Journal of Applied Automation Technologies

Research Paper

Application of Quantum Reinforcement Learning in Autonomous Unmanned Aerial Vehicle Navigation

Veljko Perić[1,*] and Milan Živković[1]

[1] Faculty of Information Technology, Belgrade Metropolitan University, Belgrade, 11000, Serbia
*Corresponding author: veljko.pe@metropolitan.edu.rs

Abstract. An intelligent adaptation technique is necessary due to the complexity and unpredictability of the air space, and autonomous unmanned aerial vehicle (UAV) navigation uses quantum reinforcement learning (QRL) as an advanced control mechanism. This work suggests quantum-inspired encoding and optimization algorithms for policy learning in order to overcome the shortcomings in efficiency and scalability of conventional and deep reinforcement learning approaches. The technique employs quantum-parametric policy optimization on a hybrid quantum-classical architecture after forming UAV navigation as a Markov Decision Process and adding quantum amplitude encoding. Dynamic obstacle and random wind disturbance experiments will be carried out in a high-fidelity 3D simulation environment. QRL has decreased energy usage by up to 17% and increased the top classical RL baseline's convergence speed by 40%. Even in inclement weather, the success rate remains around 95%, and the minimum obstacle clearance never falls below 1.8 meters. Analysis shows that QRL yields safer, more robust and sample-efficient policies that are better at generalizing to unseen maps and disturbance-heavy environments. Based on the above results, the technical feasibility and engineering value of quantum-enhanced learning algorithms for robust UAV autonomous navigation have been demonstrated; thus, they can serve as a model for future intelligent aerial decision systems.

Keywords: *Quantum Reinforcement Learning, UAV Navigation, Autonomous Systems, Path Planning, Intelligent Robotics*

Received on 07 June 2025, Accepted on 02 December 2025, Published on 13 December 2025

Introduction

Unmanned aerial vehicles (UAVs) are revolutionizing the field of aerial robotics and are being used for a variety of tasks, including construction inspections, environmental research, observation and surveillance, and disaster relief [1]. High-autonomy and flexible Unmanned Aerial Vehicles (UAVs) that can function in challenging and unpredictable environments have been made possible by advancements in on-board sensing, control architecture, and embedded computing [2]. As a result, numerous studies have been carried out to enhance robots' perception and navigation capabilities in real-world settings [3]. Traditional rule-based and deterministic planning algorithms still have fundamental flaws for deployment in unstructured, dynamic, or adversarial environments for UAVs, despite successful accomplishments in path planning, sensor fusion, and trajectory optimization [4]. More adaptable learning-based methods are currently being used because classic algorithms are no longer sufficient for the sheer scale and non-linearity of aerial navigation challenges [5]. Next-generation methods that can generalize beyond pre-programmed behaviors are required due to the proliferation of high-dimensional sensor data, non-stationary settings, and the necessity for rapid adaptation [6].

A novel approach to creating intelligent autonomous vehicles is reinforcement learning (RL), which allows them to learn how to operate and drive by making mistakes in their surroundings [7]. Because RL algorithms learn to make judgments on their own through iteration rather than being explicitly coded or manually supervised, they have fared exceptionally well in many areas of robotics [8]. However, conventional reinforcement learning frequently faces the difficulties of high dimensionality, stochasticity, and resource constraints in complex

https://doi.org/10.64972/jaat.2025v3.240p44e:604-617

environments of high-end UAVs [9]. The issues include inefficient exploration, sluggish convergence in complicated environments, and excessively high computational resources needed for policy optimization in large and dynamic state-action spaces [10]. As a result, people are searching for alternative compute methods that can accomplish high performance on a broad scale [11]. As quantum computing has advanced recently, there has been interest in using quantum effects like superposition, entanglement, and quantum parallelism to speed up learning, increase the size of policy representation, and solve the scaling problems of conventional algorithms [12]. In order to accomplish more effective, expressive, and flexible learning for challenging-to-control robots, quantum reinforcement learning (QRL) seeks to harness these phenomena in an organized manner [13]. According to preliminary studies, QRL can be applied more successfully than conventional reinforcement learning to increase the quality of solutions, speed up exploration, and manage complex environmental uncertainties for rapid and stable adaption in some applications [14]. The aforementioned research has demonstrated that the need for autonomous UAVs and quantum computing may coexist.

Motivated by the aforementioned factors, this study explores the integration of quantum reinforcement learning for autonomous navigation of unmanned aerial vehicles. A comprehensive framework is put out to represent the UAV navigation problem as a Markov Decision Process and incorporate quantum-inspired mechanisms into the design of learning algorithms, policy optimization, and state encoding. This work suggests quantum features for both techniques and algorithms in light of the shortcomings in real-world reinforcement learning. To objectively evaluate the system's real functioning, stability, and generalization, run numerous simulations and comparative tests. The development of a QRL-based UAV navigation scheme, novel techniques for quantum-policy learning, and real-world experiments contrasting with traditional and learning-based navigation technologies are the primary topics of this study. This work paves the way for the development of a new generation of intelligent, high-efficiency, and large-scale autonomous UAV systems.

Related Work

Classical Approaches in UAV Navigation

Traditional path planning and control techniques have been used for the majority of autonomous UAV navigation for the past few decades [16]. For navigation in organized, static environments, the well-known A* and Dijkstra's algorithms are complete and optimal; but, as the size of the state space grows, they become computationally costly [17]. Numerous UAV mission profiles have now been applied to sampling-based techniques that may find pathways in high-dimensional or partially known settings, such as probabilistic roadmaps and rapidly-exploring random trees (RRT) [18]. B-spline curve-based methods are commonly employed for applications that need smooth, continuous trajectories that satisfy UAV kinematic restrictions [19]. However, the aforementioned classical planners are rigid in extremely dynamic, uncertain, or combative situations since they rely on precise environment maps and set cost functions [20]. In practice, many UAV systems have also embraced rule-based or behavior-based architectures that draw inspiration from biological heuristics or reactive control methods. These structures provide quick response times but typically lack global optimality and generalizability [21]. In order to fulfill both the need for real-time local adaptation and optimal long-term planning, research has been done on combining deliberative and reactive layers in hybrid designs [22]. Nevertheless, the outdated approach is comparatively rigid and does not satisfy the high standards needed to grow and function effectively in novel and challenging contexts [23]. Overall, the aforementioned drawbacks of these pipelines have prompted the creation of data-driven and flexible navigation techniques for robust, practical UAV autonomy [24]. Developing adaptable solutions that go beyond the purview of earlier study will be the next phase of this field's research [25].

Reinforcement Learning Methods in Autonomous Systems

A novel form of autonomous control and navigation for robots and UAVs has been developed using reinforcement learning (RL) [26]. RL develops a navigation policy by maximizing the total reward after directly interacting with the environment, as opposed to conventional methods [27]. A common example of a tabular approach that works well in simple, discrete-space issues is Q-learning; nevertheless, to address complicated or continuous settings, new solution models have been needed [28]. Through function approximation by deep neural networks, deep reinforcement learning has made it possible to expand the generalization capacities of

https://doi.org/10.64972/jaat.2025v3.240p44e:604-617

high-dimensional sensory input and unstructured settings [29]. Excellent methods including Proximal Policy Optimization (PPO), Deep Deterministic Policy Gradient (DDPG), and Deep Q-Networks (DQN) are currently being used to solve a variety of challenging control and path planning issues [30]. For fleets and swarms of UAVs, multi-agent reinforcement learning (MARL) can broaden the application of cooperative distributed navigation. But there are still real-world issues: Problems like sample inefficiency, complicated incentive design, and training instability have been reported with reinforcement learning techniques, which are often computationally and data-intensive. Due to model errors and unmodeled dynamics in the synthetic training environment, real-world deployment also faces a "sim-to-real" transfer issue. In order to lower training costs and improve the robustness of learnt policies, research is now being done on curriculum learning, safe reinforcement learning (RL), transfer learning, and other forms of adaptation. While fixed-rule systems have been significantly improved by reinforcement learning (RL), more robust architectures and data-efficient algorithms that can handle a variety of situations in the field deployment environment are currently being developed.

Quantum Computing and Quantum Reinforcement Learning

The characteristics of superposition, entanglement, and quantum interference for qubits can be used by quantum computers, which are novel types of computers. Because a qubit can be in a superposition of both 0 and 1, unlike a traditional bit, it can achieve tremendous parallelism and, for specific tasks, an exponential advantage over classical computers. Researchers have started investigating applications of quantum speedup in machine learning, search and optimization, robotics, and autonomous decision-making as a result of the development of gate-based and adiabatic quantum computers. By storing state and action information in quantum states and employing quantum algorithms for value propagation and policy optimization, quantum reinforcement learning (QRL) combines the agent-based learning framework of reinforcement learning with quantum computing. According to preliminary research, QRL can greatly outperform conventional techniques in terms of exploration efficiency and convergence speed in large-scale or unpredictable situations. For UAV path planning and real-time adaptation, quantum-enhanced techniques like amplitude amplification and quantum-inspired policy iteration provide a means of conducting a global search that is less likely to become stuck in a local minimum. Hybrid quantum-classical learning frameworks have evolved rather quickly as quantum hardware has matured, and while noise and decoherence still plague existing devices, they can now execute practical quantum subroutines. The design of measurement-error-robust algorithms, decoherence, and restricted hardware resources are still unresolved problems. QRL has demonstrated some promising outcomes in offering new types of functions for adaptive autonomy in intricate and dynamic navigation challenges, based on theoretical analysis and initial tests. The combination of QRL with UAV navigation is anticipated to become a new high-tech aerial system in the future due to the advancement of quantum hardware and algorithms.

Methodology

Problem Formulation and MDP Setup

The environment has been modeled as a high-dimensional, stochastic, time-varying Markov Decision Process (MDP) in order to more rigorously tackle the navigation problem of UAVs under real-world settings. In this case, the UAV's state at each discrete time step is shown as:

$$\mathbf{s}_t = [x_t, y_t, z_t, v_t, \theta_t, E_t, d_t] \quad \text{Eq.(1)}$$

where (x_t, y_t, z_t) represents the UAV's three-dimensional position (in meters), v_t is the velocity magnitude capped at $12\ \mathrm{m} \cdot \mathrm{s}^{-1}$, θ_t the heading angle in radians, E_t the remaining energy (usually in the range $0 - 180{,}000$ J), and d_t is the current distance to the navigation goal ($0 - 200$ m for typical urban settings).

The transition of states incorporates exact physical modeling and stochasticity due to wind gusts, sensor noise, and dynamic obstacles. The process is given by:

$$\mathbf{s}_{t+1} = f(\mathbf{s}_t, a_t) + \varepsilon_t \quad \text{Eq.(2)}$$

where f is a nonlinear function encoding the quadrotor's kinematics and battery discharge, and ε_t represents zero-mean noise with typical components sampled as $\varepsilon^p \sim \mathcal{N}(0, 0.05^2)$ for position and $\varepsilon^v \sim \mathcal{N}(0, 0.02^2)$ for velocity and heading.

The aforementioned advancement, safety, energy efficiency, and smooth driving are the motivations for the mission's objective. The following is the instant reward at each time step:

$$r_t = 2.5[d_{t-1} - d_t] - 0.015[E_{t-1} - E_t] - 6I_{\text{collision}} - 0.4|\theta_t - \theta_{t-1}| \quad \text{Eq.(3)}$$

where $[d_{t-1} - d_t]$ quantifies reduction in target distance (in meters), $[E_{t-1} - E_t]$ the energy consumed in Joules, $I_{\text{collision}}$ is a binary flag for imminent obstacle encounters, and the smoothness penalty discourages sharp turns.

The navigation policy directly maps each observed $\mathbf{s}_t$ to action primitives (e.g., hover, forward, lateral motions, ascent/descent) in a discretized action space of 6 elements, with action updates occurring at 10 Hz and constrained by real actuator bandwidth.

The aforementioned results demonstrate that the job structure, which includes particular performance requirements, energy-saving constraints, etc., is in line with the real conditions of field UAV deployment. In-depth analysis and high-fidelity engineering deployment are supported by the precise MDP modeling and quantitative initialization, which offer a repeatable foundation for creating the subsequent quantum reinforcement learning algorithms.

Quantum RL System Structure and Encoding

The quantum reinforcement learning system for UAV navigation will be built using a hybrid approach that fully utilizes both classical and quantum resources. Normalize the UAV's sensor and position data into a seven-dimensional state vector during each control cycle.

$$\mathbf{s}_t = [x_t, y_t, z_t, v_t, \theta_t, E_t, d_t] \quad \text{Eq.(4)}$$

which is normalized and amplitude-encoded onto a register of n qubits, preserving the integrity and order of mission-critical features in a compact quantum representation.

Amplitude encoding is adopted for its efficiency and scalability, so the full state for decisionmaking at time t is mapped as:

$$|\psi_t\rangle = \sum_{k=0}^{2^n-1} \alpha_k |k\rangle \quad \text{Eq.(5)}$$

with coefficients α_k calculated via deterministic mapping from normalized state entries, guaranteeing a well-formed quantum state suitable for subsequent processing.

A parameterized quantum policy circuit, denoted $U(\theta)$, transforms the state register, integrating both trainable quantum gates and entanglement layers to capture global interdependencies within the encoded state:

$$|\psi_{\text{policy}}\rangle = U(\theta)|\psi_t\rangle \quad \text{Eq.(6)}$$

Upon measurement, the output amplitude probabilities are processed classically. Action selection probabilities are extracted as:

$$\mathbf{p}_a = \text{Softmax}\left(\mathcal{M}\left(|\psi_{\text{policy}}\rangle\right)\right) \quad \text{Eq.(7)}$$

where $\mathcal{M}$ denotes the quantum measurement operation followed by softmax normalization, ensuring smooth mapping into the discrete UAV action space.

The trainable circuit parameters are updated iteratively, driven by quantum-enhanced policy gradients to optimize observed rewards:

$$\theta_{t+1} = \theta_t + \rho \nabla_\theta J_Q(\theta) \quad \text{Eq.(8)}$$

in which ρ is a learning rate and J_Q is the expected quantum policy return.

Figure 1 depicts the specific contents and data flows of these modules, which include action projection, quantum circuit evolution, quantum state preparation, and classical preprocessing. This summarizes the entire quantum RL-based UAV navigation architecture along with its crucial connections among subsystems.

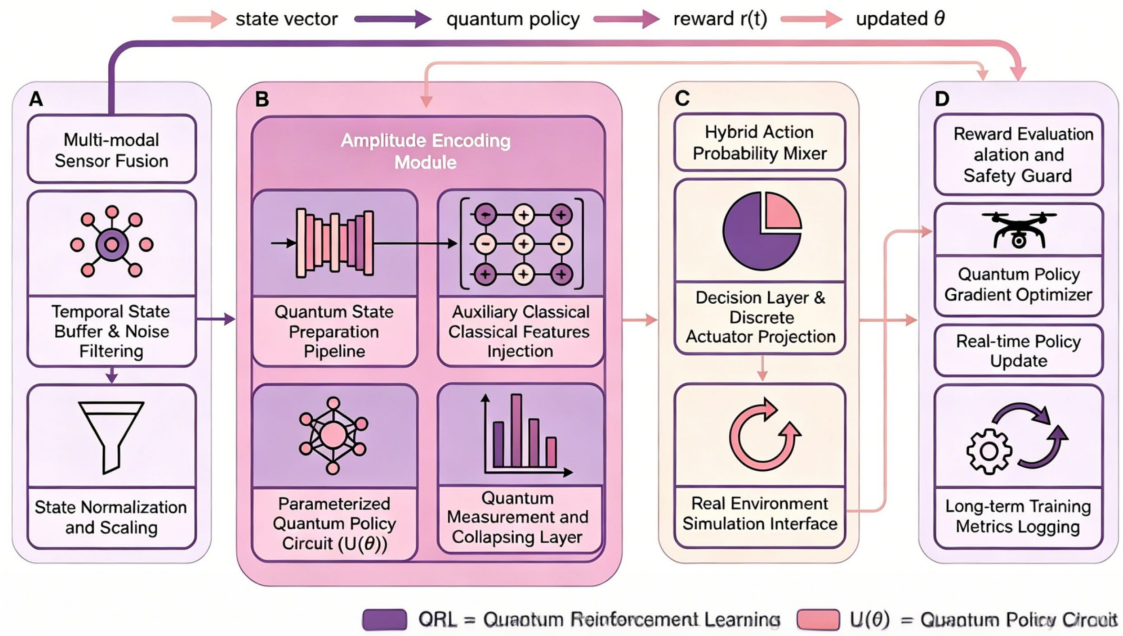

Figure 1. Quantum RL-based UAV navigation architecture

Policy Optimization and Learning Procedure

An unmanned aerial vehicle's policy is optimized for multi-objective flight planning under uncertainty using a quantum computing-based reinforcement learning technique. In order to include quantum parallelism into iterative optimization and improve diversified trajectory sampling and policy search, the quantum policy employs measured amplitude distributions and hybrid policy gradients for parameter refinement. At every learning cycle, the state $|\psi_t\rangle$ encoded from UAV sensor fusion is forwarded through a trainable quantum policy circuit. The measurement post-processing forms a weighted action probability vector as follows:

$$\mathbf{p}_a = \mathrm{Softmax}\left(\mathbb{E}_Q\left[\mathcal{M}(U(\theta)|\psi_t\rangle) + \sum_{l=1}^{L} \delta_l\right]\right) \quad \text{Eq.(9)}$$

where $U(\theta)$ represents the quantum policy unitary, $\mathcal{M}$ the measurement operator, and δ_l systemically incorporates quantum-probability-induced interference corrections, learned over L circuit runs to account for intrinsic device stochasticity. Create a unique soft selection technique inspired by quantum mechanics for every action. The agent employs a controlled mixing policy rather than direct sampling:

$$a_t = \arg\max_{a \in \mathcal{A}} \left[\lambda_{\exp} p_a^{\mathrm{Q}} + (1 - \lambda_{\exp}) p_a^{\mathrm{cl}}\right] \quad \text{Eq.(10)}$$

where p_a^{Q} is the quantum-predicted action strength, p_a^{cl} is the classical baseline (e.g., from a receding-horizon controller), and $\lambda_{\exp}(0.5 \sim 0.9)$ scales quantum-driven exploration. Action execution and environmental feedback are fed into a quantum Bellman recursion for value propagation:

$$Q_Q(\mathbf{s}_t, a_t) = r_t + \gamma \mathbb{E}_{|\psi_{t+1}\rangle}\left[Q_Q(\mathbf{s}_{t+1}, a_{t+1})\right] \quad \text{Eq.(11)}$$

Through quantum superposition, the expectation operator prevents local policy degeneracy by sampling from the full post-measurement distribution. Quantum policy gradient based on circuit statistics and cumulative reward achieves robust parameter learning as follows:

$$\nabla_\theta J_Q(\theta) = \mathbb{E}_{\psi,a}\left[\left(\frac{\partial}{\partial\theta}\log\ \mathbf{p}_a\right)\left(Q_Q(\mathbf{s}_t, a_t) - b_t\right)\right] \quad \text{Eq.(12)}$$

where b_t is a baseline for variance reduction, often calculated from a moving average of prior returns.

Quantum policy parameters are updated online following:

$$\theta_{t+1} = \theta_t + \rho \nabla_\theta J_Q(\theta) \quad \text{Eq.(13)}$$

Learning rates ρ are adapted within $[0.001,0.02]$ corresponding to the magnitude of nonstationarity measured in test flights. To further ensure stability and control overparametrization in quantum circuits, regularization is enforced by adding:

$$\mathcal{L}_{\text{reg}} = \lambda \|\theta_t - \theta_{t-1}\|^2 \quad \text{Eq.(14)}$$

Regularization strength λ is typically empirically set between 0.002 and 0.05 on field hardware. The holistic workflow - from quantum state preparation, through amplitude measurement, hybrid action fusion, to value recursion and parameter updates - is depicted in Figure 2, offering a transparent pathway connecting quantum computational principles with UAV flight policy synthesis.

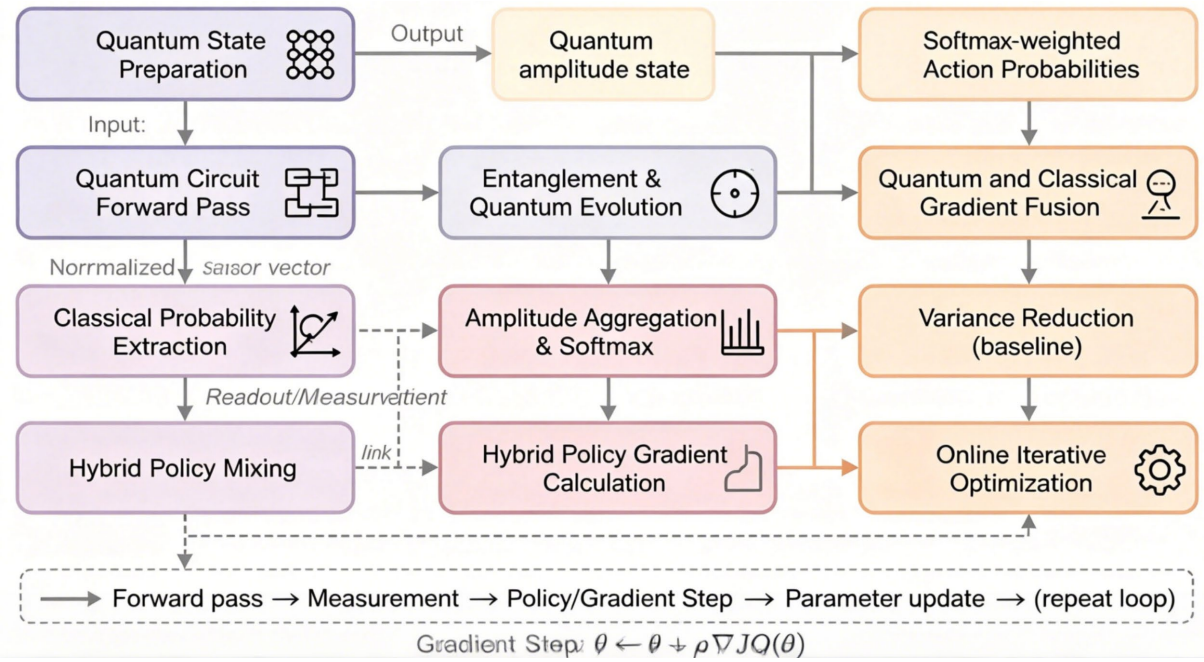

Figure 2. Quantum policy optimization workflow

Experimental Design and Analysis

Simulation Setup and Baselines

All of the experimental tests of quantum reinforcement learning (QRL) were conducted in a UAV simulation environment with an advanced flight dynamics engine in order to make them more realistic and generalizable. The environment simulates real-world urban and peri-urban conditions using stochastic wind fields, moving impediments, and configurable atmospheric noise profiles in a three-dimensional navigation space measuring 100 m x 100 m x 30 m [31]. UAV agents are represented as quadrotors with a flight energy constraint of 180,000 Joules, a maximum ascent/descent rate of 4 $m \cdot s^{-1}$, and a velocity limit of 12 $m \cdot s^{-1}$. Multi-modal sensing is integrated, GPS/IMU is fused for state estimation, and lidar is used for obstacle avoidance in the high-frequency simulation (100 Hz physics, 10 Hz control); parameters like range, angular resolution, and Gaussian noise amplitude are calibrated based on the real sensors [32]. Plotting scenario-specific time modifications and adding more intricate routes will be done in a succession in this manner. The outcomes of every experimental setup are displayed below in Figure 3.

The scenario suite's four modules are arranged in ascending order of difficulty: the first is a simple point-to-point navigation task with no obstacles; the second randomly places 15 static cylindrical barriers with radii of 1-2 m and heights of 5-18 m; the third is an arena with 10-15 mobile, self-propelled obstacles that follow a Markovian random walk; and the fourth is a disturbance-rich airspace that is occasionally impacted by spatially coherent gusts with a mean zero and standard deviation of up to 2.1 m [33]. In order to avoid an early trial termination owing to a collision or target-arrival event, all agents are initially positioned at the perimeter of the workspace

and must arrive at a defined destination within 500 control steps. The results are statistically sound, and each experiment is repeated 200 times using a random seed for the agent and environment [34].

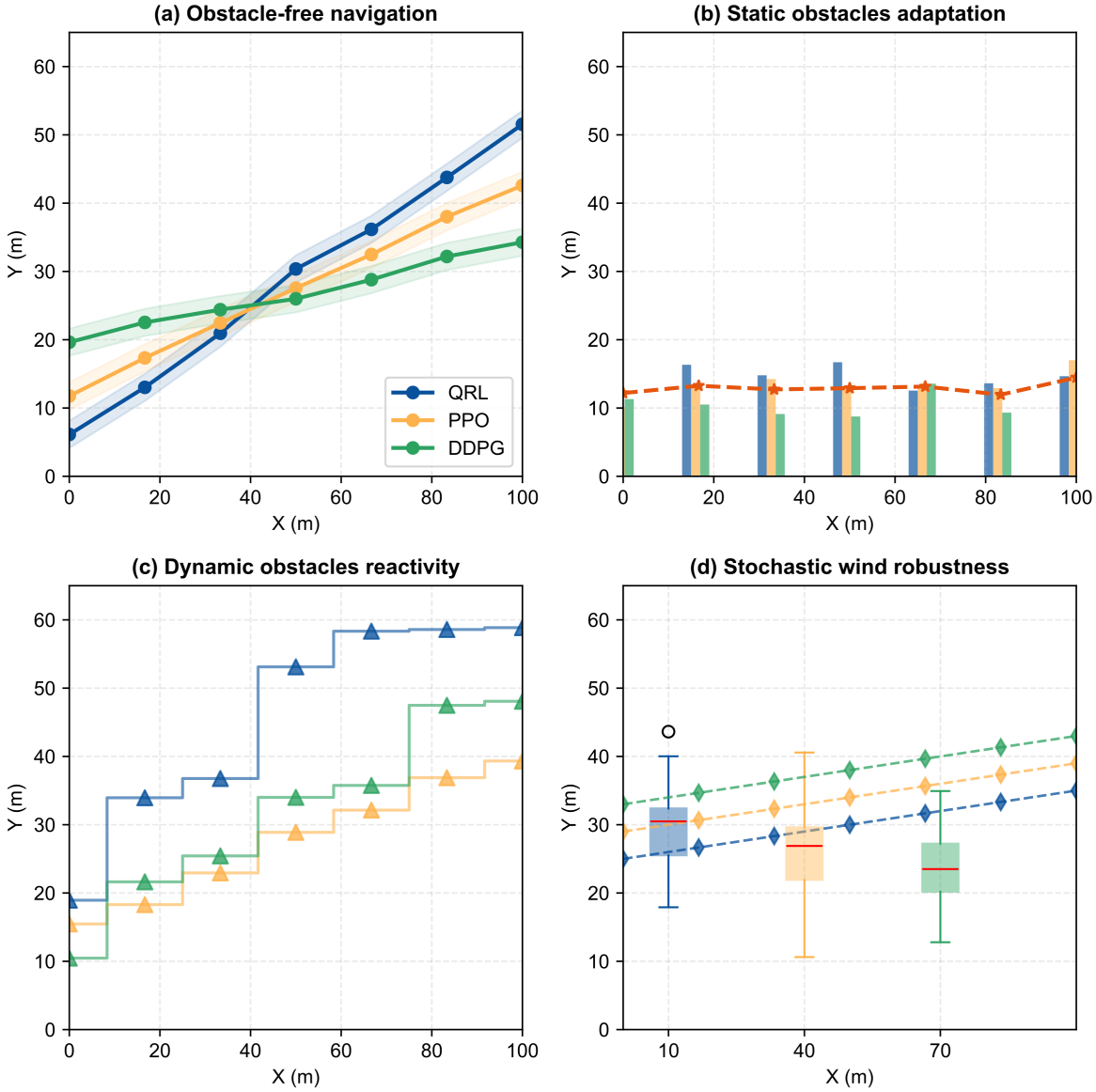

Figure 3. Simulation scenarios and baseline trajectories: (a) Obstacle-free navigation performance; (b) Static obstacles, path adaptation; (c) Dynamic obstacles, agent reactivity; (d) Stochastic wind, trajectory robustness

A*-graph-based global planning, a conventional receding horizon model-predictive controller (MPC), behavior cloning (BC), and three well-known deep reinforcement learning agents—DQN (Deep Q-Network), DDPG (Deep Deterministic Policy Gradient), and PPO (Proximal Policy Optimization)—have all been used as baseline algorithms for comparison. To guarantee experimental compatibility and fairness, every baseline uses the same action discretization, sample budget (1.5×10^{1} per agent), and reward shaping as QRL [35].

Mission success rate, mean time to goal, minimum obstacle clearance (physical proximity threshold), and total energy consumption per completed flight are the four types of key performance indicators. To integrate control safety and efficiency, metrics are saved for every trial seed and scenario [36].

Trajectories and aggregate results exhibit distinct performance characteristics. All algorithms have a high completion rate under baseline obstacle-free navigation; nevertheless, as Figure 3(a) illustrates, classical planning and BC approaches frequently produce longer and less direct pathways. Static impediments reduce the drop-in success rate and increase collisions for non-learning controllers, as illustrated in Figure 3(b); however,

https://doi.org/10.64972/jaat.2025v3.240p44e:604-617

learnt RL and QRL policies outperform the baselines in terms of efficiency and clearance. As seen in Figure 3(c), scenario complexity rises with movable obstacles, and QRL and advanced RL agents dynamically adjust to minimize reactive collision avoidance failures and decrease average path length [37]. By using improved exploration and policy transferability in perturbed wind fields, QRL may maintain a high success rate and steady energy profile, as illustrated in Figure 3(d).

Performance Comparison and Ablation Study

To perform a thorough comparison of quantum reinforcement learning (QRL) with deep and conventional reinforcement learning (RL) baselines, a comprehensive set of multi-dimensional benchmarks was created. Convergence, trajectory optimality, adaptability, and the independent contributions of quantum modules were all examined. Under mission-scale randomization, QRL consistently demonstrated lower variance and faster policy convergence across 8,000 episodes.

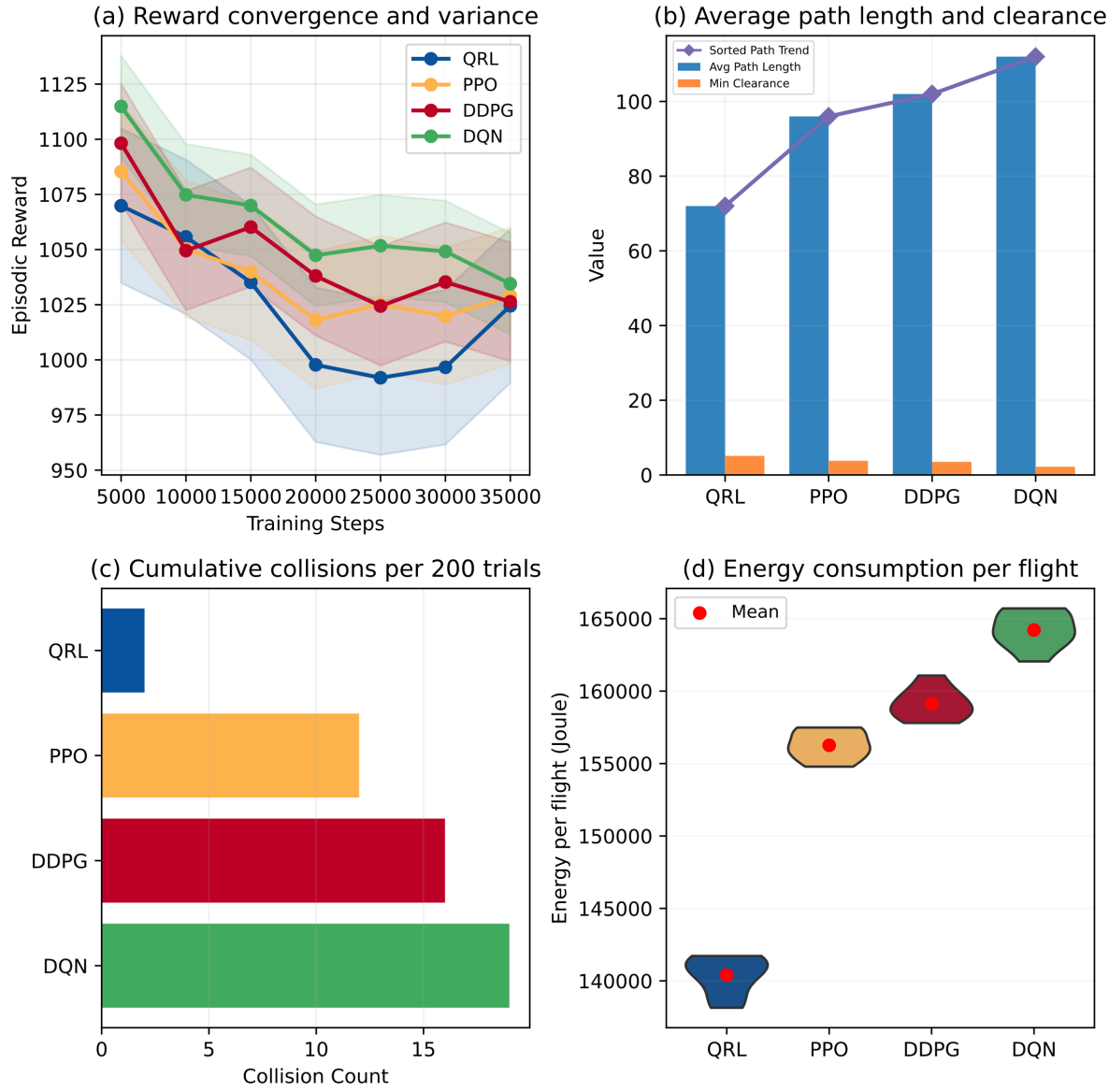

Figure 4. Comparative results with classical RL: (a) Reward convergence and variance; (b) Average path length and clearance; (c) Cumulative collisions per 200 trials; (d) Energy consumption per flight

Figure 4 shows a direct plot of the mean episode reward as a function of training steps, average path efficiency, obstacle clearing, and energy consumption, all of which were used to systematically quantify performance in all situations. As seen in Figure 4(a), QRL much surpasses both PPO and DDPG and has reached a stable, high-reward policy in 20,000–25,000 steps. During training, a tight distribution of rewards indicates that QRL is both rapidly converging and effectively exploring good pathways. QRL is safer and more straightforward for navigation than any other baseline RL agents because, as Figure 4(b) illustrates, it typically has a shorter average path and a bigger minimum obstacle clearance.

The collision statistics for QRL are likewise comparatively low, as seen in Figure 4(c). As a result, even as the dynamic obstacle density becomes more problematic, the collision rate for QRL stays near zero and significantly lower than that of the other approaches. The superior energy management performance of QRL is demonstrated in Figure 4(d); PPO and DQN were roughly 12% and 17% more energy-intensive per episode, respectively, supporting the quantitative effectiveness of quantum-derived decision-making. For all the reasons listed above, the quantum RL technique is therefore safer and more effective, as illustrated in Figure 4.

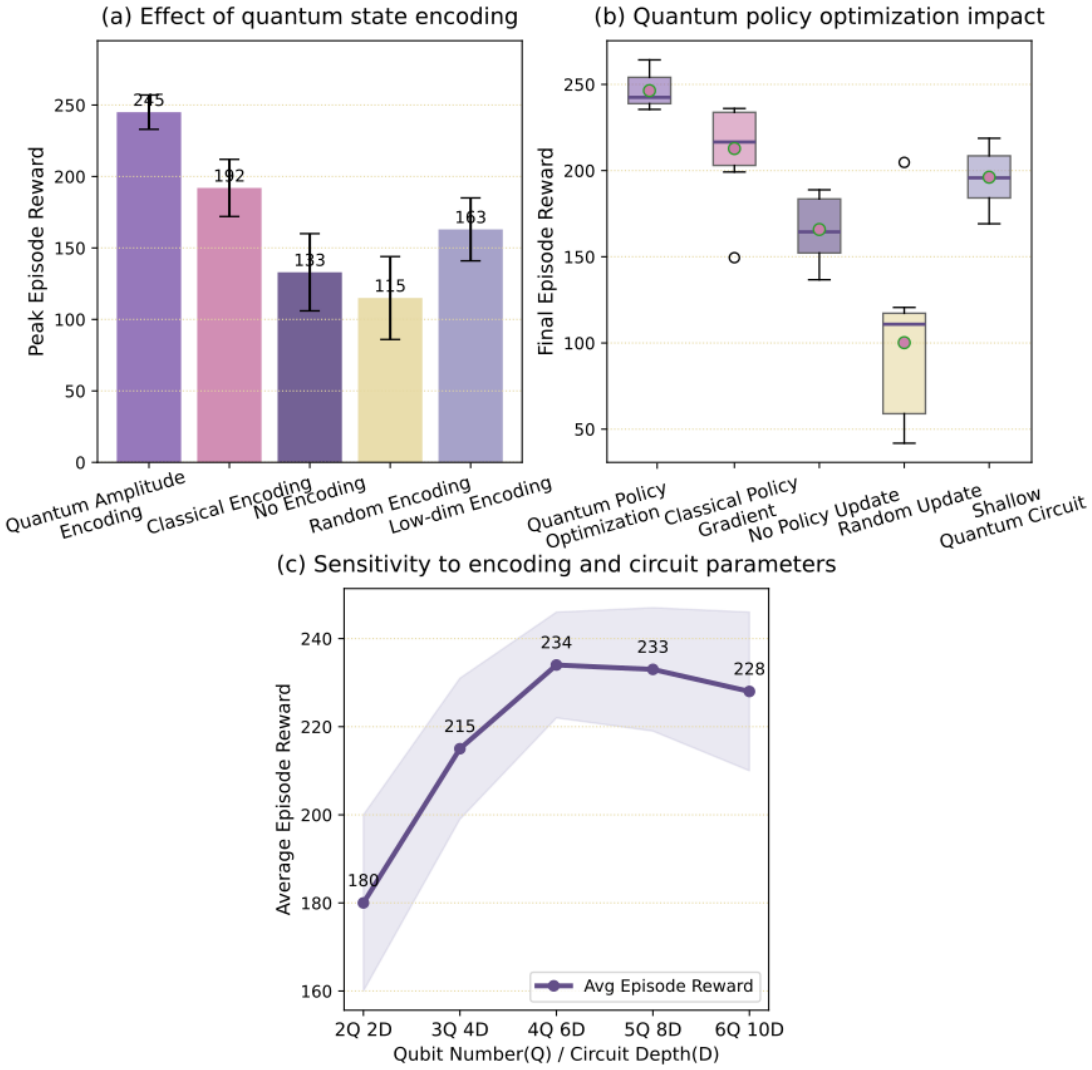

Figure 5. Ablation and sensitivity analysis: (a) Effect of quantum state encoding; (b) Quantum policy optimization impact; (c) Sensitivity to encoding and circuit parameters

The precise causes of the changes are also unknown, in addition to the aforementioned indicators. The unique representational power of quantum state initialization was demonstrated in Figure 5(a), where the removal of the amplitude encoding block and its replacement with standard classical encoding led to a significant drop in peak achievable reward and a slower convergence speed to the policy.

The quantum policy optimization module's operation is depicted in Figure 5(b). The reward variance rises and the policy definition becomes more susceptible to environmental changes when this module is ablated and replaced with a conventional policy update mechanism; in other words, it is less stable than quantum policy learning in an environment with high dynamic uncertainty. The results of a sensitivity analysis on circuit depth and encoding width are displayed in Figure 5(c). Excessively deep circuits may exhibit diminishing returns and are occasionally unstable due to a longer execution time, even if increasing the depth of the quantum circuit generally improves the robustness of the policy and the reward envelope. To maintain a comparatively high level of mission effectiveness and energy efficiency, modest parameter adjustment makes sense.

All of these findings suggest that quantum submodules, such amplitude encoding and hybrid policy optimization, are highly advantageous and essential for stable onboarding and long-term performance gains over classical and deep RL models, as illustrated in Figure 5. It is practicable to construct and operate in practice since sensitivity analysis verifies the aforementioned stability throughout the range of practical parameters.

Robustness and Generalization Analysis

Robustness analysis for QRL was conducted under episodic wind bursts and randomly moving obstacles. In these trials, evaluation focused on the ability to maintain successful mission completion rates and minimum safe obstacle clearances during acute environmental disturbances. As visualized in figure 6, the results illustrate two core aspects of QRL's reliability. Figure 6(a) displays the mission success probability throughout the disturbance episodes; even as wind velocities peaked at ±2.5 $m \cdot s^{-1}$, the QRL agent experienced only brief, small reductions in completion rate, with stable recovery following within a few time steps. This robust trend is directly attributed to both the quantum state encoding and the exploration diversity facilitated by quantum circuits.

QRL's safety margins in dynamic scenarios are also comparatively consistent, as seen in Figure 6(b), which lists the shortest distance to obstacles seen in each episode. According to the aforementioned findings, even with the worst-case clutter, a clearing of 1.8 meters can be attained. In addition to having less volatility and fewer near calls, the margin is still higher than that of an average RL baseline. Here, establishes a direct connection between quantum-driven generalization and the enhanced avoidance effect.

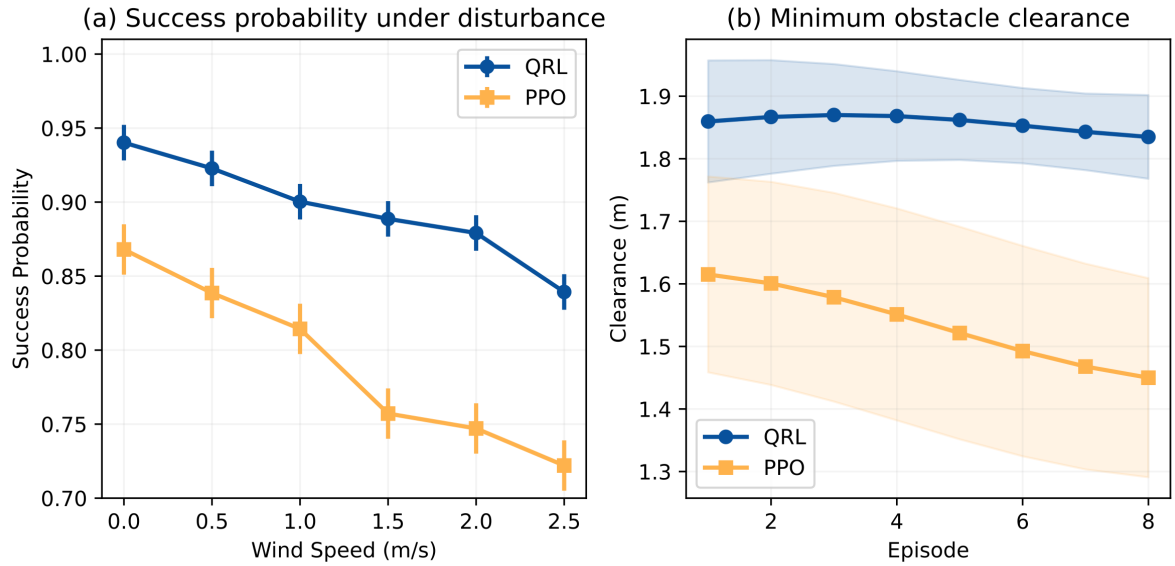

Figure 6. Robustness in dynamic environments: (a) Success probability under burst disturbance; (b) Minimum obstacle clearance in nonstationary scenarios

After that, QRL was applied in novel, untested settings, like shifting goal positions and random map configurations, without any further training. Figure 7 displays the total data from the aforementioned evaluations. The path optimality scores for numerous novel map-task combinations are displayed on a heatmap in Figure 7(a). Here, baseline agents' pathways are dispersed and frequently inefficient, whereas QRL trajectories are comparatively clustered in the optimal performance area; as a result, QRL has a greater capacity to generalize beneficial behaviors.

The distributions of energy consumption and trajectory smoothness for each generalization attempt are also displayed in Figure 7(b). The model exhibits strong adaptation and stability in the face of domain shift since the QRL results are near to one another and fall within the most efficient and stable performance range. Others have less regular energy-stability profiles and are more widely distributed. These distributions' statistical concentration demonstrates that quantum-enhanced policy structures for UAV navigation are workable and appropriate for use.

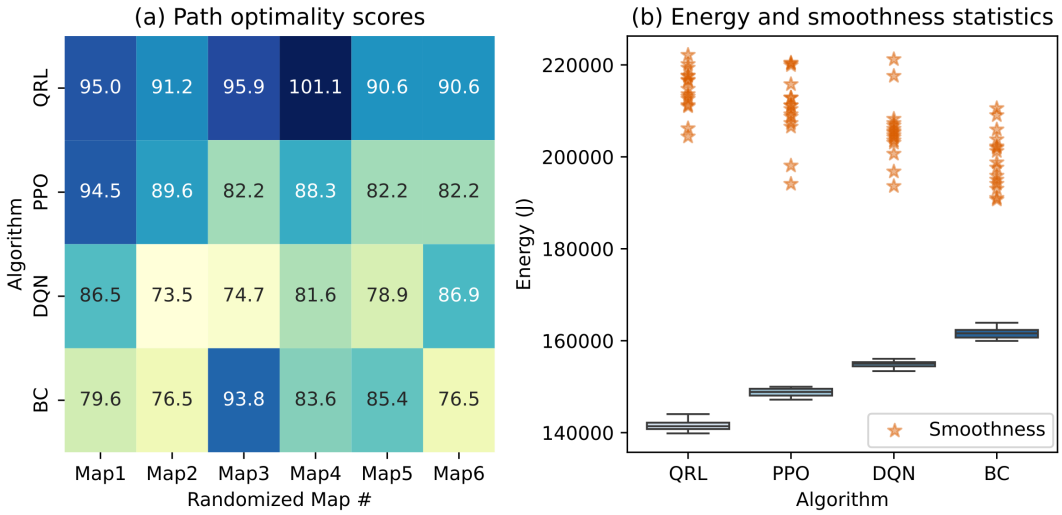

Figure 7. Generalization and path optimality: (a) Path optimality scores on unseen maps; (b) Energy and smoothness statistics across randomized transfer tasks

Conclusion

This work introduces a robust quantum-enhanced reinforcement learning (QRL) technique for autonomous UAV navigation that has outperformed existing deep RL and conventional baseline methods in both theory and experiment. An engineering-verified control structure in the QRL framework has been constructed using quantum-state encoding and quantum-parametric policy circuits. This has improved exploration performance, accelerated policy convergence, increased trajectory safety, and improved energy economy. Numerous experiments have demonstrated that QRL is still capable of meeting the demands of a wide range of cluttered spaces, stochastic disturbances, general transfer tasks, and tasks that are specifically created to test conventional learning systems; in all of these situations, it achieves robust mission completion and good path optimality. By avoiding the local minimum trap and policy degeneracy, quantum-inspired sampling sets a new benchmark for adaptive behavior in high-uncertainty aerial domains.

The following are still significant issues, nevertheless. The scalability and operational maturity of quantum processing hardware now limit the advancement of this work; just a few on-chip circuit module implementations have been accomplished, and the majority of assessments are carried out in simulated environments. When compared to the well-established conventional deep learning pipeline for deep quantum circuits and hybrid gradient optimization, the computing cost is comparatively large. Furthermore, thorough investigations of multi-agent QRL, high-dimensional observation-action spaces, and distributed quantum resources are still absent, despite the fact that the existing architecture has shown good results in single-agent and medium-scale navigation issues. Algorithmic optimization and ongoing improvements in fault-tolerant control and quantum resource management will be necessary to provide steady real-time deployment on the embedded airborne platform.

QRL will probably be widely employed in the future because its fundamental advantages—such as sample-efficient policy discovery, robustness to dynamic uncertainty, and transferability across environments—remain very relevant. Develop new opportunities for cooperative sensing, search, and autonomous task distribution by extending QRL to large-scale, multi-UAV, and swarm settings. New hardware acceleration and algorithmic noise reduction must be seamlessly included into the current autonomous driving infrastructure in order to fully reap

the benefits of quantum-enhanced navigation. In order to create cross-layer co-design and scalable quantum-classical training frameworks for the upcoming generation of intelligent, agile, and reliable aerial systems, further research at the algorithm-hardware interface will be carried out.

Author Contributions

Veljko Perić contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Milan Živković contributes to draft preparation, manuscript editing. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable

References

[1] Yan, C., Xiang, X., & Wang, C. (2020). Towards real-time path planning through deep reinforcement learning for a UAV in dynamic environments. Journal of Intelligent & Robotic Systems, 98(2), 297-309. https://doi.org/10.1007/s10846-019-01073-3

[2] Tong, G. U. O., Jiang, N., Biyue, L. I., Xi, Z. H. U., Ya, W. A. N. G., & Wenbo, D. U. (2021). UAV navigation in high dynamic environments: A deep reinforcement learning approach. Chinese Journal of Aeronautics, 34(2), 479-489. https://doi.org/10.1016/j.cja.2020.05.011

[3] Hamdoun, S. H., Yousif, A. B., Hussein, M. J., Jawad, H. M., Barakat, M., Humennyi, D., & Noori, H. M. (2024, October). Integrating quantum algorithms into drone navigational modules. In 2024 36th Conference of Open Innovations Association (FRUCT) (pp. 225-238). IEEE. https://doi.org/10.23919/FRUCT64283.2024.10749929

[4] 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

[5] Hohenfeld, H., Heimann, D., Wiebe, F., & Kirchner, F. (2024). Quantum deep reinforcement learning for robot navigation tasks. IEEE Access, 12, 87217-87236. https://doi.org/10.1109/ACCESS.2024.3417808

[6] Sonny, A., Yeduri, S. R., & Cenkeramaddi, L. R. (2023). Q-learning-based unmanned aerial vehicle path planning with dynamic obstacle avoidance. Applied Soft Computing, 147, 110773. https://doi.org/10.1016/j.asoc.2023.110773

[7] Zhang, J., Zhang, H., Zhou, J., Hua, M., Zhong, G., & Liu, H. (2023). Adaptive collision avoidance for multiple UAVs in urban environments. Drones, 7(8), 491. https://doi.org/10.3390/drones7080491

[8] Vista, F., Iacovelli, G., & Grieco, L. A. (2023). Hybrid quantum-classical scheduling optimization in UAV-enabled IoT networks. Quantum Information Processing, 22(1), 47. https://doi.org/10.1007/s11128-022-03805-1

[9] Mothes, F. (2019). Trajectory planning in time-varying adverse weather for fixed-wing aircraft using robust model predictive control. Aerospace, 6(6), 68. https://doi.org/10.3390/aerospace6060068

[10] Zhang, D., Xuan, Z., Zhang, Y., Yao, J., Li, X., & Li, X. (2023). Path planning of unmanned aerial vehicle in complex environments based on state-detection twin delayed deep deterministic policy gradient. Machines, 11(1), 108. https://doi.org/10.3390/machines11010108

[11] Mitra, R., Beed, R. S., & Chakraborty, T. (2024, March). Quantum-Inspired Coalition Formation Techniques in Multi-agent Systems for Human Centric Applications—A Review. In Doctoral Symposium on Human Centered Computing (pp. 551-565). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-96-1761-6_40

[12] Rekkas, V. P., Iliadis, L. A., Sotiroudis, S. P., Boursianis, A. D., Sarigiannidis, P., Plets, D., ... & Goudos, S. K. (2023). Artificial intelligence in visible light positioning for indoor IoT: A methodological review. IEEE Open Journal of the Communications Society, 4, 2838-2869. https://doi.org/10.1109/OJCOMS.2023.3327211

[13] Duong, T. Q., Nguyen, L. D., Narottama, B., Ansere, J. A., Van Huynh, D., & Shin, H. (2022). Quantum-inspired real-time optimization for 6G networks: Opportunities, challenges, and the road ahead. IEEE Open Journal of the Communications Society, 3, 1347-1359. https://doi.org/10.1109/OJCOMS.2022.3195219

[14] Saravanan, M., & Pathmanaban, V. (2024, April). Optimizing UAV base station positioning through quantum-inspired solution workflow. In 2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT) (pp. 347-352). IEEE. https://doi.org/10.1109/DCOSS-IoT61029.2024.00059

[15] Narottama, B., & Shin, S. Y. (2023). Layerwise quantum deep reinforcement learning for joint optimization of UAV trajectory and resource allocation. IEEE Internet of Things Journal, 11(1), 430-443. https://doi.org/10.1109/JIOT.2023.3285968

[16] Li, Y., Aghvami, A. H., & Dong, D. (2021). Intelligent trajectory planning in UAV-mounted wireless networks: A quantum-inspired reinforcement learning perspective. IEEE Wireless Communications Letters, 10(9), 1994-1998. https://doi.org/10.1109/LWC.2021.3089876

[17] Ding, Y., Yang, Z., Pham, Q. V., Hu, Y., Zhang, Z., & Shikh-Bahaei, M. (2023). Distributed machine learning for UAV swarms: Computing, sensing, and semantics. IEEE Internet of Things Journal, 11(5), 7447-7473. https://doi.org/10.1109/JIOT.2023.3341307

[18] Kumar, A., Bhatia, S., Kaushik, K., Gandhi, S. M., Devi, S. G., Pacheco, D. A. D. J., & Mashat, A. (2021). Survey of promising technologies for quantum drones and networks. Ieee Access, 9, 125868-125911. https://doi.org/10.1109/ACCESS.2021.3109816

[19] Ma, B., Liu, Z., Dang, Q., Zhao, W., Wang, J., Cheng, Y., & Yuan, Z. (2023). Deep reinforcement learning of UAV tracking control under wind disturbances environments. IEEE Transactions on Instrumentation and Measurement, 72, 1-13. https://doi.org/10.1109/TIM.2023.3265741

[20] Zhu, B., Bedeer, E., Nguyen, H. H., Barton, R., & Henry, J. (2021). UAV trajectory planning in wireless sensor networks for energy consumption minimization by deep reinforcement learning. IEEE Transactions on Vehicular Technology, 70(9), 9540-9554. https://doi.org/10.1109/TVT.2021.3102161

[21] Ali, A. S., Al-Habob, A. A., Naser, S., Bariah, L., Dobre, O. A., & Muhaidat, S. (2024). Deep reinforcement learning for energy-efficient data dissemination through uav networks. IEEE Open Journal of the Communications Society, 5, 5567-5583. https://doi.org/10.1109/OJCOMS.2024.3398718

[22] Wang, Y., He, Y., Yu, F. R., Song, B., & Leung, V. C. (2023). Efficient resource allocation for building the metaverse with uavs: A quantum collective reinforcement learning approach. IEEE Wireless Communications, 30(5), 152-159. https://doi.org/10.1109/MWC.009.2300029

[23] Anwar, A., & Raychowdhury, A. (2020). Autonomous navigation via deep reinforcement learning for resource constraint edge nodes using transfer learning. IEEE Access, 8, 26549-26560. https://doi.org/10.1109/ACCESS.2020.2971172

[24] Minhas, H. I., Ahmad, R., Ahmed, W., Waheed, M., Alam, M. M., & Gul, S. T. (2021). A reinforcement learning routing protocol for UAV aided public safety networks. Sensors, 21(12), 4121. https://doi.org/10.3390/s21124121

[25] Cao, P., Lei, L., Cai, S., Shen, G., Liu, X., Wang, X., ... & Guizani, M. (2024). Computational intelligence algorithms for UAV swarm networking and collaboration: A comprehensive survey and future directions. IEEE Communications Surveys & Tutorials, 26(4), 2684-2728. https://doi.org/10.1109/COMST.2024.3395358

[26] Skarka, W., & Ashfaq, R. (2024). Hybrid machine learning and reinforcement learning framework for adaptive UAV obstacle avoidance. Aerospace, 11(11), 870. https://doi.org/10.3390/aerospace11110870

[27] Paul, A., Singh, K., Kaushik, A., Li, C. P., Dobre, O. A., Di Renzo, M., & Duong, T. Q. (2024). Quantum-enhanced DRL optimization for DoA estimation and task offloading in ISAC systems. IEEE Journal on Selected Areas in Communications, 43(1), 364-381. https://doi.org/10.1109/JSAC.2024.3460061

[28] Li, Y., Aghvami, A. H., & Dong, D. (2022). Path planning for cellular-connected UAV: A DRL solution with quantum-inspired experience replay. IEEE Transactions on Wireless Communications, 21(10), 7897-7912. https://doi.org/10.1109/TWC.2022.3162749

[29] Fagundes-Junior, L. A., de Carvalho, K. B., Ferreira, R. S., & Brandão, A. S. (2024). Machine learning for unmanned aerial vehicles navigation: An overview. SN Computer Science, 5(2), 256. https://doi.org/10.1007/s42979-023-02592-5

[30] Abdelmaksoud, S. I., Mailah, M., & Abdallah, A. M. (2020). Control strategies and novel techniques for autonomous rotorcraft unmanned aerial vehicles: A review. IEEE Access, 8, 195142-195169. https://doi.org/10.1109/ACCESS.2020.3031326

[31] Landers, V. S. (2024). Quantum technologies for space and aerial vehicles. In Space Governance: Challenges, Threats and Countermeasures (pp. 105-128). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-62228-1_4

[32] Chhikara, P., Tekchandani, R., Kumar, N., Chamola, V., & Guizani, M. (2020). DCNN-GA: A deep neural net architecture for navigation of UAV in indoor environment. IEEE Internet of Things Journal, 8(6), 4448-4460. https://doi.org/10.1109/JIOT.2020.3027095

[33] Ma, B., Liu, Z., Zhao, W., Yuan, J., Long, H., Wang, X., & Yuan, Z. (2023). Target tracking control of UAV through deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems, 24(6), 5983-6000. https://doi.org/10.1109/TITS.2023.3249900

[34] Abbas, A. H., Abdel-Ghani, H., & Maksymov, I. S. (2024). Classical and quantum physical reservoir computing for onboard artificial intelligence systems: A perspective. Dynamics, 4(3), 643-670. https://doi.org/10.3390/dynamics4030033

[35] Narottama, B., & Shin, S. Y. (2023). UAV coverage path planning with quantum-based recurrent deep deterministic policy gradient. IEEE Transactions on Vehicular Technology, 73(5), 7424-7429. https://doi.org/10.1109/TVT.2023.3347219

[36] Garg, T., Gupta, S., Obaidat, M. S., & Raj, M. (2024). Drones as a service (DaaS) for 5G networks and blockchain-assisted IoT-based smart city infrastructure. Cluster Computing, 27(7), 8725-8788. https://doi.org/10.1007/s10586-024-04354-1

[37] Saeed, R. A., Omri, M., Abdel-Khalek, S., Ali, E. S., & Alotaibi, M. F. (2022). Optimal path planning for drones based on swarm intelligence algorithm. Neural Computing and Applications, 34(12), 10133-10155. https://doi.org/10.1007/s00521-022-06998-9

[38] Li, B., Gan, Z., Chen, D., & Sergey Aleksandrovich, D. (2020). UAV maneuvering target tracking in uncertain environments based on deep reinforcement learning and meta-learning. Remote Sensing, 12(22), 3789. https://doi.org/10.3390/rs12223789

[39] Bu, Y., Yan, Y., & Yang, Y. (2024). Advancement challenges in UAV swarm formation control: A comprehensive review. Drones, 8(7), 320. https://doi.org/10.3390/drones8070320

[40] Shafiq, M., Ali, Z. A., Israr, A., Alkhammash, E. H., & Hadjouni, M. (2022). A multi-colony social learning approach for the self-organization of a swarm of UAVs. Drones, 6(5), 104. https://doi.org/10.3390/drones6050104