上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

Communication-Efficient Collaborative Perception

Yue Hu (胡悦)

饮 水 思 源 · 爱 国 荣 校

01

Introduction & Background
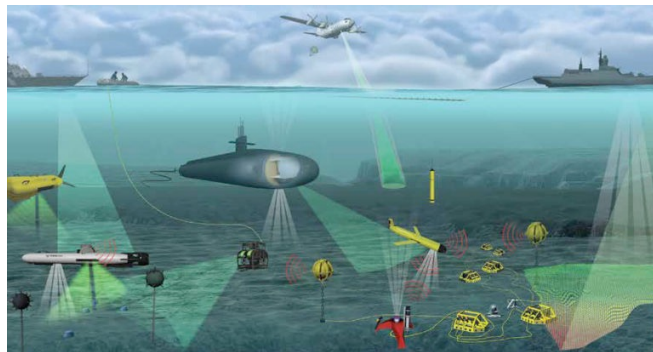
Introduction & Background

➤ **Autonomous system has lots of real-world applications.**
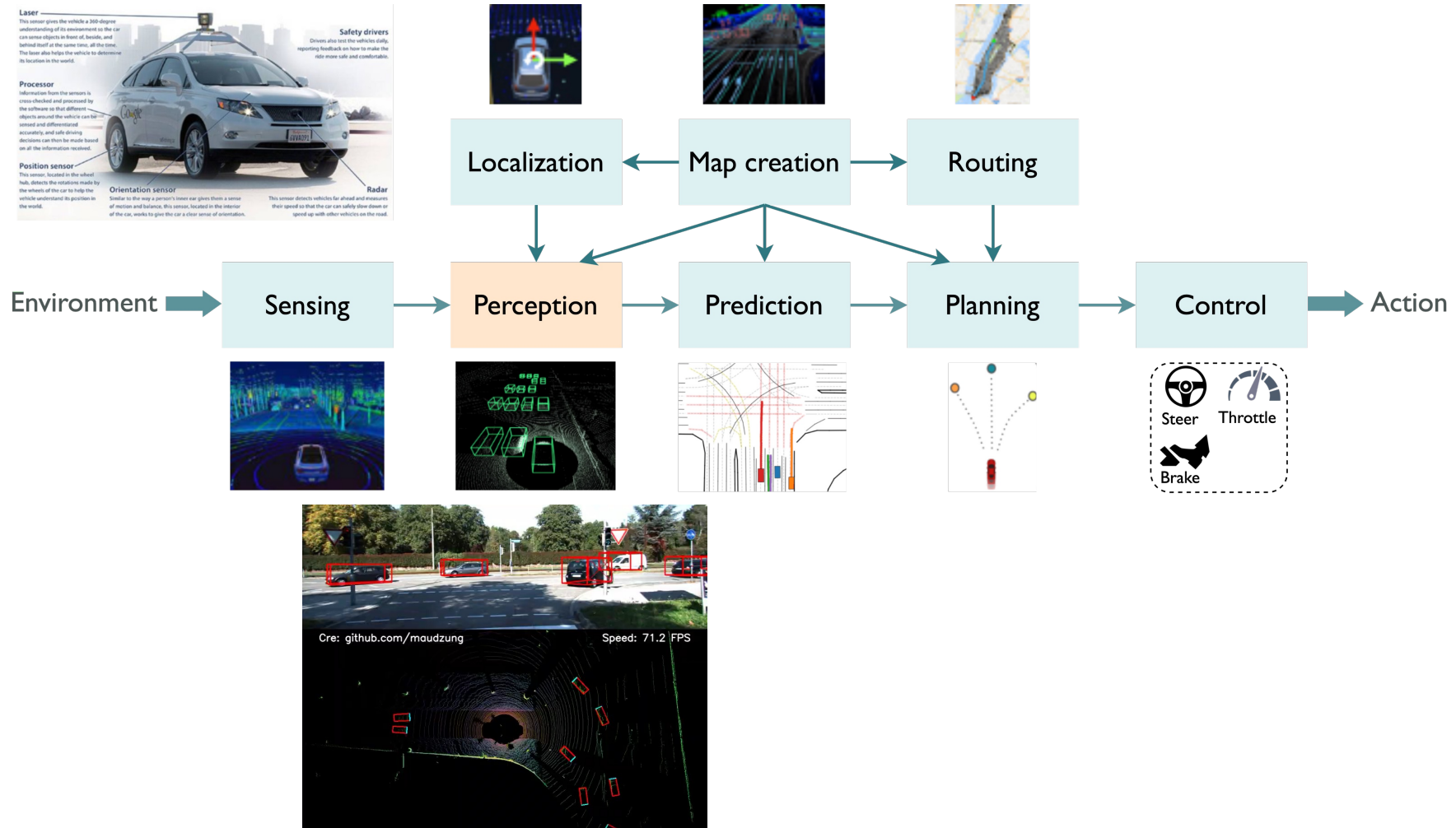

autonomous vehicles

autonomous aerial vehicles

autonomous underwater vehicles

autonomous robotics

Introduction & Background

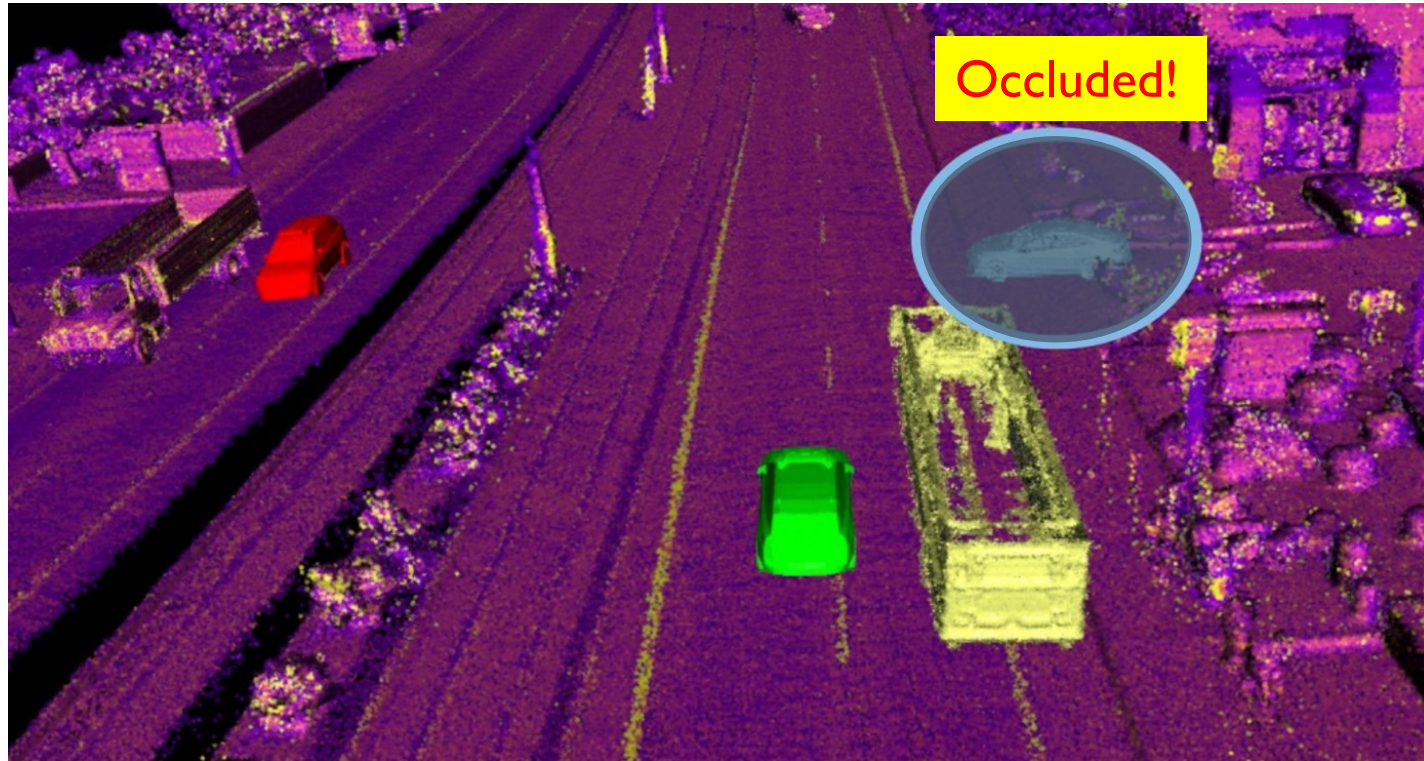➤ **Perception is the front stage of the autonomous system.**

| Localization | ← | Map creation | → | Routing |

| Environment | → | Sensing | → | Perception | → | Prediction | → | Planning | → | Control | → | Action |

Cre: github.com/maudzung   Speed: 71.2 FPS

Introduction & Background
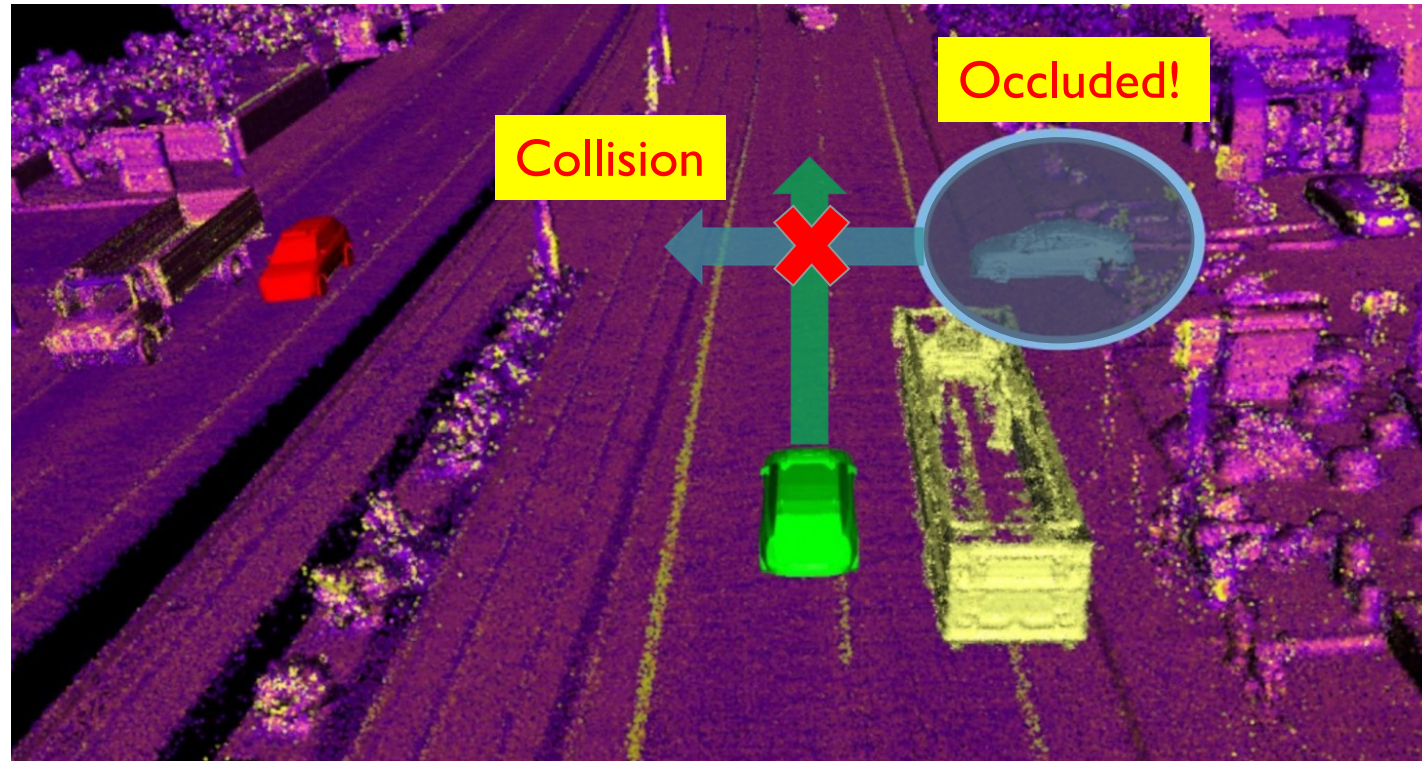
➤ **Single-agent perception**

Single agent *encounters limited visibility* *(occlusion)* *and long-range issues*

Introduction & Background

➢ **Single-agent perception**

Single agent *encounters limited visibility (occlusion) and long-range issues*
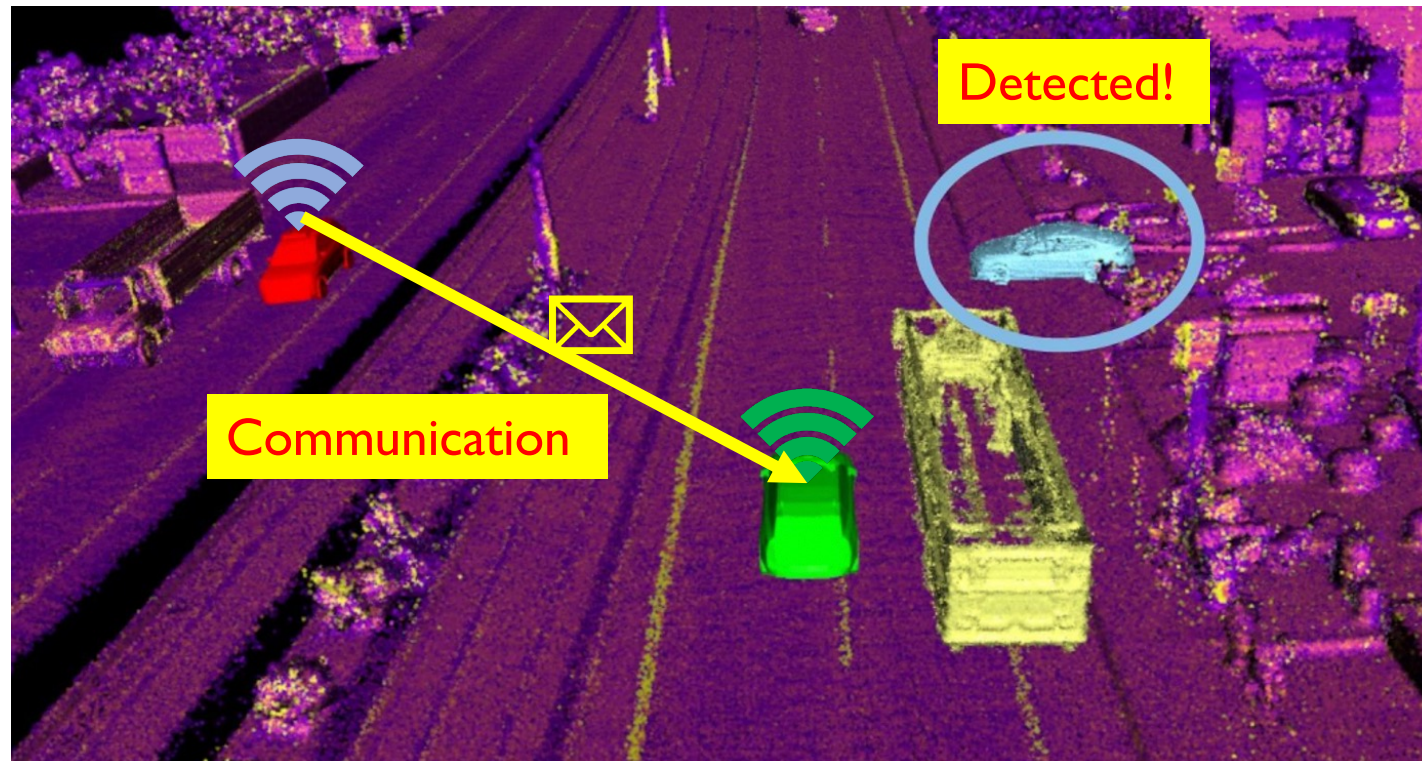
Introduction & Background
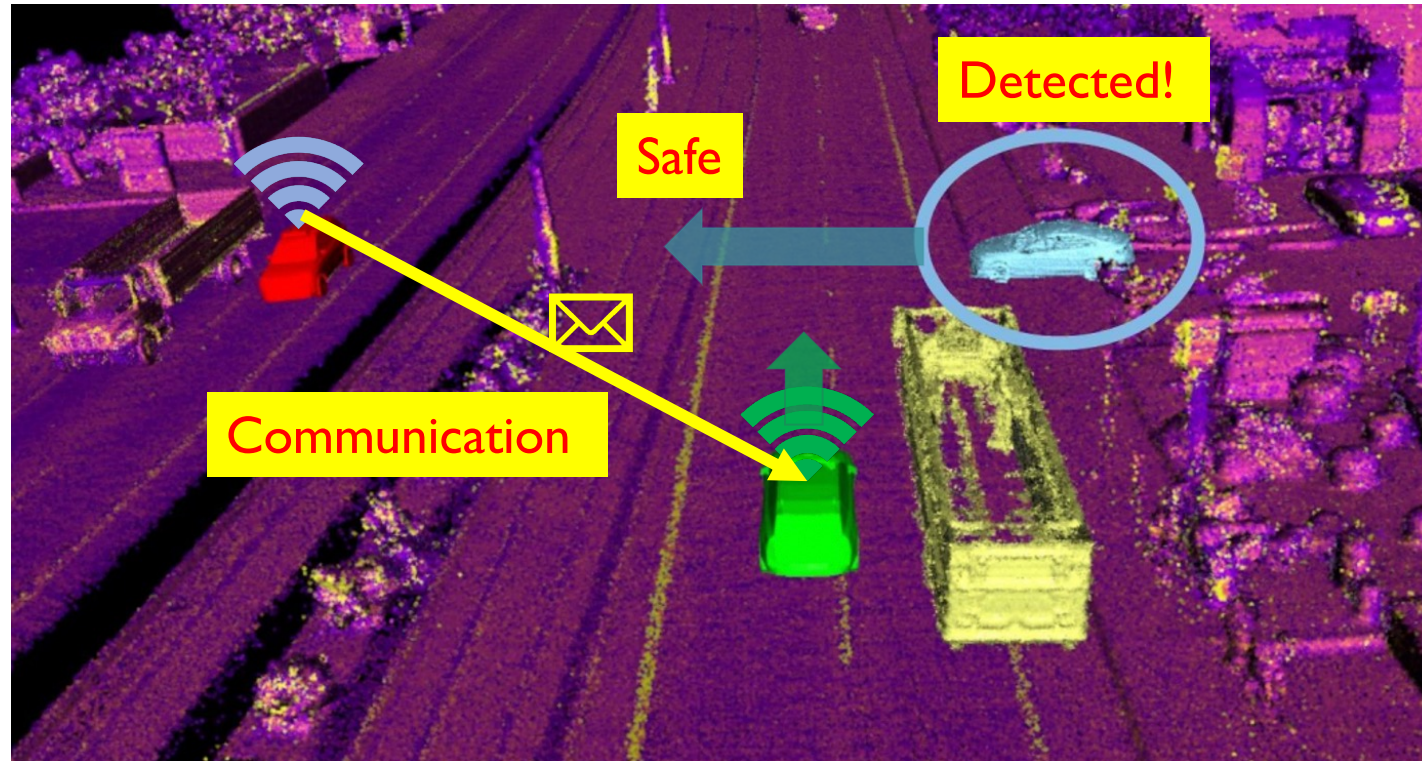
➢ **Multi-agent collaborative perception**

Collaborative agent *fundamentally overcomes single-agent limitations, achieving* *holistic view!*

Introduction & Background

➤ **Multi-agent collaborative perception**
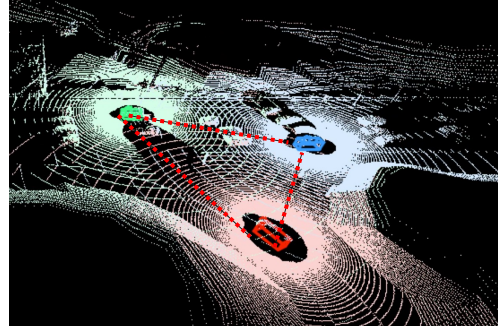
Collaborative agent fundamentally overcomes single-agent limitations, achieving holistic view!

Introduction & Background

➢ **Multi-agent collaborative perception**

Collaboration Perception
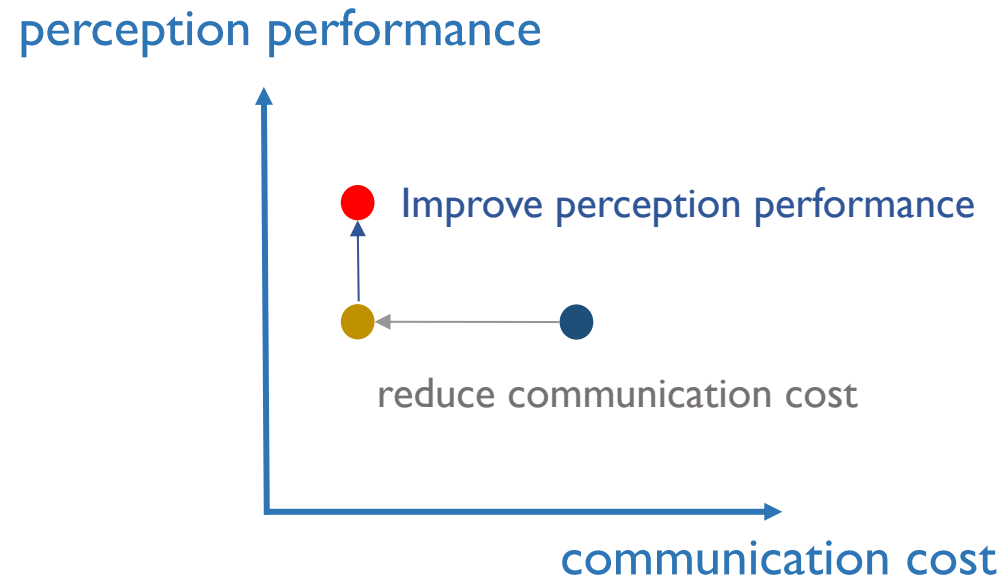
Single-Agent
Perception

Communication

Limited

Introduction & Background

➢ **Multi-agent collaborative perception**

Key challenge: Trade-off between communication cost and perception performance
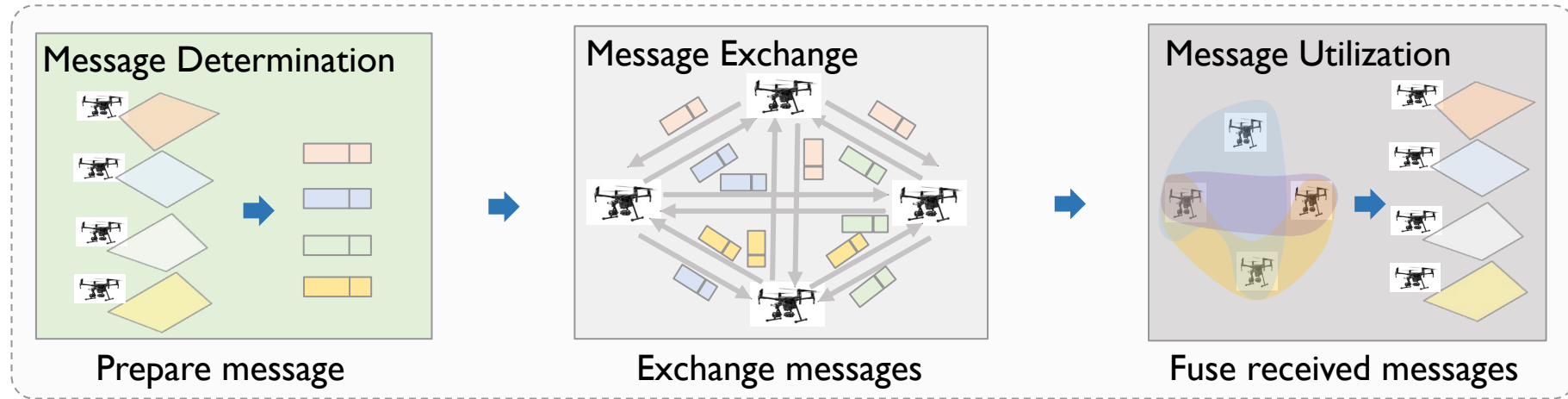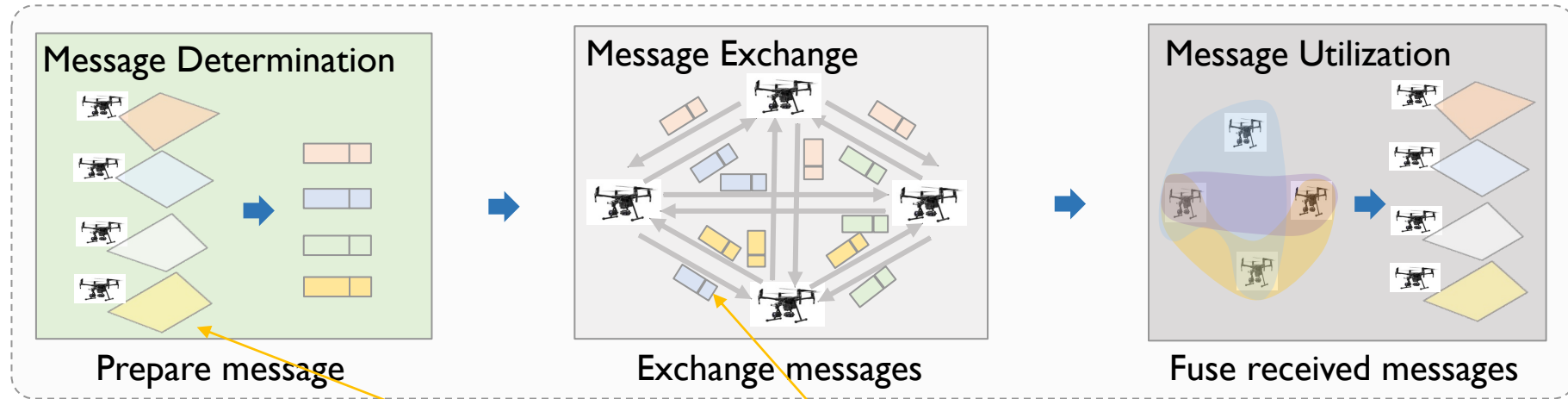
02

Collaborative Perception Formulation

Collaborative Perception Formulation

➢ Collaborative perception: improve each agent's perception capacity through information exchange

Collaborative Perception Formulation

➤ Collaborative perception: improve each agent's perception capacity through information exchange

➤ Formulation

$$\arg\max_{\theta,\mathcal{P}} \sum_{i=1}^{N} g\left(\Phi_\theta\left(\{\mathcal{X}_i^{(t)}\}_{t=1}^{T}, \{\mathcal{P}_{j\to i}^{(t)}\}_{j=1,t=1}^{N,T}\right), \{\widehat{\mathcal{T}}_i^{(t)}\}_{t=1}^{T}\right),$$

Collaborative Perception Formulation

➢ Collaborative perception: improve each agent's perception capacity through information exchange

Message Determination

Prepare message

Message Exchange

Exchange messages

Message Utilization

Fuse received messages

Observation

Messages

➢ Formulation

$$\arg\max_{\theta,\mathcal{P}} \sum_{i=1}^{N} g\left(\Phi_\theta\left(\{\mathcal{X}_i^{(t)}\}_{t=1}^T, \{\mathcal{P}_{j\rightarrow i}^{(t)}\}_{j=1,t=1}^{N,T}\right), \{\widehat{\mathcal{T}}_i^{(t)}\}_{t=1}^T\right),$$

Perceptual output

Perceptual targets

Perception network

Collaborative Perception Formulation

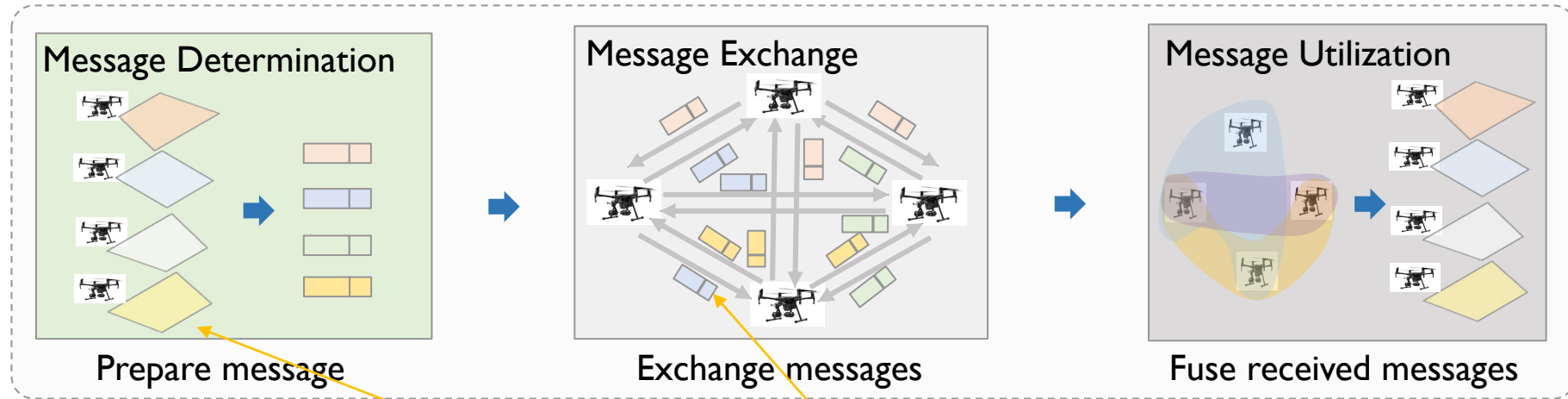➢ Collaborative perception: improve each agent's perception capacity through information exchange

Message Determination
Prepare message

Message Exchange
Exchange messages

Message Utilization
Fuse received messages

Observation

Messages

➢ Formulation

$$\arg\max_{\theta, \mathcal{P}} \sum_{i=1}^{N} g\left(\Phi_\theta\left(\{\mathcal{X}_i^{(t)}\}_{t=1}^{T}, \{\mathcal{P}_{j\to i}^{(t)}\}_{j=1,t=1}^{N,T}\right), \{\widehat{\mathcal{T}}_i^{(t)}\}_{t=1}^{T}\right)$$
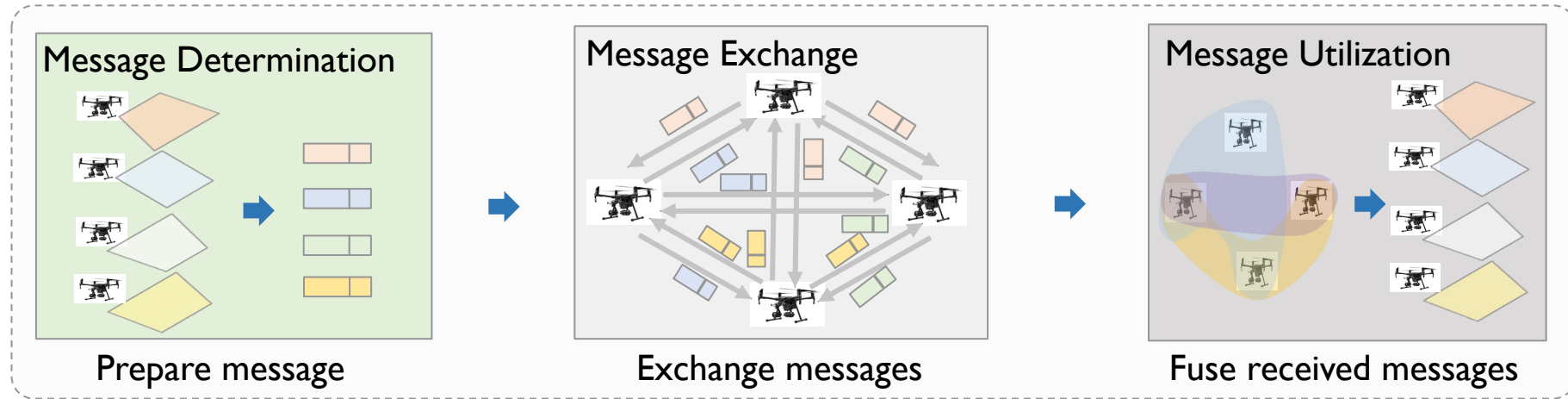
Perception
performance

Perception
evaluation metric

Perception
network

Collaborative Perception Formulation

➤ Collaborative perception: improve each agent's perception capacity through information exchange

Message Determination	Message Exchange	Message Utilization
Prepare message	Exchange messages	Fuse received messages

➤ Formulation

Observation   Messages   Perceptual targets
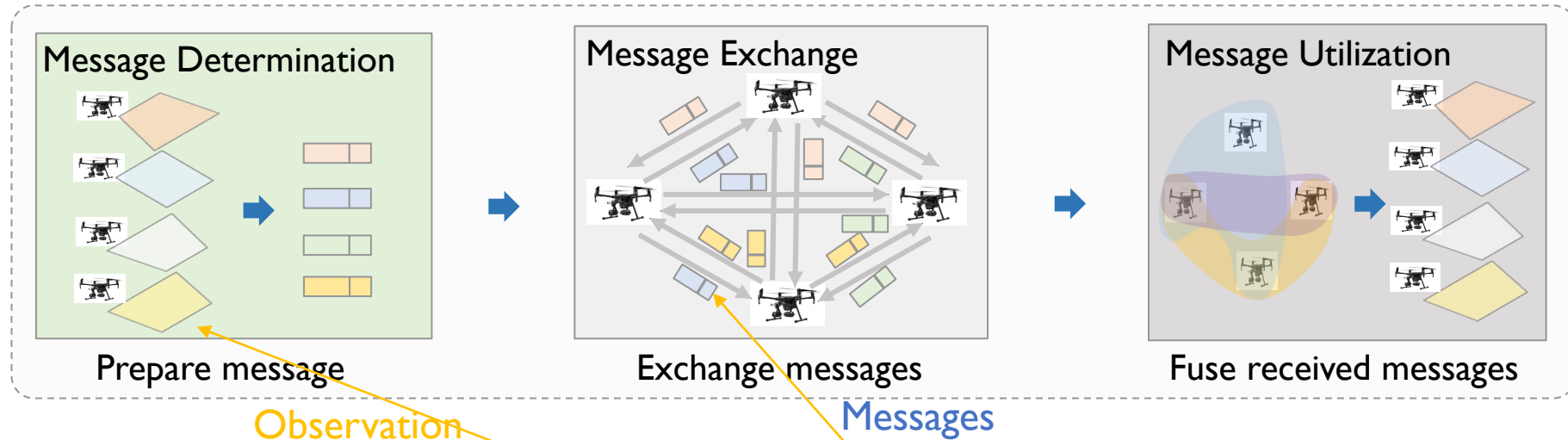
$$\arg\max_{\theta, \mathcal{P}} \sum_{i=1}^{N} g\left(\Phi_\theta\left(\{\mathcal{X}_i^{(t)}\}_{t=1}^T, \{\mathcal{P}_{j\to i}^{(t)}\}_{j=1,t=1}^{N,T}\right), \{\widehat{\mathcal{T}}_i^{(t)}\}_{t=1}^T\right)$$

Perception performance

Optimize the message to
maxmize perception performance

Collaborative Perception Formulation

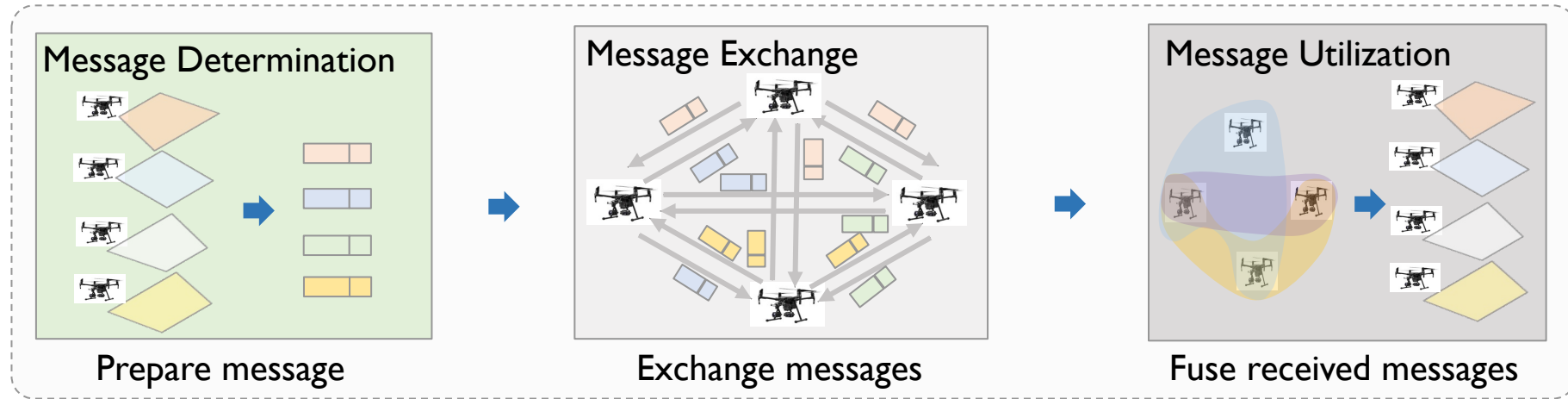➢ Collaborative perception: improve each agent's perception capacity through information exchange

Message Determination	Message Exchange	Message Utilization
Prepare message	Exchange messages	Fuse received messages

➢ Formulation

Observation   Messages   Perceptual targets

$$\arg\max_{\theta,\mathcal{P}} \sum_{i=1}^{N} g\left(\Phi_\theta\left(\{\mathcal{X}_i^{(t)}\}_{t=1}^{T}, \{\mathcal{P}_{j\to i}^{(t)}\}_{j=1,t=1}^{N,T}\right), \{\widehat{\mathcal{T}}_i^{(t)}\}_{t=1}^{T}\right)$$

Perception performance

Optimize the message to maxmize perception performance with communication budget constraint

$$\text{s.t.} \sum_{t=1}^{T} \sum_{i,j=1,i\neq j}^{N} b\left(\mathcal{P}_{j\to i}^{(t)}\right) \leq B,$$

Communication cost   Communication budget

Communication constraint

Collaborative Perception Formulation

➢ **Task-specific communication objective:** optimizing messages to fill each agent's *specific perception task demand* within the communication budget

Observation Messages Perceptual targets

Perception performance

$$\arg\max_{\theta,\mathcal{P}} \sum_{i=1}^{N} g\left(\Phi_{\theta}\left(\{\mathcal{X}_i^{(t)}\}_{t=1}^{T}, \{\mathcal{P}_{j\to i}^{(t)}\}_{j=1,t=1}^{N,T}\right), \{\widehat{\mathcal{T}}_i^{(t)}\}_{t=1}^{T}\right)$$

Perception evaluation metric Perception network

Communication constraint

$$\text{s.t.} \sum_{t=1}^{T} \sum_{i,j=1,i\neq j}^{N} b\left(\mathcal{P}_{j\to i}^{(t)}\right) \leq B,$$

Communication cost

Challenges:
1) Non-differentiable perception metric and hard bandwidth constraint → unable to directly optimize
2) Diverse input and numerous perception network parameters → unable to enumerate the large searching space

Collaborative Perception Formulation

➤ **Task-specific communication objective:** optimizing messages to fill each agent's *specific perception task demand* within the communication budget

Observation   Messages   Perceptual targets

$$\arg\max_{\theta, \mathcal{P}} \sum_{i=1}^{N} g\left(\Phi_{\theta}\left(\{\mathcal{X}_i^{(t)}\}_{t=1}^{T}, \{\mathcal{P}_{j\to i}^{(t)}\}_{j=1,t=1}^{N,T}\right), \{\widehat{\mathcal{T}}_i^{(t)}\}_{t=1}^{T}\right)$$

Perception performance

Perception evaluation metric   Perception network

$$\text{s.t.} \sum_{t=1}^{T} \sum_{i,j=1,i\neq j}^{N} b(\mathcal{P}_{j\to i}^{(t)}) \leq B,$$

Communication constraint

Communication cost

Solution: pragmatic communication

Key idea: acknowledge the pragmatic significance of the data and to develop pragmatic messages that retain only the data necessary for the collaborator's downstream task.

03

Pragmatic Communication Solution

[1] Where2comm: Communication-Efficient Collaborative Perception via Spatial Confidence Maps, NeurIPS, 2022
[2] Pragmatic Communication in Multi-Agent Collaborative Perception, TPAMI, 2024 Submitted
[3] Communication- Efficient Collaborative Perception via Information Filling with Codebook, CVPR, 2024
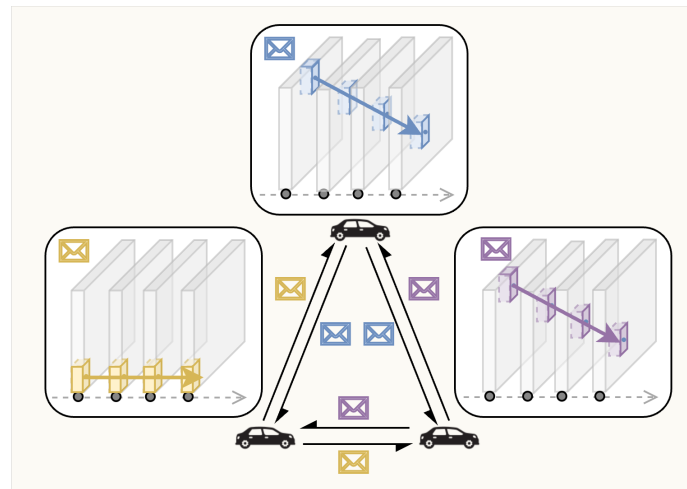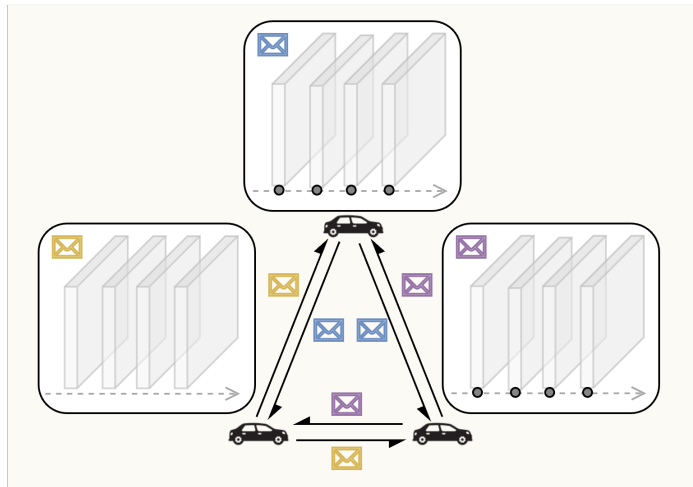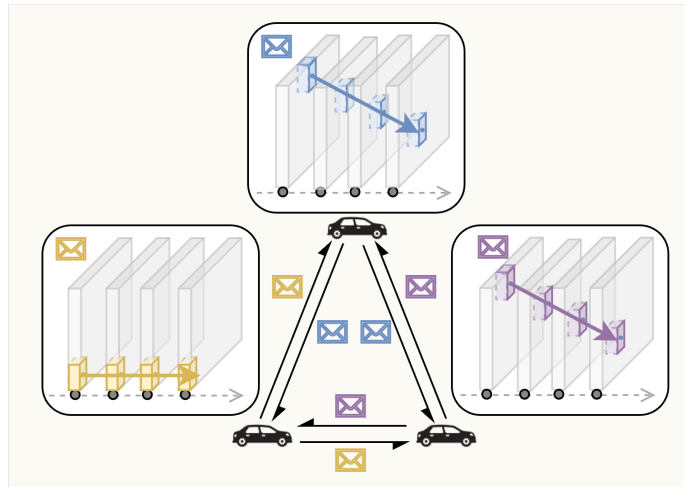
Pragmatic Communication Solution

➤ **Pragmatic communication**

Key idea: acknowledge the pragmatic significance of the data and to develop pragmatic messages that retain only the data necessary for the collaborator's downstream task.

Shannon communication
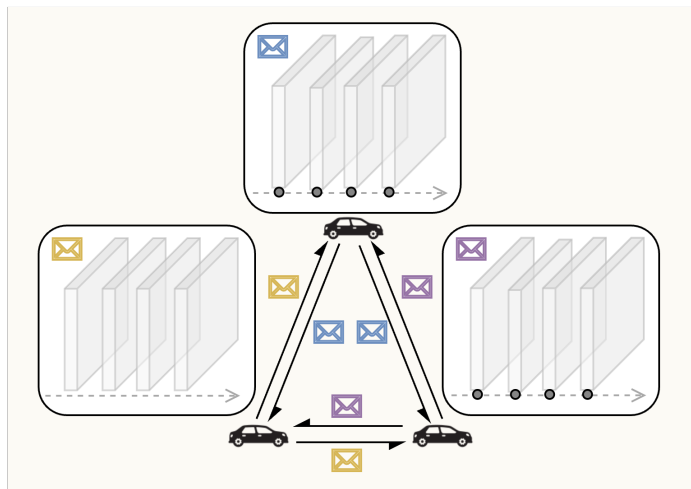
Pragmatic significance

Pragmatic message

Pragmatic Communication Solution

➢ **Pragmatic communication**

Key idea: acknowledge the pragmatic significance of the data and to develop pragmatic messages that retain only the data necessary for the collaborator's downstream task.

Shannon communication

Pragmatic significance

Pragmatic message

What to collaborate?

How to collaborate?

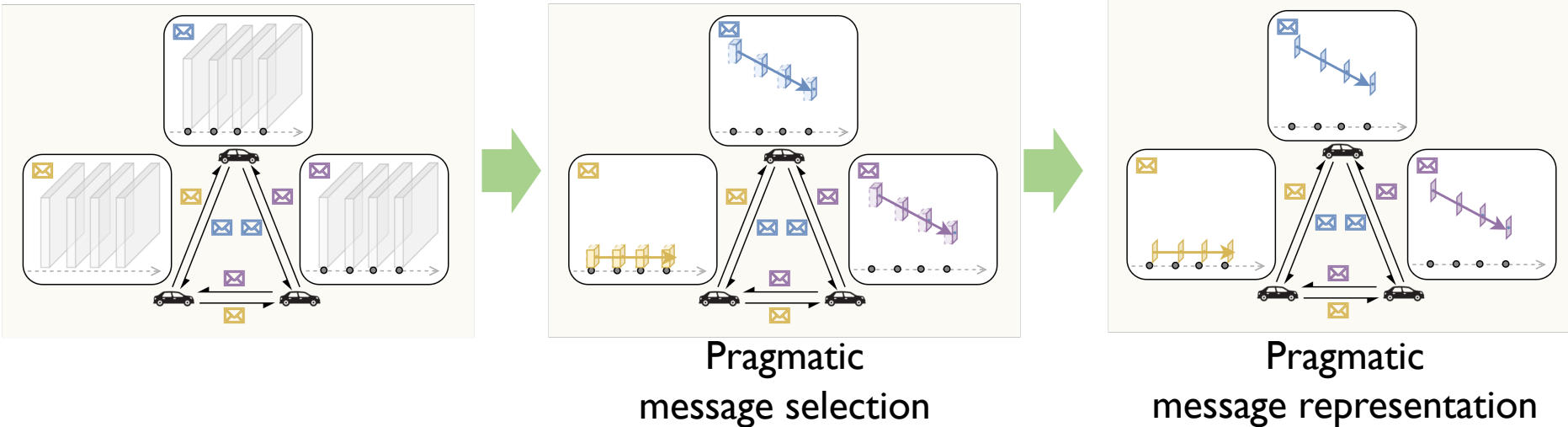Who to collaborate?
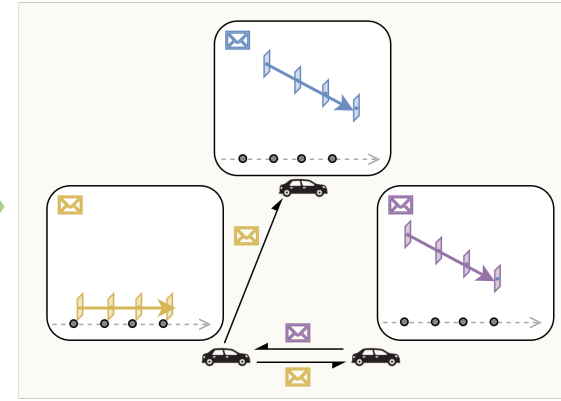
Pragmatic Communication Solution

➤ **Pragmatic message**

$$\mathcal{P} = \Psi_{\text{who}} \left(\Psi_{\text{represent}} \left(\Psi_{\text{select}} (\mathcal{X}) \right) \right)$$

Shannon communication   What to collaborate?
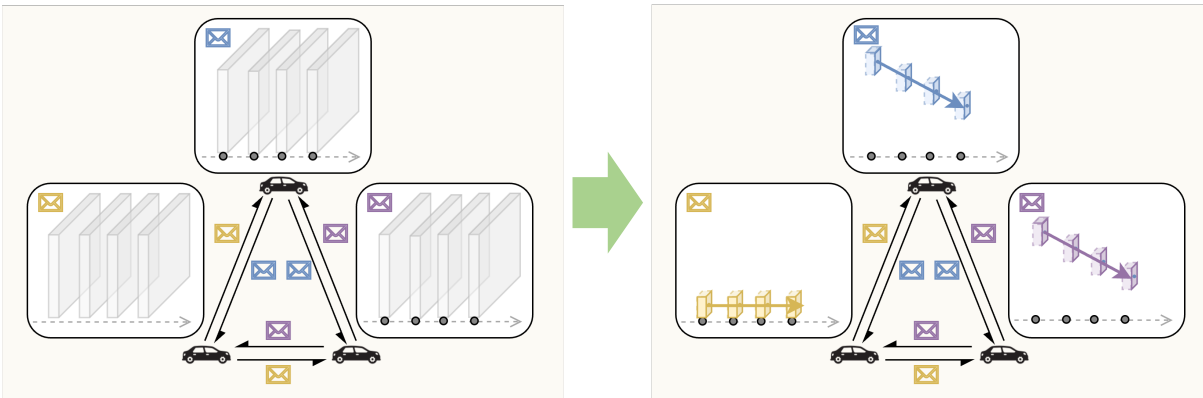
Pragmatic
message selection

➤ **Reformulated objective**

Selection matrix   Pragmatic significance   Communication cost   Analytical solution

$$\max_{\mathbf{M}} \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} \mathbf{M}_{i \to j} \odot \mathbf{C}_i, \quad \text{s.t.} \quad \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} |\mathbf{M}_{i \to j}| \leq b_1, \mathbf{M}_{i \to j} \in \{0, 1\}^{H \times W}$$

Pragmatic Communication Solution

➤ **Pragmatic message**

$$\mathcal{P} = \Psi_{\text{who}} \left(\Psi_{\text{represent}} \left(\Psi_{\text{select}} \left(\mathcal{X} \right) \right) \right)$$

Shannon communication   What to collaborate?   How to collaborate?

Pragmatic message selection

Pragmatic message representation

➤ **Reformulated objective**

Pragmatic significance   Pragmatic representation

$$\mathbf{D}^* = \arg \min_{\mathbf{D}} \sum_{\mathcal{F} \in F} \sum_{h,w} \min_{\ell} \left(\Psi(\mathbf{D}_{[\ell]}) + \left\| \mathcal{F}_{[h,w]} - \mathbf{D}_{[\ell]} \right\|_2^2 \right)$$

Original representation

Gradient decent

Pragmatic Communication Solution

➢ **Pragmatic message**

$$\mathcal{P} = \boxed{\Psi_{\text{who}}} \left(\Psi_{\text{represent}} \left(\Psi_{\text{select}} \left(\mathcal{X} \right) \right) \right)$$

Shannon communication   What to collaborate?   How to collaborate?   Who to collaborate?
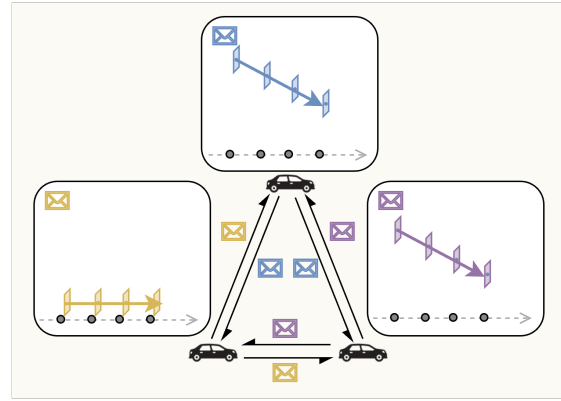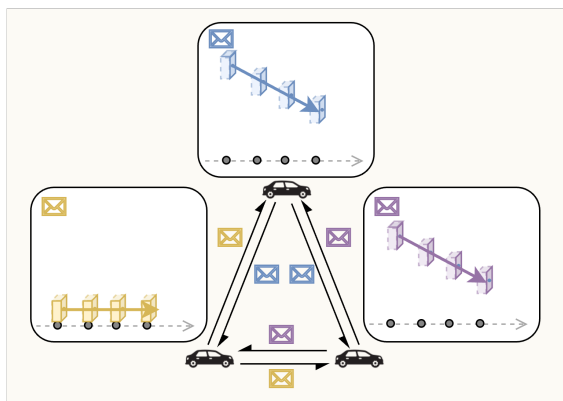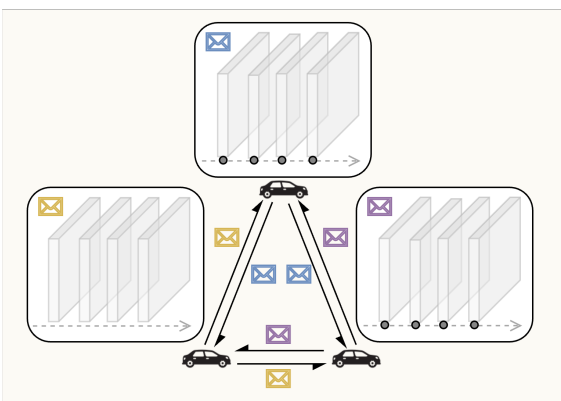
Pragmatic message selection   Pragmatic message representation   Pragmatic collaborator selection

➢ **Reformulated objective**

Selection matrix   Pragmatic significance   Information demand   Communication cost   Analytical solution

$$\{\mathbf{M}^*_{i \to j}\}_{i,j} = \underset{\mathbf{M}}{\arg\max} \sum_{j=1}^{N} f_{\min} \left(\mathbf{C}_j + \sum_{i=1, i \neq j}^{N} \mathbf{M}_{i \to j} \odot \mathbf{C}_i, u \right) \quad \text{where} \quad \sum_{i,j=1, j \neq i}^{N} \mathbf{M}_{i \to j} \leq b, \mathbf{M}_{i \to j} \in \{0,1\}^{H \times W}$$

Pragmatic Communication Solution

➢ **Pragmatic message**
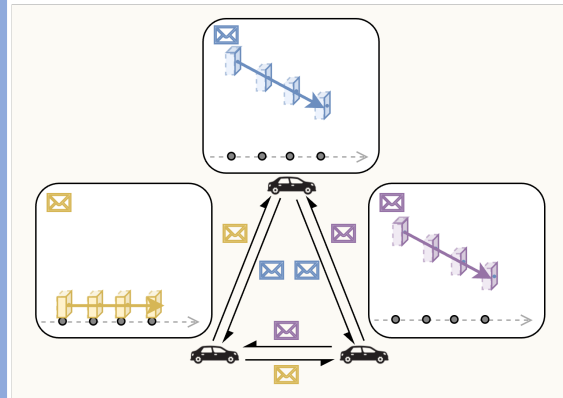
$$\mathcal{P} = \Psi_{\text{who}} \left(\Psi_{\text{represent}} \left(\Psi_{\text{select}} \left(\mathcal{X} \right) \right) \right)$$
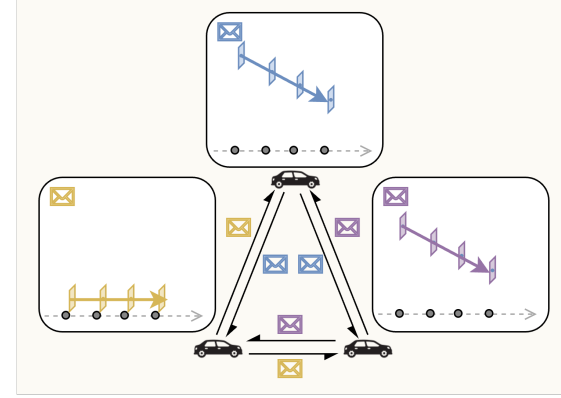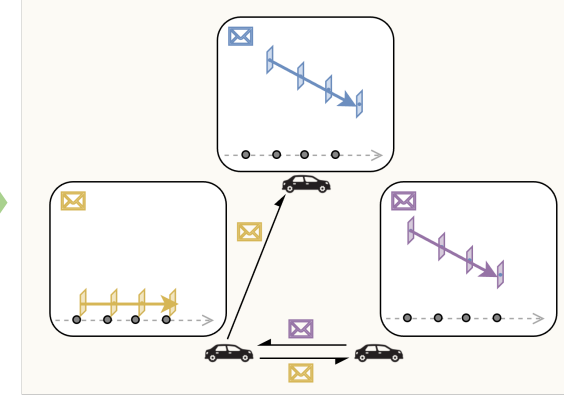

Shannon communication
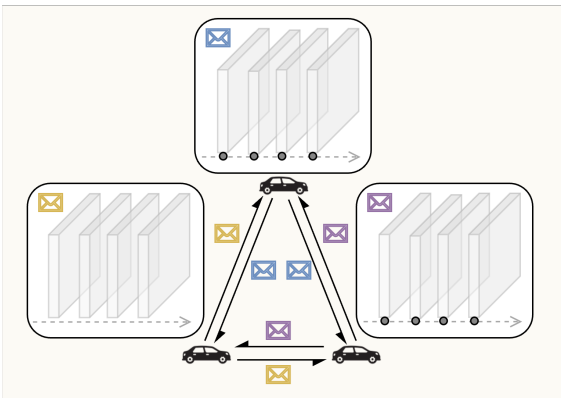
What to collaborate?

Pragmatic
message selection

How to collaborate?

Pragmatic
message representation

Who to collaborate?
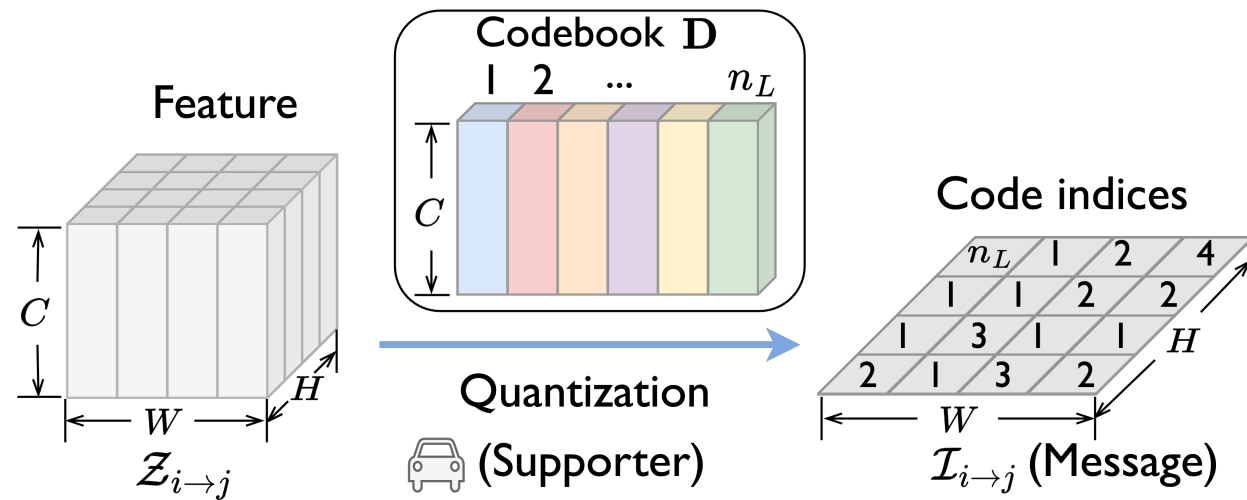
Pragmatic
collaborator selection

Pragmatic communication

Pragmatic messages retain only the data necessary for the collaborator's downstream task.

Pragmatic Message Representation

Core idea: leverage symbol-based communication (analogy to human language)

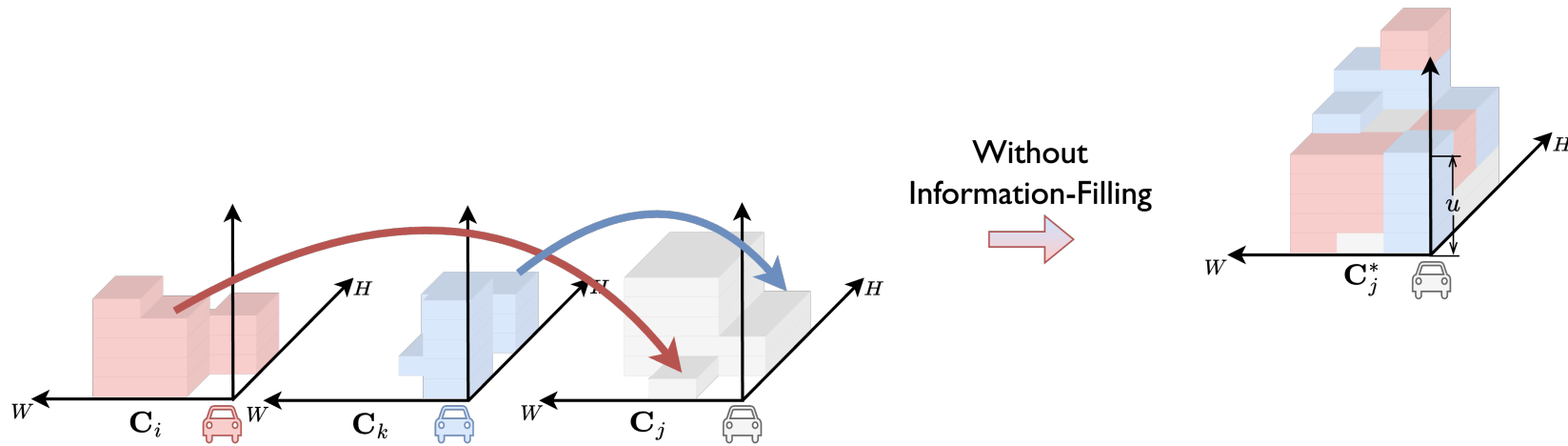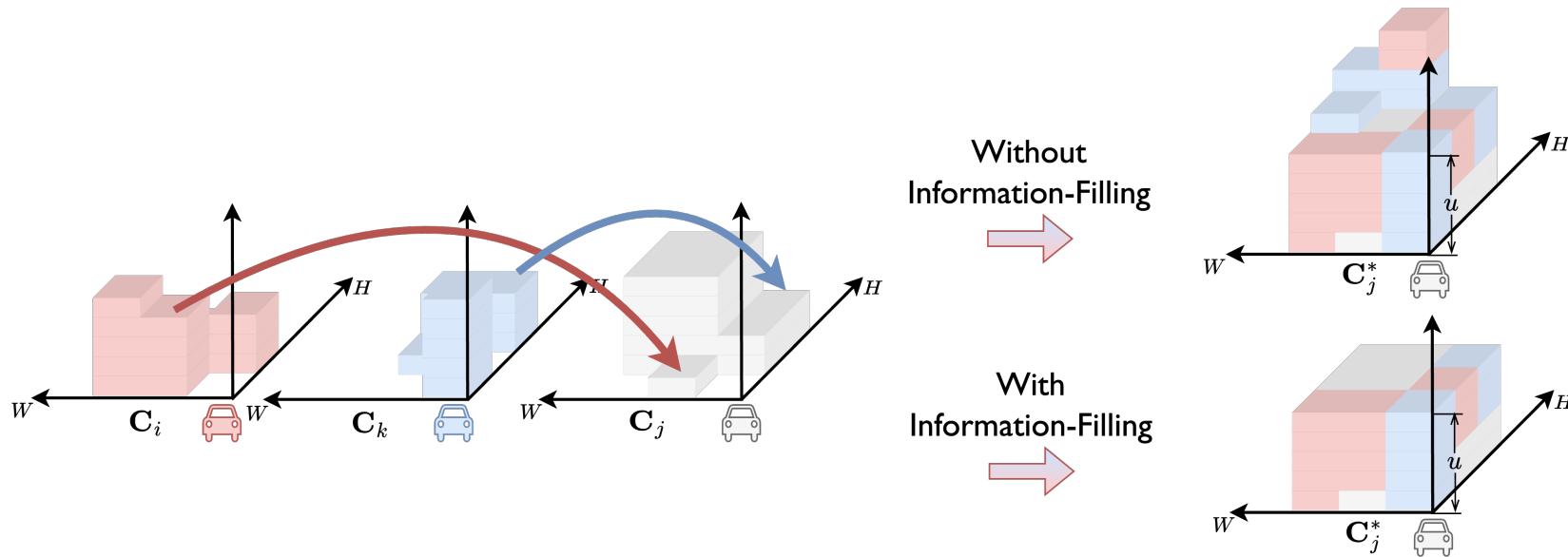High-dimentional feature vector

Integer code index

Pragmatic Collaborator Selection

Core idea: collectively fill each agent's information demand

Without
Information-Filling

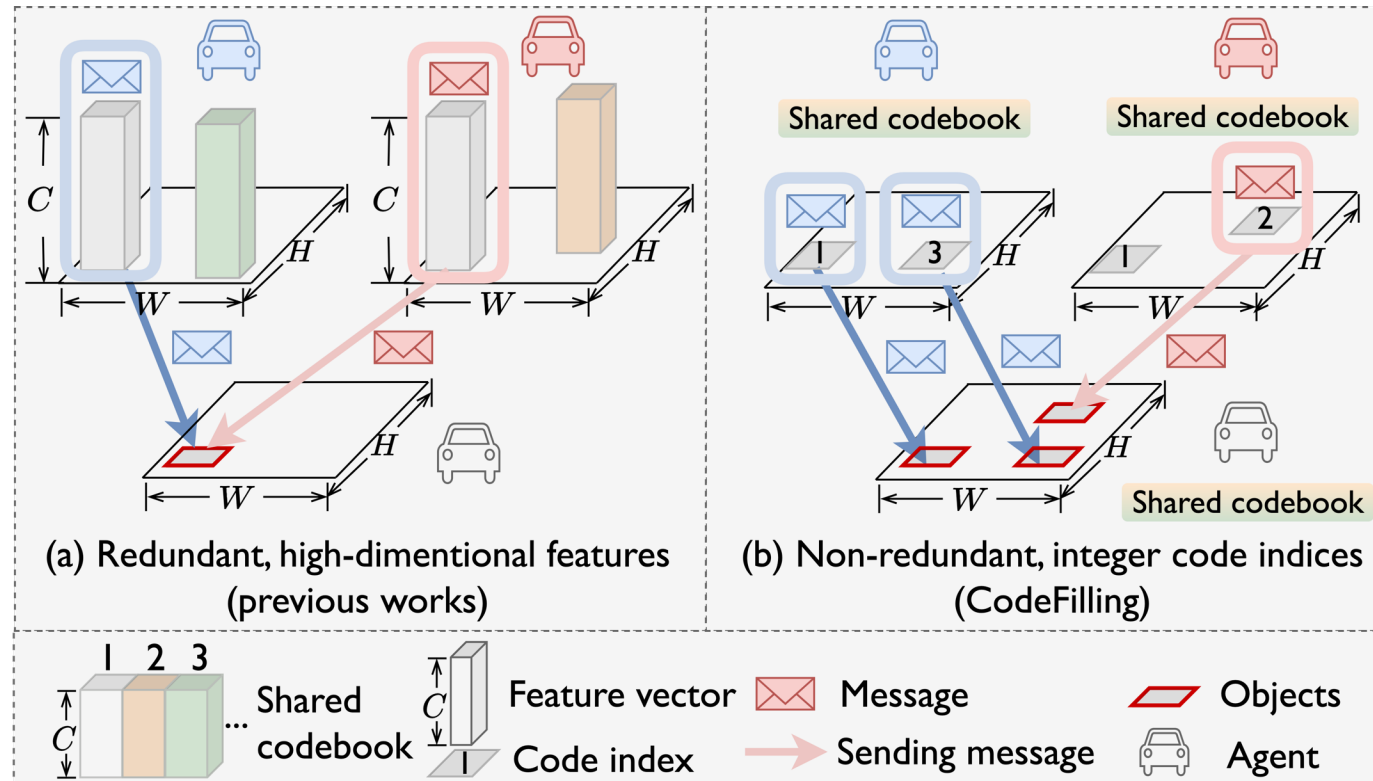$\mathbf{C}_i$ $\mathbf{C}_k$ $\mathbf{C}_j$ $\mathbf{C}_j^*$

Pragmatic Collaborator Selection

Core idea: collectively fill each agent's information demand

Without Information-Filling

With Information-Filling

Avoid information overflow.

Pragmatic Collaborator Selection

Core idea: collectively fill each agent's information demand

(a) Redundant, high-dimensional features (previous works)

(b) Non-redundant, integer code indices (CodeFilling)

Shared codebook · Feature vector · Message · Objects
Code index · Sending message · Agent

Reallocates the bandwidth wasted in redundant information to more beneficial information.

CodeFilling: Architecture
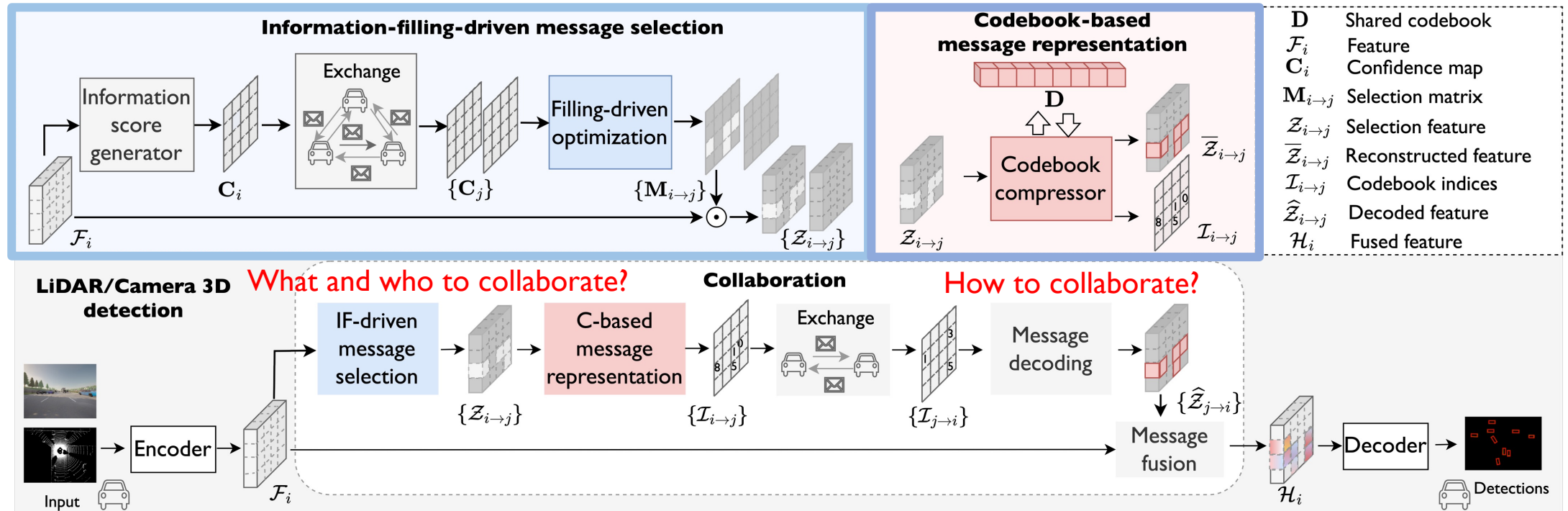
- Information-filling-driven message selection selects the **optimal collaborator set** at each spatial region to offer the most beneficial message.

- Codebook-based message representation represent the high-dimensional feature vector with integer code index.
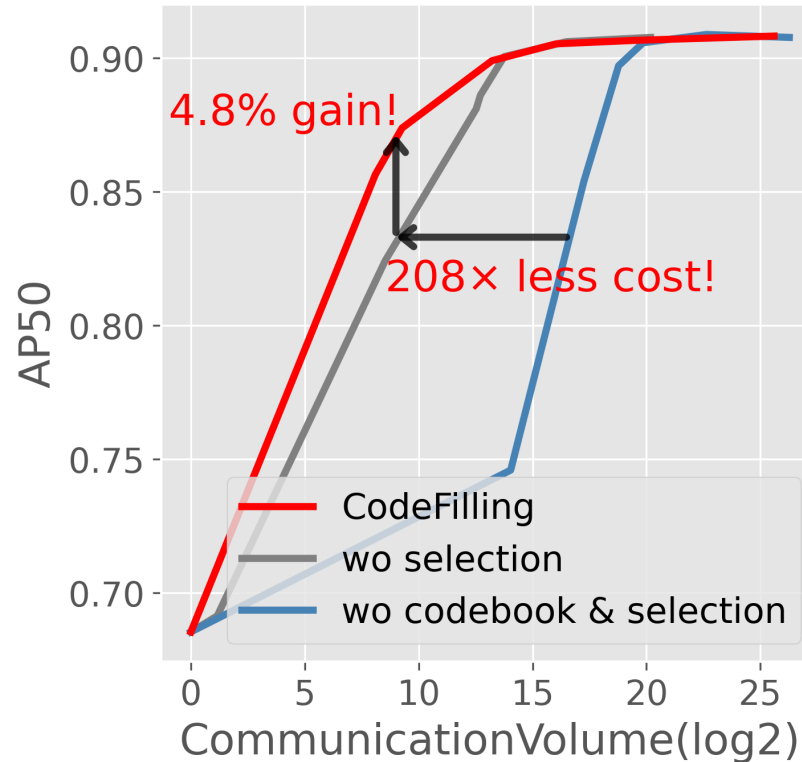
CodeFilling: Experimental results

➢ Compared to Where2comm, CodeFilling reduces communication cost under both homogeneous and heterogenous scenarios, and camera and LiDAR modality.
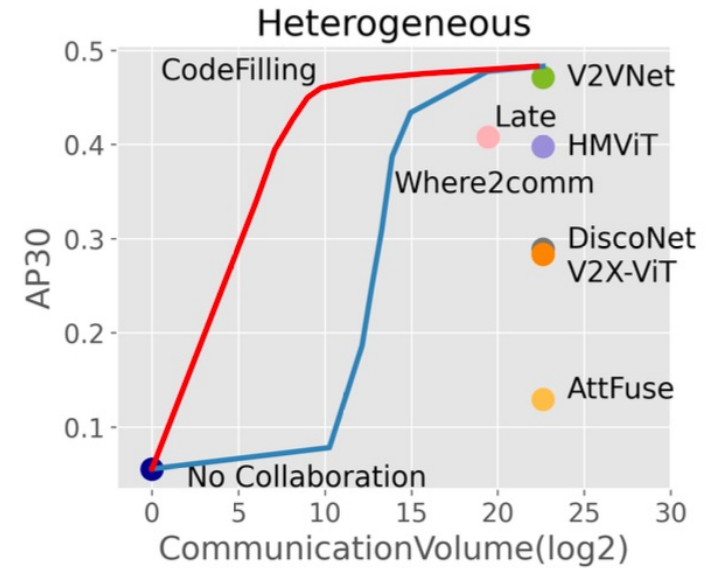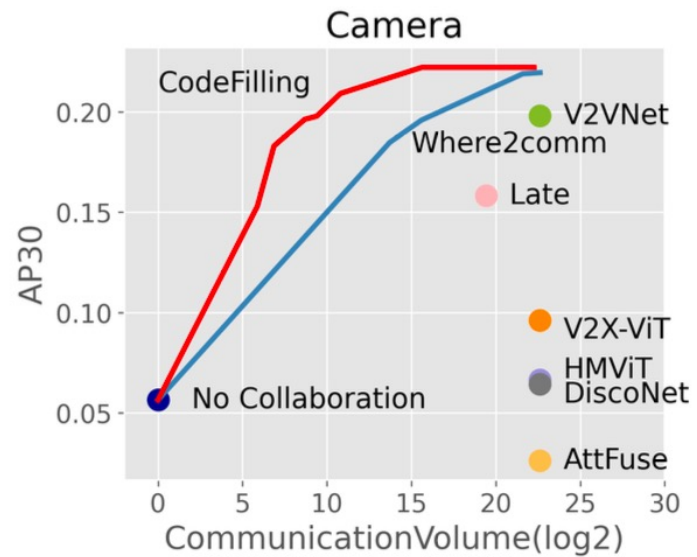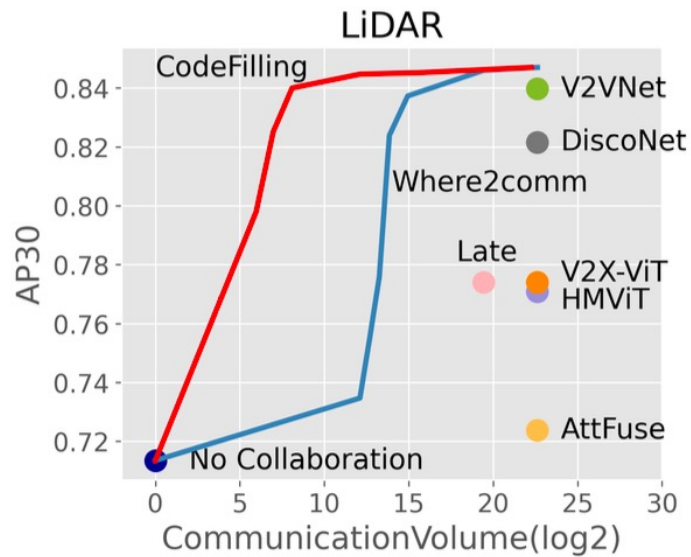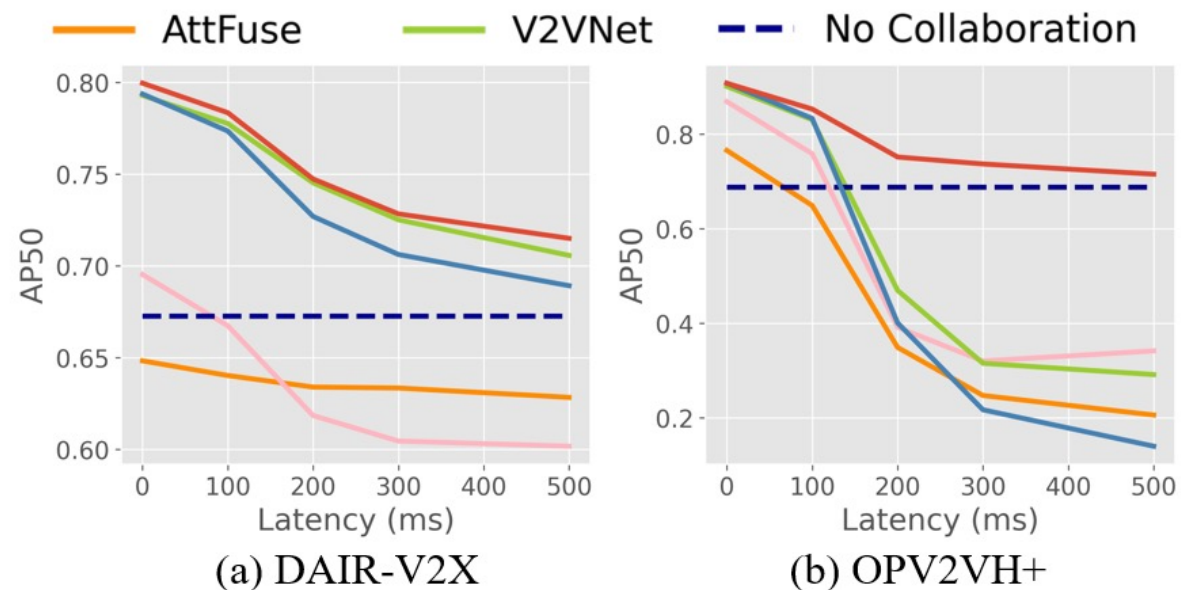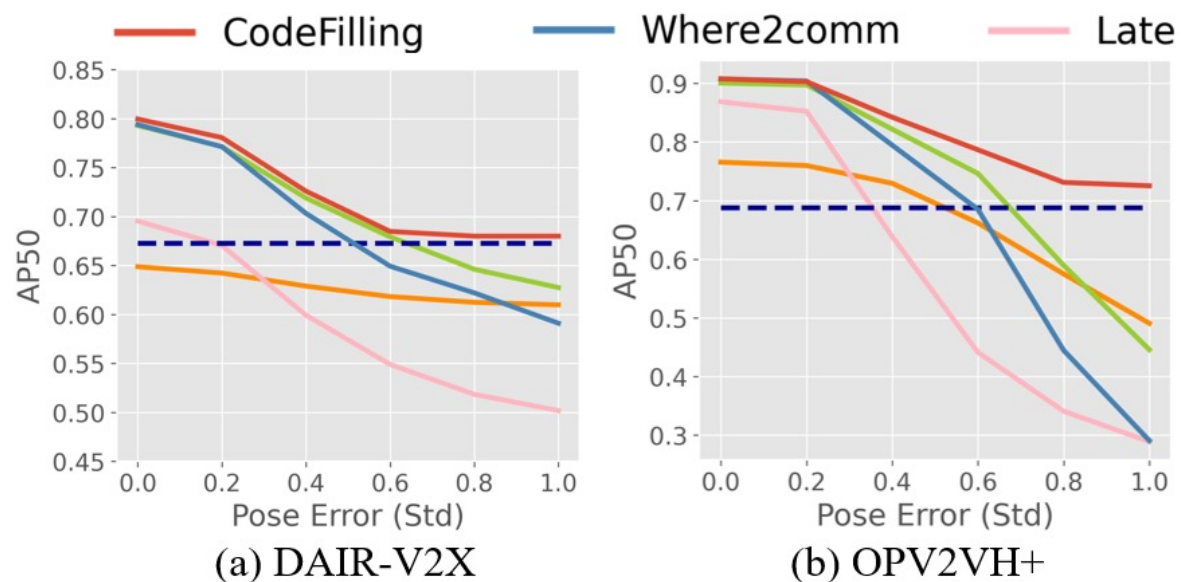
CodeFilling: Experimental results

➢ Compared to Where2comm, CodeFilling can further reduce the communication cost by 208 times.
➢ Pragmatic collaborator selection can improve the perception performance by 4.8% with the same communication cost.

CodeFilling: Experimental results

➤ CodeFilling is robust to pose error and communication latency.

(a) DAIR-V2X   (b) OPV2VH+   (a) DAIR-V2X   (b) OPV2VH+

CodeFilling: Experimental results

➢ Qualitative visualization of collaboration in CodeFilling.

(a) Ego detections   (b) Ego feature   (c) Ego information score   (d) Collaborator score 1   (e) Collaborator score 2

(f) Selection matrix 1   (g) Selection matrix 2   (h) Filled confidence   (i) Codebook-based message   (j) Collaborative detections
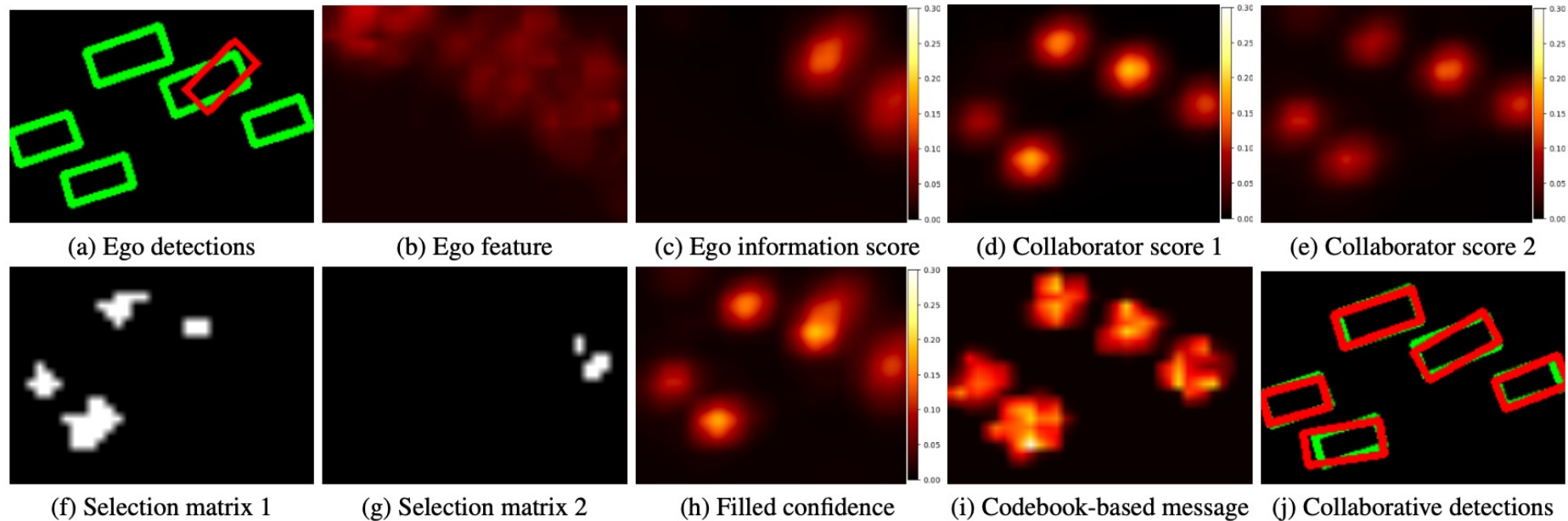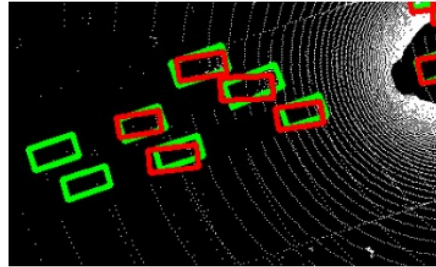
Figure 10. Visualization of collaboration in CodeFilling. Green and red denote ground truth and detection, respectively.

CodeFilling: Experimental results

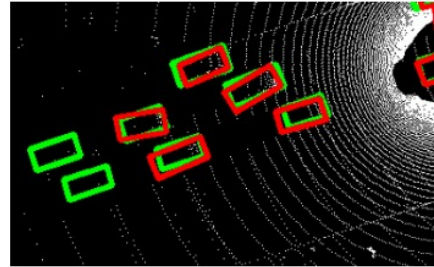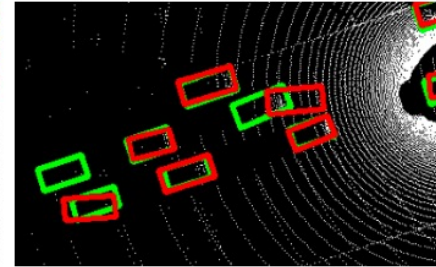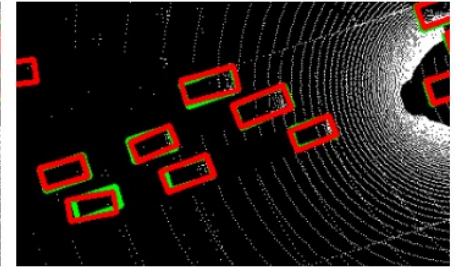➢ CodeFilling qualitatively outperforms the state-of-the-art methods.
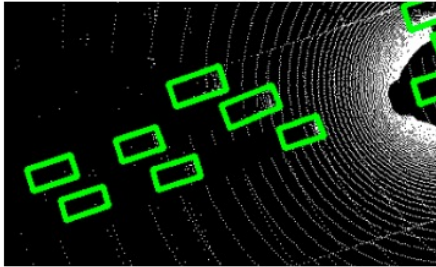
(a) No Colla   (b) V2X-ViT   (c) HMViT   (d) Where2comm   (e) CodeFilling