Adaptive PID Tuning for Quadcopters in Dynamic Environments Using Advantage Actor-Critic Reinforcement Learning

Kartik Maski
School of Computer Engineering and Technology
MIT World Peace University
Pune, India
kartikmaski29032004@gmail.com

Swamil Randive
School of Computer Engineering and Technology
MIT World Peace University
Pune, India
swamilrandive47@gmail.com

Kushagra Singh
School of Computer Engineering and Technology
MIT World Peace University
Pune, India
kushagraa.n@gmail.com

Udgeeth Deglurkar
School of Electrical and Electronics Engineering
MIT World Peace University
Pune, India
udgeethd@gmail.com

Anuja Barmecha
School of Computer Engineering and Technology
MIT World Peace University
Pune, India
anuja.barmecha@gmail.com

Asmita Verma
School of Computer Engineering and Technology
MIT World Peace University
Pune, India
asmitavermaa21@gmail.com

***Abstract*—**Achieving precise and adaptive control of quadcopter systems is essential for aerial robotics, autonomous navigation, and dynamic environment interaction. This study investigates the implementation of the Advantage Actor-Critic (A2C) algorithm, a prominent deep reinforcement learning (DRL) technique, for the online optimization of Proportional-Integral-Derivative (PID) controllers. The proposed framework autonomously adjusts PID gains through real-time interaction with a simulated environment, facilitating adaptive control under highly dynamic and uncertain conditions. Rigorous validation was conducted using simulated flight paths defined by discrete waypoints, designed to emulate complex operational scenarios. These scenarios incorporated varying payloads, external disturbances such as environmental noise, and significant mass uncertainties, illustrating the robustness of the approach. By utilizing a continuous state-action space, the A2C algorithm ensures effective adaptation and stability within the multi-input multi-output (MIMO) dynamics inherent to quadcopter control. The experimental results underscore the potential of DRL-enhanced control mechanisms for real-time, noise-resilient operation, offering a pathway toward increased autonomy, reliability, and efficiency in aerial systems.

***Keywords*—**Quadcopter control, Proportional Integral Derivative tuning, Advantage Actor-Critic (A2C), Deep reinforcement learning, MIMO systems, Adaptive control, Environmental noise tolerance, Dynamic payload management, Autonomous navigation.

I. Introduction

Quadcopters have emerged as a vital technology across various industrial domains due to their simple mechanical structure, low cost, vertical take-off and landing capabilities, and high maneuverability. These characteristics make them suitable for applications such as precision agriculture, search and rescue operations, infrastructure inspection, and aerial surveillance. However, the highly nonlinear dynamics of quadcopters, coupled with actuator interactions and susceptibility to external disturbances, present significant challenges in achieving stable and precise control [1]. As a result, the development of efficient attitude control systems for accurate path following and navigation remains a crucial research focus.

Proportional-Integral-Derivative (PID) controllers are extensively used in control systems due to their simplicity, efficiency, and ease of implementation. Despite their widespread adoption, PID controllers face inherent limitations when applied to complex, nonlinear systems [2]. Their performance is heavily influenced by accurate parameter tuning, which becomes increasingly challenging in the presence of external disturbances, parameter uncertainties, and dynamic operating conditions. Conventional offline PID tuning methods are often inadequate for systems requiring real-time adaptability and robustness, as they fail to dynamically adjust to changing environments and uncertainties [3].

To address these limitations, adaptive control strategies have been proposed, including Fuzzy Logic, Neural Networks (NNs), and Reinforcement Learning (RL) [4]. NNs, with their capability to approximate complex nonlinear functions,

and RL, which allows agents to learn optimal control strategies through interaction with their environment, have demonstrated significant potential in adaptive PID tuning [5]. RL-based methods are particularly advantageous as they provide dynamic optimization without relying on precise system modeling or human intervention, making them suitable for high-order nonlinear systems. Techniques such as Q-learning and Deep Deterministic Policy Gradient (DDPG) have been employed for PID tuning [6] [7]. However, these methods often require significant computational resources and can struggle with generalization in real-world conditions involving disturbances and uncertainties.

Recent advancements in RL, particularly Actor-Critic-based frameworks, have introduced robust solutions for adaptive PID control [8] [9]. These methods integrate NNs to approximate control policies and value functions, enabling continuous action outputs that surpass the limitations of traditional approaches. Among them, the Advantage Actor-Critic (A2C) algorithm has gained prominence for its synchronous updates, efficiency, and adaptability in dynamic environments. Unlike DDPG, which relies on offline training and is sensitive to disturbances, A2C facilitates real-time learning and optimization, making it particularly suitable for quadcopter control applications. The algorithm's ability to dynamically tune PID parameters enhances its robustness to variable payloads, environmental disturbances, and noise, ensuring precise trajectory tracking and improved stability [10].

This paper presents a novel A2C-based framework for online PID tuning in quadcopters, designed to overcome the challenges posed by nonlinear dynamics and uncertain environments. The proposed approach is validated through simulations on virtual paths defined by discrete waypoints, demonstrating its effectiveness under conditions of variable payloads, noise, and external disturbances. By eliminating the need for extensive computational resources and offline training, this method offers a practical and scalable solution for real-time quadcopter control.

II. Theoretical Framework

This research examines the control mechanisms of a quadcopter drone featuring four motors (M_1, M_2, M_3 and M_4) positioned symmetrically at the corners of a rigid rectangular frame. The thrust generated by these motors is crucial for the drone's stability and navigation. The drone operates in a six-degrees-of-freedom system, encompassing three translational movements (x, y, z) and three rotational movements: roll (ϕ), pitch (Θ), and yaw (ψ). In this system, z represents altitude, while ϕ, Θ, and ψ define the drone's orientation within an inertial reference frame. Fig. 1 illustrates the correlation between motor thrust and the drone's movement.

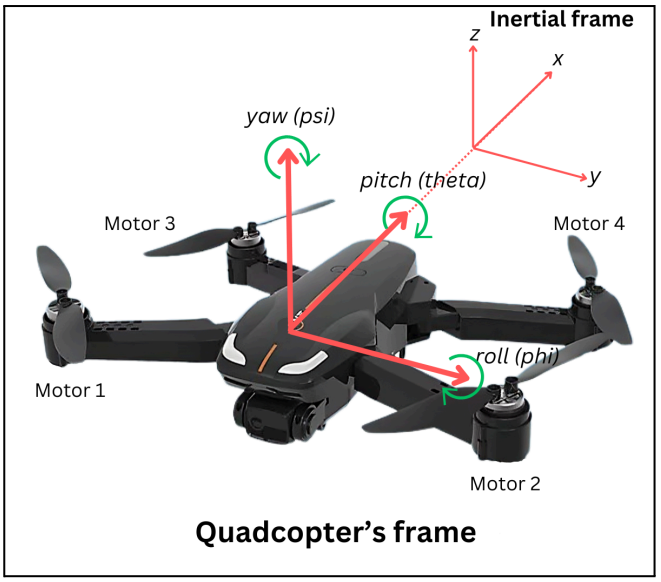

Fig. 1. Quadcopter dynamics with 4 motors

To achieve stable flight and trajectory tracking, a robust control mechanism is required to manage these degrees of freedom. The quadcopter's control system must ensure precise adjustment of the thrust generated by the four motors to counteract disturbances, follow the desired trajectory, and maintain stability in the presence of uncertainties.

The control inputs of the quadcopter, denoted as u_1, u_2, u_3 and u_4, are critical for managing its altitude, roll, pitch, and yaw, respectively. Each of these control inputs directly corresponds to a specific dynamic aspect of the quadcopter's motion. Specifically, u_1 is responsible for the total thrust T, which affects the altitude or vertical position of the quadcopter. Similarly, u_2, u_3 and u_4 correspond to the torques τ_ϕ, τ_Θ, and τ_ψ, which control the rotational motions roll (ϕ), pitch (Θ), and yaw (ψ), respectively. This relationship can be expressed as:

$$u_1 = T, \quad u_2 = \tau_\phi, \quad u_3 = \tau_\Theta, \quad u_4 = \tau_\psi \tag{1}$$

To achieve precise control and stability, each control input u_i is designed as a combination of static and dynamic gain components [11]. The static gain component, $k_{i,\,static}$, is designed to ensure baseline stability under nominal operating conditions. This gain acts as a fixed contribution that provides a fundamental level of control, addressing the basic dynamic requirements of the system.

The dynamic gain component, $k_{i,\,dynamic}$, is adaptable and crucial for responding to changes in system dynamics or external disturbances. It accounts for variations such as environmental factors, load changes, or unmodeled sensor effects, allowing the quadcopter to maintain stability and accurately follow its intended path. Thus, the control input for each degree of freedom can be expressed as:

$$u_i = k_{i,\,static} + k_{i,\,dynamic} \quad (2)$$

The total gain k_i for each input u_i is determined as a combination of proportional, integral, and derivative (PID) terms to handle different dynamic scenarios [7]. This can be written as:

$$k_i = k_{i,P}\, e_i + k_{i,I} \int e_i\, dt + k_{i,D} \frac{de_i}{dt} \quad (3)$$

where e_i is the error between the desired and actual state for the i-th degree of freedom, $k_{i,P}$, $k_{i,I}$ and $k_{i,D}$ are the proportional, integral, and derivative gains, respectively.

This two-tiered control approach ensures robust and adaptable operation [11]. Static gains address basic stability and steady-state performance requirements, while dynamic gains actively compensate for disturbances and uncertainties, enabling precise path following and improved disturbance rejection. This method effectively manages the quadcopter's inherent nonlinear and interconnected dynamics, ensuring precise control over its six degrees of freedom.

By fine-tuning the dynamic gain components, this control strategy enhances the quadcopter's performance in path following, disturbance handling, and overall stability. This approach enables the quadcopter to achieve accurate control over its six degrees of freedom, ensuring reliable operation in various conditions.

III. Network Architecture

The proposed hybrid control framework integrates the Actor-Critic (A2C) reinforcement learning approach with traditional PID control to dynamically optimize the PID gains for quadcopter stabilization [11] [12]. The hybrid structure leverages the adaptability of reinforcement learning to address the limitations of static PID controllers in highly dynamic environments, achieving robust and responsive control.

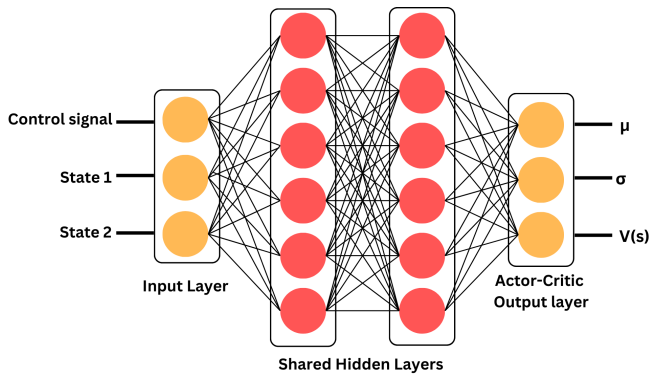

Fig. 2. Hybrid Actor-Critic Network Architecture for Policy and Value Estimation

The Actor-Critic network as specified in Fig. 2. comprises two interconnected components: the actor and the critic. The actor predicts the control action distribution, parameterized by its mean (μ) and standard deviation (σ), while the critic evaluates the value of the current state. This dual-output structure enables simultaneous optimization of the control policy and state-value estimation. The actor outputs μ and σ through separate fully connected layers, where σ is constrained to positive values using a bounded transformation, ensuring numerical stability. The critic outputs a scalar value representing the state-value function, which is instrumental in computing the temporal difference (TD) error used to guide policy updates.

To enable dynamic tuning of the PID gains (K_p, K_i, K_d), the framework introduces a neural-network-driven adaptation mechanism [11]. The PID gains are expressed as a combination of static components and adaptive adjustments influenced by the actor network. Mathematically, this is represented as:

$$PID\ Gains = K_{static} + K_{domain} \cdot tanh(Wx + b) \quad (4)$$

where $Wx + b$ denotes the linear transformation applied to the state vector x by the network. The $tanh$ activation function ensures bounded outputs, preventing excessive gain adjustments.

The learning process in the hybrid architecture is driven by a composite loss function, incorporating contributions from the actor, critic, and entropy regularization. The actor loss penalizes deviations between the predicted control signal and the desired output. This deviation is quantified by the error $\epsilon_\mu = \mu - y_d$, where y_d is the desired control signal. To ensure a well-calibrated action distribution, the model also considers the variance deviation $\epsilon_\sigma = 0.1 \cdot (\sigma - |\epsilon_\mu|)$. These terms combine in the actor loss, expressed as:

$$Loss_{actor} = w_1 \cdot (0.1 + |\delta|) \cdot \epsilon_\mu^2 + w_2 \cdot \epsilon_\sigma^2 \quad (5)$$

where δ is a TD error and w is the constant weighting factor.

The Critic loss minimizes the TD error, which represents the difference between the predicted and observed rewards in the environment [10]. It is given by:

$$\delta = r + \gamma V(s') - V(s) \quad (6)$$

where r is the immediate reward, $V(s)$ and $V(s')$ are the value functions of the current and next states, respectively, and γ is the discount factor. The critic loss is defined as:

$$Loss_{critic} = w_3 \cdot \delta^2 \quad (7)$$

To encourage exploration during training, the framework incorporates an entropy regularization term. This term

maximizes the uncertainty in the action distribution, preventing premature convergence to suboptimal policies. The entropy term is given by:

$$Entropy = w_4 \cdot \sqrt{2\Pi e\sigma^2} \tag{8}$$

The total loss function is the sum of the actor loss, critic loss, and entropy term, expressed as:

$$Loss_{Total} = Loss_{Actor} + Loss_{critic} + Entropy \tag{9}$$

During operation, the Actor-Critic network predicts the control action distribution based on the current state, dynamically adjusting the PID gains. The generated control signal is applied to the quadcopter, and the resulting state transitions are used to compute rewards, driving the optimization process. The total loss is used for backpropagation and parameter updates, enabling the network to continually improve its control policy. This hybrid approach effectively combines the interpretability and robustness of PID controllers with the adaptability of reinforcement learning, ensuring efficient and stable quadcopter control in dynamic environments.

IV. Experiments and Evaluation

To evaluate the system's performance, we analyzed the linear positions (X, Y), altitude, and angular orientations roll (ϕ), pitch (Θ), and yaw (ψ) along different paths. Additionally, we assessed the model's learning process by tracking rewards and loss over iterations under ideal conditions and also by introducing certain realistic gaussian noise disturbances.

To simulate real-world operational conditions, two key uncertainties were introduced in the system: variable payloads and Gaussian noise. These uncertainties emulate the challenges encountered due to dynamic load variations and sensor inaccuracies, respectively [13] [14]. Variable payloads introduce changes in system mass over time to simulate operational scenarios such as loading, unloading, or dynamic mass distribution. The system mass m is defined as a piecewise function, varying across distinct intervals within the simulation duration p. Mathematically, the mass m is represented as:

$$m = \begin{cases} m_0 & \text{if} 0 \le k < 0.2p, \\ 2m_0 & \text{if} 0.2p \le k < 0.4p, \\ 3m_0 & \text{if} 0.4p \le k < 0.6p, \\ 2m_0 & \text{if} 0.6p \le k < 0.8p, \\ m_0 & \text{if} 0.8p \le k \le p. \end{cases} \tag{10}$$

where m_0 is the base mass of the system, p is the total duration of the simulation, and k is the current simulation step. This function represents periodic mass changes, simulating scenarios such as payload drops or additions.

Gaussian noise was introduced to emulate sensor inaccuracies typically observed in real-world systems. Noise was added to both angular and linear sensor measurements to reflect the stochastic errors in devices such as gyroscopes, accelerometers, GPS, or altimeters. The noisy signals were computed as:

$$X_{noisy} = X_{ideal} + N(0, \sigma^2) \tag{11}$$

where X_{ideal} represents the true sensor value, $N(0, \sigma^2)$ is Gaussian noise with zero mean and variance σ^2, and σ denotes the standard deviation associated with the sensor type. These uncertainties were programmatically implemented, with payload variations applied at specific intervals of the simulation and noise continuously added to sensor measurements. This approach ensured a dynamic and realistic testing environment, facilitating the evaluation of system control algorithms under the combined effects of varying loads and sensor inaccuracies.

A. Results for a Specified Point in Space

For a specified stationary point in space, the system's tracking and stabilization capabilities were evaluated. The desired coordinates (X_d, Y_d) and the actual state (X, Y) were compared. Under ideal conditions, as shown in Fig. 3(a), the quadcopter smoothly converged to the target position. Altitude control and PID gain responses, depicted in Fig. 3(b), demonstrated effective stabilization.

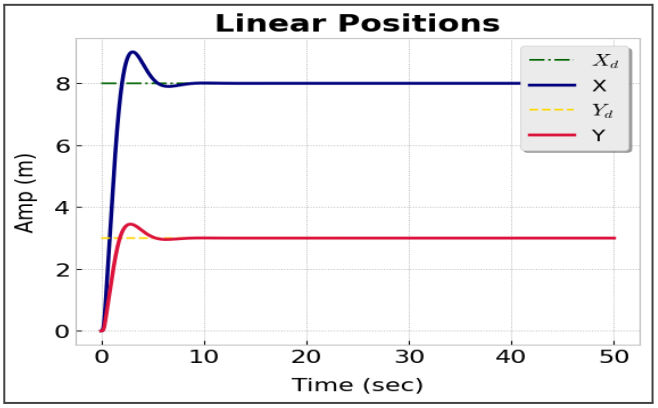

Fig. 3(a). Trajectory Tracking—Achieved X and Y Positions Under Ideal Conditions

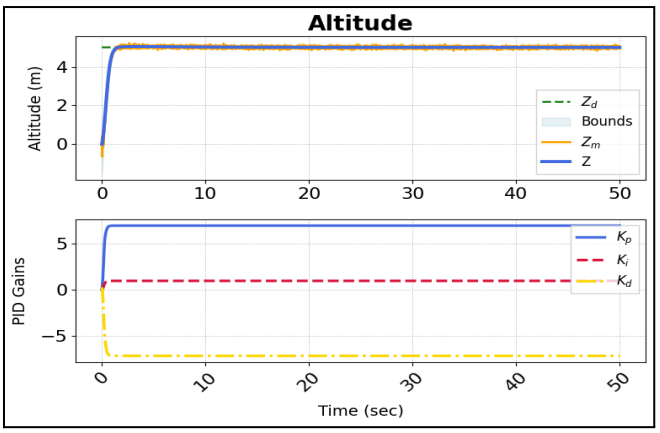

Fig. 3(b). Altitude Control and PID Gain Response for a Stationary Point

Fig. 3(c) further illustrate the system's performance under dynamic payload variations and combined payload and noise uncertainties to attain desired altitude (Z_d) as specified in Eq.(10) and Eq.(11), highlighting its robustness and adaptability in challenging conditions.

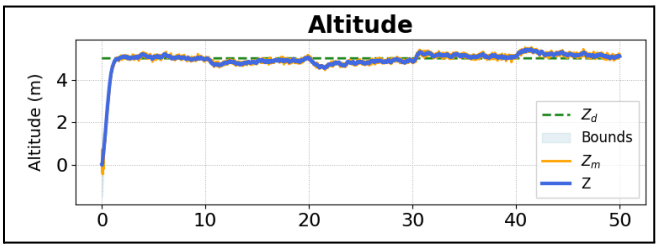

Fig. 3(c). Altitude Control Under Combined Payload Variations and Noise Uncertainties

This shows that altitude Z aligns precisely with the desired setpoint Z_d, as depicted in the altitude control figure, it highlights the effectiveness of the proposed method in achieving accurate and stable convergence of the quadcopter to its target altitude.

B. Results for a Circular Path

The circular trajectory path was evaluated to analyze the quadcopter's performance under variable payload conditions. Effective altitude control, as demonstrated in Fig. 4(a), was achieved using optimized PID gains, ensuring stability throughout the trajectory.

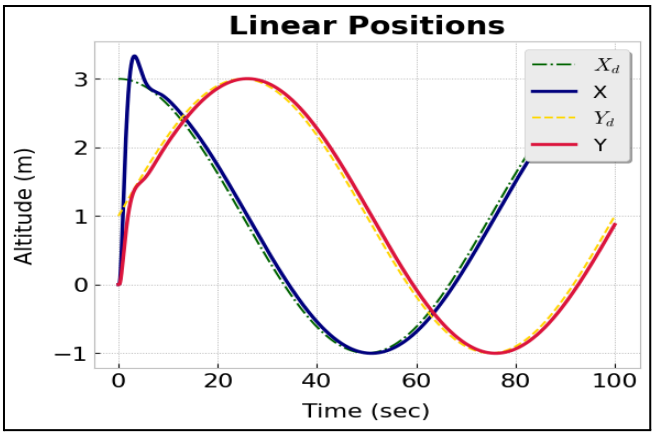

Fig. 4(a). Altitude control during circular trajectory under variable payload conditions

The roll and pitch angles, presented in Fig. 4(b) and Fig. 4(c), respectively, illustrate the system's ability to maintain orientation and stabilize the platform while following the circular path.

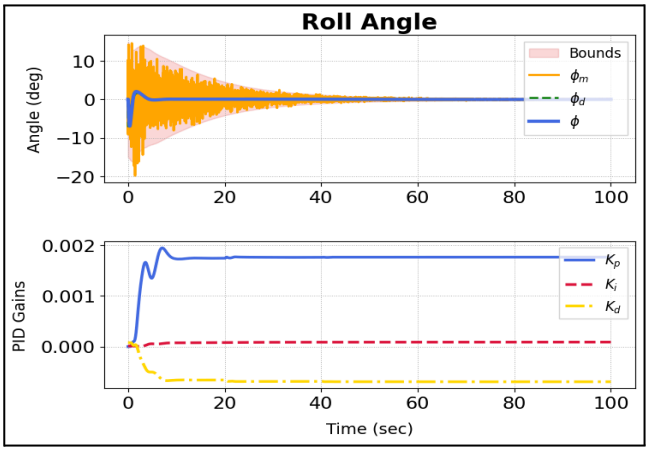

Fig. 4(b). Roll angle stabilization for circular trajectory with PID control

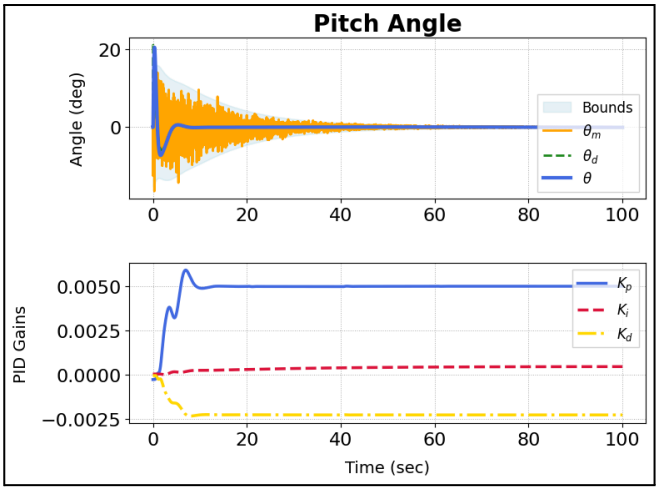

Fig. 4(c). Pitch angle stabilization for circular trajectory with PID control

C. Results for a Square Path

The square trajectory path was designed to evaluate the quadcopter's control precision and stability during dynamic maneuvers, including sharp turns and yaw adjustments, with dynamic payload variations as defined in Eq. (10). To further assess accuracy, the path was tested under Gaussian noise disturbances as outlined in Eq. (11). As depicted in Fig. 5(a), the system demonstrated robust control, effectively stabilizing sharp yaw angles despite the combined challenges of variable payloads and noise, leveraging well-tuned PID gains.

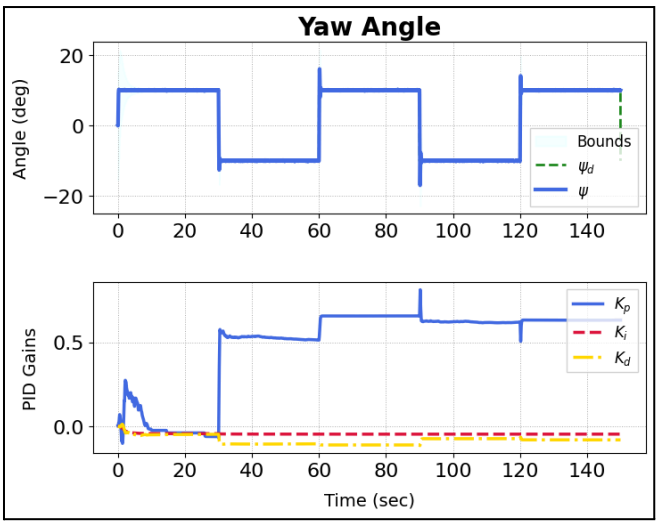

Fig. 5(a). Stability control under sharp yaw angle adjustments

The reward progression in Fig. 5(b), highlights the system's capacity to achieve optimal control for altitude, roll, pitch, and yaw. The consistent reward trends across these parameters confirm the efficacy of the adaptive strategy in minimizing trajectory errors and maintaining accuracy.

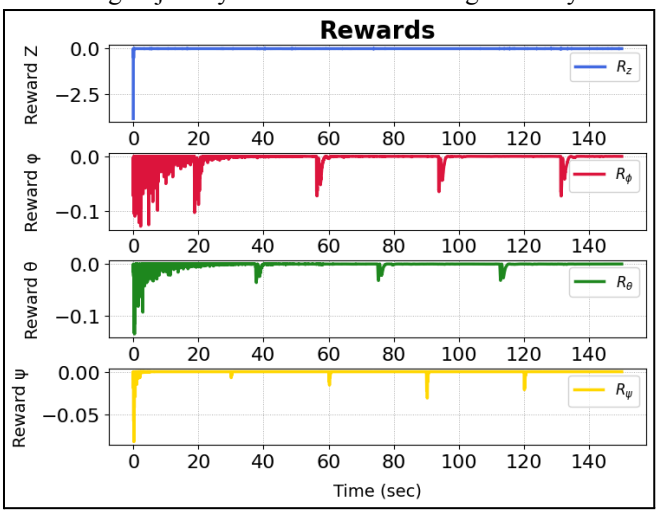

Fig. 5(b). Reward progression for altitude, roll, pitch, and yaw control

V. Conclusion

This investigation demonstrates the efficacy of integrating the Advantage Actor-Critic (A2C) deep reinforcement learning algorithm with PID controllers for precise and adaptive quadcopter control. Through autonomous optimization of PID gains in real time, the proposed framework successfully addressed dynamic operational scenarios characterized by variable payloads, sensor noise, and mass uncertainties. Simulated experiments on diverse flight paths, including stationary points, circular trajectories, and square trajectories, validated the system's robustness in achieving stability and precision under challenging conditions. The findings highlight the potential of DRL-enhanced control mechanisms to enable noise-resilient, adaptive, and efficient operation, thus advancing the field of autonomous aerial robotics.

VI. References

[1] D. Wang, Q. Pan, Y. Shi, J. Hu and C. Zhao, "Efficient Nonlinear Model Predictive Control for Quadrotor Trajectory Tracking: Algorithms and Experiment," in *IEEE Transactions on Cybernetics*, vol. 51, no. 10, pp. 5057-5068, Oct. 2021, doi: 10.1109/TCYB.2020.3043361. https://ieeexplore.ieee.org/document/9329094

[2] Ezhil, V. R. S., Sriram, B. S. R., Vijay, R. C., Yeshwant, S., Sabareesh, R., Dakkshesh, G., & Raffik, R. (2022). Investigation on PID controller usage on Unmanned Aerial Vehicle for stability control. *Materials Today Proceedings*, *66*, 1313–1318. https://doi.org/10.1016/j.matpr.2022.05.134

[3] Santoso, F., Garratt, M. A., & Anavatti, S. G. (2017). State-of-the-Art intelligent flight control systems in unmanned aerial vehicles. *IEEE Transactions on Automation Science and Engineering*, *15*(2), 613–627. https://doi.org/10.1109/tase.2017.2651109

[4] Sönmez, S., Rutherford, M., & Valavanis, K. (2024). A survey of Offline- and Online-Learning-Based Algorithms for Multirotor UAVs. *Drones*, *8*(4), 116. https://doi.org/10.3390/drones8040116

[5] Bari, S., Hamdani, S. S. Z., Khan, H. U., Rehman, M. U., & Khan, H. (2019). Artificial Neural Network based Self-Tuned PID Controller for flight control of quadcopter. *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, 1–5. https://doi.org/10.1109/ceet1.2019.8711864

[6] Koch, W., Mancuso, R., West, R., & Bestavros, A. (2019b). Reinforcement learning for UAV attitude control. *ACM Transactions on Cyber-Physical Systems*, *3*(2), 1–21. https://doi.org/10.1145/3301273

[7] Alrubyli, Y., & Bonarini, A. (2022). Using Q-Learning to automatically tune quadcopter PID controller online for fast altitude stabilization. *2022 IEEE International Conference on Mechatronics and Automation (ICMA)*, 514–519. https://doi.org/10.1109/icma54519.2022.9856292

[8] Wang, L., Wang, K., Pan, C., Xu, W., Aslam, N., & 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*, *7*(1), 73–84. https://doi.org/10.1109/tccn.2020.3027695

[9] Lin, J., Wang, L., Gao, F., Shen, S., & Zhang, F. (2019). Flying through a narrow gap using neural network: an end-to-end planning and control approach. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3526–3533. https://doi.org/10.1109/iros40897.2019.8967944

[10] R. Mukhopadhyay, S. Bandyopadhyay, A. Sutradhar and P. Chattopadhyay, "Performance Analysis of Deep Q Networks and Advantage Actor Critic Algorithms in Designing Reinforcement Learning-based Self-tuning PID Controllers," *2019 IEEE Bombay Section Signature Conference (IBSSC)*, Mumbai, India, 2019, pp. 1-6, doi:

10.1109/IBSSC47189.2019.8973068.
https://ieeexplore.ieee.org/abstract/document/8973068
[11] Sharifi and A. Alasty, "Self-Tuning PID Control via a Hybrid Actor-Critic-Based Neural Structure for Quadcopter Control," *30th Annual International Conference of Iranian Society of Mechanical Engineers (ISME2022)*, Tehran, Iran, May 2022. https://arxiv.org/pdf/2307.01312
[12] V. van Veldhuizen, "Autotuning PID control using Actor-Critic Deep Reinforcement Learning," Bachelor's thesis, University of Amsterdam, 2020. https://arxiv.org/pdf/2212.00013
[13] Lee, J., Xuan-Mung, N., Nguyen, N. P., & Hong, S. K. (2021). Adaptive altitude flight control of quadcopter under ground effect and time-varying load: theory and experiments. *Journal of Vibration and Control*, *29*(3–4), 571–581. https://doi.org/10.1177/10775463211050169
[14] Nascimento, R. G. D., Fricke, K., & Viana, F. (2020). Quadcopter Control Optimization through Machine Learning. *AIAA SCITECH 2022 Forum*. https://doi.org/10.2514/6.2020-1148