Multi-Agent Adversarial Pursuit-Evasion using Deep Reinforcement Learning for UAV Interception

Soham Karandikar, Mechanical Engineering
Mentor: Dr. Kunal Garg, Assistant Professor
School for Engineering of Matter, Transport and Energy

Objective and Research Question

Autonomous Interception is crucial for drone defense and robotic security systems. This project trains an intelligent "pursuer" agent to intercept a moving "intruder" using Proximal Policy Optimization (PPO), a Deep Reinforcement Learning algorithm. Through careful reward function design, the agent learns to balance speed and precision for successful capture.

Problem Formulation

Pursuer must intercept intruder before target is breached.

Environment Specifications:

A 2D environment of 300 x 300 units shown in Fig 1.
Target: A 30-unit radius "safe zone" at the origin.
Interception Zone: A 100-unit radius area around target.
Pursuer: Starts randomly inside the Interception Zone.
Intruder: Starts randomly outside the Interception Zone and controlled by a simple proportional controller.

State and Observation Space:

- Position and Velocity of the Pursuer.
- Relative Position and Velocity of the Intruder.

Action Space:

- Acceleration of the Pursuer. (a_x, a_y)

Constraints:

- Max velocity: 50 units/s
- Capture radius: 10 units

Methodology

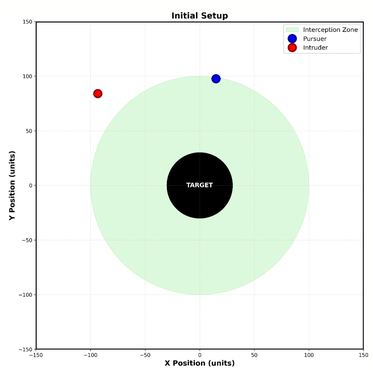
Reinforcement Learning Framework: PPO with 3-layer Actor-Critic architecture implementing LayerNorm.

Training Configuration:

Episodes: 5000 training episodes.
Timestep: dt = 0.01 s, max 2000 steps/episode.
Hyperparameters: $LR = 10^{-3}$, Entropy Coeff = 0.01

Rewards Setup

Distance-based rewards: Penalty if distance between pursuer and intruder increases.
Velocity alignment rewards: Reward if pursuer's velocity is aligned towards the intruder.
Relative speed rewards: Penalty for excessive relative velocity when close to the intruder.
Terminal rewards: High reward (+1000) for capture, high penalty (-1000) for intruder reaching the target.



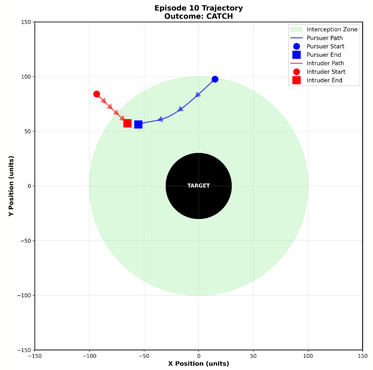


Fig 1. Interception Environment (Initial). **Fig 2.** Interception Trajectory.

Results and Conclusions

Training Success: The pursuer agent successfully learns an effective policy to intercept the intruder. (Fig. 2)
Learning: Training plots show initial improvement in the first 1000 episodes, followed by a plateau with no further gains. (Fig. 3)
Key Findings: Agent performance is critically sensitive to the design of the reward function. Different reward strategies produce vastly different agent behaviors, and a 'dense' reward, providing continuous feedback.

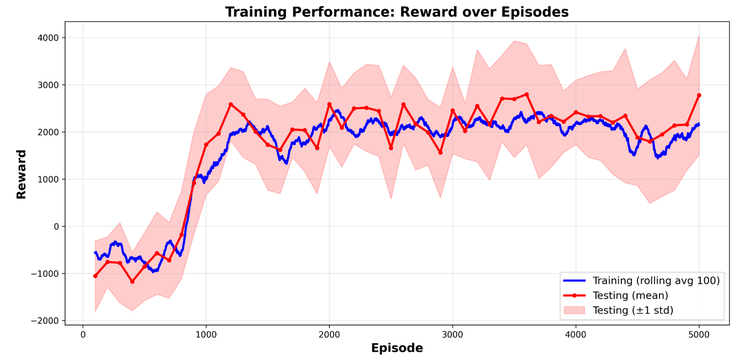


Fig 3. Pursuer Rewards over episodes.

Future Work

Multi-Agent RL: Upgrade intruder to RL agent; train both agents together.
PPO Comparison: Benchmark PPO vs MARL algorithms in adversarial tasks.

MORE Masters Opportunity for Research in Engineering

Acknowledgements

I thank Dr. Kunal Garg and Mr. Anandsingh Chauhan for their mentorship and guidance. This research was conducted at the School for Engineering of Matter, Transport and Energy, ASU

ASU Ira A. Fulton Schools of Engineering
Arizona State University