



**International Journal of Computer Applications in Technology**

ISSN online: 1741-5047 - ISSN print: 0952-8091

<https://www.inderscience.com/ijcat>

---

**A novel leader replacement-based unmanned aerial vehicle flocking scheme**

Junling Shi, Guoyu Zhu, Aihua Men, Guiying Meng

**DOI:** [10.1504/IJCAT.2025.10072643](https://doi.org/10.1504/IJCAT.2025.10072643)

**Article History:**

|                   |                  |
|-------------------|------------------|
| Received:         | 02 November 2023 |
| Last revised:     | 08 March 2024    |
| Accepted:         | 12 April 2024    |
| Published online: | 27 August 2025   |

## A novel leader replacement-based unmanned aerial vehicle flocking scheme

Junling Shi\* and Guoyu Zhu

School of Computer Science,  
Shenyang Aerospace University,  
Shenyang, China  
Email: jlshi@sau.edu.cn  
Email: zhuguoyu2108@163.com

Aihua Men\*

School of Mathematics and Computer Science,  
Chifeng University,  
Chifeng, China  
Email: cfxymah@126.com  
\*Corresponding authors

Guiying Meng

School of Computer Science,  
Shenyang Aerospace University,  
Shenyang, China  
Email: mengguiying@163.com

**Abstract:** The flocking motion is a fundamental and crucial operation in multi-UAV systems, encompassing navigation and obstacle avoidance. However, the unknown and random environment poses a great challenge to traditional Unmanned Aerial Vehicle (UAV) flocking and navigation control methods. This paper uses Reinforcement Learning (RL) techniques to achieve navigation and obstacle avoidance for a swarm of UAVs in unknown environments. The RL algorithm employed in this study is MAPPO, which has been integrated with an attention mechanism to form the ATT-MAPPO algorithm. This incorporation of the attention mechanism enables the agent to effectively filter out irrelevant information and focus its attention solely on crucial features relevant to the task at hand, thereby significantly enhancing decision-making accuracy and efficiency. The leader replacement strategy effectively balances the energy distribution within the UAV flocking, thereby significantly extending the overall flight range of the UAV. Finally, we demonstrate the scalability and adaptability of ATT-MAPPO in a simulation experiment.

**Keywords:** flocking; multi-agent reinforcement learning; unmanned aerial vehicles; leader replacement.

**Reference** to this paper should be made as follows: Shi, J., Zhu, G., Men, A. and Meng, G. (2025) 'A novel leader replacement-based unmanned aerial vehicle flocking scheme', *Int. J. Computer Applications in Technology*, Vol. 76, Nos. 1/2, pp.106–114.

**Biographical notes:** Junling Shi received the PhD degree in computer application technologies from Northeastern University, Shenyang, China in 2019. She is currently an associate professor of the School of Computer Science, Shenyang Aerospace University, China. Her research interests include wireless power transfer, edge computing offloading and flocking in UAV-network and vehicular network.

Guoyu Zhu received his master's degree in computer technologies from Shenyang Aerospace University, Shenyang, China in 2024. He currently serves as an engineer at a technology company. His research interests include flocking of UAVs and reinforcement learning.

Aihua Men received her master's degree in software engineering from Northeastern University, Shenyang, China in 2009. She is currently an associate professor of School of Mathematics and Computer Science, Chifeng University, China. Her research interests include software engineering and database.

Guiying Meng received her master's degree in computer application technologies from Northeastern University, Shenyang, China in 2006. She is currently a lecturer of the School of Computer Science, Shenyang Aerospace University, China. Her research interests include software engineering and industrial Internet of things.

## 1 Introduction

In recent times, Unmanned Aerial Vehicle (UAV) swarms have been extensively utilised in various demanding missions such as countering terrorism, safeguarding borders and conducting search and rescue operations (Chipade and Panagou, 2021; Ribeiro et al., 2021; Gassara and Rodriguez, 2021; Zhu et al., 2023a, 2023b, 2023c). The challenge lies in ensuring collision-free flocking, which is a fundamental control problem encountered in these applications. The primary objective of UAV flocking is to achieve target point navigation cohesively while ensuring autonomous navigation and obstacle avoidance. There are two predominant forms of UAV swarm leader-following (Yan et al., 2021, 2022) and leaderless (Wang et al., 2022; Shen et al., 2021). Efficient cooperation among the members of the UAV swarm is crucial for accomplishing their designated tasks.

The autonomy level of UAV swarms has been extensively researched, utilising conventional methods like consensus theory (Vicsek et al., 1995). Model predictive control that relies on precise UAV kinematic models (Luo and Duan, 2017) and environmental disturbance models (Zhao et al., 2018), which can be difficult to obtain in practical scenarios. To address this limitation, an alternative approach that has garnered significant attention is model-agnostic Deep Reinforcement Learning (DRL) (He 2023a, 2023b; Zhu et al., 2021, 2022). By not necessitating accurate environmental models, Reinforcement Learning (RL) algorithms can automatically adapt to new environments without modifying the underlying model itself. This enables them to discover innovative behaviours and showcase advantages compared to fixed rule-based systems.

Wang et al. (2020) proposed an approach for multi-UAV path planning based on the enhanced  $Q$ -learning (DQN) algorithm, which is primarily employed to address problems involving sequential decision-making with discrete control signals. Subsequently, a method called Deterministic Policy Gradient (DPG) (Lillicrap et al., 2015) is introduced to tackle continuous control signals. Following this, numerous researchers have applied DRL in UAV navigation tasks. For instance, the Deep Deterministic Policy Gradient (DDPG) algorithm is utilised in Koch et al. (2019) to govern UAV flight actions and the outcomes demonstrate that the trained policy surpasses PID controllers. In Hwangbo et al. (2017), model-free policy gradient methods are employed to regulate rotary-wing UAVs' tasks, resulting in a policy capable of tracking UAV trajectories that can be implemented on real UAVs. Moreover, Wang and Gu (2009) employed a RL leader-follower algorithm to achieve the collective motion of UAVs.

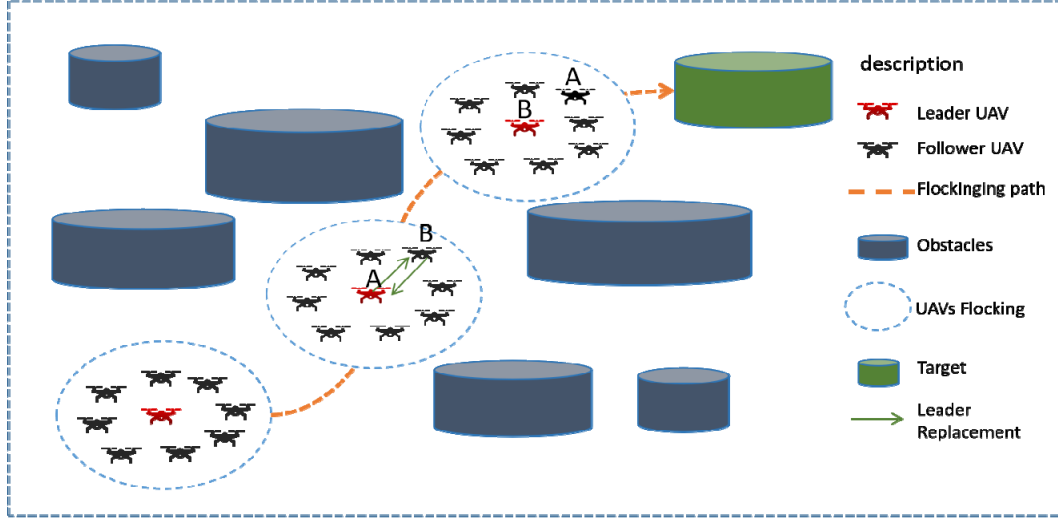
The majority of existing RL algorithms are based on the leader-follower framework (Wang et al., 2018; Yan et al., 2020; Abichandani et al., 2021). The current UAV flocking algorithm based on the leader-follower framework, however, fails to address the issue of power consumption disparity resulting from different internal roles. As a result of distinct tasks within the UAV swarm, UAVs exhibit dissimilar power requirements. While the leader necessitates decision-making and consequently consumes more power, followers do not need to process road information for path planning, resulting in lower power consumption. This significant disparity leads to situations where leaders run out of power while followers still possess sufficient energy during flight, thereby greatly reducing the overall utilisation efficiency of the UAV swarm. the utilisation efficiency of UAV swarm power, we propose a RL flocking algorithm that incorporates both leader following and role switching. The main contributions are summarised as follows:

- We investigate the application of RL in UAV teams, with a specific focus on advancements in intelligent leader replacement and leader-follower group navigation capabilities. A timing algorithm is employed to determine the optimal time for leader replacement, aiming to achieve power balance within the UAV team and enhance overall efficiency.
- RL-based leader-follower navigation function facilitates dynamic strategy adjustment, thereby enhancing the collaborative action efficiency and accuracy of UAV teams in complex environments.
- Attention mechanism enhances the efficacy and performance of RL MAPPO algorithm by focusing on policy update-related information, prioritising challenges in UAV team navigation and replacement.

## 2 System model and problem statement

### 2.1 System model

In this section, we introduce the system framework and represent the flocking motion as a sequential decision problem. To optimise the training of DRL, and to enhance the flocking movement and navigation capabilities of UAVs in a variety of scenarios, we present a system model consisting of two key components: the environmental module and the UAV module, as depicted in Figure 1.

**Figure 1** UAV swarms maintain formation and avoid obstacles while navigating toward a target point (see online version for colours)

The environmental component consists of cost-effective and small-sized UAVs along with a specific task. Each UAV is equipped with multiple sensors, facilitating real-time detection of its environment, such as the speed and angular velocity of nearby UAVs, positions of obstacles and locations of target points.

To effectively train the collective movement and avoidance of obstacles by multiple UAVs, we have designed an experimental environment that specifically caters to the requirements of UAVs with continuous action. In this experimental setup, rotary-wing UAVs exhibit seamless motion that closely replicates real-world physics, as mentioned in a previous study (Wang et al., 2019). When establishing the motion model for UAVs, our main point of reference is another research paper (Wang et al., 2022).

## 2.2 Problem statement

Firstly, our primary objective is to optimise the overall flight path of all UAVs with distance travelled while ensuring uniform power consumption across all UAVs. Our objective is to optimise the flight path of the UAV swarm by achieving a balanced distribution of battery levels among all UAVs. However, conventional algorithms that maintain a fixed leader result in disproportionate battery consumption, with the leader experiencing higher energy depletion compared to followers, leading to suboptimal flight paths. By introducing dynamic leadership changes, we observe a dispersion in battery consumption which ultimately leads to an overall increase in flight efficiency. The current power of  $UAV_i$  is denoted as  $p_i$ ,  $p_{ave}$  is the average power consumption of all UAVs. Therefore, we propose the following metrics:

$$\min \frac{\sum_{i=1}^N |p_i - p_{ave}|}{N} \quad (1)$$

Additionally, it is crucial to adhere to boid rules during the flight, including (i) Collision avoidance: prevent collisions with the surrounding environment; (ii) Match: strive to

synchronise the velocity with that of the neighbouring entities; (iii) Flocking centralisation: aim for positioning within the immediate vicinity.

## 3 Algorithm

This section initially introduces the modelling problem as a Markov model. Subsequently, we formulate a customised reward function to cater to the requirements of the swarming problem. Finally, we present a comprehensive description of MAPPO with the incorporation of the attention mechanism and replacement leader algorithm.

### 3.1 Reinforcement learning for flocking motion problem

The problem of swarm movement in UAVs is described as a Markov model, where the MDP satisfies the Markov property that the next state depends on both the current state and the action chosen by the decision maker. Treating each UAV as an individual decision-maker, it is essential for flocking motion to adhere to the Markov property. Consequently, swarming and navigation control of UAVs are approached as sequential decision problems requiring step-by-step control of each UAV to achieve coordinated swarming motion.

To accomplish the collective movement and navigation objectives of a UAV swarm, the UAVs are categorised into two distinct roles: leaders and followers, each with their unique reward functions. The leaders primarily focus on navigation and path planning functionalities, while the followers encompass both following the leader and maintaining swarm capabilities with other UAVs. In their respective reward functions, both roles share common functionalities of reaching target points and avoiding obstacles.

**Target approaching reward:** the target-approaching reward is specifically designed to guide UAVs in leadership roles, primarily directing them towards designated target points. It determines a positive reward by calculating the

difference between the distance from the previous position of the leading UAV to the target point and its current distance from that same target point. A higher value indicates superior performance.  $\langle x_t^l, y_t^l \rangle$  is the position of the leader UAV.  $\langle x_{tar}, y_{tar} \rangle$  is the position of the target point. The target approaching reward  $r_1$  is defined as:

$$r_1 = w_1 \frac{\sqrt{(x_{t-1}^l - x_{tar})^2 + (y_{t-1}^l - y_{tar})^2}}{\sqrt{(x_t^l - x_{tar})^2 + (y_t^l - y_{tar})^2}} \quad (2)$$

The value of  $w_1$  is positive.

Flocking reward: the flocking reward is a distinctive value assigned to the follower UAV, comprising two components: ensuring that the current UAV maintains an appropriate distance from the leader UAV and also ensuring that this UAV maintains an appropriate distance from other follower UAVs.  $\langle x_t^i, y_t^i \rangle$  is the position of the UAV<sub>*i*</sub>. The safe distance between UAVs is denoted as  $d_1$ . The maximum communication distance between UAVs is denoted as  $d_3$ . the maximum communication distance between UAVs is denoted as  $d_2$ . The flocking reward  $r_2$  and  $r_3$  are defined as:

$$r_2 = \begin{cases} w_2, \sqrt{(x_t^i - x_t^l)^2 + (y_t^i - y_t^l)^2} < d_1 \\ w_3, d_1 \leq \sqrt{(x_t^i - x_t^l)^2 + (y_t^i - y_t^l)^2} \leq d_3 \\ w_4, \sqrt{(x_t^i - x_t^l)^2 + (y_t^i - y_t^l)^2} > d_3 \end{cases} \quad (3)$$

$$r_3 = \begin{cases} w_5, \sqrt{(x_t^i - x_t^j)^2 + (y_t^i - y_t^j)^2} < d_2, & i \neq j, \forall j \in N \\ w_6, d_1 \leq \sqrt{(x_t^i - x_t^j)^2 + (y_t^i - y_t^j)^2} \leq d_3, & i \neq j, \forall j \in N \end{cases} \quad (4)$$

The values of  $w_3$  and  $w_6$  are positive. The value of  $w_2$ ,  $w_4$  and  $w_5$  are negative.

Obstacle avoidance reward: this reward is specifically designed to provide guidance for UAVs and enable them to maintain a safe distance from stationary obstacles, with collisions being considered adverse incidents that lead to the termination of the training session and subsequent penalties.  $\langle x_m, y_m \rangle$  is the position of the obstacle. The obstacle avoidance reward  $r_4$  is defined as:

$$r_5 = w_7, \sqrt{(x_t^i - x_m)^2 + (y_t^i - y_m)^2} \leq d_2, \quad \forall m \in O \quad (5)$$

$$0 \quad \text{otherwise}$$

The value of  $w_7$  is negative.

Target point reaching reward: when the guiding UAV and the following UAV reach the target point, they will receive a shared reward value as a termination condition, and each UAV will receive a substantial reward value to signify the completion of training. The target point is abstracted as a circular area with a radius of  $d_4$ .

$$r_5 = w_8, \sqrt{(x_t^i - x_{tar})^2 + (y_t^i - y_{tar})^2} \leq d_4 \quad (6)$$

$$0 \quad \text{otherwise}$$

where  $w_8$  is a positive constant.

After each action executed by the UAV at time  $t$ , the leader's cumulative reward is computed as the summation of equations (2), (5) and (6), while the combined reward function is formulated as follows:

$$r_t^l = r_1 + r_4 + r_5. \quad (7)$$

The cumulative reward of the follower is computed as the summation of equations (3), (4), (5) and (6), while the combined reward function is defined as follows:

$$r_t^f = r_2 + r_3 + r_4 + r_5. \quad (8)$$

### 3.2 ATT-MAPPO

MAPPO is an advanced algorithm that combines DRL and policy optimisation to train UAVs for navigation, flocking movement, leader replacement and obstacle avoidance. It effectively handles scenarios with multiple agents by enabling collaborative learning and dynamic strategy adjustment. Additionally, MAPPO integrates centralised training and decentralised execution techniques for better policy development.

The MAPPO algorithm employs an RNN network. In comparison to a standard RNN, the integration of RNN with an attention mechanism confers several advantages and facilitates a more adaptable approach for information extraction. By incorporating the attention mechanism, the model can assign varying weights to different positions within input sequences during information processing, thereby constructing a weighted sum vector that enables focused attention on pertinent information for the current task at hand. This weighted sum vector methodology offers enhanced flexibility in extracting information and facilitates improved adaptation across diverse tasks and data sets.

The time complexity of the MAPPO algorithm is primarily determined by two factors: the number of policy optimisation iterations and the amount of sampled data in each iteration. During each iteration, MAPPO estimates the advantage function by sampling multiple trajectories and utilises the PPO algorithm to update the policy. Additionally, for each trajectory, estimating the value function is required by the MAPPO algorithm. Therefore, this algorithm's time complexity is directly proportional to the number of sampled trajectories. Assuming that  $N$  parallel trajectories are generated with a length denoted as  $T$ , an approximate estimation for total time complexity can be expressed as  $O(N * T * I)$ , where  $I$  represents the number of iterations.

We incorporate the attention mechanism into the MAPPO algorithm, thereby enhancing its performance. The ATT-MAPPO algorithm is summarised in Algorithm 1.

**Algorithm 1:** ATT-MAPPO Algorithm

---

```

1: Initialise  $\theta$ , the parameters for policy  $\pi$  and  $\phi$ , the
   parameters for critic  $V$ ;
2: Set learning rate  $\alpha$ ;
3: while  $step \leq step_{max}$  do
4:   data buffer  $D = \{\}$ ;
5:   for  $i = 1$  to  $batch_{size}$  do
6:      $\tau = \text{empty list}$ ;
7:     initialise  $h_{0,\pi}^{(1)}, \dots, h_{0,\pi}^{(n)}$  actor RNN states;
8:     initialise  $h_{0,V}^{(1)}, \dots, h_{0,V}^{(n)}$  critic RNN states;
9:     for  $t = 1$  to  $T$  do
10:      for all agents do
11:         $p_t^{(a)}, h_{t,\pi}^{(a)} = \pi(o_t^a, h_{t-1,\pi}^{(a)}; \theta)$ ;
12:         $v_t^{(a)} \sim p_t^{(a)}$ ;
13:         $v_t^{(a)}, h_{t,V}^{(a)} = V(s_t^a, h_{t-1,V}^{(a)}; \phi)$ ;
14:      end for
15:      Execute actions  $u_t$ , observe  $r_t, s_{t+1}, o_{t+1}$ ;
16:       $\tau \leftarrow [\tau, [s_t, o_t, h_{t,\pi}, h_{t,V}, u_t, r_t, s_{t+1}, o_{t+1}]]$ ;
17:    end for
18:    Computer advantage estimate  $\hat{A}$  via GAE on  $\tau$ 
    using PopArt;
19:    Computer reward-to-go  $\hat{R}$  on  $\tau$  and normalise
    with PopArt;
20:    Split trajectory  $\tau$  into chunks of length  $L$ ;
21:    for  $l = 0, 1, \dots, T/L$  do
22:       $D = D \cup (\tau[l:L+T], \hat{A}[l:L+L], \hat{R}[l:L+L])$ ;
23:    end for
24:  end for
25:  for mini-batch  $k = 1, \dots, k$  do
26:     $b \leftarrow$  random mini-batch from  $D$  with all
    agent data
27:    for each data chunk  $c$  in the mini-batch  $b$  do
28:      update RNN hidden states for  $\pi$  and  $V$ 
      from first hidden state in data chunk;
29:    end for
30:  end for
31:  Update  $\theta$  on  $L(\theta)$  with data  $b$ ;
32:  Update  $\phi$  on  $L(\phi)$  with data  $b$ ;
33: end while

```

---

To ensure a balanced power distribution within the swarm, priority should be given to selecting a leader with optimal current power levels. This choice guarantees that the leader possesses sufficient energy reserves for undertaking leadership responsibilities while maximising overall swarm longevity.  $p_i$  is the current power of the  $UAV_i$ .

### 3.3 Leader replacement

In the UAV flocking algorithm, the leader UAV plays a crucial role in gathering information about the surrounding

environment and making important decisions regarding path-planning tasks. However, this responsibility comes at a cost of higher power consumption compared to the following UAVs. This increased power usage can lead to a reduction in the maximum flight distance of the entire swarm.

To address this issue and ensure optimal performance of all UAVs within the swarm, it becomes necessary to implement a leader replacement algorithm. The objective of this algorithm is to balance power consumption among multiple UAVs and extend their overall service life.

By periodically replacing the leader UAV with another member from within the flock, we distribute the burden of collecting information and performing path-planning tasks more evenly across all UAVs. This not only reduces individual power consumption but also prevents any single UAV from being overburdened or experiencing premature battery drain.

The leader replacement algorithm ensures that each member of the swarm gets an opportunity to contribute as a leader while also conserving energy for prolonged operation. By achieving a balanced distribution of responsibilities, we enhance efficiency and longevity for all participating UAVs.

Moreover, extending their overall service life has significant benefits beyond just reducing power consumption. It allows for longer mission durations without interruptions due to depleted batteries or limited flight range. This translates into improved productivity and effectiveness in various applications such as surveillance missions, search-and-rescue operations or even package delivery services.

To identify the most suitable leader UAV, it is imperative to calculate the inter-UAV distances within the UAV swarm. These distances serve as indicators of relevance and proximity among the UAVs. Subsequently, by determining the closest UAV to all others can select an appropriate leader capable of effectively coordinating and directing cooperative efforts within the entire swarm.  $l_i$  is the average distance between  $UAV_i$  and other UAVs. The equation is as follows:

$$l_i = \frac{\sum_{j=1}^N \sqrt{(x_i^j - x_t^j)^2 + (y_i^j - y_t^j)^2}}{N}, j \neq i \quad (9)$$

Finally, by calculating each UAV's average neighbour distance and performing a weighted sum considering their respective power capacities, they can best be suited for assuming leadership duties. The present study defines a comprehensive evaluation criterion for leader change  $c_i$  as follows:

$$c_i = -\phi_1 * l_i + \phi_2 * p_i. \quad (10)$$

The weights of  $l_i$  and  $p_i$  are denoted as  $w_1$  and  $w_2$ .

In summary, the implementation of an efficient leader replacement algorithm in the UAV swarming system enables enhanced resource utilisation by effectively balancing power consumption across multiple UAVs. This approach not only extends the overall system's service life but also enhances its operational efficiency.

## 4 Experiments

In this section, we conduct simulations to evaluate the effectiveness of the UAV flocking algorithm. Initially, we

introduce the simulation environment and define relevant parameters. Subsequently, we compare the convergence of ATT-MAPPO, MAPPO and MADDPG algorithms. Finally, we compare the performance of ATT-MAPPO with other existing methods.

#### 4.1 Convergence comparison

To verify the effectiveness of our algorithm, we compare the convergence of ATT-MAPPO with MAPPO and MADDPG by displaying the curves of the average reward with varying episodes as shown in Figure 2. Simulation parameters are presented in Table 1. Aside from differences in network architecture, all other simulation parameters are identical for ATT-MAPPO, MAPPO and MADDPG. The results show that, regardless of the number of UAVs, ATT-MAPPO performs better than MAPPO and MADDPG after convergence. Additionally, the ATT-MAPPO achieves convergence faster than MAPPO and MADDPG during training and exhibits more stable convergence performance upon completion of training.

**Figure 2** Convergence comparison between ATT-MAPPO, MAPPO and MADDPG (see online version for colours)

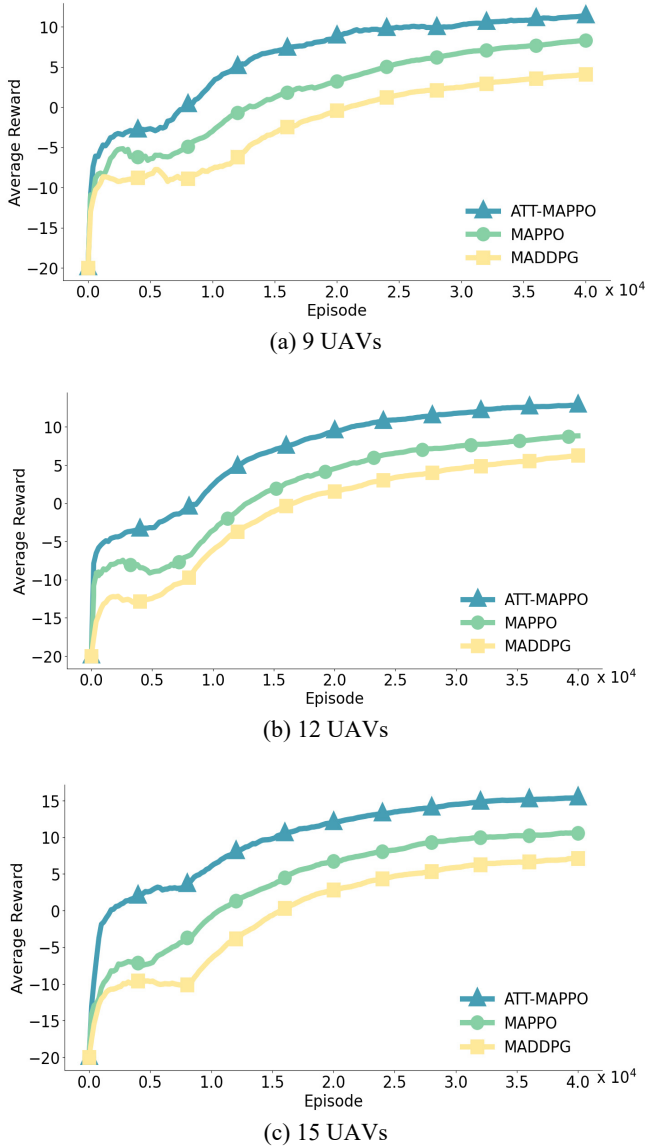


**Table 1** Simulation parameters

| Parameter                                       | Value       |
|-------------------------------------------------|-------------|
| The maximum speed of the UAV                    | 6 m/s       |
| The minimum speed of the UAV                    | 2 m/s       |
| The maximum roll of the UAV                     | $3\pi / 4$  |
| The minimum roll of the UAV                     | $-1\pi / 4$ |
| Power consumption per metre of the leader UAV   | 0.13 %      |
| Power consumption per metre of the follower UAV | 0.1 %       |
| The safe distance with neighbours               | 3 m         |
| The communication distance                      | 20 m        |
| The safe distance with obstacles                | 3 m         |
| The number of obstacles                         | 4           |
| State dim                                       | 48          |
| action dim                                      | 2           |
| Replay buffer size                              | 100,000     |
| Batch size                                      | 256         |
| The discount rate                               | 0.99        |

The ATT-MAPPO outperforms the traditional MAPPO algorithm due to its incorporation of an attention mechanism, which enhances the model's ability to focus on and accurately learn crucial information. By incorporating the importance weight and truncated advantage function, the MAPPO algorithm can effectively leverage previous empirical data and enhance sampling efficiency. In contrast, the MADDPG algorithm is susceptible to sample correlation during training, leading to suboptimal sampling efficiency.

#### 4.2 Performance comparison of ATT-MAPPO

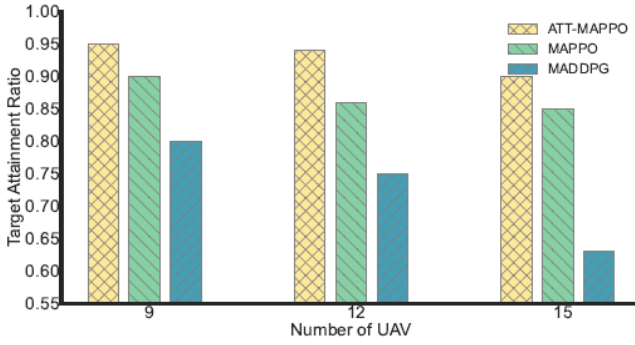
**Target attainment ratio (TAR):** to evaluate the efficiency of the flocking algorithm in reaching the designated region, we employ TAR as a performance measure. The TAR represents the proportion of UAVs that successfully reach the target area out of all launched UAVs, with a higher value indicating the superior performance of the UAV flocking algorithm.

**Average travel distance (ATD):** the aforementioned measurement is employed to calculate the mean travel distance required by UAVs to reach their designated target. This metric quantifies the average distance covered by the UAVs during their journey toward the destination. It should be noted that energy consumption is directly proportional to the distance travelled, thus a lower average travel distance indicates improved algorithmic energy efficiency.

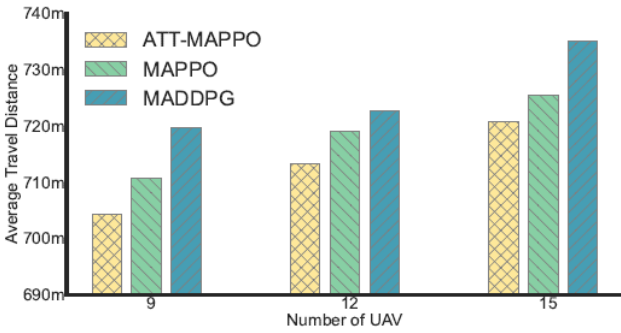
In Figure 3(a), as the number of UAVs increases, the TAR of the ATT-MAPPO algorithm gradually decreases, while both MAPPO and MADDPG algorithms exhibit more pronounced performance degradation. The exceptional performance achieved by ATT-MAPPO can be attributed to its utilisation of attention mechanisms during the training process. As shown in Figure 3(b), all algorithms demonstrate an increasing trend in their indicators with an increasing number of UAVs. However, when comparing the same number of UAVs, ATT-MAPPO outperforms MAPPO and

MADDPG in terms of efficiency for flocking motion task by consuming less energy and reducing travel time.

**Figure 3** Convergence comparison between ATT-MAPPO, MAPPO and MADDPG (see online version for colours)



(a) TAR



(a) ATD

#### 4.3 Performance comparison of leader replacement

**Battery-level consistency (BLC):** BLC refers to the requirement for maintaining stable and uniform power distribution among a swarm of UAVs. In our experiments, BLC is calculated using equation (11), which involves computing the difference between each UAV's residual electric command and the average residual of the entire swarm, normalised by their average speed. A smaller value indicates a higher level of coordination and unity regarding residual electric power within the UAV swarm.

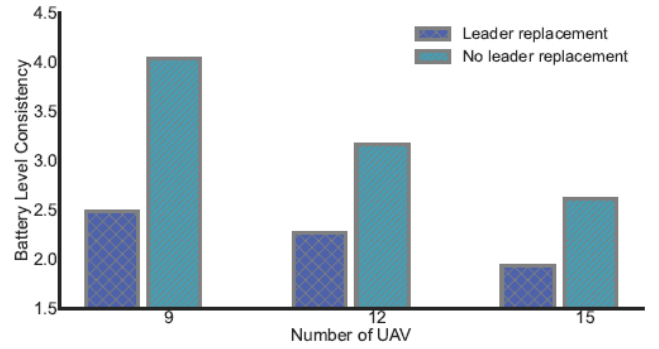
$$BLC = \frac{\sum_{i=1}^N \frac{|p_i - p_{ave}|}{p_{ave}}}{N}. \quad (11)$$

**Maximum flight distance (MFD):** MFD comparison will be conducted between the UAV swarm with leader replacement function and the one with no leader replacement. The MFD of the UAV swarm is defined when a UAV within the swarm exhausts its power. A larger maximum flight distance of the UAV swarm indicates a better performance.

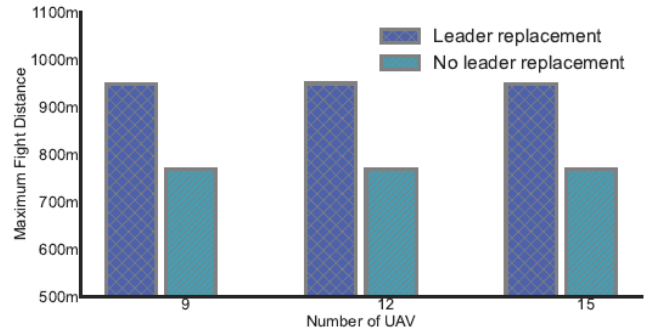
In Figure 4, it has been observed that incorporating leader replacement into the algorithm significantly enhances the performance of UAV swarms compared to the conventional UAV swarm algorithm. By incorporating the leader replacement function into the algorithm, a significant improvement in the performance of the UAV swarm is

observed compared to the traditional UAV swarm algorithm. The test results demonstrate that equipping the swarm with a strong leader and an efficient transition algorithm can greatly enhance overall performance when validated with different numbers of UAVs (9, 12 and 15 UAVs). Ensuring a balanced distribution and effective utilisation of surplus power after mission completion or during flight when leadership is depleted can extend the flight range of the entire UAV swarm and maximise individual UAV efficiency. The proposed algorithm consistently maintains power balance within the UAV swarm at all times, unlike the traditional algorithm which leads to non-uniform power distribution within the UAV swarm thereby reducing overall flight life.

**Figure 4** Convergence comparison between leader replacement and no leader replacement (see online version for colours)



(a) BLC



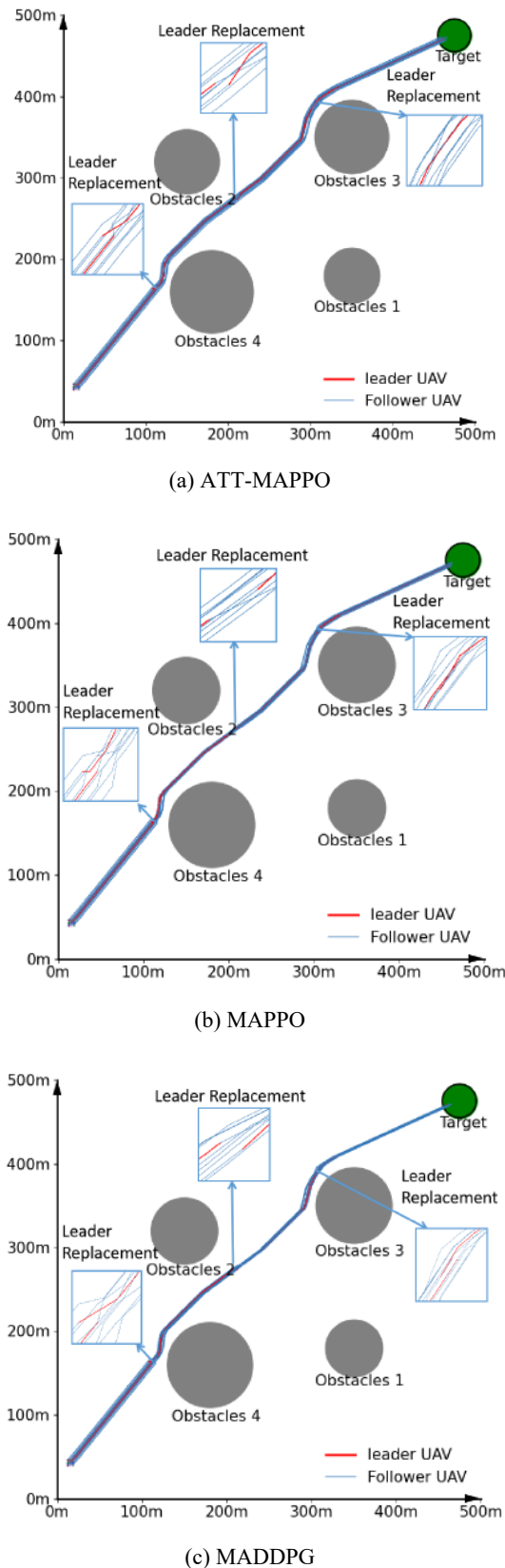
(b) MFD

#### 4.4 The trajectories of flocking algorithm

The effectiveness of our flocking algorithm is clearly demonstrated in Figure 5(a). It has been determined leader replacement occurs after completing each flight cycle within the UAV swarm. The anticipated timeframe for reaching the target position is approximately three cycles, necessitating three leader replacements. By implementing three leader replacements during the swarm flight, we consistently maintain the flocking algorithm with each replacement. This outcome is particularly gratifying as ensuring robust coordination and stability holds paramount importance in UAV swarm flights. In Figures 5(b) and 5(c), it can be observed that both the MAPPO and MADDPG fail to maintain the original flocked state after leader replacement, while the ATT-MAPPO algorithm introduces an attention

mechanism to prioritise important information, resulting in improved training results.

**Figure 5** The trajectories generated by ATT-MAPPO, MAPPO and MADDPG (see online version for colours)



## 5 Conclusion

In this paper, we propose a navigation algorithm called ATT-MAPPO for UAV flocking with replaceable leaders. By incorporating the attention mechanism and MAPPO, we achieve fused motion of the UAV swarm and enable dynamic leader swapping. The inclusion of the attention mechanism enhances the training performance of the MAPPO algorithm. Additionally, our leader-following mechanism ensures power distribution among all UAVs in the swarm is balanced, thereby increasing flight distance. Experimental results demonstrate that ATT-MAPPO outperforms both MAPPO and MADDPG algorithms. Furthermore, when considering average with leader and exchange function exhibits superior performance compared to other algorithms.

## References

- Abichandani, P., Speck, C., Bucci, D., McIntyre, W. and Lobo, D. (2021) 'Implementation of decentralized reinforcement learning-based multi-quadrotor flocking', *IEEE Access*, Vol. 9, pp.132491–132507.
- Chipade, V.S. and Panagou, D. (2021) 'Multiagent planning and control for swarm herding in 2-d obstacle environments under bounded inputs', *IEEE Transactions on Robotics*, Vol. 37, No. 6, pp.1956–1972.
- Gassara, A. and Rodriguez, I.B. (2021) 'Describing correct uavs cooperation architectures applied on an anti-terrorism scenario', *Journal of Information Security and Applications*, Vol. 58. Doi: 10.1016/j.jisa.2021.102775.
- He, Q., Feng, Z., Fang, H., Wang, X., Zhao, L., Yao, Y. and Yu, K. (2023a) 'A blockchain-based scheme for secure data offloading in healthcare with deep reinforcement learning', *IEEE/ACM Transactions on Networking*, Vol. 32, No. 1, pp.65–80.
- He, Q., Wang, Y., Wang, X., Xu, W., Li, F., Yang, K. and Ma, L. (2023b) 'Routing optimization with deep reinforcement learning in knowledge defined networking', *IEEE Transactions on Mobile Computing*, Vol. 23, No. 2, pp.1444–1455.
- Hwangbo, J., Sa, I., Siegwart, R. and Hutter, M. (2017) 'Control of a quadrotor with reinforcement learning', *IEEE Robotics and Automation Letters*, Vol. 2, No. 4, pp.2096–2103.
- Koch, W., Mancuso, R., West, R. and Bestavros, A. (2019) 'Reinforcement learning for UAV attitude control', *ACM Transactions on Cyber-Physical Systems*, Vol. 3, No. 2, pp.1–21.
- Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. and Wierstra, D. (2015) 'Continuous control with deep reinforcement learning', *arXiv:1509.02971v6 [cs.LG]*, pp.1–14.
- Luo, Q. and Duan, H. (2017) 'Distributed UAV flocking control based on homing pigeon hierarchical strategies', *Aerospace Science and Technology*, Vol. 70, pp.257–264.
- Ribeiro, R.G., Cota, L.P., Euzébio, T.A.M., Ramrez, J.A. and Guimarães, F.G. (2021) 'Unmanned-aerial-vehicle routing problem with mobile charging stations for assisting search and rescue missions in post-disaster scenarios', *IEEE transactions on systems, man, and cybernetics: Systems*, Vol. 52, No. 11, pp.6682–6696.

- Shen, G., Lei, L., Li, Z., Cai, S., Zhang, L., Cao, P. and Liu, X. (2021) 'Deep reinforcement learning for flocking motion of multi-UAV systems: learn from a digital twin', *IEEE Internet of Things Journal*, Vol. 9, No. 13, pp.11141–11153.
- Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. and Shochet, O. (1995) 'Novel type of phase transition in a system of self-driven particles', *Physical Review*, Vol. 75, No. 6. Doi: 10.1103/PhysRevLett.75.1226.
- Wang, C., Wang, J. and Zhang, X. (2018) 'A deep reinforcement learning approach to flocking and navigation of UAVs in large-scale complex environments', *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, IEEE, pp.1228–1232.
- Wang, C., Wang, J., Shen, Y. and Zhang, X. (2019) 'Autonomous navigation of UAVs in large-scale complex environments: a deep reinforcement learning approach', *IEEE Transactions on Vehicular Technology*, Vol. 68, No. 3, pp.2124–2136.
- Wang, L., Wang, K., Pan, C., Xu, W., Aslam, N. and Hanzo, L. (2020) 'Multi-agent deep reinforcement learning-based trajectory planning for multi-uav assisted mobile edge computing', *IEEE Transactions on Cognitive Communications and Networking*, Vol. 7, No. 1, pp.73–84.
- Wang, W., Wang, L., Wu, J., Tao, X. and Wu, H. (2022) 'Oracle-guided deep reinforcement learning for large-scale multi-uavs flocking and navigation', *IEEE Transactions on Vehicular Technology*, Vol. 71, No. 10, pp.10280–10292.
- Wang, Z. and Gu, D. (2009) 'Distributed leader-follower flocking control', *Asian Journal of Control*, Vol. 11, No. 4, pp.396–406.
- Yan, C., Low, K.H., Xiang, X., Hu, T. and Shen, L. (2022) 'Attention-based population-invariant deep reinforcement learning for collision-free flocking with a scalable fixed-wing UAV swarm', *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.13730–13736.
- Yan, C., Wang, C., Xiang, X., Lan, Z. and Jiang, Y. (2021) 'Deep reinforcement learning of collision-free flocking policies for multiple fixed-wing uavs using local situation maps', *IEEE Transactions on Industrial Informatics*, Vol. 18, No. 2, pp.1260–1270.
- Yan, P., Bai, C., Zheng, H. and Guo, J. (2020) 'Flocking control of uav swarms with deep reinforcement learning approach', *Proceedings of the 3rd International Conference on Unmanned Systems (ICUS)*, IEEE, pp.592–599.
- Zhao, H., Liu, H., Leung, Y-W. and Chu, X. (2018) 'Self-adaptive collective motion of swarm robots', *IEEE Transactions on Automation Science and Engineering*, Vol. 15, No. 4, pp.1533–1545.
- Zhu, F., Chen, X., Chen, S., Zheng, W. and Ye, W. (2023a) 'Relative margin induced support vector ordinal regression', *Expert Systems with Applications*. Doi: 10.1016/j.eswa.2023.120766.
- Zhu, F., Chen, X., Gao, X., Ye, W., Zhao, H. and Vasilakos, A.V. (2023b) 'Constraint-weighted support vector ordinal regression to resist constraint noises', *Information Sciences*, Vol. 649. Doi: 10.1016/j.ins.2023.119644.
- Zhu, F., Gao, J., Yang, J. and Ye, N. (2022) 'Neighborhood linear discriminant analysis', *Pattern Recognition*, Vol. 123. Doi: 10.1016/j.patcog.2021.108422.
- Zhu, F., Ning, Y., Chen, X., Zhao, Y. and Gang, Y. (2021) 'On removing potential redundant constraints for svor learning', *Applied Soft Computing*, Vol. 102. Doi: 10.1016/j.asoc.2020.106941.
- Zhu, F., Zhang, W., Chen, X., Gao, X. and Ye, N. (2023c) 'Large margin distribution multi-class supervised novelty detection', *Expert Systems with Applications*, Vol. 224. Doi: 10.1016/j.eswa.2023.119937.