Multi-UAV Energy Consumption Minimization using Deep Reinforcement Learning: An Age of Information Approach

Jeena Kim[1], Seunghyun Park[2], and Hyunhee Park[1]

[1]Myongji University - Natural Science Campus
[2]Hansung University

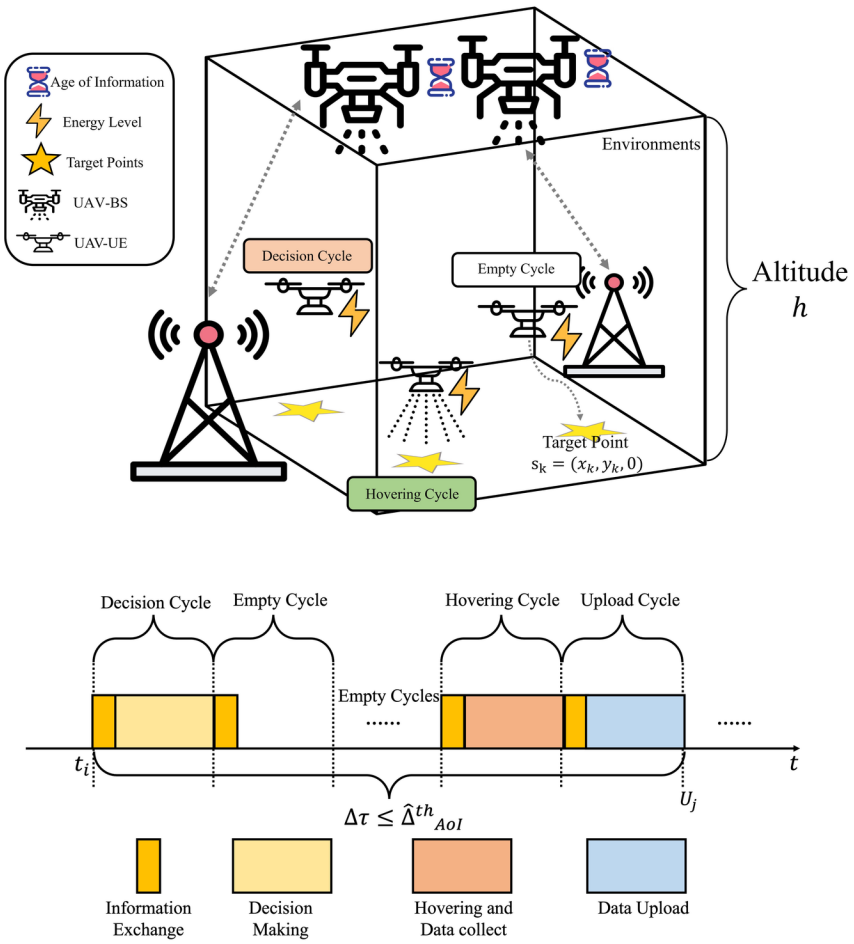
April 25, 2024

Abstract

This letter introduces an innovative approach for minimizing energy consumption in multi-UAV (Unmanned Aerial Vehicles) networks using Deep Reinforcement Learning (DRL), with a focus on optimizing the Age of Information (AoI) in disaster environments. We propose a hierarchical UAV deployment strategy that facilitates cooperative trajectory planning, ensuring timely data collection and transmission while minimizing energy consumption. By formulating the inter-UAV network path planning problem as a Markov Decision Process (MDP), we apply a Deep Q-Network (DQN) strategy to enable real-time decision-making that accounts for dynamic environmental changes, obstacles, and UAV battery constraints. Our extensive simulation results, conducted in both rural and urban scenarios, demonstrate the effectiveness of employing a memory access approach within the DQN framework, significantly reducing energy consumption up to 33.25\% in rural settings and 74.20\% in urban environments compared to non-memory approaches. By integrating AoI considerations with energy-efficient UAV control, this work offers a robust solution for maintaining fresh data in critical applications, such as disaster response, where ground-based communication infrastructures are compromised. The use of replay memory approach, particularly the online history approach, proves crucial in adapting to changing conditions and optimizing UAV operations for both data freshness and energy consumption.

Hosted file

`Multi-UAV Energy Consumption Minimization using Deep Reinforcement Learning: An Age of Information Appr`
available at https://authorea.com/users/774379/articles/869338-multi-uav-energy-consumption-minimization-using-deep-reinforcement-learning-an-age-of-information-approach

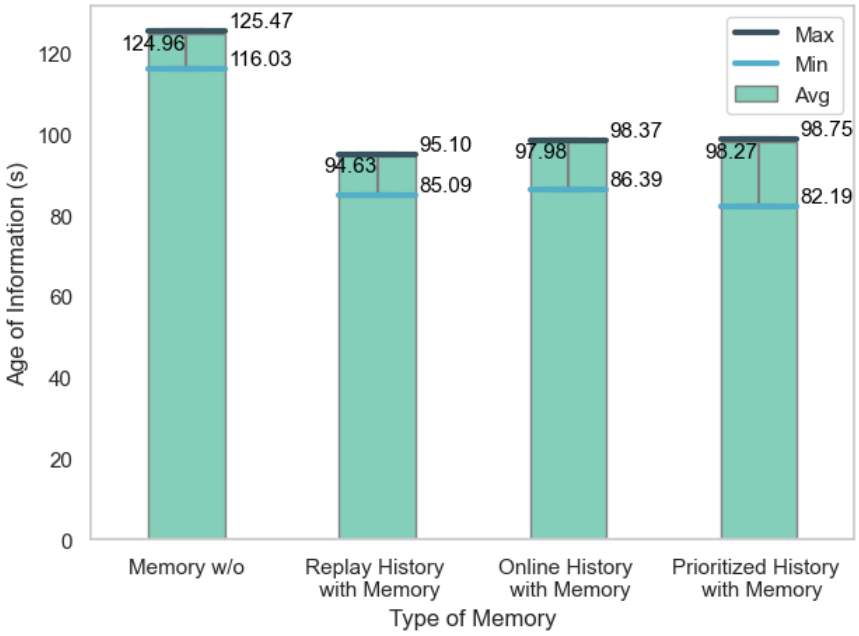


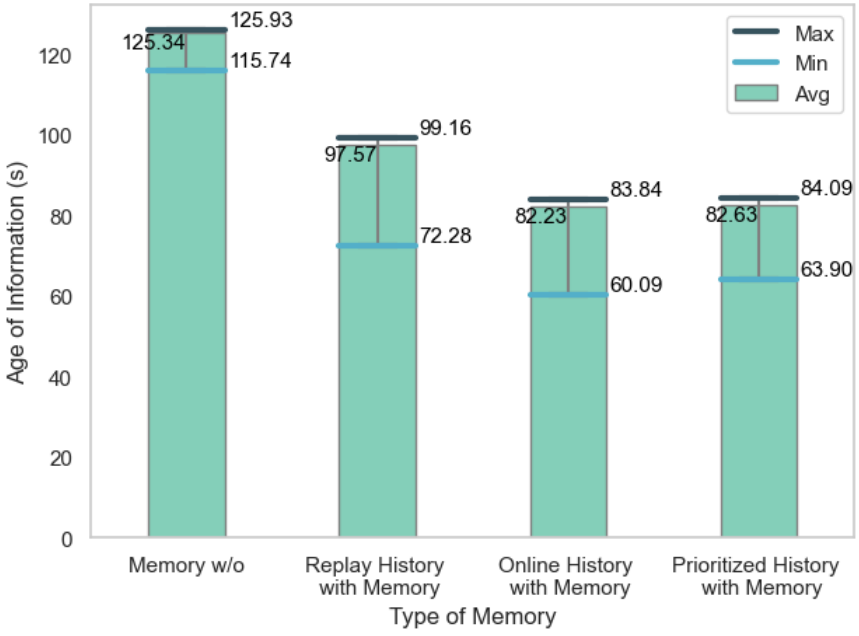
Map	Map Size (m^2)	Obstacle	Initial Position
Rural	1600×1600	4	(800,800)
Urban	800×800	8	(400,400)

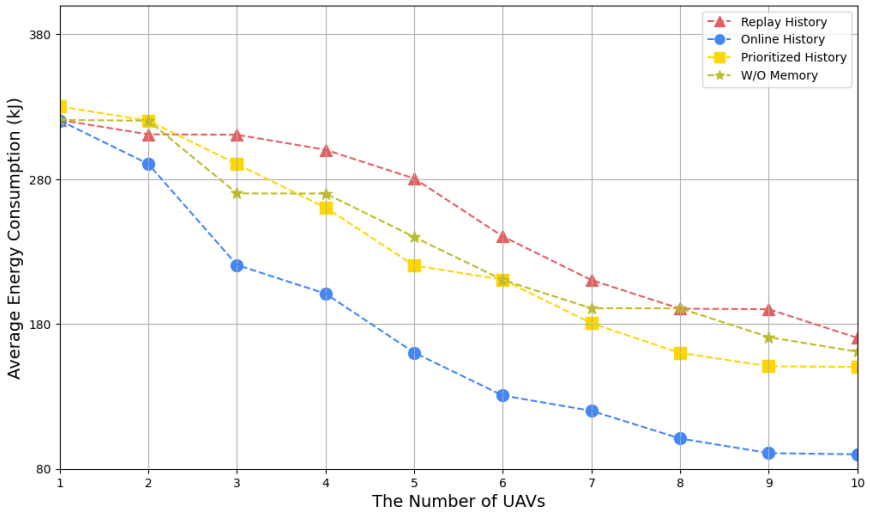
Table 1: Map Characteristics

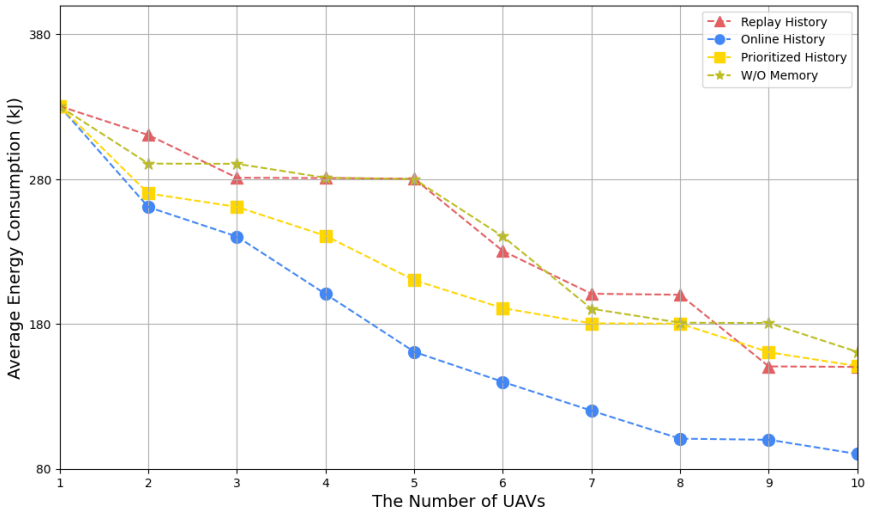
Map	Value
Episode	1000
Learning rate (α)	0.0005
Discount factor (γ)	0.99
Mini-batch size	32
Size of memory ($\mathcal{M}$)	10000

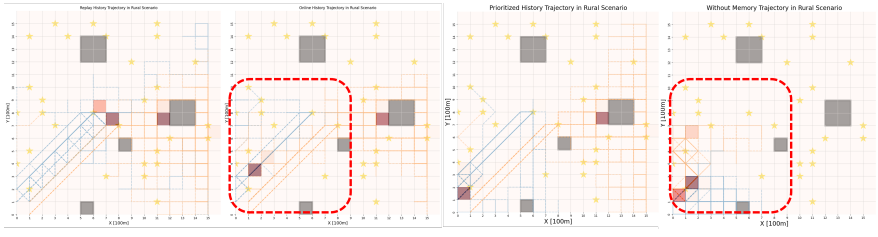
Table 2: Simulation Hyperparameter Values

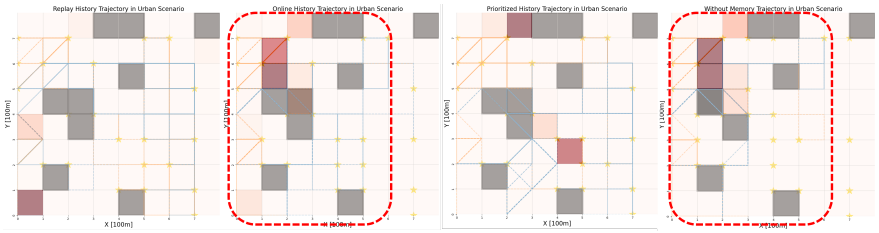












Multi-UAV Energy Consumption Minimization using Deep Reinforcement Learning: An Age of Information Approach

Jeena Kim, Seunghyun Park, Hyunhee Park

This letter introduces an innovative approach for minimizing energy consumption in multi-UAV (Unmanned Aerial Vehicles) networks using Deep Reinforcement Learning (DRL), with a focus on optimizing the Age of Information (AoI) in disaster environments. We propose a hierarchical UAV deployment strategy that facilitates cooperative trajectory planning, ensuring timely data collection and transmission while minimizing energy consumption. By formulating the inter-UAV network path planning problem as a Markov Decision Process (MDP), we apply a Deep Q-Network (DQN) strategy to enable real-time decision-making that accounts for dynamic environmental changes, obstacles, and UAV battery constraints. Our extensive simulation results, conducted in both rural and urban scenarios, demonstrate the effectiveness of employing a memory access approach within the DQN framework, significantly reducing energy consumption up to 33.25% in rural settings and 74.20% in urban environments compared to non-memory approaches. By integrating AoI considerations with energy-efficient UAV control, this work offers a robust solution for maintaining fresh data in critical applications, such as disaster response, where ground-based communication infrastructures are compromised. The use of replay memory approach, particularly the online history approach, proves crucial in adapting to changing conditions and optimizing UAV operations for both data freshness and energy consumption.

Introduction: In recent years, the use of Unmanned Aerial Vehicles(UAVs) has expanded extensively across civilian, commercial, and military domains. [1-4] In particular, in environments where ground-based stations are unreliable, UAVs can act as aerial base stations to provide communications during disasters. [5-9] In a disaster environment that is dynamically changing and requires real-time status updates, it is important to maintain the freshness of the collected data for immediate response and action. Age of Information (AoI) [10] is a metric to measure the freshness of data, which considers the overall temporal aspect of data from its generation to its delivery to the end user. In [11], the authors aim to minimize AoI in UAV systems using traditional dynamic programming (DP) algorithms and ant colonies (AC). In [12], the authors present an explicit formulation for the average AoI in Hamiltonian and non-Hamiltonian cycles using a graph-theoretical approach, and provide a mechanism to improve the AoI on a given flight path by creating new cycles around specific IoT devices. However, as the number of constraints increases, optimizing the trajectory design of UAVs while minimizing AoI becomes more complex and leads to an NP-hard problem. To alleviate these problems, deep reinforcement learning has been proposed as an approach to address a variety of different constraints. In particular, In [13], a deep Q-network is applied to optimize UAV scouting in an edge computing environment, taking into account energy efficiency and the Age of Information (AoI) context. Specifically, in [14] the authors explore the problem of path design to minimize the AoI through cooperative sensing and transmission in the cellular Internet of UAVs, introduce a scheduling method, and propose a composite action actor-critic (CA2C) algorithm based on deep reinforcement learning to address this. Deep reinforcement learning algorithms are trained to take the experience gained as episodes increase and store it in an experience replay memory so that the agent can take actions that can increase the rewards it can obtain in the future. In this work, we explore whether these properties of the experiential replay memory ensure AoI while reducing the energy consumption of UAVs in disaster environments. The main contributions of this thesis can be summarized as follows.

- First, we propose a hierarchical UAV deployment structure based on their respective roles for cooperative trajectory planning in disaster environments.
- Second, we propose a scheduling method to ensure AoI while minimizing energy consumption. To this end, we define the inter-UAV network path planning problem as an Markov Decision Process (MDP) and apply DQN to support real-time decision making.
- Finally, we conduct extensive experimental analysis to evaluate the performance of the proposed approach. Using the average AoI performance metric values, we conduct a simulation analysis to find the appropriate parameters for the learning model. The results suggest that the UAV AoI and energy consumption can be optimized.

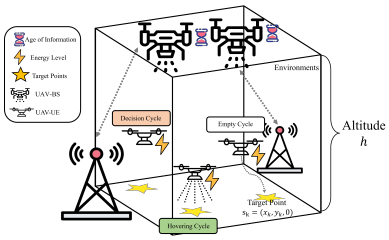


Fig. 1. Overview of System Model

System Model: In this letter, we address disaster scenarios occurring in rural and urban areas of size $N \times N\ m^2$. As illustrated in Fig. 1, our model incorporates a hierarchical UAV structure consisting of UAV-Base Station(BS) linked to operational ground base stations and I UAV-User Equipment(UE) deployed to collect data over dispersed target points. The locations where data is generated, positioned at $s_k \in \mathbb{R}^2$ for $1 \leq k \leq K$, are expected to initiate data production at $t_1 \leq 0$. Upon data generation at the kth Target Point, the UAV-BS identifies the genesis location of data. Positioned at $q_b = (x_b, y_b, h_b) \in \mathbb{R}^3$ for $b = 1, 2, \ldots, B$, each UAV-BS operates at the maximum altitude $h_{\max}$ and acts as a relay node covering area R_b. Subsequent to data generation, the UAV-BS designates the nearest UAV-UE to the target point. During any given time slot t, only a single UAV-UE is allowed to navigate to the target point $s_k$location. To determine the UAV-UE positioned at the minimum distance to the target point, we calculate the Euclidean distance as follows:

$$d_{3D}(t) = \sqrt{(x_i(t) - x_k(t))^2 + (y_i(t) - y_k(t))^2 + (h_i(t) - h_k(t))^2} \tag{1}$$

We assume the channel model between the i-th UAV-UE and the UAV-BS encompasses both large-scale and small-scale fading. Based on a 3D map simulated in [15], it precisely discerns the presence of a Line-of-Sight (LoS) or Non-Line-of-Sight (NLoS) connection. To minimize path loss, the UAV-UE must fly below altitude $h_i < h_{\max}$. If the distance from the UAV-UE's location s_k does not exceed the safety distance δ, the UAV-UE can directly transmit the collected data to the UAV-BS. This safety distance δ, as defined in equation 2, indicates that the data uploading location falls within the UAV-BS's coverage area. Otherwise, the UAV-UE must re-navigate within a safe distance to the area covered by the UAV-BS R_b.

$$\min_{1 \leq b \leq B} \{\|s_k - q_b\|\} \leq \delta \quad \text{for } 1 \leq k \leq K \tag{2}$$

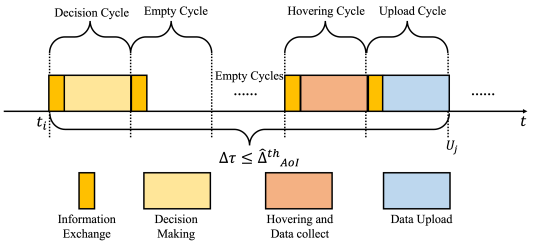


Fig. 2. Scheduling Cycle

Problem Formulation: We introduce AoI-Energy aware Scheduling to efficiently coordinate UAVs under contextual constraints, namely trajectory design, AoI, and energy consumption constraints. As depicted in Fig. 2, it consists of five stages. The duration of each cycle is represented by $\tau(t)$.

1) Information Exchange Cycle: Each cycle begins with the exchange of information between UAV-BS and UAV-UE. The information exchanged includes the current position of the i-th UAV-UE at time slot t, denoted as $q_{(i,t)} = (x_i, y_i, h_i)$, and the energy consumption during the previous cycle, represented as $E^{\text{cmp}}_{i,\tau(t-1)}$. The energy consumption used for flying the UAV-UE is denoted by $p_{\text{move}}(t) =$

$\sqrt{(\Delta x + \Delta y + \Delta h)}$. The variable $o_i(t) \in \{0, 1\}$ indicates whether there is a collision with obstacles. The energy consumption for hovering and uploading is represented by $P_h(t)$ and $\hat{P}_{(b,i)}(t)$, respectively. The AoI, calculated as $\Delta_{\text{AoI}}\tau_\tau^i$, is included. The total duration of 5 cycles is denoted by $\Delta\tau(t)$, and $U_i(t-1)$ represents the last upload time. This information forms the basis for the next cycle decisions.

$$E_{\text{cmp}} = \frac{1}{I}\sum_{i=1}^{I}\big(p_{\text{move}}(t)\cdot o_i(t)\cdot\tau_e(t) + P_h(t)\cdot\tau_h + \hat{P}_{(b,i)}(t)\cdot\tau_{tx}(t)\big) \quad (3)$$

$$\Delta\text{AoI}_i = \Delta\tau_i(t) - U_j(t-1) \quad (4)$$

2) Decision Cycle: The decision cycle begins when the UAV-BS identifies the location of the requested target point, $s_k(t)$, and, considering the current position and state of each UAV-UE, selects the UAV-UE that is closest to the requested target point. The selected UAV-UE must adhere to the energy constraint equation $\epsilon_i(t) = \frac{e_i^{\text{cmp}}(t)}{e_i^{\text{max}}(t)}$, where the current energy $\Delta e_i^{\text{cmp}}(t)$ and the maximum energy capacity $e_i^{\text{max}}(t)$ must meet the condition. If the selected UAV-UE does not satisfy the condition, the UAV-BS must reselect a new UAV-UE that is the closest within its area R_b.

$$\epsilon_i(t) = \begin{cases} 1, & \text{if } \Delta {e_i}^{\text{cmp}}(t) \geq {e_i}^{\text{max}} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

3) Empty Cycle: An empty cycle represents the state in which the UAV-UE is en route to the target point $s_k(t)$ but has not yet arrived. During this phase, the UAV-BS continuously monitors the UAV-UE and considers the estimated flight time $\tilde{\tau}_e$. If necessary, the path of the UAV-UE can be adjusted to minimize energy consumption and ensure the AoI.

4) Hovering Cycle: The hovering cycle occurs when the UAV-UE reaches the designated target point $s_k(t)$ and stops for data collection. At this time, the remaining energy of the i-th UAV-UE must satisfy the energy constraint condition $\epsilon_i(t)$

5) Upload Cycle: After the UAV-UE completes the hovering cycle, it starts the upload cycle. During this cycle, the UAV-UE can upload the collected data to the UAV-BS. It must transmit the data to a UAV-BS that covers the area R_b, which is within the transmission range of the UAV-UE. After the transmission is complete, the UAV-UE can record the time step U_j indicating the completion of all cycles. Subsequently, the UAV-UE flies to an area R_b within its transmission range for re-upload.

$$\zeta_i(t) = \begin{cases} \tau(t+1) = \tau_d, & \text{if } q_i \in R_b \text{ and } U_j \\ \tau(t+1) = \tau_{tx}, & \text{otherwise} \end{cases} \quad (6)$$

The UAV-UE selects a UAV-BS that covers an area R_b within the transmission distance.

$$\mu_i(t) = \begin{cases} 1, & \text{if } \sum_{j=1}^{U_j}\Delta\tau_i(t) \leq \hat{\Delta}_{\text{AoI}}^{\text{th}} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

To ensure the AoI for the i-th UAV-UE, the scheduling total duration $\tau(T)$ must not exceed the threshold $\hat{\Delta}_{\text{AoI}}^{\text{th}}$ as stipulated in Equation (7). If the total scheduling duration exceeds $\hat{\Delta}_{\text{AoI}}^{\text{th}}$, the data will be discarded. The variable j represents the time it takes for the UAV to perform the selected task, and U_j denotes the total time until the task is completed, which is used in the calculation of AoI.

To solve the problem, we apply the Deep Q-Network (DQN) [16], which is a combination of deep neural networks and reinforcement learning algorithms. A DQN can be defined as a MDP represented by a tuple $< S, A, R, S_{t+1} >$. An agent decides on an action a in a given state s. The agent receives a reward R and builds a policy π that takes into account a discount factor γ for the cumulative future reward. The proposed DQN approach consists of:

1 A deep neural network to reduce the dimensionality of the state space used to extract contextual features.
2 An experience replay memory to store the state transitions observed by the UAV-BS agent and the UAV-UE agent.
3 A reinforcement learning framework to find the optimal trajectory policy by solving constraints (9-11) to have a unique target area for each UAV-UE.

State: The state can be represented as $S_i(t) = [q_i(t), e_i^{\text{cmp}}(t), c_i(t)]$, which represents three key elements at time t. The position of the UAV-UE, $q_i(t) = (x_i(t), y_i(t), h_i(t))$, accurately tracks the spatial location of the UAV and is used to plan the next movement. $e_i^{\text{cmp}}(t)$ represents the current energy level of the UAV-UE, which can be expressed as the remaining operational energy $e_i^{\text{cmp}}(t) \in \mathbb{R}$. This directly impacts the sustainable operation and mission execution capability of the UAV. Lastly, $c_i(t)$ indicates the current cycle in which the UAV-UE is located. The possible states include {"Decision", "Empty", "Hovering", "Transmission"}, and this information is used to determine the next action of the UAV.

Action: Action is defined by the following equation (12), which describes the mobility of the UAV-UE in a given state. If it is hovering, it does not move.

$$A_i(t) = \begin{cases} q_i(t+1) = \Delta x + \Delta y + \Delta h, \text{Moving} \\ q_i(t+1) = (x_i, y_i, h_i), \text{Hovering} \end{cases} \quad (8)$$

Reward: When the learning agent, namely the UAV-UE, executes action $a_i(t)$, it transitions to a new state $s_i(t+1)$ and receives an immediate reward $r_i(t)$ associated with the state transition $s_i(t), a_i(t), s_i(t+1)$. The reward can be defined as follows in equation (13), where ϵ_i represents the energy constraint, and $r_{\text{cmp}}(t)$ signifies the reward for saving energy. The energy reward $r_{\text{energy}}^i = \Delta e_i(t)$ is defined by $\Delta e_i(t) = e_i(t) - \Delta e_i(t-1)$, which represents the energy consumed due to action a_t^i. $\mu_i(t)$ indicates the AoI constraint, and $r_{\text{AoI}}(t) = \Delta U_i(t)$ is expressed as $\Delta U_i(t) = U_i(t) - \Delta U_i(t-1)$. This provides a higher reward for the UAV-UE's continuous upload of fresh data. Lastly, $o_i(t)$ indicates whether there is a collision with obstacles.

$$R_i(t) = \epsilon_i(t) \times r_{\text{cmp}}(t) + \mu_i(t) \times r_{\text{AoI}}(t) + o_i(t) \quad (9)$$

The learning agent, UAV-UE, aims to maximize future rewards over T time slots as defined in equation (14). $\gamma = [0, 1]$ reflects the balance between the importance of immediate and future rewards, allowing convergence to the optimal policy π^{opt}.

$$\hat{R}(s, a, t) = \sum_{t_0=0}^{T} \gamma^{t_0} \times r_i(t - t_0) \quad (10)$$

Therefore, we can update the Q-function to derive the optimal policyπ^{opt}. as follows.

$$Q_{t'}(s, a) = Q_t(s, a) + \alpha\left[R + \gamma \max_{a'} q(s', a') - Q_t(s, a)\right] \quad (11)$$

Here, α is the learning rate that regulates the speed of the Q-function update. Additionally, $t' = t + 1$, and a' represents all actions considered during the maximization process.

Simulation Results: We propose a Replay Memory-based approach to find the appropriate AoI $\Delta_{\text{AoI}}^{\text{th}}$ within the proposed method, ensuring AoI through the use of replay memory. Initially, Replay Memory represents the (s, a, r, s_{t+1}) obtained by the agent interacting with the environment during the learning process. It exists in the following types:

- **Replay history:** Stores all past experiences and randomly selects them for learning, contributing to the learning process.
- **Online history:** Stores real-time or the most recent experiences, contributing to immediate learning.
- **Prioritized history:** Selects experiences for learning based on their importance, contributing to the learning process by choosing specific experiences.

Map	Map Size (m^2)	Obstacle	Initial Position
Rural	1600×1600	4	(800,800)
Urban	800×800	8	(400,400)

Table 1: Map Characteristics

Map	Value
Episode	1000
Learning rate (α)	0.0005
Discount factor (γ)	0.99
Mini-batch size	32
Size of memory ($\mathcal{M}$)	10000

Table 2: Simulation Hyperparameter Values

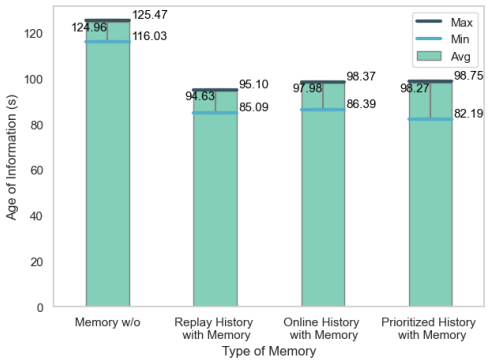


Fig. 3. Age of Information in Rural Scenario

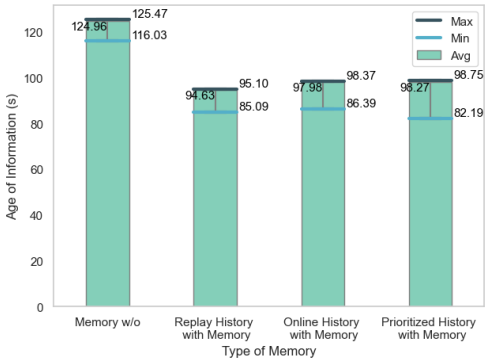


Fig. 4. Age of Information in Urban Scenario

We conducted experiments in a rural scenario characterized by few obstacles and a relatively large area, and an urban scenario with many obstacles and a relatively small area, in order to test UAVs in various environments. The following Fig. 3 and Fig. 4 represents the AoI results according to each memory access approach, facilitating the search for the appropriate $\Delta^{\text{th}}_{\text{AoI}}$. In the rural scenario of Fig. 3, the lowest average AoI was observed to be 94.63 seconds when applying the priority history memory access approach, which was 30.33 seconds shorter than the approach without memory usage. In the urban scenario of Fig. 4, the application of the online history memory access approach resulted in the lowest average AoI of 82.23 seconds, which was a reduction of 43.11 seconds compared to the non-memory approach.

After setting the average AoI to $\Delta^{\text{th}}_{\text{AoI}}$ in each scenario, we proceeded with energy consumption experiments. Fig. 5 indicates that, in the rural scenario with 5 UAVs deployed, the online history memory access approach shows the lowest energy consumption, which is up to 33.25% lower compared to the non-memory approach. Similarly, Fig. 6 shows that in the urban scenario, also with 5 UAVs, the online history approach results in the lowest energy consumption, showing up to a 74.20% reduction compared to the non-memory approach. These results suggest that the online history approach can adapt in real-time to relatively dynamic environments. On the other hand, both the replay history memory access method and the non-memory approach show comparatively higher energy consumption.

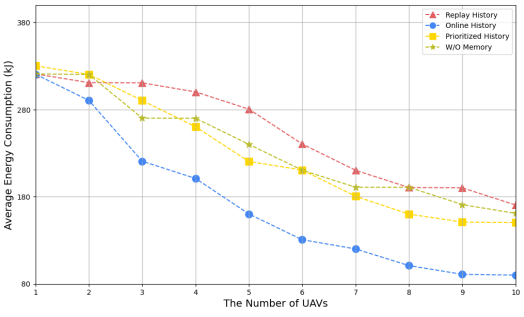


Fig. 5. Energy Consumption in Rural Scenario

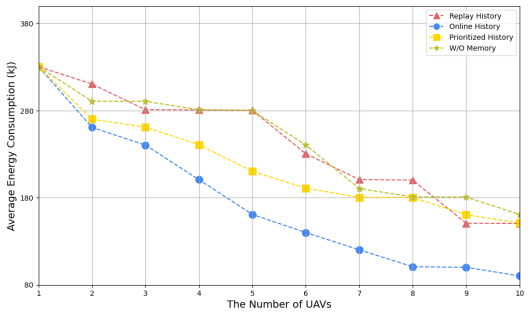


Fig. 6. Energy Consumption in Urban Scenario

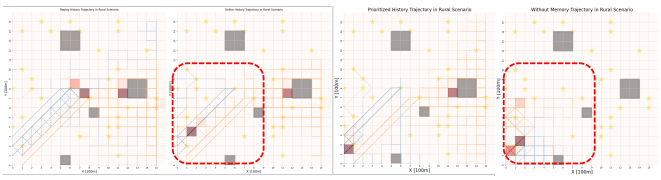


Fig. 7. Trajectory Visualization in Rural Scenario

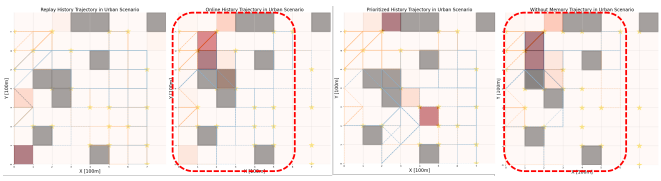


Fig. 8. Trajectory Visualization in Urban Scenario

To examine the energy consumption of UAVs based on different memory access approaches, we visualized the trajectories of two UAVs as shown in Fig. 7 and Fig. 8. Fig. 7 illustrates that in the rural scenario with the online history memory access approach applied, the UAVs fly in divided areas, suggesting that they reach the target points and collect data. In contrast, without the memory access approach, the UAVs overlap in their flight paths and fail to reach the target points. Therefore, the results indicate that the absence of a memory access approach leads to increased energy consumption due to overlapping flight paths and a failure to occupy distinct flying zones.

Conclusion: In this letter, we propose a hierarchical deployment structure and an energy consumption minimization scheduling method centered around the AoI for the efficient operation of UAVs. The results of applying a memory access approach-based DQN demonstrated that the online history approach reduces energy consumption by up to 33.25% in rural scenarios and up to 74.20% in urban scenarios. This shows that the proposed method is optimized for collecting relatively fresh data and minimizing energy consumption.

Acknowledgment: This work was supported in part by the National Research Foundation of Korea (NRF) and This research was financially supported by Hansung University for Seunghyun Park.

E-mail: jeena0627@gmail.com, sp@hansung.ac.kr, hhpark@mju.ac.kr

References

1 S.R. Sabuj, A. Ahmed, Y. Cho, K. Lee, H. Jo, "Cognitive UAV-aided URLLC and mMTC services: Analyzing energy efficiency and latency," *IEEE Access*, vol. 9, pp. 5011–5027 (2020).

2 B. Li, Z. Fei, Y. Zhang, "UAV communications for 5G and beyond: Recent advances and future trends," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2241–2263 (2018).

3 Y. Zeng, Q. Wu, R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proceedings of the IEEE*, vol. 107, no. 12, pp. 2327–2375 (2019).

4 L. Gupta, R. Jain, G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1123–1152 (2015).

5 L.A. binti Burhanuddin, X. Liu, Y. Deng, U. Challita, A. Zahemszky, "QoE optimization for live video streaming in UAV-to-UAV communications via deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 5358–5370 (2022).
6 M. Erdelj, E. Natalizio, "UAV-assisted disaster management: Applications and open issues," in *2016 International Conference on Computing, Networking and Communications (ICNC)*, pp. 1–5 (2016).
7 N. Zhao, W. Lu, M. Sheng, et al., "UAV-assisted emergency networks in disasters," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 45–51 (2019).
8 K.G. Panda, S. Das, D. Sen, W. Arif, "Design and deployment of UAV-aided post-disaster emergency network," *IEEE Access*, vol. 7, pp. 102985–102999 (2019).
9 Y. Gao, L. Xiao, F. Wu, D. Yang, Z. Sun, "Cellular-connected UAV trajectory design with connectivity constraint: A deep reinforcement learning approach," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 3, pp. 1369–1380 (2021).
10 İ. Kahraman, A. Köse, M. Koca, E. Anarım, "Age of Information in Internet of Things: A Survey," *IEEE Internet of Things Journal* (2023).
11 H. Hu, K. Xiong, G. Qu, Q. Ni, P. Fan, K.B. Letaief, "AoI-minimal trajectory planning and data collection in UAV-assisted wireless powered IoT networks," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 1211–1223 (2020).
12 G. Ahani, D. Yuan, Y. Zhao, "Age-optimal UAV scheduling for data collection with battery recharging," *IEEE Communications Letters*, vol. 25, no. 4, pp. 1254–1258 (2020).
13 S.F. Abedin, M.S. Munir, N.H. Tran, Z. Han, C.S. Hong, "Data freshness and energy-efficient UAV navigation optimization: A deep reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 9, pp. 5994–6006 (2020).
14 J. Hu, H. Zhang, L. Song, R. Schober, H.V. Poor, "Cooperative internet of UAVs: Distributed trajectory design by multi-agent deep reinforcement learning," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 6807–6821 (2020).
15 Y. Wang, Z. Gao, J. Zhang, et al., "Trajectory design for UAV-based Internet of Things data collection: A deep reinforcement learning approach," *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 3899–3912 (2021).
16 M.P. Bonecker, Y. Zhu, "Deep Q-network based decision making for autonomous driving," in *2019 3rd International Conference on Robotics and Automation Sciences (ICRAS)*, pp. 154–160 (2019).