

# Multidimensional Swarm Flight Approach for Chasing Unauthorized UAVs Leveraging Asynchronous Deep Learning

Tae-Won Ban <sup>1</sup>, Member, IEEE, Kyu-Min Kang <sup>2</sup>, and Bang Chul Jung <sup>3</sup>, Senior Member, IEEE

**Abstract**—This article introduces a novel uncrewed aerial vehicles (UAV) chasing system designed to track and chase unauthorized UAVs, significantly enhancing their neutralization effectiveness. The system utilizes a multidimensional swarm flight strategy, employing deep reinforcement learning (DRL) to dynamically adapt the tracking unit's movements based on the received signal strength indicators emitted by unauthorized UAVs. Asynchronous learning techniques involving multiple agents are implemented to expedite the system's learning process. A key feature of our approach is the coordinated use of a swarm of UAVs, which circumvents the considerable size burden associated with mounting multiple antennas on a single UAV. We further refine the asynchronous DRL framework by integrating advanced channel modeling techniques, such as spatial correlation and Doppler shift, to augment the robustness and adaptability of the system. Performance evaluations confirm the system's efficacy under varying channel conditions and operational scenarios. Key contributions include the integration of tracking and chasing functionalities into a unified system, the employment of realistic channel models to enhance system adaptability, and a comprehensive analysis of the relationship between channel sampling frequency and chasing performance. This research advances the field of UAV regulation and control, offering an effective solution to the escalating security challenges posed by unauthorized UAVs.

**Index Terms**—Anti-UAV, asynchronous learning, multiagent deep reinforcement learning (DRL), uncrewed aerial vehicle (UAV) chasing, UAV.

## I. INTRODUCTION

### A. Background and Motivation

ORIGINALLY developed primarily for military applications, uncrewed aerial vehicles (UAVs) have expanded into a wide array of civilian applications, including agriculture,

Received 16 April 2025; revised 24 September 2025; accepted 1 November 2025. Date of publication 19 November 2025; date of current version 14 January 2026. This work was supported by Basic Research Program-Individual Research (Global Matching Grants-Sweden) from the National Research Foundation of Korea (NRF) funded by the Korean Government (Ministry of Science and ICT) in the year 2024 under Grant RS-2024-00436923. (Corresponding author: Bang Chul Jung.)

Tae-Won Ban is with the Department of Artificial Intelligence and Information Engineering, Gyeongsang National University, Jinju 52828, South Korea (e-mail: twban35@gnu.ac.kr).

Kyu-Min Kang is with the Radio Research Division, Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea (e-mail: kmkang@etri.re.kr).

Bang Chul Jung is with the Department of Electrical and Computer Engineering, Ajou University, Suwon 16499, South Korea (e-mail: bcjung@ajou.ac.kr). Digital Object Identifier 10.1109/JSYST.2025.3629737

environmental monitoring, disaster response, and film production [1]. Supported by significant advancements in UAV technology [2], [3], these developments include miniaturization, enabling the creation of smaller, more versatile UAVs that are easily integrated into various daily tasks.

In addition, UAVs increasingly incorporate artificial intelligence (AI) to enhance their autonomy, significantly improving operational efficiency and reducing the need for human oversight. Further advancements in battery technology and propulsion systems have extended UAV endurance and range, supporting missions that require longer flight times and greater distances. These technological improvements have spurred rapid growth in the UAV market, projected to increase from 26.2 billion USD in 2022 to 38.3 billion USD in 2027, at a compound annual growth rate of 7.9% [4].

However, the increasing deployment of UAVs across various sectors raises concerns about security, privacy, and airspace regulation. Unauthorized UAVs pose risks ranging from accidental airspace intrusions to deliberate threats such as espionage and terrorism. To mitigate these risks, robust anti-UAV technologies are needed to ensure airspace safety and compliance [5]. The development of anti-UAV systems has mainly focused on identifying [6], [7], detecting [8], [9], [10], [11], and tracking [12], [13], [14], [15], [16], [17], [18], [19] unauthorized UAVs. However, most conventional approaches are limited to localizing unauthorized UAVs rather than actively chasing them to support efficient neutralization.

### B. Problem Statement and Objective

Despite significant progress in UAV tracking technologies, existing methods generally focus on monitoring unauthorized UAVs without providing real-time chasing and interception capabilities. Traditional approaches often suffer from limitations in dynamic environments where UAVs exhibit unpredictable motion and evade detection. In particular, traditional feedback-based control methods such as PID control or optimal control techniques typically rely on predictable system dynamics or cooperative control laws, which assume that the target follows a known motion model or provides feedback. Thus, they are not suitable for unauthorized UAV chasing systems. This article addresses these limitations by proposing a novel multidimensional swarm flight approach for actively chasing unauthorized UAVs, leveraging deep reinforcement learning (DRL). Unlike

conventional tracking methods, the proposed approach enables UAVs to pursue unauthorized UAVs in real time, significantly improving security measures. The primary objectives of this work are as follows.

- 1) Achieve real-time pursuit of unauthorized UAVs, beyond just tracking their locations.
- 2) Optimize control strategies for coordinated UAV swarms, ensuring rapid and precise pursuit.
- 3) Reduce tracking errors and computational complexity by utilizing asynchronous learning methods.
- 4) Improve resilience to environmental variations by incorporating spatial correlation and Doppler effects into the control framework.

These objectives collectively enhance the effectiveness of UAV interception systems, minimizing security threats while ensuring efficient airspace management.

### C. Previous Works

The Federal Aviation Administration recently introduced a rule known as remote identification, aimed at increasing accountability in UAV operations by requiring UAVs to broadcast sensitive data, such as their identity [6], [7]. However, this measure can be circumvented by rogue or unauthorized UAVs.

Detecting unauthorized UAVs is crucial for anti-UAV systems [8]. Primary detection techniques [9], [10], [11] include the following:

- 1) radar systems that analyze radio waves reflected by UAVs;
- 2) acoustic sensors that identify UAVs based on rotor noise signatures;
- 3) visual or image detection methods using cameras and infrared sensors, often enhanced by deep neural networks.

Tracking involves continuously monitoring the trajectory of detected UAVs using various technologies such as radar, acoustic sensors, visual analysis, GPS, and RF signal tracking, to maintain precise control over airspace. He et al. [12] proposed a method combining coherent and noncoherent processing to enhance micro-UAV tracking under poor signal-to-noise ratios. An et al. [13] developed a novel direction-of-arrival estimator that improves tracking of unresolved scattering centers. Jin et al. [14] examined the effects of location errors on radar-based UAV detection and introduced an adaptive beam control scheme to counteract jamming inaccuracies.

Significant advancements in UAV tracking include the following.

- 1) Background-aware correlation filters by Galoogahi et al. [15] that adjust tracking models according to environmental changes.
- 2) Continuous convolution operators by Danelljan et al. [16] for fine-grained UAV tracking.
- 3) A learning response interference suppression correlation filter by Li et al. [17] to enhance consistency and minimize distractions during UAV tracking.

Li et al. [18] also proposed a computationally efficient tracking algorithm validated using fixed-wing UAV video data, and a general architecture for UAV-to-UAV detection and tracking

within a space-air-ground integrated network was introduced to provide extensive, cost-effective tracking [19].

Meanwhile, the defense of specific regions, such as boundaries or perimeters, has been extensively studied [20], [21], [22]. These approaches primarily focus on blocking entry rather than actively pursuing and neutralizing threats. In contrast, our study adopts an active pursuit approach to enhance interception efficiency. A pursuit strategy for evaders has been explored in robotics [23], sharing similarities with our research as it investigates evader pursuit rather than static defense. However, Pierson et al. [23] limit pursuers to tracking evaders only within their Voronoi neighborhood, restricting their situational awareness. In contrast, our study leverages communication signals to track and pursue unauthorized UAVs beyond local constraints, enabling a more efficient and adaptable interception strategy.

### D. Contributions of This Work

To overcome the limitations of conventional tracking methods, this article presents a novel UAV swarm-based *chasing* system that bridges the gap between passive tracking and active interception. Relative to prior multiagent DRL and asynchronous learning works that typically assume vision/full-state sensing and fixed formations, we explicitly formulate *received signal strength indicator (RSSI)-driven 3-D swarm control with single-antenna micro-UAVs and learnable swarm geometry*. The key contributions are as follows.

- 1) *From tracking to pursuit*: Development of an integrated tracking-and-chasing system using asynchronous DRL to actively pursue and intercept unauthorized UAVs, rather than just estimating their locations. This represents a shift from passive surveillance to closed-loop interception.
- 2) *Single-antenna swarm with learnable geometry*: Introduction of a swarm flight strategy leveraging multiple UAVs, each equipped with a single antenna, to avoid the spatial-correlation limitations of multi-antenna single-airframe designs. Crucially, we treat the *inter-UAV spacing* as a *control action* that is learned online, enabling geometry adaptation that mitigates correlation and improves RF-based localization during the chase.
- 3) *Asynchronous multihead A3C for 3-D control*: A new asynchronous, *multihead* A3C formulation decomposes control into four orthogonal policies ( $d, x, y, z$ ), trained without replay for stability and low memory pressure. This decoupling makes the controller lightweight at inference time and suitable for edge deployment.
- 4) *Channel-aware learning and sampling analysis*: Implementation of realistic RF channel modeling (spatial correlation and Doppler) and a systematic analysis of how the *channel sampling rate* interacts with correlation and swarm geometry to affect steps-to-capture and total time. This coupling of RF sampling with DRL control has not been quantified in prior UAV chasing/tracking studies.
- 5) *Comparative evaluation and feasibility*: Head-to-head comparisons against an RSSI-max greedy baseline and a genie-aided (optimal, brute-force) oracle clarify where learning helps and how far we are from the ideal

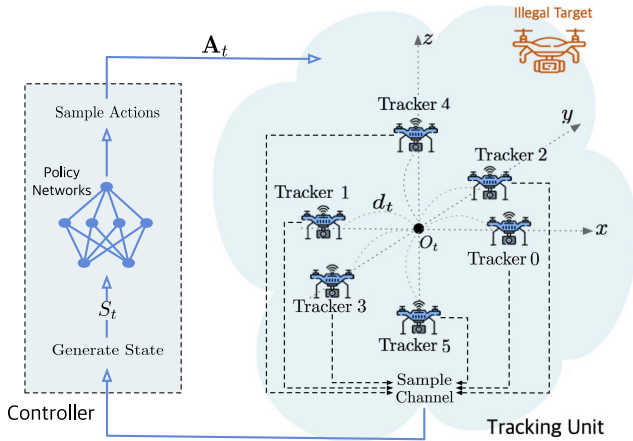


Fig. 1. Proposed system based on swarm flight using DRL for chasing unauthorized UAVs.

bound. We also provide an analytic multiply–accumulate (MAC)/FLOP count and device-level latency (CPU/GPU) showing 2.9~5.5 ms per decision and a compact model, demonstrating real-time feasibility without vision pipelines.

- 6) *Empirical robustness*: Empirical validation under diverse channel conditions and operating regimes shows consistent gains in interception reliability and competitiveness in end-to-end time.

By integrating swarm intelligence with a problem-specific asynchronous multihead DRL controller and learnable swarm geometry, the proposed system transforms UAV-based security operations from passive monitoring into proactive intervention, while remaining computationally lightweight and deployable in real time.

### E. Article Organization

The rest of this article is organized as follows. Section II describes the system architecture, including swarm flight methodologies and DRL techniques. Section III details the control framework, emphasizing UAV swarm coordination strategies. Section IV presents numerical results evaluating the system’s performance across varying operational conditions. Finally, Section V concludes this article.

## II. SYSTEM MODEL

### A. Overall System Architecture

In this article, we introduce a system for chasing unauthorized UAVs based on swarm flight using DRL. While identifying unauthorized UAVs is beyond the scope of our work, Section I-C discusses various methods for UAV identification, such as remote ID systems. In addition, in critical areas such as nuclear power plants or high-security zones, any UAV entering a predefined restricted airspace can be classified as unauthorized. The system comprises a controller and a tracking unit consisting of six compact UAVs, as illustrated in Fig. 1. The swarm structure of our tracking UAVs is modeled as a *rectangular bipyramid* to support 3-D pursuit, where each UAV is positioned at a vertex of

the shape in 3-D space with two UAVs along each of the  $x$ -,  $y$ -, and  $z$ -axes. Each UAV is denoted as  $u$ ,  $0 \leq u \leq 5$ . At a specific time  $t$ , the center point of the tracking UAVs is denoted as  $O_t$  and its coordinates are  $(x_t, y_t, z_t)$ . The distance between  $O_t$  and each UAV is represented as  $d_t$ .

The tracking UAVs intercept communication signals that are exchanged between an illegal target UAV and its controller that refers to the entity responsible for operating the unauthorized UAV. This can typically be a ground control station, a remote pilot using a handheld transmitter, or a network-based control system issuing commands through a wireless communication link. The tracking UAVs then measure RSSI values at a given sampling frequency  $F$  [samples/s]. Subsequently, the RSSI values are transmitted to the controller through a cellular network or wireless local area network (WLAN). The controller utilizes the RSSI values to adjust both the movement of the tracking UAVs and the spacing between the tracking UAVs.

If the tracking unit is equipped with multiple RF antennas on a single UAV for system simplicity, the increased spatial correlation among antennas could lead to reduced diversity in the RSSI values measured by each antenna [24]. Consequently, the unit’s ability to accurately predict the direction of the target may be compromised, impairing chasing performance. Furthermore, such a configuration might struggle to operate reliably in 3-D spaces due to challenges in multidimensional arrangement. Given these considerations, we decided to configure the tracking unit with multiple compact UAVs, each equipped with a single RF receiver, accepting increased system complexity as a tradeoff for enhanced performance and reliability, particularly in 3-D spaces.

The controller employs a policy-gradient DRL approach, specifically the asynchronous advantage actor-critic (A3C) model, to manage the movements of the tracking unit. A3C is selected for its asynchronous training, which accelerates learning while eliminating the need for experience replay, making it computationally efficient for UAVs. Its training paradigm further reduces memory overhead by operating without experience replay. In addition, A3C is more robust to environmental variations and requires minimal hyperparameter tuning, ensuring stable performance in dynamic UAV pursuit tasks.

The controller leverages RSSI values from tracking UAVs to make real-time decisions. Initially, the control unit constructs the input state  $S_t$  based on these RSSI values. This state is then processed by policy networks, which stochastically determine the optimal spacing between tracking UAVs and the unit’s movements along the  $x$ -,  $y$ -, and  $z$ -axes. This approach enhances tracking precision and improves responsiveness to dynamic target movements.

### B. Channel Model

We posit that an unauthorized UAV communicates with its controller via a wireless network. For a given distance  $d$  between such an illegal UAV and a tracking UAV, the RSSI value measured at the tracking UAV is defined as

$$\eta(d) = P_{tx} - L(d) + \xi \text{ [dBm]}. \quad (1)$$

Here,  $P_{tx}$  represents the transmit power utilized by the unauthorized UAV for communication, and  $L(d)$  is the path loss corresponding to distance  $d$ . The path loss  $L(d)$  is further detailed as

$$L(d) = L_0 + 10n \log_{10} \left( \frac{d}{d_0} \right) \text{ [dB]} \quad (2)$$

where  $d_0$ ,  $L_0$ , and  $n$  denote the reference distance, the path loss at that reference distance, and the path loss exponent, respectively. The term  $\xi$  in (1) denotes small-scale fading.

We model small-scale fading as Rician, which is appropriate for air-to-air UAV links where a dominant line-of-sight component is typical. The envelope is thus Rician with a  $K$ -factor that qualitatively captures the LOS-to-multipath power ratio. Our simulator also includes spatial correlation across the swarm and Doppler dynamics, producing time-varying, correlated RSSI representative of airborne pursuit. Using six spatially separated single-antenna UAVs further provides spatial diversity that mitigates multipath bias. If operations move to more cluttered environments, the same framework can reduce the  $K$ -factor toward the Rayleigh case, without changing the controller or training procedure.

In this article, we thoroughly explore the complex dynamics of wireless communication channels. Initially, we employ a correlation matrix framework to quantify the spatial correlation among antennas mounted on each tracking UAV. Spatial correlation characterizes the interdependence of signal strengths received at different locations, which primarily arises due to environmental factors such as terrain and the specific configuration of antenna deployment [25]. In this context, we define the spatial correlation matrix  $\mathbf{C}$  as [26]

$$\mathbf{C} = [c_{ij}]_{0 \leq i, j \leq 5} \quad (3)$$

where  $c_{ij}$  represents the correlation coefficient between two tracking UAVs  $i$  and  $j$ , defined by

$$c_{ij} = \begin{cases} 1 & \text{if } i = j \\ \rho & \text{otherwise.} \end{cases} \quad (4)$$

Unlike traditional spatial correlation matrices that exhibit varying correlation coefficients between antenna elements, our approach maintains a uniform correlation coefficient  $\rho$  across all tracking UAVs. This uniformity stems from the symmetrical structure of the tracking UAVs, which resemble a square pyramid.

Furthermore, we enhance the realism of our channel model by accounting for Doppler shift, which affects channel behavior over time. Doppler shift, represented by  $f_d$ , is defined as

$$f_d = \frac{v}{c} f_c \quad (5)$$

where  $v$ ,  $c$ , and  $f_c$  denote the relative velocity of tracking UAVs, the speed of light, and the carrier frequency, respectively [27], [28].

### III. CONTROL FRAMEWORK AND ASYNCHRONOUS TRAINING USING TRACKING CONTROLLER

#### A. Forward Pipeline for the Control Framework

As outlined in Section II, this article adopts the A3C model, leveraging its significant benefits, including asynchronous training that boosts learning speed and stability through the use of multiple agents. It combines value-based and policy-based methods to develop robust policies, effectively manages high-dimensional state and action spaces with deep neural networks, and exhibits reduced sensitivity to hyperparameter settings. These features make the A3C model widely applicable and efficient across diverse tasks and environments [29].

The controller proposed in this study utilizes the A3C model as depicted in Fig. 1 and is equipped with four distinct policy networks. Each of the four policy networks is uniquely parametrized by  $\theta^n$ , specializing in different tracking dimensions where  $n \in \{d, x, y, z\}$ .  $\theta^d$  adjusts the spacing between the tracking UAVs, as indicated by  $d_t$  in Fig. 1, to ensure optimal positioning for pursuing an unauthorized target. Meanwhile,  $\theta^x$ ,  $\theta^y$ , and  $\theta^z$  manage the movements of the tracking unit along the  $x$ -,  $y$ -, and  $z$ -axes, respectively, enabling precise maneuverability in 3-D space.

1) *State*: The input state  $S_t$  at any given time  $t$  is constructed using the RSSI values collected by each tracking UAV. The matrix representation of  $S_t$  is defined as

$$S_t = [\eta_{iu}]_{0 \leq i \leq L-1, 0 \leq u \leq 5} \quad (6)$$

where  $\eta_{iu}$  represents the  $i$ th RSSI sample measured by the  $u$ th UAV. The variable  $L$  indicates the total number of RSSI samples measured by each UAV within this interval, and the UAV index  $u$  runs from 0 to 5, assuming there are six tracking UAVs in the tracking unit as depicted in Fig. 1.

Each RSSI sample  $\eta_{iu}$  is obtained at a sampling frequency  $F$  [samples/s], which determines how often RSSI measurements are recorded within the given interval. The arrangement of these RSSI values in matrix is depicted in Fig. 2. This formulation implies that each row in matrix  $S_t$  corresponds to the synchronized RSSI samples taken by all UAVs at a single sampling instance, while each column represents a series of RSSI samples from a single UAV across all sampling instances.

2) *Policy and Value Functions*: As depicted in Fig. 2, input state  $S_t$  is processed by four separate neural networks, which consist of long short-term memory (LSTM) layers and fully connected layers. LSTMs are particularly well-suited for tasks where understanding the temporal dynamics of input sequences is crucial. They are capable of learning long-term dependencies in the data, avoiding the vanishing gradient problem common in standard recurrent neural networks. Following the LSTM layers, the final hidden state of the LSTM layers is then passed through fully connected (dense) layers. Each of these fully connected layers uses the rectified linear unit activation function.

Each network produces two critical outputs: a policy  $\pi(a|S_t; \theta^n)$  and a value function  $V(S_t; \theta^n)$ . The policy output specifies the probability distribution across possible actions, directing the agent's decision making by indicating the most favorable action based on the current state  $S_t$ . Simultaneously,

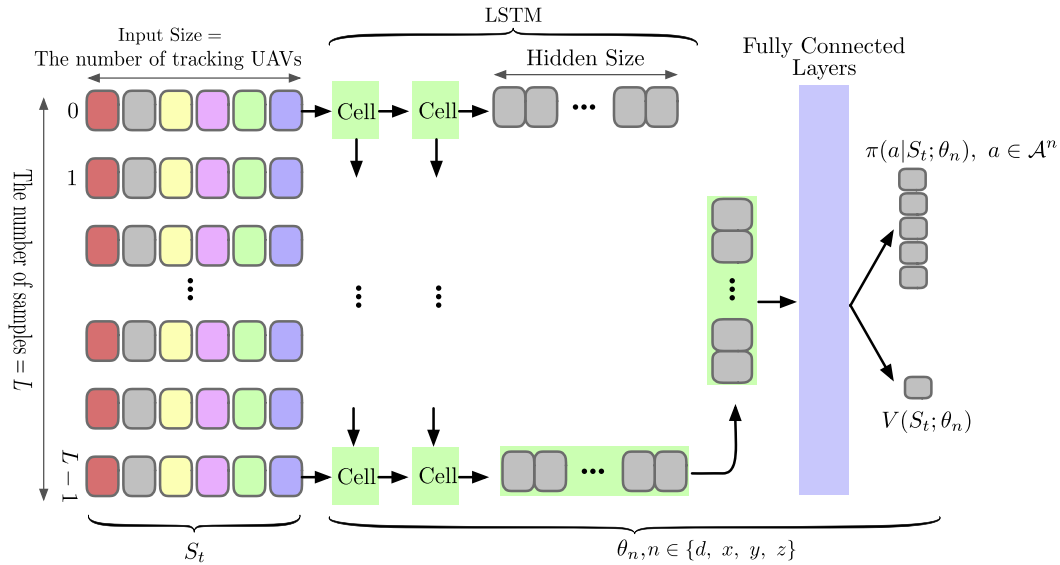


Fig. 2. Architecture of the proposed neural network.

the value output provides an estimation of the expected cumulative reward from taking a specific action in the given state, thereby informing the agent about potential future benefits. This design ensures that the agent remains adaptive and responsive to changes in the environment, enhancing its ability to make informed decisions that consider both present circumstances and future possibilities.

3) *Actions*: An action  $A_t^n$  is stochastically sampled based on the policy  $\pi(a|S_t; \theta^n)$ . This action is drawn from its designated action space  $\mathcal{A}^n$ . Specifically,  $\mathcal{A}^d = \{1, 2, 3, 4, 5\}$  [m] and  $\mathcal{A}^x, \mathcal{A}^y, \mathcal{A}^z = \{-4, -2, 0, 2, 4\}$  [m]. In the proposed model, four action spaces are designed to be orthogonal to each other, underscoring the fact that actions generated by different policy networks do not interfere with each other. This orthogonality ensures that adjustments in one dimension do not inadvertently affect the performance in the other dimensions. As a result, the proposed tracking system can execute complex maneuvers with precision, as the implementation of one action does not necessitate compensatory adjustments in the others.

### B. Training of the Proposed Model

The tracking controller adjusts the positions of the tracking UAVs to carry out a determined set of actions,  $\mathbf{A}_t = [A_t^d, A_t^x, A_t^y, A_t^z]$ . After implementing these actions, the controller observes the subsequent state  $S_{t+1}$  and collects associated rewards,  $\mathbf{R}_{t+1} = [R_{t+1}^d, R_{t+1}^x, R_{t+1}^y, R_{t+1}^z]$ . Utilizing these observations and rewards, the controller then computes losses and updates the parameters of the neural networks to minimize the losses.

1) *Rewards*: In this study, we have developed a specialized reward structure tailored for four distinct neural networks, each designed to generate orthogonal actions within a multiagent DRL framework. This unique reward system is critical for aligning the learning process with the individual objectives of each

policy network, thereby enhancing the tracking performance of our UAV chasing system.

For each axis-aligned action  $A_t^n$  where  $n \in \{x, y, z\}$ , the reward  $R_{t+1}^n$  is computed based on the decrease in the absolute difference in the corresponding-axis coordinates between the tracker's center point and the target UAV. Specifically, the reward is calculated as

$$R_{t+1}^n = |O_t[n] - U[n]| - |O_{t+1}[n] - U[n]| \quad (7)$$

for each  $n$  axis coordinate. In addition, the reward for the action  $A_t^d$  is defined as the reduction in the Euclidean distance between the tracker and the target UAV

$$R_{t+1}^d = |O_t - U| - |O_{t+1} - U| \quad (8)$$

where  $|O_t - U|$  denotes the Euclidean distance between the tracking center point  $O_t$  and the target UAV  $U$ .

These tailored rewards ensure that the learning outcomes are directly tied to the effectiveness of the UAV's tracking ability, promoting a more efficient and targeted learning process across different control dimensions.

2) *Losses*: In this study, the loss function is defined with three primary components: policy loss, value loss, and an entropy bonus. These components are strategically combined to optimize the policy and value function networks across the update interval from  $t$  to  $t+T-1$ . Each neural network, indexed by  $n$  from the set  $\{d, x, y, z\}$  and parametrized by  $\theta^n$ , processes a sequence of  $T$  observations  $((S_t, A_t^n, R_{t+1}^n, S_{t+1}), \dots, (S_{t+T-1}, A_{t+T-1}^n, R_{t+T}^n, S_{t+T}))$  over  $T$  time steps starting from time  $t$ .

The policy loss  $L_p^n$  is formulated to encourage actions that are expected to yield higher returns. It utilizes the advantage function, which measures the relative benefit of selected actions against a baseline expectation. This is accomplished by calculating the difference between the temporal-difference (TD) target and the current value estimate. The advantage function at time

$t$  is defined as

$$\text{ADV}_t^n = G_t^n - V(S_t; \theta^n) \quad (9)$$

where  $G_t^n$  represents the TD target at state  $S_t$ , and  $V(S_t; \theta^n)$  is the estimated value of the current state. Typically,  $G_t^n$  is calculated using the value of the subsequent state  $V(S_{t+1}; \theta^n)$  combined with the immediate reward, but in this study, we compute it using the discounted sum of future rewards

$$G_t^n = \sum_{k=1}^{T-t+1} \gamma^{k-1} R_{t+k}^n \quad (10)$$

where  $\gamma$  is the discount factor, prioritizing immediate rewards and diminishing the importance of future rewards. If the final state is terminal,  $R_{T+1}^n = 0$ ; otherwise,  $R_{T+1}^n = V(S_T; \theta^n)$ , indicating the absence of future rewards beyond the terminal state. This method of calculating  $G_t^n$ , which incorporates a series of future rewards, provides a comprehensive assessment of potential outcomes, enhancing the robustness and accuracy of the advantage function. This approach is particularly effective when  $T$ , the length of the trajectory, is long. It allows the system to account for the extended consequences of actions, ensuring that the advantage function reflects the cumulative impact over the entire trajectory.

The policy loss is subsequently calculated as

$$L_P^n = -\frac{1}{T} \sum_{t=0}^{T-1} \log \pi(A_t^n | S_t; \theta^n) \cdot \text{ADV}_t^n. \quad (11)$$

The value loss  $L_V^n$  is crucial for aligning the predictions of the value function with the actual returns observed. To improve model stability and accuracy, the value loss is calculated using a TD approach that directly leverages the next-step value predictions to form a more responsive learning update.  $L_V^n$  is defined using the absolute difference (L1-loss) between the predicted value at a state and the TD target. This can be mathematically expressed as

$$L_V^n = \frac{1}{T} \sum_{t=0}^{T-1} |R_{t+1}^n + \gamma V(S_{t+1}; \theta^n) - V(S_t; \theta^n)| \quad (12)$$

where  $V(S_T; \theta^n) = 0$  if  $S_T$  is a terminal state.

To encourage exploration and prevent premature convergence to suboptimal policies, an entropy bonus, denoted as  $H^n$ , is incorporated into the loss calculation. This entropy bonus measures the uncertainty of the policy distribution at each time step and is calculated as

$$H^n = -\frac{1}{T} \sum_{t=0}^{T-1} \sum_{a \in A^n} \pi(a|S_t; \theta^n) \log \pi(a|S_t; \theta^n). \quad (13)$$

Combining all these components, the total loss over an update interval  $T$  is formulated as

$$L^n = L_P^n + L_V^n - \beta H^n \quad (14)$$

where  $\beta$  is a coefficient that adjust the relative importance of entropy bonus and  $H^n$  is subtracted from the loss to promote a more exploratory and diverse policy behavior.

3) *Shared Learning and Asynchronous Update*: Based on the loss formulation in (14), each agent independently computes the gradients for their local parameters across four neural networks, denoted as  $\nabla \theta = \{\nabla \theta^n | n \in d, x, y, z\}$ . These gradients are then transmitted to a central global agent, which asynchronously updates the network parameters by incorporating the experiences gathered from all local agents. This shared learning approach enhances the efficiency of policy and value function estimation by leveraging the collective insights of multiple agents. Furthermore, the asynchronous update mechanism eliminates the need for strict synchronization, allowing for more continuous and scalable training cycles.

The overall architecture and training algorithm for the proposed model are depicted in Fig. 3 and detailed in Algorithm 1, respectively. This framework integrates  $N$  local agents and one global agent, all operating in a multithreaded or multiprocessor environment to enable shared learning and asynchronous updates effectively. An additional auxiliary agent is incorporated to assess intermediate training outcomes periodically. Should these outcomes meet predefined performance criteria, training may be terminated early, thereby enhancing the efficiency and effectiveness of the process.

#### IV. NUMERICAL RESULTS AND DISCUSSION

In this section, we evaluate the performance of the UAV chasing system, as described in Section III. The analysis considers various channel sampling frequencies and spatial correlation coefficients, denoted as  $F$  and  $\rho$ , respectively. The target UAV is positioned at a distance of 100 m from the tracking system's center and is assumed to be randomly located on the surface of a sphere with a radius of 100 m, ensuring  $|O_0 - U| = 100$  m.

Key parameters related to the channel model, neural networks, and training processes are detailed in Table I. The relative velocity of the tracking UAVs is set at 0.1 m/s, while the frequency emitted by the target UAV is 3 GHz. The resulting Doppler shift, calculated according to (5), is 1 Hz. In addition, the minimum distance threshold for successful tracking and chasing is specified as  $th$  in Table I.

Fig. 4 presents examples of small-scale fading in the channels between each of the six tracking UAVs and a target UAV, under a Doppler shift condition of  $f_d = 1$  Hz. The influence of channel sampling frequency,  $F$ , on the temporal resolution of the channel's characteristics is highlighted. As depicted in Fig. 4(a), a higher sampling frequency ( $F = 20$  Hz) provides finer resolution, capturing detailed fluctuations in the fading process. Conversely, a lower sampling frequency ( $F = 10$  Hz) as shown in Fig. 4(b), results in more pronounced variations between samples due to longer intervals, which can capture significant changes but at the cost of longer channel sampling periods. Furthermore, a low spatial correlation coefficient [ $\rho = 0.1$  in Fig. 4(a)] indicates less correlation among the channel paths, suggesting a greater diversity in path characteristics. In contrast, a high spatial correlation coefficient [ $\rho = 0.9$  in Fig. 4(b)] indicates highly correlated channel conditions, likely due to similar physical or environmental factors.

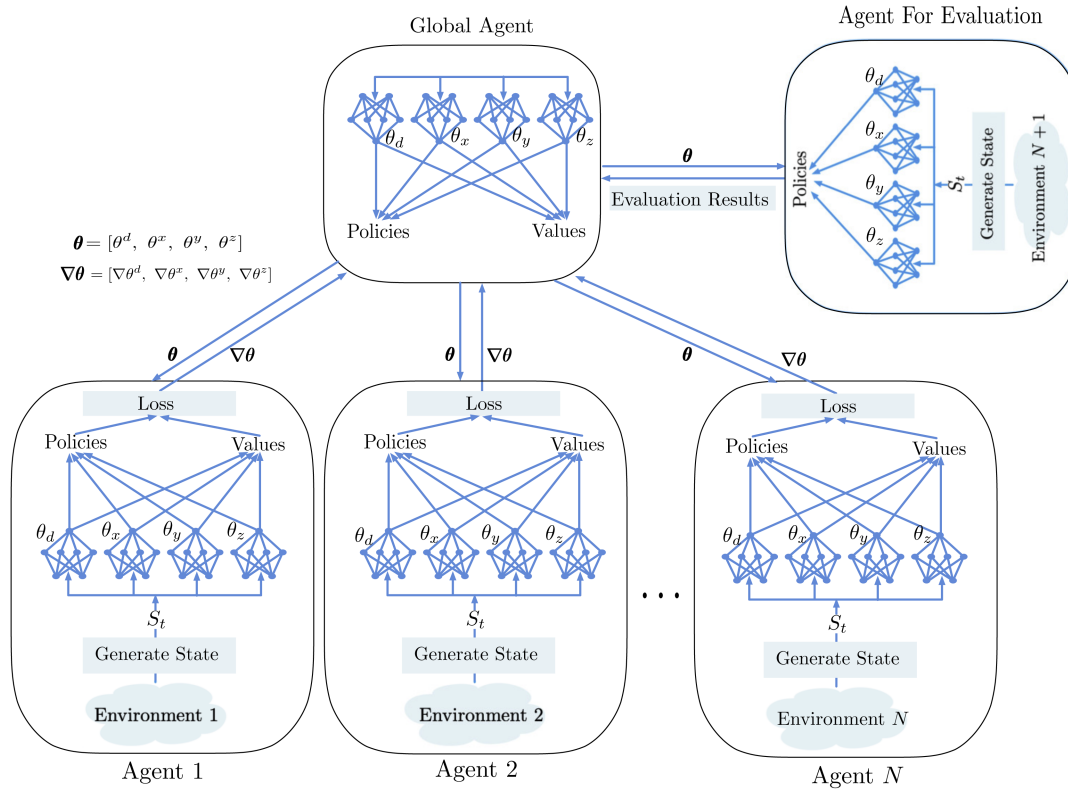


Fig. 3. Training architecture using shared learning and asynchronous updates.

The proposed neural networks, denoted as  $\theta^n, n = \{d, x, y, z\}$ , have been asynchronously trained by eight local agents, as detailed in Table I. Fig. 5 displays the average rewards obtained over 100 episodes from these agents, illustrating a progressive improvement in all networks. Notably, the reward for  $\theta^d$  remains relatively stable and consistently higher compared to those for  $\theta^x, \theta^y$ , and  $\theta^z$ . Consistent with standard policy-gradient theory, A3C is on-policy, and under typical stochastic-approximation conditions (bounded rewards/gradients, suitably small step sizes, and entropy regularization), actor-critic updates are well behaved and converge to a stationary point of the performance objective [29]. Empirically, Fig. 5 shows steady, monotonic improvement, and plateauing of the average training reward across episodes without signs of divergence, supporting practical convergence.

Direct comparison with prior UAV tracking methods is difficult because our work introduces a new *chasing* platform: an RSSI-driven swarm of multiple UAVs, each with a single RF antenna, with dynamic inter-UAV spacing and asynchronous DRL for interception. Most existing works target different tasks (detection, localization, and vision tracking), use different sensors and action spaces, and optimize different success criteria. To provide a fair reference point, we implemented two task-aligned nonlearning baselines in the same simulator, using the same observations (RSSIs), action spaces, termination criteria, and episode budgets.

- 1) *RSSI-max greedy*: Select the column with the largest mean RSSI from the  $L \times 6$  matrix (columns ordered  $[x^+, x^-, y^+, y^-, z^+, z^-]$ ) and move the center one grid

increment along that axis; spacing is kept unchanged (myopic, snapshot-based).

- 2) *Genie-aided (oracle, idealized)*: Assumes perfect, instantaneous knowledge of the true target and tracker states and exhaustively evaluates all discrete actions. It then selects the action that succeeds in one step if possible; else maximizes the immediate decrease in center-to-target distance; else minimizes the center displacement. This oracle is not implementable in real systems and serves only as an unattainable lower bound on required steps.

The tracking system's performance was assessed over 10 000 episodes across sampling frequencies  $F \in \{10, 20, 50, 100\}$  and spatial correlation coefficients  $\rho \in \{0.1, 0.5, 0.9\}$  using disjoint test sets. Figs. 6–8 summarize the results.

Fig. 6(a)–(c) shows the cumulative distribution of movements required by the proposed policy to reach the 2-m success criterion. Across all settings, success occurs well within 500 movements, indicating robustness. As spatial correlation increases from  $\rho = 0.1$  to  $\rho = 0.9$ , the curves steepen. Lower correlation ( $\rho = 0.1$ ) yields less coherent channels and more variable measurements, making disambiguation harder; higher correlation ( $\rho = 0.9$ ) simplifies disambiguation and reduces movements. Sampling frequency shows a similar tradeoff: higher  $F$  produces closely spaced, partly redundant observations, whereas lower  $F$  increases the intersample interval and often provides more diverse, informative snapshots. Consistent with this, decreasing  $F$  shifts the distributions leftward (fewer movements) for all  $\rho$ .

With the baselines overlaid, the proposed policy generally lies to the left of the RSSI-max greedy tracker for every  $(F, \rho)$ ,

**Algorithm 1:** Training Algorithm of the Proposed Model.

```

1 Generate local networks
2 for  $ep = 0$  to  $EPISODES$  do
3   Initialize environment and state  $S_t$ 
4    $done \leftarrow \text{False}$ 
5    $cnt = 0$ 
6   while not  $done$  do
7     Clear memory
8     Synchronize local networks with global
       networks
9     for  $t = 0$  to  $T - 1$  do
10      forall  $n \in \{d, x, y, z\}$  do
11        | Select  $A_t^n$  from  $\pi(a|S_t; \theta^n)$ 
12         $\mathbf{A}_t = [A_t^d, A_t^x, A_t^y, A_t^z]$ 
13        Execute  $\mathbf{A}_t$ 
14        Observe  $\mathbf{R}_{t+1}$  and  $S_{t+1}$ 
15        Store  $(S_t, \mathbf{A}_t, \mathbf{R}_{t+1}, S_{t+1})$ 
16         $S_t \leftarrow S_{t+1}$ 
17        if  $cnt \geq MAX\_STEPS$  then
18          |  $done \leftarrow \text{True}$  // Failure
19          if  $|O_{t+1} - U| \leq th$  then
20            |  $done \leftarrow \text{True}$  // Success
21          if  $done$  then
22            | break
23           $cnt = cnt + 1$ 
24        forall  $n \in \{d, x, y, z\}$  do
25           $G_{t+1}^n = 0$  if terminal, else  $G_{t+1}^n = V(S_{t+1}; \theta^n)$ 
26           $L_P^n = 0; L_V^n = 0; H^n = 0$ 
27          for  $k = t$  to  $0$  do
28            Calculate the discounted sum of
              rewards and advantage:
29             $G_k^n = R_{k+1}^n + \gamma G_{k+1}^n$ 
30             $ADV_k^n = G_k^n - V(S_k; \theta^n)$ 
31            Calculate the policy (actor) loss:
32             $L_P^n = L_P^n - ADV_k^n \cdot \log \pi(A_k^n | S_k; \theta^n)$ 
33            Calculate the critic loss:
34             $L_V^n = L_V^n + |ADV_k^n|$ 
35            Calculate the entropy:
36             $H^n = H^n - \sum_{a \in \mathcal{A}^n} \pi(a | S_k; \theta^n)$ 
               $\times \log \pi(a | S_k; \theta^n)$ 
37            Calculate the overall loss:
38             $L^n = \frac{1}{t+1} [L_P^n + L_V^n - \beta H^n]$ 
39            Calculate the gradients  $\nabla \theta^n(L^n)$ 
40            Transfer  $[\nabla \theta^d, \nabla \theta^x, \nabla \theta^y, \nabla \theta^z]$  to the global agent
41          forall  $n \in \{d, x, y, z\}$  do
42            |  $\theta^n \leftarrow \theta^n - lr \cdot \nabla \theta^n$  // Global agent

```

indicating a higher probability of finishing within any given movement budget. The genie-aided oracle appears as the tight leftmost reference because it chooses the best action with perfect state access; it is channel-agnostic and uses no RSSI sampling, so it serves only as an unattainable lower bound on steps (around a few tens of movements). In short, the proposed controller stochastically dominates the greedy baseline across

TABLE I  
PARAMETERS

Parameters		Values
$P_{tx}$	Transmit power of target UAV	23 dBm
$d_0$	Reference distance of path loss	1 m
$L_0$	Reference path loss at $d_0$	30 dB
$n$	Path loss exponent	2.6
$K$	$K$ -factor of Rician distribution	3
$f_d$	Doppler spread	1 Hz
$L$	The length of input states $S_t$	50
$T$	Update interval	5
$\gamma$	Discounting factor	0.99
$\beta$	Coefficient for entropy bonus	0.01
$th$	Min. required distance btw tracking and target UAVs required for success	2 m
MAX_STEPS	Max. steps allowable for success	500
Number of LSTM layers		3
Number of features in the hidden state of LSTM		128
Number of linear layers		2
Number of features of linear layers		128
Optimizer		Adam
Learning rate		$10^{-5}$
Number of local agents		8

$\rho \in \{0.1, 0.5, 0.9\}$  and  $F \in \{10, 20, 50, 100\}$ , while the oracle clarifies the remaining gap to idealized performance.

Fig. 7 reports the 90th percentile movements, highlighting near worst-case behavior. For the proposed controller, required movements drop sharply as  $\rho$  increases: across  $F \in \{10, 20, 50, 100\}$  the 90th-percentile count falls from about 42–182 moves at  $\rho = 0.1$  to 44–166 at  $\rho = 0.5$ , and further to 26–49 at  $\rho = 0.9$ . This reflects that more coherent channels make RSSI-based disambiguation easier. The sampling frequency  $F$  shows the expected tradeoff: lower  $F$  generally leads to fewer movements because successive snapshots are more diverse and thus more informative, whereas higher  $F$  can yield redundant observations and extra short hops. Modest nonmonotonicity appears in some regimes (for example, at  $\rho = 0.1$  the 50-Hz case is worse than 100 Hz) due to stochastic fading, the discrete action grid, and the learned approximate policy.

Against the baselines, the proposed policy consistently requires fewer movements than RSSI-max greedy across all  $(F, \rho)$ . Representative gaps include: at  $\rho = 0.5$ ,  $F = 50$  Hz, 57 versus 159 moves (−64%); at  $\rho = 0.9$ ,  $F = 100$  Hz, 49 versus 108 (−55%); and at  $\rho = 0.1$ ,  $F = 10$  Hz, 42 versus 112 (−63%). The genie-aided oracle remains near 22 moves for all settings because it assumes perfect state knowledge and exhaustive action evaluation; it serves only as an idealized lower bound

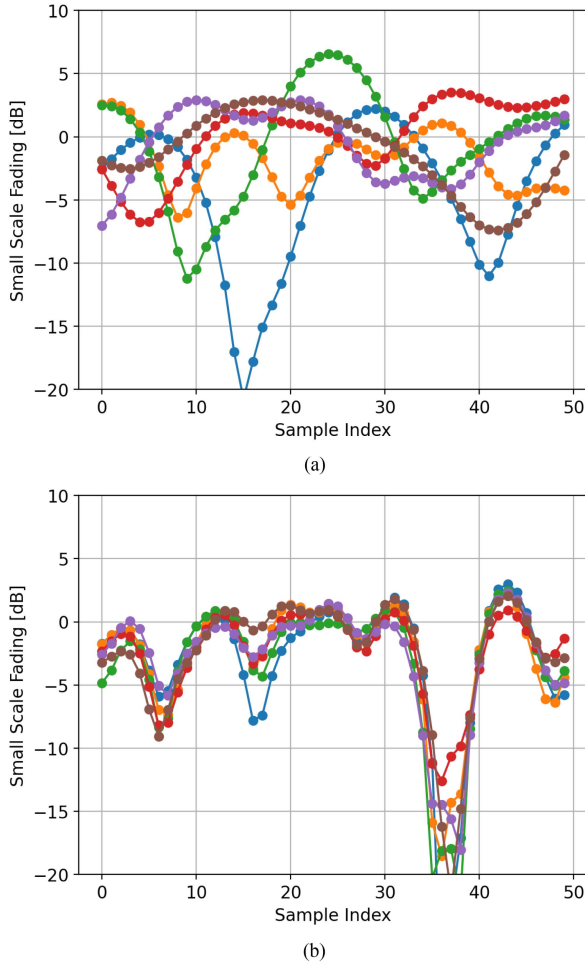


Fig. 4. Examples of small-scale fading of six channels between each tracking UAV and target UAV when  $f_d = 1$  Hz. (a)  $F = 20$  Hz and  $\rho = 0.1$ . (b)  $F = 10$  Hz and  $\rho = 0.9$ .

and clarifies the remaining headroom relative to implementable schemes.

Fig. 8 reports the end-to-end tracking time (channel observation plus travel) for  $v = 2$  m/s and  $v = 5$  m/s, with bars decomposed into observation time (lighter) and travel time (darker). The total time for  $M$  movements at sampling rate  $F$  and platform speed  $v$  is

$$\tau(M, F) = M \frac{L}{F} + \sum_{m=1}^M \frac{d_m}{v} \text{ [s]}. \quad (15)$$

Increasing the platform speed from 2 to 5 m/s shortens the travel component and therefore the total. Varying the sampling rate  $F$  introduces opposing effects: larger  $F$  reduces the per-decision observation cost, while denser measurements may increase the number of moves  $M(F, \rho)$  by encouraging short hops. Consequently, the dependence on  $F$  is not strictly monotonic. At higher spatial correlation ( $\rho$  large), a larger  $F$  often lowers the total because observation cost shrinks markedly while  $M$  grows modestly; at lower  $\rho$ , a smaller  $F$  can be preferable because each sample carries more information and  $M$  remains small.

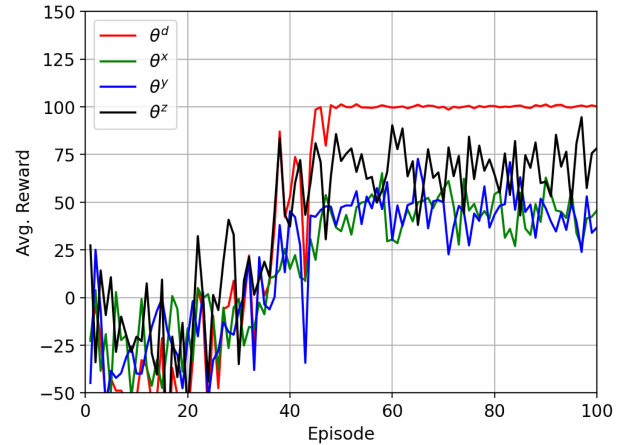


Fig. 5. Average training reward when  $F = 100$  Hz and  $\rho = 0.5$ .

For  $v = 2$  m/s, the proposed controller shows the expected tradeoffs across  $\rho$  and  $F$ . As channels become more coherent (larger  $\rho$ ), totals drop substantially (e.g., at  $F=100$  Hz: 326  $\rightarrow$  315  $\rightarrow$  130 s when  $\rho$  goes 0.1  $\rightarrow$  0.5  $\rightarrow$  0.9). The dependence on  $F$  is not monotone because larger  $F$  lowers per-decision observation time but can increase the number of moves, consistent with (15). Representative gaps versus the greedy baseline show sizable reductions where sampling is costly or channels are coherent: at  $\rho=0.1$ ,  $F=10$  Hz, the totals are 247 s (proposed) versus 586 s (greedy), a  $\sim 58\%$  reduction; at  $\rho=0.5$ ,  $F=50$  Hz, they are 168 s versus 263 s ( $\sim 36\%$  lower), and at  $F=20$  Hz, 170 s versus 361 s ( $\sim 53\%$  lower); at  $\rho=0.9$ , the proposed method is lower for all  $F$  (e.g.,  $F=100$  Hz: 130 s versus 142 s,  $\sim 9\%$  lower;  $F=50$  Hz: 122 s versus 173 s,  $\sim 30\%$  lower). By contrast, at high  $F$  and moderate/low  $\rho$ , greedy can be competitive or slightly better (e.g.,  $\rho=0.5$ ,  $F=100$  Hz: 315 s versus 254 s).

For  $v = 5$  m/s, totals drop across the board because the travel component shrinks, while the same nonmonotone  $F$  effect remains [again consistent with (15)]. The proposed controller is competitive or superior in most regimes: at  $\rho=0.1$ , it is lower at  $F=20, 50, 100$  Hz (e.g., 169 s versus 361 s at 20 Hz,  $\sim 53\%$  lower; 197 s versus 216 s at 50 Hz,  $\sim 9\%$  lower; 158 s versus 527 s at 100 Hz,  $\sim 70\%$  lower), and only slightly above greedy at  $F=10$  Hz (191 s versus 184 s). At  $\rho=0.5$ , the proposed controller is lower at  $F=10, 20, 50$  Hz (e.g., 179 s versus 461 s at 10 Hz,  $\sim 61\%$  lower; 117 s versus 299 s at 20 Hz,  $\sim 61\%$  lower; 91 s versus 184 s at 50 Hz,  $\sim 51\%$  lower) and essentially tied at  $F=100$  Hz (154 s versus 152 s). At  $\rho=0.9$ , the proposed method dominates across all  $F$  (e.g., 62 s versus 85 s at 100 Hz,  $\sim 27\%$  lower; 69 s versus 121 s at 50 Hz,  $\sim 43\%$  lower; 98 s versus 213 s at 20 Hz,  $\sim 54\%$  lower). Overall, the trends align with (15): higher  $\rho$  simplifies disambiguation (fewer moves), higher  $v$  compresses travel time, and varying  $F$  shifts the balance between observation cost and the number of moves.

In practice, selecting  $F$  should account jointly for  $v$  and  $\rho$ : slower platforms and low  $\rho$  benefit from smaller  $F$  (more informative samples, fewer moves), whereas fast platforms and high  $\rho$  benefit from larger  $F$  (minimal observation cost). Across these conditions, the proposed policy typically shortens total

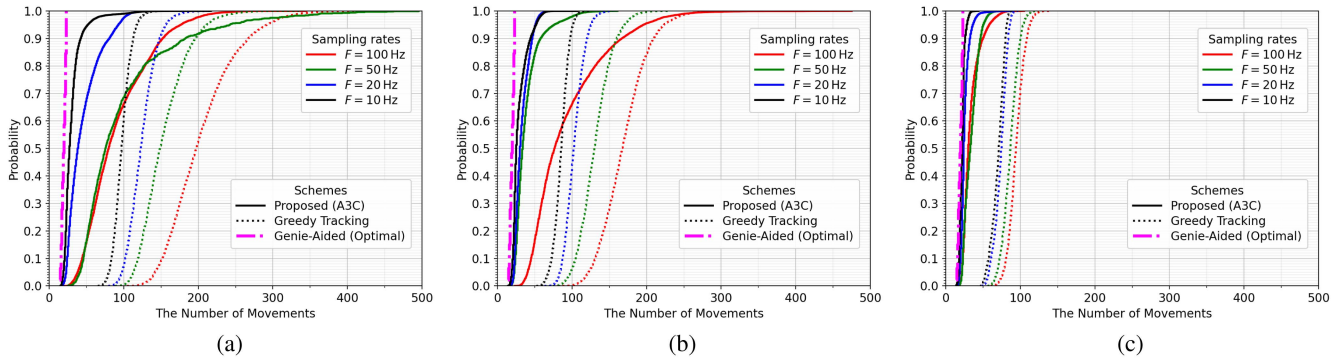


Fig. 6. CDF of the number of movements required to successfully track the target UAV for spatial correlation coefficients  $\rho \in \{0.1, 0.5, 0.9\}$ . (a)  $\rho = 0.1$ . (b)  $\rho = 0.5$ . (c)  $\rho = 0.9$ .

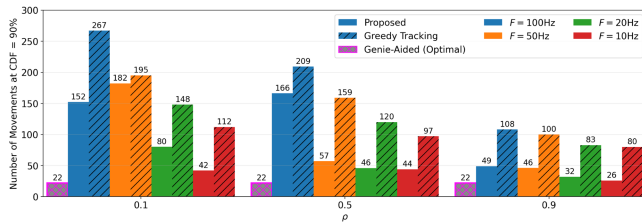


Fig. 7. Number of movements at the 90th percentile.

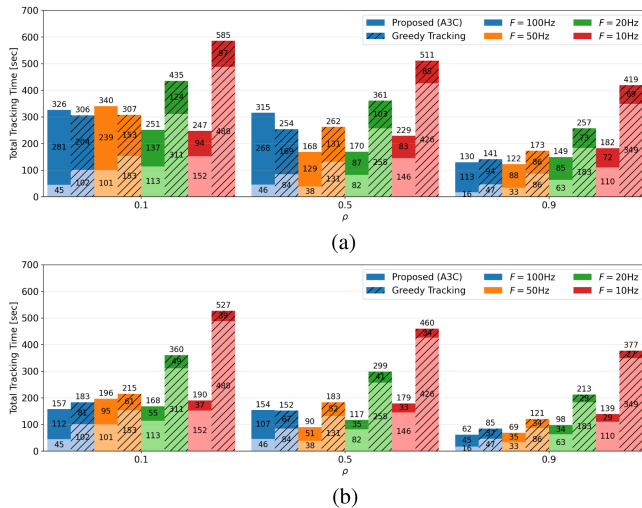


Fig. 8. Average end-to-end tracking time, split into channel observation (light) and travel (dark). (a)  $v = 2$  m/s. (b)  $v = 5$  m/s.

tracking time relative to greedy, while the unattainable genie oracle indicates the remaining headroom for any RSSI-driven controller.

**Complexity and real-time feasibility:** The controller operates on compact numeric RSSI sequences (six features per time step), not images or video, which keeps memory and compute budgets small. The policy has 367 722 trainable parameters (model state  $\approx 1.40$  MB with 32-bit floats). The forward cost is dominated by the stacked LSTMs. For an LSTM with input size  $I$  and hidden size  $H$ , the MAC cost per time step and layer is  $4(IH + H^2)$  (four gates). This standard count follows common LSTM formulations in [30, ch. 10].

TABLE II  
END-TO-END INFERENCE LATENCY BY DEVICE AND THE CORRESPONDING THROUGHPUT (INFERENCES PER SECOND)

Category	Device	Latency [ms]	Inferences/s
GPU	NVIDIA RTX 4090	2.9	344.8
	NVIDIA RTX 3090	4.7	212.8
CPU	Intel Core i9-13900 (64 GB RAM)	4.1	243.9
	AMD Ryzen 9 5950X (64 GB RAM)	5.5	181.8

Using  $L=50$  steps,  $H=128$ , and three layers with  $(I_1, I_2, I_3) = (6, 128, 128)$  yields

$$\text{Layer 1: } 4(6 \cdot 128 + 128^2) = 68,608$$

$$\text{Layer 2: } 4(128^2 + 128^2) = 131,072$$

$$\text{Layer 3: } 4(128^2 + 128^2) = 131,072$$

so one decision requires  $50 \times (68,608 + 131,072 + 131,072) = 16,537,600$  MAC ( $\approx 16.5$  M). Two small fully connected heads (policy/value) add  $\ll 1$  M MAC, giving a total of  $\approx 17$  M MAC per decision. Using the common convention  $1 \text{ MAC} \approx 2 \text{ FLOPs}$ , this corresponds to  $\approx 34$  M FLOPs per decision.

These analytic counts are consistent with the measured device-level latency in Table II: 2.9–5.5 ms per decision across commodity GPUs and modern desktop CPUs, i.e., at least  $\sim 180$  inferences/s. In the overall control loop, the dominant contributors to elapsed time are radio-side acquisition of RSSI samples and the physical motion and actuation of the UAVs—both hardware- and platform-dependent rather than algorithmic. In short, the proposed controller is computationally lightweight and real-time capable on CPUs and GPUs; policy inference is not the bottleneck compared with sensing and vehicle dynamics.

Although our experiments target a single unauthorized UAV, the architecture extends directly to multiple targets by instantiating parallel six-UAV swarm cells, each running the same A3C controller on its own RSSI stream. This requires only system-level orchestration (target-cell assignment, inter-cell de-confliction, and basic RF separation) rather than changes to the learning algorithm.

To facilitate reproducibility, the source code is publicly available at [31].

## V. CONCLUSION

This article proposed an innovative anti-UAV system designed to track and chase unauthorized UAVs using a sophisticated multidimensional swarm flight strategy, supported by the latest advancements in asynchronous DRL techniques. By enabling tracking UAV swarms to closely chase unauthorized target UAVs, this system significantly enhances the effectiveness of neutralizing such UAVs while concurrently minimizing the risk of collateral damage.

The core innovation of our system lies in the seamless integration of tracking and chasing functionalities by capturing communication signals between unauthorized UAVs and their controllers, traditionally handled as separate operations. A swarm flight strategy was also introduced to address the large-scale similarity issues of signals received by multiple antennas mounted on a single tracking UAV.

We have incorporated advanced channel modeling that accounts for spatial correlation and Doppler shifts, significantly enhancing the proposed system's responsiveness and adaptability to variable environmental dynamics. This modeling is crucial as it ensures the system maintains optimal performance under various operational conditions, as demonstrated by extensive simulations involving different channel sampling frequencies and spatial correlation coefficients.

Empirical results have consistently shown that the proposed system can achieve successful tracking and chasing within a defined criterion, irrespective of the sampling frequency and spatial correlation coefficients. This confirms the system's robust adaptability and reliability, establishing it as a valuable tool for both surveillance and active monitoring applications. The choice of channel sampling frequency ( $F$ ) is crucial and should be carefully adjusted based on the tracking UAVs' velocity ( $v$ ) and spatial correlation coefficient ( $\rho$ ) to optimize the tracking efficiency of our system.

While our approach primarily relies on RSSIs, it can be extended to handle autonomous UAVs that do not emit RF signals by incorporating multisensor fusion techniques without altering the fundamental system structure. This enhancement will further enhance the system's adaptability and expand its operational scope for broader UAV interception scenarios.

## REFERENCES

- [1] *IEEE Standard for Drone Applications Framework*, IEEE Standard 1936.1-2021, pp. 1–28, 2021.
- [2] O. A. Martinez and M. Cardona, "State of the art and future trends on unmanned aerial vehicle," in *Proc. Int. Conf. Res. Intell. Comput. Eng.*, 2018, pp. 1–6.
- [3] P. Čisar, R. Pinter, S. M. Čisar, and M. Gligorijević, "Principles of anti-drone defense," in *Proc. 11th IEEE Int. Conf. Cogn. Infocommun.*, 2020, pp. 19–26.
- [4] UAV Market, Sep. 2022. [Online]. Available: <https://www.marketsandmarkets.com/Market-Reports/unmanned-aerial-vehicles-uav-market-662.html>
- [5] C. Xu, X. Liao, J. Tan, H. Ye, and H. Lu, "Recent research progress of unmanned aerial vehicle regulation policies and technologies in urban low altitude," *IEEE Access*, vol. 8, pp. 74175–74194, 2020.
- [6] P. Tedeschi, F. A. Al Nuaimi, A. I. Awad, and E. Natalizio, "Privacy-aware remote identification for unmanned aerial vehicles: Current solutions, potential threats, and future directions," *IEEE Trans. Ind. Inform.*, vol. 20, no. 2, pp. 1069–1080, Feb. 2024.
- [7] Federal Aviation Administration, *Remote Identification of Drones*, Federal Aviation Administration (FAA), 2023. [Online]. Available: [https://www.faa.gov/uas/getting\\_started/remote\\_id](https://www.faa.gov/uas/getting_started/remote_id)
- [8] J. Sadovskis and A. Aboltins, "Modern methods for UAV detection, classification, and tracking," in *Proc. IEEE 63th Int. Sci. Conf. Power Elect. Eng. Riga Tech. Univ.*, 2022, pp. 1–7.
- [9] Y. Zhang, C. Wu, T. Zhang, Y. Liu, and Y. Zheng, "Self-attention guidance and multiscale feature fusion-based UAV image object detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
- [10] J. You, Z. Ye, J. Gu, and J. Pu, "UAV-Pose: A dual capture network algorithm for low altitude UAV attitude detection and tracking," *IEEE Access*, vol. 11, pp. 129144–129155, 2023.
- [11] H. Lee et al., "CNN-Based UAV detection and classification using sensor fusion," *IEEE Access*, vol. 11, pp. 68791–68808, 2023.
- [12] M. He, X. Fang, D. Huang, and Z. Zhang, "A hybrid integration method for low-observable micro-UAV trajectory tracking by 2D MIMO radar," in *Proc. 38th Youth Academic Annu. Conf. Chin. Assoc. Automat.*, 2023, pp. 809–813.
- [13] Q. An, C. Yeh, Y. Lu, Y. He, and J. Yang, "Time-varying angle estimation of multiple unresolved extended targets via monopulse radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 4, pp. 4650–4667, Aug. 2024.
- [14] W.-C. Jin, K. Kim, and J.-W. Choi, "Adaptive beam control considering location inaccuracy for anti-UAV systems," *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 2320–2331, Feb. 2024.
- [15] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1135–1143.
- [16] M. Danelljan, A. Robinson, F. Shahbaz Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *Proc. 14th Eur. Conf. Comput. Vis.*, Oct. 11–14, 2016, pp. 472–488.
- [17] Y. Li, H. Zhang, Y. Yang, H. Liu, and D. Yuan, "RISTrack: Learning response interference suppression correlation filters for UAV tracking," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
- [18] J. Li, D. H. Ye, M. Kolsch, J. P. Wachs, and C. A. Bouman, "Fast and robust UAV to UAV detection and tracking from video," *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 3, pp. 1519–1531, Jul./Aug. 2022.
- [19] J. Li et al., "Secure and efficient UAV tracking in space-air-ground integrated network," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10682–10695, Aug. 2023.
- [20] D. Shishika and V. Kumar, "A review of multi agent perimeter defense games," in *Decision and Game Theory for Security*, Q. Zhu, J. S. Baras, R. Poovendran, and J. Chen, Eds. Cham, Switzerland: Springer, 2020, pp. 472–485.
- [21] D. Shishika, J. Paulos, and V. Kumar, "Cooperative team strategies for multi-player perimeter-defense games," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 2738–2745, Apr. 2020.
- [22] S. L. Smith, S. D. Bopardikar, and F. Bullo, "A dynamic boundary guarding problem with translating targets," in *Proc. IEEE 48th Conf. Decis. Control, 28th Chin. Control Conf.*, 2009, pp. 8543–8548.
- [23] A. Pierson, Z. Wang, and M. Schwager, "Intercepting rogue robots: An algorithm for capturing multiple evaders with multiple pursuers," *IEEE Robot. Automat. Lett.*, vol. 2, no. 2, pp. 530–537, Apr. 2017.
- [24] 3rd Generation Partnership Project (3GPP), "Evolved universal terrestrial radio access (E-UTRA); user equipment (UE) radio transmission and reception," 3rd Generation Partnership Project (3GPP), Tech. Specification 36.101 V17.13.0, 2024.
- [25] T. Brown, P. Kyritsi, and E. De Carvalho, *Practical Guide to MIMO Radio Channel: With MATLAB Examples*. New York, NY, USA: Wiley, 2012.
- [26] MathWorks, "Propagation channel models," May 2024. [Online]. Available: <https://kr.mathworks.com/help/iteug/propagation-channel-models.html>
- [27] T. S. Rappaport, *Wireless Communications: Principles and Practices*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.
- [28] MathWorks, "Filter input signal through MIMO multipath fading channel," May 2024. [Online]. Available: <https://kr.mathworks.com/help/comm/ref/comm.mimochannel-system-object.html#d126e189388>
- [29] V. Mnih et al., "Asynchronous methods for deep reinforcement learning," in *Proc. 33rd Int. Conf. Mach. Learn.*, Jun. 20–22, 2016, pp. 1928–1937.
- [30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

- [31] T.-W. Ban, "UAV chasing system using asynchronous deep reinforcement learning," 2025. Accessed: Feb. 11, 2025. [Online]. Available: [https://github.com/twban/uav\\_chasing\\_a3c](https://github.com/twban/uav_chasing_a3c)



**Tae-Won Ban** (Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from the Department of Electronic Engineering, Kyungpook National University, Daegu, South Korea, in 1998 and 2000, respectively, and the Ph.D. degree in electrical and computer engineering from the Department of Electrical and Electronic Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2010.

He was a Researcher and a Network Engineer with Korea Telecom (KT), from 2000 to 2012. With KT, he researched 3G WCDMA, LTE, and Femto systems. He was also responsible for traffic engineering and spectrum strategy. He is currently a Professor with the Department of Artificial Intelligence and Information Engineering, Gyeongsang National University, Jinju, South Korea. His current research interests include radio resource management for mobile communication systems and deep reinforcement learning.



**Kyu-Min Kang** received the B.S., M.S., and Ph.D. degrees in electronic and electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, South Korea, in 1997, 1999, and 2003, respectively.

Since 2003, he has been with Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea, as a Principal Researcher. His current research interests include counter uncrewed aerial vehicle technology, spectrum sharing, cognitive radio networks, digital signal processing, and high-speed digital transmission systems.



**Bang Chul Jung** (Senior Member, IEEE) received the B.S. degree in electronics engineering from Aju University, Suwon, South Korea, in 2002, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Korea Advanced Institute for Science and Technology (KAIST), Daejeon, South Korea, in 2004 and 2008, respectively.

He was a Senior Researcher/Research Professor with the KAIST Institute for Information Technology Convergence, Daejeon, from 2009 to 2010. He is currently a Professor with the Department of Electrical and Computer Engineering, Aju University. His research interests include wireless communication systems, the Internet of Things (IoT) communications, statistical signal processing, information theory, interference management, radio resource management, spectrum-sharing techniques, and machine learning.

Dr. Jung was the recipient of the Fifth IEEE Communication Society Asia-Pacific Outstanding Young Researcher Award, in 2011; the Bronze Prize in Intel Student Paper Contest, in 2005; the First Prize in the KAIST's Invention Idea Contest, in 2008; and the Bronze Prize in Samsung Humantech Paper Contest, in 2009. He was selected as a Winner of the Haedong Young Scholar Award, in 2015, sponsored by the Haedong Foundation and given by the Korean Institute of Communications and Information Sciences (KICS). He has been selected as a Winner of the 29th Science and Technology Best Paper Award, in 2019, sponsored by the Korean Federation of Science and Technology Societies. He served as an Associate Editor for *IEEE Vehicular Technology Magazine*, from 2020 to 2022, and is also a Senior Editor for *IEEE Vehicular Technology Magazine*.