ICRAMCS 2026

THE EIGHTH EDITION OF THE INTERNATIONAL CONFERENCE ON RESEARCH IN APPLIED MATHEMATICS AND COMPUTER SCIENCE

April 23-24-25, 2026 | Marrakech, Morocco

Impact of Training Architecture on SAC and TD3 Performance in Multi-Agent Swarm Systems

Communication Info

Authors:

Samuel OYEDOTUN[1]

Joshua Ojo NEHINBE[2]

John Temitope OGBITI[3]

[1]Spring University, Benin, Nigeria

[2]Edo State University, Iyamho, Nigeria

[3]Edo State University, Iyamho, Nigeria

Keywords:

((1) Agent

(2) Swarm

(3) Reinforcement Learning

(4) Twin Delayed Deep Deterministic Policy Gradient (TD3)

(5) Soft Actor-Critic (SAC)

Abstract

Scalability and stability remain central challenges in Deep Reinforcement Learning (DRL) for swarm robotics particularly as agent population and state-action dimensionality increase [1, 2]. This study investigates how training architecture and algorithmic choice influence performance, robustness and convergence behavior in a centralized multi-agent swarm control environment. Using Python programming language, we compare Soft Actor-Critic (SAC) [3] and Twin Delayed Deep Deterministic Policy Gradient (TD3) [4] under structured hyperparameter grids, deterministic seeding, vectorized execution and progressive environment scaling (3-8 agents with increasing workspace size and stochastic perturbations). Evaluation shows that both algorithms achieve comparable mean rewards (e.g.−5.04 average return in representative runs) but exhibit distinct learning dynamics across controlled experiments. SAC proves stronger asymptotic robustness and adaptability under stochastic scaling and maintains competitive efficiency (mean normalized return -4.83). However, it shows higher reward variance (about 0.97) due to entropy-driven exploration. TD3 yields more consistent trajectories and lower variability across seeds but is more sensitive to hyperparameter configuration and increasing swarm dimensionality. Convergence analysis indicates that performance plateaus are gradual rather than sharply stabilized. These highlight the importance of training-time allocation (about 1450s per configuration) and architectural design. Statistical evaluation using bootstrap confidence intervals, effect size analysis and ANOVA confirms that performance differences are systematic, though moderate due to trade-off between exploration-driven robustness (SAC) and deterministic stability (TD3). These findings create a reproducible benchmark for centralized swarm DRL and provide principled guidance for selecting training architectures and continuous-control algorithms in scalable and uncertainty-prone robotic deployments [5].

References

[1] Orr, J., Dutta, A., Multi-Agent Deep Reinforcement Learning for Multi-Robot Applications: A Survey, *Sensors*, 23, 2023, 3625.

[2] Zhu, Y., Wan Hasan, W.Z., Harun Ramli, H.R., Norsahperi, N.M.H., Kassim, M.S.M., Yao, Y., Deep Reinforcement Learning of Mobile Robot Navigation in Dynamic Environment: A Review, *Sensors*, 25, 2025, 3394.

[3] Tang, C., Abbatematteo, B., Hu, J., Chandra, R., Martín-Martín, R., Stone, P., Deep Reinforcement Learning for Robotics: A Survey of Real-World Successes, *Annual Review of Control, Robotics, and Autonomous Systems*, 8, 2025, 153–188

[4] Aina, K.O., Ha, S., Deep Reinforcement Learning for Multi-Agent Coordination, *Artificial Life and Robotics*, 2025, (in press).

[5] Ali, M.A., Maqsood, A., Athar, U. & Khanzada, H.R. (2025). Comparative Evaluation of Reinforcement Learning Algorithms for Multi-Agent UAV Path Planning in 2D and 3D Environments. Drones, 9(6), 438. Offers a comparative study of DRL algorithms, including SAC, across simulated multi-agent UAV tasks