

自主式智能通信计算一体化技术

汇报人：杨照辉

浙江大学信息与电子工程学院

汇报日期：2025年9月

目录

研究背景和意义

自主式智能边缘网络一体化融合架构

边缘通感算控一体化任务协同优化部署

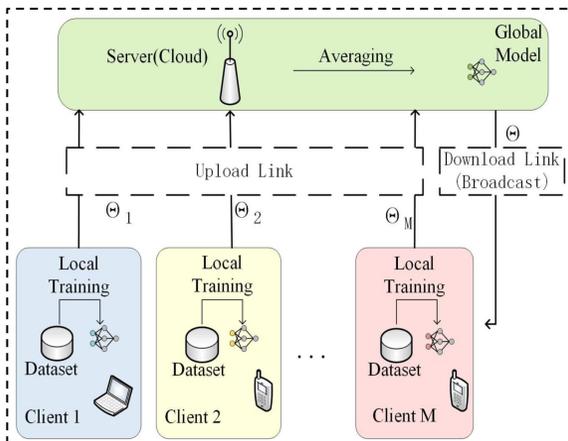
自主式智能通算应用平台

总结与展望

无线AI领域研究现状

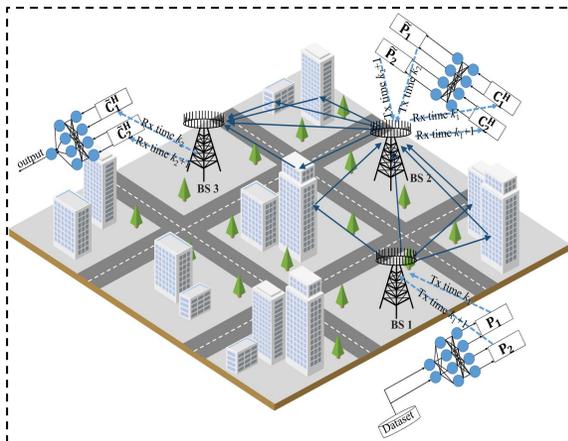
无线联邦学习

新型无线联邦学习架构包括异步、量化联邦学习架构等，实现了存在无线传输差错和延迟或带宽受限条件下的**可靠训练**



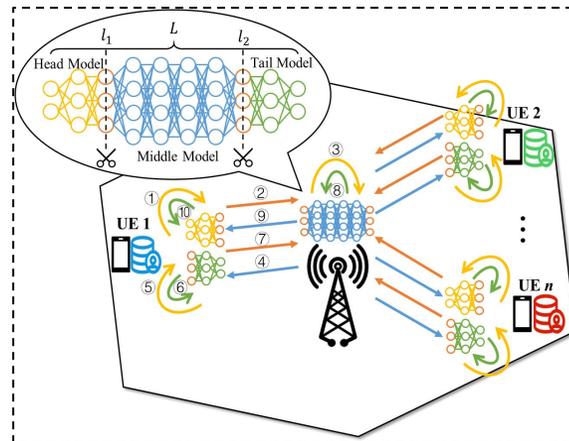
无线拆分学习

前馈计算和反向传播过程利用信道互易性有效避免了信道估计，利用星座图映射函数替代激活函数，实现**高效的神经网络计算**



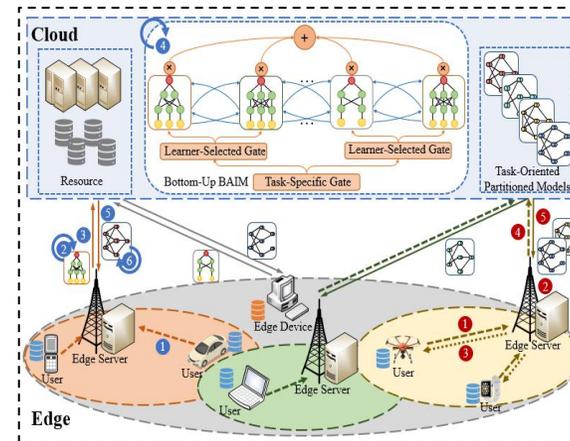
无线并行计算

结合编码计算的神经网络张量并行、流水线并行拆分训练机制，实现神经网络在多个不同算力和运力设备间的**高效分布式训练**



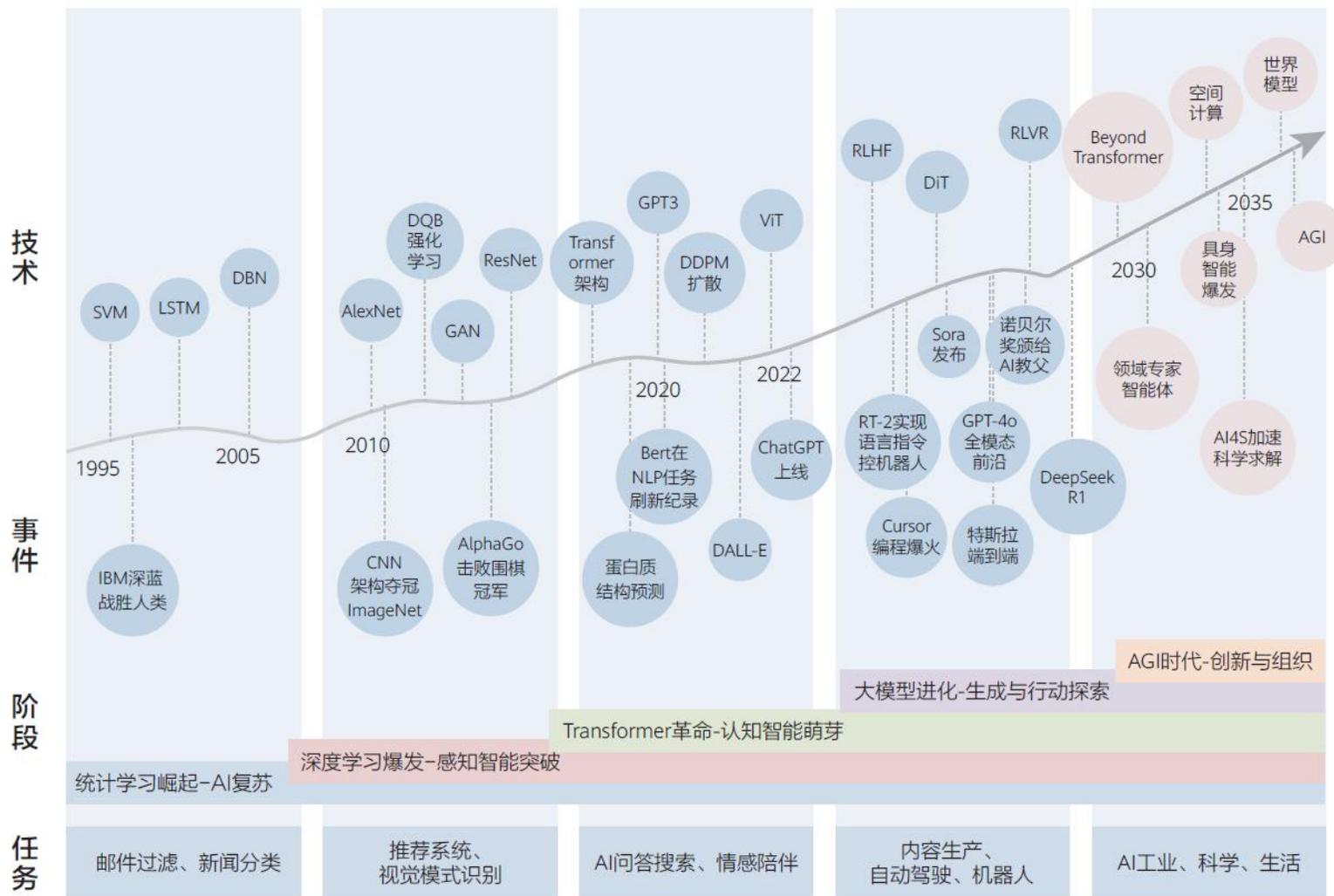
无线大模型

基于大小模型混合部署、云边协同训练的无线大模型实现框架以及面向任务的拆分聚合方案，实现**大模型的高效训练、持续微调和灵活部署**



- W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar and M. Debbah, "Edge Learning for B5G Networks With Distributed Signal Processing: Semantic Communication, Edge Computing, and Wireless Sensing," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 1, pp. 9-39, Jan. 2023.

人工智能的演进历程



人工智能的演进历程

●数据来源：华为《与AI共进：通往智能世界2035的十大技术跃迁》

■ 人工智能演进历程

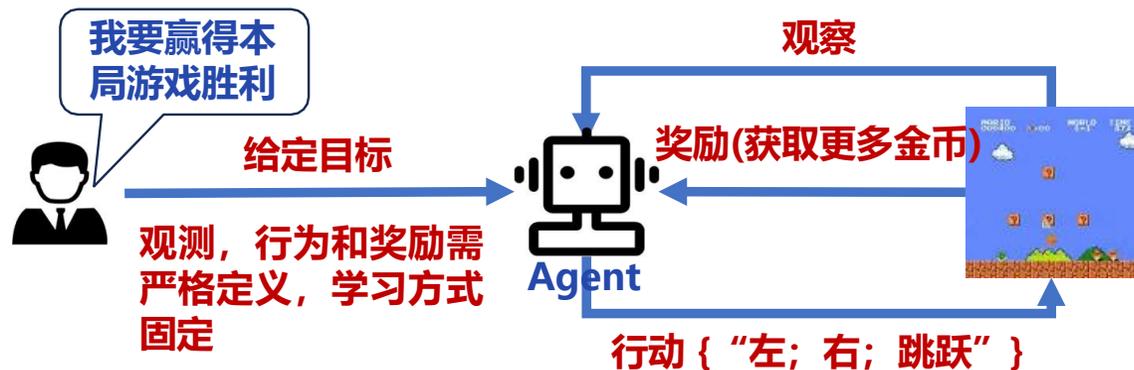
- ✓ 任务多样化：跨领域应用，解决复杂问题
- ✓ 技术融合化：深度学习、强化学习与多模态结合
- ✓ 自主智能化：自主学习、决策与适应能力提升

Agentic AI概述

定义

Agentic AI 是一种由**多个智能体**协同构成的系统，具备持久记忆、自主计划与目标驱动能力，能够在**复杂、动态环境**中持续学习和适应，标志着从被动响应任务向**主动理解、规划和执行任务**的智能跃升

AI Agent



大模型



Agentic AI



- R. Sapkota, K. I. Roumeliotis, and M. Karkee. "AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges, 2025." URL <https://arxiv.org/abs/2505.10468>.
- F. Jiang, et al. "From Large AI Models to Agentic AI: A Tutorial on Future Intelligent Communications." arXiv preprint arXiv:2505.22311 (2025).

自主式智能边缘网络 (Agentic AI-RAN)

自主式智能 (Agentic AI)

- **主动性**: 主动识别目标并采取行动, 无需外部指令触发
- **自治性**: 在预设框架内独立决策和执行任务, 减少人为干预
- **适应性**: 动态调整策略以应对环境变化或新信息

自主式智能边缘网络 (Agentic AI-RAN)

- 通过在**边缘深度部署AI和算力**, 并使**AI模型架构与通信网络架构和计算架构紧密协同**
- 从而使边缘网络**具备支撑端侧设备高效执行智能通感算控一体化协同任务的关键能力**
- 为未来**自主式智能物联网 (Agentic AIoT) 和具身智能**等新兴应用场景提供关键支撑

应用愿景：自主式智能边缘网络为“万物智联”赋能

通算一体化网络是实现“万物智联”的技术基石，自主式智能物联网作为其关键应用形态，拥有巨大的产业价值和广阔的应用前景

□ 截至2024年，中国物联网市场规模突破**4.01万亿元**，年复合增长率(CAGR)达**15%**（中研普华产业研究院《2025-2030年中国物联网与RFID行业竞争分析及发展前景预测报告》）

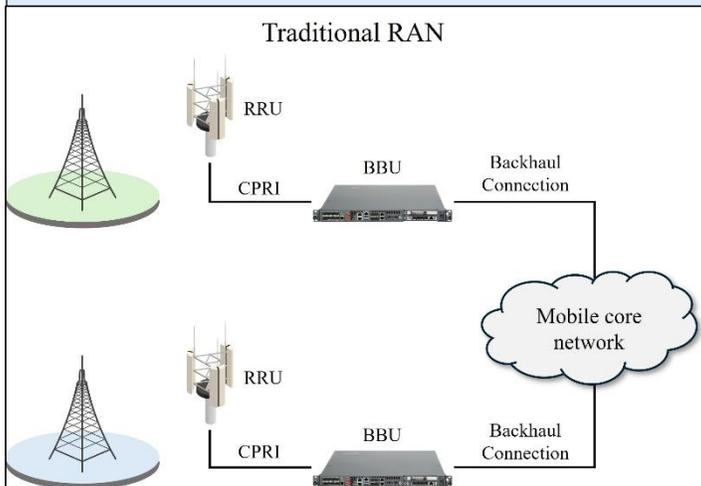
□ 中国移动发布《**AI+智慧城市安全解决方案白皮书**》《**中国物联网产业创新白皮书**》，为垂直行业赋能



当前网络存在的问题

功能冗余

传统的无线接入网（RAN）采用“烟囱式”架构，不同功能模块之间难以复用，当引入新业务或能力时往往需要重复建设，造成了功能上的冗余和资源浪费



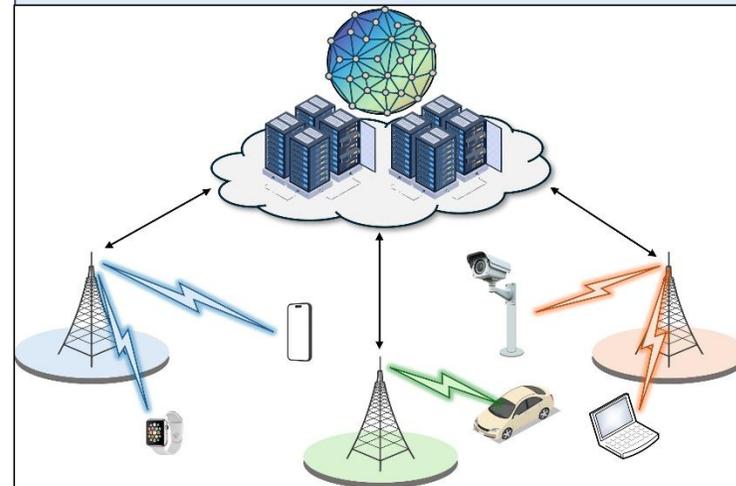
灵活定制能力差

传统RAN架构的紧耦合架构采用专用硬件体系，通信与计算资源相解耦且缺乏资源动态调度机制，导致硬件资源难以灵活适配多样化业务需求



智能程度低

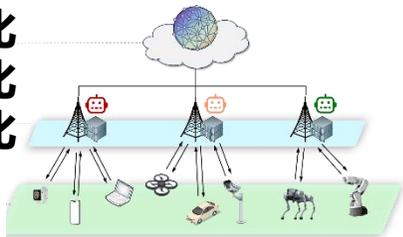
现有网络的AI能力多为外挂式，依赖云端集中式AI优化架构，边缘节点缺乏分布式智能决策能力，难以满足超低时延场景需求，无法实现全网资源的协同调度与自适应闭环



关键科学问题：自主式智能通信计算一体化协同

网络特征

模块化
分立化
定制化



研究挑战

网络智能化程度低、自主适配能力弱，多任务一体化协同难

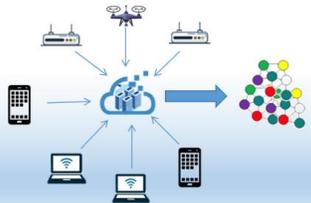


关键科学问题

关键科学问题1：

如何设计**自主式智能边缘网络一体化融合架构**

多任务
通算资源受限



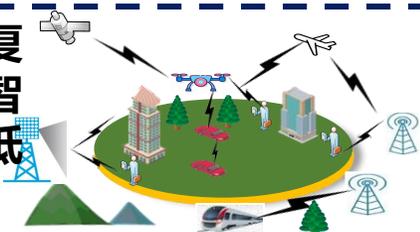
通算一体化任务需求复杂、资源多样，协同优化部署难



关键科学问题2：

如何实现**边缘通算一体化任务协同优化部署**

场景复杂、智能化低



接入复杂，空口智能化程度低



关键科学问题3：

如何构建**自主式智能通算应用平台**

目录

研究背景和意义

自主式智能边缘网络一体化融合架构

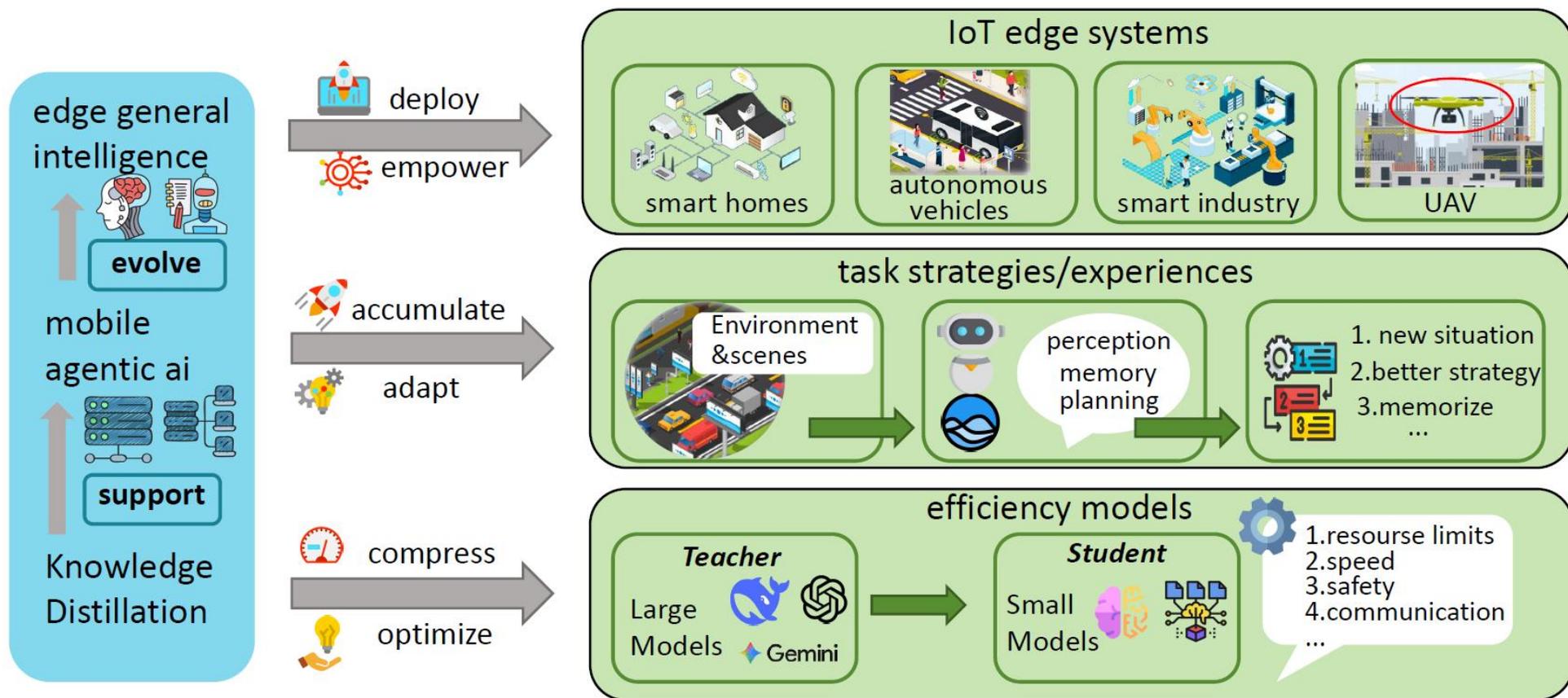
边缘通感算控一体化任务协同优化部署

自主式智能通算应用平台

总结与展望

自主式智能边缘网络一体化融合架构

- 知识蒸馏将大型模型压缩为轻量级模型以便部署，使移动自主人工智能能够积累经验并调整策略，推动了边缘通用智能的发展，赋能无人机、自动驾驶汽车和智能医疗等物联网边缘系统。
- 考虑无线传输能力的智能体决策调度， or MEC + AI + Knowledge (Memory) + Control



Agentic AI辅助的无人机语义通信系统

人工智能



无线通信



无人机技术

✓ 配备无人机的自主系统

智能监控



灾害响应

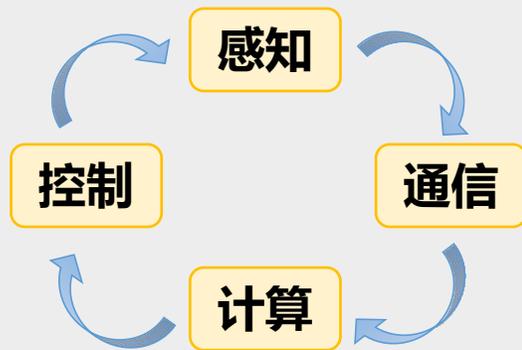


物流投递



Agentic AI辅助的无人机语义通信系统

配备无人机的自主系统



关键瓶颈

- ❑ 将高维传感器数据从无人机传输至地面决策实体的过程
- ❑ 系统自主处理智能任务的能力
- ❑ 无人机有限资源下的系统能耗优化

技术方案

语义通信

旨在提取并传输源数据中嵌入的与任务相关的核心意义
可大幅减少传输数据量，从而提升频谱和能量效率

Agentic AI

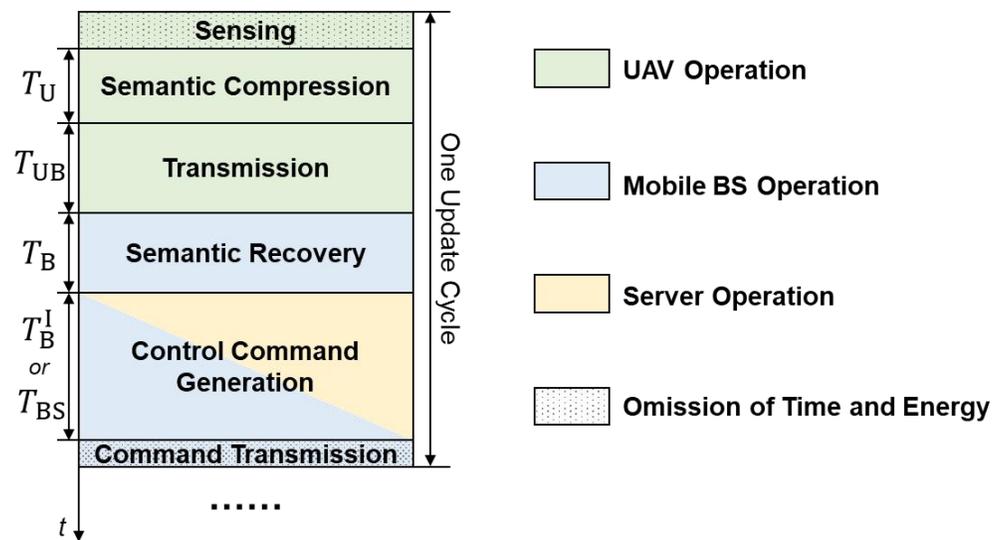
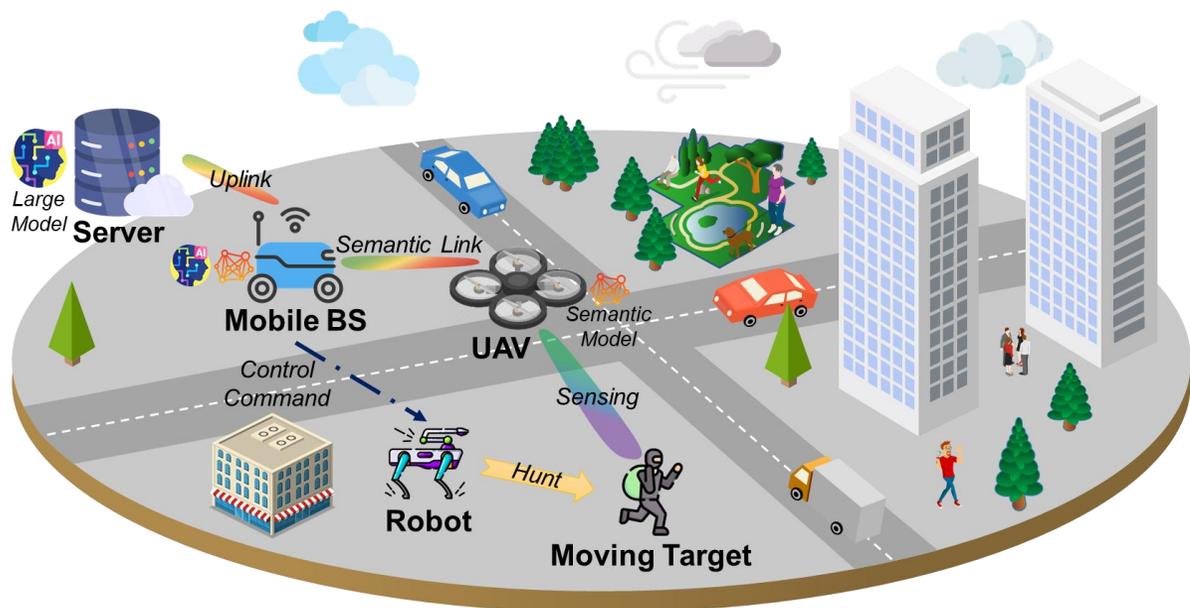
基于大模型的 Agentic AI，为这些自主系统提供了“大脑”
使其能够基于接收的语义信息进行复杂的推理和决策

联合优化

提出感知-通信-计算-控制一体化方案
联合优化 Agentic AI决策变量、语义压缩率、无人机和基站传输功率，最小化系统能耗

Agentic AI辅助的无人机语义通信系统：系统模型

- **组成元素：**无人机、移动基站、云服务器、机器人
- **工作流程：**无人机感知→语义压缩→数据传输→语义恢复→智能决策→控制机器人
- **任务卸载：**智能决策任务可由移动基站或云服务器执行



Agentic AI辅助的无人机语义通信系统：无人机通感算模型

■ 感知模型

$$q = e^{-\xi d_{UT}} \xrightarrow{\text{参考}}$$

■ 语义计算模型

$$O(\rho) = -\kappa_1 D \ln(\rho)$$

$$T_U = \frac{-\kappa_1 D \ln(\rho)}{f_U} \quad E_U = \tau T_U f_U^3$$

■ 通信模型

$$G(\mathbf{L}_U) = \frac{G_0}{d_{UB}^2} \rightarrow \text{LoS为主, 平方项约等于2}$$

$$T_{UB} = \frac{D\rho}{B_U \log\left(1 + \frac{p_U G(\mathbf{L}_U)}{B_U N_0}\right)} \quad E_{UB} = p_U T_{UB}$$

6578

IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 19, NO. 10, OCTOBER 2020

Age of Information in a Cellular Internet of UAVs: Sensing and Communication Trade-Off Design

Shuhang Zhang¹, Graduate Student Member, IEEE, Hongliang Zhang¹, Member, IEEE,
Zhu Han¹, Fellow, IEEE, H. Vincent Poor², Life Fellow, IEEE,
and Lingyang Song¹, Fellow, IEEE

We assume that it takes the UAV t_0 time slots to sense the data of a task for once. When the UAV perform video based data sensing such as precision agriculture [4], the successful sensing probability satisfies the model introduced in [23] and [24]. When the UAV performs data sensing for task n 's i -th update for one time, the successful sensing probability is

$$p_n^i(t) = e^{-\xi d_n^i(t)}, \quad (2)$$

where ξ is a parameter evaluating the sensing performance. Note that an unsuccessful data sensing severely delays the data update at the BS.⁴ Therefore, the UAV may repeatedly sense the data of a task for multiple times to achieve a higher successful sensing probability. Let ω_n^i be the number of data sensing for task n 's i -th update, and the successful sensing probability of this update can be shown as

$$\mathcal{P}_n^i = 1 - (1 - p_n^i(t))^{\omega_n^i}. \quad (3)$$

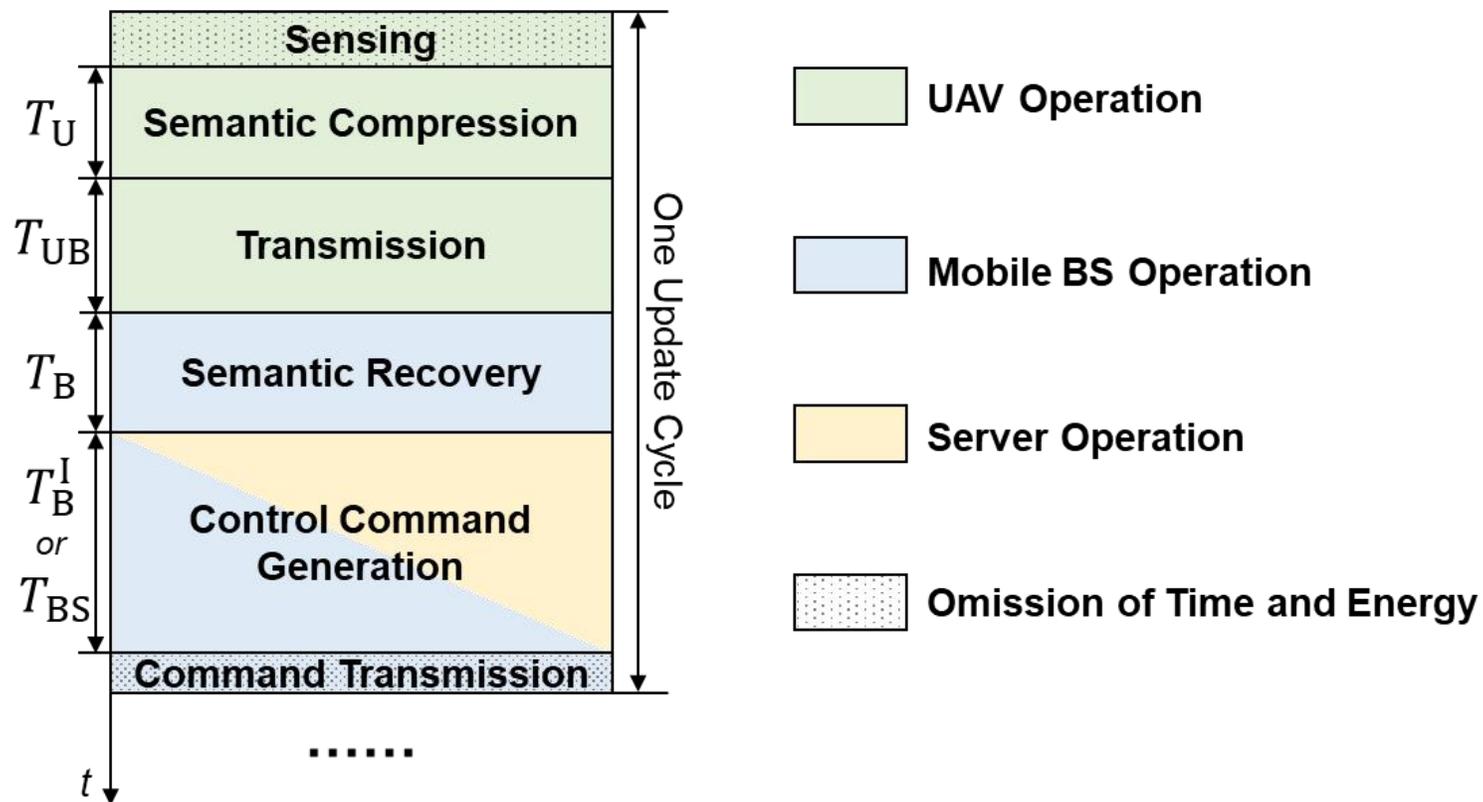
Agentic AI辅助的无人机语义通信系统：基站语义恢复模型

■ 计算时间

$$T_B = \frac{-\kappa_2 D \ln(\rho)}{f_B}$$

■ 计算能耗

$$E_B = \tau T_B f_B^3$$



Agentic AI辅助的无人机语义通信系统：控制指令生成模型

■ Case 1: 由移动基站生成 ($a = 1$)

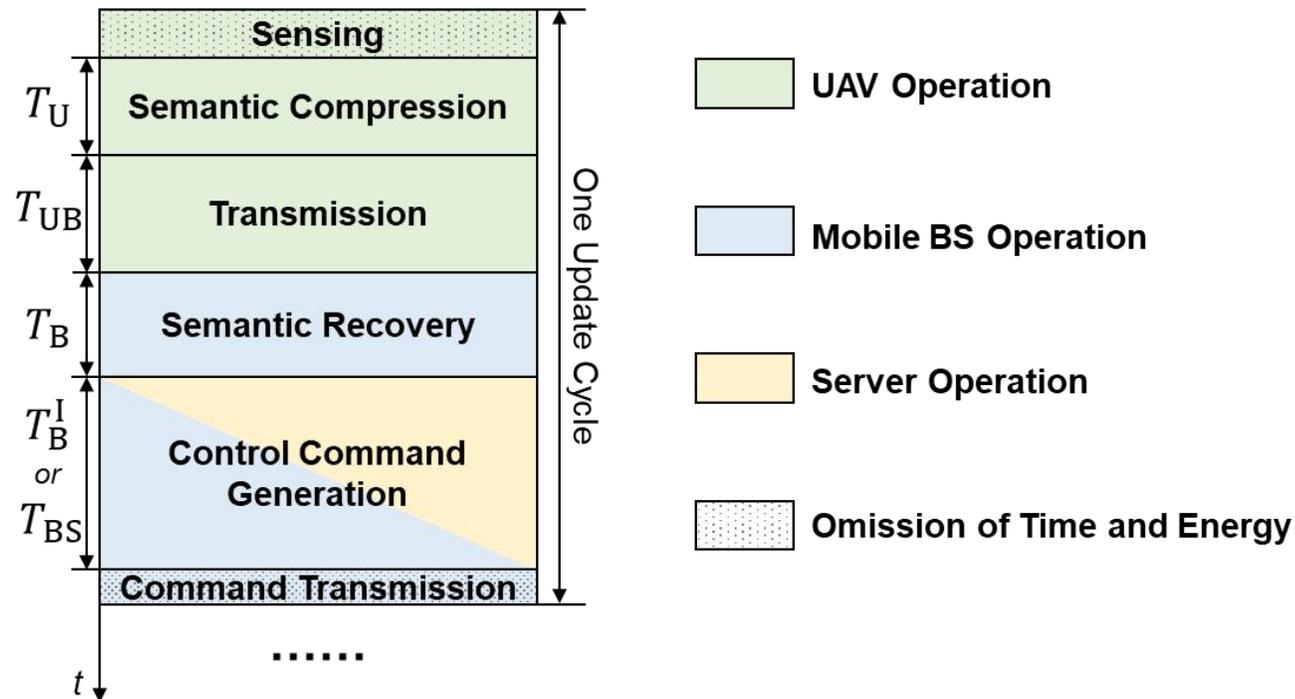
- 基站直接使用内嵌的大模型进行推理

$$T_B^I = \frac{\kappa_3 D}{f_B} \quad E_B^I = \tau T_B^I f_B^3$$

■ Case 2: 由云端服务器生成 ($a = 0$)

- 基站需要把恢复的信息上传到服务器

$$T_{BS} = \frac{D}{B_B \log \left(1 + \frac{p_B G_{BS}}{B_B N_0} \right)} \quad E_{BS} = p_B T_{BS}$$



为了简化问题:

- 服务器的计算时间和能耗可忽略, 因为服务器计算能力强大, 且有电源接入
- 控制指令的传输时间和能耗可忽略, 因为其通常体积很小

Agentic AI辅助的无人机语义通信系统：问题建模

● 总能耗 $E_{\text{tot}} = E_U + E_{UB} + E_B + aE_B^I + (1 - a)E_{BS}$

● 优化问题

$$\min_{a, \mathbf{L}_U, \rho, p_U, p_B} E_{\text{tot}} \quad (17)$$

$$\text{s.t.} \quad T_U + T_{UB} + T_B + aT_B^I + (1 - a)T_{BS} \leq T_{\text{th}}, \quad (17a)$$

$$q \geq q_{\text{th}}, \quad (17b)$$

$$H_{\min} \leq H_U \leq H_{\max}, \quad (17c)$$

$$\rho_{\text{th}} \leq \rho \leq 1, \quad (17d)$$

$$0 < p_U \leq p_U^{\max}, \quad (17e)$$

$$0 < p_B \leq p_B^{\max}, \quad (17f)$$

$$a \in \{0, 1\}. \quad (17g)$$

目标：最小化一个update cycle内的总能耗

优化变量：Agentic AI 推理任务卸载变量

无人机位置

语义压缩率

无人机发射功率

移动基站发射功率

约束：总时延约束

感知QoS约束

无人机高度约束

语义压缩率约束

无人机发射功率约束

移动基站发射功率约束

a的0-1约束

Agentic AI辅助的无人机语义通信系统：最优无人机位置

可以直观地看到，无人机位置在满足感知QoS约束的前提下，离移动基站越近越好，这既节约能耗，又可以缩短时延。

$$\begin{aligned} \min_{\mathbf{L}_U} \quad & d_{UB} \\ \text{s.t.} \quad & d_{UT} \leq \frac{-\ln(q_{th})}{\xi}, \\ & H_{min} \leq H_U \leq H_{max}. \end{aligned}$$

- 令 $D_{max} = \frac{-\ln(q_{th})}{\xi}$ ，假设 $H_{min} \leq D_{max}$ 以保证问题可行
- 运用几何学的知识，可以得到最优无人机位置的闭式解

Theorem 1 *The optimal height is always the minimum allowable height:*

$$H_U^* = H_{min}. \quad (19)$$

The optimal horizontal coordinates (x_U^, y_U^*) are determined by the following two cases:*

1) *Case 1 ($x_T^2 + y_T^2 \leq D_{max}^2 - H_{min}^2$): In this case, the optimal location is directly above the mobile BS at the minimum height, which is*

$$\mathbf{L}_U^* = (0, 0, H_{min}). \quad (20)$$

2) *Case 2 ($x_T^2 + y_T^2 > D_{max}^2 - H_{min}^2$): In this case, the optimal location is on the horizontal boundary of the service area, in the direction of the target, which is*

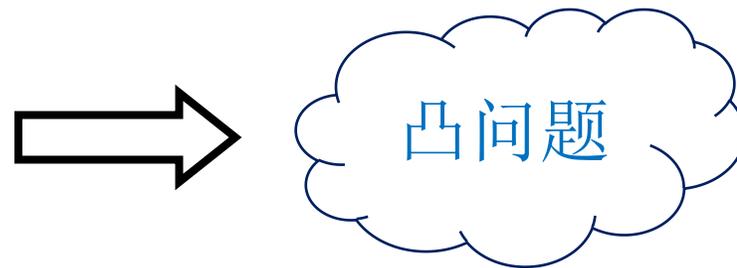
$$x_U^* = x_T \left(1 - \frac{\sqrt{D_{max}^2 - H_{min}^2}}{\sqrt{x_T^2 + y_T^2}} \right), \quad (21)$$

$$y_U^* = y_T \left(1 - \frac{\sqrt{D_{max}^2 - H_{min}^2}}{\sqrt{x_T^2 + y_T^2}} \right), \quad (22)$$

$$H_U^* = H_{min}. \quad (23)$$

Agentic AI辅助的无人机语义通信系统：最优语义压缩率

$$\begin{aligned} \min_{\rho} \quad & A\rho - F \ln(\rho) \\ \text{s.t.} \quad & K_1\rho + K_2 \ln(\rho) \leq T, \\ & \rho_{\text{th}} \leq \rho \leq 1, \end{aligned}$$



求导可得闭式解： $\rho_0 = \frac{F}{A}$

$$\rho^*(p_U, p_B) = \min \{ \max \{ \rho_0, \rho_{\text{th}}, \rho_a \}, 1, \rho_b \}$$

其中 ρ_a, ρ_b 是超越方程 $K_1\rho + K_2 \ln(\rho) = T$ 的两个根，可用朗博W函数求解：

$$\rho_a = \frac{K_2}{K_1} W_0 \left(\frac{K_1}{K_2} e^{\frac{T}{K_2}} \right) \quad \rho_b = \frac{K_2}{K_1} W_{-1} \left(\frac{K_1}{K_2} e^{\frac{T}{K_2}} \right)$$

其中

$$A \triangleq \frac{p_U D}{B_U \log_2 \left(1 + \frac{p_U G(\mathbf{L}_U^*)}{B_U N_0} \right)}, \quad F \triangleq D(\tau_U \kappa_1 f_U^2 + \tau_B \kappa_2 f_B^2),$$

$$K_1 \triangleq \frac{D}{B_U \log_2 \left(1 + \frac{p_U G(\mathbf{L}_U^*)}{B_U N_0} \right)}, \quad K_2 \triangleq -D \left(\frac{\kappa_1}{f_U} + \frac{\kappa_2}{f_B} \right),$$

$$T \triangleq T_{\text{th}} - \left(\frac{a\kappa_3 D}{f_B} + \frac{(1-a)D}{B_B \log_2 \left(1 + \frac{p_B G_{\text{BS}}}{B_B N_0} \right)} \right).$$

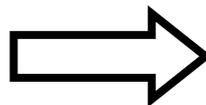
Agentic AI辅助的无人机语义通信系统：最优发射功率

$$\min_{p_U, p_B} \frac{p_U D \rho^*(p_U, p_B)}{B_U \log \left(1 + \frac{p_U G(\mathbf{L}_U^*)}{B_U N_0} \right)} + \frac{p_B D(1-a)}{B_B \log \left(1 + \frac{p_B G_{BS}}{B_B N_0} \right)} \quad (26)$$

$$\text{s.t. } T_U + T_{UB} + T_B + aT_B^I + (1-a)T_{BS} \leq T_{th}, \quad (26a)$$

$$0 < p_U \leq p_U^{\max}, \quad (26b)$$

$$0 < p_B \leq p_B^{\max}. \quad (26c)$$



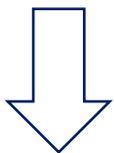
非凸问题

其中 $\rho^*(p_U, p_B) = \min \{ \max \{ \rho_0, \rho_{th}, \rho_a \}, 1, \rho_b \}$

- 由于问题中 $\rho^*(p_U, p_B)$ 的表达式较为复杂，很难对目标函数进行求导。
- 因此，采用二维搜索的方法对此问题进行求解。

Agentic AI辅助的无人机语义通信系统：最优任务卸载

考虑到推理任务卸载变量 a 只有两个取值，0或1，只需要在两种情况下对上述问题进行求解，然后比较两个解的优劣即可。



- 令 V_0 为 $a = 0$ 对应的解， V_1 为 $a = 1$ 对应的解，则最优推理任务卸载为

$$a^* = \begin{cases} 0, & \text{if } V_0 < V_1, \\ 1, & \text{otherwise.} \end{cases}$$

总体算法流程

Algorithm 1 Solving Problem (17) with Optimal Solution

- 1: Obtain the optimal UAV location \mathbf{L}_U^* using Theorem 1.
- 2: Obtain the optimal semantic compression ratio $\rho^*(p_U, p_B)$ using Theorem 2.
- 3: **for** $a \in \{0, 1\}$ **do**
- 4: Solve problem (26) via two-dimensional search.
- 5: Store the minimum total energy as V_a .
- 6: **end for**
- 7: Compare V_0 and V_1 to determine a^* and the corresponding optimal variables.

可以得到原问题的**全局最优解**!

Agentic AI辅助的无人机语义通信系统：仿真结果

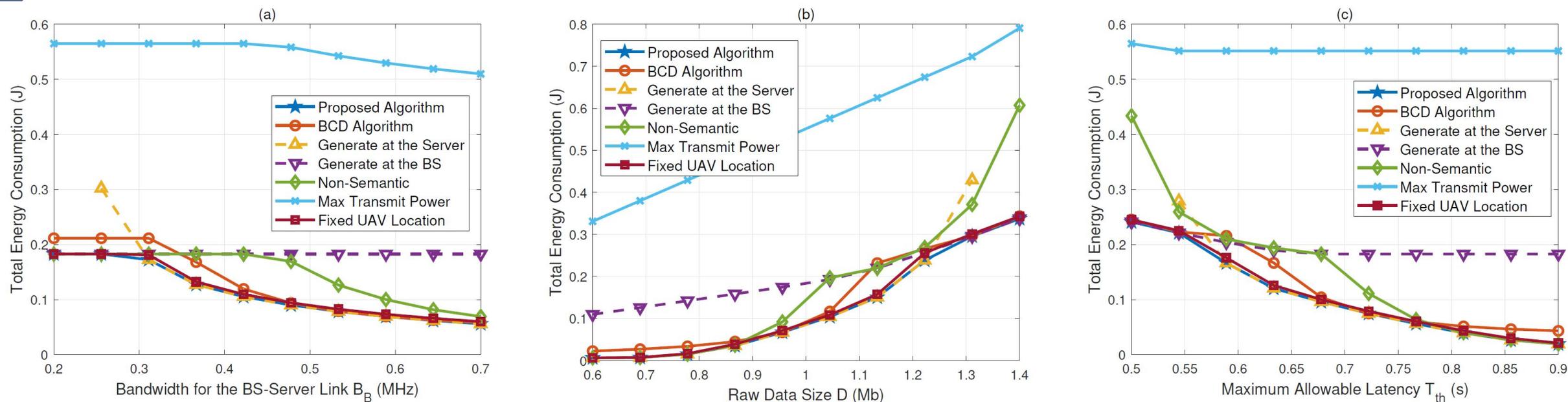


Fig. 3. Total energy consumption versus: (a) Bandwidth for the BS-server link B_B , (b) Raw data size D , (c) Maximum allowable latency T_{th} .

- 在所有模拟场景中，所提算法始终实现**最低能耗**，验证了其在联合资源分配中的卓越性能
- “Fixed UAV Location” 方案始终存在性能差距，凸显了**优化无人机部署**带来的节能效益
- “BCD Algorithm” 偶尔会收敛到局部最优解，这进一步证明了本方案在获取**全局最优解**方面的优势
- 随着基站-服务器带宽增加，**向服务器卸载的能耗成本降低**，使其成为更优策略
- “Non-Semantic” 方案在大数据量或严格时延约束下表现欠佳，证实了**语义通信**带来的显著效率提升

目录

研究背景和意义

自主式智能边缘网络一体化融合架构

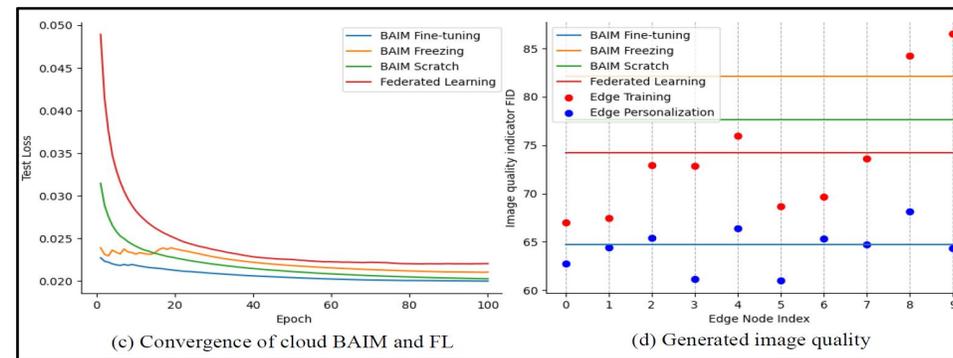
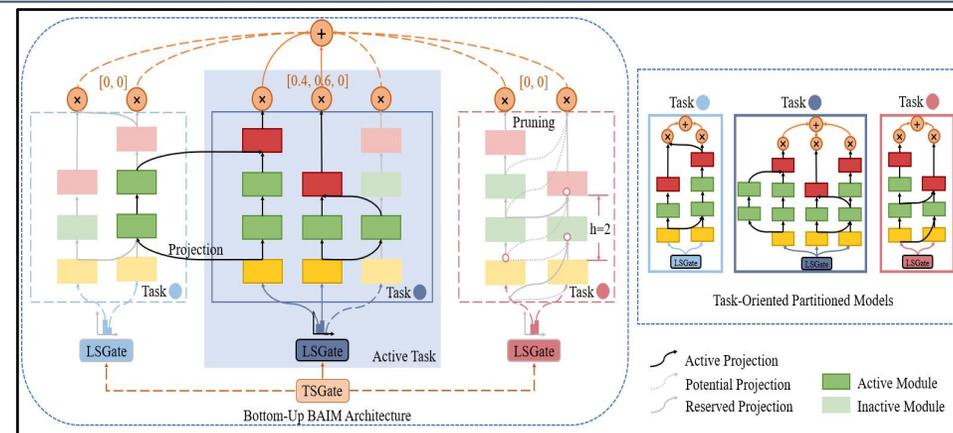
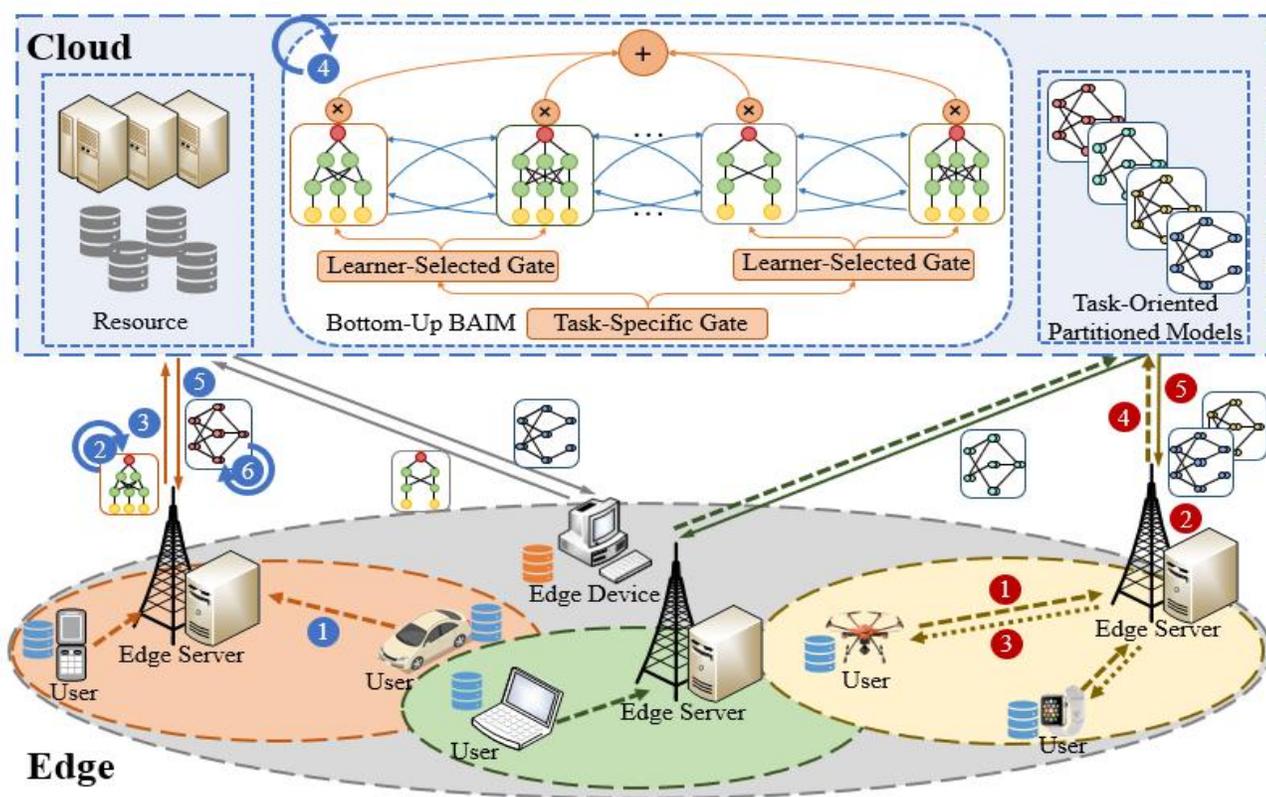
边缘通感算控一体化任务协同优化部署

自主式智能通算应用平台

总结与展望

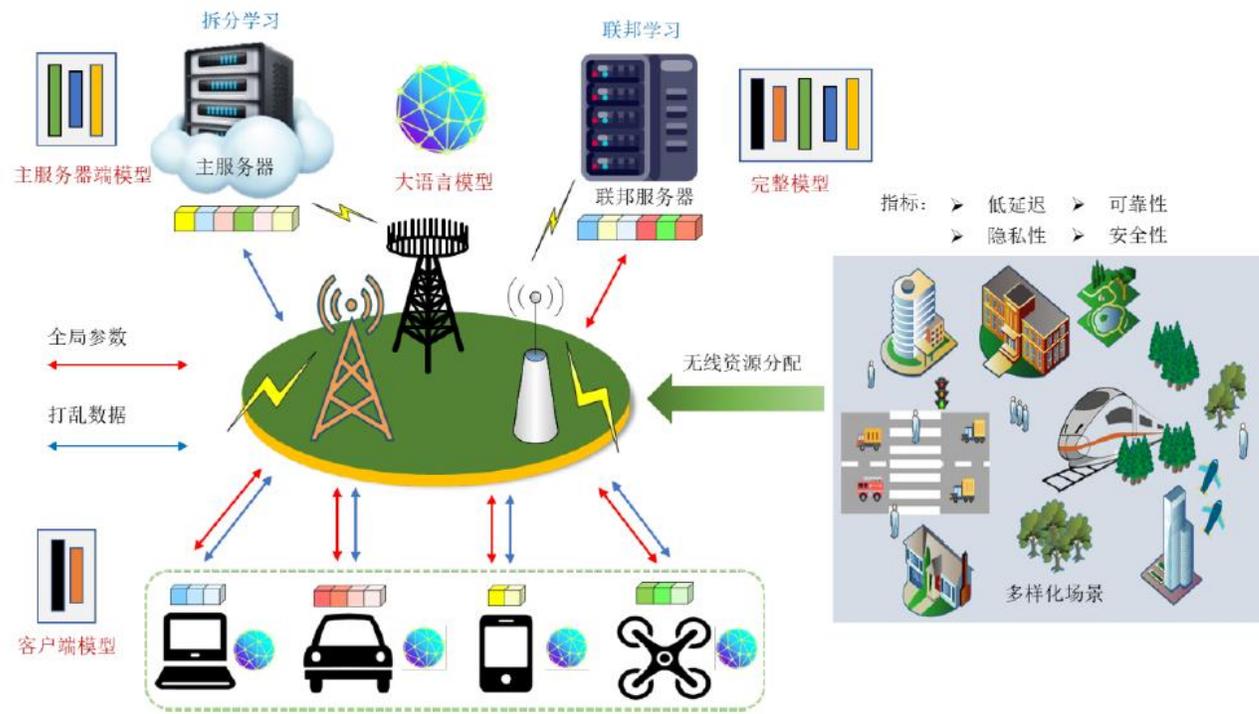
边缘通感算控一体化任务协同优化部署——云边大小模型协同训练部署

基于MoE架构的云边混合模型协同训练部署： 云端大模型以MoE方式实现对边侧/端侧小模型的专家聚合和（按任务）个性化拆分下发，通过优化利用分布式数据和分布式算力，实现了云边协同大模型高效训练、持续演进、灵活部署，相同数据和通信时间内Loss降低50%，大幅度提升通信计算效能

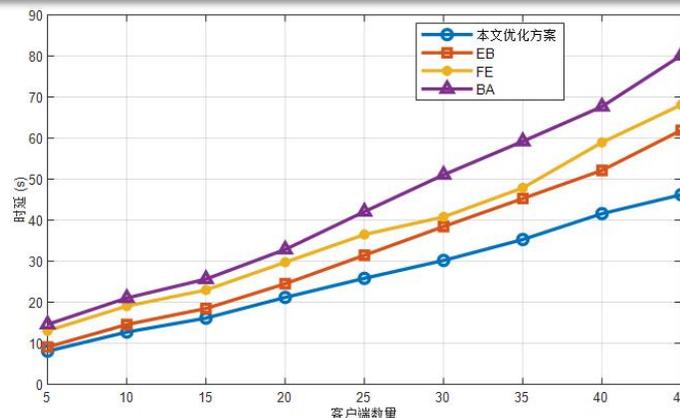


边缘通感算控一体化任务协同优化部署——云边大小模型协同训练部署

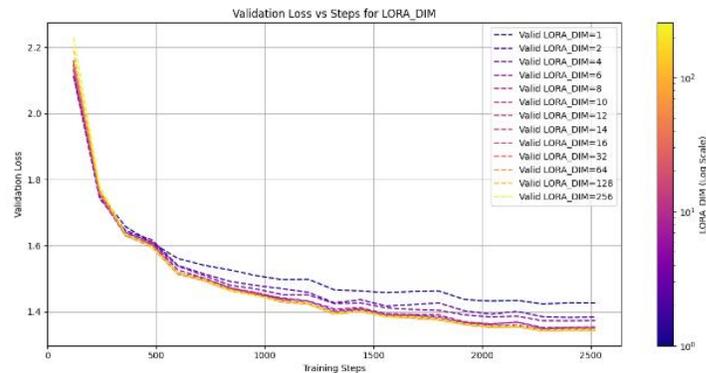
- 设计了结合联邦学习、拆分学习和低秩自适应技术的大模型分布式部署方法
- 提出了联合通信和计算的大模型部署资源分配方案



面向6G网络的拆分联邦大模型微调方法



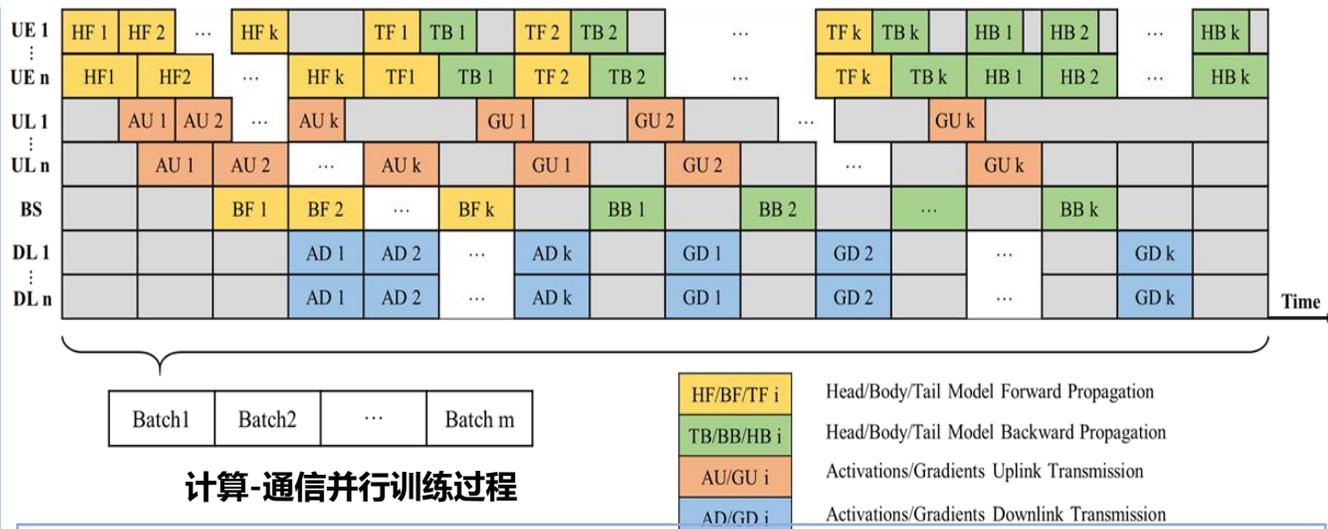
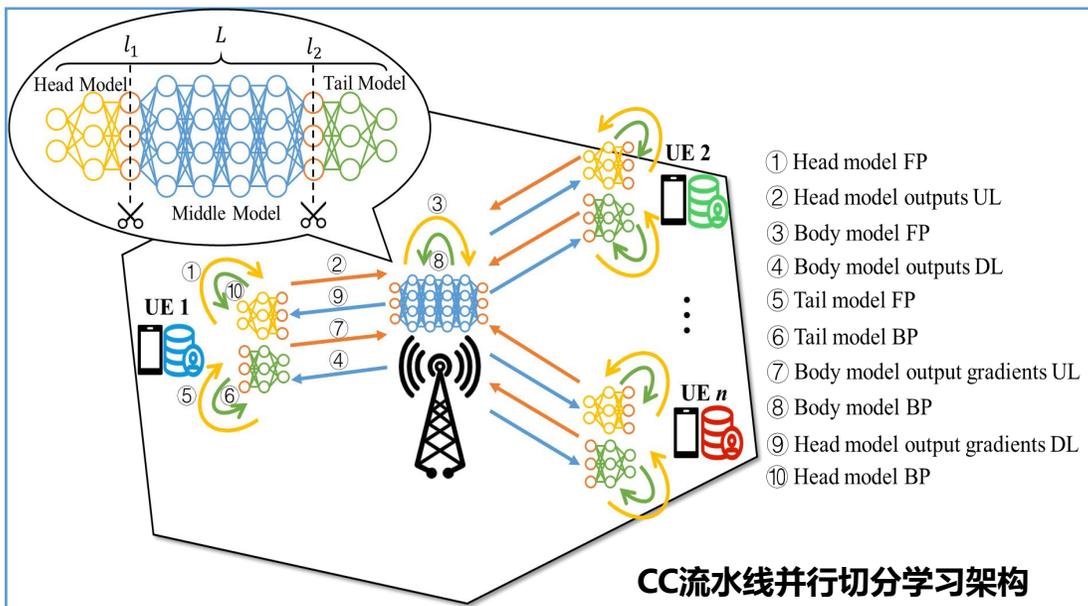
不同客户端数量时的最小训练时延



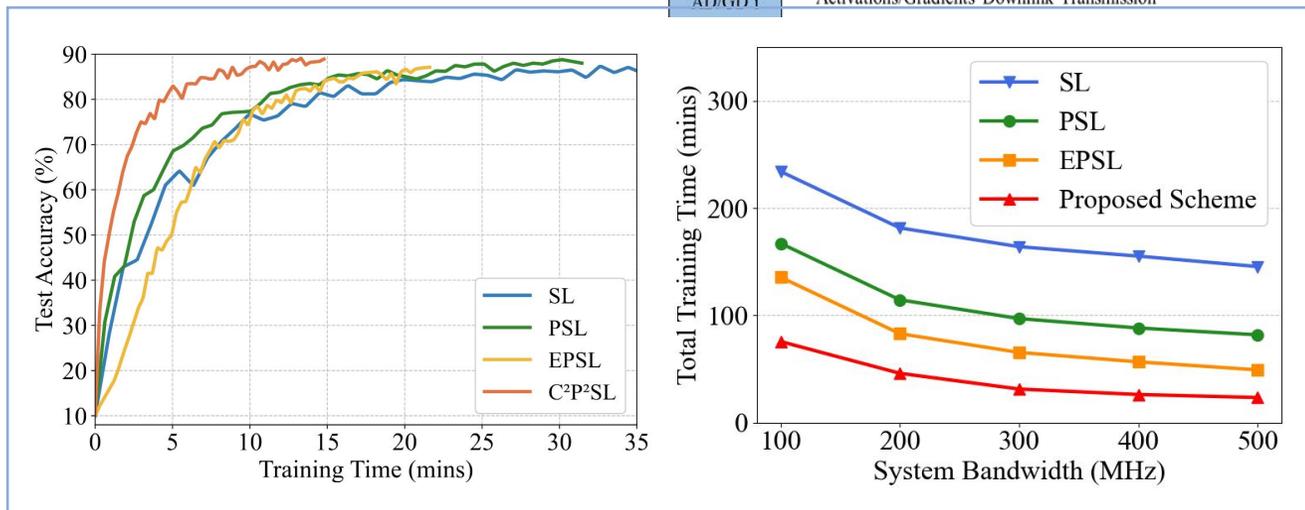
不同LoRA秩数时微调任务收敛情况

边缘通感算控一体化任务协同优化部署——通信-计算流水线并行C²P²

适配无线网络边、端算力和多用户通信能力的通信-计算流水线并行拆分学习 (C²P²SL) 架构：优化神经网络拆分机制和系统通信、计算资源分配策略，大幅提升并行效率，延迟降低超2/3

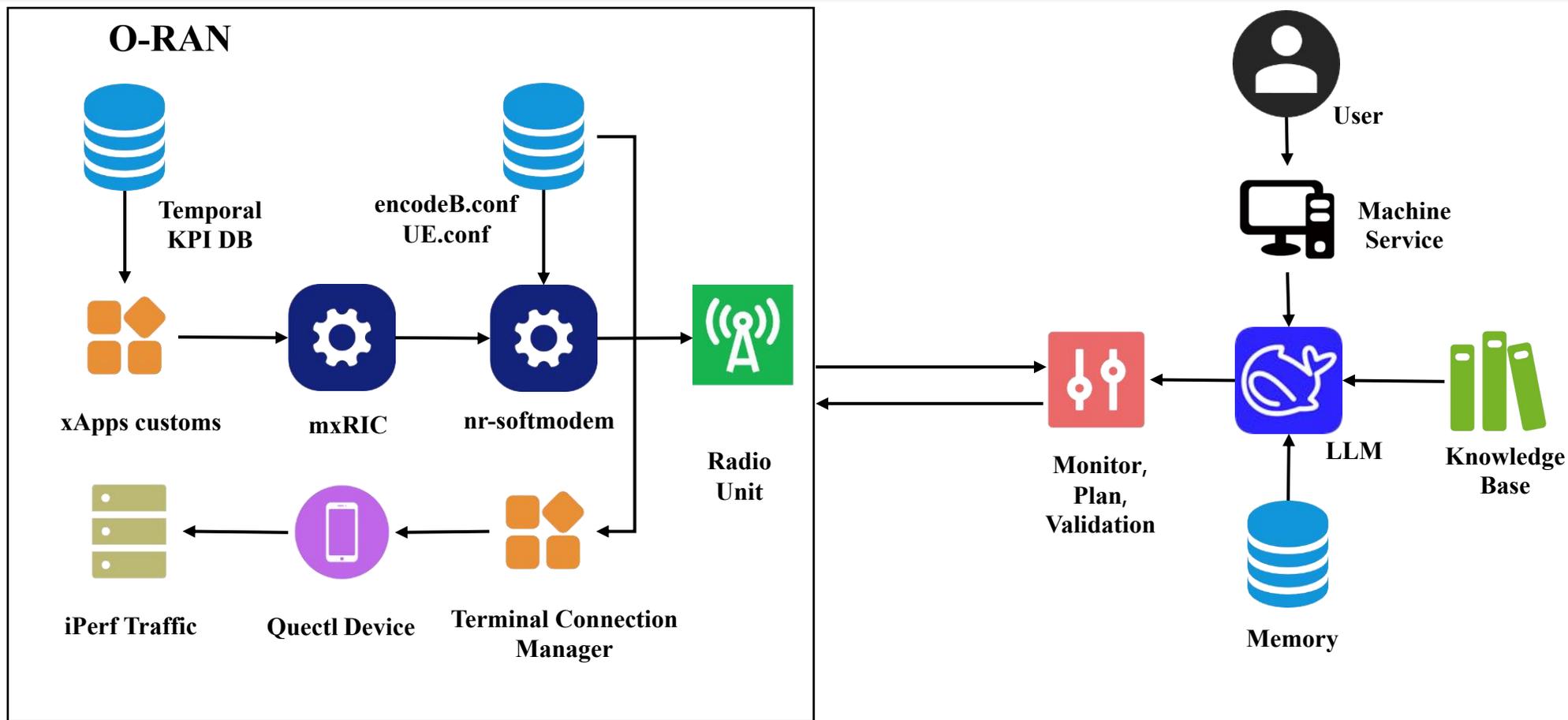


- C. Liu, Z. Zhang, et al, "Split Learning over Wireless Network via Communication and Computing Parallelization," submitted.
- S. Ji, Z. Zhang, et al, "Coded Parallelism for Distributed Deep Learning," 2023 IEEE International Symposium on Information Theory (ISIT), 2023, pp. 2410-2415. (Student Travel Award)



边缘通感算控一体化任务协同优化部署——AI Agent和LLM驱动的系统框架

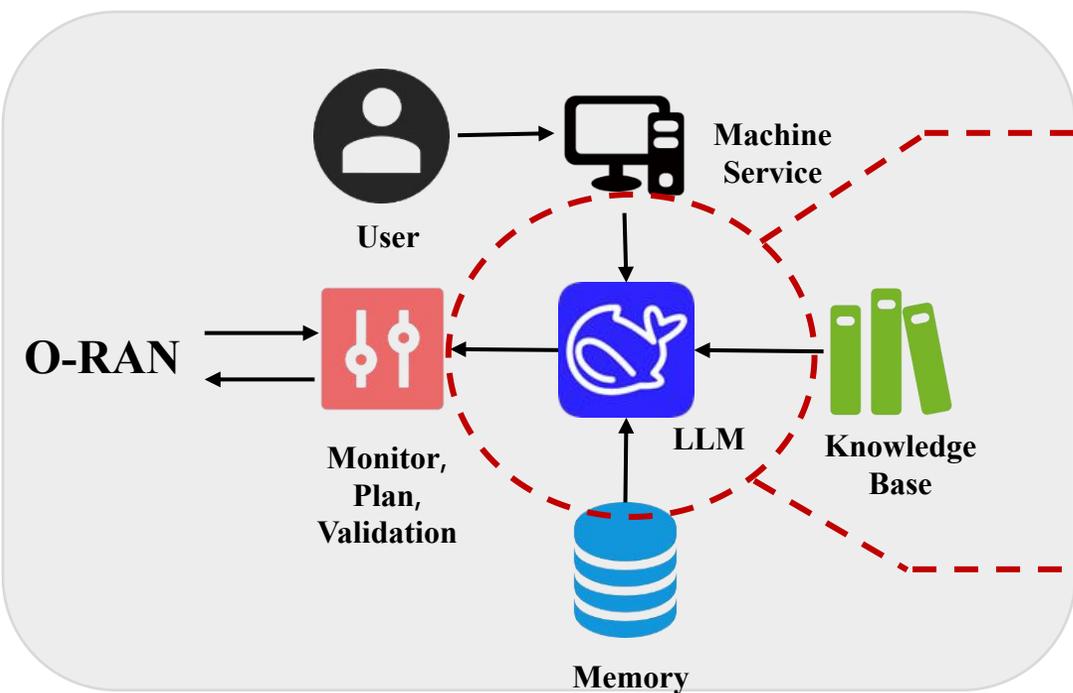
□ 利用AI Agent和LLM驱动的系统框架，实现边缘接入网络的**实时监测**、**智能规划**与**闭环优化**。通过融合LLM推理能力和领域知识库，系统可在动态环境中**自主权衡**多项性能指标，**动态调整配置参数**，**自适应**提升网络性能与资源利用效率。



边缘通感算控一体化任务协同优化部署——多智能体框架

利用 **CAMEL-AI** 团队构建的**多智能体框架**构建具备图像理解、生成与编辑能力的智能体系统，系统将结合 LLM 推理与图像工具，在动态环境中**解析**视觉数据、**生成**场景内容，并闭环**优化**网络配置，实现图像驱动的智能感知与自适应调度。

AI Agent 与 LLM 协同驱动的网络自优化系统



基于 LLM 的多模态 Agent 协同推理与工具调用核心架构图



目录

研究背景和意义

自主式智能边缘网络一体化融合架构

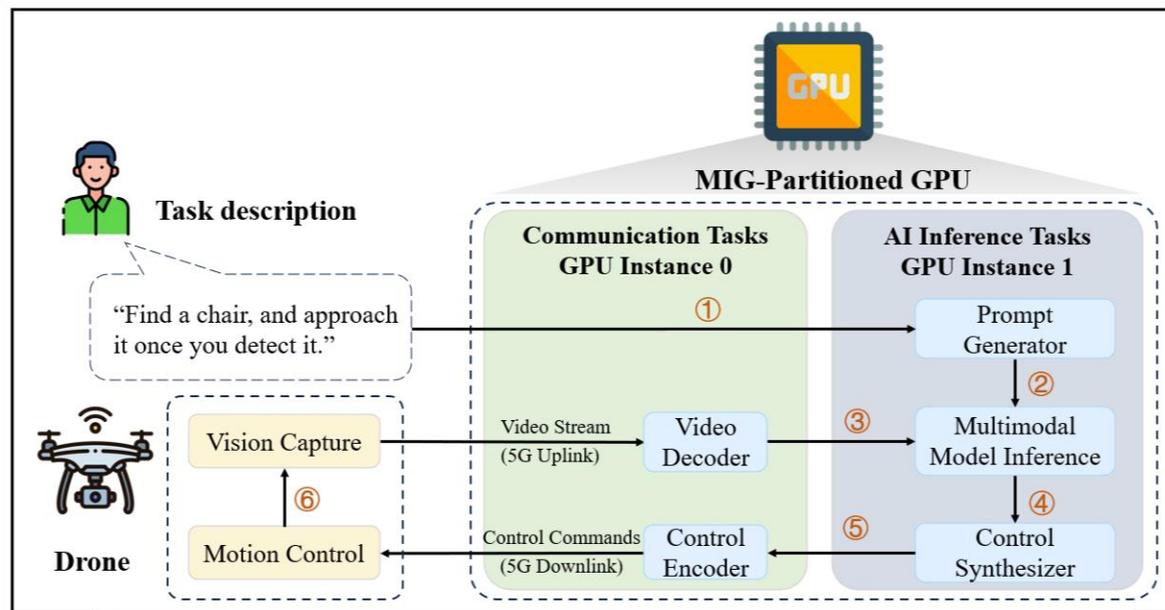
边缘通感算控一体化任务协同优化部署

自主式智能通算应用平台

总结与展望

自主式智能通算应用平台—Agentic AI-RAN

基于通用高性能GPU和通用算子库的自主式智能边缘网络（Agentic AI-RAN）节点：在同一边缘通用算力单元上部署高速宽带通信收发算法、DeepSeek-VL大模型，实现了端侧视觉信息实时感知、高速实时传输接入、边侧智能处理决策、精准反馈控制全过程一体化协同，完整实现了边缘网络智能通感算控一体化功能，初步验证了未来Agentic AI-RAN和Agentic AIoT的基本愿景



- ① High-level task instruction from user
- ② Prompt text generated for model inference
- ③ Visual input decoded from RTSP stream
- ④ Action plan inferred from multimodal model
- ⑤ Low-level commands generated for execution
- ⑥ Closed-loop feedback to vision

Prompt content:

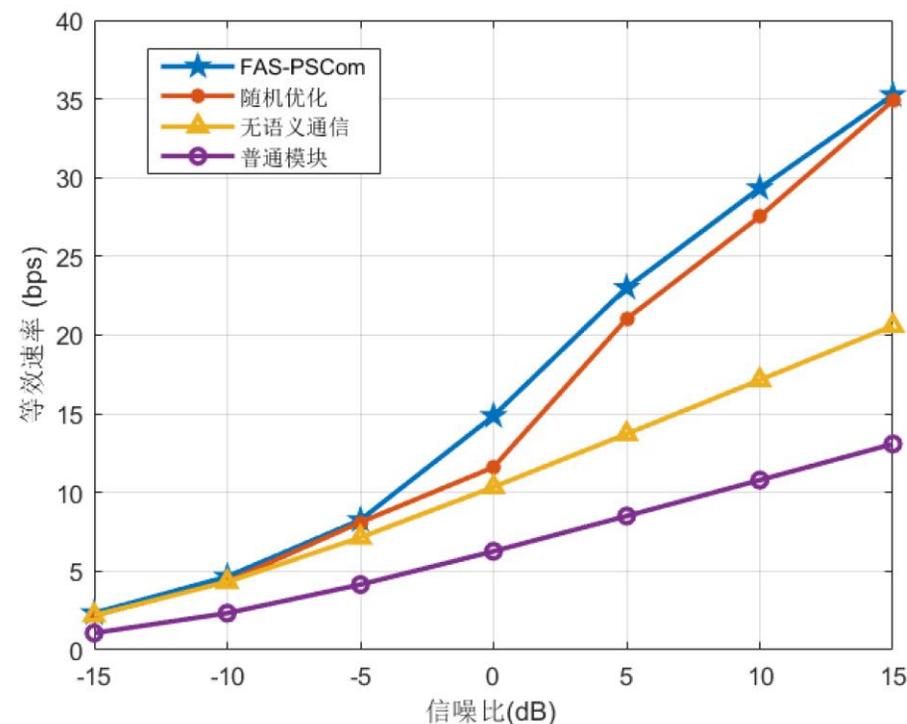
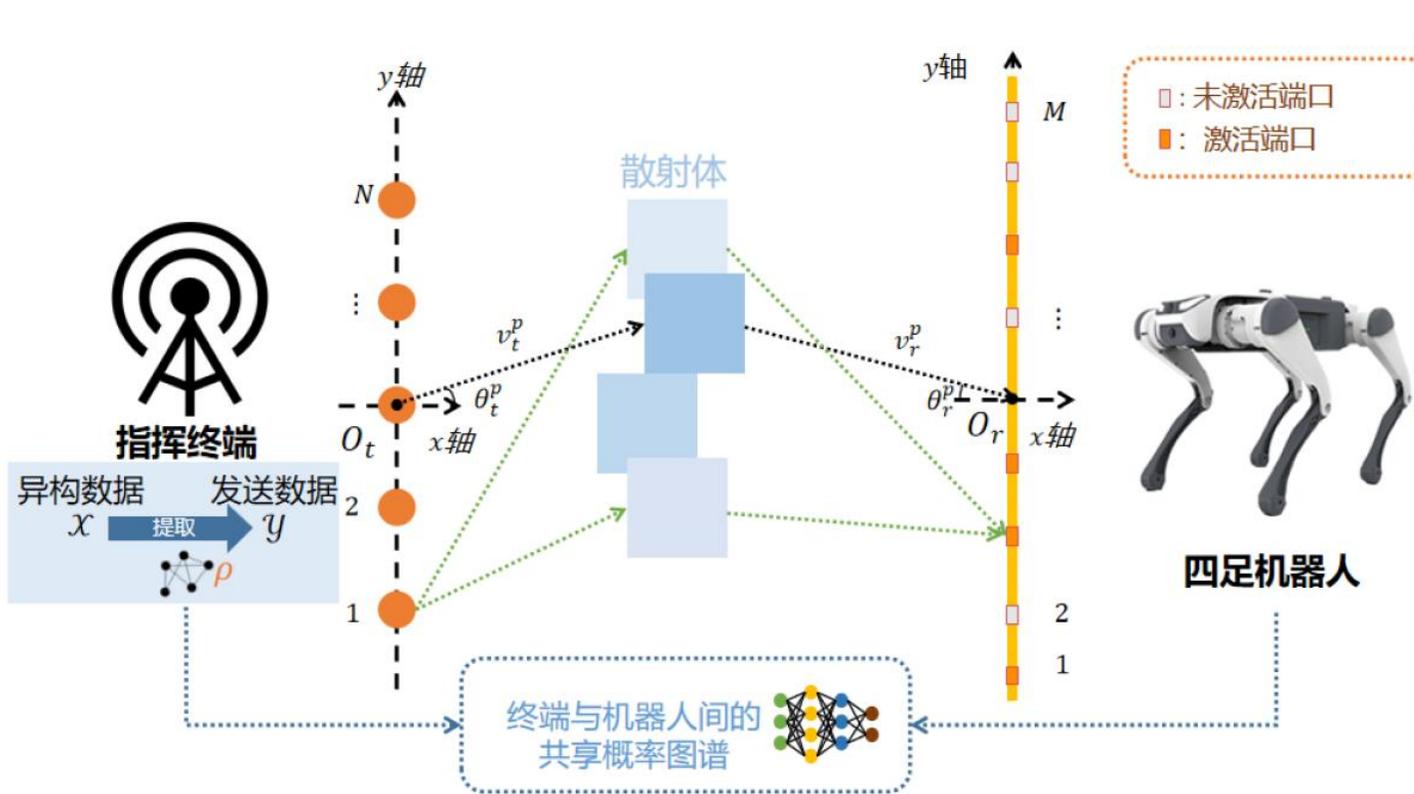
“Generate a navigation plan based on scene understanding and task intent.”

Action plan:

Precondition: control authority acquired
Action: turn right 30°, move forward 1m

自主式智能通算应用平台—四足机器人的端云协同

□ 基于平面波的像素阵列流体天线，传输的多模态数据（文本、声音、视频等）经过语义提取手段用高度凝练的知识图谱形式统一表示，有效降低了带宽压力和数据冗余，为实际应用中的大范围、复杂场景下的四足机器人协作作业提供了坚实而灵活的通信保障。



目录

研究背景和意义

自主式智能边缘网络一体化融合架构

边缘通感算控一体化任务协同优化部署

自主式智能通算应用平台

总结与展望

总结

- 自主式智能边缘网络一体化融合架构，在边缘深度部署AI和算力，并使AI模型架构与通信网络架构和计算架构紧密协同
- 边缘通感算控一体化任务协同优化部署，通信-计算并行拆分学习和无线大模型联邦低秩分解
- 自主式智能通算应用平台，任务驱动式智能交互
- 探索未来探索电磁效应启发的自主式智能通算机理

**非常感谢大家的聆听！
敬请批评指正！**