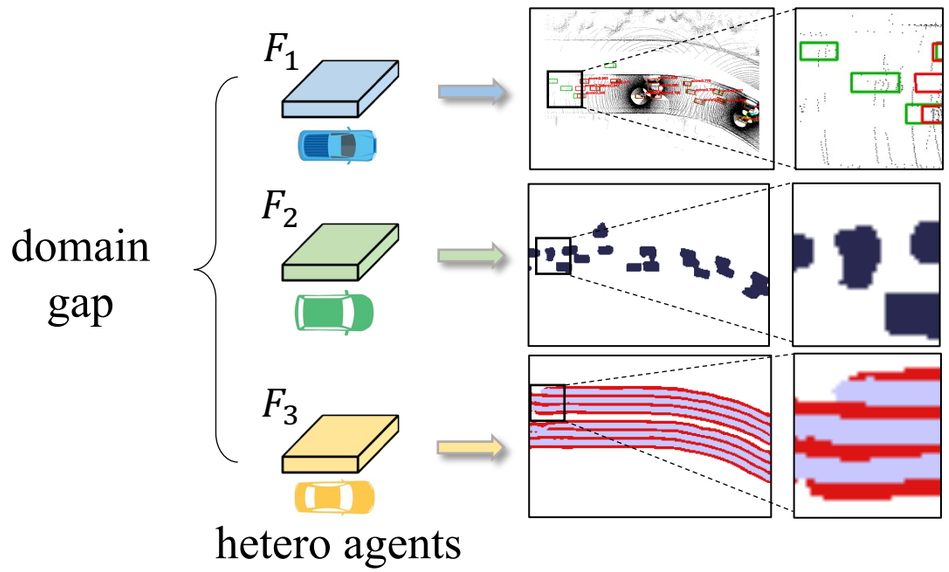
Step 1: Training Protocol Network



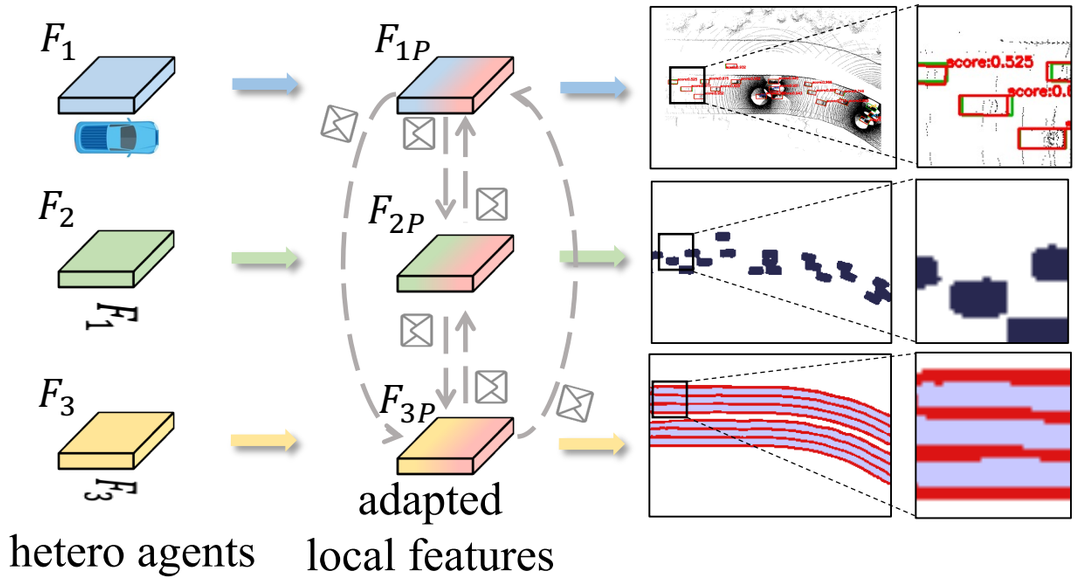
Step 2: Training Local Adapters and Reverters



(I) Before **CFA** (Non-collaborative)

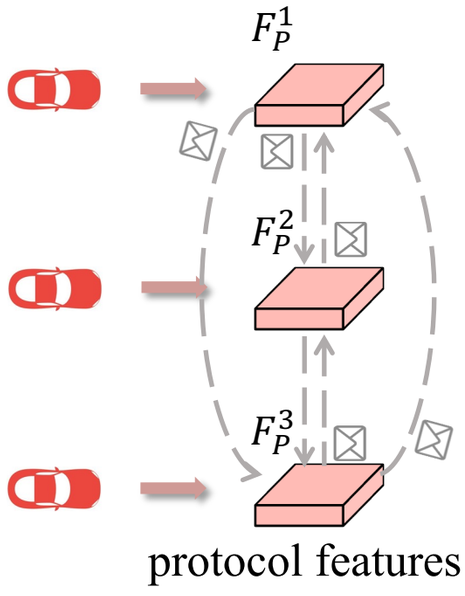


(III) After **CFA** (heterogeneously collaborative)

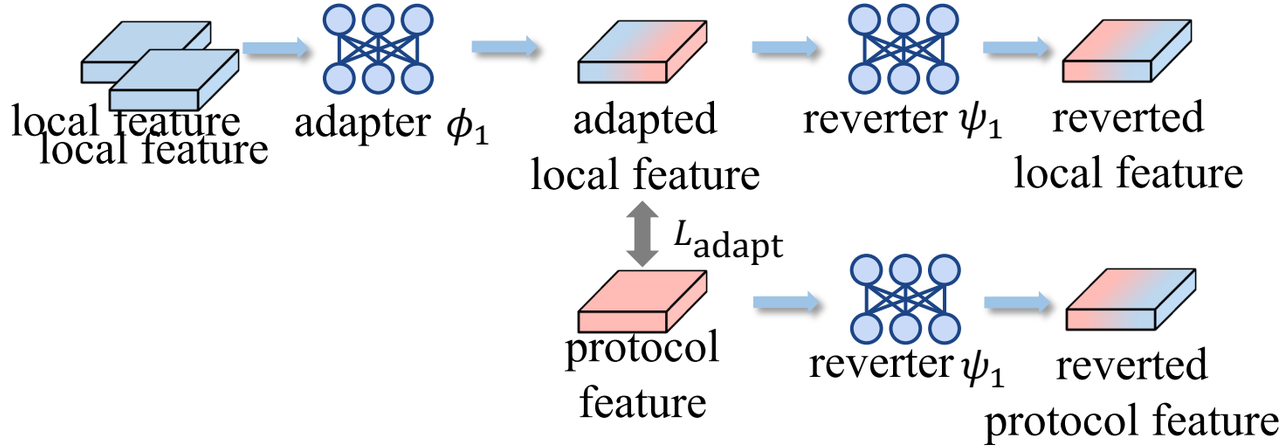


(II) **Collaborative Feature Alignment (CFA)**

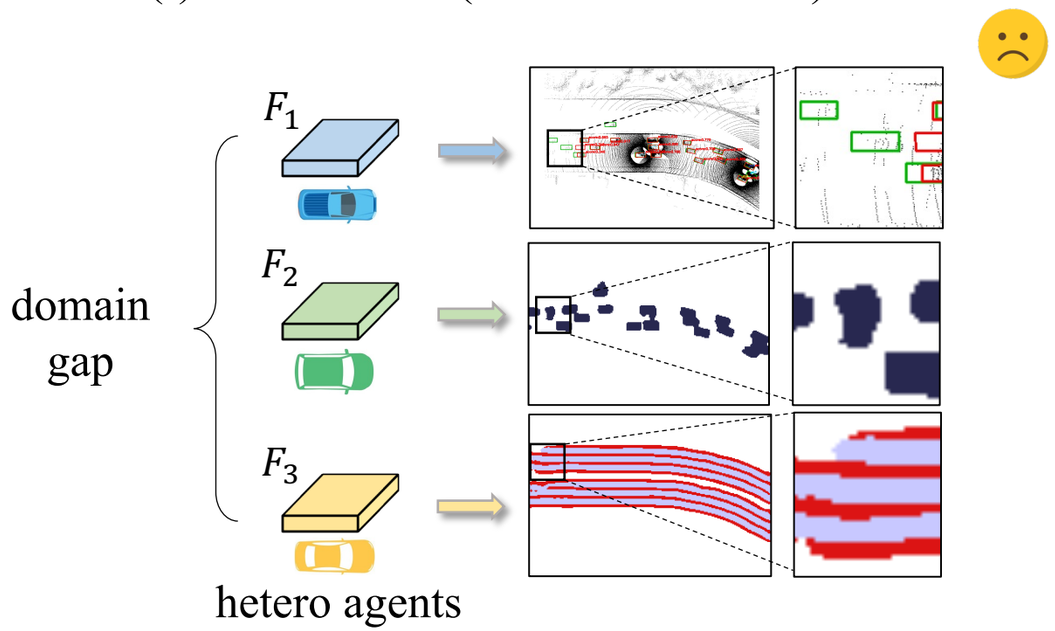
Step 1: Training Protocol Network



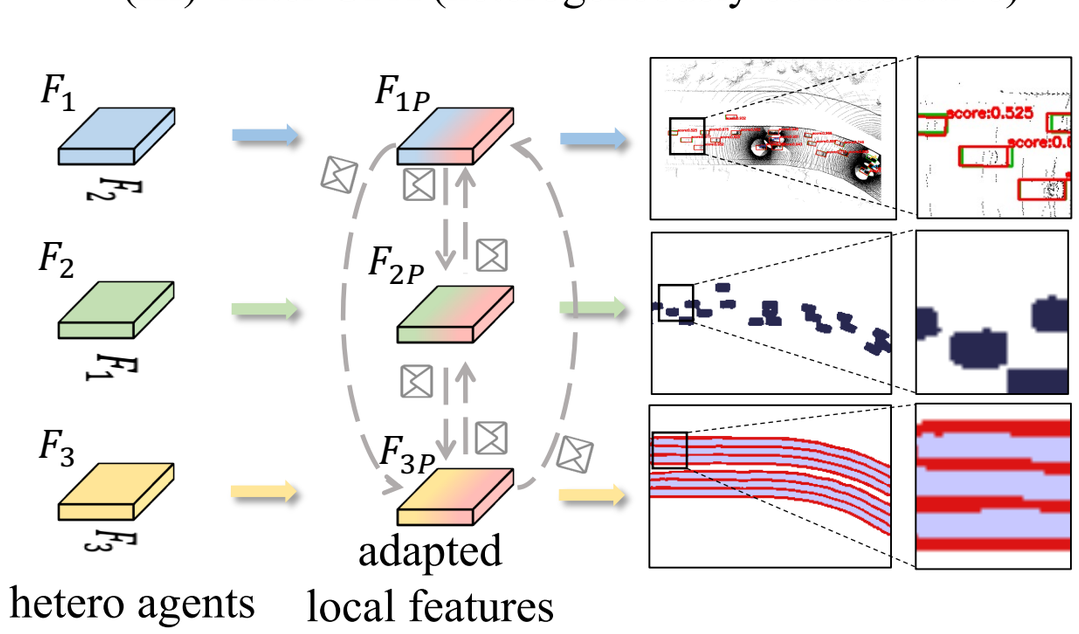
Step 2: Training Local Adapters and Reverters



(I) Before **CFA** (Non-collaborative)

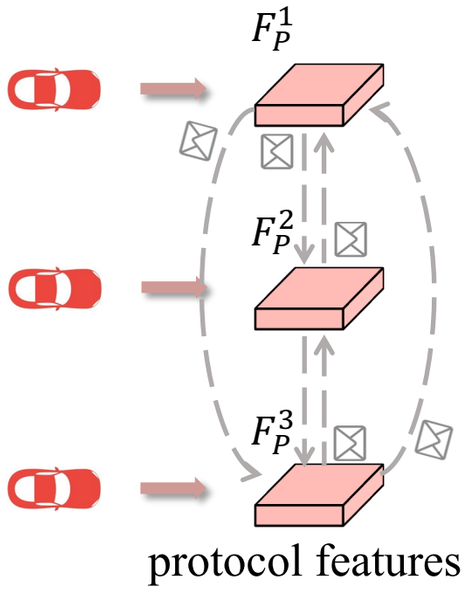


(III) After **CFA** (heterogeneously collaborative)

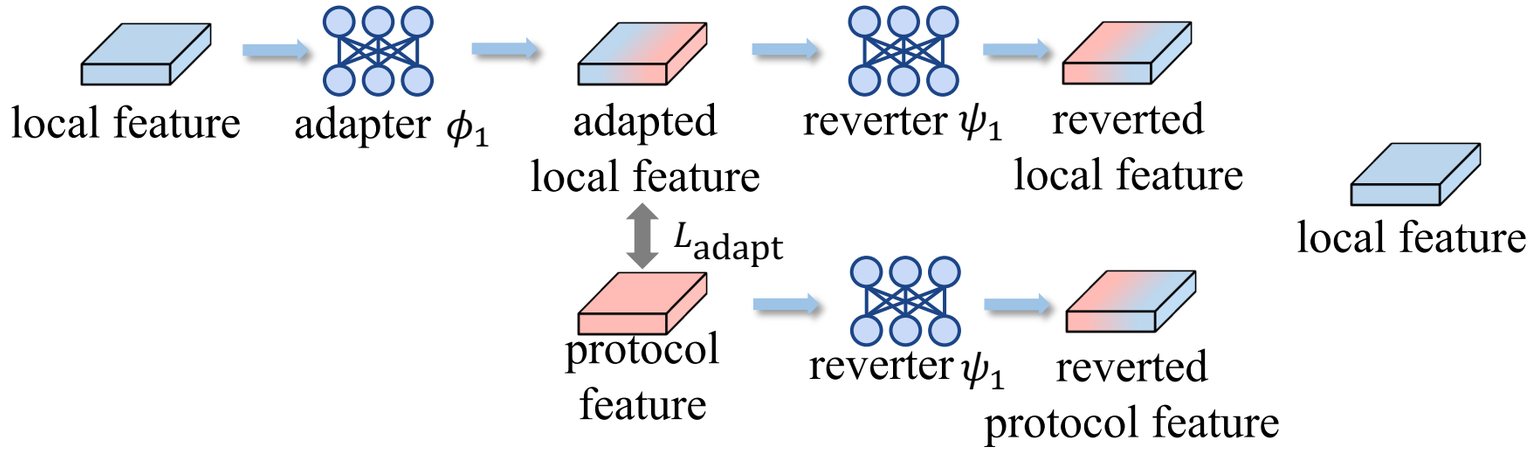


(II) **Collaborative Feature Alignment (CFA)**

Step 1: Training Protocol Network



Step 2: Training Local Adapters and Reverters



(I) Before **CFA** (Non-collaborative)

domain gap

F_1

F_2

F_3

hetero agents

(III) After **CFA** (heterogeneously collaborative)

F_1 F_2 F_3

hetero agents

F_{1P} F_{2P} F_{3P}

adapted local features

(II) **Collaborative Feature Alignment (CFA)**

Step 1: Training Protocol Network

F_P^1 F_P^2 F_P^3

protocol features

Step 2: Training Local Adapters and Reverters

local feature → adapter ϕ_1 → adapted local feature → reverter ψ_1 → reverted local feature

L_{revert}

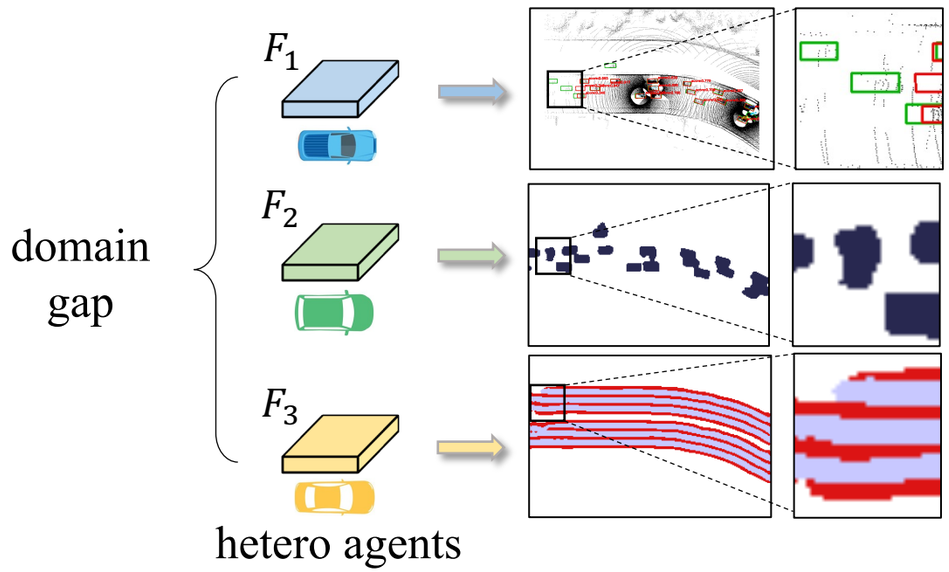
L_{adapt}

local feature

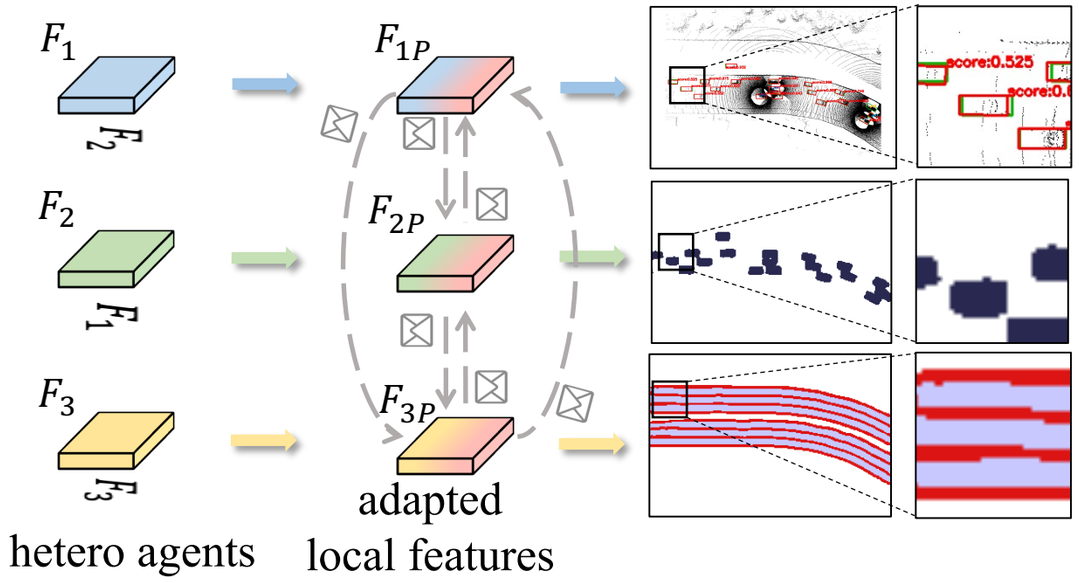
L_{revert}

protocol feature → reverter ψ_1 → reverted protocol feature

(I) Before **CFA** (Non-collaborative)

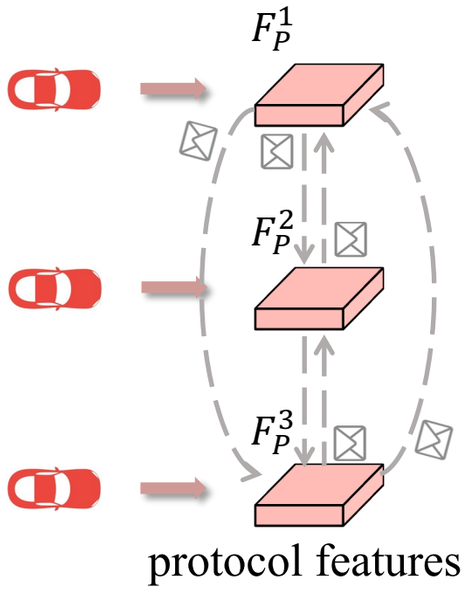


(III) After **CFA** (heterogeneously collaborative)

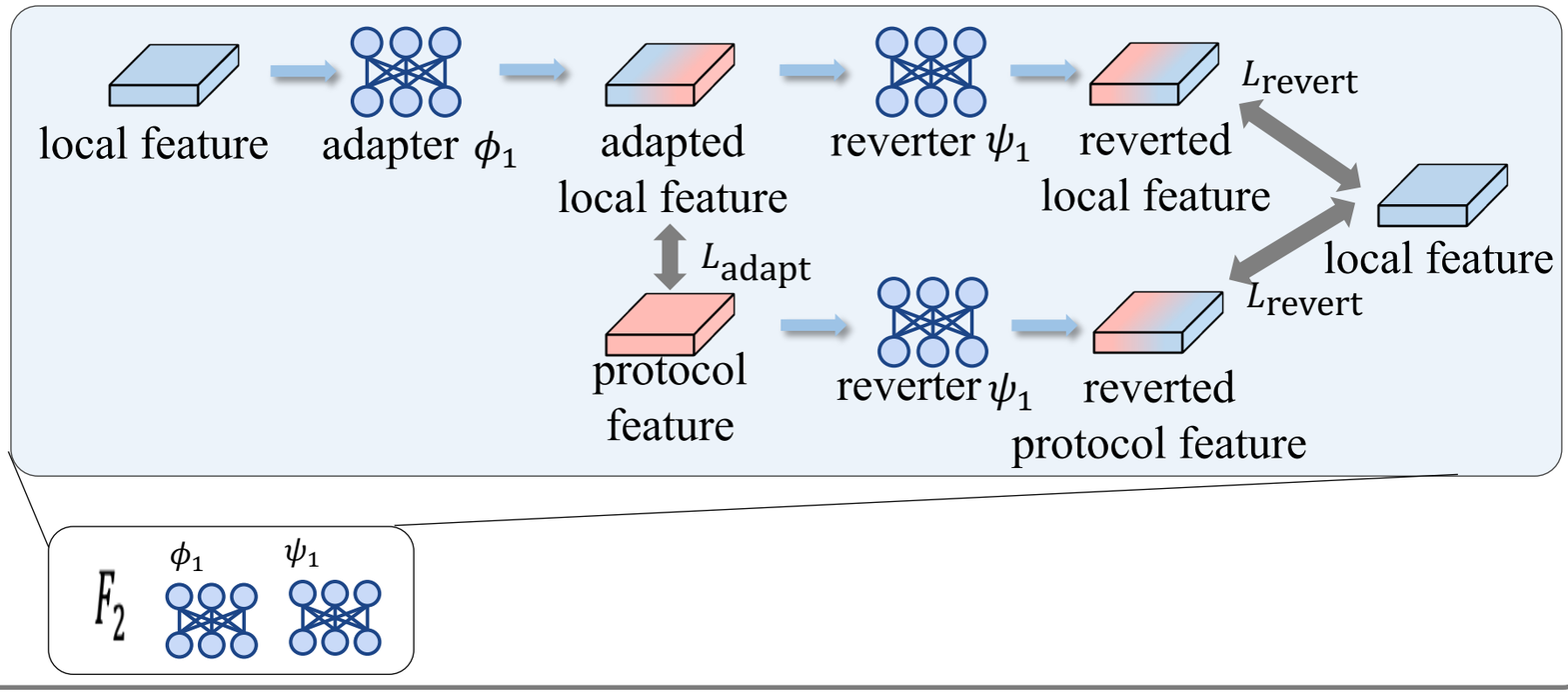


(II) **Collaborative Feature Alignment (CFA)**

Step 1: Training Protocol Network



Step 2: Training Local Adapters and Reverters



(I) Before **CFA** (Non-collaborative)

domain gap

F_1

F_2

F_3

hetero agents

(III) After **CFA** (heterogeneously collaborative)

F_1 F_2 F_3

hetero agents

F_{1P} F_{2P} F_{3P}

adapted local features

score:0.525

(II) Collaborative Feature Alignment (CFA)

Step 1: Training Protocol Network

F_P^1 F_P^2 F_P^3

protocol features

Step 2: Training Local Adapters and Reverters

local feature → adapter ϕ_1 → adapted local feature → reverter ψ_1 → reverted local feature

L_{adapt}

protocol feature → reverter ψ_1 → reverted protocol feature

L_{revert} L_{revert}

local feature

F_2 ϕ_1 ψ_1

F_1 ϕ_2 ψ_2

F_3 ϕ_3 ψ_3

(I) Before **CFA** (Non-collaborative)

domain gap

F_1

F_2

F_3

hetero agents

(III) After **CFA** (heterogeneously collaborative)

F_1 F_2 F_3

hetero agents

F_{1P} F_{2P} F_{3P}

adapted local features

(II) **Collaborative Feature Alignment (CFA)**

Step 1: Training Protocol Network

F_P^1 F_P^2 F_P^3

protocol features

Step 2: Training Local Adapters and Reverters

local feature → adapter ϕ_1 → adapted local feature → reverter ψ_1 → reverted local feature

L_{revert}

local feature

L_{adapt}

L_{revert}

Off-line Training

protocol feature → reverter ψ_1 → reverted protocol feature

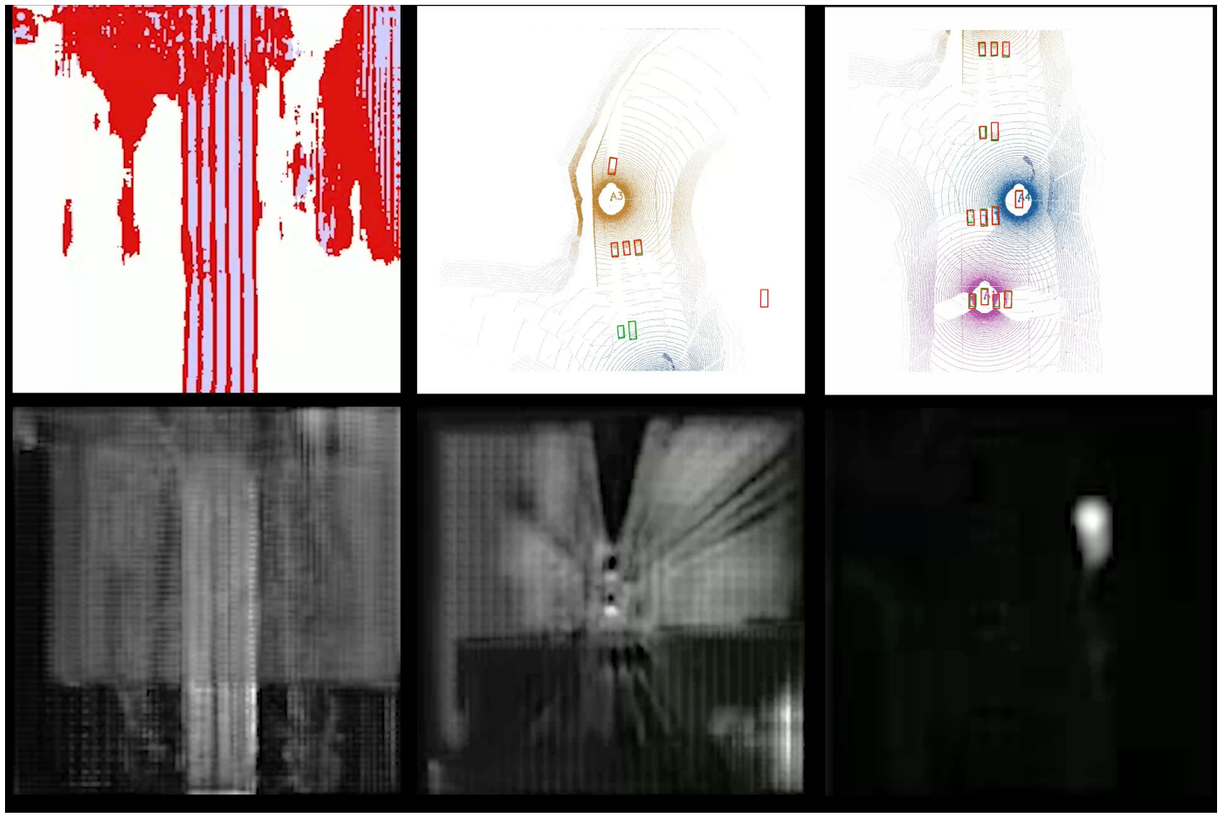
F_2 ϕ_1 ψ_1

F_1 ϕ_2 ψ_2

F_3 ϕ_3 ψ_3

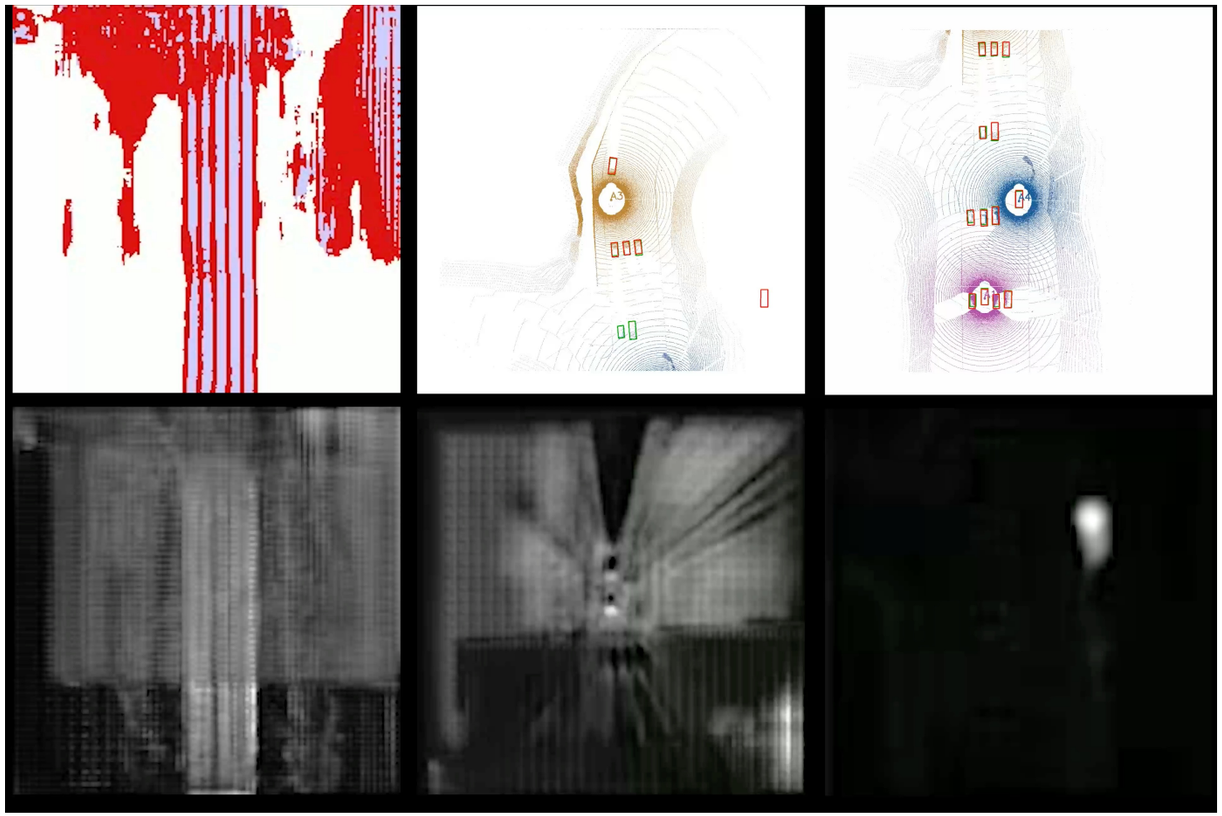
Visualization

Before CFA

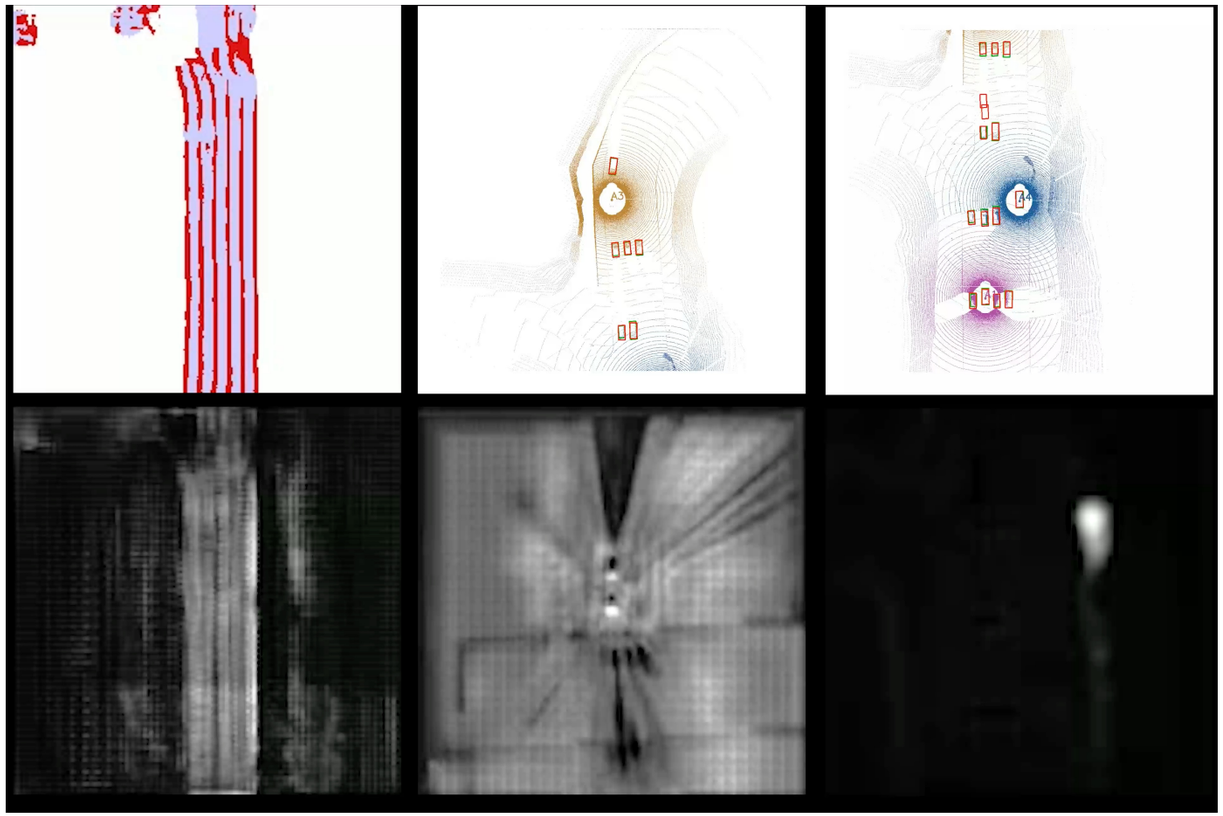


Visualization

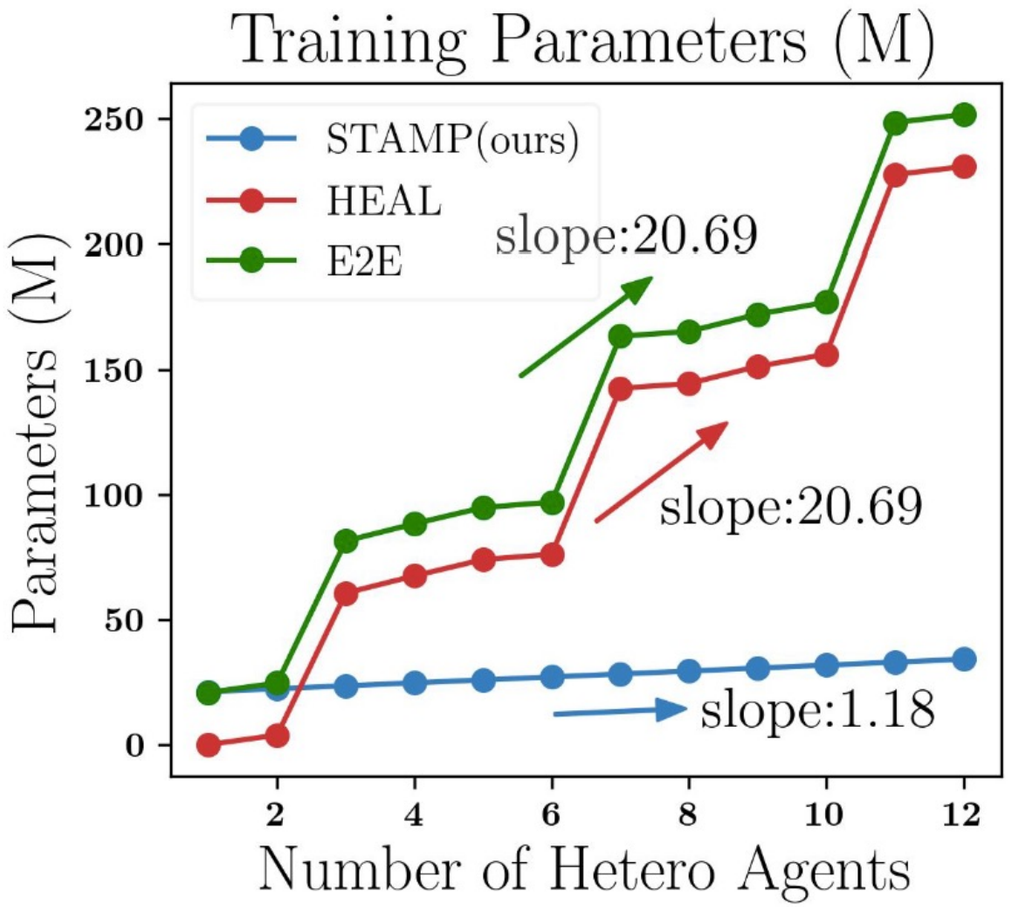
Before CFA

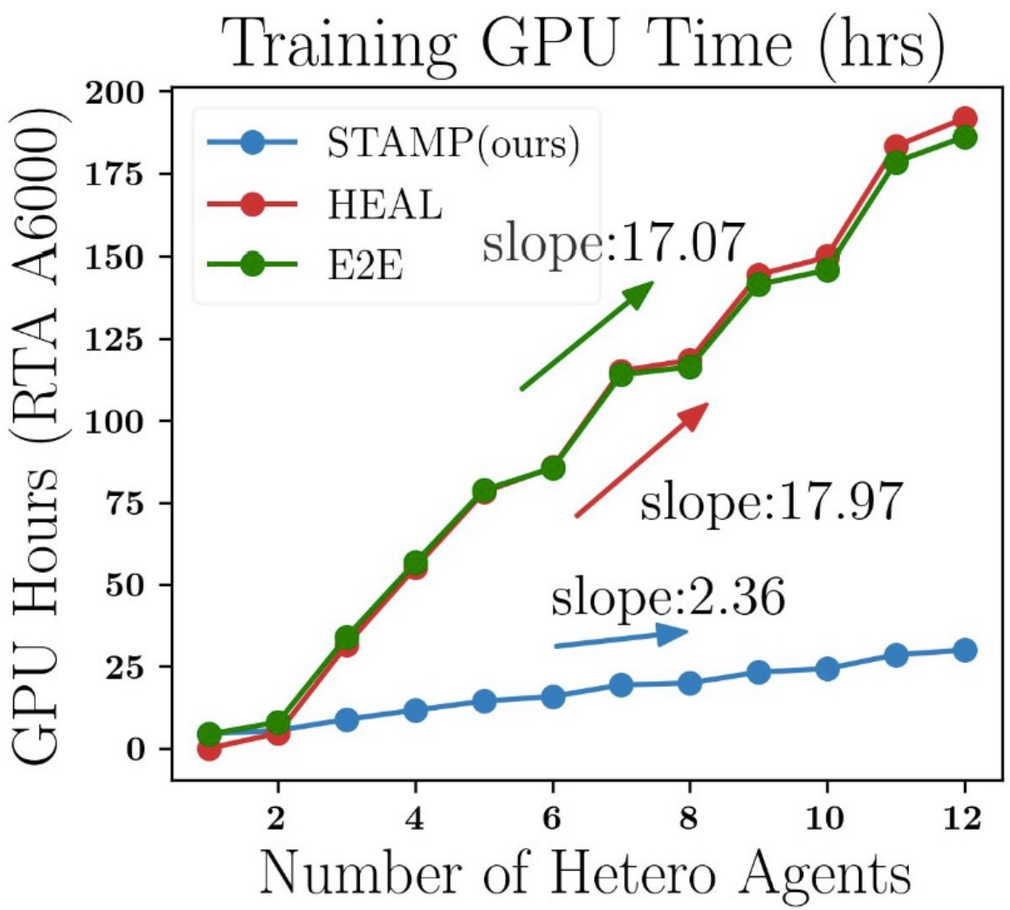


After CFA



Efficiency


Training Parameters (M)
STAMP(ours)
HEAL
E2E
slope:20.69
slope:20.69
slope:1.18
Parameters (M)
250
200
150
100
50
0
2
4
6
8
10
12
Number of Hetero Agents


Training GPU Time (hrs)
STAMP(ours)
HEAL
E2E
slope:17.07
slope:17.97
slope:2.36
GPU Hours (RTA A6000)
200
175
150
125
100
75
50
25
0
2
4
6
8
10
12
Number of Hetero Agents