Learning-Based Strategies for UAV Swarm Target Defense: An Imitation and Reinforcement Learning Perspective Inspired by the Hawk–Pigeon Game

Trebert Hugo, Morin Eliot, Iakovlev Mikhaïl
Collective Behaviour 2025/26 , First report

I. Introduction

Unmanned aerial vehicle (UAV) swarms have recently gained prominence in surveillance, search-and-rescue, and defense applications. Coordinating multiple UAVs in dynamic or adversarial environments poses significant challenges in navigation, communication, and decision-making, particularly in **target defense**, where defending UAVs must protect a target from coordinated attackers.

The ongoing war between Ukraine and Russia has illustrated the crucial role of UAVs in modern warfare. Both sides have extensively used drones for reconnaissance, precision strikes, and swarm tactics, showing that unmanned systems are now **central assets in high-intensity conflicts** [1]. This real-world context highlights the urgent need for **robust and adaptive swarm defense strategies** capable of handling uncertainty and evolving threats. The Hawk–Pigeon Game provides a theoretical foundation for studying this interaction. In this model, **hawks** (defenders) attempt to intercept **pigeons** (attackers) before they reach a target, formulated as a **differential game** balancing interception and penetration objectives. Analytical solutions offer mathematically optimal pursuit and evasion strategies but rely on idealized assumptions—perfect information, deterministic dynamics, and few agents. As the number of UAVs grows, these models become intractable and fail to adapt to realistic uncertainty or adversarial variability.

This report proposes to extend the Hawk–Pigeon framework using **machine learning (ML)** to enable UAVs to learn cooperative defense and attack strategies from data or experience. Specifically, we combine **imitation learning**, which reproduces expert trajectories, and **reinforcement learning (RL)**, which refines them through interaction. This hybrid approach aims to bridge analytical control and adaptive learning, offering a scalable and data-driven solution to UAV swarm defense.

II. Review of Concepts and Existing Models

A. Analytical Control and the Hawk–Pigeon Game

In the original *Hawk–Pigeon Game Tactics for UAV Swarm Target Defense* paper [2], UAV dynamics are modeled by differential equations defining each agent's motion as a function of position, velocity, and control input :

$$\dot{x}_i = f_i(x_i, u_i) \tag{1}$$

where x_i represents the state vector (e.g., position, velocity) and u_i the control input applied by UAV i. The interaction between hawks and pigeons is formulated as a **differential game** in which both groups aim to optimize opposite objectives : defenders attempt to minimize interception time, while attackers try to reach the target as quickly as possible:

$$\min_{u_h} \max_{u_p} J(x, u_h, u_p) \tag{2}$$

Here, J denotes the cost function representing the conflicting goals of both sides. Solving this optimization problem yields **closed-form optimal control laws** for each group:

$$u^*_{\text{hawk}}(t) = g(x(t)), \quad u^*_{\text{pigeon}}(t) = h(x(t)) \tag{3}$$

These laws describe mathematically optimal actions under perfect information and idealized conditions. The advantages of analytical control include interpretability, theoretical guarantees, and computational efficiency. However, this framework faces several limitations:

- It assumes **perfect state knowledge** and deterministic dynamics.
- It scales poorly with the number of agents.
- It lacks robustness to noise, communication delays, and environment uncertainty.

Therefore, analytical control is valuable as a **baseline** but insufficient for real-world swarm defense applications.

B. Learning-Based Control

Recent advances in **machine learning** have introduced new ways to model **decision-making under uncertainty**. Two complementary approaches are particularly relevant: **Imitation Learning (IL)** and **Reinforcement Learning (RL)**.

Imitation Learning (IL): Imitation Learning enables an agent to learn control policies by **mimicking expert**

behavior rather than solving equations analytically. Given a dataset of **expert demonstrations** $\mathcal{D} = \{(x, u^*)\}$, where x represents the system state and u^* the expert control, a neural policy $f_\theta(x)$ is trained to approximate the expert's decisions by minimizing the **mean squared error**:

$$\mathcal{L}(\theta) = \mathbb{E}_{(x,u^*)\sim\mathcal{D}} \left[\| f_\theta(x) - u^* \|^2 \right] \tag{4}$$

This approach is **efficient** and **stable**, as it directly leverages existing analytical or simulated data. However, it inherits the expert's limitations and cannot surpass its performance unless combined with additional learning.

Reinforcement Learning (RL): In contrast, Reinforcement Learning allows an agent to learn through experience by **interacting with its environment** and receiving **rewards**. The objective is to find a policy π_θ that maximizes the **expected cumulative reward**:

$$\pi_\theta^* = \arg\max_{\pi_\theta} \mathbb{E}_{\tau\sim\pi_\theta} \left[\sum_{t=0}^{T} \gamma^t r_t \right] \tag{5}$$

where τ denotes a trajectory of states and actions, r_t is the reward at time step t, and γ is a **discount factor**. RL offers **adaptability** and can discover **emergent cooperative behaviors**, but training can be **unstable** and **computationally expensive**.

C. Planned Methodology

Our current work builds on the analytical foundations of the **Hawk–Pigeon Game** while introducing **learning-based methods** to overcome its limitations.

1) Environment Setup: We developed a **2D simulation environment** inspired by the Hawk–Pigeon framework. Each UAV follows the dynamic equation $\dot{x}_i = f_i(x_i, u_i)$, integrated with a **fourth-order Runge–Kutta (RK4)** scheme. The **converter module** plays a key role in translating the outputs of the learning model, typically accelerations or heading changes, into concrete state updates such as position and velocity. This conversion step allows us to directly use the network's **continuous control predictions** within the physical simulation. The **dynamics module** then applies these converted actions to propagate each UAV's motion and detect potential collisions or boundary violations. Visualization tools are implemented to display trajectories of hawks, pigeons, and the target, allowing qualitative analysis of **pursuit–evasion behaviors**. Finally, visualization tools are implemented to display the trajectories of hawks, pigeons, and the target, enabling qualitative analysis of **pursuit–evasion interactions**.

This **modular setup** establishes the foundation for training, testing, and comparing both analytical and learning-based control policies.

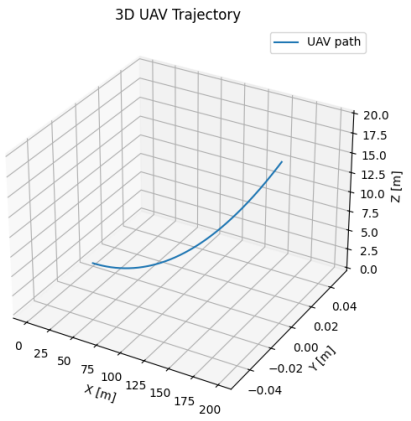

Fig. 1. Example trajectory using basic equations in the 3D simulation environment.

2) Imitation Learning Phase: **Imitation Learning** enables an agent to learn control policies by mimicking expert behavior rather than solving equations analytically [3]. In the first stage, we plan to collect **expert data** using the analytical Hawk–Pigeon control laws. For each simulation, we will record **state vectors** (positions, velocities, distances) and corresponding **optimal control actions** u^*. The policy network $f_\theta(x)$ will be trained by minimizing the **loss function** $\mathcal{L}(\theta)$ defined above. To improve generalization, we will apply **domain randomization** (varying number of agents, noise, and initial conditions) and **curriculum learning** (progressively increasing difficulty).

3) Reinforcement Learning Phase: **Reinforcement Learning (RL)** allows an agent to learn through interaction with the environment and optimize its **expected cumulative reward** [4]. Once the imitation model converges, we will fine-tune it through **multi-agent reinforcement learning**. Defenders (hawks) will receive positive rewards for successful interceptions and negative rewards if the target is reached. Attackers (pigeons) will have inverse objectives. The RL algorithm will optimize the expected return as defined by π_θ^*, allowing the agents to adapt to unseen or stochastic conditions. The **imitation-trained policy** will serve as an initialization, stabilizing RL training and enabling **faster convergence**.

4) Evaluation Plan: Performance will be evaluated through: Success rate: percentage of simulations where the target remains safe.

$$R_{\text{success}} = \frac{N_{\text{success}}}{N_{\text{total}}} \tag{6}$$

Interception time: average duration to neutralize attackers.

$$T_{\text{avg}} = \frac{1}{N_{\text{success}}} \sum_{i=1}^{N_{\text{success}}} T_i \tag{7}$$

Energy efficiency: cumulative control effort.

$$E_{\text{mean}} = \frac{1}{N} \sum_{i=1}^{N} \int_0^T \|u_i(t)\|^2 dt \tag{8}$$

Generalization: ability to perform under unseen configurations.

We will compare learned policies to classical analytical controls and simple heuristics like pure pursuit or proportional navigation.

III. Discussion and Expected Outcomes

Although no experimental results are presented yet, this research aims to demonstrate the benefits of combining analytical and learning-based methods for UAV swarm defense. Analytical control offers interpretability and mathematical guarantees, but fails to scale or adapt.

Imitation learning can replicate these optimal behaviors efficiently, serving as a fast and generalizable policy approximation.

Reinforcement learning can further refine these behaviors, enabling adaptation to dynamic and uncertain environments.

We expect the hybrid approach to yield: Faster computation of control actions compared to solving the game analytically in real-time.

Better scalability to multiple agents and noisy environments.

Emergent cooperation among defenders through experience-based learning.

Future work will include extending the simulation to 3D, testing more advanced network architectures (e.g., graph neural networks for multi-agent interaction), and exploring sim-to-real transfer.

References

[1] Hudson Institute, "The impact of drones on the battlefield: Lessons of the russia–ukraine war from a french perspective," November 2025.

[2] X. Guo, Y. Wang, and Z. Li, "Hawk-pigeon game tactics for uav swarm target defense," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 4, pp. 3312–3324, 2022.

[3] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, pp. 627–635, 2011.

[4] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.