

Distributed multi-UAV cooperation in UAV-assisted mobile crowd sensing based on information exchange

Guisong Yang¹, Mengdie Liu¹, Xingyu He^{2}*

¹School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China

²College of Publishing, University of Shanghai for Science and Technology, Shanghai, China

*Corresponding Author. Email: sherri_he@163.com

Abstract. In UAV-assisted Mobile Crowd Sensing, ground workers are responsible for basic data collection, while Unmanned Aerial Vehicles (UAVs) can perform supplementary sensing in areas that are difficult to cover in a timely manner. To address the limited efficiency of distributed multi-UAV cooperative supplementary sensing in the absence of a central controller, this paper proposes an information-exchange-based distributed multi-UAV cooperative sensing algorithm, termed IE-DDQN. The proposed method formulates the problem as a decentralized partially observable Markov decision process, enables inter-UAV information exchange through short-range Device-to-Device (D2D) links, and employs a neighbor information pooling module to aggregate multi-source high-dimensional messages into low-dimensional representations, which are then combined with local observations for decision making. Experimental results demonstrate that the proposed method outperforms pure-worker and pure-UAV schemes in terms of weighted data value per unit composite cost, task completion rate, and task completion time.

Keywords: UAV-assisted mobile crowd sensing, multi-UAV cooperative sensing, information exchange, decentralized decision making, multi-agent reinforcement learning

1. Introduction

Mobile Crowd Sensing (MCS) leverages ground workers equipped with smart devices to collect environmental data, offering flexible deployment and wide coverage. It has been widely used in environmental monitoring, traffic surveillance, and public safety applications [1-3]. However, as task areas become larger and sensing scenarios more complex, relying solely on ground workers is often insufficient to achieve satisfactory spatiotemporal coverage and data quality, especially in areas that are difficult, dangerous, or time-consuming to access.

To address this limitation, UAV-assisted MCS has emerged as an important architecture. Owing to their mobility and line-of-sight advantages, UAVs can perform supplementary sensing in regions that are difficult for ground workers to reach, thereby improving sensing coverage and efficiency. Existing studies on UAV-assisted MCS mainly focus on task allocation and trajectory planning. For example, Wang et al. [4] optimize the amount and freshness of collected data under energy constraints, while Mao et al. [5] propose a

hierarchical deep reinforcement learning framework for global task allocation and local path planning. Although these methods improve UAV coverage efficiency and energy utilization, most of them treat ground sensing resources as cost-free auxiliary units and fail to explicitly capture the trade-off between the labor cost of ground workers and the energy consumption of UAVs in a unified framework.

In practical scenarios such as urban air quality monitoring and transmission line inspection, sensing systems often exhibit a ground-air hybrid sensing structure, in which ground workers remain at fixed collection points to provide basic sensing data, while multiple UAVs perform cross-regional supplementary sensing for uncovered or high-priority tasks [6]. In such a structure, ground collection incurs labor costs, whereas UAV-based supplementary sensing introduces additional energy consumption. Therefore, optimizing only UAV trajectories and energy efficiency is insufficient for evaluating the overall sensing performance under different divisions of labor between ground collection and aerial supplementary sensing.

A key challenge in this setting is to achieve efficient cooperation among multiple UAVs. Since each UAV can only access local observations and limited neighboring information, the problem naturally exhibits multi-agent, dynamic, and partially observable characteristics. Existing MARL-based methods often rely on centralized training, global state visibility, or direct concatenation of neighboring UAV states, which may lead to high communication overhead, poor scalability, and limited robustness in resource-constrained or communication-limited environments [7-10].

To address these issues, this paper studies a typical UAV-assisted MCS scenario in which ground workers continuously collect part of the required data at their task locations, while the remaining data are released as task requests to multiple UAVs for supplementary collection. Ground workers are modeled as fixed basic sensing nodes and do not participate in UAV decision-making. Under this setting, we aim to maximize the ratio of weighted data value to composite cost by jointly considering task priority, worker cost, and UAV energy consumption. Accordingly, we formulate the multi-UAV cooperative supplementary sensing problem as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP) and propose an information-exchange-based distributed multi-UAV cooperative sensing algorithm, termed IE-DDQN. The proposed method enables UAVs to exchange neighboring information through short-range Device-to-Device (D2D) links and introduces a neighbor information pooling module to aggregate multi-source high-dimensional messages into compact features for decision making. Experimental results show that the proposed method achieves better weighted data value per unit composite cost, task completion rate, and task completion time than pure-worker and pure-UAV schemes.

2. Related work

2.1. UAV-assisted mobile crowd sensing

To overcome the coverage gaps and uneven energy consumption caused by relying solely on ground users, a series of studies on UAV-assisted MCS have emerged in recent years. Zhou et al. [11] considers the collaboration between UAVs and mobile users to complete sensing tasks, jointly modeling task assignment and UAV trajectory planning as an energy-constrained optimization problem, and reducing system energy consumption through an approximation algorithm while ensuring the task completion rate. Pilloni et al. [12] focuses on the crowdsensing scenario involving multiple UAVs, constructs a task-UAV matching model, comprehensively considers task value, execution cost, and capability constraints, and achieves approximately optimal task allocation through heuristic search. Gao et al. [13] further extends tasks to a multi-subtask structure, proposes a UAV-assisted multi-task allocation method, and integrates task dependencies, flight constraints, and energy limitations into a mixed-integer programming framework. Regarding the issue of

coverage quality, Liu et al. [14] introduces coverage-aware indicators and designs a task allocation method with coverage constraints in UAV-assisted MCS scenarios to improve the sensing quality of key areas.

Overall, most of the aforementioned works treat ground participants as anonymous users, lacking explicit modeling of the continuous collection capability of long-term resident workers and labor costs. There are also few studies that quantitatively characterize the relationship between "worker collection ratio and UAV supplementary collection capability" under a unified system objective.

2.2. Multi-UAV cooperation and multi-agent reinforcement learning

In multi-UAV systems, as the scale of tasks and the complexity of the environment increase, traditional methods based on integer programming, heuristic search, or centralized optimization have gradually exposed problems such as high computational complexity and insufficient scalability. Skaltsis et al. [15] reviews multi-UAV task allocation methods and points out that centralized solving is difficult to adapt to large-scale systems in dynamic environments and large-scale systems. In recent years, MARL has gradually become an important tool for solving the problems of multi-UAV collaborative control and task allocation. Liu et al. [16] models the dynamic task allocation of heterogeneous multi-UAVs as a multi-agent decision-making problem and uses value function decomposition to achieve a balance between task revenue and energy consumption; Du et al. [17] utilizes hierarchical reinforcement learning to optimize task scheduling and motion control in a hierarchical manner, effectively improving system throughput and energy efficiency in data collection scenarios. Frattolillo et al. [18] summarizes deep reinforcement learning methods in multi-UAV systems from a review perspective, emphasizing that in multi-agent scenarios, it is necessary to take into account scalability, training stability, and communication overhead, and points out that constructing a low-dimensional and robust representation of neighbor information is one of the key issues in current research.

Existing multi-UAV collaborative work based on multi-agent reinforcement learning mostly assumes that tasks are completely undertaken by UAVs, without incorporating the long-term perception ability of ground workers, labor costs, and the division of labor between them and UAVs into a unified reinforcement learning framework; at the same time, neighbor UAV information is usually realized through simple concatenation, making the observation dimension grow linearly with the number of UAVs, which is difficult to adapt to changes in scale and link failures.

3. Methodology

3.1. System model

3.1.1. Ground-air hybrid sensing scenario

As shown in Figure 1, this paper considers a ground-air hybrid sensing scenario under UAV-assisted MCS and constructs a multi-UAV collaborative sensing system model in this scenario. In this hybrid sensing scenario, the system can be abstracted into three layers: the MCS platform layer, the multi-UAV sensing layer, and the ground worker sensing layer. The MCS platform layer is responsible for maintaining and updating the task pool, aggregating task requests reported by ground workers, and uniformly managing and distributing task information such as location, remaining data volume, priority, and completion status. The platform only provides task information services and does not perform real-time control over the actions of each UAV. The multi-UAV sensing layer consists of N UAVs, denoted as the set of UAVs $U = \{u_1, \dots, u_i, \dots, u_N\}$. UAVs autonomously decide on target tasks within the mission area and fly to designated sensing points to perform data collection. The ground worker sensing layer is composed of several ground workers who are permanently stationed at key locations and their corresponding collection tasks. The set of ground workers is

denoted as $W = \{w_1, \dots, w_k, \dots, w_M\}$. For the convenience of modeling, it is assumed that each ground worker w_k is responsible for only one fixed task τ_k , and there is a one-to-one correspondence between ground workers and tasks. Therefore, the set of data collection tasks is denoted as $T = \{\tau_1, \dots, \tau_k, \dots, \tau_M\}$, where $|T| = M$, which is consistent with the size of the ground worker set. Each task $\tau_k \in T$ corresponds to a spatial position $p^{\tau_k} = (x^{\tau_k}, y^{\tau_k})$, a total data volume d_{total} , and a priority pr^{τ_k} . Ground workers are responsible for continuously collecting basic data locally. When their own collection capacity is insufficient to complete the task within the specified time limit, they will report the remaining data requirements to the platform in the form of task requests, which will be supplemented by multiple UAVs.

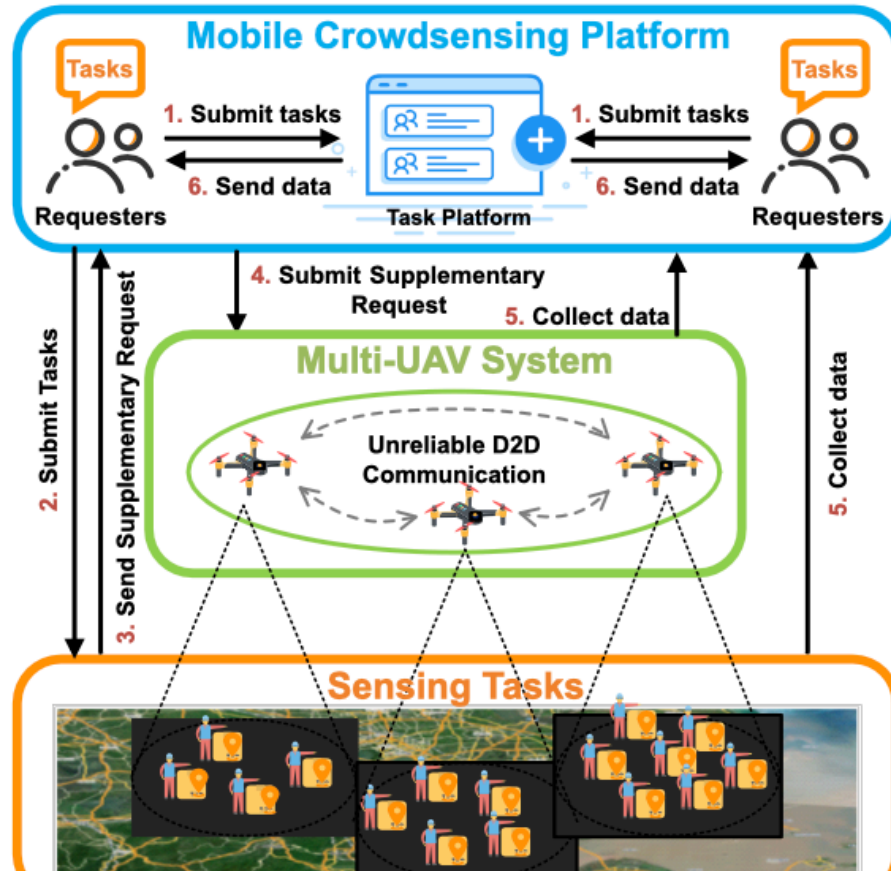


Figure 1. Ground-air hybrid sensing scenario for UAV-assisted MCS

The system adopts discrete time slot scheduling, denoted as $t = \{0, 1, \dots, T - 1\}$, and the duration of each time slot is δ_t . UAVs perform tasks in consecutive time slots until the maximum duration T is reached. The two-dimensional position of UAV u_i , at time t is denoted as $q_t^{u_i} = (x_t^{u_i}, y_t^{u_i})$, and its movement is constrained by the maximum flight speed v_{max} ; when the distance between the UAV and the task location is less than the sensing radius r_s , it can collect the remaining data of the task. It is assumed that UAVs have sufficient storage space to temporarily store the collected data, and energy consumption mainly comes from flight and hovering sensing. To simplify the modeling, it is assumed that all UAVs start from the same initial position at the center of the mission area at time t_0 ; they have sufficient power within the maximum duration T , and mid-way charging and return scheduling after energy exhaustion are not considered.

3.1.2. Data collection and cost model

In the considered system, ground workers and UAVs jointly complete the sensing tasks. Ground workers provide stable local sensing at fixed task locations, while UAVs perform cross-regional supplementary sensing. To quantify the division of labour between ground collection and aerial supplementary sensing, we introduce a data-sharing coefficient $\alpha \in [0, 1]$. Accordingly, the data demand of each task is decomposed into a ground-collected part and a UAV-collected part as

$$d_{total}^W = \alpha \cdot d_{total} = \sum_{k=1}^M d^{w_k} \tag{1}$$

$$d_{total}^U = (1 - \alpha) \cdot d_{total} = \sum_{i=1}^N d^{u_i} \tag{2}$$

Here, d_{total}^W denotes the amount of data completed by ground workers, and d_{total}^U denotes the amount of data to be completed by UAVs. A larger α implies stronger reliance on ground collection, whereas a smaller α indicates that more data are assigned to UAVs.

The sensing value contributed by ground workers is evaluated by the priority-weighted amount of collected data. It is defined as

$$V_W = \sum_{k=1}^M pr^{\tau_k} \cdot d^{w_k} \tag{3}$$

Meanwhile, the labour cost of ground workers is modelled as a linear function of the collected data volume, i.e.,

$$C_W = \lambda_W \cdot \sum_{k=1}^M d^{w_k} = \lambda_W \cdot \alpha \cdot d_{total} \tag{4}$$

where $\lambda_W > 0$ is the unit labor cost coefficient.

For UAV supplementary sensing, a mutual exclusion mechanism is adopted to avoid redundant collection. Specifically, at any time, each task can be collected by at most one UAV. Let $I_t^{u_i, \tau_k} \in \{0, 1\}$ denote whether UAV u_i collects data from task τ_k at time slot t . Then the remaining data demand of task τ_k is updated as

$$d_{t+1}^{u_i} = \max(d_t^{u_i} - r \cdot \delta_t \cdot I_t^{u_i, \tau_k}, 0) \tag{5}$$

Here, r is the UAV sensing rate. Once the remaining data demand becomes zero, the task is regarded as completed. To characterize the timeliness of UAV sensing, the value of UAV-collected data is discounted according to task completion time. Thus, the sensing value obtained by UAVs is written as

$$V^U = d_t^{\tau_k} \cdot pr^{\tau_k} \cdot e^{-\beta t_{finish}^{\tau_k}}, \tau_k \in F \tag{6}$$

where $t_{finish}^{\tau_k}$ denotes the completion time of task τ_k , $F = \{k | t_{finish}^{\tau_k} > 0\}$ is the completed task set, and β is the discount coefficient. The total system sensing value is then given by

$$V_{total} = V_W + V^U \tag{7}$$

The energy consumption of UAVs includes flight energy and communication energy. Let $E_{u_i}^{fly}(t)$ and $E_{u_i}^{comm}(t)$ denote the flight and communication energy consumption of UAV u_i at slot t , respectively. Then the slot-wise energy consumption is written as

$$E_{u_i}^t = E_{u_i}^{fly}(t) + E_{u_i}^{comm}(t) \tag{8}$$

and the total UAV energy consumption over the whole task cycle is

$$E_{total} = \sum_{i=1}^N \sum_{t=0}^{T-1} E_{u_i}^t \tag{9}$$

3.2. Problem formulation

Based on the above system model, the objective is to maximize the sensing utility of the ground-air hybrid sensing system under worker cost and UAV energy consumption constraints. To evaluate different divisions of labour and cooperative sensing strategies in a unified manner, we define the system utility as the ratio of weighted data value to composite cost:

$$U = \frac{V_{total}}{C_W + \theta \cdot E_{total}/10^3} \quad (10)$$

where θ is the energy cost coefficient of UAVs. Accordingly, the optimization problem is formulated as $maxU$ subject to the data collection dynamics, mutual exclusion constraints, sensing range constraints, and UAV motion constraints defined in the system model. Since multiple UAVs must make distributed decisions in a dynamic and partially observable environment, the problem is highly nonlinear and difficult to solve using conventional centralized optimization methods. Therefore, we further reformulate it as a decentralized Partially Observable Markov Decision Process (Dec-POMDP) and solve it using a distributed multi-UAV cooperative sensing algorithm.

3.3. Dec-POMDP formulation

The considered multi-UAV cooperative supplementary sensing problem is formulated as a Dec-POMDP, represented by the tuple $\langle N, S, A, P, R, \Omega, O, \gamma \rangle$, where N is the set of agents, S is the global state space, A is the joint action space, P denotes the state transition function, R is the reward function, Ω is the joint observation space, O is the observation function, and γ is the discount factor.

The global state contains both UAV states and task states, and can be written as

$$s_t = (\{s_t^{u_i}\}_{i=1}^N, \{\{s_t^{\tau_k}\}_{k=1}^M\}) \quad (11)$$

where $s_t^{u_i} = [x_t^{u_i}, y_t^{u_i}, b_t^{u_i}, \kappa_t^{u_i}, \rho_t^{u_i}]$ includes the position, residual energy, assigned task, and execution progress of each UAV, while $s_t^{\tau_k} = [x_t^{\tau_k}, y_t^{\tau_k}, d_{total}^{\tau_k}, pr^{\tau_k}, u_i^{\tau_k}, \rho_t^{\tau_k}, \sigma_t^{\tau_k}]$ includes task location, remaining data demand, priority, assignment status, and completion status.

Since each UAV cannot directly access the complete global state, the problem is inherently partially observable. The local observation of UAV u_i at time t consists of three parts, namely its own state, task information released by the platform, and communication information received from neighbouring UAVs. Thus, the local observation is expressed as

$$o_t^{u_i} = [o_t^{u_i}(self), o_t^{u_i}(tasks), o_t^{u_i}(comms)] \quad (12)$$

where $o_t^{u_i}(self) = [x_t^{u_i}, y_t^{u_i}, b_t^{u_i}, \kappa_t^{u_i}]$ represents the self-state information of the UAV u_i ; $o_t^{u_i}(tasks) = \{\{[x_t^{\tau_k}, y_t^{\tau_k}, \sigma_t^{\tau_k}]\}_{k=1}^M\}$ represents the position and completion status information of all data collection tasks stably issued by the platform. To improve the coordination of distributed decision-making, each UAV broadcasts its position at the previous time slot and the selected task to execute through a short-range D2D link at each time slot. After neighbouring UAVs receive this information, they generate low-dimensional interaction information $o_t^{u_i}(comms)$ through the neighbouring UAV information pooling module.

At each decision slot, each UAV u_i must select a discrete action from the action space $A_i = \{a_0, a_1, \dots, a_{m-1}, a_{continue}, a_{drop}, a_{wait}\}$, which respectively represent selecting a certain task, continuing the current task, abandoning, and hovering to wait for the next action selection. To reduce invalid exploration and concurrent conflicts, we introduce an action mask:

$$A_{valid}^{u_i}(t) = \{a \in A | Mask_t^{u_i}(a) = 1\} \quad (13)$$

We have considered the following action constraints: tasks that have been completed cannot be selected; when there is no current task, the action of "continuing the current task" cannot be selected.

To guide cooperative supplementary sensing, the reward function jointly considers task completion, execution continuity, regional consistency, and task conflicts. The immediate reward of UAV u_i at time t is defined as

$$R_t^{u_i} = \begin{cases} R_{complete}^{u_i} + R_{continue}^{u_i} + R_{operate}^{u_i}, & \text{if not conflict} \\ R_{invalid}^{u_i}, & \text{if conflicts} \end{cases} \quad (14)$$

where $R_{complete}^{u_i}$ encourages the early completion of tasks with large data volumes and high priorities through weighted gains with time decay:

$$R_{complete}^{u_i} = C_0 \left(1 + \frac{d_{total}^{\tau_k} - d_{mid}}{d_{mid}} \cdot \frac{pr^{\tau_k}}{5} \right) \quad (15)$$

where d_{mid} is the median used to normalize the task data volume, and the constant C_0 is the basic completion reward. Secondly, to encourage UAVs to form continuous collection trajectories in space, a reward $R_{continue}^{u_i}$ is given when the distance between tasks is short ($d_{last \rightarrow \tau_k} \leq D_{max}$), prompting UAVs to prioritize completing nearby unfinished tasks, thereby naturally forming the behavior of continuous execution within a partition. To balance simplicity and implementability, regional consistency, load balancing, and path efficiency are combined into $R_{operate}^{u_i}$. This term tends to keep UAVs within their assigned areas, suppress excessive crowding in local areas, and encourage the selection of nearby tasks through distance normalization preferences. If multiple UAVs choose the same task, resulting in a conflict, they will receive a strong penalty $R_{invalid}^{u_i}$. Finally, to avoid extreme values from affecting training stability and gradient scales, we clip and scale the immediate reward of each UAV:

$$R_t^{u_i} \leftarrow clip(R_t^{u_i}, R_{min}, R_{max}) \quad (16)$$

3.4. Neighbour information pooling module

Directly concatenating the complete states of neighbouring UAVs causes the observation dimension to grow linearly with the number of agents, leading to high communication overhead and poor scalability. To address this issue, we design a neighbour information pooling module that transforms multi-source neighbouring messages into a fixed-dimensional communication feature.

At each time slot, UAV u_i receives a message from neighboring UAV $u_{i'}$, denoted by $m_t^{u_{i'}}$. For each neighbor, UAV u_i maintains a sliding window of length H to store the most recent messages. The windowed message sequence is written as

$$H_t^{u_{i'}} = \{m_{t-l}^{u_{i'}}\}_{l=0}^{H-1} \quad (17)$$

To highlight the geometric and semantic information related to task selection, for each message within the window, we first normalize and concatenate the sender's last position and task-related features using $g(\cdot)$ to encode structured features:

$$f_t^{u_{i'}} = g(m_t^{u_{i'}}) = \left[\underbrace{x_{t-1}^{\tau_k}, y_{t-1}^{\tau_k}, d_{total}^{\tau_k}, pr^{\tau_k}}_{\text{last_action_task}}, \underbrace{x_{t-1}^{u_i}, y_{t-1}^{u_i}}_{\text{last_position}} \right] \quad (18)$$

Subsequently, the temporal aggregation operator $\phi_H(\cdot)$ is applied to the feature sequence of the same neighbouring UAV within the window $H_t^{u_{i'}}$, and a 12-dimensional pooled feature is generated by combining

"current features + historical averages". The observation of interaction information from UAV u_i to neighbouring UAV $u_{i'}$ can be denoted as $z_t^{u_i, u_{i'}}$, and thus $o_t^{u_i}(comms)$ is:

$$z_t^{u_i, u_{i'}} = \phi_H(\{f_{t-l}^{u_{i'}}\}_{l=0}^{H-1}) = [f_t^{u_{i'}}, \frac{1}{H} \sum_{l=0}^{H-1} f_{t-l}^{u_{i'}}] \in R^{12} \quad (19)$$

$$o_t^{u_i}(comms) = \{z_t^{u_i, u_{i'}} | u_{i'} \in U_{-u_i}\} \quad (20)$$

where U_{-u_i} represents other UAVs in the UAV set except u_i . Finally, $o_t^{u_i}(comms)$, as the communication-related observation component, together with its own state observation $o_t^{u_i}(self)$ and task information $o_t^{u_i}(tasks)$, constitutes the input of IE-DDQN, realizing the modeling and utilization of the historical behavior information of neighboring UAVs.

3.5. Proposed IE-DDQN algorithm

Based on the aforementioned system modeling and Dec-POMDP modeling, this paper designs a distributed IE-DDQN collaborative scheduling algorithm based on information interaction for the ground-air hybrid sensing scenario in UAV-assisted MCS. The illustration of the IE-DDQN algorithm architecture is shown in Figure 2.

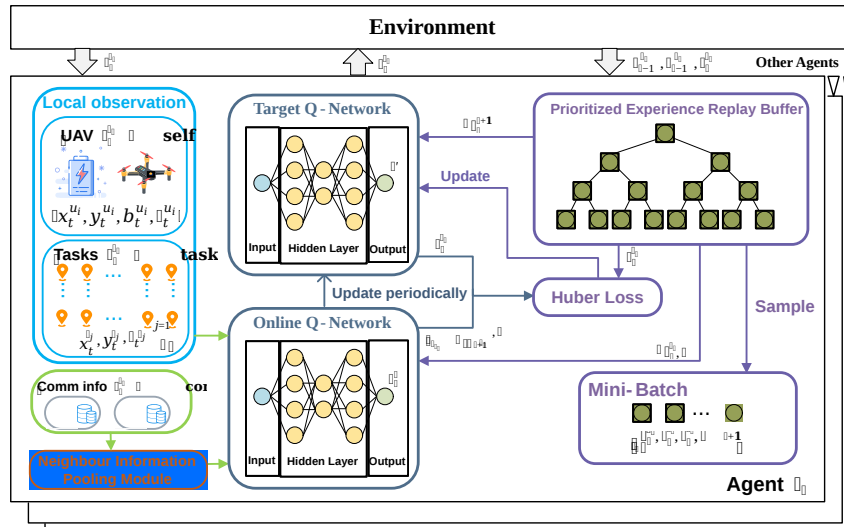


Figure 2. Illustration of the IE-DDQN algorithm architecture

There is no central controller in the system. Each UAV u_i independently maintains the online Q-network parameters θ_{u_i} , target Q-network parameters $\bar{\theta}_{u_i}$, local prioritized experience replay buffer D_{u_i} , and neighboring UAVs' historical cache $H_t^{u_{i'}}$. It completes action selection and parameter update based on its own observations and the pooled results of neighboring UAVs' information. Multi-UAVs only broadcast their position and task selection from the previous moment through short-range D2D information interaction, achieving distributed collaborative decision-making under lightweight information exchange.

At the start of each training episode, the platform generates an initial task pool based on task requests reported by ground workers, including information such as task location, total data volume, and priority, and broadcasts it to all UAVs. Meanwhile, it resets the position, battery level, and current task status of each UAV, and clears or initializes the local experience replay buffer D_{u_i} and neighboring UAVs' historical cache $H_t^{u_{i'}}$. After that, the UAVs continuously interact with the environment within a given time, repeatedly performing information exchange, local decision-making, and local learning.

At any decision slot t , UAV u_i first processes the latest communication messages from neighboring UAVs. Each UAV periodically broadcasts its position information and the number of the task it executed from the previous moment. UAV u_i writes the message $m_t^{u_{i'}}$ from neighboring UAV $u_{i'}$ into a sliding window $H_t^{u_{i'}}$, and performs feature mapping and temporal aggregation on the message sequence within the window to obtain the behavioral characteristics of the neighboring UAV over a period of time. Subsequently, a fixed-dimensional neighboring UAV information vector $o_t^{u_i}(\text{comms})$ is obtained as the observation component related to information interaction. Combining its own position, battery level, current task status, and the task pool information broadcast by the platform, UAV u_i constructs the local enhanced observation vector $\hat{o}_t^{u_i}$ at the current moment, thereby supporting local decision-making with the help of neighboring UAVs' historical behaviors and global task information without relying on the global state.

During the action selection phase, the UAV u_i first constructs an action mask $A_{\text{valid}}^{u_i}(t)$ based on the task completion status and the degree of interaction information with neighboring UAVs, excluding actions corresponding to completed tasks and those being executed by neighboring UAVs. Subsequently, it explores randomly with a probability of ε ; otherwise, it selects the action that maximizes the current network output $Q_{\theta_{u_i}}(\hat{o}_t^{u_i}, a)$. After the environment returns the immediate reward $R_t^{u_i}$ and the next observation $\hat{o}_{t+1}^{u_i}$, the transition sample $(\hat{o}_t^{u_i}, a_t^{u_i}, R_t^{u_i}, \hat{o}_{t+1}^{u_i})$ is written into the local prioritized experience replay buffer D_{u_i} after determining its priority based on the Temporal Difference (TD) error, providing samples for subsequent parameter updates.

To suppress Q-value overestimation and improve training stability, IE-DDQN adopts a dual-network structure and a prioritized experience replay mechanism. The target calculation is as follows:

$$y_t^{u_i} = R_t^{u_i} + \gamma Q_{\bar{\theta}_{u_i}}(\hat{o}_{t+1}^{u_i}, \arg Q_{\theta_{u_i}}(\hat{o}_{t+1}^{u_i}, a')) \quad (21)$$

where θ_{u_i} and $\bar{\theta}_{u_i}$ represent the parameters of the online network and the target network, respectively, and γ is the discount factor. To enhance training stability, the loss function adopts a weighted form based on the Huber function:

$$L_{u_i}(\theta_{u_i}) = E_{h \sim D_{u_i}}[\omega_h^{u_i} \cdot H(y_h^{u_i} - Q_{\theta_{u_i}}(\hat{o}_h^{u_i}, a_h^{u_i}))] \quad (22)$$

where $H(\cdot)$ is the Huber function, which maintains a quadratic form when the error is small and switches to a linear penalty when the error is large, thereby suppressing gradient instability caused by abnormal samples. $\omega_h^{u_i}$ is the importance sampling weight, determined by the sampling probability of Prioritized Experience Replay (PER), to highlight key experiences and accelerate convergence. h denotes the h -th experience sample sampled from the replay buffer. After each update, the target network maintains stable iteration through soft update $\bar{\theta}_{u_i} \leftarrow \tau \theta_{u_i} + (1 - \tau) \bar{\theta}_{u_i}$.

Overall, the proposed IE-DDQN algorithm, on the one hand, fully utilizes the interaction of task and position information between UAVs, and on the other hand, maintains the scalability of the algorithm when the number of UAVs changes and the task scale expands, providing a stable and easily deployable method for distributed multi-UAV collaboration in UAV-assisted MCS scenarios.

4. Experimental results and analysis

To verify the effectiveness and robustness of the proposed IE-DDQN multi-UAV collaborative perception method, this paper conducts a series of comparative experiments in the UAV-assisted MCS scenario to evaluate the system performance under different perception structures and different UAV data collection ratios. The key parameters of the experiments are set as shown in Table 1. All experiments use the weighted data value per unit comprehensive cost as the system utility index, and simultaneously count the task completion

time and the overall task completion rate, so as to conduct a comprehensive analysis of the algorithm from three dimensions: revenue, timeliness, and cost.

Table 1. Illustration of the IE-DDQN algorithm architecture

Parameters	Notations	Values
Number of ground workers and tasks	M	40
Number of UAVs	N	4
Duration of each episode	T	400s
Duration of each time slot	δ_t	1s
Data collection radius	r_0	8m
Learning Rate	l_r	3×10^{-4}
Discount factor	γ	0.95
Batch Size	B	64
Replay Buffer Size	$ D $	50,000
Exploration decay rate	ε_{decay}	0.9995
Soft Update Parameter	τ	0.005

During the training phase, multiple UAVs run the IE-DDQN algorithm in parallel, collecting experience through interaction with the experimental environment and updating the local Q-network. Specifically, the time average of the tail segment of each training curve is first calculated, and then the mean and standard deviation are statistically computed over 10 groups of mutually independent random seeds to reduce the impact of accidental fluctuations in a single experiment on the results. It should be specifically noted that PER is mainly positioned as a training accelerator here: it accelerates the efficiency of gradient updates and shortens the convergence time by prioritizing the sampling of experience samples with larger TD errors, but it does not change the upper limit of the long-term steady-state performance of the algorithm. Therefore, PER is enabled by default in all experiments, so that the comparison of algorithm performance focuses more on long-term stability and robustness rather than mere convergence speed.

4.1. Convergence analysis of multi-UAV cooperation

This section first examines the training convergence characteristics of IE-DDQN under a fixed division of labor ratio between ground collection and aerial supplementary collection (with a 50% proportion of UAV data collection). Figure 3 shows the average reward per UAV accumulated per episode during the training process of multiple UAVs, where each curve corresponds to one UAV.

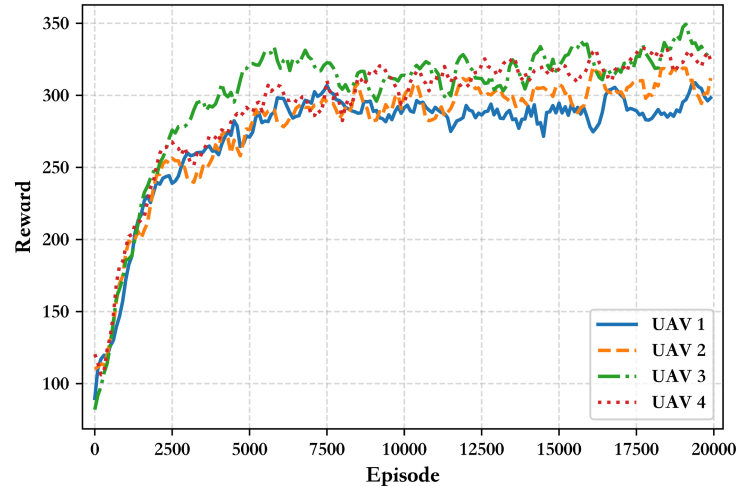


Figure 3. Training convergence reward curve of IE-DDQN

As can be seen from Figure 3, in the early stages of training, the reward levels of each UAV were low and fluctuated significantly, reflecting that the strategy was still in the stage of random exploration. At this time, task selection and flight trajectories had not yet formed a stable pattern. The neighbour information pooling module effectively reduces communication overhead and computational complexity through low-dimensional aggregation of multi-source interaction messages, enabling each UAV to efficiently utilize the historical behaviour characteristics of neighbouring UAVs to optimize decision-making strategies. As the number of training episodes increases, IE-DDQN gradually learns to select high-value, low-conflict tasks under the joint guidance of local states and neighbour information. Approximately after 2,000 episodes, each curve began to rise rapidly and stabilized around 10,000 episodes. Finally, the average reward per UAV stabilized around 300, with small differences between different UAVs, indicating that the proposed algorithm can achieve good load balancing and strategy consistency under the distributed training framework.

4.2. Analysis of different perception-execution modes

After verifying that IE-DDQN exhibits good convergence in ground-air hybrid sensing scenarios, this section further compares its performance with two single-agent sensing schemes: pure ground workers and pure UAVs. This paper focuses on the collaborative relationship between ground collection and air supplementary collection under the goals of unified modeling and system optimization. However, existing reinforcement learning methods are mostly designed for centralized scheduling or different task models, making it difficult to directly and fairly compare them in the current problem scenario. Therefore, we consider a comparative analysis of the following three typical sensing execution modes:

- a) Pure ground worker sensing: relying solely on ground workers to perform all sensing tasks, with UAVs not participating in the data collection process;
- b) Pure UAV sensing: all sensing tasks are independently completed by UAVs, corresponding to scenarios in traditional UAV-assisted MCS where the contributions of ground workers are ignored;
- c) Ground-air hybrid sensing: using the IE-DDQN algorithm to achieve distributed scheduling, where ground workers are responsible for the regular collection of basic data in their stationed areas, and multiple UAVs undertake cross-regional supplementary collection of remaining tasks.

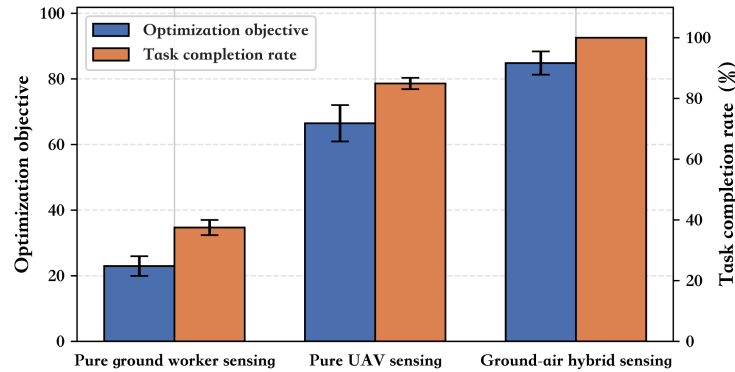


Figure 4. Optimization objective and task completion rate under different perception-execution modes

Figure 4 compares the system optimization objective values and task completion rates of the three schemes under the same resource constraints. The left vertical axis represents the weighted data value corresponding to the unit comprehensive cost, and the right vertical axis represents the task completion rate. It can be seen from Figure 4 that the pure ground worker perception scheme is limited by the movement speed and spatial distribution of ground workers, making it difficult to timely cover tasks far away from their resident areas, resulting in a low overall system benefit. Although the pure UAV perception scheme has greatly improved the task coverage capability, and its corresponding optimization objective value and task completion rate are significantly better than those of the pure worker scheme, since UAVs need to take full responsibility for all data collection tasks, the energy consumption during flight and hovering increases sharply, which limits the improvement of data value under unit comprehensive cost and makes it difficult to achieve the optimal balance between efficiency and cost.

In the ground-air hybrid perception scheme, ground workers give priority to completing regular perception tasks in their stationed areas and adjacent surrounding areas, while UAVs focus on high-value tasks and long-distance supplementary collection tasks. The task completion rate of this scheme is close to 100%, indicating that this collaborative structure can complete most perception tasks within a limited time. At the same time, its optimization objective value is the best among the three schemes, achieving a better balance among data value, worker cost and UAV energy consumption. The above results show that by adopting the ground-air hybrid perception structure combined with the distributed IE-DDQN multi-UAV collaborative scheduling strategy, the performance of this scheme is significantly better than the single-agent perception of pure workers or pure UAVs.

4.3. Analysis of the impact of different UAV data collection ratios

On the basis of determining the adoption of a ground-air hybrid perception structure and keeping the number of workers and drones fixed, the proportion of data collection undertaken by drones has a significant impact on system efficiency. In this section, under the premise that the number of workers and drones remains unchanged, the proportion of data collection undertaken by drones is gradually increased from 30% to 70%. For each configuration, the IE-DDQN algorithm is independently trained, and the system optimization target value after the algorithm converges is calculated. The experimental results are shown in Figure 5 and 6, where the error bars represent the standard deviation of multiple repeated experiments.

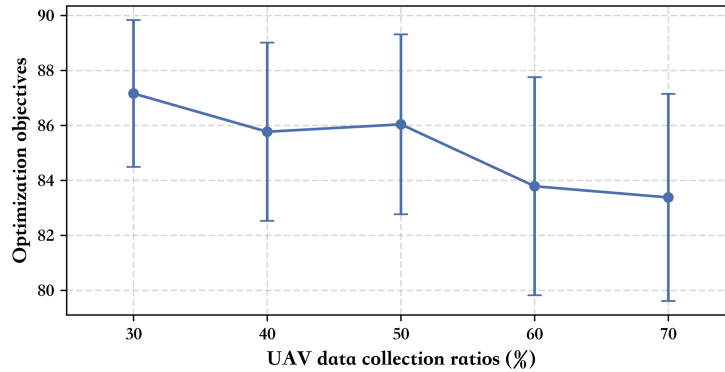


Figure 5. Optimization objective under different UAV data collection ratios

Figure 5 shows the changes in the optimization objective under different UAV data collection ratios. It can be seen from the results that when the UAV data collection ratio is within the range of [30%, 70%], the optimization objective remains at a relatively high level of 83-87, which is significantly better than the schemes using only workers or only UAVs. This indicates that regardless of the fluctuation in the proportion of data volume undertaken by UAVs, as long as the ground-air hybrid sensing mode is adopted, the overall efficiency of the system can stabilize on a "high-yield platform" and will not decline significantly due to small adjustments in the division of labor ratio. Further analysis shows that with the increase in the UAV data collection ratio, the optimization objective shows a slow downward trend. When the UAV data collection ratio is low (e.g., 30%), ground workers can complete most routine collection tasks nearby, while UAVs focus on cross-regional supplementary collection and high-priority tasks. This ensures coverage capability while effectively controlling the energy consumption for flight and hovering, thus achieving the highest weighted data value per unit comprehensive cost. However, when the UAV data collection ratio continues to increase, more tasks that could have been completed on-site by workers are transferred to UAVs. The marginal benefit of the additional flight energy consumption gradually decreases, and the extra flight energy consumption accumulates continuously, eventually leading to a slight decline in the optimization objective. Overall, a moderate UAV data collection ratio (e.g., 40%-50%) maintains high yields while avoiding unnecessary energy consumption, making it a relatively economical configuration range.

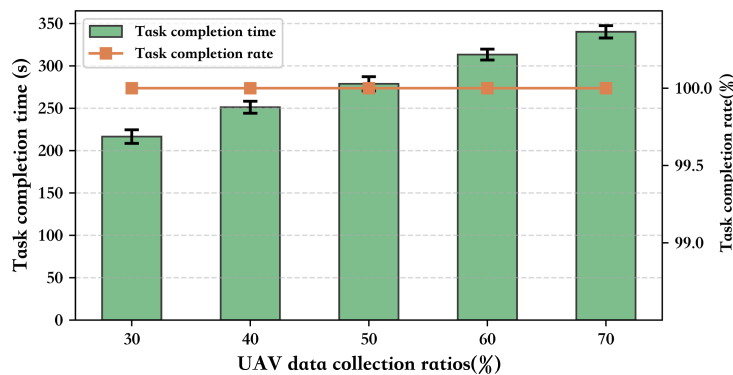


Figure 6. Task completion time and task completion rate under different UAV data collection ratios

To further analyze the impact of the UAV data collection ratio on system timeliness and reliability, Figure 6 presents the task completion time and task completion rate under different ratios. The bar chart in the figure

corresponds to the task completion time, the line chart corresponds to the task completion rate, and the error bars also represent the standard deviation of multiple experiments.

As can be seen from Figure 6, as the proportion of data collected by UAVs increases from 30% to 70%, the overall task completion time shows an upward trend. That is, with the total amount of data fixed, assigning more tasks to UAVs will prolong the overall completion time. This is because some tasks that were originally quickly completed by workers near their stationed positions are transferred to UAVs, which need to take off, fly, and hover, resulting in an increase in the overall maneuvering time and trajectory length. On the other hand, the task completion rate remains stable at nearly 100% under all configurations and hardly changes with the proportion of data collected by UAVs. This indicates that under the scheduling of IE-DDQN, the system can ensure a task completion rate close to 100% within a wide range of UAV data collection ratios, showing good reliability.

Overall, the change in the proportion of data collected by UAVs is mainly reflected in the dynamic trade-off between cost and timeliness: moderately increasing the proportion of data collected by UAVs can enhance the cross-regional supplementary collection capability, but an excessively high proportion will increase energy consumption and time costs. The task completion rate remains close to 100% across the entire proportion range, which further verifies the effectiveness and robustness of the proposed ground-air hybrid perception and IE-DDQN algorithm. A moderate proportion of data collected by UAVs (such as 40% - 50%) can achieve a more balanced multi-objective trade-off among benefits, costs, and timeliness.

5. Conclusion

This paper focuses on the ground-air hybrid sensing scenario under UAV-assisted MCS, constructing a system utility model that includes task data value, ground worker costs, and UAV energy consumption, and models multi-UAV collaborative scheduling as a Dec-POMDP. To achieve efficient distributed multi-UAV collaboration, a neighboring UAV information pooling module and an information interaction-driven distributed multi-UAV collaborative sensing algorithm, IE-DDQN, are designed, which is suitable for scenarios with insufficient cellular network coverage or damaged infrastructure. Experiments show that IE-DDQN has good convergence and robustness, and significantly outperforms the pure worker and pure UAV baseline schemes in core indicators such as weighted data value per unit comprehensive cost, task completion rate, and task completion time. Moreover, within the range of 30% to 70% UAV data collection ratio, it always maintains a task completion rate close to 100%.

Future work will introduce more typical distributed multi-UAV collaborative algorithms for system comparison under a unified Dec-POMDP modeling framework, and further verify the applicability of the proposed method in more complex sensing scenarios.

References

- [1] Guo, X., Huang, F., Yang, D., Tu, C., Yu, Z., & Guo, W. (2023). Spatiotemporal fracture data inference in sparse mobile crowdsensing: A graph-and attention-based approach. *IEEE/ACM Transactions on Networking*, 32(2), 1631-1644.
- [2] Li, M., Ma, M., Wang, L., Yang, B., Wang, T., & Sun, J. (2022). Multitask-oriented collaborative crowdsensing based on reinforcement learning and blockchain for intelligent transportation system. *IEEE Transactions on Industrial Informatics*, 19(9), 9503-9514.
- [3] Rubio-Aparicio, J., & Santa, J. (2022). An embedded crowdsensing unit for mobile urban pollution monitoring. *IEEE Communications Magazine*, 61(1), 90-96.

-
- [4] Wang, H., Liu, C. H., Yang, H., Wang, G., & Leung, K. K. (2023). Ensuring threshold AoI for UAV-assisted mobile crowdsensing by multi-agent deep reinforcement learning with transformer. *IEEE/ACM Transactions on Networking*, 32(1), 566-581.
- [5] Mao, X., Wu, G., Fan, M., Cao, Z., & Pedrycz, W. (2024). DL-DRL: A double-level deep reinforcement learning approach for large-scale task scheduling of multi-UAV. *IEEE Transactions on Automation Science and Engineering*, 22, 1028-1044.
- [6] Sun, L., Wang, J., Wan, L., Li, K., Wang, X., & Lin, Y. (2024). Human-UAV interaction assisted heterogeneous UAV swarm scheduling for target searching in communication denial environment. *IEEE Transactions on Automation Science and Engineering*, 22, 4457-4472.
- [7] Li, B., Fei, Z., & Zhang, Y. (2018). UAV communications for 5G and beyond: Recent advances and future trends. *IEEE Internet of Things Journal*, 6(2), 2241-2263.
- [8] Deng, L., Gong, W., Liwang, M., Li, L., Zhang, B., & Li, C. (2024). Towards intelligent mobile crowdsensing with task state information sharing over edge-assisted UAV networks. *IEEE Transactions on Vehicular Technology*, 73(8), 11773-11788.
- [9] Yan, S., Feng, J., & Pan, F. (2024). A distributed task allocation method for multi-UAV systems in communication-constrained environments. *Drones*, 8(8), 342.
- [10] Dai, Z., Liu, C. H., Han, R., Wang, G., Leung, K. K., & Tang, J. (2021). Delay-sensitive energy-efficient UAV crowdsensing by deep reinforcement learning. *IEEE Transactions on Mobile Computing*, 22(4), 2038-2052.
- [11] Zhou, Z., Feng, J., Gu, B., Ai, B., Mumtaz, S., Rodriguez, J., & Guizani, M. (2018). When mobile crowd sensing meets UAV: Energy-efficient task assignment and route planning. *IEEE Transactions on Communications*, 66(11), 5526-5538.
- [12] Pilloni, V., Ning, H., & Atzori, L. (2021). Task allocation among connected devices: Requirements, approaches, and challenges. *IEEE Internet of Things Journal*, 9(2), 1009-1023.
- [13] Gao, H., Feng, J., Xiao, Y., Zhang, B., & Wang, W. (2022). A UAV-assisted multi-task allocation method for mobile crowd sensing. *IEEE Transactions on Mobile Computing*, 22(7), 3790-3804.
- [14] Liu, X., Wang, Y., Gao, H., Ngai, E. C., Zhang, B., Wang, C., & Wang, W. (2024). A coverage-aware task allocation method for UAV-assisted mobile crowd sensing. *IEEE Transactions on Vehicular Technology*, 73(7), 10642-10654.
- [15] Skaltsis, G. M., Shin, H. S., & Tsourdos, A. (2023). A review of task allocation methods for UAVs. *Journal of Intelligent & Robotic Systems*, 109(4), 76.
- [16] Liu, D., Dou, L., Zhang, R., Zhang, X., & Zong, Q. (2022). Multi-agent reinforcement learning-based coordinated dynamic task allocation for heterogeneous UAVs. *IEEE Transactions on Vehicular Technology*, 72(4), 4372-4383.
- [17] Du, T., Gui, X., & Sheng, T. (2025). Multi-agent Deep Reinforcement Learning-Based Hierarchical Scheduling in Heterogeneous UAVs Enabled Vehicular Networks. *IEEE Internet of Things Journal*, 12(24): 54938-54954.
- [18] Frattolillo, F., Brunori, D., & Iocchi, L. (2023). Scalable and cooperative deep reinforcement learning approaches for multi-UAV systems: A systematic review. *Drones*, 7(4), 236.